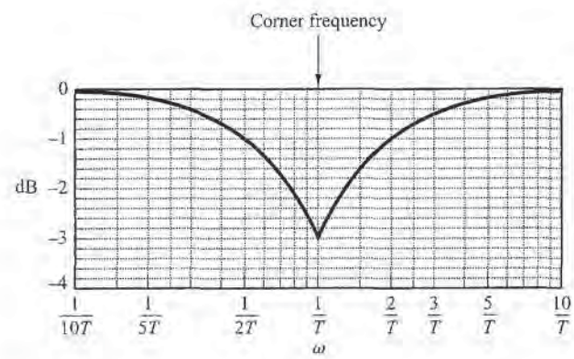


Figure 8–7
Log-magnitude error in the asymptotic expression of the frequency-response curve of $1/(1 + j\omega T)$.

Thus, the error at one octave below or above the corner frequency is approximately equal to −1 dB. Similarly, the error at one decade below or above the corner frequency is approximately −0.04 dB. The error in decibels involved in using the asymptotic expression for the frequency-response curve of $1/(1 + j\omega T)$ is shown in Figure 8–7. The error is symmetric with respect to the corner frequency.

Since the asymptotes are quite easy to draw and are sufficiently close to the exact curve, the use of such approximations in drawing Bode diagrams is convenient in establishing the general nature of the frequency-response characteristics quickly with a minimum amount of calculation and may be used for most preliminary design work. If accurate frequency-response curves are desired, corrections may easily be made by referring to the curve given in Figure 8–7. In practice, an accurate frequency-response curve can be drawn by introducing a correction of 3 dB at the corner frequency and a correction of 1 dB at points one octave below and above the corner frequency and then connecting these points by a smooth curve.

Note that varying the time constant T shifts the corner frequency to the left or to the right, but the shapes of the log-magnitude and the phase-angle curves remain the same.

The transfer function $1/(1 + j\omega T)$ has the characteristics of a low-pass filter. For frequencies above $\omega = 1/T$, the log magnitude falls off rapidly toward $-\infty$. This is essentially due to the presence of the time constant. In the low-pass filter, the output can follow a sinusoidal input faithfully at low frequencies. But as the input frequency is increased, the output cannot follow the input because a certain amount of time is required for the system to build up in magnitude. Thus, at high frequencies, the amplitude of the output approaches zero and the phase angle of the output approaches −90°. Therefore, if the input function contains many harmonics, then the low-frequency components are reproduced faithfully at the output, while the high-frequency components are attenuated in amplitude and shifted in phase. Thus, a first-order element yields exact, or almost exact, duplication only for constant or slowly varying phenomena.

An advantage of the Bode diagram is that for reciprocal factors—for example, the factor $1 + j\omega T$—the log-magnitude and the phase-angle curves need only be changed in sign, since

$$20 \log|1 + j\omega T| = -20 \log\left|\frac{1}{1 + j\omega T}\right|$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

and

$$\underline{/1 + j\omega T} = \tan^{-1}\omega T = -\underline{/\frac{1}{1 + j\omega T}}$$

The corner frequency is the same for both cases. The slope of the high-frequency asymptote of $1 + j\omega T$ is 20 dB/decade, and the phase angle varies from 0° to 90° as the frequency ω is increased from zero to infinity. The log-magnitude curve, together with the asymptotes, and the phase-angle curve for the factor $1 + j\omega T$ are shown in Figure 8–8.

To draw a phase curve accurately, we have to locate several points on the curve. The phase angles of $(1 + j\omega T)^{\mp 1}$ are

$$\mp 45° \quad \text{at} \quad \omega = \frac{1}{T}$$

$$\mp 26.6° \quad \text{at} \quad \omega = \frac{1}{2T}$$

$$\mp 5.7° \quad \text{at} \quad \omega = \frac{1}{10T}$$

$$\mp 63.4° \quad \text{at} \quad \omega = \frac{2}{T}$$

$$\mp 84.3° \quad \text{at} \quad \omega = \frac{10}{T}$$

For the case where a given transfer function involves terms like $(1 + j\omega T)^{\mp n}$, a similar asymptotic construction may be made. The corner frequency is still at $\omega = 1/T$, and the asymptotes are straight lines. The low-frequency asymptote is a horizontal straight line

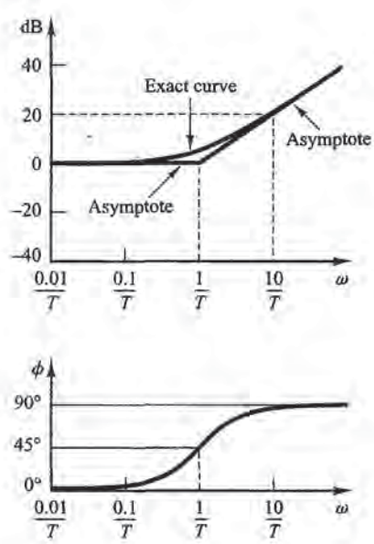


Figure 8–8
Log-magnitude curve, together with the asymptotes, and phase-angle curve for $1 + j\omega T$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

at 0 dB, while the high-frequency asymptote has the slope of $-20n$ dB/decade or $20n$ dB/decade. The error involved in the asymptotic expressions is n times that for $(1 + j\omega T)^{\mp 1}$. The phase angle is n times that of $(1 + j\omega T)^{\mp 1}$ at each frequency point.

Quadratic Factors $[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]^{\mp 1}$. Control systems often possess quadratic factors of the form

$$G(j\omega) = \frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2} \tag{8-7}$$

If $\zeta > 1$, this quadratic factor can be expressed as a product of two first-order factors with real poles. If $0 < \zeta < 1$, this quadratic factor is the product of two complex-conjugate factors. Asymptotic approximations to the frequency-response curves are not accurate for a factor with low values of ζ. This is because the magnitude and phase of the quadratic factor depend on both the corner frequency and the damping ratio ζ.

The asymptotic frequency-response curve may be obtained as follows: Since

$$20 \log \left| \frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2} \right| = -20 \log \sqrt{\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2}$$

for low frequencies such that $\omega \ll \omega_n$, the log magnitude becomes

$$-20 \log 1 = 0 \text{ dB}$$

The low-frequency asymptote is thus a horizontal line at 0 dB. For high frequencies such that $\omega \gg \omega_n$, the log magnitude becomes

$$-20 \log \frac{\omega^2}{\omega_n^2} = -40 \log \frac{\omega}{\omega_n} \text{ dB}$$

The equation for the high-frequency asymptote is a straight line having the slope -40 dB/decade since

$$-40 \log \frac{10\omega}{\omega_n} = -40 - 40 \log \frac{\omega}{\omega_n}$$

The high-frequency asymptote intersects the low-frequency one at $\omega = \omega_n$ since at this frequency

$$-40 \log \frac{\omega_n}{\omega_n} = -40 \log 1 = 0 \text{ dB}$$

This frequency, ω_n, is the corner frequency for the quadratic factor considered.

The two asymptotes just derived are independent of the value of ζ. Near the frequency $\omega = \omega_n$, a resonant peak occurs, as may be expected from Equation (8–7). The damping ratio ζ determines the magnitude of this resonant peak. Errors obviously exist in the approximation by straight-line asymptotes. The magnitude of the error depends on the value of ζ. It is large for small values of ζ. Figure 8–9 shows the exact log-magnitude curves, together with the straight-line asymptotes and the exact

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

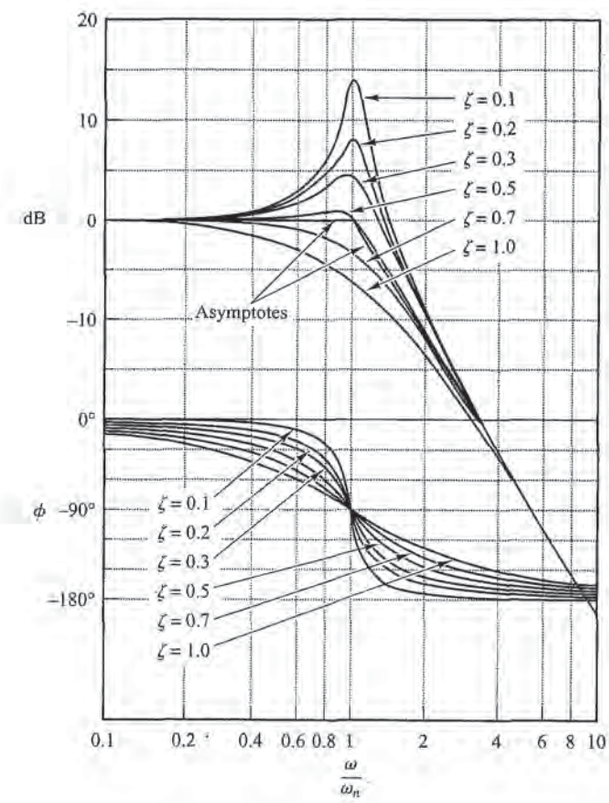


Figure 8–9
Log-magnitude curves, together with the asymptotes, and phase-angle curves of the quadratic transfer function given by Equation (8–7).

phase-angle curves for the quadratic factor given by Equation (8–7) with several values of ζ. If corrections are desired in the asymptotic curves, the necessary amounts of correction at a sufficient number of frequency points may be obtained from Figure 8–9.

The phase angle of the quadratic factor $[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]^{-1}$ is

$$\phi = \angle \frac{1}{1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2} = -\tan^{-1}\left[\frac{2\zeta\frac{\omega}{\omega_n}}{1 - \left(\frac{\omega}{\omega_n}\right)^2}\right] \tag{8–8}$$

The phase angle is a function of both ω and ζ. At $\omega = 0$, the phase angle equals 0°. At the corner frequency $\omega = \omega_n$, the phase angle is −90° regardless of ζ, since

$$\phi = -\tan^{-1}\left(\frac{2\zeta}{0}\right) = -\tan^{-1}\infty = -90°$$

At $\omega = \infty$, the phase angle becomes −180°. The phase-angle curve is skew symmetric about the inflection point—the point where $\phi = -90°$. There are no simple ways to sketch such phase curves. We need to refer to the phase-angle curves shown in Figure 8–9.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The frequency-response curves for the factor

$$1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2$$

can be obtained by merely reversing the sign of the log magnitude and that of the phase angle of the factor

$$\frac{1}{1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2}$$

To obtain the frequency-response curves of a given quadratic transfer function, we must first determine the value of the corner frequency ω_n and that of the damping ratio ζ. Then, by using the family of curves given in Figure 8–9, the frequency-response curves can be plotted.

The Resonant Frequency ω_r and the Resonant Peak Value M_r. The magnitude of

$$G(j\omega) = \frac{1}{1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2}$$

is

$$|G(j\omega)| = \frac{1}{\sqrt{\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2}} \tag{8–9}$$

If $|G(j\omega)|$ has a peak value at some frequency, this frequency is called the *resonant* frequency. Since the numerator of $|G(j\omega)|$ is constant, a peak value of $|G(j\omega)|$ will occur when

$$g(\omega) = \left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2 \tag{8–10}$$

is a minimum. Since Equation (8–10) can be written

$$g(\omega) = \left[\frac{\omega^2 - \omega_n^2(1 - 2\zeta^2)}{\omega_n^2}\right]^2 + 4\zeta^2(1 - \zeta^2) \tag{8–11}$$

the minimum value of $g(\omega)$ occurs at $\omega = \omega_n\sqrt{1 - 2\zeta^2}$. Thus the resonant frequency ω_r is

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2}, \qquad \text{for } 0 \le \zeta \le 0.707 \tag{8–12}$$

As the damping ratio ζ approaches zero, the resonant frequency approaches ω_n. For $0 < \zeta \le 0.707$, the resonant frequency ω_r is less than the damped natural frequency $\omega_d = \omega_n\sqrt{1 - \zeta^2}$, which is exhibited in the transient response. From Equation (8–12), it can be seen that for $\zeta > 0.707$, there is no resonant peak. The magnitude $|G(j\omega)|$ decreases monotonically with increasing frequency ω. (The magnitude is less than 0 dB for all values of $\omega > 0$. Recall that, for $0.7 < \zeta < 1$, the step response is oscillatory, but the oscillations are well damped and are hardly perceptible.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

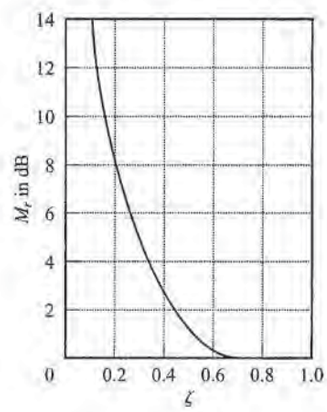


Figure 8–10
M_r-versus-ζ curve for the second-order system $1/[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]$.

The magnitude of the resonant peak, M_r, can be found by substituting Equation (8–12) into Equation (8–9). For $0 \le \zeta \le 0.707$,

$$M_r = |G(j\omega)|_{\max} = |G(j\omega_r)| = \frac{1}{2\zeta\sqrt{1-\zeta^2}} \tag{8–13}$$

For $\zeta > 0.707$,

$$M_r = 1 \tag{8–14}$$

As ζ approaches zero, M_r approaches infinity. This means that if the undamped system is excited at its natural frequency, the magnitude of $G(j\omega)$ becomes infinity. The relationship between M_r and ζ is shown in Figure 8–10.

The phase angle of $G(j\omega)$ at the frequency where the resonant peak occurs can be obtained by substituting Equation (8–12) into Equation (8–8). Thus, at the resonant frequency ω_r,

$$\underline{/G(j\omega_r)} = -\tan^{-1}\frac{\sqrt{1-2\zeta^2}}{\zeta} = -90^\circ + \sin^{-1}\frac{\zeta}{\sqrt{1-\zeta^2}}$$

General Procedure for Plotting Bode Diagrams. MATLAB provides an easy way to plot Bode diagrams. (See Section 8–3.) Here, however, we consider the case where we want to draw Bode diagrams manually without using MATLAB.

First rewrite the sinusoidal transfer function $G(j\omega)H(j\omega)$ as a product of basic factors discussed above. Then identify the corner frequencies associated with these basic factors. Finally, draw the asymptotic log-magnitude curves with proper slopes between the corner frequencies. The exact curve, which lies close to the asymptotic curve, can be obtained by adding proper corrections.

The phase-angle curve of $G(j\omega)H(j\omega)$ can be drawn by adding the phase-angle curves of individual factors.

The use of Bode diagrams employing asymptotic approximations requires much less time than other methods that may be used for computing the frequency response of a transfer function. The ease of plotting the frequency-response curves for a given transfer function and the ease of modification of the frequency-response curve as compensation is added are the main reasons why Bode diagrams are very frequently used in practice.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 8–3 Draw the Bode diagram for the following transfer function:

$$G(j\omega) = \frac{10(j\omega + 3)}{(j\omega)(j\omega + 2)[(j\omega)^2 + j\omega + 2]}$$

Make corrections so that the log-magnitude curve is accurate.

To avoid any possible mistakes in drawing the log-magnitude curve, it is desirable to put $G(j\omega)$ in the following normalized form, where the low-frequency asymptotes for the first-order factors and the second-order factor are the 0-dB line:

$$G(j\omega) = \frac{7.5\left(\frac{j\omega}{3} + 1\right)}{(j\omega)\left(\frac{j\omega}{2} + 1\right)\left[\frac{(j\omega)^2}{2} + \frac{j\omega}{2} + 1\right]}$$

This function is composed of the following factors:

$$7.5, \qquad (j\omega)^{-1}, \qquad 1 + j\frac{\omega}{3}, \qquad \left(1 + j\frac{\omega}{2}\right)^{-1}, \qquad \left[1 + j\frac{\omega}{2} + \frac{(j\omega)^2}{2}\right]^{-1}$$

The corner frequencies of the third, fourth, and fifth terms are $\omega = 3$, $\omega = 2$, and $\omega = \sqrt{2}$, respectively. Note that the last term has the damping ratio of 0.3536.

To plot the Bode diagram, the separate asymptotic curves for each of the factors are shown in Figure 8–11. The composite curve is then obtained by algebraically adding the individual curves, also shown in Figure 8–11. Note that when the individual asymptotic curves are added at each frequency, the slope of the composite curve is cumulative. Below $\omega = \sqrt{2}$, the plot has the slope of −20 dB/decade. At the first corner frequency $\omega = \sqrt{2}$, the slope changes to −60 dB/decade and continues to the next corner frequency $\omega = 2$, where the slope becomes −80 dB/decade. At the last corner frequency $\omega = 3$, the slope changes to −60 dB/decade.

Once such an approximate log-magnitude curve has been drawn, the actual curve can be obtained by adding corrections at each corner frequency and at frequencies one octave below and above the corner frequencies. For first-order factors $(1 + j\omega T)^{\mp 1}$, the corrections are ±3 dB at the corner frequency and ±1 dB at the frequencies one octave below and above the corner frequency. Corrections necessary for the quadratic factor are obtained from Figure 8–9. The exact log-magnitude curve for $G(j\omega)$ is shown by a dashed curve in Figure 8–11.

Note that any change in the slope of the magnitude curve is made only at the corner frequencies of the transfer function $G(j\omega)$. Therefore, instead of drawing individual magnitude curves and adding them up, as shown, we may sketch the magnitude curve without sketching individual curves. We may start drawing the lowest-frequency portion of the straight line (that is, the straight line with the slope −20 dB/decade for $\omega < \sqrt{2}$). As the frequency is increased, we get the effect of the complex-conjugate poles (quadratic term) at the corner frequency $\omega = \sqrt{2}$. The complex-conjugate poles cause the slopes of the magnitude curve to change from −20 to −60 dB/decade. At the next corner frequency, $\omega = 2$, the effect of the pole is to change the slope to −80 dB/decade. Finally, at the corner frequency $\omega = 3$, the effect of the zero is to change the slope from −80 to −60 dB/decade.

For plotting the complete phase-angle curve, the phase-angle curves for all factors have to be sketched. The algebraic sum of all phase-angle curves provides the complete phase-angle curve, as shown in Figure 8–11.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

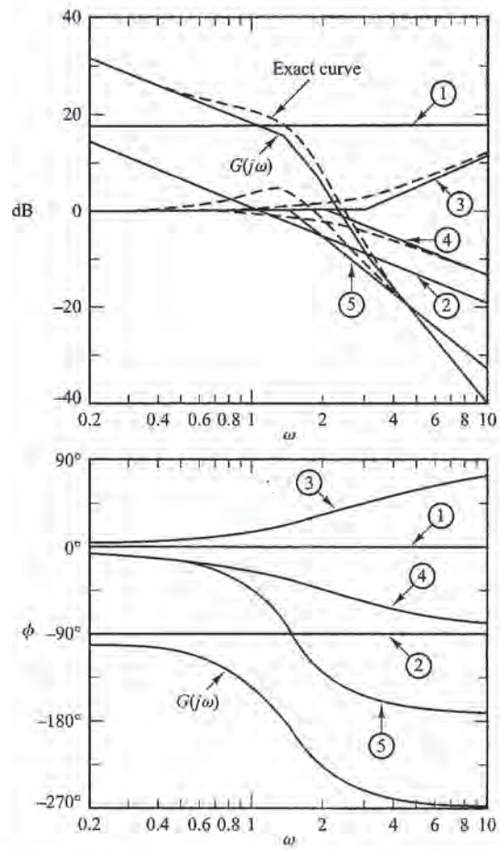


Figure 8–11
Bode diagram of the system considered in Example 8–3.

Minimum-Phase Systems and Nonminimum-Phase Systems. Transfer functions having neither poles nor zeros in the right-half s plane are minimum-phase transfer functions, whereas those having poles and/or zeros in the right-half s plane are nonminimum-phase transfer functions. Systems with minimum-phase transfer functions are called *minimum-phase* systems, whereas those with nonminimum-phase transfer functions are called *nonminimum-phase* systems.

For systems with the same magnitude characteristic, the range in phase angle of the minimum-phase transfer function is minimum among all such systems, while the range in phase angle of any nonminimum-phase transfer function is greater than this minimum.

It is noted that for a minimum-phase system, the transfer function can be uniquely determined from the magnitude curve alone. For a nonminimum-phase system, this is not the case. Multiplying any transfer function by all-pass filters does not alter the magnitude curve, but the phase curve is changed.

Consider as an example the two systems whose sinusoidal transfer functions are, respectively,

$$G_1(j\omega) = \frac{1 + j\omega T}{1 + j\omega T_1}, \qquad G_2(j\omega) = \frac{1 - j\omega T}{1 + j\omega T_1}, \qquad 0 < T < T_1$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

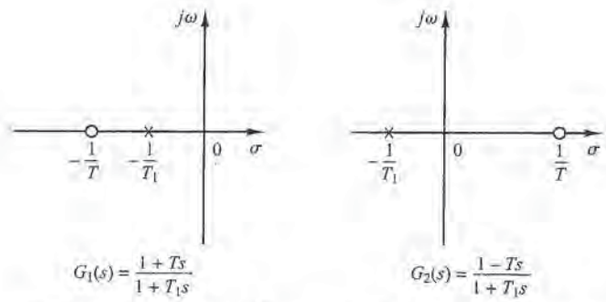


Figure 8–12
Pole–zero configurations of a minimum-phase system $G_1(s)$ and nonminimum-phase system $G_2(s)$.

The pole–zero configurations of these systems are shown in Figure 8–12. The two sinusoidal transfer functions have the same magnitude characteristics, but they have different phase-angle characteristics, as shown in Figure 8–13. These two systems differ from each other by the factor

$$G(j\omega) = \frac{1 - j\omega T}{1 + j\omega T}$$

The magnitude of the factor $(1 - j\omega T)/(1 + j\omega T)$ is always unity. But the phase angle equals $-2\tan^{-1}\omega T$ and varies from 0° to −180° as ω is increased from zero to infinity.

As stated earlier, for a minimum-phase system, the magnitude and phase-angle characteristics are uniquely related. This means that if the magnitude curve of a system is specified over the entire frequency range from zero to infinity, then the phase-angle curve is uniquely determined, and vice versa. This, however, does not hold for a nonminimum-phase system.

Nonminimum-phase situations may arise in two different ways. One is simply when a system includes a nonminimum-phase element or elements. The other situation may arise in the case where a minor loop is unstable.

For a minimum-phase system, the phase angle at $\omega = \infty$ becomes $-90°(q - p)$, where p and q are the degrees of the numerator and denominator polynomials of the transfer function, respectively. For a nonminimum-phase system, the phase angle at $\omega = \infty$ differs from $-90°(q - p)$. In either system, the slope of the log-magnitude curve at $\omega = \infty$ is equal to $-20(q - p)$ dB/decade. It is therefore possible to detect whether the system is minimum phase by examining both the slope of the high-frequency asymptote of the log-magnitude curve and the phase angle at $\omega = \infty$. If the slope of the log-magnitude curve as ω approaches infinity is $-20(q - p)$ dB/decade and the phase angle at $\omega = \infty$ is equal to $-90°(q - p)$, then the system is minimum phase.

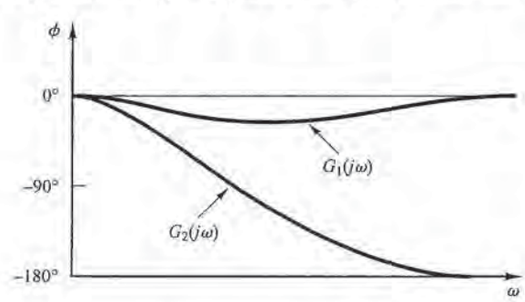


Figure 8–13
Phase-angle characteristics of the systems $G_1(s)$ and $G_2(s)$ shown in Figure 8–12.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Nonminimum-phase systems are slow in responding because of their faulty behavior at the start of a response. In most practical control systems, excessive phase lag should be carefully avoided. In designing a system, if fast speed of response is of primary importance, we should not use nonminimum-phase components. (A common example of nonminimum-phase elements that may be present in control systems is transport lag or dead time.)

It is noted that the techniques of frequency-response analysis and design to be presented in this and the next chapter are valid for both minimum-phase and nonminimum-phase systems.

Transport Lag. Transport lag, which is also called dead time, is of nonminimum-phase behavior and has an excessive phase lag with no attenuation at high frequencies. Such transport lags normally exist in thermal, hydraulic, and pneumatic systems.

Consider the transport lag given by

$$G(j\omega) = e^{-j\omega T}$$

The magnitude is always equal to unity since

$$|G(j\omega)| = |\cos\omega T - j\sin\omega T| = 1$$

Therefore, the log magnitude of the transport lag $e^{-j\omega T}$ is equal to 0 dB. The phase angle of the transport lag is

$$\begin{aligned} \underline{/G(j\omega)} &= -\omega T \qquad \text{(radians)} \\ &= -57.3\,\omega T \qquad \text{(degrees)} \end{aligned}$$

The phase angle varies linearly with the frequency ω. The phase-angle characteristic of transport lag is shown in Figure 8–14.

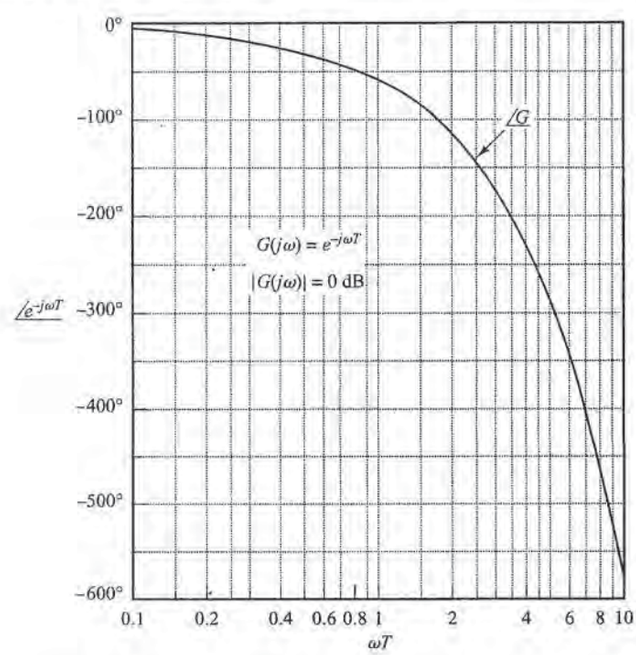


Figure 8–14
Phase-angle characteristic of transport lag.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 8–4 Draw the Bode diagram of the following transfer function:

$$G(j\omega) = \frac{e^{-j\omega L}}{1 + j\omega T}$$

The log magnitude is

$$20 \log|G(j\omega)| = 20 \log|e^{-j\omega L}| + 20 \log\left|\frac{1}{1 + j\omega T}\right|$$

$$= 0 + 20 \log\left|\frac{1}{1 + j\omega T}\right|$$

The phase angle of $G(j\omega)$ is

$$\underline{/G(j\omega)} = \underline{/e^{-j\omega L}} + \underline{\left/\frac{1}{1 + j\omega T}\right.}$$

$$= -\omega L - \tan^{-1}\omega T$$

The log-magnitude and phase-angle curves for this transfer function with $L = 0.5$ and $T = 1$ are shown in Figure 8–15.

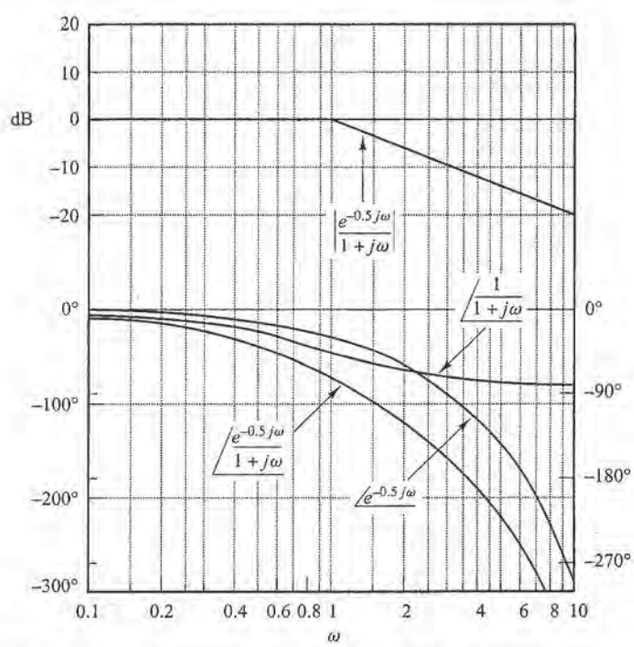


Figure 8–15
Bode diagram for the system $e^{-j\omega L}/(1 + j\omega T)$ with $L = 0.5$ and $T = 1$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

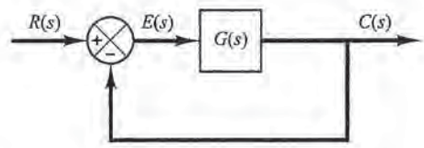


Figure 8–16
Unity-feedback control system.

Relationship between System Type and Log-Magnitude Curve. Consider the unity-feedback control system. The static position, velocity, and acceleration error constants describe the low-frequency behavior of type 0, type 1, and type 2 systems, respectively. For a given system, only one of the static error constants is finite and significant. (The larger the value of the finite static error constant, the higher the loop gain is as ω approaches zero.)

The type of the system determines the slope of the log-magnitude curve at low frequencies. Thus, information concerning the existence and magnitude of the steady-state error of a control system to a given input can be determined from the observation of the low-frequency region of the log-magnitude curve.

Determination of Static Position Error Constants. Consider the unity-feedback control system shown in Figure 8–16. Assume that the open-loop transfer function is given by

$$G(s) = \frac{K(T_a s + 1)(T_b s + 1)\cdots(T_m s + 1)}{s^N(T_1 s + 1)(T_2 s + 1)\cdots(T_p s + 1)}$$

or

$$G(j\omega) = \frac{K(T_a j\omega + 1)(T_b j\omega + 1)\cdots(T_m j\omega + 1)}{(j\omega)^N(T_1 j\omega + 1)(T_2 j\omega + 1)\cdots(T_p j\omega + 1)}$$

Figure 8–17 shows an example of the log-magnitude plot of a type 0 system. In such a system, the magnitude of $G(j\omega)$ equals K_p at low frequencies, or

$$\lim_{\omega \to 0} G(j\omega) = K = K_p$$

It follows that the low-frequency asymptote is a horizontal line at 20 log K_p dB.

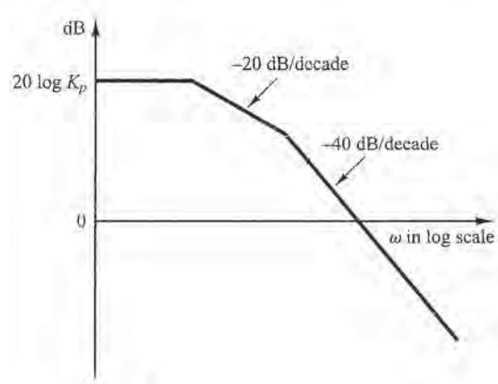


Figure 8–17
Log-magnitude curve of a type 0 system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Determination of Static Velocity Error Constants. Consider the unity-feedback control system shown in Figure 8–16. Figure 8–18 shows an example of the log-magnitude plot of a type 1 system. The intersection of the initial −20-dB/decade segment (or its extension) with the line $\omega = 1$ has the magnitude $20 \log K_v$. This may be seen as follows: In a type 1 system

$$G(j\omega) = \frac{K_v}{j\omega}, \qquad \text{for } \omega \ll 1$$

Thus,

$$20 \log \left| \frac{K_v}{j\omega} \right|_{\omega=1} = 20 \log K_v$$

The intersection of the initial −20-dB/decade segment (or its extension) with the 0-dB line has a frequency numerically equal to K_v. To see this, define the frequency at this intersection to be ω_1; then

$$\left| \frac{K_v}{j\omega_1} \right| = 1$$

or

$$K_v = \omega_1$$

As an example, consider the type 1 system with unity feedback whose open-loop transfer function is

$$G(s) = \frac{K}{s(Js + F)}$$

If we define the corner frequency to be ω_2 and the frequency at the intersection of the −40-dB/decade segment (or its extension) with 0-dB line to be ω_3, then

$$\omega_2 = \frac{F}{J}, \qquad \omega_3^2 = \frac{K}{J}$$

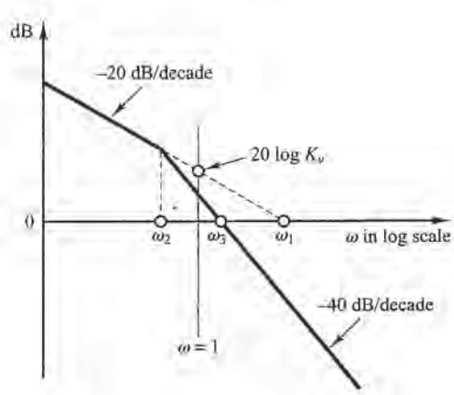


Figure 8–18
Log-magnitude curve of a type 1 system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Since

$$\omega_1 = K_v = \frac{K}{F}$$

it follows that

$$\omega_1 \omega_2 = \omega_3^2$$

or

$$\frac{\omega_1}{\omega_3} = \frac{\omega_3}{\omega_2}$$

On the Bode diagram,

$$\log \omega_1 - \log \omega_3 = \log \omega_3 - \log \omega_2$$

Thus, the ω_3 point is just midway between the ω_2 and ω_1 points. The damping ratio ζ of the system is then

$$\zeta = \frac{F}{2\sqrt{KJ}} = \frac{\omega_2}{2\omega_3}$$

Determination of Static Acceleration Error Constants. Consider the unity-feedback control system shown in Figure 8–16. Figure 8–19 shows an example of the log-magnitude plot of a type 2 system. The intersection of the initial −40-dB/decade segment (or its extension) with the $\omega = 1$ line has the magnitude of 20 log K_a. Since at low frequencies

$$G(j\omega) = \frac{K_a}{(j\omega)^2}, \qquad \text{for } \omega \ll 1$$

it follows that

$$20 \log \left| \frac{K_a}{(j\omega)^2} \right|_{\omega=1} = 20 \log K_a$$

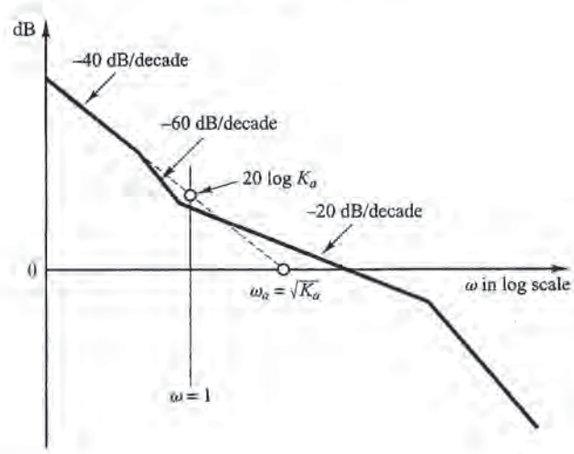


Figure 8–19
Log-magnitude curve of a type 2 system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The frequency ω_a at the intersection of the initial −40-dB/decade segment (or its extension) with the 0-dB line gives the square root of K_a numerically. This can be seen from the following:

$$20 \log \left| \frac{K_a}{(j\omega_a)^2} \right| = 20 \log 1 = 0$$

which yields

$$\omega_a = \sqrt{K_a}$$

8–3 PLOTTING BODE DIAGRAMS WITH MATLAB

The command bode computes magnitudes and phase angles of the frequency response of continuous-time, linear, time-invariant systems.

When the command bode (without left-hand arguments) is entered in the computer, MATLAB produces a Bode plot on the screen. Most commonly used bode commands are

```
bode(num,den)
bode(num,den,w)
bode(A,B,C,D)
bode(A,B,C,D,w)
bode(sys)
```

When invoked with left-hand arguments, such as

```
[mag,phase,w] = bode(num,den,w)
```

bode returns the frequency response of the system in matrices mag, phase, and w. No plot is drawn on the screen. The matrices mag and phase contain magnitudes and phase angles of the frequency response of the system, evaluated at user-specified frequency points. The phase angle is returned in degrees. The magnitude can be converted to decibels with the statement

```
magdB = 20*log10(mag)
```

Other Bode commands with left-hand arguments are

```
[mag,phase,w] = bode(num,den)
[mag,phase,w] = bode(num,den,w)
[mag,phase,w] = bode(A,B,C,D)
[mag,phase,w] = bode(A,B.C,D,w)
[mag,phase,w] = bode(A,B,C,D,iu,w)
[mag,phase,w] = bode(sys)
```

To specify the frequency range, use the command logspace(d1,d2) or logspace (d1,d2,n). logspace(d1,d2) generates a vector of 50 points logarithmically equally spaced between decades 10^{d1} and 10^{d2}. (50 points include both endpoints. There are 48 points between the endpoints.) To generate 50 points between 0.1 rad/sec and 100 rad/sec. enter the command

```
w = logspace(-1,2)
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

logspace(dl,d2,n) generates n points logarithmically equally spaced between decades 10^{d1} and 10^{d2}. (n points include both endpoints.) For example, to generate 100 points between 1 rad/sec and 1000 rad/sec, enter the following command:

```
w = logspace(0,3,100)
```

To incorporate the user-specified frequency points when plotting Bode diagrams, the bode command must include the frequency vector w, such as bode(num,den,w) and [mag,phase,w] = bode(A,B,C,D,w).

EXAMPLE 8–5 Consider the following transfer function:

$$G(s) = \frac{25}{s^2 + 4s + 25}$$

Plot a Bode diagram for this transfer function.

When the system is defined in the form

$$G(s) = \frac{\text{num}(s)}{\text{den}(s)}$$

use the command bode(num,den) to draw the Bode diagram. [When the numerator and denominator contain the polynomial coefficients in descending powers of s, bode(num,den) draws the Bode diagram.] MATLAB Program 8–1 shows a program to plot the Bode diagram for this system. The resulting Bode diagram is shown in Figure 8–20.

MATLAB Program 8–1

```
num = [0 0 25];
den = [1 4 25];
bode(num,den)
title('Bode Diagram of G(s) = 25/(s^2 + 4s + 25)')
```

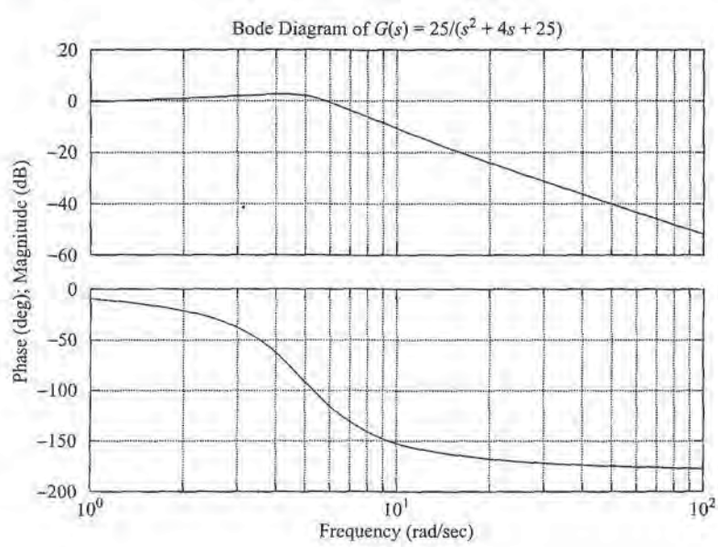


Figure 8–20
Bode diagram of $G(s) = \dfrac{25}{s^2 + 4s + 25}$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 8–6 Consider the system shown in Figure 8–21. The open-loop transfer function is

$$G(s) = \frac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}$$

Plot a bode diagram.

MATLAB Program 8–2 plots a Bode diagram for the system. The resulting plot is shown in Figure 8–22. The frequency range in this case is automatically determined to be from 0.01 to 10 rad/sec.

MATLAB Program 8–2

```
num = [0 9 1.8 9];
den = [1 1.2 9 0];
bode(num,den)
title('Bode Diagram of G(s) = 9(s^2 + 0.2s + 1)/[s(s^2 + 1.2s + 9)]')
```

If it is desired to plot the Bode diagram from 0.01 to 1000 rad/sec, enter the following command:

```
w = logspace(-2,3,100)
```

This command generates 100 points logarithmically equally spaced between 0.01 and 100 rad/sec. (Note that such a vector w specifies the frequencies in radians per second at which the frequency response will be calculated.)

If we use the command

```
bode(num,den,w)
```

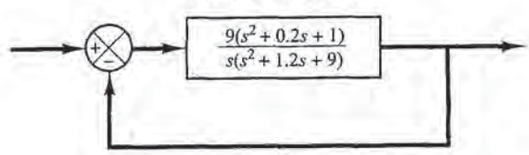


Figure 8–21
Control system.

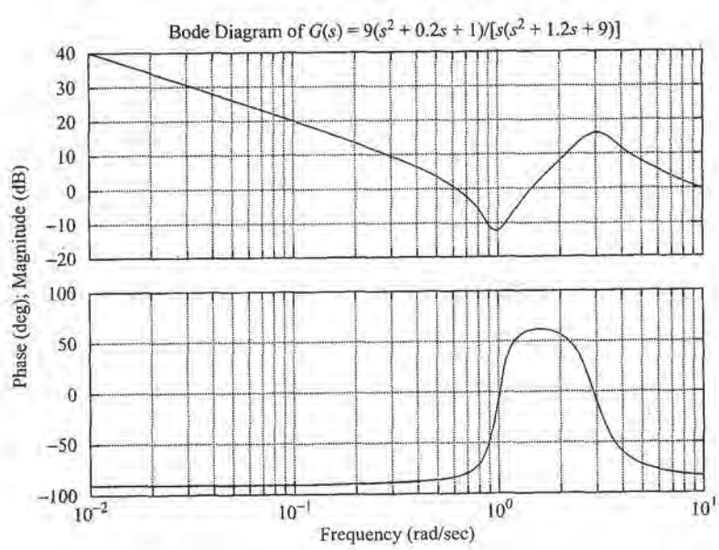


Figure 8–22
Bode diagram of
$G(s) = \dfrac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

then the frequency range is as the user specified, but the magnitude range and phase-angle range will be automatically determined. See MATLAB Program 8–3 and the resulting plot in Figure 8–23.

To specify the magnitude range and phase-angle range, use the following command:

```
[mag,phase,w] = bode(num,den,w)
```

The matrices mag and phase contain the magnitudes and phase angles of the frequency response evaluated at the user-specified frequency points. The phase angle is returned in degrees. The magnitude can be converted to decibels with the statement

```
magdb = 20*log10(mag)
```

MATLAB Program 8–3

```
num = [0 9 1.8 9];
den = [1 1.2 9 0];
w = logspace(-2,3,100);
bode(num,den,w)
title('Bode Diagram of G(s) = 9(s^2 + 0.2s + 1)/[s(s^2 + 1.2s + 9)]')
```

If we wish to specify the magnitude range to be, for example, between −50 dB and +50 dB, then enter lines at −50 dB and +50 dB in the plot by specifying dBmax (maximum magnitude) and dBmin (minimum magnitude) as follows:

```
dBmax = 50*ones(1,100);
dBmin = -50*ones(1,100);
```

Then enter the following semilog plot command:

```
semilogx(w,magdB,'o',w,magdB,'-',w,dBmax,'--',w,dBmin,':')
```

(Note that the number of dBmax points and that of dBmin points must be equal to the number of frequency points in w. In this example, all numbers are 100.) Then the screen will show the magnitude curve magdB with 'o' marks.

The range for the magnitude is normally a multiple of 5, 10, 20, or 50 dB. (There are exceptions.) For the present case, the range for the magnitude will be from −50 dB to +50 dB.

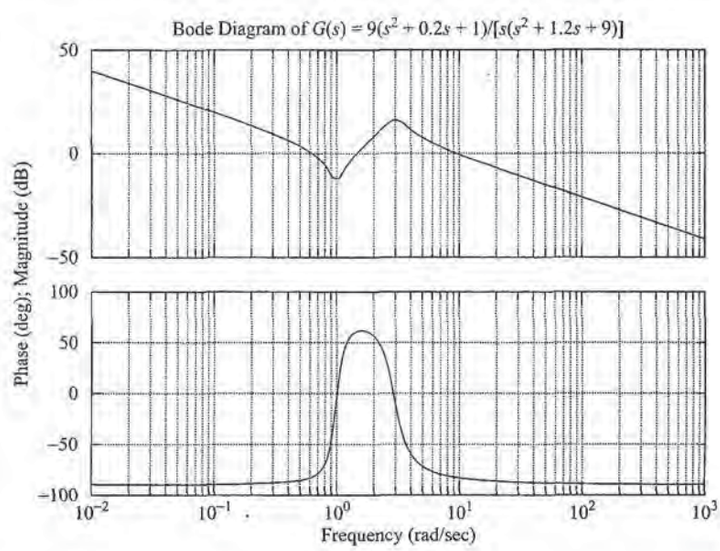


Figure 8–23
Bode diagram of
$$G(s) = \frac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}.$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

For the phase angle, if we wish to specify the range to be, for example, between −150° and +150°, we enter lines at −150° and +150° in the program by specifying pmax (maximum phase angle) and pmin (minimum phase angle) as follows:

```
pmax = 150*ones(1,100);
pmin = -150*ones(1,100);
```

Then enter the semilog plot command:

```
semilogx(w,phase,'o',w,phase,'-',w,pmax,'--',w,pmin,':')
```

(The number of pmax points and that of pmin points must be equal to the number of frequency points in w.) The screen will show the phase curve.

The range for the phase angle is normally a multiple of 5°, 10°, 50°, or 100°. (There are exceptions.) For the present case, the range for the phase angle will be from −150° to +150°.

MATLAB Program 8–4 produces the Bode diagram for the system such that the frequency range is from 0.01 to 1000 rad/sec, the magnitude range is from −50 to +50 dB (the magnitude range is a multiple of 50 dB), and the phase-angle range is from −150° to +150° (the phase-angle range is a multiple of 50°). Figure 8–24 shows the Bode diagram obtained by use of MATLAB Program 8–4.

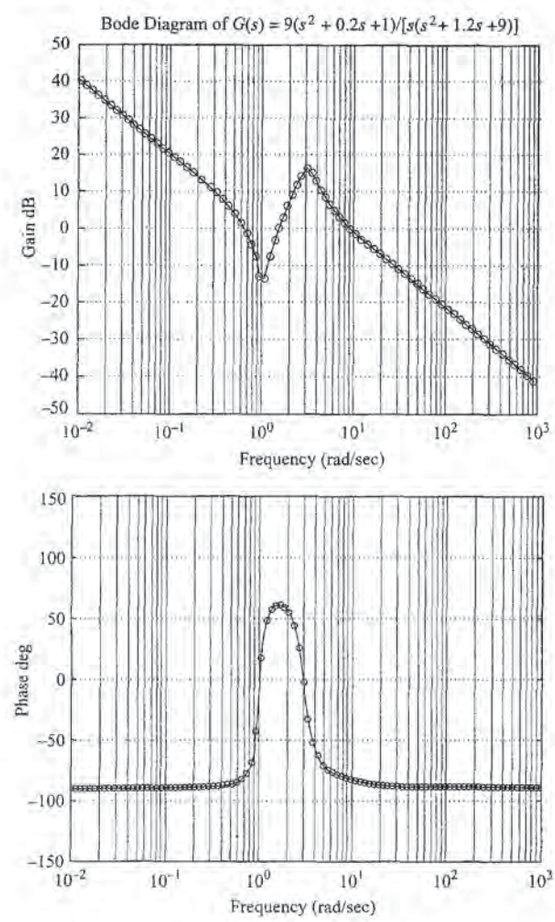


Figure 8–24
Bode diagram of
$$G(s) = \frac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}.$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 8–4

```
% ---------- Bode diagram ----------

% ***** Enter the numerator and denominator of the transfer
% function *****

num = [0  9  1.8  9];
den = [1  1.2  9  0];

% ***** Specify the frequency range and enter the command
% [mag,phase,w] = bode(num,den,w) *****

w = logspace(-2,3,100);
[mag,phase,w] = bode(num,den,w);

% Convert mag to decibels *****

magdB = 20*log10(mag);

% ***** Specify the range for magnitude. For the system
% considered, the magnitude range should include -50 dB
% and +50 dB.  Enter dBmax and dBmin in the program and
% draw dBmax line and dBmin line.  To plot the magdB curve
% enter the following dBmax, dBmin, and semilogx command *****

dBmax = 50*ones(1,100);
dBmin = -50*ones(1,100);
semilogx(w,magdB,'o',w,magdB,'-',w,dBmax,'-',w,dBmin,'-')

% ***** Enter grid, title, xlabel, and ylabel *****

grid
title('Bode Diagram of G(s) = 9(s^2+0.2s+1)/[s(s^2+1.2s+9)]')
xlabel('Frequency (rad/sec)')
ylabel('Gain dB')

% ***** Next, we shall plot the phase angle curve *****

% ***** Specify the range for phase angle. For the system
% considered, the phase angle range should include -150 degrees
% and +150 degrees. Enter pmax and pmin in the program and
% draw pmax line and pmin line. To plot the phase curve, enter
% the following pmax, pmin, and semilog command *****

pmax = 150*ones(1,100);
pmin = -150*ones(1,100);
semilogx(w,phase,'o',w,phase,'-',w,pmax,'-',w,pmin,'-')

% ***** Enter grid, xlabel, and ylabel *****

grid
xlabel('Frequency (rad/sec)')
ylabel('Phase deg')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Obtaining Bode Diagrams of Systems Defined in State Space. Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{y}$ = output vector (m-vector)
$\mathbf{u}$ = control vector (r-vector)
$\mathbf{A}$ = state matrix ($n \times n$ matrix)
$\mathbf{B}$ = control matrix ($n \times r$ matrix)
$\mathbf{C}$ = output matrix ($m \times n$ matrix)
$\mathbf{D}$ = direct transmission matrix ($m \times r$ matrix)

A Bode diagram for this system may be obtained by entering the command

bode(A,B,C,D)

or others listed at the beginning of this section.

The command bode(A,B,C,D) produces a series of Bode plots, one for each input of the system, with the frequency range automatically determined. (More points are used when the response is changing rapidly.)

The command bode(A,B,C,D,iu), where iu is the ith input of the system, produces the Bode diagrams from the input iu to all the outputs $(y_1, y_2, \ldots, y_m)$ of the system, with a frequency range automatically determined. (The scalar iu is an index into the inputs of the system and specifies which input is to be used for plotting Bode diagrams). If the control vector $\mathbf{u}$ has three inputs such that

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

then iu must be set to either 1, 2, or 3.

If the system has only one input u, then either of the following commands may be used:

bode(A,B,C,D)

or

bode(A,B,C,D,1)

EXAMPLE 8–7 Consider the following system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -25 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 25 \end{bmatrix} u$$
$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

This system has one input u and one output y. By using the command

bode(A,B,C,D)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

and entering MATLAB Program 8–5 into the computer, we obtain the Bode diagram shown in Figure 8–25.

MATLAB Program 8–5

```
A = [0 1;-25 -4];
B = [0;25];
C = [1 0];
D = [0];
bode(A,B,C,D)
title('Bode Diagram')
```

If we replace the command bode(A,B,C,D) in MATLAB Program 8–5 with

```
bode(A,B,C,D,1)
```

then MATLAB will produce the Bode diagram identical to that shown in Figure 8–25.

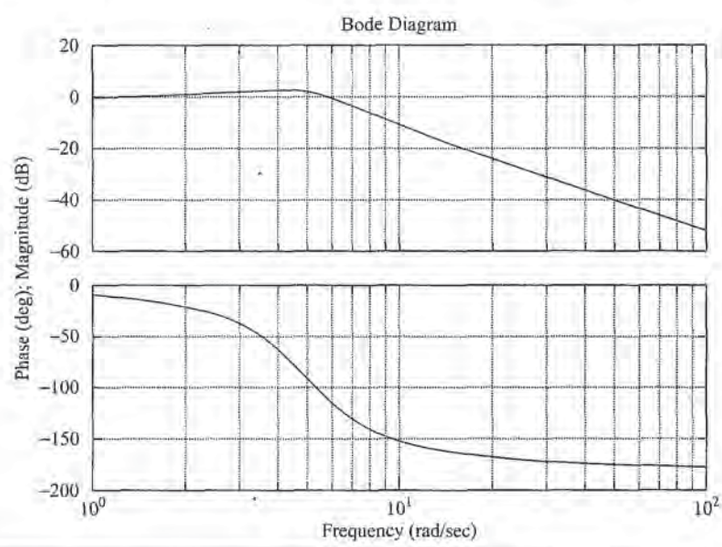


Figure 8–25
Bode diagram of the system considered in Example 8–7.

8–4 POLAR PLOTS

The polar plot of a sinusoidal transfer function $G(j\omega)$ is a plot of the magnitude of $G(j\omega)$ versus the phase angle of $G(j\omega)$ on polar coordinates as ω is varied from zero to infinity. Thus, the polar plot is the locus of vectors $|G(j\omega)| \underline{/G(j\omega)}$ as ω is varied from zero to infinity. Note that in polar plots a positive (negative) phase angle is measured counterclockwise (clockwise) from the positive real axis. The polar plot is often called the Nyquist plot. An example of such a plot is shown in Figure 8–26. Each point on the polar plot of $G(j\omega)$ represents the terminal point of a vector at a particular value of ω. In the polar plot, it is important to show the frequency graduation of the locus. The projections of $G(j\omega)$ on the real and imaginary axes are its real and imaginary components.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

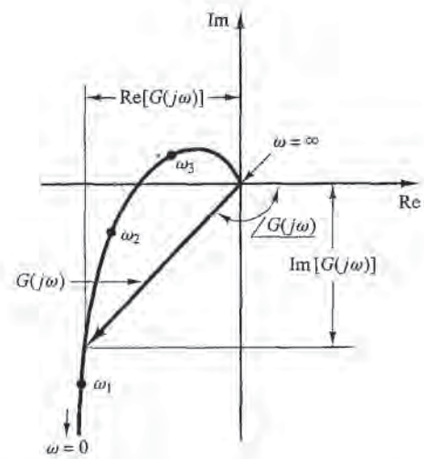


Figure 8–26
Polar plot.

MATLAB may be used to obtain a polar plot $G(j\omega)$ or to obtain $|G(j\omega)|$ and $\angle G(j\omega)$ accurately for various values of ω in the frequency range of interest. (See Section 8–5.)

An advantage in using a polar plot is that it depicts the frequency-response characteristics of a system over the entire frequency range in a single plot. One disadvantage is that the plot does not clearly indicate the contributions of each individual factor of the open-loop transfer function.

Integral and Derivative Factors $(j\omega)^{\mp 1}$. The polar plot of $G(j\omega) = 1/j\omega$ is the negative imaginary axis since

$$G(j\omega) = \frac{1}{j\omega} = -j\frac{1}{\omega} = \frac{1}{\omega}\angle -90^\circ$$

The polar plot of $G(j\omega) = j\omega$ is the positive imaginary axis.

First-Order Factors $(1 + j\omega T)^{\mp 1}$. For the sinusoidal transfer function

$$G(j\omega) = \frac{1}{1 + j\omega T} = \frac{1}{\sqrt{1 + \omega^2 T^2}}\angle -\tan^{-1}\omega T$$

the values of $G(j\omega)$ at $\omega = 0$ and $\omega = 1/T$ are, respectively,

$$G(j0) = 1\angle 0^\circ \quad \text{and} \quad G\left(j\frac{1}{T}\right) = \frac{1}{\sqrt{2}}\angle -45^\circ$$

If ω approaches infinity, the magnitude of $G(j\omega)$ approaches zero and the phase angle approaches -90°. The polar plot of this transfer function is a semicircle as the frequency ω is varied from zero to infinity, as shown in Figure 8–27(a). The center is located at 0.5 on the real axis, and the radius is equal to 0.5.

To prove that the polar plot of the first-order factor $G(j\omega) = 1/(1 + j\omega T)$ is a semicircle, define

$$G(j\omega) = X + jY$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

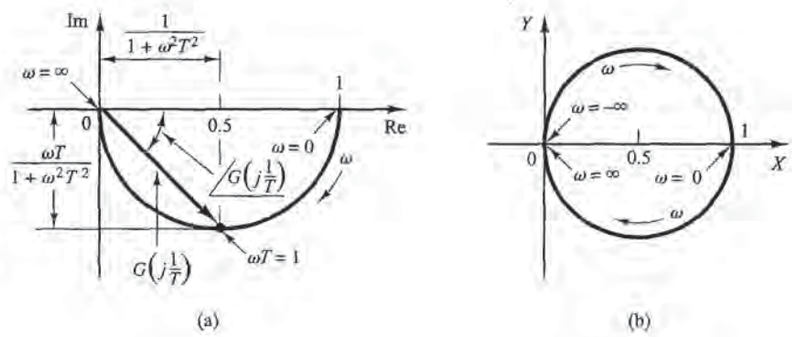


Figure 8–27
(a) Polar plot of $1/(1 + j\omega T)$; (b) plot of $G(j\omega)$ in X-Y plane.

where

$$X = \frac{1}{1 + \omega^2T^2} = \text{real part of } G(j\omega)$$

$$Y = \frac{-\omega T}{1 + \omega^2T^2} = \text{imaginary part of } G(j\omega)$$

Then we obtain

$$\left(X - \frac{1}{2}\right)^2 + Y^2 = \left(\frac{1}{2}\frac{1 - \omega^2T^2}{1 + \omega^2T^2}\right)^2 + \left(\frac{-\omega T}{1 + \omega^2T^2}\right)^2 = \left(\frac{1}{2}\right)^2$$

Thus, in the X-Y plane $G(j\omega)$ is a circle with center at $X = \frac{1}{2}$, $Y = 0$ and with radius $\frac{1}{2}$, as shown in Figure 8–27(b). The lower semicircle corresponds to $0 \le \omega \le \infty$, and the upper semicircle corresponds to $-\infty \le \omega \le 0$.

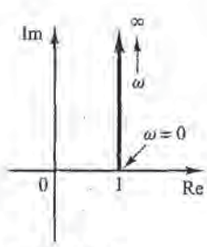


Figure 8–28
Polar plot of $1 + j\omega T$.

The polar plot of the transfer function $1 + j\omega T$ is simply the upper half of the straight line passing through point (1,0) in the complex plane and parallel to the imaginary axis, as shown in Figure 8–28. The polar plot of $1 + j\omega T$ has an appearance completely different from that of $1/(1 + j\omega T)$.

Quadratic Factors $[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]^{\mp 1}$. The low- and high-frequency portions of the polar plot of the following sinusoidal transfer function

$$G(j\omega) = \frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}, \qquad \text{for } \zeta > 0$$

are given, respectively, by

$$\lim_{\omega \to 0} G(j\omega) = 1\underline{/0^\circ} \quad \text{and} \quad \lim_{\omega \to \infty} G(j\omega) = 0\underline{/-180^\circ}$$

The polar plot of this sinusoidal transfer function starts at $1\underline{/0^\circ}$ and ends at $0\underline{/-180^\circ}$ as ω increases from zero to infinity. Thus, the high-frequency portion of $G(j\omega)$ is tangent to the negative real axis.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

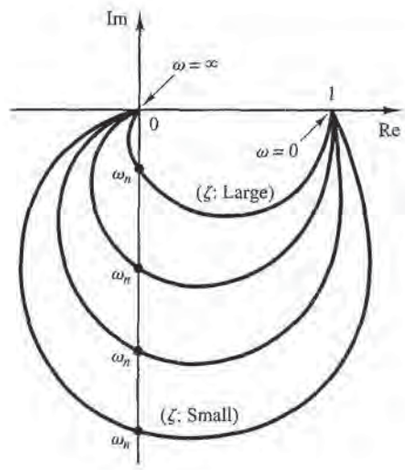


Figure 8–29
Polar plots of $\dfrac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}$, for $\zeta > 0$.

Examples of polar plots of the transfer function just considered are shown in Figure 8–29. The exact shape of a polar plot depends on the value of the damping ratio ζ, but the general shape of the plot is the same for both the underdamped case $(1 > \zeta > 0)$ and overdamped case $(\zeta > 1)$.

For the underdamped case at $\omega = \omega_n$, we have $G(j\omega_n) = 1/(j2\zeta)$, and the phase angle at $\omega = \omega_n$ is $-90°$. Therefore, it can be seen that the frequency at which the $G(j\omega)$ locus intersects the imaginary axis is the undamped natural frequency ω_n. In the polar plot, the frequency point whose distance from the origin is maximum corresponds to the resonant frequency ω_r. The peak value of $G(j\omega)$ is obtained as the ratio of the magnitude of the vector at the resonant frequency ω_r to the magnitude of the vector at $\omega = 0$. The resonant frequency ω_r is indicated in the polar plot shown in Figure 8–30.

For the overdamped case, as ζ increases well beyond unity, the $G(j\omega)$ locus approaches a semicircle. This may be seen from the fact that, for a heavily damped system, the characteristic roots are real and one is much smaller than the other. Since, for sufficiently large ζ, the effect of the larger root (larger in the absolute value) on the response becomes very small, the system behaves like a first-order one.

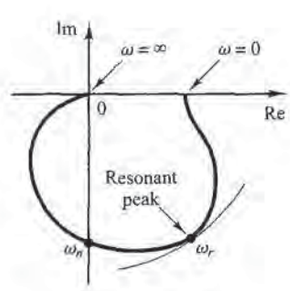


Figure 8–30
Polar plot showing the resonant peak and resonant frequency ω_r.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

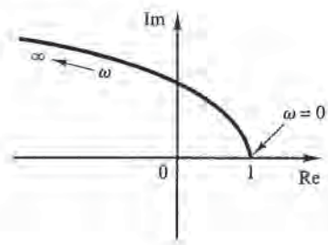


Figure 8–31
Polar plot of
$1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2$, for $\zeta > 0$.

Next, consider the following sinusoidal transfer function:

$$G(j\omega) = 1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2$$
$$= \left(1 - \frac{\omega^2}{\omega_n^2}\right) + j\left(\frac{2\zeta\omega}{\omega_n}\right)$$

The low-frequency portion of the curve is

$$\lim_{\omega\to 0} G(j\omega) = 1\underline{/0^\circ}$$

and the high-frequency portion is

$$\lim_{\omega\to\infty} G(j\omega) = \infty\underline{/180^\circ}$$

Since the imaginary part of $G(j\omega)$ is positive for $\omega > 0$ and is monotonically increasing, and the real part of $G(j\omega)$ is monotonically decreasing from unity, the general shape of the polar plot of $G(j\omega)$ is as shown in Figure 8–31. The phase angle is between 0° and 180°.

EXAMPLE 8–8 Consider the following second-order transfer function:

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

Sketch a polar plot of this transfer function.

Since the sinusoidal transfer function can be written

$$G(j\omega) = \frac{1}{j\omega(1 + j\omega T)} = -\frac{T}{1 + \omega^2T^2} - j\frac{1}{\omega(1 + \omega^2T^2)}$$

the low-frequency portion of the polar plot becomes

$$\lim_{\omega\to 0} G(j\omega) = -T - j\infty = \infty\underline{/-90^\circ}$$

and the high-frequency portion becomes

$$\lim_{\omega\to\infty} G(j\omega) = 0 - j0 = 0\underline{/-180^\circ}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

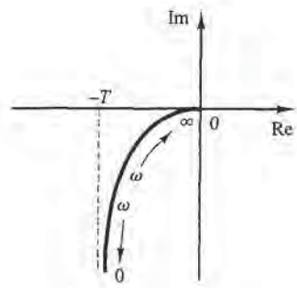


Figure 8–32
Polar plot of $1/[j\omega(1 + j\omega T)]$.

The general shape of the polar plot of $G(j\omega)$ is shown in Figure 8–32. The $G(j\omega)$ plot is asymptotic to the vertical line passing through the point $(-T, 0)$. Since this transfer function involves an integrator $(1/s)$, the general shape of the polar plot differs substantially from those of second-order transfer functions that do not have an integrator.

EXAMPLE 8–9 Obtain the polar plot of the following transfer function:

$$G(j\omega) = \frac{e^{-j\omega L}}{1 + j\omega T}$$

Since $G(j\omega)$ can be written

$$G(j\omega) = \left(e^{-j\omega L}\right)\left(\frac{1}{1 + j\omega T}\right)$$

the magnitude and phase angle are, respectively,

$$|G(j\omega)| = |e^{-j\omega L}| \cdot \left|\frac{1}{1 + j\omega T}\right| = \frac{1}{\sqrt{1 + \omega^2 T^2}}$$

and

$$\underline{/G(j\omega)} = \underline{/e^{-j\omega L}} + \underline{\left/\frac{1}{1 + j\omega T}\right.} = -\omega L - \tan^{-1}\omega T$$

Since the magnitude decreases from unity monotonically and the phase angle also decreases monotonically and indefinitely, the polar plot of the given transfer function is a spiral, as shown in Figure 8–33.

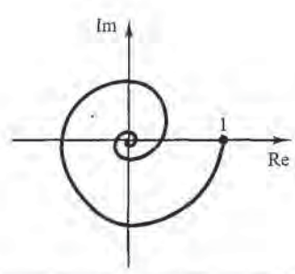


Figure 8–33
Polar plot of $e^{-j\omega L}/(1 + j\omega T)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

General Shapes of Polar Plots. The polar plots of a transfer function of the form

$$G(j\omega) = \frac{K(1 + j\omega T_a)(1 + j\omega T_b)\cdots}{(j\omega)^{\lambda}(1 + j\omega T_1)(1 + j\omega T_2)\cdots}$$

$$= \frac{b_0(j\omega)^m + b_1(j\omega)^{m-1} + \cdots}{a_0(j\omega)^n + a_1(j\omega)^{n-1} + \cdots}$$

where $n > m$ or the degree of the denominator polynomial is greater than that of the numerator, will have the following general shapes:

1. *For $\lambda = 0$ or type 0 systems:* The starting point of the polar plot (which corresponds to $\omega = 0$) is finite and is on the positive real axis. The tangent to the polar plot at $\omega = 0$ is perpendicular to the real axis. The terminal point, which corresponds to $\omega = \infty$, is at the origin, and the curve is tangent to one of the axes.
2. *For $\lambda = 1$ or type 1 systems:* the $j\omega$ term in the denominator contributes $-90°$ to the total phase angle of $G(j\omega)$ for $0 \leq \omega \leq \infty$. At $\omega = 0$, the magnitude of $G(j\omega)$ is infinity, and the phase angle becomes $-90°$. At low frequencies, the polar plot is asymptotic to a line parallel to the negative imaginary axis. At $\omega = \infty$, the magnitude becomes zero, and the curve converges to the origin and is tangent to one of the axes.
3. *For $\lambda = 2$ or type 2 systems:* The $(j\omega)^2$ term in the denominator contributes $-180°$ to the total phase angle of $G(j\omega)$ for $0 \leq \omega \leq \infty$. At $\omega = 0$, the magnitude of $G(j\omega)$ is infinity, and the phase angle is equal to $-180°$. At low frequencies, the polar plot is asymptotic to a line parallel to the negative real axis. At $\omega = \infty$, the magnitude becomes zero, and the curve is tangent to one of the axes.

The general shapes of the low-frequency portions of the polar plots of type 0, type 1, and type 2 systems are shown in Figure 8–34. It can be seen that, if the degree of the denominator polynomial of $G(j\omega)$ is greater than that of the numerator, then the $G(j\omega)$

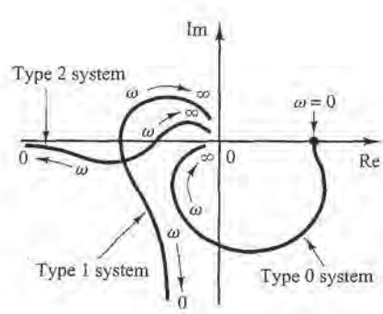
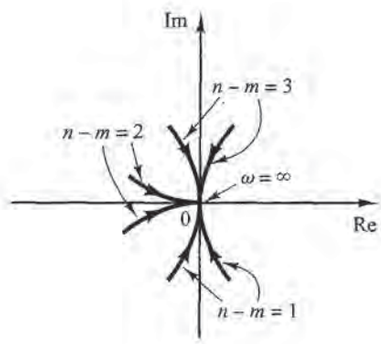


Figure 8–34
Polar plots of type 0, type 1, and type 2 systems.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565



$$G(j\omega) = \frac{b_o(j\omega)^m + \cdots}{a_o(j\omega)^n + \cdots}$$

Figure 8–35
Polar plots in the high-frequency range.

loci converge to the origin clockwise. At $\omega = \infty$, the loci are tangent to one or the other axes, as shown in Figure 8–35.

Note that any complicated shapes in the polar plot curves are caused by the numerator dynamics, that is, by the time constants in the numerator of the transfer function. Figure 8–36 shows examples of polar plots of transfer functions with numerator dynamics. In analyzing control systems, the polar plot of $G(j\omega)$ in the frequency range of interest must be accurately determined.

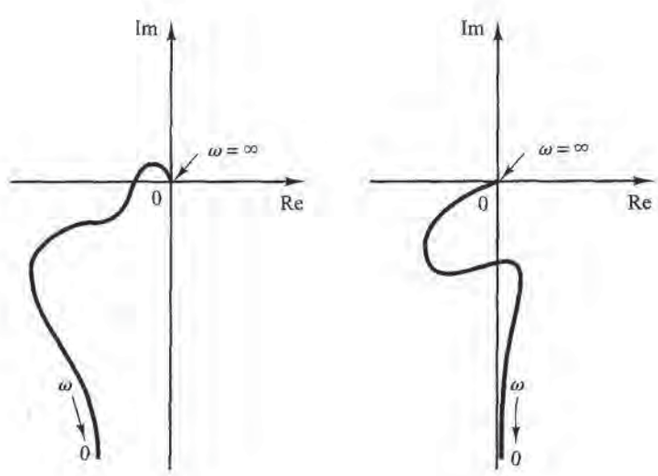


Figure 8–36
Polar plots of transfer functions with numerator dynamics.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Table 8–1 Polar Plots of Simple Transfer Functions

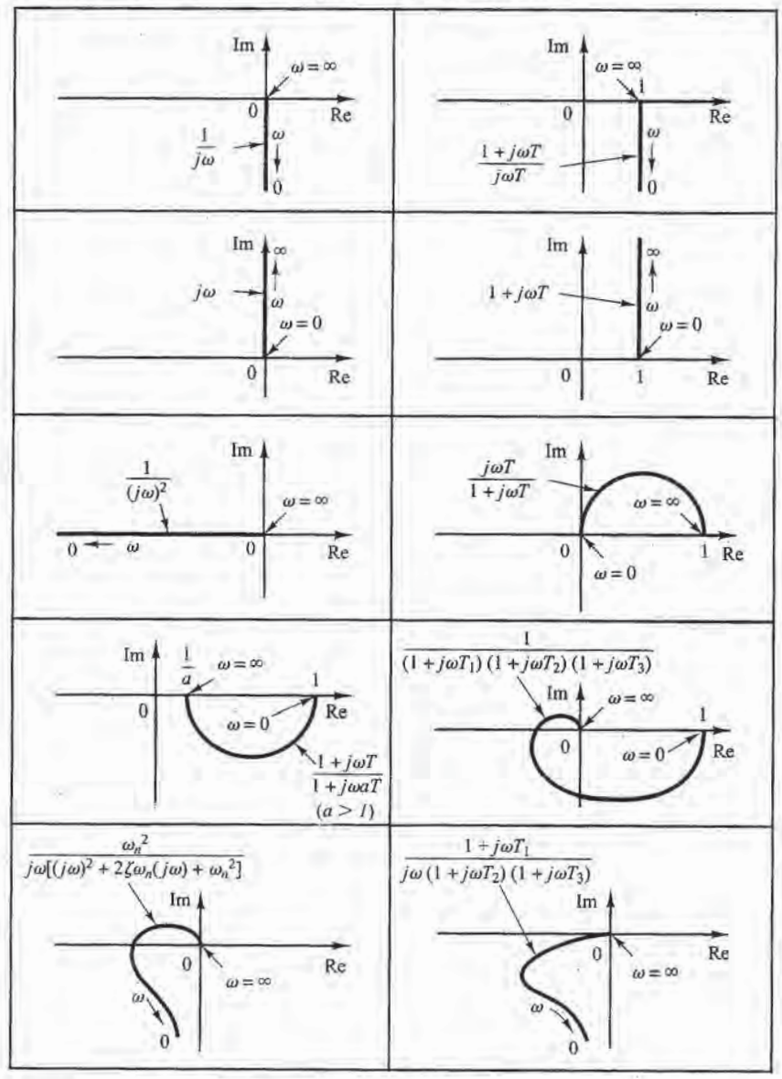


Table 8–1 shows sketches of polar plots of several transfer functions.

8–5 DRAWING NYQUIST PLOTS WITH MATLAB

Nyquist plots, just like Bode diagrams, are commonly used in the frequency-response representation of linear, time-invariant, feedback control systems. Nyquist plots are polar plots, while Bode diagrams are rectangular plots. One plot or the other may be more

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

convenient for a particular operation, but a given operation can always be carried out in either plot.

The MATLAB command nyquist computes the frequency response for continuous-time, linear, time-invariant systems. When invoked without left-hand arguments, nyquist produces a Nyquist plot on the screen.

The command

```
nyquist(num,den)
```

draws the Nyquist plot of the transfer function

$$G(s) = \frac{\text{num}(s)}{\text{den}(s)}$$

where num and den contain the polynomial coefficients in descending powers of s. Other commonly used nyquist commands are

```
nyquist(num,den,w)
nyquist(A,B,C,D)
nyquist(A,B,C,D,w)
nyquist(A,B,C,D,iu,w)
nyquist(sys)
```

The command involving the user-specified frequency vector w, such as

```
nyquist(num,den,w)
```

calculates the frequency response at the specified frequency points in radians per second.

When invoked with left-hand arguments such as

```
[re,im,w] = nyquist(num,den)
[re,im,w] = nyquist(num,den,w)
[re,im,w] = nyquist(A,B,C,D)
[re,im,w] = nyquist(A,B,C,D,w)
[re,im,w] = nyquist(A,B,C,D,iu,w)
[re,im,w] = nyquist(sys)
```

MATLAB returns the frequency response of the system in the matrices re, im, and w. No plot is drawn on the screen. The matrices re and im contain the real and imaginary parts of the frequency response of the system, evaluated at the frequency points specified in the vector w. Note that re and im have as many columns as outputs and one row for each element in w.

EXAMPLE 8–10 Consider the following open-loop transfer function:

$$G(s) = \frac{1}{s^2 + 0.8s + 1}$$

Draw a Nyquist plot with MATLAB.

Since the system is given in the form of the transfer function, the command

```
nyquist(num,den)
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

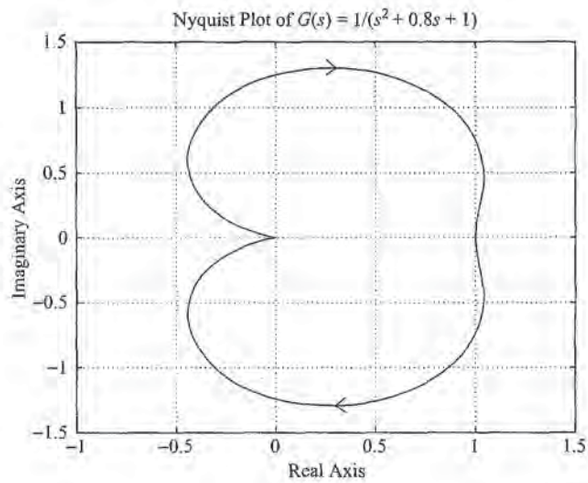


Figure 8–37
Nyquist plot of
$G(s) = \dfrac{1}{s^2 + 0.8s + 1}$.

may be used to draw a Nyquist plot. MATLAB Program 8–6 produces the Nyquist plot shown in Figure 8–37. In this plot, the ranges for the real axis and imaginary axis are automatically determined.

MATLAB Program 8–6

```
num = [0  0  1];
den = [1  0.8  1];
nyquist(num,den)
grid
title('Nyquist Plot of G(s) = 1/(s^2 + 0.8s + 1)')
```

If we wish to draw the Nyquist plot using manually determined ranges, for example, from −2 to 2 on the real axis and from −2 to 2 on the imaginary axis, enter the following command into the computer:

```
v = [-2  2  -2  2];
axis(v);
```

or, combining these two lines into one,

```
axis([-2  2  -2  2]);
```

See MATLAB Program 8–7 and the resulting Nyquist plot shown in Figure 8–38.

MATLAB Program 8–7

```
% ---------- Nyquist plot ----------

num = [0  0  1];
den = [1  0.8  1];
nyquist(num,den)
v = [-2  2  -2  2]; axis(v)
grid
title('Nyquist Plot of G(s) = 1/(s^2 + 0.8s + 1)')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

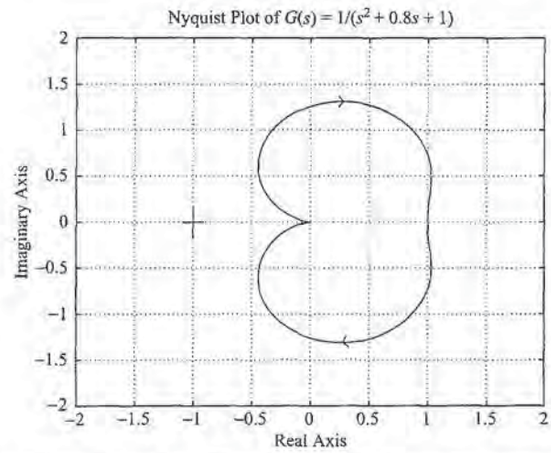


Figure 8–38
Nyquist plot of
$G(s) = \dfrac{1}{s^2 + 0.8s + 1}$.

Caution. In drawing a Nyquist plot, where a MATLAB operation involves "Divide by zero," the resulting Nyquist plot may be erroneous. For example, if the transfer function $G(s)$ is given by

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

then the MATLAB command

```
num = [0 0 1];
den = [1 1 0];
nyquist(num,den)
```

produces an erroneous Nyquist plot. An example of an erroneous Nyquist plot is shown in Figure 8–39. If such an erroneous Nyquist plot appears on the computer, then it can be corrected if we specify the axis(v). For example, if we enter the axis command

```
v = [-2 2 -5 5]; axis(v)
```

in the computer, then a correct Nyquist plot can be obtained. See Example 8–11.

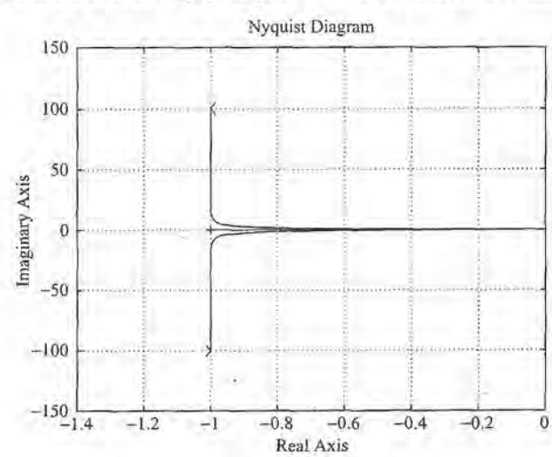


Figure 8–39
Erroneous Nyquist plot.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 8–11 Draw a Nyquist plot for the following $G(s)$:

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

MATLAB Program 8–8 will produce a correct Nyquist plot on the computer, even though a warning message "Divide by zero" may appear on the screen. The resulting Nyquist plot is shown in Figure 8–40.

MATLAB Program 8–8

```
% ---------- Nyquist plot----------
num = [0 0 1];
den = [1 1 0];
nyquist(num,den)
v = [-2 2 -5 5]; axis(v)
grid
title('Nyquist Plot of G(s) = 1/[s(s + 1)]')
```

Notice that the Nyquist plot shown in Figure 8–40 includes the loci for both $\omega > 0$ and $\omega < 0$. If we wish to draw the Nyquist plot for only the positive frequency region ($\omega > 0$), then we need to use the command

```
[re,im,w]=nyquist(num,den,w)
```

A MATLAB program using this nyquist command is shown in MATLAB Program 8–9. The resulting Nyquist plot is presented in Figure 8–41.

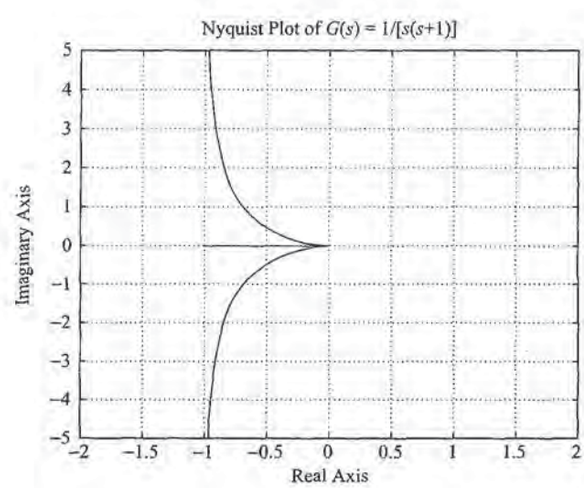


Figure 8–40
Nyquisit plot of $G(s) = \dfrac{1}{s(s+1)}$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 8–9

```
% ---------- Nyquist plot----------
num = [0  0  1];
den = [1  1  0];
w = 0.1:0.1:100;
[re,im,w] = nyquist(num,den,w);
plot(re,im)
v = [-2  2  -5  5]; axis(v)
grid
title('Nyquist Plot of G(s) = 1/[s(s + 1)]')
xlabel('Real Axis')
ylabel('Imag Axis')
```

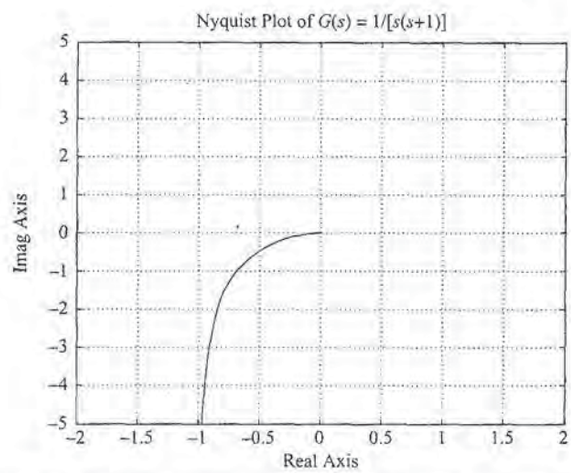


Figure 8–41
Nyquist plot of $G(s) = \dfrac{1}{s(s+1)}$ for $\omega > 0$.

Drawing Nyquist Plots of a System Defined in State Space. Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{y}$ = output vector (m-vector)
$\mathbf{u}$ = control vector (r-vector)
$\mathbf{A}$ = state matrix ($n \times n$ matrix)
$\mathbf{B}$ = control matrix ($n \times r$ matrix)
$\mathbf{C}$ = output matrix ($m \times n$ matrix)
$\mathbf{D}$ = direct transmission matrix ($m \times r$ matrix)

Nyquist plots for this system may be obtained by the use of the command

```
nyquist(A,B,C,D)
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

This command produces a series of Nyquist plots, one for each input and output combination of the system. The frequency range is automatically determined.

The command

```
nyquist(A,B,C,D,iu)
```

produces Nyquist plots from the single input iu to all the outputs of the system, with the frequency range determined automatically. The scalar iu is an index into the inputs of the system and specifies which input to use for the frequency response.

The command

```
nyquist(A,B,C,D,iu,w)
```

uses the user-supplied frequency vector w. The vector w specifies the frequencies in radians per second at which the frequency response should be calculated.

EXAMPLE 8–12 Consider the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -25 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 25 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + [0]u$$

Draw a Nyquist plot.

This system has a single input u and a single output y. A Nyquist plot may be obtained by entering the command

```
nyquist(A,B,C,D)
```

or

```
nyquist(A,B,C,D,1)
```

MATLAB Program 8–10 will provide the Nyquist plot. (Note that we obtain the identical result by using either of these two commands.) Figure 8–42 shows the Nyquist plot produced by MATLAB Program 8–10.

MATLAB Program 8–10

```
A = [0 1;-25 -4];
B = [0;25];
C = [1 0];
D = [0];
nyquist(A,B,C,D)
grid
title('Nyquist Plot')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

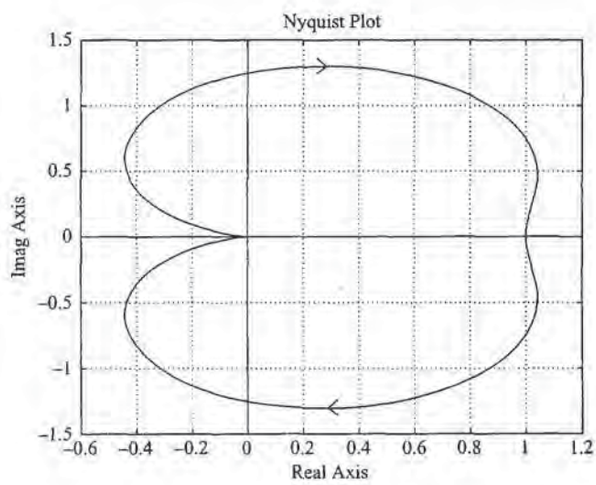


Figure 8–42
Nyquist plot of system considered in Example 8–12.

EXAMPLE 8–13 Consider the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 6.5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

This system involves two inputs and two outputs. There are four sinusoidal output–input relationships: $Y_1(j\omega)/U_1(j\omega)$, $Y_2(j\omega)/U_1(j\omega)$, $Y_1(j\omega)/U_2(j\omega)$, and $Y_2(j\omega)/U_2(j\omega)$. Draw Nyquist plots for the system. (When considering input u_1, we assume that input u_2 is zero, and vice versa.)

The four individual Nyquist plots can be obtained by the use of the command

```
nyquist(A,B,C,D)
```

MATLAB Program 8–11 produces the four Nyquist plots. They are shown in Figure 8–43.

MATLAB Program 8–11

```
A = [-1 -1;6.5 0];
B = [1 1;1 0];
C = [1 0;0 1];
D = [0 0;0 0];
nyquist(A,B,C,D)
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

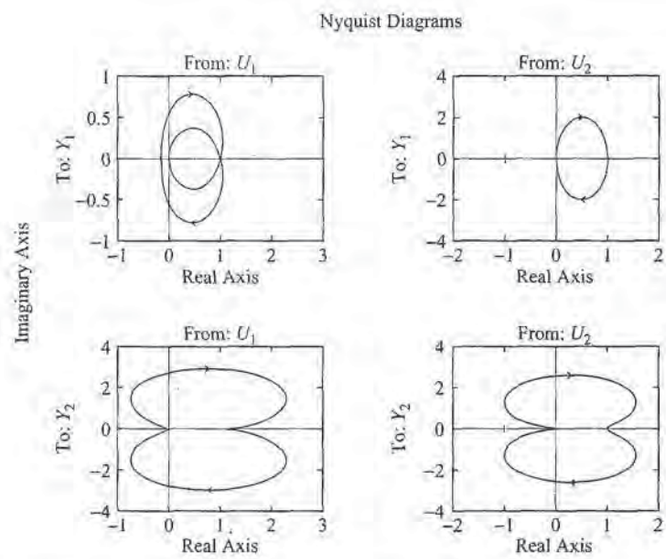


Figure 8–43
Nyquist plot of system considered in Example 8–13.

8–6 LOG-MAGNITUDE-VERSUS-PHASE PLOTS

Another approach to graphically portraying the frequency-response characteristics is to use the log-magnitude-versus-phase plot, which is a plot of the logarithmic magnitude in decibels versus the phase angle or phase margin for a frequency range of interest. [The phase margin is the difference between the actual phase angle ϕ and $-180°$; that is, $\phi - (-180°) = 180° + \phi$.] The curve is graduated in terms of the frequency ω. Such log-magnitude-versus-phase plots are commonly called Nichols plots.

In the Bode diagram, the frequency-response characteristics of $G(j\omega)$ are shown on semilog paper by two separate curves, the log-magnitude curve and the phase-angle curve, while in the log-magnitude-versus-phase plot, the two curves in the Bode diagram are combined into one. In the manual approach the log-magnitude-versus-phase plot can easily be constructed by reading values of the log magnitude and phase angle from the Bode diagram. Notice that in the log-magnitude-versus-phase plot a change in the gain constant of $G(j\omega)$ merely shifts the curve up (for increasing gain) or down (for decreasing gain), but the shape of the curve remains the same.

Advantages of the log-magnitude-versus-phase plot are that the relative stability of the closed-loop system can be determined quickly and that compensation can be worked out easily.

The log-magnitude-versus-phase plot for the sinusoidal transfer function $G(j\omega)$ and that for $1/G(j\omega)$ are skew symmetrical about the origin since

$$\left|\frac{1}{G(j\omega)}\right| \text{ in dB} = -|G(j\omega)| \text{ in dB}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

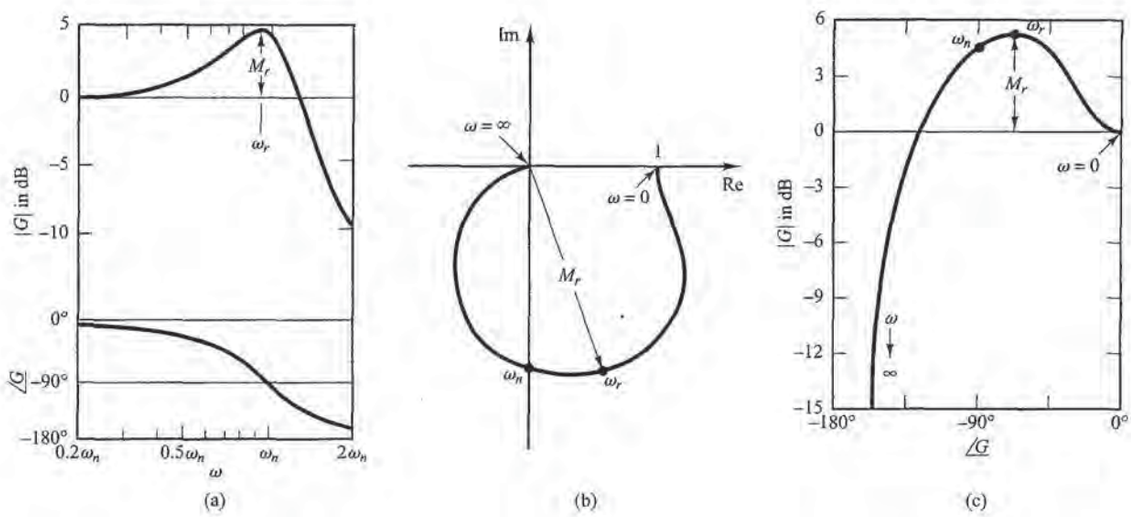


Figure 8–44

Three representations of the frequency response of $\dfrac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}$, for $\zeta > 0$.

(a) Bode diagram; (b) polar plot; (c) log-magnitude-versus-phase plot.

and

$$\underline{\left/\frac{1}{G(j\omega)}\right.} = -\underline{/G(j\omega)}$$

Figure 8–44 compares frequency-response curves of

$$G(j\omega) = \frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}$$

in three different representations. In the log-magnitude-versus-phase plot, the vertical distance between the points $\omega = 0$ and $\omega = \omega_r$, where ω_r is the resonant frequency, is the peak value of $G(j\omega)$ in decibels.

Since log-magnitude and phase-angle characteristics of basic transfer functions have been discussed in detail in Sections 8–2 and 8–3, it will be sufficient here to give examples of some log-magnitude-versus-phase plots. Table 8–2 shows such examples. (However, more on Nichols charts will be discussed in Section 8–10.)

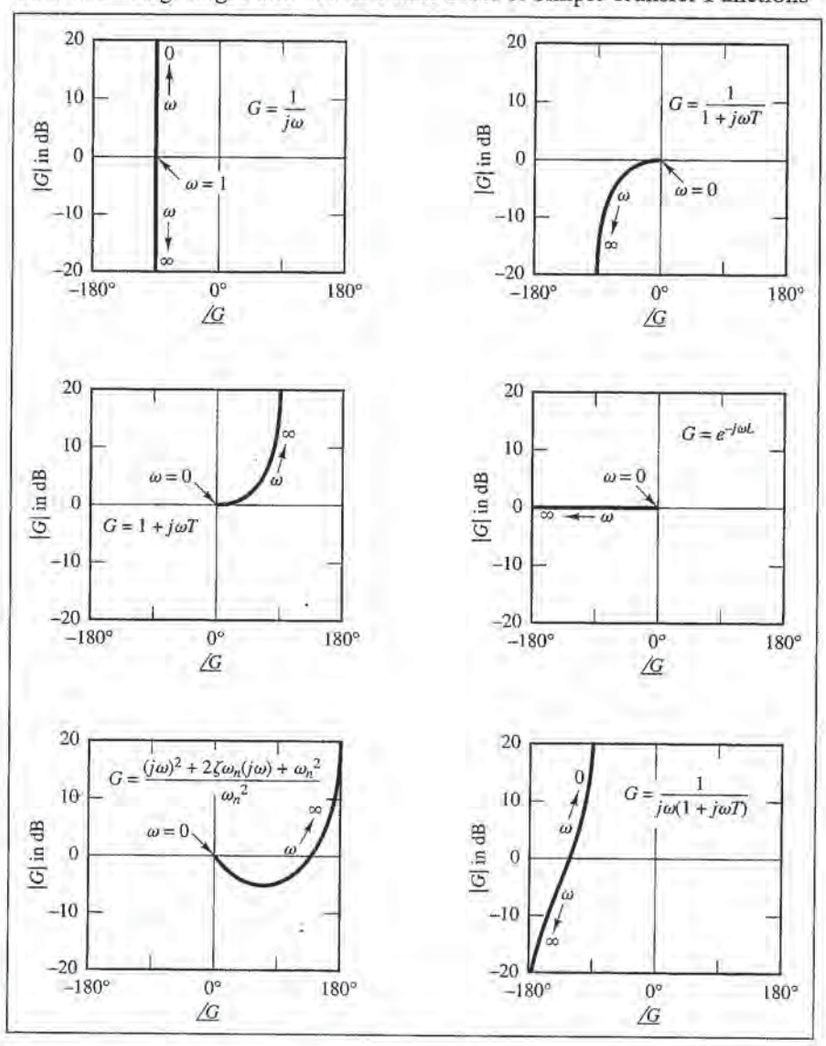
8–7 NYQUIST STABILITY CRITERION

The Nyquist stability criterion determines the stability of a closed-loop system from its open-loop frequency response and open-loop poles.

This section presents mathematical background for understanding the Nyquist stability criterion. Consider the closed-loop system shown in Figure 8–45. The closed-loop transfer function is

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Table 8–2 Log-Magnitude-versus-Phase Plots of Simple Transfer Functions



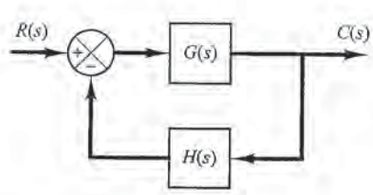


Figure 8–45
Closed-loop system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

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

For stability, all roots of the characteristic equation

$$1 + G(s)H(s) = 0$$

must lie in the left-half s plane. [It is noted that, although poles and zeros of the open-loop transfer function $G(s)H(s)$ may be in the right-half s plane, the system is stable if all the poles of the closed-loop transfer function (that is, the roots of the characteristic equation) are in the left-half s plane.] The Nyquist stability criterion relates the open-loop frequency response $G(j\omega)H(j\omega)$ to the number of zeros and poles of $1 + G(s)H(s)$ that lie in the right-half s plane. This criterion, derived by H. Nyquist, is useful in control engineering because the absolute stability of the closed-loop system can be determined graphically from open-loop frequency-response curves, and there is no need for actually determining the closed-loop poles. Analytically obtained open-loop frequency-response curves, as well as those experimentally obtained, can be used for the stability analysis. This is convenient because, in designing a control system, it often happens that mathematical expressions for some of the components are not known; only their frequency-response data are available.

The Nyquist stability criterion is based on a theorem from the theory of complex variables. To understand the criterion, we shall first discuss mappings of contours in the complex plane.

We shall assume that the open-loop transfer function $G(s)H(s)$ is representable as a ratio of polynomials in s. For a physically realizable system, the degree of the denominator polynomial of the closed-loop transfer function must be greater than or equal to that of the numerator polynomial. This means that the limit of $G(s)H(s)$ as s approaches infinity is zero or a constant for any physically realizable system.

Preliminary Study. The characteristic equation of the system shown in Figure 8–45 is

$$F(s) = 1 + G(s)H(s) = 0$$

We shall show that, for a given continuous closed path in the s plane that does not go through any singular points, there corresponds a closed curve in the $F(s)$ plane. The number and direction of encirclements of the origin of the $F(s)$ plane by the closed curve play a particularly important role in what follows, for later we shall correlate the number and direction of encirclements with the stability of the system.

Consider, for example, the following open-loop transfer function:

$$G(s)H(s) = \frac{2}{s-1}$$

The characteristic equation is

$$\begin{aligned} F(s) &= 1 + G(s)H(s) \\ &= 1 + \frac{2}{s-1} = \frac{s+1}{s-1} = 0 \end{aligned} \tag{8–15}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The function $F(s)$ is analytic everywhere in the s plane except at its singular points. For each point of analyticity in the s plane, there corresponds a point in the $F(s)$ plane. For example, if $s = 2 + j1$, then $F(s)$ becomes

$$F(2 + j1) = \frac{2 + j1 + 1}{2 + j1 - 1} = 2 - j1$$

Thus, point $s = 2 + j1$ in the s plane maps into point $2 - j1$ in the $F(s)$ plane.

Thus, as stated previously, for a given continuous closed path in the s plane, which does not go through any singular points, there corresponds a closed curve in the $F(s)$ plane.

For the characteristic equation $F(s)$ given by Equation (8–15), the conformal mapping of the lines $\omega = 0, \pm 1, \pm 2$ and the lines $\sigma = 0, \pm 1, \pm 2$ [see Figure 8–46(a)] yield circles in the $F(s)$ plane, as shown in Figure 8–46(b). Suppose that representative point s traces out a contour in the s plane in the clockwise direction. If the contour in the s plane encloses the pole of $F(s)$, there is one encirclement of the origin of the $F(s)$ plane by the locus of $F(s)$ in the counterclockwise direction. [See Figure 8–47(a).] If the contour in the s plane encloses the zero of $F(s)$, there is one encirclement of the origin of the $F(s)$ plane by the locus of $F(s)$ in the clockwise direction. [See Figure 8–47(b).] If the contour in the s plane encloses both the zero and the pole or if the contour encloses neither the zero nor the pole, then there is no encirclement of the origin of the $F(s)$ plane by the locus of $F(s)$. [See Figures 8–47(c) and (d).]

From the foregoing analysis, we can say that the direction of encirclement of the origin of the $F(s)$ plane by the locus of $F(s)$ depends on whether the contour in the s plane encloses a pole or a zero. Note that the location of a pole or zero in the s plane, whether in the right-half or left-half s plane, does not make any difference, but the enclosure of a pole or zero does. If the contour in the s plane encloses equal numbers of poles and zeros, then the corresponding closed curve in the $F(s)$ plane does not encircle the origin of the $F(s)$ plane. The foregoing discussion is a graphical explanation of the mapping theorem, which is the basis for the Nyquist stability criterion.

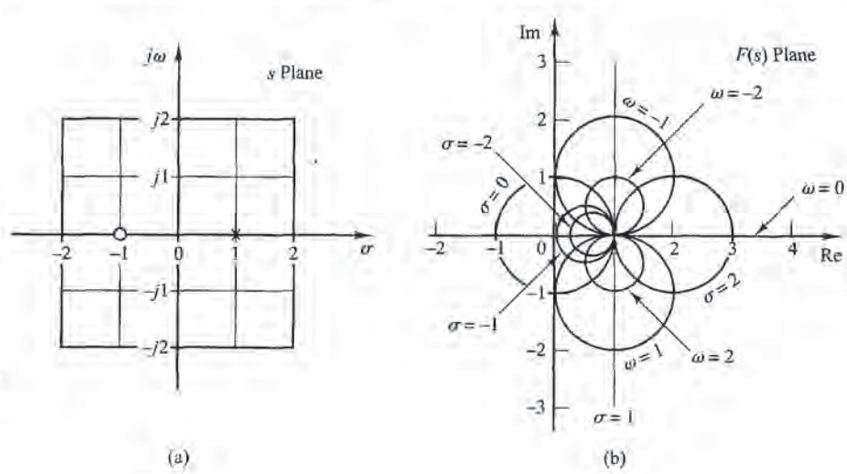


Figure 8–46
Conformal mapping of the s-plane grids into the $F(s)$ plane, where $F(s) = (s + 1)/(s - 1)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

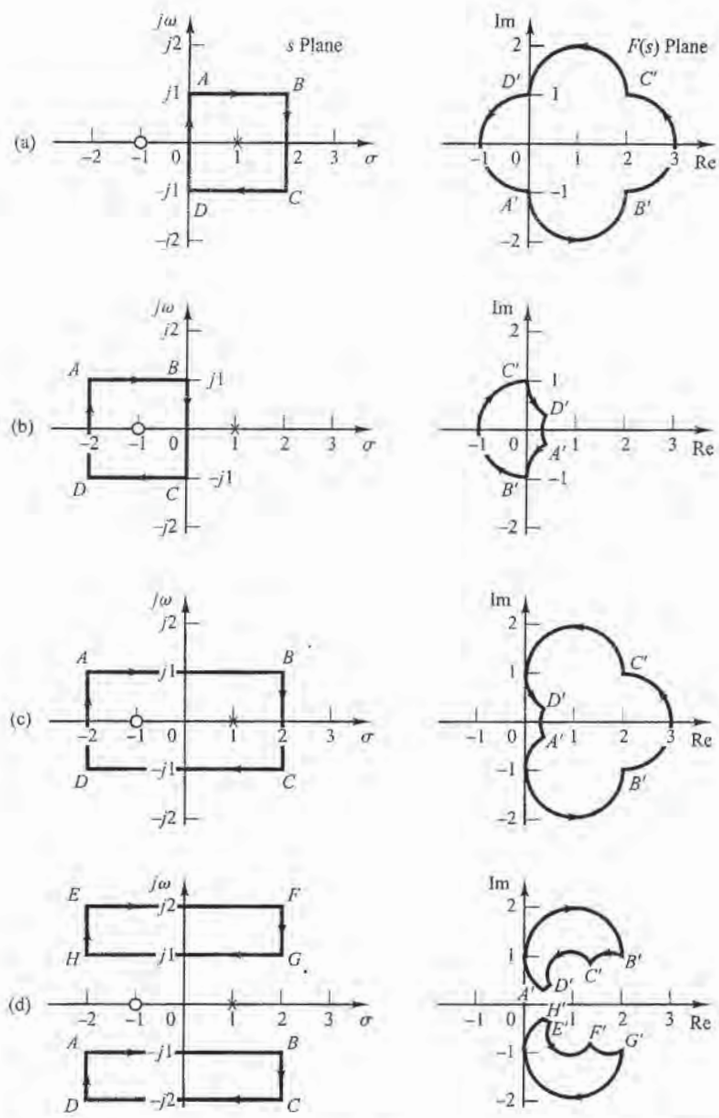


Figure 8–47
Closed contours in the s plane and their corresponding closed curves in the $F(s)$ plane, where $F(s) = (s + 1)/(s - 1)$.

Mapping Theorem. Let $F(s)$ be a ratio of two polynomials in s. Let P be the number of poles and Z be the number of zeros of $F(s)$ that lie inside some closed contour in the s plane, with multiplicity of poles and zeros accounted for. Let the contour be such that it does not pass through any poles or zeros of $F(s)$. This closed contour in the s plane is then mapped into the $F(s)$ plane as a closed curve. The total number N of clockwise encirclements of the origin of the $F(s)$ plane, as a representative point s traces out the entire contour in the clockwise direction, is equal to $Z - P$. (Note that by this mapping theorem, the numbers of zeros and of poles cannot be found—only their difference.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

We shall not present a formal proof of this theorem here, but leave the proof to Problem **A–8–6**. Note that a positive number N indicates an excess of zeros over poles of the function $F(s)$ and a negative N indicates an excess of poles over zeros. In control system applications, the number P can be readily determined for $F(s) = 1 + G(s)H(s)$ from the function $G(s)H(s)$. Therefore, if N is determined from the plot of $F(s)$, the number of zeros in the closed contour in the s plane can be determined readily. Note that the exact shapes of the s-plane contour and $F(s)$ locus are immaterial so far as encirclements of the origin are concerned, since encirclements depend only on the enclosure of poles and/or zeros of $F(s)$ by the s-plane contour.

Application of the Mapping Theorem to the Stability Analysis of Closed-Loop Systems. For analyzing the stability of linear control systems, we let the closed contour in the s plane enclose the entire right-half s plane. The contour consists of the entire $j\omega$ axis from $\omega = -\infty$ to $+\infty$ and a semicircular path of infinite radius in the right-half s plane. Such a contour is called the Nyquist path. (The direction of the path is clockwise.) The Nyquist path encloses the entire right-half s plane and encloses all the zeros and poles of $1 + G(s)H(s)$ that have positive real parts. [If there are no zeros of $1 + G(s)H(s)$ in the right-half s plane, then there are no closed-loop poles there, and the system is stable.] It is necessary that the closed contour, or the Nyquist path, not pass through any zeros and poles of $1 + G(s)H(s)$. If $G(s)H(s)$ has a pole or poles at the origin of the s plane, mapping of the point $s = 0$ becomes indeterminate. In such cases, the origin is avoided by taking a detour around it. (A detailed discussion of this special case is given later.)

If the mapping theorem is applied to the special case in which $F(s)$ is equal to $1 + G(s)H(s)$, then we can make the following statement: If the closed contour in the s plane encloses the entire right-half s plane, as shown in Figure 8–48, then the number of right-half plane zeros of the function $F(s) = 1 + G(s)H(s)$ is equal to the number of poles of the function $F(s) = 1 + G(s)H(s)$ in the right-half s plane plus the number of clockwise encirclements of the origin of the $1 + G(s)H(s)$ plane by the corresponding closed curve in this latter plane.

Because of the assumed condition that

$$\lim_{s\to\infty}[1 + G(s)H(s)] = \text{constant}$$

the function of $1 + G(s)H(s)$ remains constant as s traverses the semicircle of infinite radius. Because of this, whether the locus of $1 + G(s)H(s)$ encircles the origin of the $1 + G(s)H(s)$ plane can be determined by considering only a part of the closed contour in the s plane, that is, the $j\omega$ axis. Encirclements of the origin, if there are any, occur only

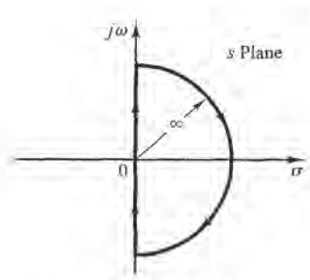


Figure 8–48
Closed contour in the s plane.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

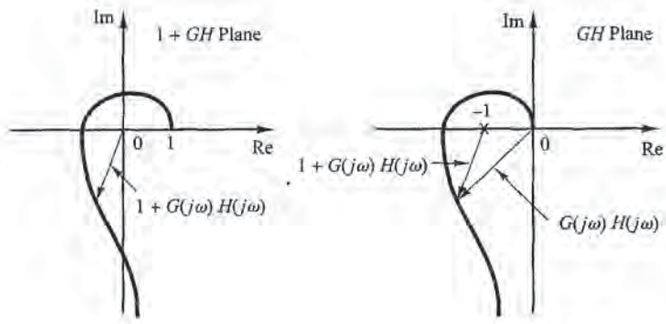


Figure 8–49
Plots of $1 + G(j\omega)H(j\omega)$ in the $1 + GH$ plane and GH plane.

while a representative point moves from $-j\infty$ to $+j\infty$ along the $j\omega$ axis, provided that no zeros or poles lie on the $j\omega$ axis.

Note that the portion of the $1 + G(s)H(s)$ contour from $\omega = -\infty$ to $\omega = \infty$ is simply $1 + G(j\omega)H(j\omega)$. Since $1 + G(j\omega)H(j\omega)$ is the vector sum of the unit vector and the vector $G(j\omega)H(j\omega)$, $1 + G(j\omega)H(j\omega)$ is identical to the vector drawn from the $-1 + j0$ point to the terminal point of the vector $G(j\omega)H(j\omega)$, as shown in Figure 8–49. Encirclement of the origin by the graph of $1 + G(j\omega)H(j\omega)$ is equivalent to encirclement of the $-1 + j0$ point by just the $G(j\omega)H(j\omega)$ locus. Thus, the stability of a closed-loop system can be investigated by examining encirclements of the $-1 + j0$ point by the locus of $G(j\omega)H(j\omega)$. The number of clockwise encirclements of the $-1 + j0$ point can be found by drawing a vector from the $-1 + j0$ point to the $G(j\omega)H(j\omega)$ locus, starting from $\omega = -\infty$, going through $\omega = 0$, and ending at $\omega = +\infty$, and by counting the number of clockwise rotations of the vector.

Plotting $G(j\omega)H(j\omega)$ for the Nyquist path is straightforward. The map of the negative $j\omega$ axis is the mirror image about the real axis of the map of the positive $j\omega$ axis. That is, the plot of $G(j\omega)H(j\omega)$ and the plot of $G(-j\omega)H(-j\omega)$ are symmetrical with each other about the real axis. The semicircle with infinite radius maps into either the origin of the GH plane or a point on the real axis of the GH plane.

In the preceding discussion, $G(s)H(s)$ has been assumed to be the ratio of two polynomials in s. Thus, the transport lag e^{-Ts} has been excluded from the discussion. Note, however, that a similar discussion applies to systems with transport lag, although a proof of this is not given here. The stability of a system with transport lag can be determined from the open-loop frequency-response curves by examining the number of encirclements of the $-1 + j0$ point, just as in the case of a system whose open-loop transfer function is a ratio of two polynomials in s.

Nyquist Stability Criterion. The foregoing analysis, utilizing the encirclement of the $-1 + j0$ point by the $G(j\omega)H(j\omega)$ locus, is summarized in the following Nyquist stability criterion:

> *Nyquist stability criterion [for a special case when $G(s)H(s)$ has neither poles nor zeros on the $j\omega$ axis.]*: In the system shown in Figure 8–45, if the open-loop transfer function $G(s)H(s)$ has k poles in the right-half s plane and $\lim_{s\to\infty} G(s)H(s) = \text{constant}$, then for stability, the $G(j\omega)H(j\omega)$ locus, as ω varies from $-\infty$ to ∞, must encircle the $-1 + j0$ point k times in the counterclockwise direction.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Remarks on the Nyquist Stability Criterion

1. This criterion can be expressed as

$$Z = N + P$$

where Z = number of zeros of $1 + G(s)H(s)$ in the right-half s plane
N = number of clockwise encirclements of the $-1 + j0$ point
P = number of poles of $G(s)H(s)$ in the right-half s plane

If P is not zero, for a stable control system, we must have $Z = 0$, or $N = -P$, which means that we must have P counterclockwise encirclements of the $-1 + j0$ point.

If $G(s)H(s)$ does not have any poles in the right-half s plane, then $Z = N$. Thus, for stability there must be no encirclement of the $-1 + j0$ point by the $G(j\omega)H(j\omega)$ locus. In this case it is not necessary to consider the locus for the entire $j\omega$ axis, only for the positive-frequency portion. The stability of such a system can be determined by seeing if the $-1 + j0$ point is enclosed by the Nyquist plot of $G(j\omega)H(j\omega)$. The region enclosed by the Nyquist plot is shown in Figure 8–50. For stability, the $-1 + j0$ point must lie outside the shaded region.

2. We must be careful when testing the stability of multiple-loop systems since they may include poles in the right-half s plane. (Note that although an inner loop may be unstable, the entire closed-loop system can be made stable by proper design.) Simple inspection of encirclements of the $-1 + j0$ point by the $G(j\omega)H(j\omega)$ locus is not sufficient to detect instability in multiple-loop systems. In such cases, however, whether any pole of $1 + G(s)H(s)$ is in the right-half s plane can be determined easily by applying the Routh stability criterion to the denominator of $G(s)H(s)$.

If transcendental functions, such as transport lag e^{-Ts}, are included in $G(s)H(s)$, they must be approximated by a series expansion before the Routh stability criterion can be applied. A few forms of series expansion of e^{-Ts} were presented in Section 6–7.

3. If the locus of $G(j\omega)H(j\omega)$ passes through the $-1 + j0$ point, then zeros of the characteristic equation, or closed-loop poles, are located on the $j\omega$ axis. This is not desirable for practical control systems. For a well-designed closed-loop system, none of the roots of the characteristic equation should lie on the $j\omega$ axis.

Special Case when $G(s)H(s)$ Involves Poles and/or Zeros on the $j\omega$ Axis. In the previous discussion, we assumed that the open-loop transfer function $G(s)H(s)$ has

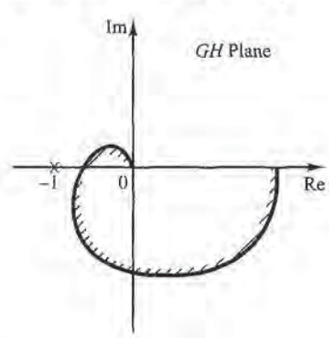


Figure 8–50
Region enclosed by a Nyquist plot.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

neither poles nor zeros at the origin. We now consider the case where $G(s)H(s)$ involves poles and/or zeros on the $j\omega$ axis.

Since the Nyquist path must not pass through poles or zeros of $G(s)H(s)$, if the function $G(s)H(s)$ has poles or zeros at the origin (or on the $j\omega$ axis at points other than the origin), the contour in the s plane must be modified. The usual way of modifying the contour near the origin is to use a semicircle with the infinitesimal radius ε, as shown in Figure 8–51. [Note that this semicircle may lie in the right-half s plane or in the left-half s plane. Here we take the semicircle in the right-half s plane.] A representative point s moves along the negative $j\omega$ axis from $-j\infty$ to $j0-$. From $s = j0-$ to $s = j0+$, the point moves along the semicircle of radius ε (where $\varepsilon \ll 1$) and then moves along the positive $j\omega$ axis from $j0+$ to $j\infty$. From $s = j\infty$, the contour follows a semicircle with infinite radius, and the representative point moves back to the starting point, $s = -j\infty$. The area that the modified closed contour avoids is very small and approaches zero as the radius ε approaches zero. Therefore, all the poles and zeros, if any, in the right-half s plane are enclosed by this contour.

Consider, for example, a closed-loop system whose open-loop transfer function is given by

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

The points corresponding to $s = j0+$ and $s = j0-$ on the locus of $G(s)H(s)$ in the $G(s)H(s)$ plane are $-j\infty$ and $j\infty$, respectively. On the semicircular path with radius ε (where $\varepsilon \ll 1$), the complex variable s can be written

$$s = \varepsilon e^{j\theta}$$

where θ varies from $-90°$ to $+90°$. Then $G(s)H(s)$ becomes

$$G(\varepsilon e^{j\theta})H(\varepsilon e^{j\theta}) = \frac{K}{\varepsilon e^{j\theta}} = \frac{K}{\varepsilon} e^{-j\theta}$$

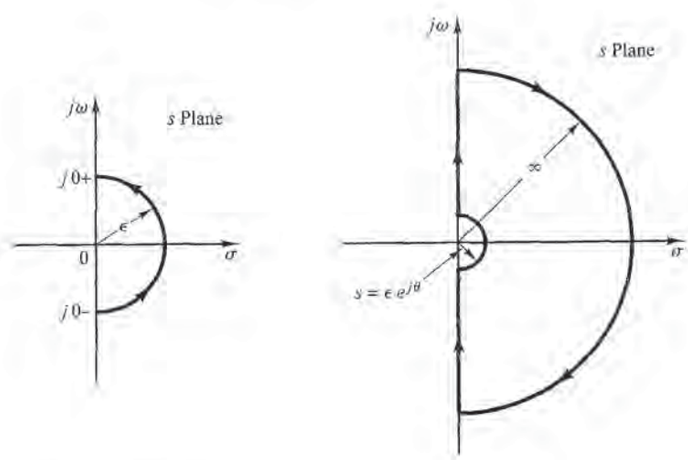


Figure 8–51
Contour near the origin of the s plane and closed contour in the s plane avoiding poles and zeros at the origin.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The value K/ε approaches infinity as ε approaches zero, and $-\theta$ varies from 90° to −90° as a representative point s moves along the semicircle in the s plane. Thus, the points $G(j0-)H(j0-) = j\infty$ and $G(j0+)H(j0+) = -j\infty$ are joined by a semicircle of infinite radius in the right-half GH plane. The infinitesimal semicircular detour around the origin in the s plane maps into the GH plane as a semicircle of infinite radius. Figure 8–52 shows the s-plane contour and the $G(s)H(s)$ locus in the GH plane. Points A, B, and C on the s-plane contour map into the respective points A', B', and C' on the $G(s)H(s)$ locus. As seen from Figure 8–52, points D, E, and F on the semicircle of infinite radius in the s plane map into the origin of the GH plane. Since there is no pole in the right-half s plane and the $G(s)H(s)$ locus does not encircle the $-1 + j0$ point, there are no zeros of the function $1 + G(s)H(s)$ in the right-half s plane. Therefore, the system is stable.

For an open-loop transfer function $G(s)H(s)$ involving a $1/s^n$ factor (where $n = 2, 3, \ldots$), the plot of $G(s)H(s)$ has n clockwise semicircles of infinite radius about the origin as a representative point s moves along the semicircle of radius ε (where $\varepsilon \ll 1$). For example, consider the following open-loop transfer function:

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

Then

$$\lim_{s \to \varepsilon e^{j\theta}} G(s)H(s) = \frac{K}{\varepsilon^2 e^{2j\theta}} = \frac{K}{\varepsilon^2} e^{-2j\theta}$$

As θ varies from −90° to 90° in the s plane, the angle of $G(s)H(s)$ varies from 180° to −180°, as shown in Figure 8–53. Since there is no pole in the right-half s plane and the locus encircles the $-1 + j0$ point twice clockwise for any positive value of K, there are two zeros of $1 + G(s)H(s)$ in the right-half s plane. Therefore, this system is always unstable.

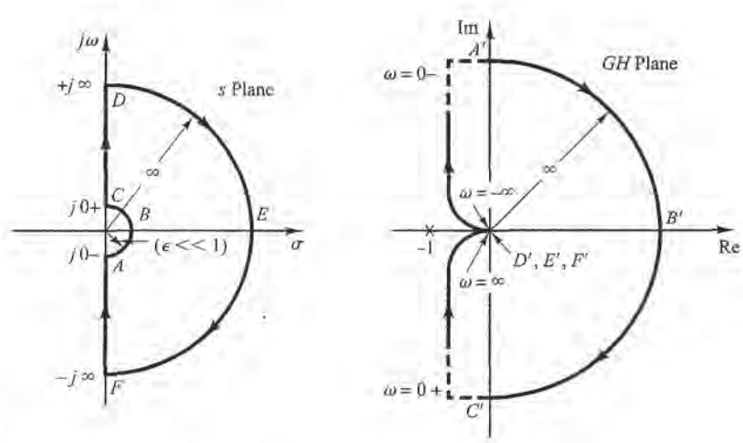


Figure 8–52
s-Plane contour and the $G(s)H(s)$ locus in the GH plane, where $G(s)H(s) = K/[s(Ts + 1)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

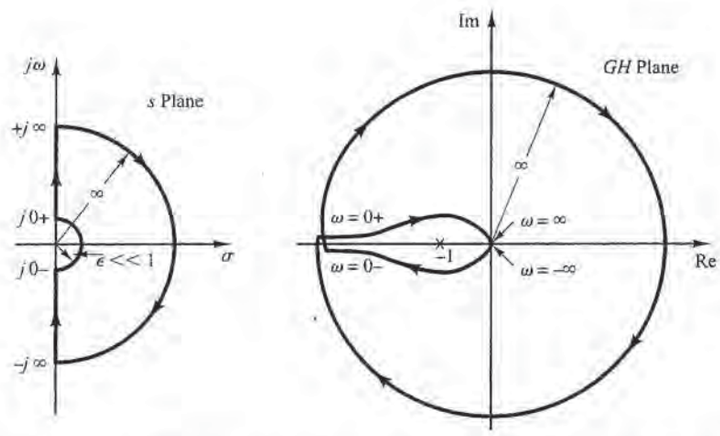


Figure 8–53
s-Plane contour and the $G(s)H(s)$ locus in the GH plane, where $G(s)H(s) = K/[s^2(Ts + 1)]$.

Note that a similar analysis can be made if $G(s)H(s)$ involves poles and/or zeros on the $j\omega$ axis. The Nyquist stability criterion can now be generalized as follows:

> *Nyquist stability criterion [for a general case when $G(s)H(s)$ has poles and/or zeros on the $j\omega$ axis.]*: In the system shown in Figure 8–45, if the open-loop transfer function $G(s)H(s)$ has k poles in the right-half s plane, then for stability the $G(s)H(s)$ locus, as a representative point s traces on the modified Nyquist path in the clockwise direction, must encircle the $-1 + j0$ point k times in the counterclockwise direction.

8–8 STABILITY ANALYSIS

In this section, we shall present several illustrative examples of the stability analysis of control systems using the Nyquist stability criterion.

If the Nyquist path in the s plane encircles Z zeros and P poles of $1 + G(s)H(s)$ and does not pass through any poles or zeros of $1 + G(s)H(s)$ as a representative point s moves in the clockwise direction along the Nyquist path, then the corresponding contour in the $G(s)H(s)$ plane encircles the $-1 + j0$ point $N = Z - P$ times in the clockwise direction. (Negative values of N imply counterclockwise encirclements.)

In examining the stability of linear control systems using the Nyquist stability criterion, we see that three possibilities can occur:

1. There is no encirclement of the $-1 + j0$ point. This implies that the system is stable if there are no poles of $G(s)H(s)$ in the right-half s plane; otherwise, the system is unstable.
2. There are one or more counterclockwise encirclements of the $-1 + j0$ point. In this case the system is stable if the number of counterclockwise encirclements is the same as the number of poles of $G(s)H(s)$ in the right-half s plane; otherwise, the system is unstable.
3. There are one or more clockwise encirclements of the $-1 + j0$ point. In this case the system is unstable.

In the following examples, we assume that the values of the gain K and the time constants (such as T, T_1, and T_2) are all positive.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 8–14 Consider a closed-loop system whose open-loop transfer function is given by

$$G(s)H(s) = \frac{K}{(T_1 s + 1)(T_2 s + 1)}$$

Examine the stability of the system.

A plot of $G(j\omega)H(j\omega)$ is shown in Figure 8–54. Since $G(s)H(s)$ does not have any poles in the right-half s plane and the $-1 + j0$ point is not encircled by the $G(j\omega)H(j\omega)$ locus, this system is stable for any positive values of K, T_1, and T_2.

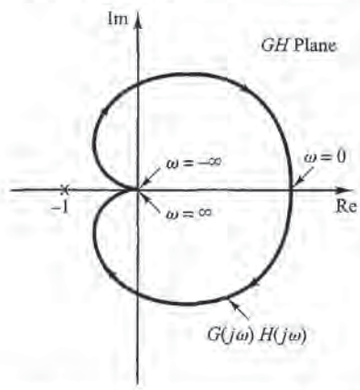


Figure 8–54 Polar plot of $G(j\omega)H(j\omega)$ considered in Example 8–14.

EXAMPLE 8–15 Consider the system with the following open-loop transfer function:

$$G(s)H(s) = \frac{K}{s(T_1 s + 1)(T_2 s + 1)}$$

Determine the stability of the system for two cases: (1) the gain K is small and (2) K is large.

The Nyquist plots of the open-loop transfer function with a small value of K and a large value of K are shown in Figure 8–55. The number of poles of $G(s)H(s)$ in the right-half s plane is zero. Therefore, for this system to be stable, it is necessary that $N = Z = 0$ or that the $G(s)H(s)$ locus not encircle the $-1 + j0$ point.

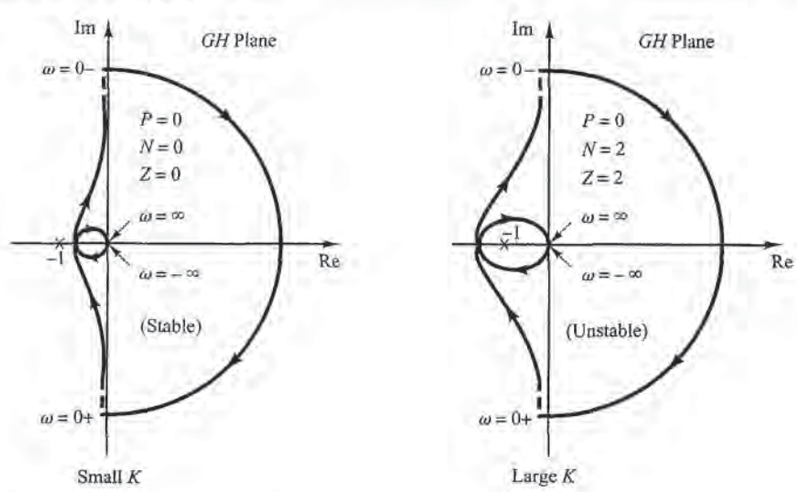


Figure 8–55 Polar plots of the system considered in Example 8–15.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

For small values of K, there is no encirclement of the $-1 + j0$ point. Hence, the system is stable for small values of K. For large values of K, the locus of $G(s)H(s)$ encircles the $-1 + j0$ point twice in the clockwise direction, indicating two closed-loop poles in the right-half s plane, and the system is unstable. (For good accuracy, K should be large. From the stability viewpoint, however, a large value of K causes poor stability or even instability. To compromise between accuracy and stability, it is necessary to insert a compensation network into the system. Compensating techniques in the frequency domain are discussed in Chapter 9.)

EXAMPLE 8–16 The stability of a closed-loop system with the following open-loop transfer function

$$G(s)H(s) = \frac{K(T_2 s + 1)}{s^2(T_1 s + 1)}$$

depends on the relative magnitudes of T_1 and T_2. Draw Nyquist plots and determine the stability of the system.

Plots of the locus $G(s)H(s)$ for three cases, $T_1 < T_2$, $T_1 = T_2$, and $T_1 > T_2$, are shown in Figure 8–56. For $T_1 < T_2$, the locus of $G(s)H(s)$ does not encircle the $-1 + j0$ point, and the closed-loop system is stable. For $T_1 = T_2$, the $G(s)H(s)$ locus passes through the $-1 + j0$ point, which indicates that the closed-loop poles are located on the $j\omega$ axis. For $T_1 > T_2$, the locus of $G(s)H(s)$ encircles the $-1 + j0$ point twice in the clockwise direction. Thus, the closed-loop system has two closed-loop poles in the right-half s plane, and the system is unstable.

EXAMPLE 8–17 Consider the closed-loop system having the following open-loop transfer function:

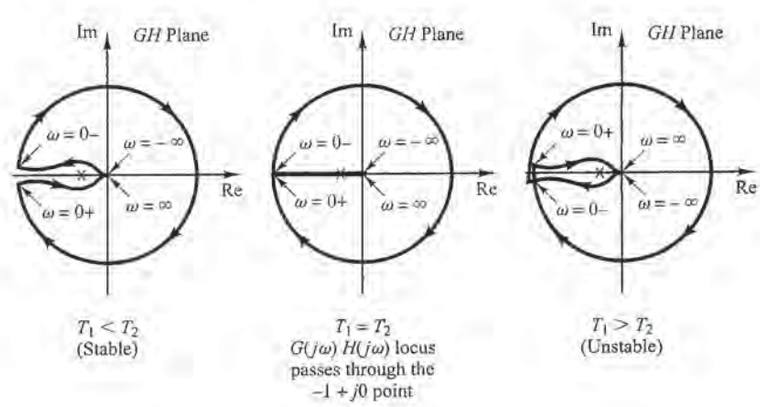
$$G(s)H(s) = \frac{K}{s(Ts - 1)}$$

Determine the stability of the system.

The function $G(s)H(s)$ has one pole ($s = 1/T$) in the right-half s plane. Therefore, $P = 1$. The Nyquist plot shown in Figure 8–57 indicates that the $G(s)H(s)$ plot encircles the $-1 + j0$ point once clockwise. Thus, $N = 1$. Since $Z = N + P$, we find that $Z = 2$. This means that the closed-loop system has two closed-loop poles in the right-half s plane and is unstable.

EXAMPLE 8–18 Investigate the stability of a closed-loop system with the following open-loop transfer function:

$$G(s)H(s) = \frac{K(s + 3)}{s(s - 1)} \qquad (K > 1)$$



$T_1 < T_2$ (Stable)

$T_1 = T_2$ $G(j\omega)H(j\omega)$ locus passes through the $-1 + j0$ point

$T_1 > T_2$ (Unstable)

Figure 8–56 Polar plots of the system considered in Example 8–16.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

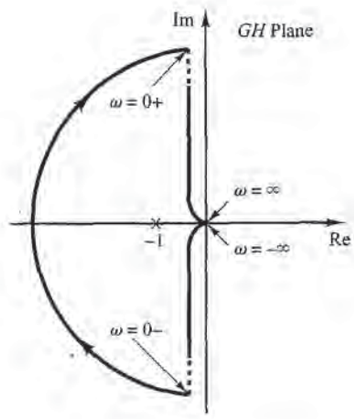


Figure 8–57
Polar plot of the system considered in Example 8–17.

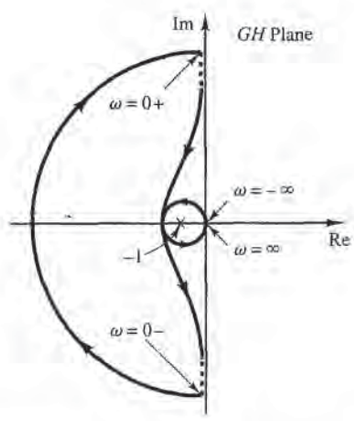


Figure 8–58
Polar plot of the system considered in Example 8–18.

The open-loop transfer function has one pole ($s = 1$) in the right-half s plane, or $P = 1$. The open-loop system is unstable. The Nyquist plot shown in Figure 8–58 indicates that the $-1 + j0$ point is encircled by the $G(s)H(s)$ locus once in the counterclockwise direction. Therefore, $N = -1$. Thus, Z is found from $Z = N + P$ to be zero, which indicates that there is no zero of $1 + G(s)H(s)$ in the right-half s plane, and the closed-loop system is stable. This is one of the examples for which an unstable open-loop system becomes stable when the loop is closed.

Conditionally Stable Systems. Figure 8–59 shows an example of a $G(j\omega)H(j\omega)$ locus for which the closed-loop system can be made unstable by varying the open-loop gain. If the open-loop gain is increased sufficiently, the $G(j\omega)H(j\omega)$ locus encloses the $-1 + j0$ point twice, and the system becomes unstable. If the open-loop gain is decreased sufficiently, again the $G(j\omega)H(j\omega)$ locus encloses the $-1 + j0$ point twice. For stable operation of the system considered here, the critical point $-1 + j0$ must not be located in the regions between OA and BC shown in Figure 8–59. Such a system that is stable only for limited ranges of values of the open-loop gain for which the $-1 + j0$ point is completely outside the $G(j\omega)H(j\omega)$ locus is a conditionally stable system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

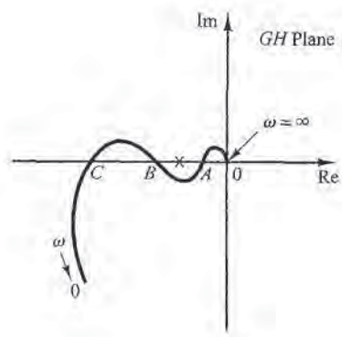


Figure 8–59
Polar plot of a conditionally stable system.

A conditionally stable system is stable for the value of the open-loop gain lying between critical values, but it is unstable if the open-loop gain is either increased or decreased sufficiently. Such a system becomes unstable when large input signals are applied, since a large signal may cause saturation, which in turn reduces the open-loop gain of the system. It is advisable to avoid such a situation.

Multiple-Loop System. Consider the system shown in Figure 8–60. This is a multiple-loop system. The inner loop has the transfer function

$$G(s) = \frac{G_2(s)}{1 + G_2(s)H_2(s)}$$

If $G(s)$ is unstable, the effects of instability are to produce a pole or poles in the right-half s plane. Then the characteristic equation of the inner loop, $1 + G_2(s)H_2(s) = 0$, has a zero or zeros in the right-half s plane. If $G_2(s)$ and $H_2(s)$ have P_1 poles here, then the number Z_1 of right-half plane zeros of $1 + G_2(s)H_2(s)$ can be found from $Z_1 = N_1 + P_1$, where N_1 is the number of clockwise encirclements of the $-1 + j0$ point by the $G_2(s)H_2(s)$ locus. Since the open-loop transfer function of the entire system is given by $G_1(s)G(s)H_1(s)$, the stability of this closed-loop system can be found from the Nyquist plot of $G_1(s)G(s)H_1(s)$ and knowledge of the right-half plane poles of $G_1(s)G(s)H_1(s)$.

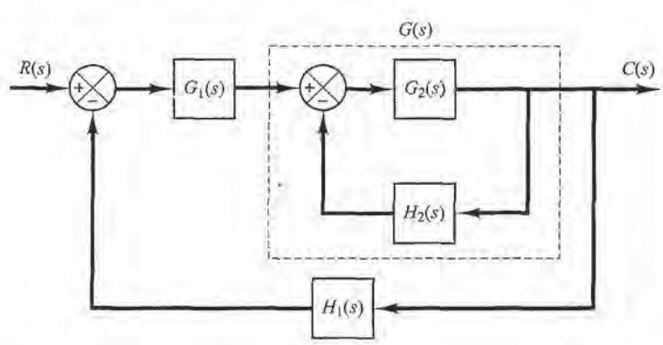


Figure 8–60
Multiple-loop system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Notice that if a feedback loop is eliminated by means of block diagram reductions, there is a possibility that unstable poles are introduced; if the feedforward branch is eliminated by means of block diagram reductions, there is a possibility that right-half plane zeros are introduced. Therefore, we must note all right-half plane poles and zeros as they appear from subsidiary loop reductions. This knowledge is necessary in determining the stability of multiple-loop systems.

EXAMPLE 8–19 Consider the control system shown in Figure 8–61. The system involves two loops. Determine the range of gain K for stability of the system by the use of the Nyquist stability criterion. (The gain K is positive.)

To examine the stability of the control system, we need to sketch the Nyquist locus of $G(s)$, where

$$G(s) = G_1(s)G_2(s)$$

However, the poles of $G(s)$ are not known at this point. Therefore, we need to examine the minor loop if there are right-half s-plane poles. This can be done easily by use of the Routh stability criterion. Since

$$G_2(s) = \frac{1}{s^3 + s^2 + 1}$$

the Routh array becomes as follows:

$$\begin{array}{ccc} s^3 & 1 & 0 \\ s^2 & 1 & 1 \\ s^1 & -1 & 0 \\ s^0 & 1 & \end{array}$$

Notice that there are two sign changes in the first column. Hence, there are two poles of $G_2(s)$ in the right-half s plane.

Once we find the number of right-half s plane poles of $G_2(s)$, we proceed to sketch the Nyquist locus of $G(s)$, where

$$G(s) = G_1(s)G_2(s) = \frac{K(s + 0.5)}{s^3 + s^2 + 1}$$

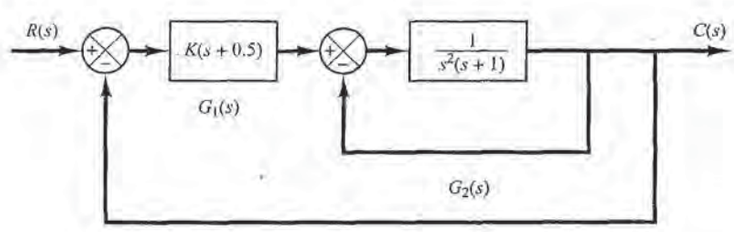


Figure 8–61
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Our problem is to determine the range of the gain K for stability. Hence, instead of plotting Nyquist loci of $G(j\omega)$ for various values of K, we plot the Nyquist locus of $G(j\omega)/K$. Figure 8–62 shows the Nyquist plot or polar plot of $G(j\omega)/K$.

Since $G(s)$ has two poles in the right-half s plane, we have $P = 2$. Noting that

$$Z = N + P$$

for stability, we require $Z = 0$ or $N = -2$. That is, the Nyquist locus of $G(j\omega)$ must encircle the $-1 + j0$ point twice counterclockwise. From Figure 8–62, we see that, if the critical point lies between 0 and -0.5, then the $G(j\omega)/K$ locus encircles the critical point twice counterclockwise. Therefore, we require

$$-0.5K < -1$$

The range of the gain K for stability is

$$2 < K$$

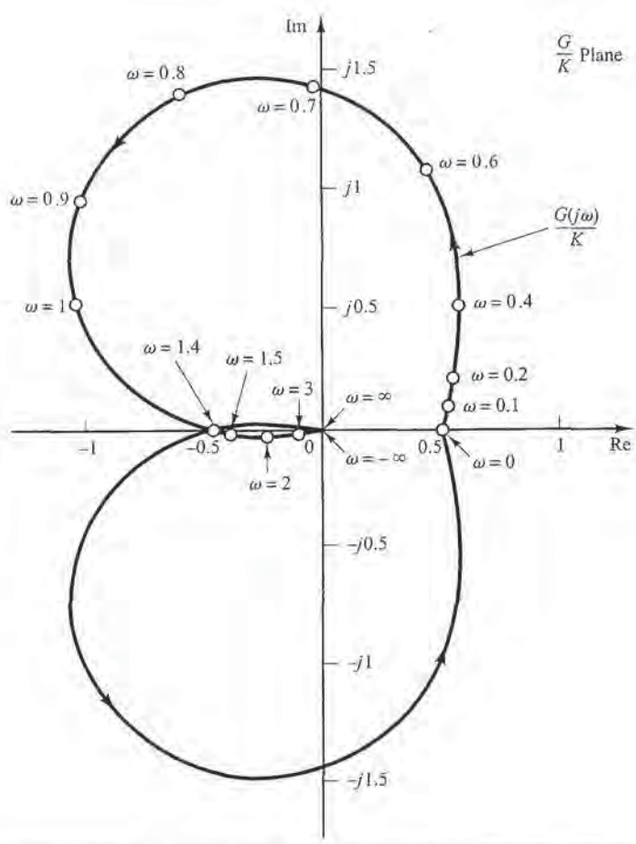


Figure 8–62
Polar plot of $G(j\omega)/K$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Nyquist Stability Criterion Applied to Inverse Polar Plots. In the previous analyses, the Nyquist stability criterion was applied to polar plots of the open-loop transfer function $G(s)H(s)$.

In analyzing multiple-loop systems, the inverse transfer function may sometimes be used in order to permit graphical analysis; this avoids much of the numerical calculation. (The Nyquist stability criterion can be applied equally well to inverse polar plots. The mathematical derivation of the Nyquist stability criterion for inverse polar plots is the same as that for direct polar plots.)

The inverse polar plot of $G(j\omega)H(j\omega)$ is a graph of $1/[G(j\omega)H(j\omega)]$ as a function of ω. For example, if $G(j\omega)H(j\omega)$ is

$$G(j\omega)H(j\omega) = \frac{j\omega T}{1 + j\omega T}$$

then

$$\frac{1}{G(j\omega)H(j\omega)} = \frac{1}{j\omega T} + 1$$

The inverse polar plot for $\omega \geq 0$ is the lower half of the vertical line starting at the point $(1, 0)$ on the real axis.

The Nyquist stability criterion applied to inverse plots may be stated as follows: For a closed-loop system to be stable, the encirclement, if any, of the $-1 + j0$ point by the $1/[G(s)H(s)]$ locus (as s moves along the Nyquist path) must be counterclockwise, and the number of such encirclements must be equal to the number of poles of $1/[G(s)H(s)]$ [that is, the zeros of $G(s)H(s)$] that lie in the right-half s plane. [The number of zeros of $G(s)H(s)$ in the right-half s plane may be determined by the use of the Routh stability criterion.] If the open-loop transfer function $G(s)H(s)$ has no zeros in the right-half s plane, then for a closed-loop system to be stable, the number of encirclements of the $-1 + j0$ point by the $1/[G(s)H(s)]$ locus must be zero.

Note that although the Nyquist stability criterion can be applied to inverse polar plots, if experimental frequency-response data are incorporated, counting the number of encirclements of the $1/[G(s)H(s)]$ locus may be difficult because the phase shift corresponding to the infinite semicircular path in the s plane is difficult to measure. For example, if the open-loop transfer function $G(s)H(s)$ involves transport lag such that

$$G(s)H(s) = \frac{Ke^{-j\omega L}}{s(Ts + 1)}$$

then the number of encirclements of the $-1 + j0$ point by the $1/[G(s)H(s)]$ locus becomes infinite, and the Nyquist stability criterion cannot be applied to the inverse polar plot of such an open-loop transfer function.

In general, if experimental frequency-response data cannot be put into analytical form, both the $G(j\omega)H(j\omega)$ and $1/[G(j\omega)H(j\omega)]$ loci must be plotted. In addition, the number of right-half plane zeros of $G(s)H(s)$ must be determined. It is more difficult to determine the right-half plane zeros of $G(s)H(s)$ (in other words, to determine whether a given component is minimum phase) than it is to determine the right-half plane poles of $G(s)H(s)$ (in other words, to determine whether the component is stable).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Depending on whether the data are graphical or analytical and whether nonminimum-phase components are included, an appropriate stability test must be used for multiple-loop systems. If the data are given in analytical form or if mathematical expressions for all the components are known, the application of the Nyquist stability criterion to inverse polar plots causes no difficulty, and multiple-loop systems may be analyzed and designed in the inverse GH plane.

EXAMPLE 8–20 Consider the control system shown in Figure 8–61. (Refer to Example 8–19.) Using the inverse polar plot, determine the range of gain K for stability.

Since

$$G_2(s) = \frac{1}{s^3 + s^2 + 1}$$

we have

$$G(s) = G_1(s)G_2(s) = \frac{K(s + 0.5)}{s^3 + s^2 + 1}$$

Hence

$$\frac{1}{G(s)} = \frac{s^3 + s^2 + 1}{K(s + 0.5)}$$

Notice that $1/G(s)$ has a pole at $s = -0.5$. It does not have any pole in the right-half s plane. Therefore, the Nyquist stability equation

$$Z = N + P$$

reduces to $Z = N$ since $P = 0$. The reduced equation states that the number Z of the zeros of $1 + [1/G(s)]$ in the right-half s plane is equal to N, the number of clockwise encirclements of the $-1 + j0$ point. For stability, N must be equal to zero, or there should be no encirclement. Figure 8–63 shows the Nyquist plot or polar plot of $K/G(j\omega)$.

Notice that since

$$\frac{K}{G(j\omega)} = \left[\frac{(j\omega)^3 + (j\omega)^2 + 1}{j\omega + 0.5}\right]\left(\frac{0.5 - j\omega}{0.5 - j\omega}\right)$$
$$= \frac{0.5 - 0.5\omega^2 - \omega^4 + j\omega(-1 + 0.5\omega^2)}{0.25 + \omega^2}$$

the $K/G(j\omega)$ locus crosses the negative real axis at $\omega = \sqrt{2}$, and the crossing point at the negative real axis is -2.

From Figure 8–63, we see that if the critical point lies in the region between -2 and $-\infty$, then the critical point is not encircled. Hence, for stability, we require

$$-1 < \frac{-2}{K}$$

Thus, the range of gain K for stability is

$$2 < K$$

which is the same result as we obtained in Example 8–19.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

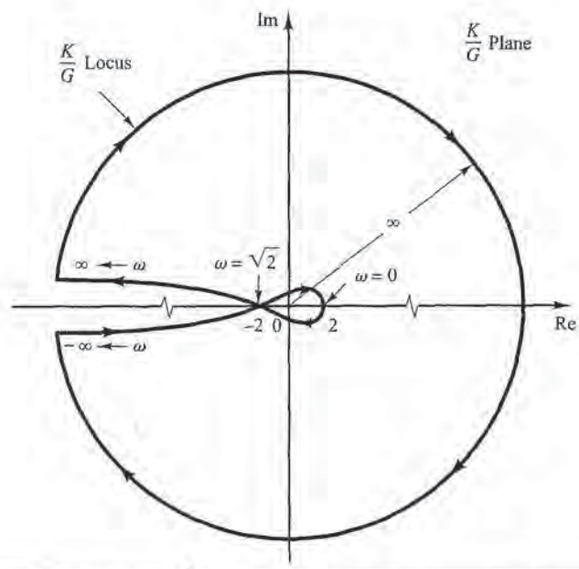


Figure 8–63
Polar plot of $K/G(j\omega)$.

Relative Stability Analysis through Modified Nyquist Plots. The Nyquist path for stability tests can be modified in order that we may investigate the relative stability of closed-loop systems.

Suppose that the s-plane contour consists of a line to the left of and parallel to the $j\omega$ axis at a distance $-\sigma_0$ (or the line $s = -\sigma_0 + j\omega$) and the semicircle of infinite radius enclosing the entire right-half s plane and that part of the left-half s plane between the lines $s = -\sigma_0 + j\omega$ and $s = j\omega$, as shown in Figure 8–64(a). If the $G(s)H(s)$ locus corresponding to this s-plane contour does not encircle the $-1 + j0$ point and $G(s)H(s)$ has no poles within the enclosed s-plane contour, then the characteristic equation does not have any zeros in the region enclosed by the modified s-plane contour. All roots of the characteristic equation lie to the left of the line $s = -\sigma_0 + j\omega$. An example of a

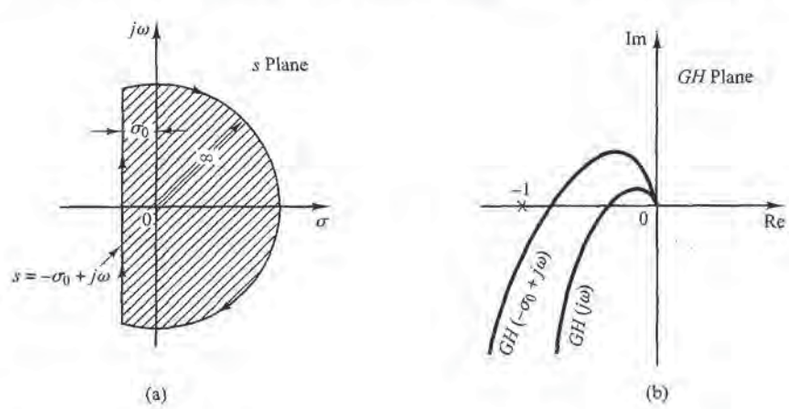


Figure 8–64
(a) Modified Nyquist path; (b) polar plots of $G(-\sigma_0 + j\omega)H(-\sigma_0 + j\omega)$ locus and $G(j\omega)H(j\omega)$ locus in the GH plane.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$G(-\sigma_0 + j\omega)H(-\sigma_0 + j\omega)$ locus, together with a $G(j\omega)H(j\omega)$ locus, is shown in Figure 8–64(b). The magnitude $1/\sigma_0$ is indicative of the time constant of the dominant closed-loop poles. If all roots lie outside the s-plane contour, all time constants of the closed-loop transfer function are less than $1/\sigma_0$.

8–9 RELATIVE STABILITY

In designing a control system, we require that the system be stable. Furthermore, it is necessary that the system have adequate relative stability.

In this section, we shall show that the Nyquist plot indicates not only whether a system is stable, but also the degree of stability of a stable system. The Nyquist plot also gives information as to how stability may be improved, if this is necessary. (See Chapter 9.)

In the following discussion, we shall assume that the systems considered have unity feedback. Note that it is always possible to reduce a system with feedback elements to a unity-feedback system, as shown in Figure 8–65. Hence, the extension of relative stability analysis for the unity-feedback system to nonunity-feedback systems is possible.

We shall also assume that, unless otherwise stated, the systems are minimum-phase systems; that is, the open-loop transfer function has neither poles nor zeros in the right-half s plane.

Relative Stability Analysis by Conformal Mapping. One of the important problems in analyzing a control system is to find all closed-loop poles or at least those closest to the $j\omega$ axis (or the dominant pair of closed-loop poles). If the open-loop frequency-response characteristics of a system are known, it may be possible to estimate the closed-loop poles closest to the $j\omega$ axis. It is noted that the Nyquist locus $G(j\omega)$ need not be an analytically known function of ω. The entire Nyquist locus may be experimentally obtained. The technique to be presented here is essentially graphical and is based on a conformal mapping of the s plane into the $G(s)$ plane.

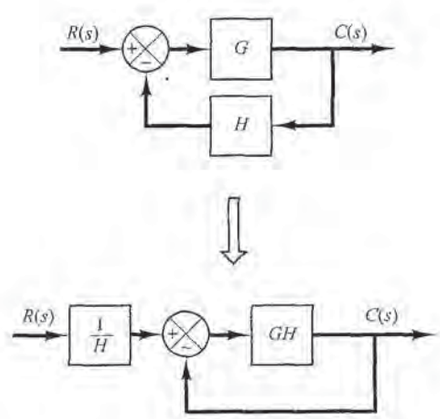


Figure 8–65
Modification of a system with feedback elements to a unity-feedback system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

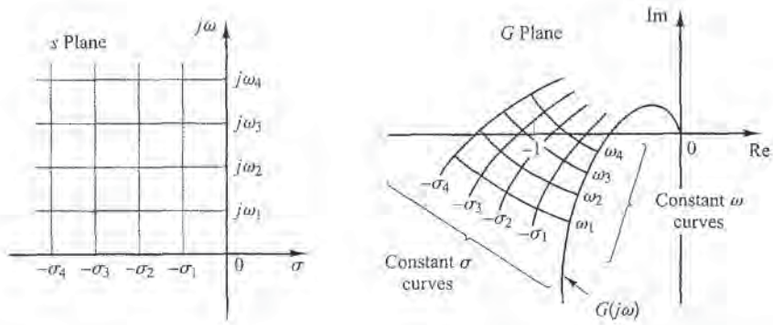


Figure 8–66
Conformal mapping of s-plane grids into the $G(s)$ plane.

Consider the conformal mapping of constant-σ lines (lines $s = \sigma + j\omega$, where σ is constant and ω varies) and constant-ω lines (lines $s = \sigma + j\omega$, where ω is constant and σ varies) in the s plane. The $\sigma = 0$ line (the $j\omega$ axis) in the s plane maps into the Nyquist plot in the $G(s)$ plane. The constant-σ lines in the s plane map into curves that are similar to the Nyquist plot and are in a sense parallel to the Nyquist plot, as shown in Figure 8–66. The constant-ω lines in the s plane map into curves, also shown in Figure 8–66.

Although the shapes of constant-σ and constant-ω loci in the $G(s)$ plane and the closeness of approach of the $G(j\omega)$ locus to the $-1 + j0$ point depend on a particular $G(s)$, the closeness of approach of the $G(j\omega)$ locus to the $-1 + j0$ point is an indication of the relative stability of a stable system. In general, we may expect that the closer the $G(j\omega)$ locus is to the $-1 + j0$ point, the larger the maximum overshoot is in the step transient response and the longer it takes to damp out.

Consider the two systems shown in Figures 8–67(a) and (b). (In Figure 8–67, the ×'s indicate closed-loop poles.) System (a) is obviously more stable than system (b) because the closed-loop poles of system (a) are located farther left than those of system (b). Figures 8–68(a) and (b) show the conformal mapping of s-plane grids into the $G(s)$ plane. The closer the closed-loop poles are located to the $j\omega$ axis, the closer the $G(j\omega)$ locus is to the $-1 + j0$ point.

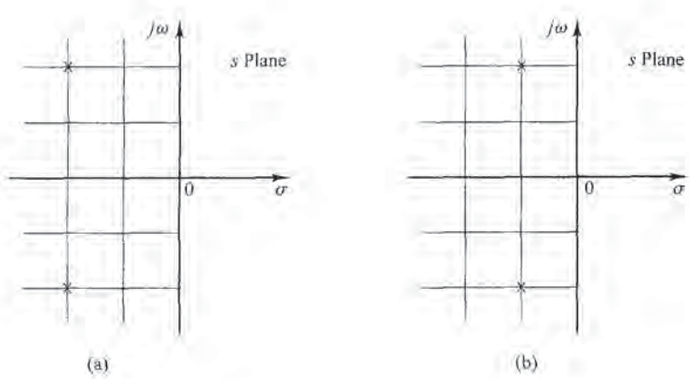


Figure 8–67
Two systems with two closed-loop poles each.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

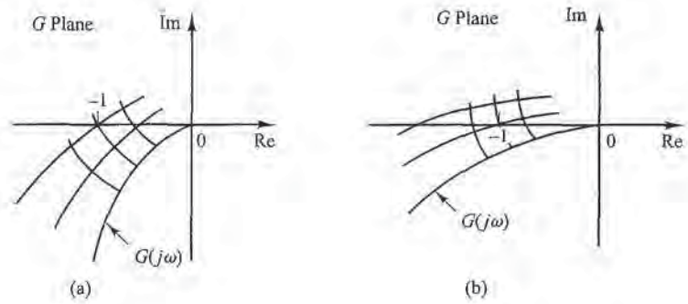


Figure 8–68
Conformal mappings of s-plane grids for the systems shown in Figure 8–67 into the $G(s)$ plane.

Phase and Gain Margins. Figure 8–69 shows the polar plots of $G(j\omega)$ for three different values of the open-loop gain K. For a large value of the gain K, the system is unstable. As the gain is decreased to a certain value, the $G(j\omega)$ locus passes through the $-1 + j0$ point. This means that with this gain value the system is on the verge of instability, and the system will exhibit sustained oscillations. For a small value of the gain K, the system is stable.

In general, the closer the $G(j\omega)$ locus comes to encircling the $-1 + j0$ point, the more oscillatory is the system response. The closeness of the $G(j\omega)$ locus to the $-1 + j0$ point can be used as a measure of the margin of stability. (This does not apply, however, to conditionally stable systems.) It is common practice to represent the closeness in terms of phase margin and gain margin.

Phase margin. The phase margin is that amount of additional phase lag at the gain crossover frequency required to bring the system to the verge of instability. The gain crossover frequency is the frequency at which $|G(j\omega)|$, the magnitude of the open-loop transfer function, is unity. The phase margin γ is 180° plus the phase angle ϕ of the open-loop transfer function at the gain crossover frequency, or

$$\gamma = 180° + \phi$$

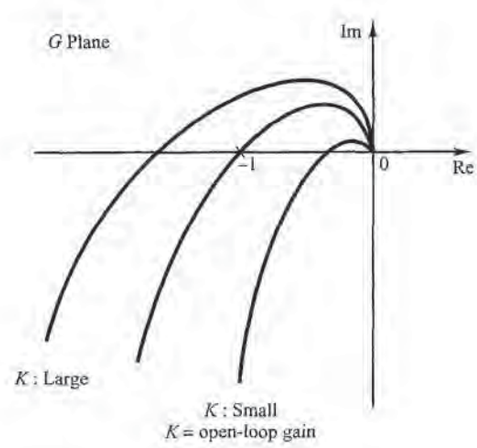


Figure 8–69
Polar plots of $\dfrac{K(1 + j\omega T_a)(1 + j\omega T_b)\cdots}{(j\omega)(1 + j\omega T_1)(1 + j\omega T_2)\cdots}$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Figures 8–70(a), (b), and (c) illustrate the phase margin of both a stable system and an unstable system in Bode diagrams, polar plots, and log-magnitude-versus-phase plots. In the polar plot, a line may be drawn from the origin to the point at which the unit circle crosses the $G(j\omega)$ locus. The angle from the negative real axis to this line is the phase

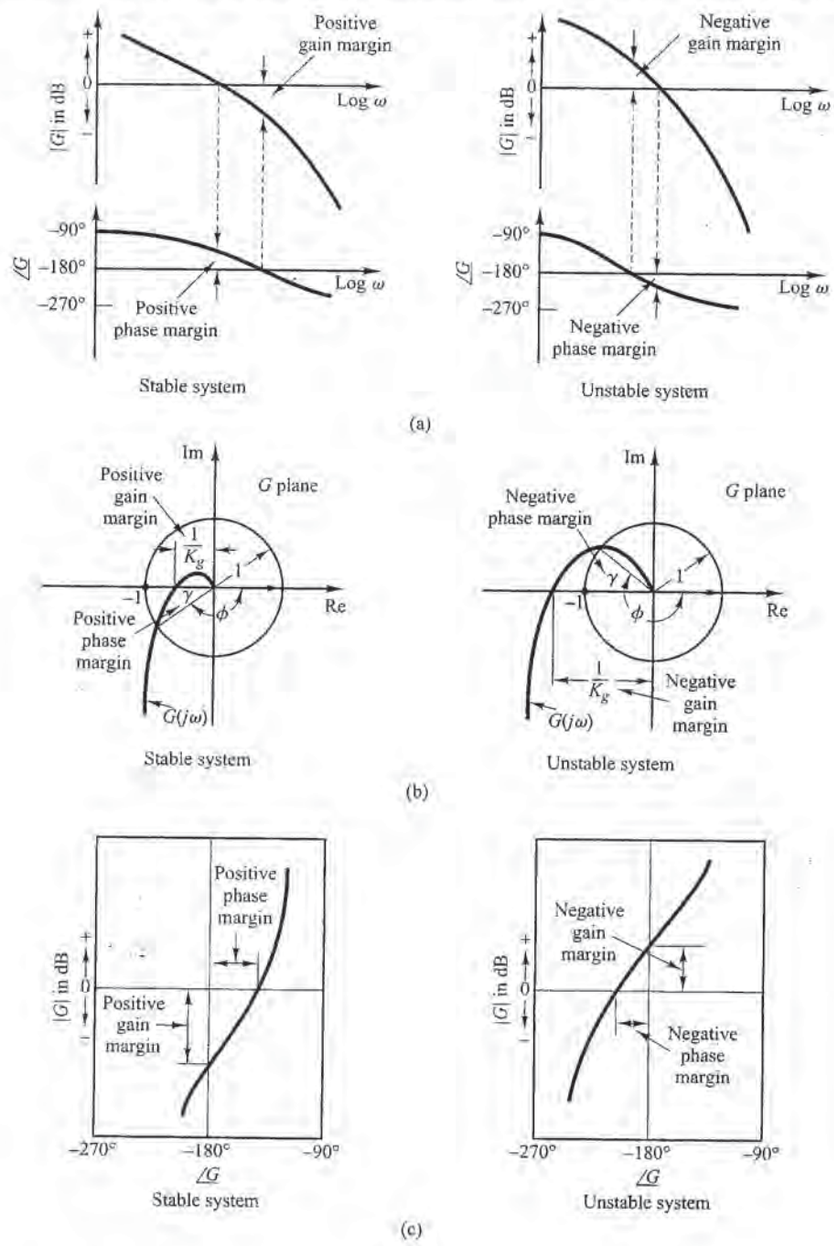


Figure 8–70
Phase and gain margins of stable and unstable systems.
(a) Bode diagrams;
(b) polar plots;
(c) log-magnitude-versus-phase plots.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

margin. The phase margin is positive for $\gamma > 0$ and negative for $\gamma < 0$. For a minimum-phase system to be stable, the phase margin must be positive. In the logarithmic plots, the critical point in the complex plane corresponds to the 0-dB and $-180°$ lines.

Gain margin: The gain margin is the reciprocal of the magnitude $|G(j\omega)|$ at the frequency at which the phase angle is $-180°$. Defining the phase crossover frequency ω_1 to be the frequency at which the phase angle of the open-loop transfer function equals $-180°$ gives the gain margin K_g:

$$K_g = \frac{1}{|G(j\omega_1)|}$$

In terms of decibels,

$$K_g \text{ dB} = 20 \log K_g = -20 \log |G(j\omega_1)|$$

The gain margin expressed in decibels is positive if K_g is greater than unity and negative if K_g is smaller than unity. Thus, a positive gain margin (in decibels) means that the system is stable, and a negative gain margin (in decibels) means that the system is unstable. The gain margin is shown in Figures 8–70(a), (b), and (c).

For a stable minimum-phase system, the gain margin indicates how much the gain can be increased before the system becomes unstable. For an unstable system, the gain margin is indicative of how much the gain must be decreased to make the system stable.

The gain margin of a first- or second-order system is infinite since the polar plots for such systems do not cross the negative real axis. Thus, theoretically, first- or second-order systems cannot be unstable. (Note, however, that so-called first- or second-order systems are only approximations in the sense that small time lags are neglected in deriving the system equations and are thus not truly first- or second-order systems. If these small lags are accounted for, the so-called first- or second-order systems may become unstable.)

It is noted that for a nonminimum-phase system with unstable open loop the stability condition will not be satisfied unless the $G(j\omega)$ plot encircles the $-1 + j0$ point. Hence, such a stable nonminimum-phase system will have negative phase and gain margins.

It is also important to point out that conditionally stable systems will have two or more phase crossover frequencies, and some higher-order systems with complicated numerator dynamics may also have two or more gain crossover frequencies, as shown in Figure 8–71. For stable systems having two or more gain crossover frequencies, the phase margin is measured at the highest gain crossover frequency.

A Few Comments on Phase and Gain Margins. The phase and gain margins of a control system are a measure of the closeness of the polar plot to the $-1 + j0$ point. Therefore, these margins may be used as design criteria.

It should be noted that either the gain margin alone or the phase margin alone does not give a sufficient indication of the relative stability. Both should be given in the determination of relative stability.

For a minimum-phase system, both the phase and gain margins must be positive for the system to be stable. Negative margins indicate instability.

Proper phase and gain margins ensure us against variations in the system components and are specified for definite positive values. The two values bound the behavior of the

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

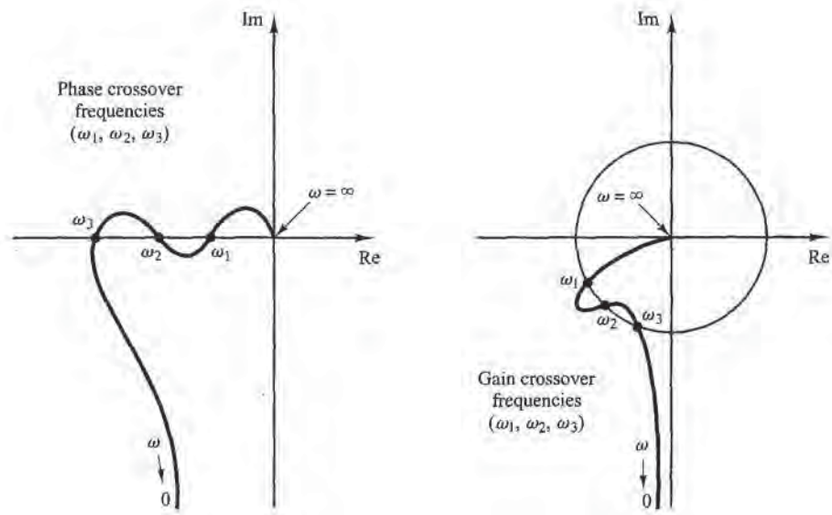


Figure 8–71
Polar plots showing more than two phase or gain crossover frequencies.

closed-loop system near the resonant frequency. For satisfactory performance, the phase margin should be between 30° and 60°, and the gain margin should be greater than 6 dB. With these values, a minimum-phase system has guaranteed stability, even if the open-loop gain and time constants of the components vary to a certain extent. Although the phase and gain margins give only rough estimates of the effective damping ratio of the closed-loop system, they do offer a convenient means for designing control systems or adjusting the gain constants of systems.

For minimum-phase systems, the magnitude and phase characteristics of the open-loop transfer function are definitely related. The requirement that the phase margin be between 30° and 60° means that in a Bode diagram the slope of the log-magnitude curve at the gain crossover frequency should be more gradual than −40 dB/decade. In most practical cases, a slope of −20 dB/decade is desirable at the gain crossover frequency for stability. If it is −40 dB/decade, the systems could be either stable or unstable. (Even if the system is stable, however, the phase margin is small.) If the slope at the gain crossover frequency is −60 dB/decade or steeper, the system is most likely unstable.

For nonminimum-phase systems, the correct interpretation of stability margins requires careful study. The best way to determine the stability of nonminimum-phase systems is to use the Nyquist diagram approach rather than Bode diagram approach.

EXAMPLE 8–21 Obtain the phase and gain margins of the system shown in Figure 8–72 for the two cases where $K = 10$ and $K = 100$.

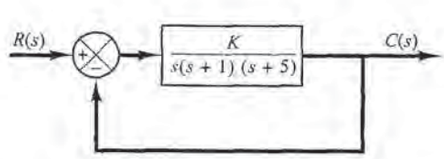


Figure 8–72
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

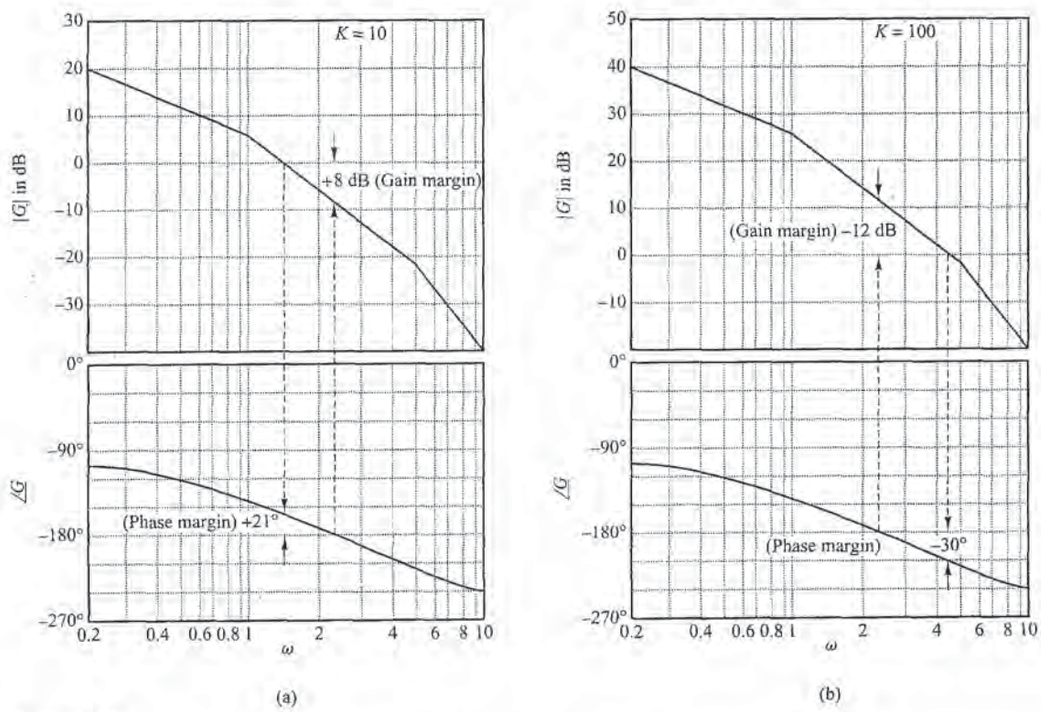


Figure 8–73
Bode diagrams of the system shown in Figure 8–72; (a) with $K = 10$ and (b) with $K = 100$.

The phase and gain margins can easily be obtained from the Bode diagram. A Bode diagram of the given open-loop transfer function with $K = 10$ is shown in Figure 8–73(a). The phase and gain margins for $K = 10$ are

$$\text{Phase margin} = 21°, \qquad \text{Gain margin} = 8 \text{ dB}$$

Therefore, the system gain may be increased by 8 dB before the instability occurs.

Increasing the gain from $K = 10$ to $K = 100$ shifts the 0-dB axis down by 20 dB, as shown in Figure 8–73(b). The phase and gain margins are

$$\text{Phase margin} = -30°, \qquad \text{Gain margin} = -12 \text{ dB}$$

Thus, the system is stable for $K = 10$, but unstable for $K = 100$.

Notice that one of the very convenient aspects of the Bode diagram approach is the ease with which the effects of gain changes can be evaluated. Note that to obtain satisfactory performance, we must increase the phase margin to $30° \sim 60°$. This can be done by decreasing the gain K. Decreasing K is not desirable, however, since a small value of K will yield a large error for the ramp input. This suggests that reshaping of the open-loop frequency-response curve by adding compensation may be necessary. Compensation techniques are discussed in detail in Chapter 9.

Obtaining Gain Margin, Phase Margin, Phase-Crossover Frequency, and Gain-Crossover Frequency with MATLAB. The gain margin, phase margin, phase-crossover frequency, and gain-crossover frequency can be obtained easily with MATLAB. The command to be used is

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

[Gm,pm,wcp,wcg] = margin(sys)

where Gm is the gain margin, pm is the phase margin, wcp is the phase-crossover frequency, and wcg is the gain-crossover frequency. For details of how to use this command, see Example 8–22.

EXAMPLE 8–22 Draw a Bode diagram of the open-loop transfer function $G(s)$ of the closed-loop system shown in Figure 8–74. Determine the gain margin, phase margin, phase-crossover frequency, and gain-crossover frequency with MATLAB.

A MATLAB program to plot a Bode diagram and to obtain the gain margin, phase margin, phase-crossover frequency, and gain-crossover frequency is shown in MATLAB Program 8–12. The Bode diagram of $G(s)$ is shown in Figure 8–75.

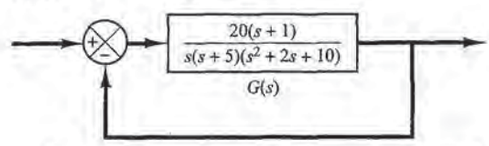


Figure 8–74
Closed-loop system.

MATLAB Program 8–12

```
num = [0 0 0 20 20];
den = conv([1 5 0],[1 2 10]);
sys = tf(num,den);
w = logspace(-1,2,100);
bode(sys,w)
[Gm,pm,wcp,wcg] = margin(sys);
GmdB = 20*log10(Gm);
[GmdB pm wcp wcg]

ans =

   9.9293  103.6573  4.0131  0.4426
```

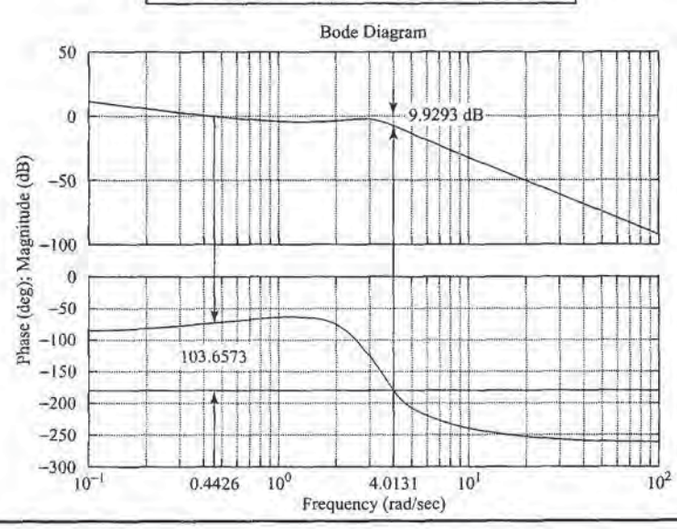


Figure 8–75
Bode diagram of $G(s)$ shown in Figure 8–74.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Resonant Peak Magnitude M_r and Resonant Frequency ω_r. Consider the standard second-order system shown in Figure 8–76. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{8–16}$$

where ζ and ω_n are the damping ratio and the undamped natural frequency, respectively. The closed-loop frequency response is

$$\frac{C(j\omega)}{R(j\omega)} = \frac{1}{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right) + j2\zeta\dfrac{\omega}{\omega_n}} = Me^{j\alpha}$$

where

$$M = \frac{1}{\sqrt{\left(1 - \dfrac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\dfrac{\omega}{\omega_n}\right)^2}}, \qquad \alpha = -\tan^{-1}\frac{2\zeta\dfrac{\omega}{\omega_n}}{1 - \dfrac{\omega^2}{\omega_n^2}}$$

As given by Equation (8–12), for $0 \le \zeta \le 0.707$, the maximum value of M occurs at the frequency ω_r, where

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2} \tag{8–17}$$

The frequency ω_r is the resonant frequency. At the resonant frequency, the value of M is maximum and is given by Equation (8–13), rewritten

$$M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \tag{8–18}$$

where M_r is defined as the *resonant peak magnitude*. The resonant peak magnitude is related to the damping of the system.

The magnitude of the resonant peak gives an indication of the relative stability of the system. A large resonant peak magnitude indicates the presence of a pair of dominant closed-loop poles with small damping ratio, which will yield an undesirable transient response. A smaller resonant peak magnitude, on the other hand, indicates the absence of a pair of dominant closed-loop poles with small damping ratio, meaning that the system is well damped.

Remember that ω_r is real only if $\zeta < 0.707$. Thus, there is no closed-loop resonance if $\zeta > 0.707$. [The value of M_r is unity only if $\zeta > 0.707$. See Equation (8–14).] Since the values of M_r and ω_r can be easily measured in a physical system, they are quite useful for checking agreement between theoretical and experimental analyses.

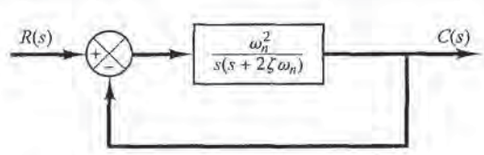


Figure 8–76
Standard second-order system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

It is noted, however, that in practical design problems the phase margin and gain margin are more frequently specified than the resonant peak magnitude to indicate the degree of damping in a system.

Correlation between Step Transient Response and Frequency Response in the Standard Second-Order System. The maximum overshoot in the unit-step response of the standard second-order system, as shown in Figure 8–76, can be exactly correlated with the resonant peak magnitude in the frequency response. Hence, essentially the same information about the system dynamics is contained in the frequency response as is in the transient response.

For a unit-step input, the output of the system shown in Figure 8–76 is given by Equation (5-12), or

$$c(t) = 1 - e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right), \qquad \text{for } t \geq 0$$

where

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \tag{8–19}$$

On the other hand, the maximum overshoot M_p for the unit-step response is given by Equation (5–21), or

$$M_p = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \tag{8–20}$$

This maximum overshoot occurs in the transient response that has the damped natural frequency $\omega_d = \omega_n\sqrt{1-\zeta^2}$. The maximum overshoot becomes excessive for values of $\zeta < 0.4$.

Since the second-order system shown in Figure 8–76 has the open-loop transfer function

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

for sinusoidal operation, the magnitude of $G(j\omega)$ becomes unity when

$$\omega = \omega_n\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}$$

which can be obtained by equating $|G(j\omega)|$ to unity and solving for ω. At this frequency, the phase angle of $G(j\omega)$ is

$$\underline{/G(j\omega)} = -\underline{/j\omega} - \underline{/j\omega + 2\zeta\omega_n} = -90^\circ - \tan^{-1}\frac{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}{2\zeta}$$

Thus, the phase margin γ is

$$\begin{aligned}\gamma &= 180^\circ + \underline{/G(j\omega)} \\ &= 90^\circ - \tan^{-1}\frac{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}{2\zeta} \\ &= \tan^{-1}\frac{2\zeta}{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}\end{aligned} \tag{8–21}$$

Equation (8–21) gives the relationship between the damping ratio ζ and the phase margin γ. (Notice that the phase margin γ is a function *only* of the damping ratio ζ.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

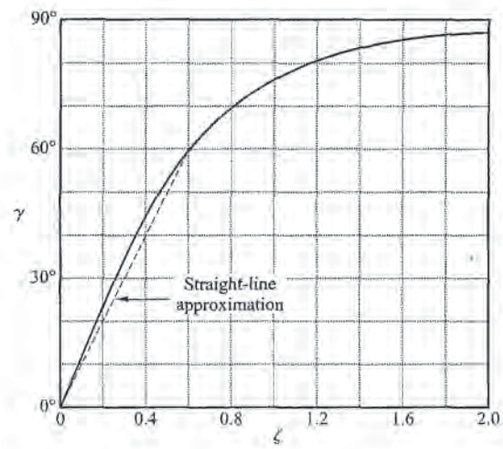


Figure 8–77
Curve γ (phase margin) versus ζ for the system shown in Figure 8–76.

In the following, we shall summarize the correlation between the step transient response and frequency response of the standard second-order system given by Equation (8–16):

1. The phase margin and the damping ratio are directly related. Figure 8–77 shows a plot of the phase margin γ as a function of the damping ratio ζ. It is noted that for the standard second-order system shown in Figure 8–76, the phase margin γ and the damping ratio ζ are related approximately by a straight line for $0 \leq \zeta \leq 0.6$, as follows:
$$\zeta = \frac{\gamma}{100}$$
Thus a phase margin of 60° corresponds to a damping ratio of 0.6. For higher-order systems having a dominant pair of closed-loop poles, this relationship may be used as a rule of thumb in estimating the relative stability in the transient response (that is, the damping ratio) from the frequency response.
2. Referring to Equations (8–17) and (8–19), we see that the values of ω_r and ω_d are almost the same for small values of ζ. Thus, for small values of ζ, the value of ω_r is indicative of the speed of the transient response of the system.
3. From Equations (8–18) and (8–20), we note that the smaller the value of ζ is, the larger the values of M_r and M_p are. The correlation between M_r and M_p as a function of ζ is shown in Figure 8–78. A close relationship between M_r and M_p can be seen for $\zeta > 0.4$. For very small values of ζ, M_r becomes very large $(M_r \gg 1)$, while the value of M_p does not exceed 1.

Correlation between Step Transient Response and Frequency Response in General Systems. The design of control systems is very often carried out on the basis of the frequency response. The main reason for this is the relative simplicity of this approach compared with others. Since in many applications it is the transient response of the system to aperiodic inputs rather than the steady-state response to sinusoidal inputs that is of primary concern, the question of correlation between transient response and frequency response arises.

For the standard second-order system shown in Figure 8–76, mathematical relationships correlating the step transient response and frequency response can be obtained

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

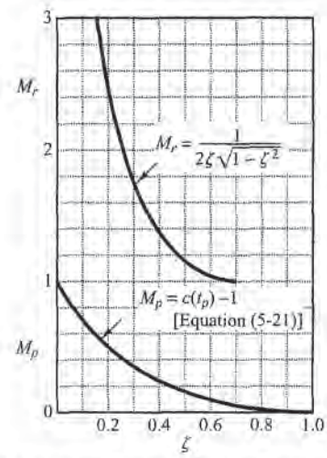


Figure 8–78
Curves M_r versus ζ and M_p versus ζ for the system shown in Figure 8–76.

easily. The time response of the standard second-order system can be predicted exactly from a knowledge of the M_r and ω_r of its closed-loop frequency response.

For nonstandard second-order systems and higher-order systems, the correlation is more complex, and the transient response may not be predicted easily from the frequency response because additional zeros and/or poles may change the correlation between the step transient response and the frequency response existing for the standard second-order system. Mathematical techniques for obtaining the exact correlation are available, but they are very laborious and of little practical value.

The applicability of the transient-response–frequency-response correlation existing for the standard second-order system shown in Figure 8–76 to higher-order systems depends on the presence of a dominant pair of complex-conjugate closed-loop poles in the latter systems. Clearly, if the frequency response of a higher-order system is dominated by a pair of complex-conjugate closed-loop poles, the transient-response–frequency-response correlation existing for the standard second-order system can be extended to the higher-order system.

For linear, time-invariant, higher-order systems having a dominant pair of complex-conjugate closed-loop poles, the following relationships generally exist between the step transient response and frequency response:

1. The value of M_r is indicative of the relative stability. Satisfactory transient performance is usually obtained if the value of M_r is in the range $1.0 < M_r < 1.4$ $(0 \text{ dB} < M_r < 3 \text{ dB})$, which corresponds to an effective damping ratio of $0.4 < \zeta < 0.7$. For values of M_r greater than 1.5, the step transient response may exhibit several overshoots. (Note that, in general, a large value of M_r corresponds to a large overshoot in the step transient response. If the system is subjected to noise signals whose frequencies are near the resonant frequency ω_r, the noise will be amplified in the output and will present serious problems.)
2. The magnitude of the resonant frequency ω_r is indicative of the speed of the transient response. The larger the value of ω_r, the faster the time response is. In other words, the rise time varies inversely with ω_r. In terms of the open-loop frequency response, the damped natural frequency in the transient response is somewhere between the gain crossover frequency and phase crossover frequency.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

3. The resonant peak frequency ω_r and the damped natural frequency ω_d for the step transient response are very close to each other for lightly damped systems.

The three relationships just listed are useful for correlating the step transient response with the frequency response of higher-order systems, provided that they can be approximated by the standard second-order system or a pair of complex-conjugate closed-loop poles. If a higher-order system satisfies this condition, a set of time-domain specifications may be translated into frequency-domain specifications. This simplifies greatly the design work or compensation work of higher-order systems.

In addition to the phase margin, gain margin, resonant peak M_r, and resonant frequency ω_r, there are other frequency-domain quantities commonly used in performance specifications. They are the cutoff frequency, bandwidth, and the cutoff rate. These will be defined in what follows.

Cutoff Frequency and Bandwidth. Referring to Figure 8–79, the frequency ω_b at which the magnitude of the closed-loop frequency response is 3 dB below its zero-frequency value is called the *cutoff frequency*. Thus

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| < \left|\frac{C(j0)}{R(j0)}\right| - 3\text{ dB}, \qquad \text{for } \omega > \omega_b$$

For systems in which $|C(j0)/R(j0)| = 0$ dB,

$$\left|\frac{C(j\omega)}{R(j\omega)}\right| < -3\text{ dB}, \qquad \text{for } \omega > \omega_b$$

The closed-loop system filters out the signal components whose frequencies are greater than the cutoff frequency and transmits those signal components with frequencies lower than the cutoff frequency.

The frequency range $0 \le \omega \le \omega_b$ in which the magnitude of the closed loop does not drop -3 dB is called the *bandwidth* of the system. The bandwidth indicates the frequency where the gain starts to fall off from its low-frequency value. Thus, the bandwidth indicates how well the system will track an input sinusoid. Note that for a given ω_n, the rise time increases with increasing damping ratio ζ. On the other hand, the bandwidth decreases with the increase in ζ. Therefore, the rise time and the bandwidth are inversely proportional to each other.

The specification of the bandwidth may be determined by the following factors:

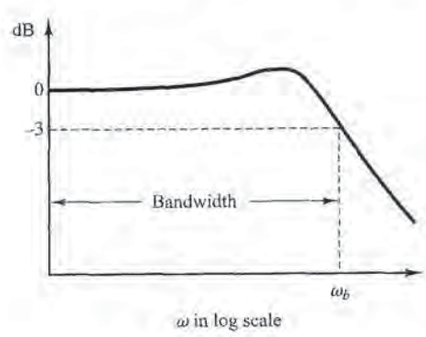


Figure 8–79
Plot of a closed-loop frequency response curve showing cutoff frequency ω_b and bandwidth.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

1. The ability to reproduce the input signal. A large bandwidth corresponds to a small rise time, or fast response. Roughly speaking, we can say that the bandwidth is proportional to the speed of response. (For example, to decrease the rise time in the step response by a factor of 2, the bandwidth must be increased by approximately a factor of 2.)
2. The necessary filtering characteristics for high-frequency noise.

For the system to follow arbitrary inputs accurately, it must have a large bandwidth. From the viewpoint of noise, however, the bandwidth should not be too large. Thus, there are conflicting requirements on the bandwidth, and a compromise is usually necessary for good design. Note that a system with large bandwidth requires high-performance components, so the cost of components usually increases with the bandwidth.

Cutoff Rate. The cutoff rate is the slope of the log-magnitude curve near the cutoff frequency. The cutoff rate indicates the ability of a system to distinguish the signal from noise.

It is noted that a closed-loop frequency response curve with a steep cutoff characteristic may have a large resonant peak magnitude, which implies that the system has a relatively small stability margin.

EXAMPLE 8–23 Consider the following two systems:

$$\text{System I:}\quad \frac{C(s)}{R(s)} = \frac{1}{s+1}, \qquad \text{System II:}\quad \frac{C(s)}{R(s)} = \frac{1}{3s+1}$$

Compare the bandwidths of these two systems. Show that the system with the larger bandwidth has a faster speed of response and can follow the input much better than the one with the smaller bandwidth.

Figure 8–80(a) shows the closed-loop frequency-response curves for the two systems. (Asymptotic curves are shown by dashed lines.) We find that the bandwidth of system I is $0 \le \omega \le 1$ rad/sec and that of system II is $0 \le \omega \le 0.33$ rad/sec. Figures 8–80(b) and (c) show, respectively, the unit-step response and unit-ramp response curves for the two systems. Clearly, system I, whose bandwidth is three times wider than that of system II, has a faster speed of response and can follow the input much better.

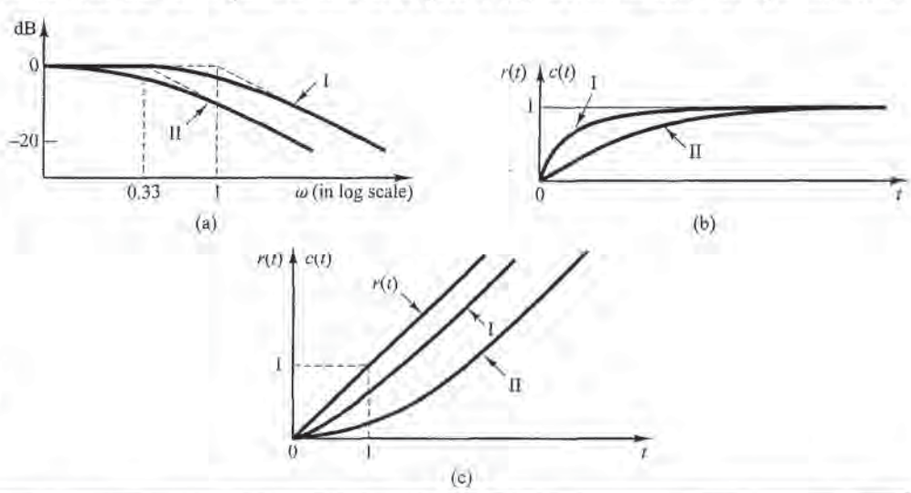


Figure 8–80 Comparison of dynamic characteristics of the two systems considered in Example 8–23. (a) Closed-loop frequency-response curves; (b) unit-step response curves; (c) unit-ramp response curves.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Approach to Get Resonant Peak, Resonant Frequency, and Bandwidth. The resonant peak is the value of the maximum magnitude (in decibels) of the closed-loop frequency response. The resonant frequency is the frequency that yields the maximum magnitude. MATLAB commands to be used for obtaining the resonant peak and resonant frequency are as follows:

```
[mag,phase,w] = bode(num,den,w);     or     [mag,phase,w] = bode(sys,w);
[Mp,k] = max(mag);
resonant_peak = 20*log10(Mp);
resonant_frequency = w(k)
```

The bandwidth can be obtained by entering the following lines in the program:

```
n = 1;
while 20*log10(mag(n)) > = -3; n = n + 1;
end
bandwidth = w(n)
```

For a detailed MATLAB program, see Example 8–24.

EXAMPLE 8–24 Consider the system shown in Figure 8–81. Using MATLAB, obtain a Bode diagram for the closed-loop transfer function. Obtain also the resonant peak, resonant frequency, and bandwidth.

MATLAB Program 8–13 produces a Bode diagram for the closed-loop system as well as the resonant peak, resonant frequency, and bandwidth. The resulting Bode diagram is shown in

MATLAB Program 8–13

```
nump = [0 0 0 1];
denp = [0.5 1.5 1 0];
sysp = tf(nump,denp);
sys = feedback(sysp,1);
w = logspace(-1,1);
bode(sys,w)
[mag,phase,w] = bode(sys,w);
[Mp,k] = max(mag);
resonant_peak = 20*log10(Mp)

resonant_peak =

   5.2388

resonant_frequency = w(k)

resonant_frequency =

   0.7906

n = 1;
while 20*log(mag(n))> = -3; n = n + 1;
end
bandwidth = w(n)

bandwidth =

   1.2649
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

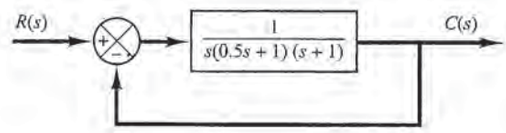


Figure 8–81
Closed-loop system.

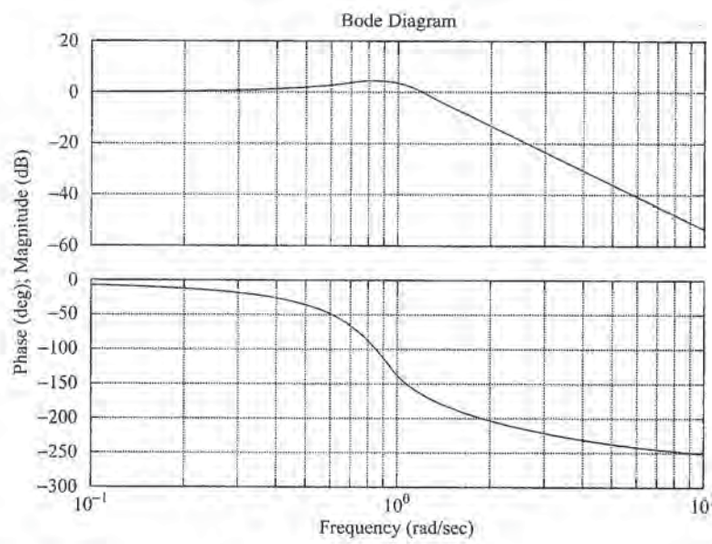


Figure 8–82
Bode diagram of the closed-loop transfer function of the system shown in Figure 8–81.

Figure 8–82. The resonant peak is obtained as 5.2388 dB. The resonant frequency is 0.7906 rad/sec. The bandwidth is 1.2649 rad/sec. These values can be verified from Figure 8–82.

8–10 CLOSED-LOOP FREQUENCY RESPONSE OF UNITY-FEEDBACK SYSTEMS

Closed-Loop Frequency Response. For a stable, unity-feedback closed-loop system, the closed-loop frequency response can be obtained easily from that of the open-loop frequency response. Consider the unity-feedback system shown in Figure 8–83(a). The closed-loop transfer function is

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

In the Nyquist or polar plot shown in Figure 8–83(b), the vector $\overrightarrow{OA}$ represents $G(j\omega_1)$, where ω_1 is the frequency at point A. The length of the vector $\overrightarrow{OA}$ is $|G(j\omega_1)|$ and the angle of the vector $\overrightarrow{OA}$ is $\underline{/G(j\omega_1)}$. The vector $\overrightarrow{PA}$, the vector from the $-1 + j0$ point to the Nyquist locus, represents $1 + G(j\omega_1)$. Therefore, the ratio of $\overrightarrow{OA}$, to $\overrightarrow{PA}$ represents the closed-loop frequency response, or

$$\frac{\overrightarrow{OA}}{\overrightarrow{PA}} = \frac{G(j\omega_1)}{1 + G(j\omega_1)} = \frac{C(j\omega_1)}{R(j\omega_1)}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

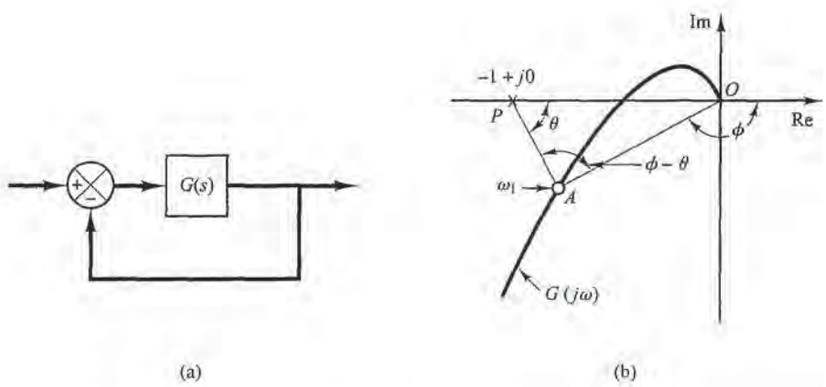


Figure 8–83
(a) Unity-feedback system;
(b) determination of closed-loop frequency response from open-loop frequency response.

The magnitude of the closed-loop transfer function at $\omega = \omega_1$ is the ratio of the magnitudes of $\overrightarrow{OA}$ to $\overrightarrow{PA}$. The phase angle of the closed-loop transfer function at $\omega = \omega_1$ is the angle formed by the vectors $\overrightarrow{OA}$ to $\overrightarrow{PA}$, that is $\phi - \theta$, shown in Figure 8–83(b). By measuring the magnitude and phase angle at different frequency points, the closed-loop frequency-response curve can be obtained.

Let us define the magnitude of the closed-loop frequency response as M and the phase angle as α, or

$$\frac{C(j\omega)}{R(j\omega)} = Me^{j\alpha}$$

In the following, we shall find the constant magnitude loci and constant phase-angle loci. Such loci are convenient in determining the closed-loop frequency response from the polar plot or Nyquist plot.

Constant-Magnitude Loci (M circles). To obtain the constant-magnitude loci, let us first note that $G(j\omega)$ is a complex quantity and can be written as follows:

$$G(j\omega) = X + jY$$

where X and Y are real quantities. Then M is given by

$$M = \frac{|X + jY|}{|1 + X + jY|}$$

and M^2 is

$$M^2 = \frac{X^2 + Y^2}{(1 + X)^2 + Y^2}$$

Hence

$$X^2(1 - M^2) - 2M^2X - M^2 + (1 - M^2)Y^2 = 0 \qquad (8\text{–}22)$$

If $M = 1$, then from Equation (8–22), we obtain $X = -\frac{1}{2}$. This is the equation of a straight line parallel to the Y axis and passing through the point $(-\frac{1}{2}, 0)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

If $M \neq 1$, Equation (8–22) can be written

$$X^2 + \frac{2M^2}{M^2 - 1} X + \frac{M^2}{M^2 - 1} + Y^2 = 0$$

If the term $M^2/(M^2 - 1)^2$ is added to both sides of this last equation, we obtain

$$\left(X + \frac{M^2}{M^2 - 1}\right)^2 + Y^2 = \frac{M^2}{(M^2 - 1)^2} \tag{8–23}$$

Equation (8–23) is the equation of a circle with center at $X = -M^2/(M^2 - 1)$, $Y = 0$ and with radius $|M/(M^2 - 1)|$.

The constant M loci on the $G(s)$ plane are thus a family of circles. The center and radius of the circle for a given value of M can be easily calculated. For example, for $M = 1.3$, the center is at $(-2.45, 0)$ and the radius is 1.88. A family of constant M circles is shown in Figure 8–84. It is seen that as M becomes larger compared with 1, the M circles become smaller and converge to the $-1 + j0$ point. For $M > 1$, the centers of the M circles lie to the left of the $-1 + j0$ point. Similarly, as M becomes smaller compared with 1, the M circle becomes smaller and converges to the origin. For $0 < M < 1$, the centers of the M circles lie to the right of the origin. $M = 1$ corresponds to the locus of points equidistant from the origin and from the $-1 + j0$ point. As stated earlier, it is a straight line passing through the point $(-\frac{1}{2}, 0)$ and parallel to the imaginary axis. (The constant M circles corresponding to $M > 1$ lie to the left of the $M = 1$ line, and those corresponding to $0 < M < 1$ lie to the right of the $M = 1$ line.) The M circles are symmetrical with respect to the straight line corresponding to $M = 1$ and with respect to the real axis.

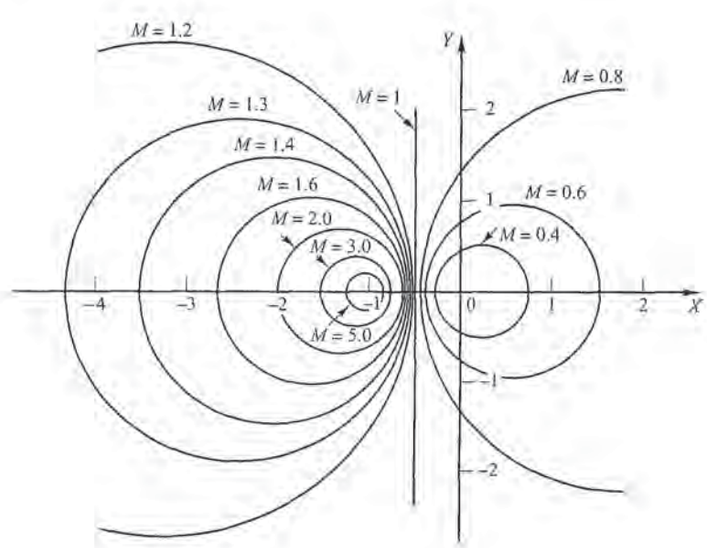


Figure 8–84
A family of constant M circles.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Constant Phase-Angle Loci (N Circles). We shall obtain the phase angle α in terms of X and Y. Since

$$\angle e^{j\alpha} = \angle \frac{X + jY}{1 + X + jY}$$

the phase angle α is

$$\alpha = \tan^{-1}\left(\frac{Y}{X}\right) - \tan^{-1}\left(\frac{Y}{1 + X}\right)$$

If we define

$$\tan\alpha = N$$

then

$$N = \tan\left[\tan^{-1}\left(\frac{Y}{X}\right) - \tan^{-1}\left(\frac{Y}{1 + X}\right)\right]$$

Since

$$\tan(A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}$$

we obtain

$$N = \frac{\dfrac{Y}{X} - \dfrac{Y}{1 + X}}{1 + \dfrac{Y}{X}\left(\dfrac{Y}{1 + X}\right)} = \frac{Y}{X^2 + X + Y^2}$$

or

$$X^2 + X + Y^2 - \frac{1}{N}Y = 0$$

The addition of $(\frac{1}{4}) + 1/(2N)^2$ to both sides of this last equation yields

$$\left(X + \frac{1}{2}\right)^2 + \left(Y - \frac{1}{2N}\right)^2 = \frac{1}{4} + \left(\frac{1}{2N}\right)^2 \tag{8–24}$$

This is an equation of a circle with center at $X = -\frac{1}{2}$, $Y = 1/(2N)$ and with radius $\sqrt{\frac{1}{4} + 1/(2N)^2}$. For example, if $\alpha = 30°$, then $N = \tan\alpha = 0.577$, and the center and the radius of the circle corresponding to $\alpha = 30°$ are found to be $(-0.5, 0.866)$ and unity, respectively. Since Equation (8–24) is satisfied when $X = Y = 0$ and $X = -1, Y = 0$ regardless of the value of N, each circle passes through the origin and the $-1 + j0$ point. The constant α loci can be drawn easily once the value of N is given. A family of constant N circles is shown in Figure 8–85 with α as a parameter.

It should be noted that the constant N locus for a given value of α is actually not the entire circle, but only an arc. In other words, the $\alpha = 30°$ and $\alpha = -150°$ arcs are parts of the same circle. This is so because the tangent of an angle remains the same if $\pm 180°$ (or multiples thereof) is added to the angle.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

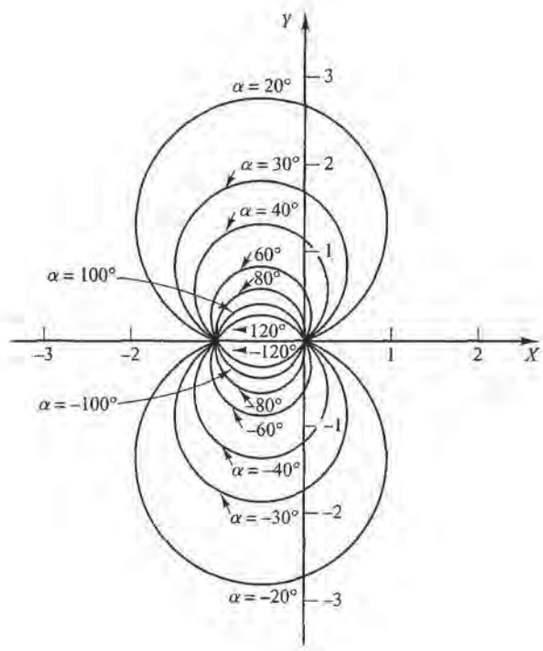


Figure 8–85
A family of constant N circles.

The use of the M and N circles enables us to find the entire closed-loop frequency response from the open-loop frequency response $G(j\omega)$ without calculating the magnitude and phase of the closed-loop transfer function at each frequency. The intersections of the $G(j\omega)$ locus and the M circles and N circles give the values of M and N at frequency points on the $G(j\omega)$ locus.

The N circles are multivalued in the sense that the circle for $\alpha = \alpha_1$ and that for $\alpha = \alpha_1 \pm 180°n$ $(n = 1, 2, \ldots)$ are the same. In using the N circles for the determination of the phase angle of closed-loop systems, we must interpret the proper value of α. To avoid any error, start at zero frequency, which corresponds to $\alpha = 0°$, and proceed to higher frequencies. The phase-angle curve must be continuous.

Graphically, the intersections of the $G(j\omega)$ locus and M circles give the values of M at the frequencies denoted on the $G(j\omega)$ locus. Thus, the constant M circle with the smallest radius that is tangent to the $G(j\omega)$ locus gives the value of the resonant peak magnitude M_r. If it is desired to keep the resonant peak value less than a certain value, then the system should not enclose the critical point $(-1 + j0$ point) and, at the same time, there should be no intersections with the particular M circle and the $G(j\omega)$ locus.

Figure 8–86(a) shows the $G(j\omega)$ locus superimposed on a family of M circles. Figure 8–86(b) shows the $G(j\omega)$ locus superimposed on a family of N circles. From these plots, it is possible to obtain the closed-loop frequency response by inspection. It is seen that the $M = 1.1$ circle intersects the $G(j\omega)$ locus at frequency point $\omega = \omega_1$. This means that at this frequency the magnitude of the closed-loop transfer function is 1.1. In Figure 8–86(a), the $M = 2$ circle is just tangent to the $G(j\omega)$ locus. Thus, there is only one point on the $G(j\omega)$ locus for which $|C(j\omega)/R(j\omega)|$ is equal to 2. Figure 8–86(c) shows the

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

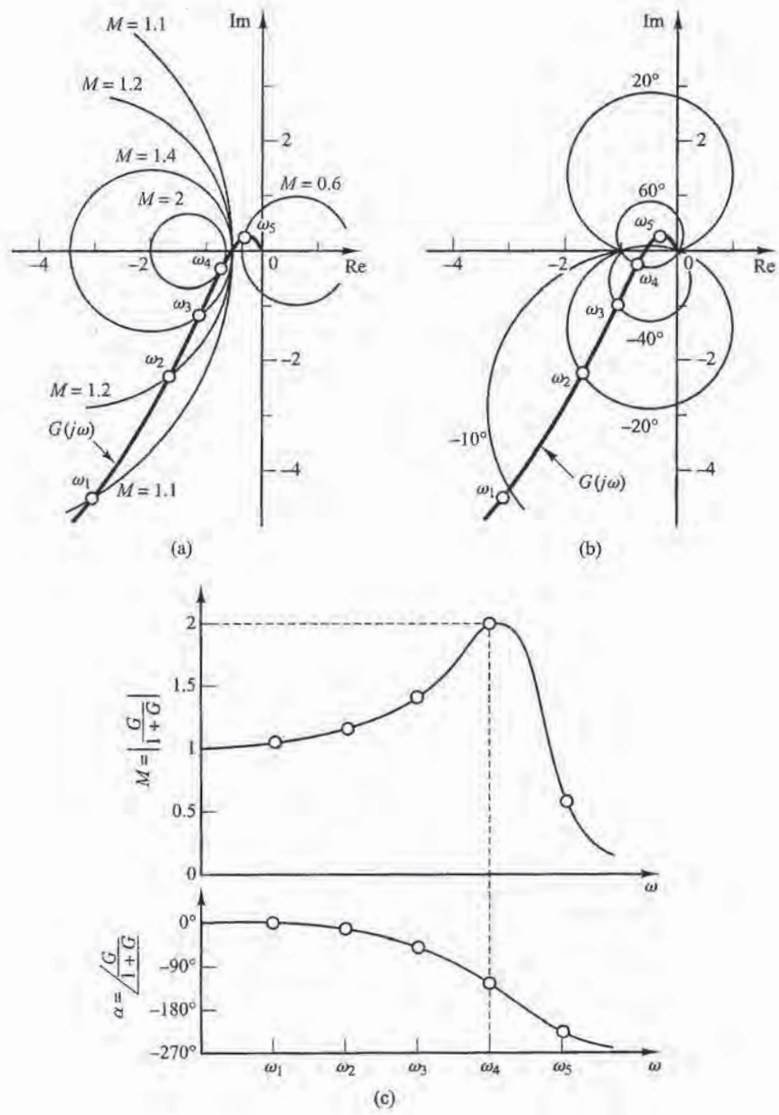


Figure 8–86
(a) $G(j\omega)$ locus superimposed on a family of M circles; (b) $G(j\omega)$ locus superimposed on a family of N circles; (c) closed-loop frequency-response curves.

closed-loop frequency-response curve for the system. The upper curve is the M-versus-frequency ω curve, and the lower curve is the phase angle α-versus-frequency ω curve.

The resonant peak value is the value of M corresponding to the M circle of smallest radius that is tangent to the $G(j\omega)$ locus. Thus, in the Nyquist diagram, the resonant peak value M_r and the resonant frequency ω_r can be found from the M-circle tangency to the $G(j\omega)$ locus. (In the present example, $M_r = 2$ and $\omega_r = \omega_4$.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Nichols Chart. In dealing with design problems, we find it convenient to construct the M and N loci in the log-magnitude-versus-phase plane. The chart consisting of the M and N loci in the log-magnitude-versus-phase diagram is called the Nichols chart. The $G(j\omega)$ locus drawn on the Nichols chart gives both the gain characteristics and phase characteristics of the closed-loop transfer function at the same time. The Nichols chart is shown in Figure 8–87, for phase angles between 0° and −240°.

Note that the critical point ($-1 + j0$ point) is mapped to the Nichols chart as the point (0 dB, −180°). The Nichols chart contains curves of constant closed-loop magnitude and phase angle. The designer can graphically determine the phase margin, gain margin, resonant peak magnitude, resonant frequency, and bandwidth of the closed-loop system from the plot of the open-loop locus, $G(j\omega)$.

The Nichols chart is symmetric about the −180° axis. The M and N loci repeat for every 360°, and there is symmetry at every 180° interval. The M loci are centered about the critical point (0 dB, −180°). The Nichols chart is useful for determining the frequency response of the closed loop from that of the open loop. If the open-loop frequency-response curve is superimposed on the Nichols chart, the intersections of the open-loop frequency-response curve $G(j\omega)$ and the M and N loci give the values of the magnitude M and phase angle α of the closed-loop frequency response at each frequency point. If the $G(j\omega)$ locus does not intersect the $M = M_r$ locus, but is tangent to it, then the resonant peak value of M of the closed-loop frequency response is given by M_r. The resonant frequency is given by the frequency at the point of tangency.

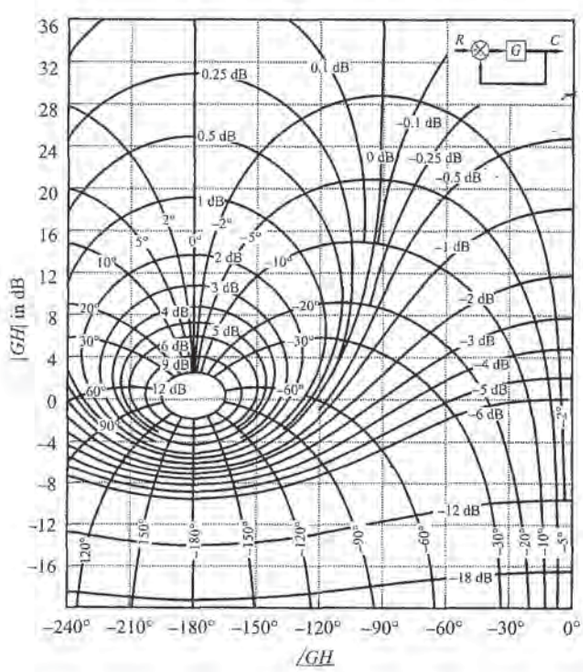


Figure 8–87
Nichols chart.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

As an example, consider the unity-feedback system with the following open-loop transfer function:

$$G(j\omega) = \frac{K}{s(s+1)(0.5s+1)}, \qquad K = 1$$

To find the closed-loop frequency response by use of the Nichols chart, the $G(j\omega)$ locus is constructed in the log-magnitude-versus-phase plane by use of MATLAB or from the Bode diagram. Figure 8–88(a) shows the $G(j\omega)$ locus together with the M and N loci. The closed-loop frequency-response curves may be constructed by reading the magnitudes and phase angles at various frequency points on the $G(j\omega)$ locus from the M and N loci, as shown in Figure 8–88(b). Since the largest magnitude contour touched by the $G(j\omega)$ locus is 5 dB, the resonant peak magnitude M_r is 5 dB. The corresponding resonant peak frequency is 0.8 rad/sec.

Notice that the phase crossover point is the point where the $G(j\omega)$ locus intersects the $-180°$ axis (for the present system, $\omega = 1.4$ rad/sec), and the gain crossover point is the point where the locus intersects the 0-dB axis (for the present system, $\omega = 0.76$ rad/sec). The phase margin is the horizontal distance (measured in degrees) between the gain crossover point and the critical point (0 dB, $-180°$). The gain margin is the distance (in decibels) between the phase crossover point and the critical point.

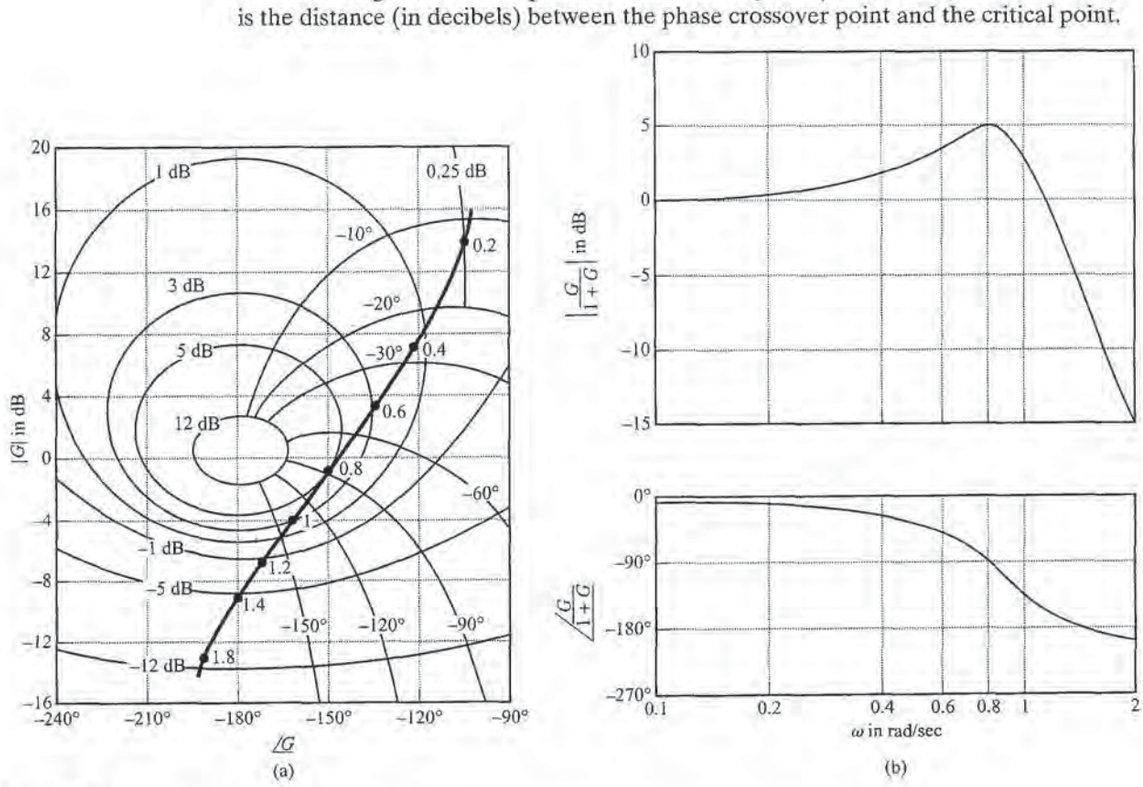


Figure 8–88
(a) Plot of $G(j\omega)$ superimposed on Nichols chart; (b) closed-loop frequency-response curves.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The bandwidth of the closed-loop system can easily be found from the $G(j\omega)$ locus in the Nichols diagram. The frequency at the intersection of the $G(j\omega)$ locus and the $M = -3$ dB locus gives the bandwidth.

If the open-loop gain K is varied, the shape of the $G(j\omega)$ locus in the log-magnitude-versus-phase diagram remains the same, but it is shifted up (for increasing K) or down (for decreasing K) along the vertical axis. Therefore, the $G(j\omega)$ locus intersects the M and N loci differently, resulting in a different closed-loop frequency-response curve. For a small value of the gain K, the $G(j\omega)$ locus will not be tangent to any of the M loci, which means that there is no resonance in the closed-loop frequency response.

EXAMPLE 8–25 Consider the unity-feedback control system whose open-loop transfer function is

$$G(j\omega) = \frac{K}{j\omega(1 + j\omega)}$$

Determine the value of the gain K so that $M_r = 1.4$.

The first step in the determination of the gain K is to sketch the polar plot of

$$\frac{G(j\omega)}{K} = \frac{1}{j\omega(1 + j\omega)}$$

Figure 8–89 shows the $M_r = 1.4$ locus and the $G(j\omega)/K$ locus. Changing the gain has no effect on the phase angle, but merely moves the curve vertically up for $K > 1$ and down for $K < 1$.

In Figure 8–89, the $G(j\omega)/K$ locus must be raised by 4 dB in order that it be tangent to the desired M_r locus and that the entire $G(j\omega)/K$ locus be outside the $M_r = 1.4$ locus. The amount of

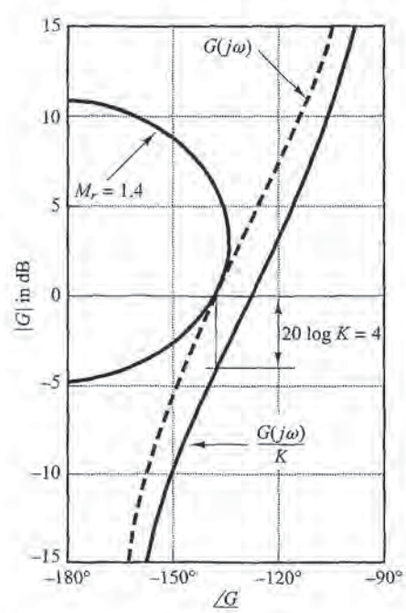


Figure 8–89
Determination of the gain K using the Nichols chart.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

vertical shift of the $G(j\omega)/K$ locus determines the gain necessary to yield the desired value of M_r. Thus, by solving

$$20 \log K = 4$$

we obtain

$$K = 1.59$$

8-11 EXPERIMENTAL DETERMINATION OF TRANSFER FUNCTIONS

The first step in the analysis and design of a control system is to derive a mathematical model of the plant under consideration. Obtaining a model analytically may be quite difficult. We may have to obtain it by means of experimental analysis. The importance of the frequency-response methods is that the transfer function of the plant, or any other component of a system, may be determined by simple frequency-response measurements.

If the amplitude ratio and phase shift have been measured at a sufficient number of frequencies within the frequency range of interest, they may be plotted on the Bode diagram. Then the transfer function can be determined by asymptotic approximations. We build up asymptotic log-magnitude curves consisting of several segments. With some trial-and-error juggling of the corner frequencies, it is usually possible to find a very close fit to the curve. (Note that if the frequency is plotted in cycles per second rather than radians per second, the corner frequencies must be converted to radians per second before computing the time constants.)

Sinusoidal-Signal Generators. In performing a frequency-response test, suitable sinusoidal-signal generators must be available. The signal may have to be in mechanical, electrical, or pneumatic form. The frequency ranges needed for the test are approximately 0.001 to 10 Hz for large-time-constant systems and 0.1 to 1000 Hz for small-time-constant systems. The sinusoidal signal must be reasonably free from harmonics or distortion.

For very low frequency ranges (below 0.01 Hz), a mechanical signal generator (together with a suitable pneumatic or electrical transducer if necessary) may be used. For the frequency range from 0.01 to 1000 Hz, a suitable electrical-signal generator (together with a suitable transducer if necessary) may be used.

Determination of Minimum-Phase Transfer Functions from Bode Diagrams. As stated previously, whether a system is minimum phase can be determined from the frequency-response curves by examining the high-frequency characteristics.

To determine the transfer function, we first draw asymptotes to the experimentally obtained log-magnitude curve. The asymptotes must have slopes of multiples of ±20 dB/decade. If the slope of the experimentally obtained log-magnitude curve changes from −20 to −40 dB/decade at $\omega = \omega_1$, it is clear that a factor $1/[1 + j(\omega/\omega_1)]$ exists in the transfer function. If the slope changes by −40 dB/decade at $\omega = \omega_2$, there must be a quadratic factor of the form

$$\frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_2}\right) + \left(j\dfrac{\omega}{\omega_2}\right)^2}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

in the transfer function. The undamped natural frequency of this quadratic factor is equal to the corner frequency ω_2. The damping ratio ζ can be determined from the experimentally obtained log-magnitude curve by measuring the amount of resonant peak near the corner frequency ω_2 and comparing this with the curves shown in Figure 8–9.

Once the factors of the transfer function $G(j\omega)$ have been determined, the gain can be determined from the low-frequency portion of the log-magnitude curve. Since such terms as $1 + j(\omega/\omega_1)$ and $1 + 2\zeta(j\omega/\omega_2) + (j\omega/\omega_2)^2$ become unity as ω approaches zero, at very low frequencies the sinusoidal transfer function $G(j\omega)$ can be written

$$\lim_{\omega \to 0} G(j\omega) = \frac{K}{(j\omega)^\lambda}$$

In many practical systems, λ equals 0, 1, or 2.

1. For $\lambda = 0$, or type 0 systems,

$$G(j\omega) = K, \qquad \text{for } \omega \ll 1$$

or

$$20 \log|G(j\omega)| = 20 \log K, \qquad \text{for } \omega \ll 1$$

The low-frequency asymptote is a horizontal line at $20 \log K$ dB. The value of K can thus be found from this horizontal asymptote.

2. For $\lambda = 1$, or type 1 systems,

$$G(j\omega) = \frac{K}{j\omega}, \qquad \text{for } \omega \ll 1$$

or

$$20 \log|G(j\omega)| = 20 \log K - 20 \log \omega, \qquad \text{for } \omega \ll 1$$

which indicates that the low-frequency asymptote has the slope −20 dB/decade. The frequency at which the low-frequency asymptote (or its extension) intersects the 0-dB line is numerically equal to K.

3. For $\lambda = 2$, or type 2 systems,

$$G(j\omega) = \frac{K}{(j\omega)^2}, \qquad \text{for } \omega \ll 1$$

or

$$20 \log|G(j\omega)| = 20 \log K - 40 \log \omega, \qquad \text{for } \omega \ll 1$$

The slope of the low-frequency asymptote is −40 dB/decade. The frequency at which this asymptote (or its extension) intersects the 0-dB line is numerically equal to $\sqrt{K}$.

Examples of log-magnitude curves for type 0, type 1, and type 2 systems are shown in Figure 8–90, together with the frequency to which the gain K is related.

The experimentally obtained phase-angle curve provides a means of checking the transfer function obtained from the log-magnitude curve. For a minimum-phase

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

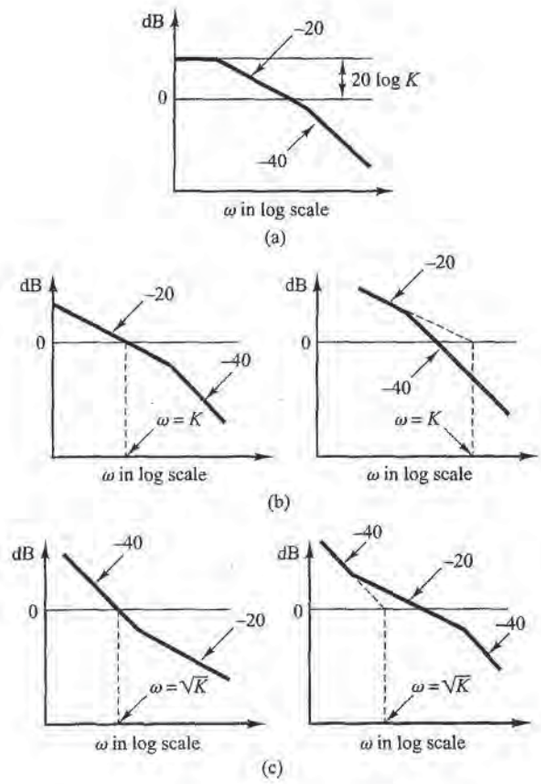


Figure 8–90
(a) Log-magnitude curve of a type 0 system; (b) log-magnitude curves of type 1 systems; (c) log-magnitude curves of type 2 systems. (The slopes shown are in dB/decade.)

system, the experimental phase-angle curve should agree reasonably well with the theoretical phase-angle curve obtained from the transfer function just determined. These two phase-angle curves should agree exactly in both the very low and very high frequency ranges. If the experimentally obtained phase angle at very high frequencies (compared with the corner frequencies) is not equal to $-90°(q - p)$, where p and q are the degrees of the numerator and denominator polynomials of the transfer function, respectively, then the transfer function must be a nonminimum-phase transfer function.

Nonminimum-Phase Transfer Functions. If, at the high-frequency end, the computed phase lag is 180° less than the experimentally obtained phase lag, then one of the zeros of the transfer function should have been in the right-half s plane instead of the left-half s plane.

If the computed phase lag differed from the experimentally obtained phase lag by a constant rate of change of phase, then transport lag, or dead time, is present. If we assume the transfer function to be of the form

$$G(s)e^{-Ts}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

where $G(s)$ is a ratio of two polynomials in s, then

$$\lim_{\omega\to\infty}\frac{d}{d\omega}\underline{/G(j\omega)e^{-j\omega T}} = \lim_{\omega\to\infty}\frac{d}{d\omega}\left[\underline{/G(j\omega)} + \underline{/e^{-j\omega T}}\right]$$

$$= \lim_{\omega\to\infty}\frac{d}{d\omega}\left[\underline{/G(j\omega)} - \omega T\right]$$

$$= 0 - T = -T$$

where we used the fact that $\lim_{\omega\to\infty}\underline{/G(j\omega)} = \text{constant}$. Thus, from this last equation, we can evaluate the magnitude of the transport lag T.

A Few Remarks on the Experimental Determination of Transfer Functions

1. It is usually easier to make accurate amplitude measurements than accurate phase-shift measurements. Phase-shift measurements may involve errors that may be caused by instrumentation or by misinterpretation of the experimental records.
2. The frequency response of measuring equipment used to measure the system output must have a nearly flat magnitude-versus-frequency curve. In addition, the phase angle must be nearly proportional to the frequency.
3. Physical systems may have several kinds of nonlinearities. Therefore, it is necessary to consider carefully the amplitude of input sinusoidal signals. If the amplitude of the input signal is too large, the system will saturate, and the frequency-response test will yield inaccurate results. On the other hand, a small signal will cause errors due to dead zone. Hence, a careful choice of the amplitude of the input sinusoidal signal must be made. It is necessary to sample the waveform of the system output to make sure that the waveform is sinusoidal and that the system is operating in the linear region during the test period. (The waveform of the system output is not sinusoidal when the system is operating in its nonlinear region.)
4. If the system under consideration is operating continuously for days and weeks, then normal operation need not be stopped for frequency-response tests. The sinusoidal test signal may be superimposed on the normal inputs. Then, for linear systems, the output due to the test signal is superimposed on the normal output. For the determination of the transfer function while the system is in normal operation, stochastic signals (white noise signals) also are often used. By use of correlation functions, the transfer function of the system can be determined without interrupting normal operation.

EXAMPLE 8–26 Determine the transfer function of the system whose experimental frequency-response curves are as shown in Figure 8–91.

The first step in determining the transfer function is to approximate the log-magnitude curve by asymptotes with slopes ±20 dB/decade and multiples thereof, as shown in Figure 8–91. We then estimate the corner frequencies. For the system shown in Figure 8–91, the following form of the transfer function is estimated:

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

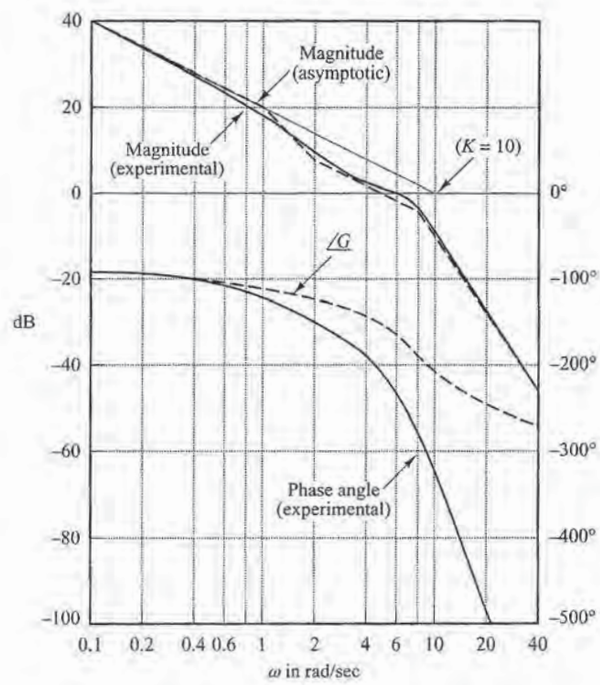


Figure 8–91
Bode diagram of a system. (Solid curves are experimentally obtained curves.)

$$G(j\omega) = \frac{K(1 + 0.5j\omega)}{j\omega(1 + j\omega)\left[1 + 2\zeta\left(j\frac{\omega}{8}\right) + \left(j\frac{\omega}{8}\right)^2\right]}$$

The value of the damping ratio ζ is estimated by examining the peak resonance near $\omega = 6$ rad/sec. Referring to Figure 8–9, ζ is determined to be 0.5. The gain K is numerically equal to the frequency at the intersection of the extension of the low-frequency asymptote and the 0-dB line. The value of K is thus found to be 10. Therefore, $G(j\omega)$ is tentatively determined as

$$G(j\omega) = \frac{10(1 + 0.5j\omega)}{j\omega(1 + j\omega)\left[1 + \left(j\frac{\omega}{8}\right) + \left(j\frac{\omega}{8}\right)^2\right]}$$

or

$$G(s) = \frac{320(s + 2)}{s(s + 1)(s^2 + 8s + 64)}$$

This transfer function is tentative because we have not examined the phase-angle curve yet.

Once the corner frequencies are noted on the log-magnitude curve, the corresponding phase-angle curve for each component factor of the transfer function can easily be drawn. The sum of these component phase-angle curves is that of the assumed transfer function. The phase-angle curve for $G(j\omega)$ is denoted by $\underline{/G}$ in Figure 8–91. From Figure 8–91, we clearly notice a discrepancy between the computed phase-angle curve and the experimentally obtained phase-angle curve. The difference between the two curves at very high frequencies appears to be a constant rate of change. Thus, the discrepancy in the phase-angle curves must be caused by transport lag.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Hence, we assume the complete transfer function to be $G(s)e^{-Ts}$. Since the discrepancy between the computed and experimental phase angles is -0.2ω rad for very high frequencies, we can determine the value of T as follows.

$$\lim_{\omega \to \infty} \frac{d}{d\omega} \underline{/G(j\omega)e^{-j\omega T}} = -T = -0.2$$

or

$$T = 0.2 \text{ sec.}$$

The presence of transport lag can thus be determined, and the complete transfer function determined from the experimental curves is

$$G(s)e^{-Ts} = \frac{320(s + 2)e^{-0.2s}}{s(s + 1)(s^2 + 8s + 64)}$$

EXAMPLE PROBLEMS AND SOLUTIONS

A–8–1. Consider a system whose closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{10(s + 1)}{(s + 2)(s + 5)}$$

(This is the same system considered in Problem **A–6–8.**) Clearly, the closed-loop poles are located at $s = -2$ and $s = -5$, and the system is not oscillatory. (The unit-step response, however, exhibits overshoot due to the presence of a zero at $s = -1$. See Figure 6–46.)

Show that the closed-loop frequency response of this system will exhibit a resonant peak, although the damping ratio of the closed-loop poles is greater than unity.

Solution. Figure 8–92 shows the Bode diagram for the system. The resonant peak value is approximately 3.5 dB. (Note that, in the absence of a zero, the second-order system with $\zeta > 0.7$ will not exhibit a resonant peak; however, the presence of a closed-loop zero will cause such a peak.)

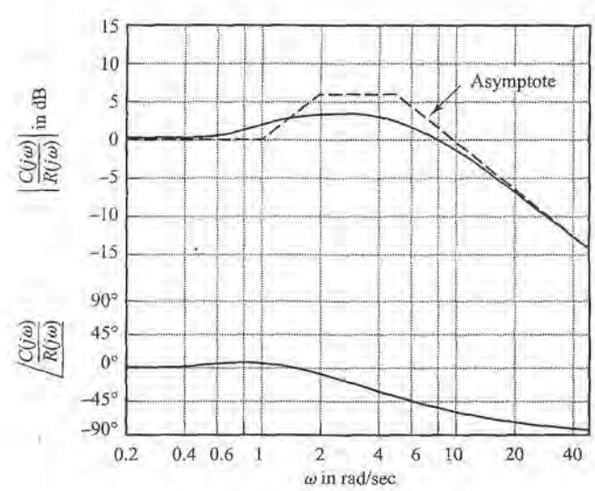


Figure 8–92
Bode diagram for $10(1 + j\omega)/[(2 + j\omega)(5 + j\omega)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–8–2. Consider the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -25 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Obtain the sinusoidal transfer functions $Y_1(j\omega)/U_1(j\omega)$, $Y_2(j\omega)/U_1(j\omega)$, $Y_1(j\omega)/U_2(j\omega)$, and $Y_2(j\omega)/U_2(j\omega)$. In deriving $Y_1(j\omega)/U_1(j\omega)$ and $Y_2(j\omega)/U_1(j\omega)$, we assume that $U_2(j\omega) = 0$. Similarly, in obtaining $Y_1(j\omega)/U_2(j\omega)$ and $Y_2(j\omega)/U_2(j\omega)$, we assume that $U_1(j\omega) = 0$.

Solution. The transfer matrix expression for the system defined by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\dot{\mathbf{y}} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

is given by

$$\mathbf{Y}(s) = \mathbf{G}(s)\mathbf{U}(s)$$

where $\mathbf{G}(s)$ is the transfer matrix and is given by

$$\mathbf{G}(s) = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D}$$

For the system considered here, the transfer matrix becomes

$$\begin{aligned} \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s & -1 \\ 25 & s+4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ &= \frac{1}{s^2 + 4s + 25} \begin{bmatrix} s+4 & 1 \\ -25 & s \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{s+4}{s^2+4s+25} & \dfrac{s+5}{s^2+4s+25} \\ \dfrac{-25}{s^2+4s+25} & \dfrac{s-25}{s^2+4s+25} \end{bmatrix} \end{aligned}$$

Hence

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} \dfrac{s+4}{s^2+4s+25} & \dfrac{s+5}{s^2+4s+25} \\ \dfrac{-25}{s^2+4s+25} & \dfrac{s-25}{s^2+4s+25} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix}$$

Assuming that $U_2(j\omega) = 0$, we find $Y_1(j\omega)/U_1(j\omega)$ and $Y_2(j\omega)/U_1(j\omega)$ as follows:

$$\frac{Y_1(j\omega)}{U_1(j\omega)} = \frac{j\omega + 4}{(j\omega)^2 + 4j\omega + 25}$$

$$\frac{Y_2(j\omega)}{U_1(j\omega)} = \frac{-25}{(j\omega)^2 + 4j\omega + 25}$$

Similarly, assuming that $U_1(j\omega) = 0$, we find $Y_1(j\omega)/U_2(j\omega)$ and $Y_2(j\omega)/U_2(j\omega)$ as follows:

$$\frac{Y_1(j\omega)}{U_2(j\omega)} = \frac{j\omega + 5}{(j\omega)^2 + 4j\omega + 25}$$

$$\frac{Y_2(j\omega)}{U_2(j\omega)} = \frac{j\omega - 25}{(j\omega)^2 + 4j\omega + 25}$$

Notice that $Y_2(j\omega)/U_2(j\omega)$ is a nonminimum-phase transfer function.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–8–3. Referring to Problem A–8–2, plot Bode diagrams for the system, using MATLAB.

Solution. MATLAB Program 8–14 produces Bode diagrams for the system. There are four sets of Bode diagrams: two for input 1 and two for input 2. These Bode diagrams are shown in Figure 8–93.

MATLAB Program 8–14

```
A = [0 1;-25 -4];
B = [1 1;0 1];
C = [1 0;0 1];
D = [0 0;0 0];
bode(A,B,C,D)
```

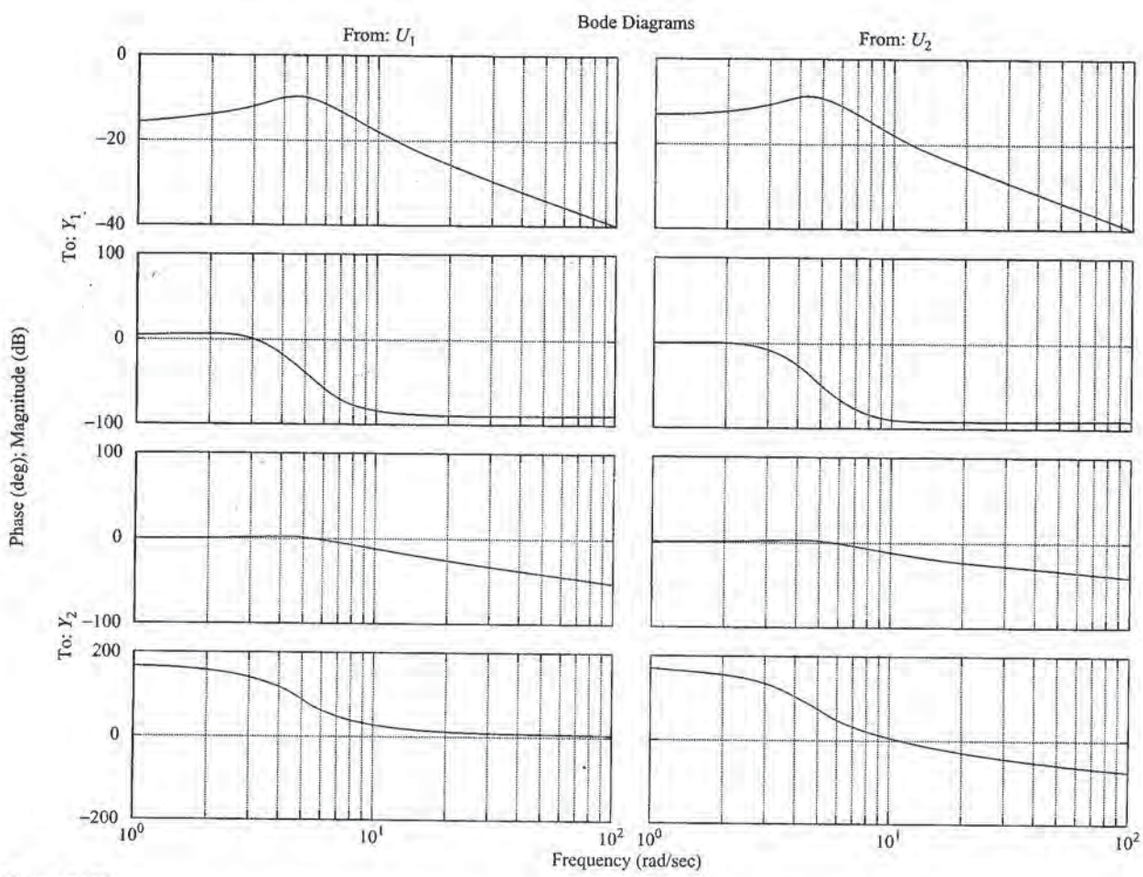


Figure 8–93
Bode diagrams.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

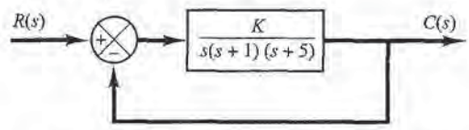


Figure 8–94
Closed-loop system.

A–8–4. Using MATLAB, plot Bode diagrams for the closed-loop system shown in Figure 8–94 for $K = 1$, $K = 10$, and $K = 20$. Plot three magnitude curves in one diagram and three phase-angle curves in another diagram.

Solution. The closed-loop transfer function of the system is given by

$$\frac{C(s)}{R(s)} = \frac{K}{s(s+1)(s+5)+K}$$
$$= \frac{K}{s^3 + 6s^2 + 5s + K}$$

Hence the numerator and denominator of $C(s)/R(s)$ are

```
num = [0 0 0 K]
den = [1 6 5 K]
```

A possible MATLAB program is shown in MATLAB Program 8–15. The resulting Bode diagrams are shown in Figures 8–95(a) and (b).

MATLAB Program 8–15

```
w = logspace(-1,2,200);
for i = 1:3;
   if i == 1; K = 1;[mag,phase,w] = bode([0 0 0 K],[1 6 5 K],w);
      mag1dB = 20*log10(mag); phase1 = phase; end;
   if i == 2; K = 10;[mag,phase,w] = bode([0 0 0 K],[1 6 5 K],w);
      mag2dB = 20*log10(mag); phase2 = phase; end;
   if i == 3; K = 20;[mag,phase,w] = bode([0 0 0 K],[1 6 5 K],w);
      mag3dB = 20*log10(mag); phase3 = phase; end;
end
semilogx(w,mag1dB,'-',w,mag2dB,'-',w,mag3dB,'-')
grid
title('Bode Diagrams of G(s) = K/[s(s + 1)(s + 5)], where K = 1, K = 10, and K = 20')
xlabel('Frequency (rad/sec)')
ylabel('Gain (dB)')
text(1.2,-31,'K = 1')
text(1.1,-8,'K = 10')
text(11,-31,'K = 20')
semilogx(w,phase1,'-',w,phase2,'-',w,phase3,'-')
grid
xlabel('Frequency (rad/sec)')
ylabel('Phase (deg)')
text(0.2,-90,'K = 1')
text(0.2,-20,'K =10')
text(1.6,-20,'K = 20')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

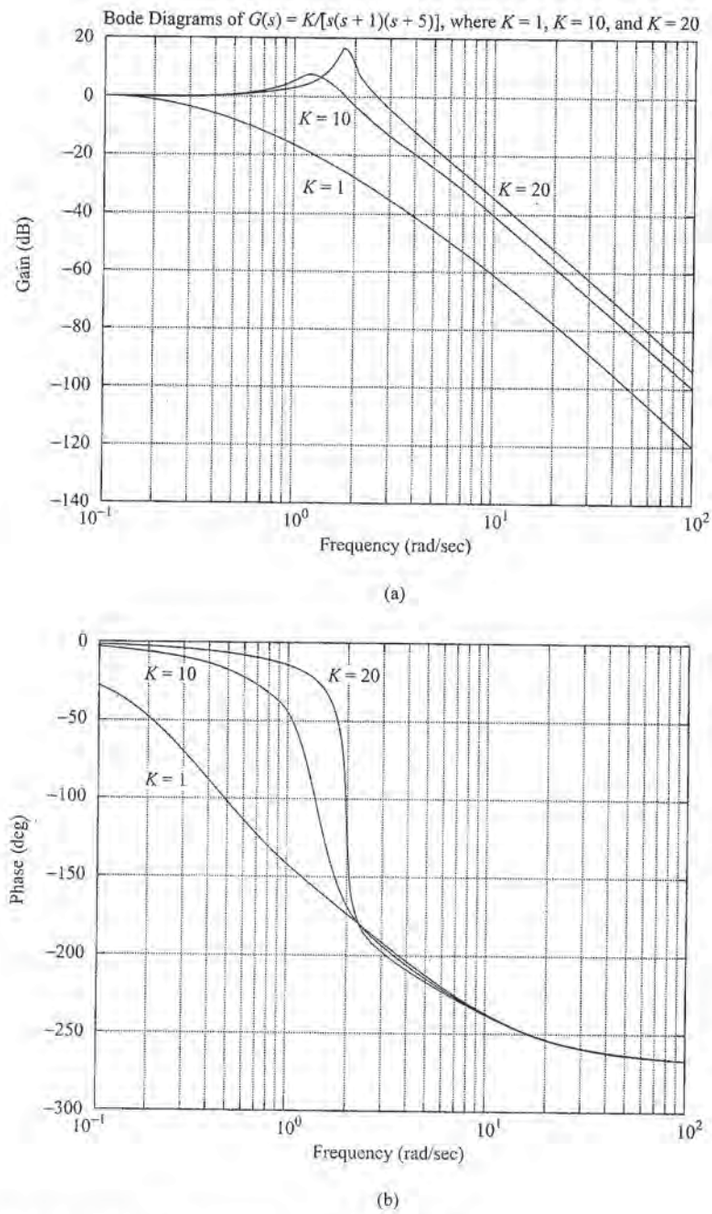


Figure 8–95
Bode diagrams: (a) Magnitude-versus-frequency curves; (b) phase-angle-versus-frequency curves.

A–8–5. Prove that the polar plot of the sinusoidal transfer function

$$G(j\omega) = \frac{j\omega T}{1 + j\omega T}, \qquad \text{for } 0 \le \omega \le \infty$$

is a semicircle. Find the center and radius of the circle.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Solution. The given sinusoidal transfer function $G(j\omega)$ can be written as follows:

$$G(j\omega) = X + jY$$

where

$$X = \frac{\omega^2 T^2}{1 + \omega^2 T^2}, \qquad Y = \frac{\omega T}{1 + \omega^2 T^2}$$

Then

$$\left(X - \frac{1}{2}\right)^2 + Y^2 = \frac{(\omega^2 T^2 - 1)^2}{4(1 + \omega^2 T^2)^2} + \frac{\omega^2 T^2}{(1 + \omega^2 T^2)^2} = \frac{1}{4}$$

Hence, we see that the plot of $G(j\omega)$ is a circle centered at (0.5,0) with radius equal to 0.5. The upper semicircle corresponds to $0 \le \omega \le \infty$, and the lower semicircle corresponds to $-\infty \le \omega \le 0$.

A–8–6. Prove the following mapping theorem: Let $F(s)$ be a ratio of polynomials in s. Let P be the number of poles and Z be the number of zeros of $F(s)$ that lie inside a closed contour in the s plane, with multiplicity accounted for. Let the closed contour be such that it does not pass through any poles or zeros of $F(s)$. The closed contour in the s plane then maps into the $F(s)$ plane as a closed curve. The number N of clockwise encirclements of the origin of the $F(s)$ plane, as a representative point s traces out the entire contour in the s plane in the clockwise direction, is equal to $Z - P$.

Solution. To prove this theorem, we use Cauchy's theorem and the residue theorem. Cauchy's theorem states that the integral of $F(s)$ around a closed contour in the s plane is zero if $F(s)$ is analytic within and on the closed contour, or

$$\oint F(s)\,ds = 0$$

Suppose that $F(s)$ is given by

$$F(s) = \frac{(s + z_1)^{k_1}(s + z_2)^{k_2}\cdots}{(s + p_1)^{m_1}(s + p_2)^{m_2}\cdots} X(s)$$

where $X(s)$ is analytic in the closed contour in the s plane and all the poles and zeros are located in the contour. Then the ratio $F'(s)/F(s)$ can be written

$$\frac{F'(s)}{F(s)} = \left(\frac{k_1}{s + z_1} + \frac{k_2}{s + z_2} + \cdots\right) - \left(\frac{m_1}{s + p_1} + \frac{m_2}{s + p_2} + \cdots\right) + \frac{X'(s)}{X(s)} \tag{8–25}$$

This may be seen from the following consideration: If $F(s)$ is given by

$$F(s) = (s + z_1)^k X(s)$$

then $F(s)$ has a zero of kth order at $s = -z_1$. Differentiating $F(s)$ with respect to s yields

$$F'(s) = k(s + z_1)^{k-1} X(s) + (s + z_1)^k X'(s)$$

Hence,

$$\frac{F'(s)}{F(s)} = \frac{k}{s + z_1} + \frac{X'(s)}{X(s)} \tag{8–26}$$

We see that by taking the ratio $F'(s)/F(s)$, the kth-order zero of $F(s)$ becomes a simple pole of $F'(s)/F(s)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

If the last term on the right-hand side of Equation (8–26) does not contain any poles or zeros in the closed contour in the s plane, $F'(s)/F(s)$ is analytic in this contour except at the zero $s = -z_1$. Then, referring to Equation (8–25) and using the residue theorem, which states that the integral of $F'(s)/F(s)$ taken in the clockwise direction around a closed contour in the s plane is equal to $-2\pi j$ times the residues at the simple poles of $F'(s)/F(s)$, or

$$\oint \frac{F'(s)}{F(s)}\,ds = -2\pi j\left(\sum \text{residues}\right)$$

we have

$$\oint \frac{F'(s)}{F(s)}\,ds = -2\pi j[(k_1 + k_2 + \cdots) - (m_1 + m_2 + \cdots)] = -2\pi j(Z - P)$$

where $Z = k_1 + k_2 + \cdots =$ total number of zeros of $F(s)$ enclosed in the closed contour in the s plane

$P = m_1 + m_2 + \cdots =$ total number of poles of $F(s)$ enclosed in the closed contour in the s plane

[The k multiple zeros (or poles) are considered k zeros (or poles) located at the same point.] Since $F(s)$ is a complex quantity, $F(s)$ can be written

$$F(s) = |F|e^{j\theta}$$

and

$$\ln F(s) = \ln|F| + j\theta$$

Noting that $F'(s)/F(s)$ can be written

$$\frac{F'(s)}{F(s)} = \frac{d\ln F(s)}{ds}$$

we obtain

$$\frac{F'(s)}{F(s)} = \frac{d\ln|F|}{ds} + j\frac{d\theta}{ds}$$

If the closed contour in the s plane is mapped into the closed contour Γ in the $F(s)$ plane, then

$$\oint \frac{F'(s)}{F(s)}\,ds = \oint_\Gamma d\ln|F| + j\oint_\Gamma d\theta = j\int d\theta = 2\pi j(P - Z)$$

The integral $\oint_\Gamma d\ln|F|$ is zero since the magnitude $\ln|F|$ is the same at the initial point and the final point of the contour Γ. Thus we obtain

$$\frac{\theta_2 - \theta_1}{2\pi} = P - Z$$

The angular difference between the final and initial values of θ is equal to the total change in the phase angle of $F'(s)/F(s)$ as a representative point in the s plane moves along the closed contour. Noting that N is the number of clockwise encirclements of the origin of the $F(s)$ plane and $\theta_2 - \theta_1$ is zero or a multiple of 2π rad, we obtain

$$\frac{\theta_2 - \theta_1}{2\pi} = -N$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

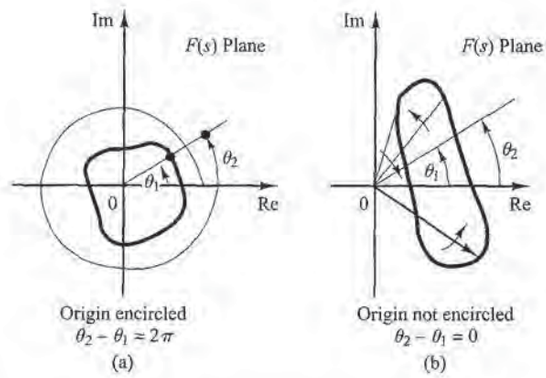


Figure 8–96
Determination of encirclement of the origin of $F(s)$ plane.

Thus, we have the relationship

$$N = Z - P$$

This proves the theorem.

Note that by this mapping theorem, the exact numbers of zeros and of poles cannot be found—only their difference. Note also that, from Figures 8–96(a) and (b), we see that if θ does not change through 2π rad, then the origin of the $F(s)$ plane cannot be encircled.

A–8–7. The Nyquist plot (polar plot) of the open-loop frequency response of a unity-feedback control system is shown in Figure 8–97(a). Assuming that the Nyquist path in the s plane encloses the entire right-half s plane, draw a complete Nyquist plot in the G plane. Then answer the following questions:

(a) If the open-loop transfer function has no poles in the right-half s plane, is the closed-loop system stable?

(b) If the open-loop transfer function has one pole and no zeros in right-half s plane, is the closed-loop system stable?

(c) If the open-loop transfer function has one zero and no poles in the right-half s plane, is the closed-loop system stable?

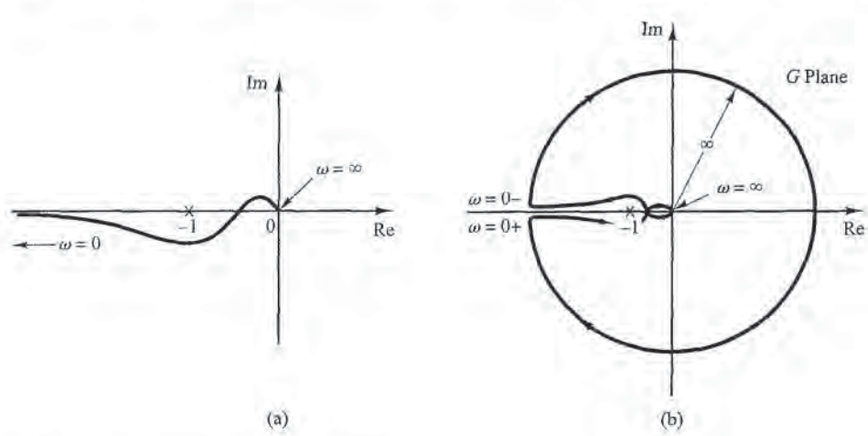


Figure 8–97
(a) Nyquist plot;
(b) complete Nyquist plot in the G plane.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Solution. Figure 8–97(b) shows a complete Nyquist plot in the G plane. The answers to the three questions are as follows:

(a) The closed-loop system is stable, because the critical point $(-1 + j0)$ is not encircled by the Nyquist plot. That is, since $P = 0$ and $N = 0$, we have $Z = N + P = 0$.

(b) The open-loop transfer function has one pole in the right-half s plane. Hence, $P = 1$. (The open-loop system is unstable.) For the closed-loop system to be stable, the Nyquist plot must encircle the critical point $(-1 + j0)$ once counterclockwise. However, the Nyquist plot does not encircle the critical point. Hence, $N = 0$. Therefore, $Z = N + P = 1$. The closed-loop system is unstable.

(c) Since the open-loop transfer function has one zero, but no poles, in the right-half s plane, we have $Z = N + P = 0$. Thus, the closed-loop system is stable. (Note that the zeros of the open-loop transfer function do not affect the stability of the closed-loop system.)

A–8–8. Is a closed-loop system with the following open-loop transfer function and with $K = 2$ stable?

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

Find the critical value of the gain K for stability.

Solution. The open-loop transfer function is

$$\begin{aligned} G(j\omega)H(j\omega) &= \frac{K}{j\omega(j\omega+1)(2j\omega+1)} \\ &= \frac{K}{-3\omega^2 + j\omega(1-2\omega^2)} \end{aligned}$$

This open-loop transfer function has no poles in the right-half s plane. Thus, for stability, the $-1 + j0$ point should not be encircled by the Nyquist plot. Let us find the point where the Nyquist plot crosses the negative real axis. Let the imaginary part of $G(j\omega)H(j\omega)$ be zero, or

$$1 - 2\omega^2 = 0$$

from which

$$\omega = \pm\frac{1}{\sqrt{2}}$$

Substituting $\omega = 1/\sqrt{2}$ into $G(j\omega)H(j\omega)$, we obtain

$$G\left(j\frac{1}{\sqrt{2}}\right)H\left(j\frac{1}{\sqrt{2}}\right) = -\frac{2K}{3}$$

The critical value of the gain K is obtained by equating $-2K/3$ to -1, or

$$-\frac{2}{3}K = -1$$

Hence,

$$K = \frac{3}{2}$$

The system is stable if $0 < K < \frac{3}{2}$. Hence, the system with $K = 2$ is unstable.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

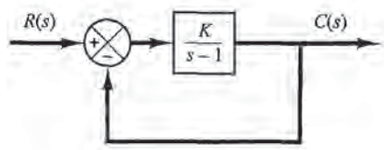


Figure 8–98
Closed-loop system.

A–8–9. Consider the closed-loop system shown in Figure 8–98. Determine the critical value of K for stability by the use of the Nyquist stability criterion.

Solution. The polar plot of

$$G(j\omega) = \frac{K}{j\omega - 1}$$

is a circle with center at $-K/2$ on the negative real axis and radius $K/2$, as shown in Figure 8–99(a). As ω is increased from $-\infty$ to ∞, the $G(j\omega)$ locus makes a counterclockwise rotation. In this system, $P = 1$ because there is one pole of $G(s)$ in the right-half s plane. For the closed-loop system to be stable, Z must be equal to zero. Therefore, $N = Z - P$ must be equal to -1, or there must be one counterclockwise encirclement of the $-1 + j0$ point for stability. (If there is no encirclement of the $-1 + j0$ point, the system is unstable.) Thus, for stability, K must be greater than unity, and $K = 1$ gives the stability limit. Figure 8–99(b) shows both stable and unstable cases of $G(j\omega)$ plots.

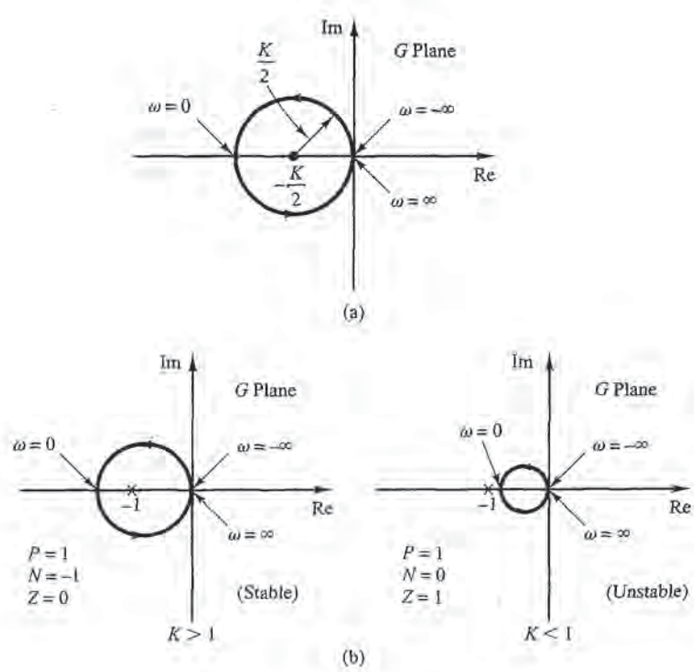


Figure 8–99
(a) Polar plot of $K/(j\omega - 1)$;
(b) polar plots of $K/(j\omega - 1)$ for stable and unstable cases.

VW EX1032 (Part 4 of 5)
VW v. MMT
IPR2020-00445

0611

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–8–10. Consider a unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{Ke^{-0.8s}}{s+1}$$

Using the Nyquist plot, determine the critical value of K for stability.

Solution. For this system,

$$\begin{aligned} G(j\omega) &= \frac{Ke^{-0.8j\omega}}{j\omega + 1} \\ &= \frac{K(\cos 0.8\omega - j\sin 0.8\omega)(1 - j\omega)}{1 + \omega^2} \\ &= \frac{K}{1+\omega^2}\left[(\cos 0.8\omega - \omega \sin 0.8\omega) - j(\sin 0.8\omega + \omega\cos 0.8\omega)\right] \end{aligned}$$

The imaginary part of $G(j\omega)$ is equal to zero if

$$\sin 0.8\omega + \omega \cos 0.8\omega = 0$$

Hence,

$$\omega = -\tan 0.8\omega$$

Solving this equation for the smallest positive value of ω, we obtain

$$\omega = 2.4482$$

Substituting $\omega = 2.4482$ into $G(j\omega)$, we obtain

$$G(j2.4482) = \frac{K}{1 + 2.4482^2}(\cos 1.9586 - 2.4482 \sin 1.9586) = -0.378K$$

The critical value of K for stability is obtained by letting $G(j2.4482)$ equal -1. Hence,

$$0.378K = 1$$

or

$$K = 2.65$$

Figure 8–100 shows the Nyquist or polar plots of $2.65e^{-0.8j\omega}/(1 + j\omega)$ and $2.65/(1 + j\omega)$. The first-order system without transport lag is stable for all values of K, but the one with a transport lag of 0.8 sec becomes unstable for $K > 2.65$.

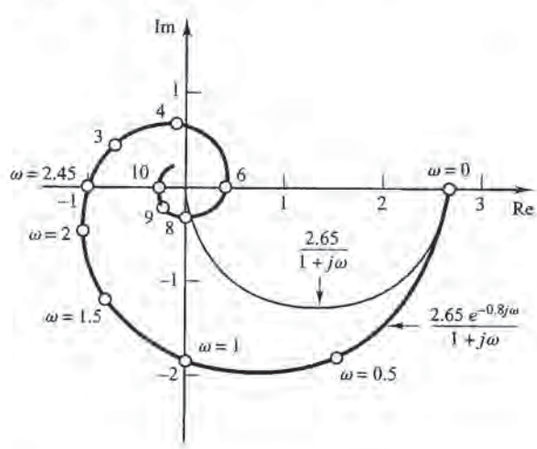


Figure 8–100
Polar plots of $2.65e^{-0.8j\omega}/(1 + j\omega)$ and $2.65/(1 + j\omega)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–8–11. Consider a unity-feedback system with the following open-loop transfer function:

$$G(s) = \frac{20(s^2 + s + 0.5)}{s(s + 1)(s + 10)}$$

Draw a Nyquist plot with MATLAB and examine the stability of the closed-loop system.

Solution. MATLAB Program 8–16 produces the Nyquist diagram shown in Figure 8–101. From this figure, we see that the Nyquist plot does not encircle the $-1 + j0$ point. Hence, $N = 0$ in the Nyquist stability criterion. Since no open-loop poles lie in the right-half s plane, $P = 0$. Therefore, $Z = N + P = 0$. The closed-loop system is stable.

MATLAB Program 8–16

```
num = [0 20 20 10];
den = [1 11 10 0];
nyquist(num,den)
v = [-2 3 -3 3]; axis(v)
grid
```

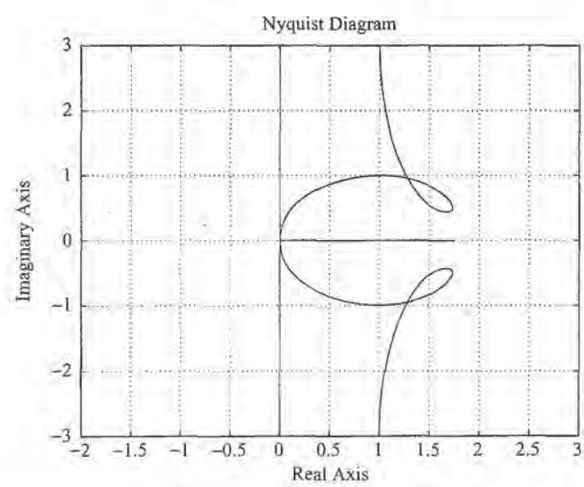


Figure 8–101
Nyquist plot of
$G(s) = \dfrac{20(s^2 + s + 0.5)}{s(s + 1)(s + 10)}$.

A–8–12. Consider the same system as discussed in Problem **A–8–11.** Draw the Nyquist plot for only the positive frequency region.

Solution. Drawing a Nyquist plot for only the positive frequency region can be done by the use of the following command:

```
[re,im,w] = nyquist(num,den,w)
```

The frequency region may be divided into several subregions by using different increments. For example, the frequency region of interest may be divided into three subregions as follows:

```
w1 = 0.1:0.1:10;
w2 = 10:2:100;
w3 = 100:10:500;
w  = [w1 w2 w3]
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 8–17 uses this frequency region. Using this program, we obtain the Nyquist plot shown in Figure 8–102.

MATLAB Program 8–17

```
num = [0  20  20  10];
den = [1  11  10  0];
w1 = 0.1:0.1:10; w2 = 10:2:100; w3 = 100:10:500;
w = [w1  w2  w3];
[re,im,w] = nyquist(num,den,w);
plot(re,im)
v = [-3  3  -5  1]; axis(v)
grid
title('Nyquist Plot of G(s) = 20(s^2 + s + 0.5)/[s(s + 1)(s + 10)]')
xlabel('Real Axis')
ylabel('Imag Axis')
```

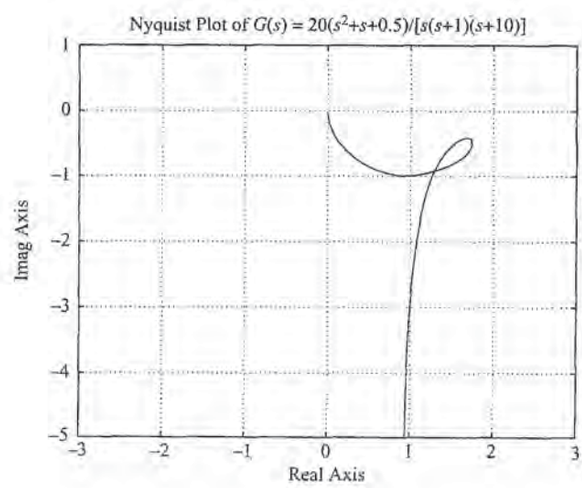


Figure 8–102
Nyquist plot for the positive frequency region.

A–8–13. Referring to Problem **A–8–12**, plot the polar locus of $G(s)$ where

$$G(s) = \frac{20(s^2 + s + 0.5)}{s(s + 1)(s + 10)}$$

Locate on the polar locus frequency points where $\omega = 0.2, 0.3, 0.5, 1, 2, 6, 10$, and 20 rad/sec. Also, find the magnitudes and phase angles of $G(j\omega)$ at the specified frequency points.

Solution. In MATLAB Program 8–17 we used the frequency vector w, which consists of three frequency subvectors: w1, w2, and w3. Instead of such a w, we may simply use the

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

frequency vector w = logscale(d_1, d_2, n). MATLAB Program 8–18 uses the following frequency vector:

w = logscale(-1,2,100)

This MATLAB program plots the polar locus and locates the specified frequency points on the polar locus, as shown in Figure 8–103.

MATLAB Program 8–18

```
num = [0 20 20 10];
den = [1 11 10 0];
ww = logspace(-1,2,100);
nyquist(num,den,ww)
v = [-2 3 -5 0]; axis(v)
grid
hold
Current plot held
w = [0.2 0.3 0.5 1 2 6 10 20];
[re,im,w] = nyquist(num,den,w);
plot(re,im,'o')
text(1.1,-4.8,'w = 0.2')
text(1.1,-3.1,'0.3')
text(1.25,-1.7,'0.5')
text(1.37,-0.4,'1')
text(1.8,-0.3,'2')
text(1.4,-1.1,'6')
text(0.77,-0.8,'10')
text(0.037,-0.8,'20')

% ----- To get the values of magnitude and phase (in degrees) of G(jw)
% at the specified w values, enter the command [mag,phase,w]
% = bode(num,den,w) ------

[mag,phase,w] = bode(num,den,w);

% ----- The following table shows the specified frequency values w and
% the corresponding values of magnitude and phase (in degrees) -----

[w mag phase]

ans =

   0.2000  4.9176  -78.9571
   0.3000  3.2426  -72.2244
   0.5000  1.9975  -55.9925
   1.0000  1.5733  -24.1455
   2.0000  1.7678  -14.4898
   6.0000  1.6918  -31.0946
  10.0000  1.4072  -45.0285
  20.0000  0.8933  -63.4385
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

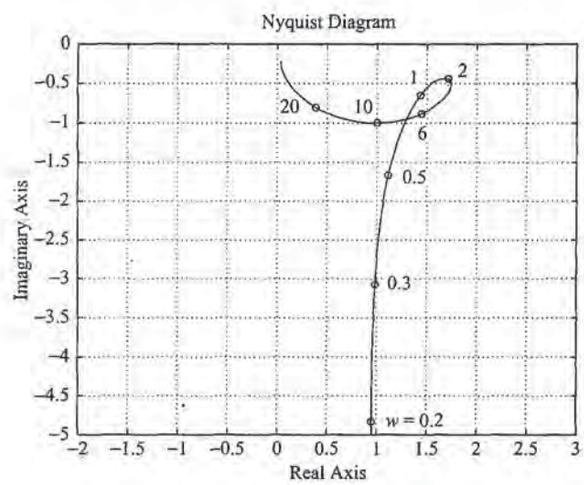


Figure 8–103
Polar plot of $G(j\omega)$ given in Problem **A–8–13**.

A–8–14. Consider a unity-feedback, positive-feedback system with the following open-loop transfer function:

$$G(s) = \frac{s^2 + 4s + 6}{s^2 + 5s + 4}$$

Draw a Nyquist plot.

Solution. The Nyquist plot of the positive-feedback system can be obtained by defining num and den as

```
num = [-1  -4  -6]
den = [1  5  4]
```

and using the command nyquist(num,den). MATLAB Program 8–19 produces the Nyquist plot, as shown in Figure 8–104.

This system is unstable, because the $-1 + j0$ point is encircled once clockwise. Note that this is a special case where the Nyquist plot passes through $-1 + j0$ point and also encircles this point once clockwise. This means that the closed-loop system is degenerate; the system behaves as if it is an unstable first-order system. See the following closed-loop transfer function of the positive-feedback system:

$$\frac{C(s)}{R(s)} = \frac{s^2 + 4s + 6}{s^2 + 5s + 4 - (s^2 + 4s + 6)}$$
$$= \frac{s^2 + 4s + 6}{s - 2}$$

MATLAB Program 8–19

```
num = [-1  -4  -6];
den = [1  5  4];
nyquist(num,den);
grid
title('Nyquist Plot of G(s) = -(s^2 + 4s + 6)/(s^2 + 5s + 4)')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

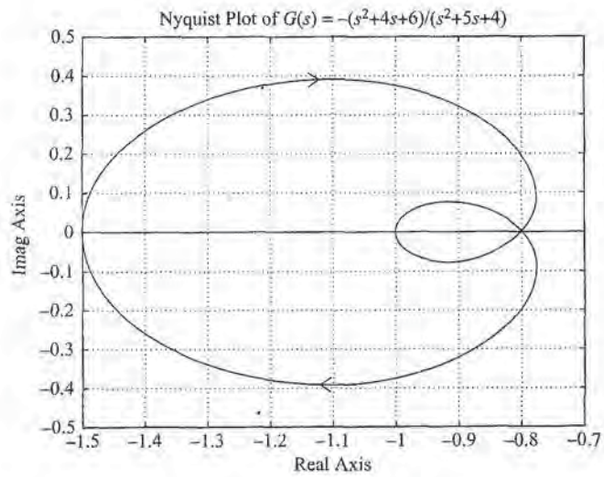


Figure 8–104
Nyquist plot for positive-feedback system.

Note that the Nyquist plot for the positive-feedback case is a mirror image about the imaginary axis of the Nyquist plot for the negative-feedback case. This may be seen from Figure 8–105, which was obtained by use of MATLAB Program 8–20. (Note that the positive-feedback case is unstable, but the negative-feedback case is stable.)

MATLAB Program 8–20

```
num1 = [1  4  6];
den1 = [1  5  4];
num2 = [-1  -4  -6];
den2 = [1  5  4];
nyquist(num1,den1);
hold on
nyquist(num2,den2);
v = [-2  2  -1  1];
axis(v);
grid
title('Nyquist Plots of G(s) and -G(s)')
text(1.0,0.5,'G(s)')
text(0.57,-0.48,'Use this Nyquist')
text(0.57,-0.61,'plot for negative')
text(0.57,-0.73,'feedback system')
text(-1.3,0.5,'-G(s)')
text(-1.7,-0.48,'Use this Nyquist')
text(-1.7,-0.61,'plot for positive')
text(-1.7,-0.73,'feedback system')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

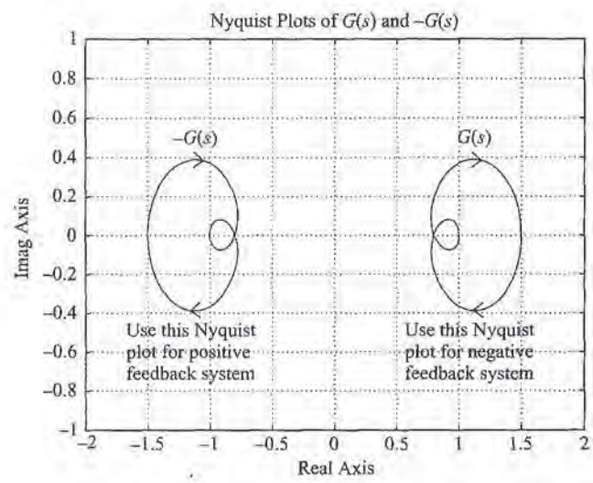


Figure 8–105
Nyquist plots for positive-feedback system and negative-feedback system.

A–8–15. Figure 8–106 shows a block diagram of a space vehicle control system. Determine the gain K such that the phase margin is 50°. What is the gain margin in this case?

Solution. Since

$$G(j\omega) = \frac{K(j\omega + 2)}{(j\omega)^2}$$

we have

$$\underline{/G(j\omega)} = \underline{/j\omega + 2} - 2\underline{/j\omega} = \tan^{-1}\frac{\omega}{2} - 180°$$

The requirement that the phase margin be 50° means that $\underline{/G(j\omega_c)}$ must be equal to −130°, where ω_c is the gain crossover frequency, or

$$\underline{/G(j\omega_c)} = -130°$$

Hence, we set

$$\tan^{-1}\frac{\omega_c}{2} = 50°$$

from which we obtain

$$\omega_c = 2.3835 \text{ rad/sec}$$

Since the phase curve never crosses the −180° line, the gain margin is +∞ dB. Noting that the magnitude of $G(j\omega)$ must be equal to 0 dB at $\omega = 2.3835$, we have

$$\left|\frac{K(j\omega + 2)}{(j\omega)^2}\right|_{\omega=2.3835} = 1$$

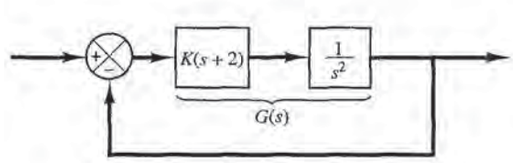


Figure 8–106
Space vehicle control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

from which we get

$$K = \frac{2.3835^2}{\sqrt{2^2 + 2.3835^2}} = 1.8259$$

This K value will give the phase margin of 50°.

A–8–16. For the standard second-order system

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

show that the bandwidth ω_b is given by

$$\omega_b = \omega_n\left(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2}\right)^{1/2}$$

Note that ω_b/ω_n is a function only of ζ. Plot a curve of ω_b/ω_n versus ζ.

Solution. The bandwidth ω_b is determined from $|C(j\omega_b)/R(j\omega_b)| = -3$ dB. Quite often, instead of -3 dB, we use -3.01 dB, which is equal to 0.707. Thus,

$$\left|\frac{C(j\omega_b)}{R(j\omega_b)}\right| = \left|\frac{\omega_n^2}{(j\omega_b)^2 + 2\zeta\omega_n(j\omega_b) + \omega_n^2}\right| = 0.707$$

Then

$$\frac{\omega_n^2}{\sqrt{(\omega_n^2 - \omega_b^2)^2 + (2\zeta\omega_n\omega_b)^2}} = 0.707$$

from which we get

$$\omega_n^4 = 0.5\left[(\omega_n^2 - \omega_b^2)^2 + 4\zeta^2\omega_n^2\omega_b^2\right]$$

By dividing both sides of this last equation by ω_n^4, we obtain

$$1 = 0.5\left\{\left[1 - \left(\frac{\omega_b}{\omega_n}\right)^2\right]^2 + 4\zeta^2\left(\frac{\omega_b}{\omega_n}\right)^2\right\}$$

Solving this last equation for $(\omega_b/\omega_n)^2$ yields

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

Since $(\omega_b/\omega_n)^2 > 0$, we take the plus sign in this last equation. Then

$$\omega_b^2 = \omega_n^2\left(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2}\right)$$

or

$$\omega_b = \omega_n\left(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2}\right)^{1/2}$$

Figure 8–107 shows a curve relating ω_b/ω_n versus ζ.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

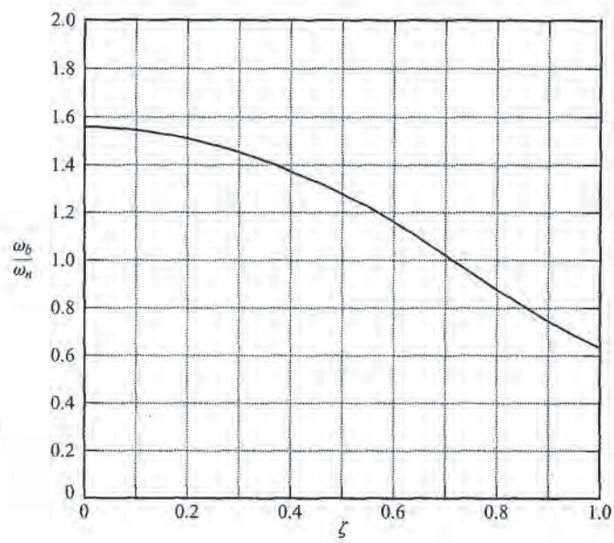


Figure 8–107
Curve of ω_b/ω_n versus ζ, where ω_b is the bandwidth.

A–8–17. Consider the control system shown in Figure 8–108. The open-loop transfer function $G(s)H(s)$ is

$$G(s)H(s) = \frac{10K}{s(s+5)(0.1s+1)}$$

Plot Nyquist diagrams of $G(s)H(s)$ for $K = 1, 7.5$, and 20.

Solution. A possible MATLAB program is shown in MATLAB Program 8–21. The resulting Nyquist diagrams are shown in Figure 8–109. From the diagrams, we see that the system is stable when $K = 1$. It is critically stable when $K = 7.5$. The system is unstable when $K = 20$.

A–8–18. Figure 8–110 shows a block diagram of a chemical reactor system. Draw a Bode diagram of $G(j\omega)$. Also, draw the $G(j\omega)$ locus on the Nichols chart. From the Nichols diagram, read magnitudes and phase angles of the closed-loop frequency response and then plot the Bode diagram of the closed-loop system $G(j\omega)/[1 + G(j\omega)]$.

Solution. Noting that

$$G(s) = \frac{80e^{-0.1s}}{s(s+4)(s+10)} = \frac{2e^{-0.1s}}{s(0.25s+1)(0.1s+1)}$$

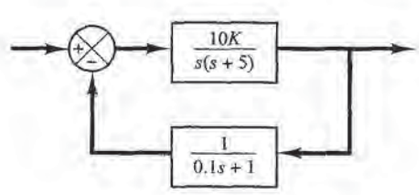


Figure 8–108
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 8–21

```
den = [0.1  1.5  5  0];
for i = 1:3;
   if i == 1; K = 1; [re1,im1] = nyquist([0  0  0  10*K],den); end;
   if i == 2; K = 7.5; [re2,im2] = nyquist([0  0  0  10*K],den); end;
   if i == 3; K = 20; [re3,im3] = nyquist([0  0  0  10*K],den); end;
end
plot(re1,im1,'-',re2,im2,'o',re2,im2,'-',re3,im3,'x',re3,im3,'-')
v = [-5  1  -5  1]; axis(v)
grid
title('Nyquist Diagrams of G(s)H(s) = 10K/[s(s + 5)(0.1s + 1)] for K = 1, 7.5, and 20')
xlabel('Real Axis')
ylabel('Imaginary Axis')
text(-0.4,-3.7,'K = 1')
text(-2.7,-2.7,'K = 7.5')
text(-4.35,-1.7,'K = 20')
```

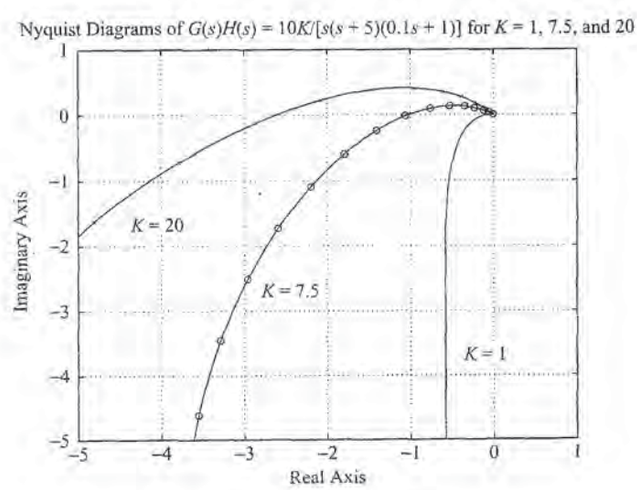


Figure 8–109
Nyquist diagrams.

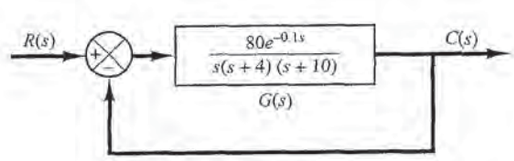


Figure 8–110
Block diagram of a chemical reactor system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

we have

$$G(j\omega) = \frac{2e^{-0.1j\omega}}{j\omega(0.25j\omega + 1)(0.1j\omega + 1)}$$

The phase angle of the transport lag $e^{-0.1j\omega}$ is

$$\underline{/e^{-0.1j\omega}} = \underline{/\cos(0.1\omega) - j\sin(0.1\omega)} = -0.1\omega \qquad \text{(rad)}$$

$$= -5.73\omega \qquad \text{(degrees)}$$

The Bode diagram of $G(j\omega)$ is shown in Figure 8–111.

Next, by reading magnitudes and phase angles of $G(j\omega)$ for various values of ω, it is possible to plot the gain-versus-phase plot on a Nichols chart. Figure 8–112 shows such a $G(j\omega)$ locus superimposed on the Nichols chart. From this diagram, magnitudes and phase angles of the closed-loop system at various frequency points can be read. Figure 8–113 depicts the Bode diagram of the closed-loop frequency response $G(j\omega)/[1 + G(j\omega)]$.

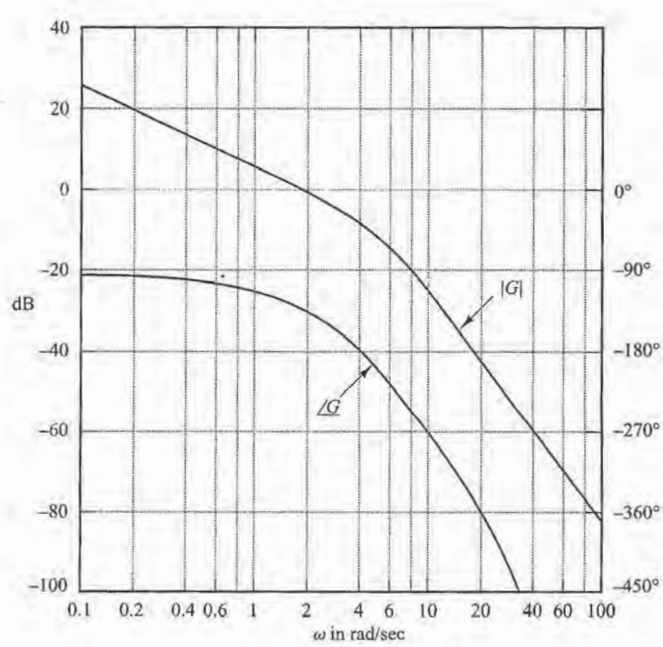


Figure 8–111
Bode diagram of $G(j\omega)$ of the system shown in Figure 8–110.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

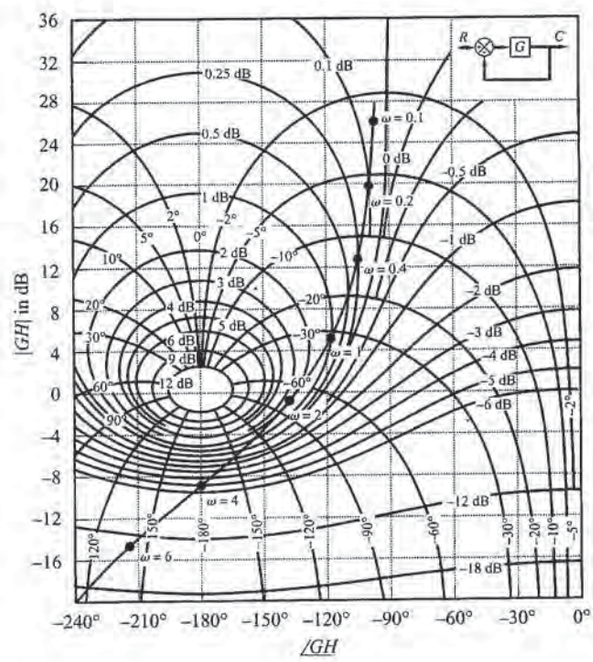


Figure 8–112
$G(j\omega)$ locus superimposed on Nichols chart (Problem **A–8–18**).

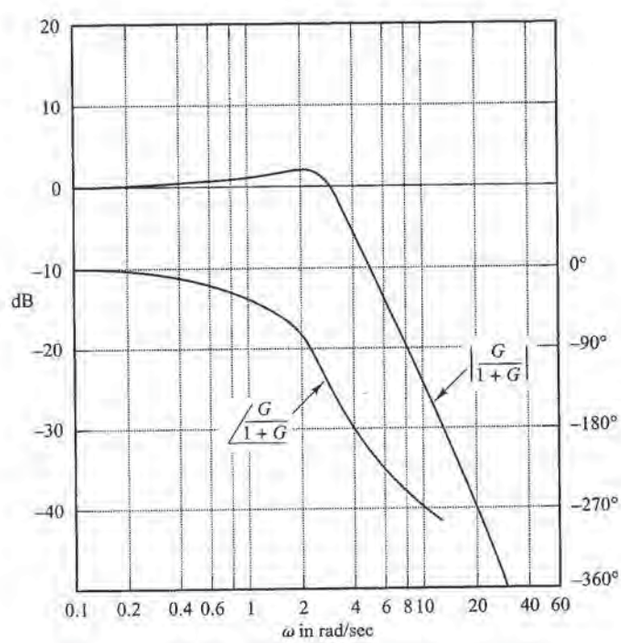


Figure 8–113
Bode diagram of the closed-loop frequency response (Problem **A–8–18**).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–8–19. A Bode diagram of the open-loop transfer function $G(s)$ of a unity-feedback control system is shown in Figure 8–114. It is known that the open-loop transfer function is minimum phase. From the diagram, it can be seen that there is a pair of complex-conjugate poles at $\omega = 2$ rad/sec. Determine the damping ratio of the quadratic term involving these complex-conjugate poles. Also, determine the transfer function $G(s)$.

Solution. Referring to Figure 8–9 and examining the Bode diagram of Figure 8–114, we find the damping ratio ζ and undamped natural frequency ω_n of the quadratic term to be

$$\zeta = 0.1, \qquad \omega_n = 2 \text{ rad/sec}$$

Noting that there is another corner frequency at $\omega = 0.5$ rad/sec and the slope of the magnitude curve in the low-frequency region is -40 dB/decade, $G(j\omega)$ can be tentatively determined as follows:

$$G(j\omega) = \frac{K\left(\dfrac{j\omega}{0.5} + 1\right)}{(j\omega)^2\left[\left(\dfrac{j\omega}{2}\right)^2 + 0.1(j\omega) + 1\right]}$$

Since, from Figure 8–114, we find $|G(j0.1)| = 40$ dB, the gain value K can be determined to be unity. Also, the calculated phase curve, $\underline{/G(j\omega)}$ versus ω, agrees with the given phase curve. Hence, the transfer function $G(s)$ can be determined to be

$$G(s) = \frac{4(2s + 1)}{s^2(s^2 + 0.4s + 4)}$$

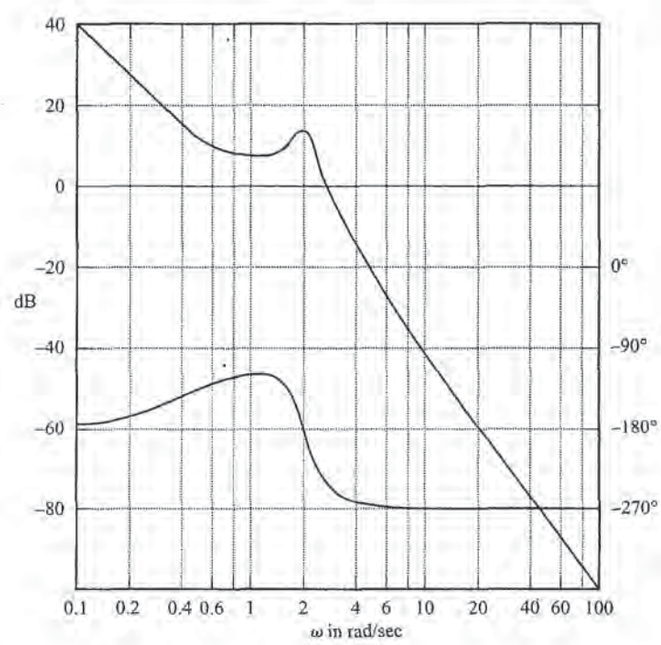


Figure 8–114
Bode diagram of the open-loop transfer function of a unity-feedback control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

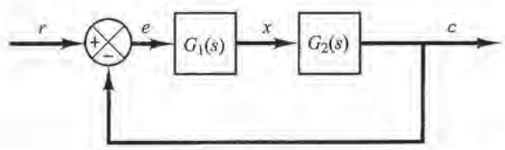


Figure 8–115
Control system.

A–8–20. A closed-loop control system may include an unstable element within the loop. When the Nyquist stability criterion is to be applied to such a system, the frequency-response curves for the unstable element must be obtained.

How can we obtain experimentally the frequency-response curves for such an unstable element? Suggest a possible approach to the experimental determination of the frequency response of an unstable linear element.

Solution. One possible approach is to measure the frequency-response characteristics of the unstable element by using it as a part of a stable system.

Consider the system shown in Figure 8–115. Suppose that the element $G_1(s)$ is unstable. The complete system may be made stable by choosing a suitable linear element $G_2(s)$. We apply a sinusoidal signal at the input. At steady state, all signals in the loop will be sinusoidal. We measure the signals $e(t)$, the input to the unstable element, and $x(t)$, the output of the unstable element. By changing the frequency [and possibly the amplitude for the convenience of measuring $e(t)$ and $x(t)$] of the input sinusoid and repeating this process, it is possible to obtain the frequency response of the unstable linear element.

PROBLEMS

B–8–1. Consider the unity-feedback system with the open-loop transfer function:

$$G(s) = \frac{10}{s+1}$$

Obtain the steady-state output of the system when it is subjected to each of the following inputs:

(a) $r(t) = \sin(t + 30°)$

(b) $r(t) = 2\cos(2t - 45°)$

(c) $r(t) = \sin(t + 30°) - 2\cos(2t - 45°)$

B–8–2. Consider the system whose closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K(T_2 s + 1)}{T_1 s + 1}$$

Obtain the steady-state output of the system when it is subjected to the input $r(t) = R\sin\omega t$.

B–8–3. Using MATLAB, plot Bode diagrams of $G_1(s)$ and $G_2(s)$ given below.

$$G_1(s) = \frac{1+s}{1+2s}$$

$$G_2(s) = \frac{1-s}{1+2s}$$

$G_1(s)$ is a minimum-phase system and $G_2(s)$ is a nonminimum-phase system.

B–8–4. Sketch the Bode diagrams of the following three transfer functions:

(a) $G(s) = \dfrac{T_1 s + 1}{T_2 s + 1} \qquad (T_1 > T_2 > 0)$

(b) $G(s) = \dfrac{T_1 s - 1}{T_2 s + 1} \qquad (T_1 > T_2 > 0)$

(c) $G(s) = \dfrac{-T_1 s + 1}{T_2 s + 1} \qquad (T_1 > T_2 > 0)$

B–8–5. Plot the Bode diagram of

$$G(s) = \frac{10(s^2 + 0.4s + 1)}{s(s^2 + 0.8s + 9)}$$

B–8–6. Given

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

show that

$$|G(j\omega_n)| = \frac{1}{2\zeta}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–8–7. Consider a unity-feedback control system with the following open-loop transfer function:

$$G(s) = \frac{s + 0.5}{s^3 + s^2 + 1}$$

This is a nonminimum-phase system. Two of the three open-loop poles are located in the right-half s plane as follows:

$$\text{Open-loop poles at } s = -1.4656$$
$$s = 0.2328 + j0.7926$$
$$s = 0.2328 - j0.7926$$

Plot the Bode diagram of $G(s)$ with MATLAB. Explain why the phase-angle curve starts from 0° and approaches +180°.

B–8–8. Sketch the polar plots of the open-loop transfer function

$$G(s)H(s) = \frac{K(T_a s + 1)(T_b s + 1)}{s^2(Ts + 1)}$$

for the following two cases:

(a) $T_a > T > 0, \quad T_b > T > 0$

(b) $T > T_a > 0, \quad T > T_b > 0$

B–8–9. The pole-zero configurations of complex functions $F_1(s)$ and $F_2(s)$ are shown in Figures 8–116(a) and (b), respectively. Assume that the closed contours in the s plane are those shown in Figures 8–116(a) and (b). Sketch qualitatively the corresponding closed contours in the $F_1(s)$ plane and $F_2(s)$ plane.

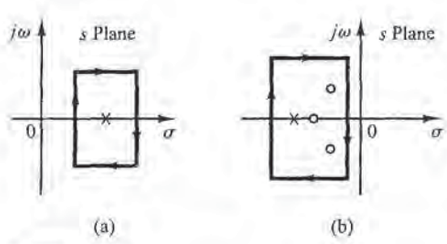


Figure 8–116
(a) s-Plane representation of complex function $F_1(s)$ and a closed contour; (b) s-Plane representation of complex function $F_2(s)$ and a closed contour.

B–8–10. Draw a Nyquist locus for the unity-feedback control system with the open-loop transfer function

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

Using the Nyquist stability criterion, determine the stability of the closed-loop system.

B–8–11. A system with the open-loop transfer function

$$G(s)H(s) = \frac{K}{s^2(T_1 s + 1)}$$

is inherently unstable. This system can be stabilized by adding derivative control. Sketch the polar plots for the open-loop transfer function with and without derivative control.

B–8–12. Consider the closed-loop system with the following open-loop transfer function:

$$G(s)H(s) = \frac{10K(s + 0.5)}{s^2(s + 2)(s + 10)}$$

Plot both the direct and inverse polar plots of $G(s)H(s)$ with $K = 1$ and $K = 10$. Apply the Nyquist stability criterion to the plots, and determine the stability of the system with these values of K.

B–8–13. Consider the closed-loop system whose open-loop transfer function is

$$G(s)H(s) = \frac{Ke^{-2s}}{s}$$

Find the maximum value of K for which the system is stable.

B–8–14. Draw a Nyquist plot of the following $G(s)$:

$$G(s) = \frac{1}{s(s^2 + 0.8s + 1)}$$

B–8–15. Consider a unity-feedback control system with the following open-loop transfer function:

$$G(s) = \frac{1}{s^3 + 0.2s^2 + s + 1}$$

Draw a Nyquist plot of $G(s)$ and examine the stability of the system.

B–8–16. Consider a unity-feedback control system with the following open-loop transfer function:

$$G(s) = \frac{s^2 + 2s + 1}{s^3 + 0.2s^2 + s + 1}$$

Draw a Nyquist plot of $G(s)$ and examine the stability of the closed-loop system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–8–17. Consider the unity-feedback system with the following $G(s)$:

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

Suppose that we choose the Nyquist path as shown in Figure 8–117. Draw the corresponding $G(j\omega)$ locus in the $G(s)$ plane. Using the Nyquist stability criterion, determine the stability of the system.

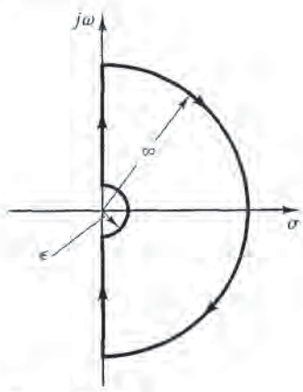


Figure 8–117
Nyquist path.

B–8–18. Consider the closed-loop system shown in Figure 8–118. $G(s)$ has no poles in the right-half s plane.

If the Nyquist plot of $G(s)$ is as shown in Figure 8–119(a), is this system stable?

If the Nyquist plot is as shown in Figure 8–119(b), is this system stable?

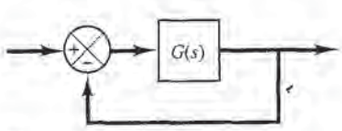


Figure 8–118
Closed-loop system.

B–8–19. A Nyquist plot of a unity-feedback system with the feedforward transfer function $G(s)$ is shown in Figure 8–120.

If $G(s)$ has one pole in the right-half s plane, is the system stable?

If $G(s)$ has no pole in the right-half s plane, but has one zero in the right-half s plane, is the system stable?

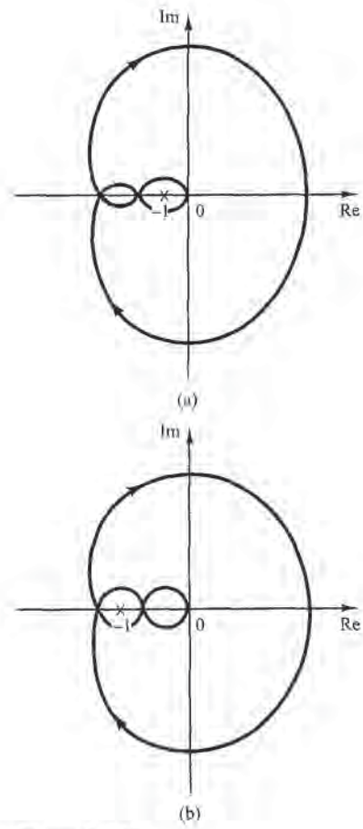


Figure 8–119
Nyquist plots.

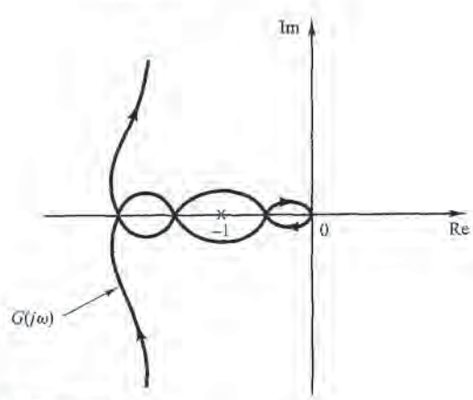


Figure 8–120
Nyquist plot.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

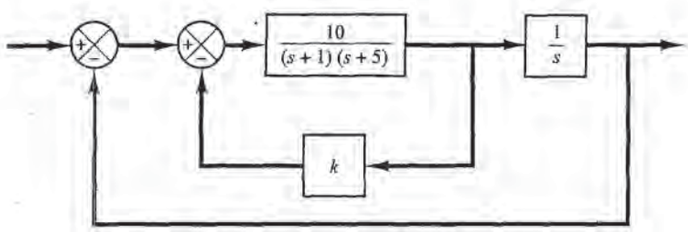


Figure 8–121
Control system.

B–8–20. Consider the unity-feedback control system with the following open-loop transfer function $G(s)$:

$$G(s) = \frac{K(s+2)}{s(s+1)(s+10)}$$

Plot Nyquist diagrams of $G(s)$ for $K = 1, 10$, and 100.

B–8–21. Consider a negative-feedback system with the following open-loop transfer function:

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

Plot the Nyquist diagram of $G(s)$. If the system were a positive-feedback one with the same open-loop transfer function $G(s)$, what would the Nyquist diagram look like?

B–8–22. Consider the control system shown in Figure 8–121. Plot Nyquist diagrams of $G(s)$, where

$$G(s) = \frac{10}{s[(s+1)(s+5)+10k]} = \frac{10}{s^3+6s^2+(5+10k)s}$$

for $k = 0.3, 0.5$, and 0.7.

B–8–23. Consider the closed-loop system shown in Figure 8–122. Plot the Bode diagrams and Nyquist diagrams of $G(s)$ for $K = 0.2, 0.5$, and 2. Also, plot the root-locus diagram of $G(s)$ and locate the closed-loop poles of the system for $K = 0.2, 0.5$, and 2. [$G(s)$ is the open-loop transfer function of the system.]

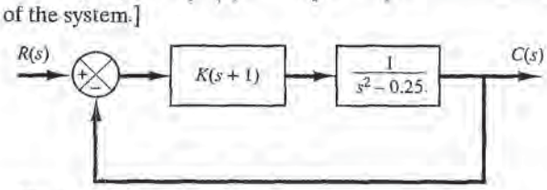


Figure 8–122
Closed-loop system.

B–8–24. Consider the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 6.5 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

There are four individual Nyquist plots involved in this system. Draw two Nyquist plots for the input u_1 in one diagram and two Nyquist plots for the input u_2 in another diagram. Write a MATLAB program to obtain these two diagrams.

B–8–25. Referring to Problem **B–8–24**, it is desired to plot only $Y_1(j\omega)/U_1(j\omega)$ for $\omega > 0$. Write a MATLAB program to produce such a plot.

If it is desired to plot $Y_1(j\omega)/U_1(j\omega)$ for $-\infty < \omega < \infty$, what changes must be made in the MATLAB program?

B–8–26. Consider the unity-feedback control system whose open-loop transfer function is

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

Determine the value of a so that the phase margin is 45°.

B–8–27. Consider the system shown in Figure 8–123. Draw a Bode diagram of the open-loop transfer function $G(s)$. Determine the phase margin and gain margin.

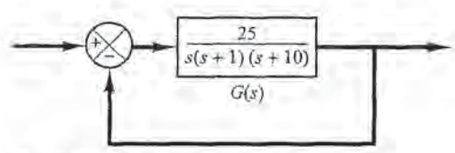


Figure 8–123
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–8–28. Consider the system shown in Figure 8–124. Draw a Bode diagram of the open-loop transfer function $G(s)$. Determine the phase margin and gain margin with MATLAB.

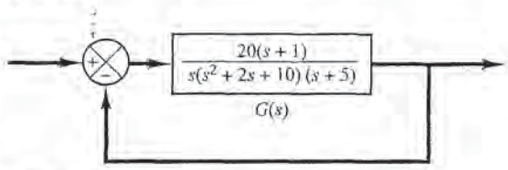


Figure 8–124
Control system.

B–8–29. Consider a unity-feedback control system with the open-loop transfer function

$$G(s) = \frac{K}{s(s^2 + s + 4)}$$

Determine the value of the gain K such that the phase margin is 50°. What is the gain margin with this gain K?

B–8–30. Consider the system shown in Figure 8–125. Draw a Bode diagram of the open-loop transfer function, and determine the value of the gain K such that the phase margin is 50°. What is the gain margin of this system with this gain K?

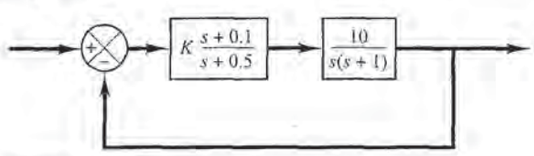


Figure 8–125
Control system.

B–8–31. Consider a unity-feedback control system whose open-loop transfer function is

$$G(s) = \frac{K}{s(s^2 + s + 0.5)}$$

Determine the value of the gain K such that the resonant peak magnitude in the frequency response is 2 dB, or $M_r = 2$ dB.

B–8–32. Figure 8–126 shows a block diagram of a process control system. Determine the range of the gain K for stability.

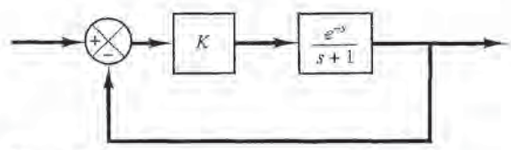


Figure 8–126
Process control system.

B–8–33. Consider a closed-loop system whose open-loop transfer function is

$$G(s)H(s) = \frac{Ke^{-Ts}}{s(s + 1)}$$

Determine the maximum value of the gain K for stability as a function of the dead time T.

B–8–34. Sketch the polar plot of

$$G(s) = \frac{(Ts)^2 - 6(Ts) + 12}{(Ts)^2 + 6(Ts) + 12}$$

Show that, for the frequency range $0 < \omega T < 2\sqrt{3}$, this equation gives a good approximation to the transfer function of transport lag, e^{-Ts}.

B–8–35. Figure 8–127 shows a Bode diagram of a transfer function $G(s)$. Determine this transfer function.

B–8–36. The experimentally determined Bode diagram of a system $G(j\omega)$ is shown in Figure 8–128. Determine the transfer function $G(s)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

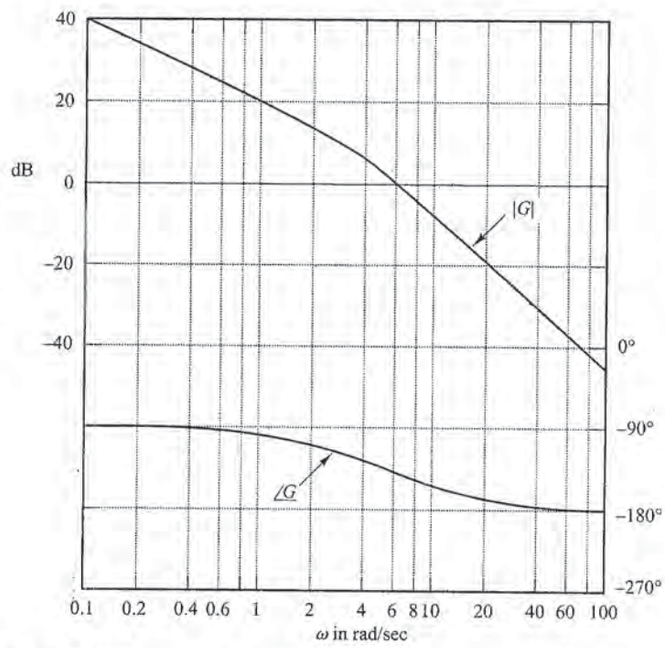


Figure 8–127
Bode diagram of a transfer function $G(s)$.

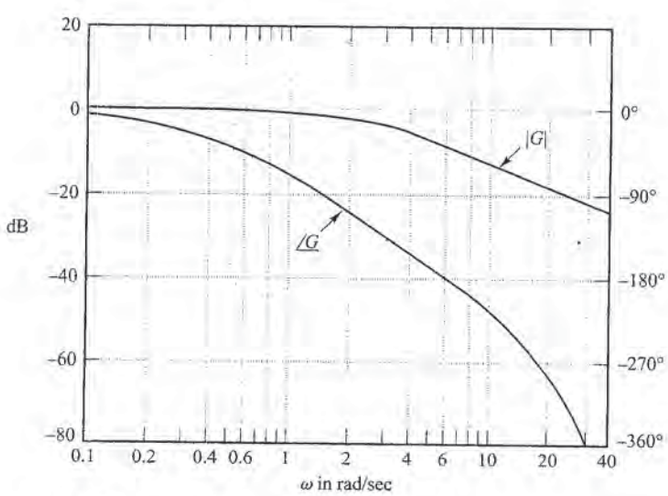


Figure 8–128
Experimentally determined Bode diagram of a system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

9

Control Systems Design by Frequency Response

9–1 INTRODUCTION

This chapter's primary objective is to present procedures for the design and compensation of single-input–single-output, linear, time-invariant control systems by the frequency-response approach.

In Chapters 6 and 7, we presented root-locus analysis and design. The root-locus method was shown to be very useful to reshape the transient response characteristics of closed-loop control systems. The root-locus approach gives us direct information on the transient response of the closed-loop system. The frequency-response approach, on the other hand, gives us this information only indirectly. However, as we shall see in this chapter, the frequency-response approach is very useful in designing control systems.

For any design problem, the designer will do well to use both approaches to the design and choose the compensator that most closely produces the desired closed-loop response.

Frequency-Response Approach to Control System Design. It is important to note that in a control system design, transient-response performance is usually most important. In the frequency-response approach, we specify the transient-response performance in an indirect manner. That is, the transient-response performance is specified in terms of the phase margin, gain margin, resonant peak magnitude (they give a rough estimate of the system damping); the gain crossover frequency, resonant frequency, bandwidth (they give a rough estimate of the speed of transient response); and static error constants (they give the steady-state accuracy). Although the correlation between the transient response and frequency response is indirect, the frequency domain specifications can be conveniently met in the Bode diagram approach.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

After the open loop has been designed by the frequency-response method, the closed-loop poles and zeros can be determined. The transient-response characteristics must be checked to see whether the designed system satisfies the requirements in the time domain. If it does not, then the compensator must be modified and the analysis repeated until a satisfactory result is obtained.

Design in the frequency domain is simple and straightforward. The frequency-response plot indicates clearly the manner in which the system should be modified, although the exact quantitative prediction of the transient-response characteristics cannot be made. The frequency-response approach can be applied to systems or components whose dynamic characteristics are given in the form of frequency-response data. Note that because of difficulty in deriving the equations governing certain components, such as pneumatic and hydraulic components, the dynamic characteristics of such components are usually determined experimentally through frequency-response tests. The experimentally obtained frequency-response plots can be combined easily with other such plots when the Bode diagram approach is used. Note also that in dealing with high-frequency noises we find that the frequency-response approach is more convenient than other approaches.

There are basically two approaches in the frequency-domain design. One is the polar plot approach and the other is the Bode diagram approach. When a compensator is added, the polar plot does not retain the original shape, and, therefore, we need to draw a new polar plot, which will take time and is thus inconvenient. On the other hand, a Bode diagram of the compensator can be simply added to the original Bode diagram, and thus plotting the complete Bode diagram is a simple matter. Also, if the open-loop gain is varied, the magnitude curve is shifted up or down without changing the slope of the curve, and the phase curve remains the same. For design purposes, therefore, it is best to work with the Bode diagram.

A common approach to the Bode diagram is that we first adjust the open-loop gain so that the requirement on the steady-state accuracy is met. Then the magnitude and phase curves of the uncompensated open loop (with the open-loop gain just adjusted) is plotted. If the specifications on the phase margin and gain margin are not satisfied, then a suitable compensator that will reshape the open-loop transfer function is determined. Finally, if there are any other requirements to be met, we try to satisfy them, unless some of them are mutually contradictory.

Information Obtainable from Open-Loop Frequency Response. The low-frequency region (the region far below the gain crossover frequency) of the locus indicates the steady-state behavior of the closed-loop system. The medium-frequency region (the region near the $-1 + j0$ point) of the locus indicates relative stability. The high-frequency region (the region far above the gain crossover frequency) indicates the complexity of the system.

Requirements on Open-Loop Frequency Response. We might say that, in many practical cases, compensation is essentially a compromise between steady-state accuracy and relative stability.

To have a high value of the velocity error constant and yet satisfactory relative stability, we find it necessary to reshape the open-loop frequency-response curve.

The gain in the low-frequency region should be large enough, and near the gain crossover frequency, the slope of the log-magnitude curve in the Bode diagram should

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

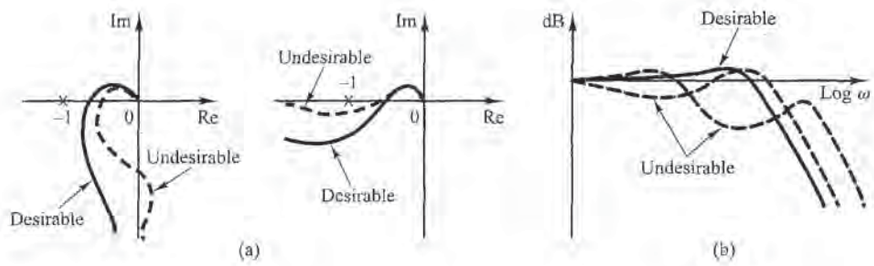


Figure 9–1
(a) Examples of desirable and undesirable open-loop frequency-response curves; (b) examples of desirable and undesirable closed-loop frequency-response curves.

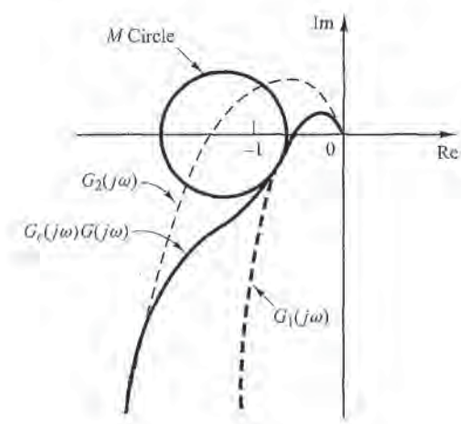


Figure 9–2
Reshaping of the open-loop frequency-response curve.

be −20 dB/decade. This slope should extend over a sufficiently wide frequency band to assure a proper phase margin. For the high-frequency region, the gain should be attenuated as rapidly as possible to minimize the effects of noise.

Examples of generally desirable and undesirable open-loop and closed-loop frequency-response curves are shown in Figure 9–1.

Referring to Figure 9–2, we see that the reshaping of the open-loop frequency-response curve may be done if the high-frequency portion of the locus follows the $G_1(j\omega)$ locus, while the low-frequency portion of the locus follows the $G_2(j\omega)$ locus. The reshaped locus $G_c(j\omega)G(j\omega)$ should have reasonable phase and gain margins or should be tangent to a proper M circle, as shown.

Basic Characteristics of Lead, Lag, and Lag–Lead Compensation. Lead compensation essentially yields an appreciable improvement in transient response and a small change in steady-state accuracy. It may accentuate high-frequency noise effects. Lag compensation, on the other hand, yields an appreciable improvement in steady-state accuracy at the expense of increasing the transient-response time. Lag compensation will suppress the effects of high-frequency noise signals. Lag–lead compensation combines the characteristics of both lead compensation and lag compensation. The use of a lead or lag compensator raises the order of the system by 1 (unless cancellation occurs between the zero of the compensator and a pole of the uncompensated open-loop trans-

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

fer function). The use of a lag–lead compensator raises the order of the system by 2 [unless cancellation occurs between zero(s) of the lag–lead compensator and pole(s) of the uncompensated open-loop transfer function], which means that the system becomes more complex and it is more difficult to control the transient response behavior. The particular situation determines the type of compensation to be used.

Outline of the Chapter. Section 9–1 has presented introductory material. Section 9–2 discusses lead compensation by the Bode diagram approach and Section 9–3 treats lag compensation by the Bode diagram approach. Section 9–4 discusses lag–lead compensation techniques based on the Bode diagram approach. Section 9–5 gives concluding comments on the frequency-response approach to the control systems design.

9–2 LEAD COMPENSATION

We shall first examine the frequency characteristics of the lead compensator. Then we present a design technique for the lead compensator by use of the Bode diagram.

Characteristics of Lead Compensators. Consider a lead compensator having the following transfer function:

$$K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}} \qquad (0<\alpha<1)$$

where α is called the attenuation factor of the lead compensator. It has a zero at $s=-1/T$ and a pole at $s=-1/(\alpha T)$. Since $0<\alpha<1$, we see that the zero is always located to the right of the pole in the complex plane. Note that for a small value of α the pole is located far to the left. The minimum value of α is limited by the physical construction of the lead compensator. The minimum value of α is usually taken to be about 0.05. (This means that the maximum phase lead that may be produced by a lead compensator is about 65°.) [See Equation (9–1).]

Figure 9–3 shows the polar plot of

$$K_c\alpha\frac{j\omega T+1}{j\omega\alpha T+1} \qquad (0<\alpha<1)$$

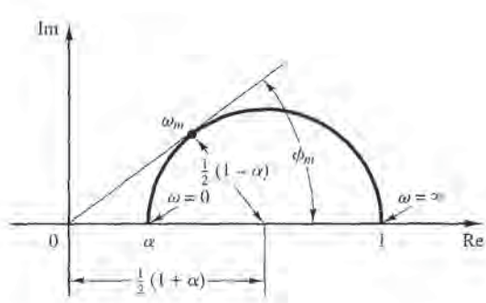


Figure 9–3
Polar plot of a lead compensator $\alpha(j\omega T+1)/(j\omega\alpha T+1)$, where $0<\alpha<1$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

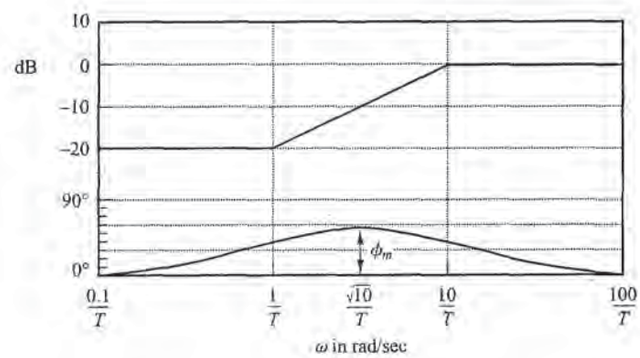


Figure 9–4
Bode diagram of a lead compensator $\alpha(j\omega T + 1)/(j\omega\alpha T + 1)$, where $\alpha = 0.1$.

with $K_c = 1$. For a given value of α, the angle between the positive real axis and the tangent line drawn from the origin to the semicircle gives the maximum phase-lead angle, ϕ_m. We shall call the frequency at the tangent point ω_m. From Figure 9–3 the phase angle at $\omega = \omega_m$ is ϕ_m, where

$$\sin\phi_m = \frac{\dfrac{1-\alpha}{2}}{\dfrac{1+\alpha}{2}} = \frac{1-\alpha}{1+\alpha} \tag{9–1}$$

Equation (9–1) relates the maximum phase-lead angle and the value of α.

Figure 9–4 shows the Bode diagram of a lead compensator when $K_c = 1$ and $\alpha = 0.1$. The corner frequencies for the lead compensator are $\omega = 1/T$ and $\omega = 1/(\alpha T) = 10/T$. By examining Figure 9–4, we see that ω_m is the geometric mean of the two corner frequencies, or

$$\log\omega_m = \frac{1}{2}\left(\log\frac{1}{T} + \log\frac{1}{\alpha T}\right)$$

Hence,

$$\omega_m = \frac{1}{\sqrt{\alpha}T} \tag{9–2}$$

As seen from Figure 9–4, the lead compensator is basically a high-pass filter. (The high frequencies are passed, but low frequencies are attenuated.)

Lead Compensation Techniques Based on the Frequency-Response Approach. The primary function of the lead compensator is to reshape the frequency-response curve to provide sufficient phase-lead angle to offset the excessive phase lag associated with the components of the fixed system.

Consider the system shown in Figure 9–5. Assume that the performance specifications are given in terms of phase margin, gain margin, static velocity error constants, and so on. The procedure for designing a lead compensator by the frequency-response approach may be stated as follows:

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

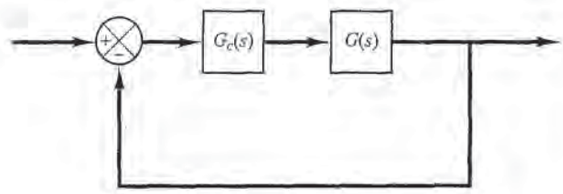


Figure 9–5
Control system.

1. Assume the following lead compensator:

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}} \qquad (0<\alpha<1)$$

Define

$$K_c\alpha = K$$

Then

$$G_c(s) = K\frac{Ts+1}{\alpha Ts+1}$$

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = K\frac{Ts+1}{\alpha Ts+1}G(s) = \frac{Ts+1}{\alpha Ts+1}KG(s) = \frac{Ts+1}{\alpha Ts+1}G_1(s)$$

where

$$G_1(s) = KG(s)$$

Determine gain K to satisfy the requirement on the given static error constant.

2. Using the gain K thus determined, draw a Bode diagram of $G_1(j\omega)$, the gain-adjusted but uncompensated system. Evaluate the phase margin.
3. Determine the necessary phase-lead angle to be added to the system. Add an additional 5° to 12° to the phase-lead angle required, because the addition of the lead compensator shifts the gain crossover frequency to the right and decreases the phase margin.
4. Determine the attenuation factor α by use of Equation (9–1). Determine the frequency where the magnitude of the uncompensated system $G_1(j\omega)$ is equal to $-20\log(1/\sqrt{\alpha})$. Select this frequency as the new gain crossover frequency. This frequency corresponds to $\omega_m = 1/(\sqrt{\alpha}T)$, and the maximum phase shift ϕ_m occurs at this frequency.
5. Determine the corner frequencies of the lead compensator as follows:

$$\text{Zero of lead compensator:} \qquad \omega = \frac{1}{T}$$

$$\text{Pole of lead compensator:} \qquad \omega = \frac{1}{\alpha T}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

6. Using the value of K determined in step 1 and that of α determined in step 4, calculate constant K_c from

$$K_c = \frac{K}{\alpha}$$

7. Check the gain margin to be sure it is satisfactory. If not, repeat the design process by modifying the pole–zero location of the compensator until a satisfactory result is obtained.

EXAMPLE 9–1 Consider the system shown in Figure 9–6. The open-loop transfer function is

$$G(s) = \frac{4}{s(s+2)}$$

It is desired to design a compensator for the system so that the static velocity error constant K_v is 20 sec^{-1}, the phase margin is at least 50°, and the gain margin is at least 10 dB.

We shall use a lead compensator of the form

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}$$

The compensated system will have the open-loop transfer function $G_c(s)G(s)$.

Define

$$G_1(s) = KG(s) = \frac{4K}{s(s+2)}$$

where $K = K_c\alpha$.

The first step in the design is to adjust the gain K to meet the steady-state performance specification or to provide the required static velocity error constant. Since this constant is given as 20 sec^{-1}, we obtain

$$K_v = \lim_{s\to 0} sG_c(s)G(s) = \lim_{s\to 0} s\frac{Ts+1}{\alpha Ts+1}G_1(s) = \lim_{s\to 0}\frac{s4K}{s(s+2)} = 2K = 20$$

or

$$K = 10$$

With $K = 10$, the compensated system will satisfy the steady-state requirement.

We shall next plot the Bode diagram of

$$G_1(j\omega) = \frac{40}{j\omega(j\omega+2)} = \frac{20}{j\omega(0.5j\omega+1)}$$

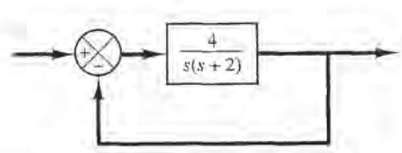


Figure 9–6
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

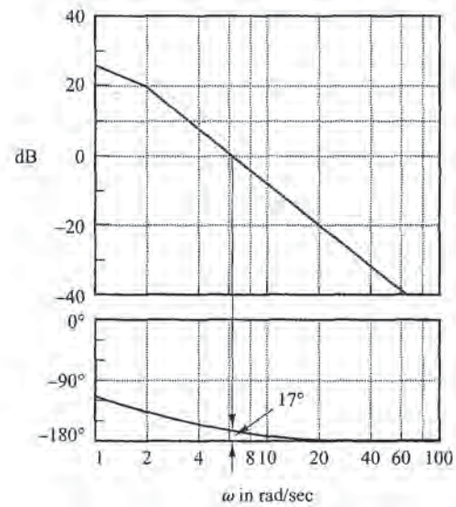


Figure 9–7
Bode diagram for $G_1(j\omega) = 10G(j\omega) = 40/[j\omega(j\omega + 2)]$

Figure 9–7 shows the magnitude and phase-angle curves of $G_1(j\omega)$. From this plot, the phase and gain margins of the system are found to be 17° and $+\infty$ dB, respectively. (A phase margin of 17° implies that the system is quite oscillatory. Thus, satisfying the specification on the steady state yields a poor transient-response performance.) The specification calls for a phase margin of at least 50°. We thus find the additional phase lead necessary to satisfy the relative stability requirement is 33°. To achieve a phase margin of 50° without decreasing the value of K, the lead compensator must contribute the required phase angle.

Noting that the addition of a lead compensator modifies the magnitude curve in the Bode diagram, we realize that the gain crossover frequency will be shifted to the right. We must offset the increased phase lag of $G_1(j\omega)$ due to this increase in the gain crossover frequency. Considering the shift of the gain crossover frequency, we may assume that ϕ_m, the maximum phase lead required, is approximately 38°. (This means that 5° has been added to compensate for the shift in the gain crossover frequency.)

Since

$$\sin\phi_m = \frac{1-\alpha}{1+\alpha}$$

$\phi_m = 38°$ corresponds to $\alpha = 0.24$. Once the attenuation factor α has been determined on the basis of the required phase-lead angle, the next step is to determine the corner frequencies $\omega = 1/T$ and $\omega = 1/(\alpha T)$ of the lead compensator. To do so, we first note that the maximum phase-lead angle ϕ_m occurs at the geometric mean of the two corner frequencies, or $\omega = 1/(\sqrt{\alpha}T)$. [See Equation (9–2).] The amount of the modification in the magnitude curve at $\omega = 1/(\sqrt{\alpha}T)$ due to the inclusion of the term $(Ts + 1)/(\alpha Ts + 1)$ is

$$\left|\frac{1 + j\omega T}{1 + j\omega\alpha T}\right|_{\omega = 1/(\sqrt{\alpha}T)} = \left|\frac{1 + j\dfrac{1}{\sqrt{\alpha}}}{1 + j\alpha\dfrac{1}{\sqrt{\alpha}}}\right| = \frac{1}{\sqrt{\alpha}}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Note that

$$\frac{1}{\sqrt{\alpha}} = \frac{1}{\sqrt{0.24}} = \frac{1}{0.49} = 6.2 \text{ dB}$$

and $|G_1(j\omega)| = -6.2$ dB corresponds to $\omega = 9$ rad/sec. We shall select this frequency to be the new gain crossover frequency ω_c. Noting that this frequency corresponds to $1/(\sqrt{\alpha}T)$, or $\omega_c = 1/(\sqrt{\alpha}T)$, we obtain

$$\frac{1}{T} = \sqrt{\alpha}\omega_c = 4.41$$

and

$$\frac{1}{\alpha T} = \frac{\omega_c}{\sqrt{\alpha}} = 18.4$$

The lead compensator thus determined is

$$G_c(s) = K_c \frac{s + 4.41}{s + 18.4} = K_c\alpha \frac{0.227s + 1}{0.054s + 1}$$

where the value of K_c is determined as

$$K_c = \frac{K}{\alpha} = \frac{10}{0.24} = 41.7$$

Thus, the transfer function of the compensator becomes

$$G_c(s) = 41.7 \frac{s + 4.41}{s + 18.4} = 10 \frac{0.227s + 1}{0.054s + 1}$$

Note that

$$\frac{G_c(s)}{K} G_1(s) = \frac{G_c(s)}{10} 10G(s) = G_c(s)G(s)$$

The magnitude curve and phase-angle curve for $G_c(j\omega)/10$ are shown in Figure 9–8. The compensated system has the following open-loop transfer function:

$$G_c(s)G(s) = 41.7 \frac{s + 4.41}{s + 18.4} \frac{4}{s(s + 2)}$$

The solid curves in Figure 9–8 show the magnitude curve and phase-angle curve for the compensated system. Note that the bandwidth is approximately equal to the gain crossover frequency. The lead compensator causes the gain crossover frequency to increase from 6.3 to 9 rad/sec. The increase in this frequency means an increase in bandwidth. This implies an increase in the speed of response. The phase and gain margins are seen to be approximately 50° and $+\infty$ dB, respectively. The compensated system shown in Figure 9–9 therefore meets both the steady-state and the relative-stability requirements.

Note that for type 1 systems, such as the system just considered, the value of the static velocity error constant K_v is merely the value of the frequency corresponding to the intersection of the extension of the initial −20-dB/decade slope line and the 0-dB line, as shown in Figure 9–8. Note also that we have changed the slope of the magnitude curve near the gain crossover frequency from −40 dB/decade to −20 dB/decade.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

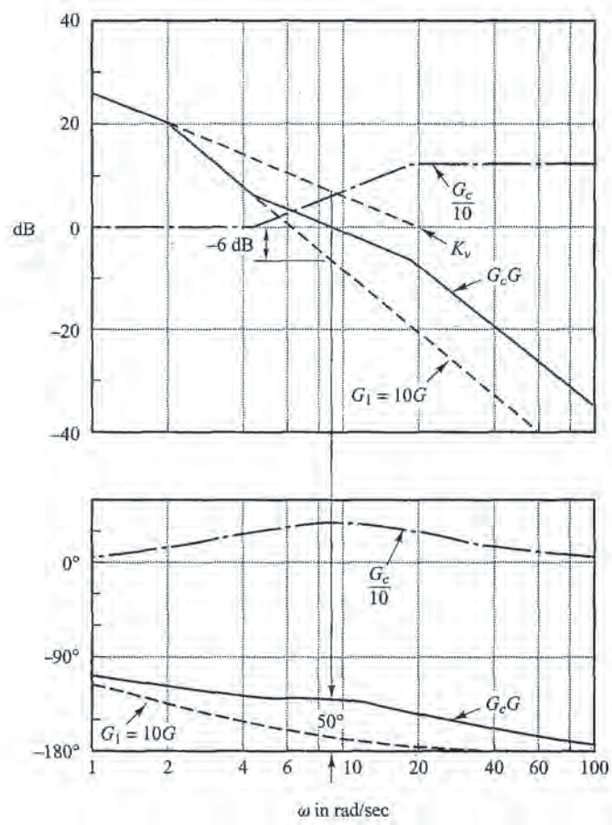


Figure 9–8
Bode diagram for the compensated system.

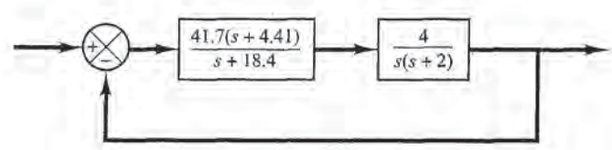


Figure 9–9
Compensated system.

Figure 9–10 shows the polar plots of the gain-adjusted but uncompensated open-loop transfer function $G_1(j\omega) = 10\,G(j\omega)$ and the compensated open-loop transfer function $G_c(j\omega)G(j\omega)$. From Figure 9–10, we see that the resonant frequency of the uncompensated system is about 6 rad/sec and that of the compensated system is about 7 rad/sec. (This also indicates that the bandwidth has been increased.)

From Figure 9–10, we find that the value of the resonant peak M_r for the uncompensated system with $K = 10$ is 3. The value of M_r for the compensated system is found to be 1.29. This clearly shows that the compensated system has improved relative stability.

Note that, if the phase angle of $G_1(j\omega)$ decreases rapidly near the gain crossover frequency, lead compensation becomes ineffective because the shift in the gain crossover frequency to the

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

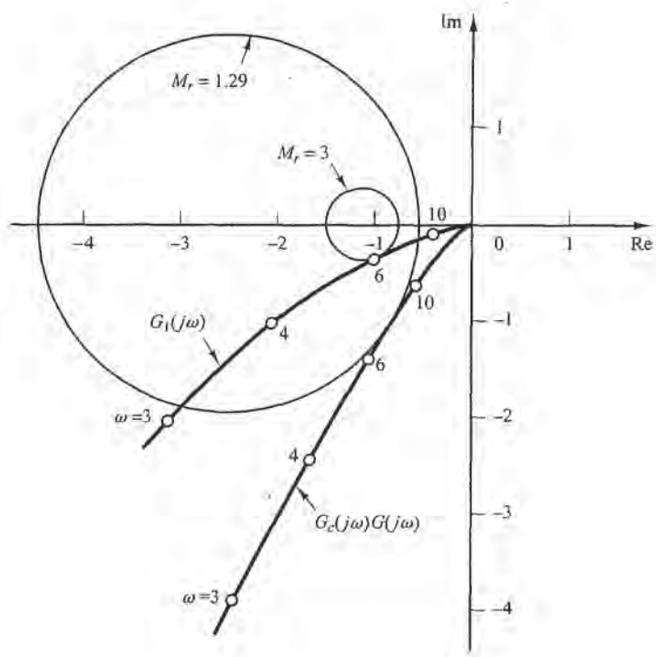


Figure 9–10
Polar plots of the gain-adjusted but uncompensated open-loop transfer function G_1 and compensated open-loop transfer function G_cG.

right makes it difficult to provide enough phase lead at the new gain crossover frequency. This means that, to provide the desired phase margin, we must use a very small value for α. The value of α, however, should not be too small (smaller than 0.05) nor should the maximum phase lead ϕ_m be too large (larger than 65°), because such values will require an additional gain of excessive value. [If more than 65° is needed, two (or more) lead networks may be used in series with an isolating amplifier.]

Finally, we shall examine the transient-response characteristics of the designed system. We shall obtain the unit-step response and unit-ramp response curves of the compensated and uncompensated systems with MATLAB. Note that the closed-loop transfer functions of the uncompensated and compensated systems are given, respectively, by

$$\frac{C(s)}{R(s)} = \frac{4}{s^2 + 2s + 4}$$

and

$$\frac{C(s)}{R(s)} = \frac{166.8s + 735.588}{s^3 + 20.4s^2 + 203.6s + 735.588}$$

MATLAB programs for obtaining the unit-step response and unit-ramp response curves are given in MATLAB Program 9–1. Figure 9–11 shows the unit-step response curves of the system before and after compensation. Also, Figure 9–12 depicts the unit-ramp response curves before and after compensation. These response curves indicate that the designed system is satisfactory.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–1

```
%*****Unit-step responses*****

num = [0  0  4];
den = [1  2  4];
numc = [0  0  166.8  735.588];
denc = [1  20.4  203.6  735.588];
t = 0:0.02:6;
[c1,x1,t] = step(num,den,t);
[c2,x2,t] = step(numc,denc,t);
plot (t,c1,'.',t,c2,'-')
grid
title('Unit-Step Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(0.4,1.31,'Compensated system')
text(1.55,0.88,'Uncompensated system')

%*****Unit-ramp responses*****

num1 = [0  0  0  4];
den1 = [1  2  4  0];
num1c = [0  0  0  166.8  735.588];
den1c = [1  20.4  203.6  735.588  0];
t = 0:0.02:5;
[y1,z1,t] = step(num1,den1,t);
[y2,z2,t] = step(num1c,den1c,t);
plot(t,y1,'.',t,y2,'-',t,t,'--')
grid
title('Unit-Ramp Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(0.89,3.7,'Compensated system')
text(2.25,1.1,'Uncompensated system')
```

It is noted that the closed-loop poles for the compensated system are located as follows:

$$s = -6.9541 \pm j8.0592$$

$$s = -6.4918$$

Because the dominant closed-loop poles are located far from the $j\omega$ axis, the response damps out quickly.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

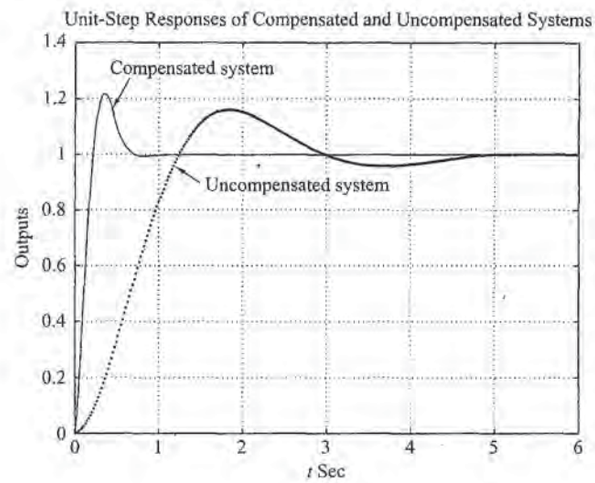


Figure 9–11
Unit-step response curves of the compensated and uncompensated systems.

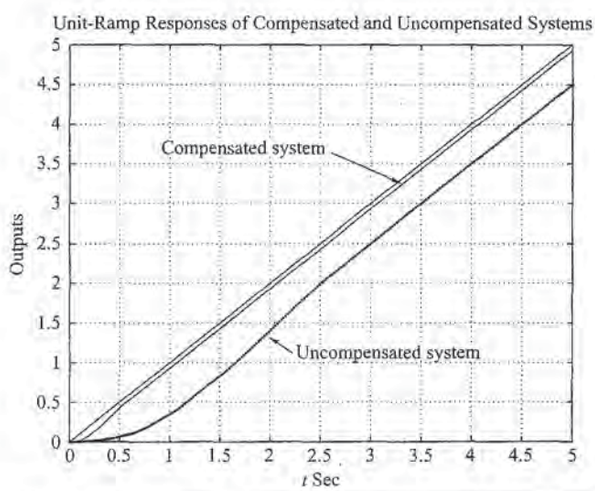


Figure 9–12
Unit-ramp response curves of the compensated and uncompensated systems.

9–3 LAG COMPENSATION

In this section we first discuss the Nyquist plot and Bode diagram of the lag compensator. Then we present lag compensation techniques based on the frequency-response approach.

Characteristics of Lag Compensators. Consider a lag compensator having the following transfer function:

$$G_c(s) = K_c\beta\frac{Ts+1}{\beta Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\beta T}} \qquad (\beta > 1)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

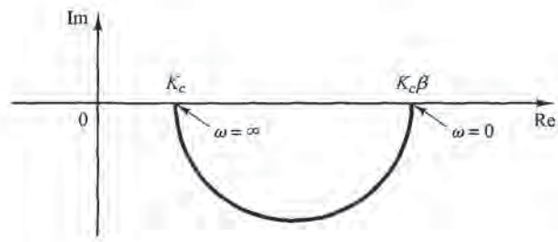


Figure 9–13
Polar plot of a lag compensator $K_c\beta(j\omega T + 1)/(j\omega\beta T + 1)$.

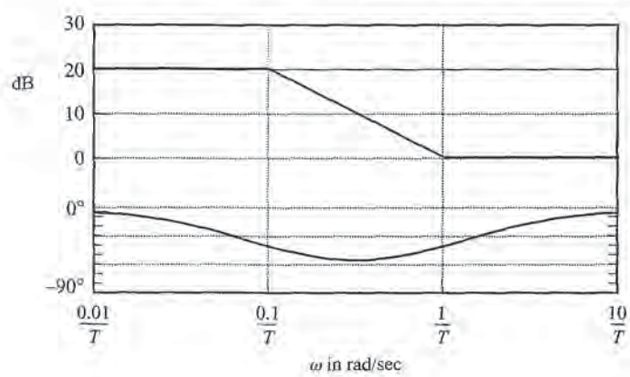


Figure 9–14
Bode diagram of a lag compensator $\beta(j\omega T + 1)/(j\omega\beta T + 1)$, with $\beta = 10$.

In the complex plane, a lag compensator has a zero at $s = -1/T$ and a pole at $s = -1/(\beta T)$. The pole is located to the right of the zero.

Figure 9–13 shows a polar plot of the lag compensator. Figure 9–14 shows a Bode diagram of the compensator, where $K_c = 1$ and $\beta = 10$. The corner frequencies of the lag compensator are at $\omega = 1/T$ and $\omega = 1/(\beta T)$. As seen from Figure 9–14, where the values of K_c and β are set equal to 1 and 10, respectively, the magnitude of the lag compensator becomes 10 (or 20 dB) at low frequencies and unity (or 0 dB) at high frequencies. Thus, the lag compensator is essentially a low-pass filter.

Lag Compensation Techniques Based on the Frequency-Response Approach. The primary function of a lag compensator is to provide attenuation in the high-frequency range to give a system sufficient phase margin. The phase-lag characteristic is of no consequence in lag compensation.

The procedure for designing lag compensators for the system shown in Figure 9–5 by the frequency-response approach may be stated as follows:

1. Assume the following lag compensator:

$$G_c(s) = K_c\beta\frac{Ts + 1}{\beta Ts + 1} = K_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}} \qquad (\beta > 1)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Define

$$K_c\beta = K$$

Then

$$G_c(s) = K\frac{Ts + 1}{\beta Ts + 1}$$

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = K\frac{Ts + 1}{\beta Ts + 1}G(s) = \frac{Ts + 1}{\beta Ts + 1}KG(s) = \frac{Ts + 1}{\beta Ts + 1}G_1(s)$$

where

$$G_1(s) = KG(s)$$

Determine gain K to satisfy the requirement on the given static velocity error constant.

2. If the gain-adjusted but uncompensated system $G_1(j\omega) = KG(j\omega)$ does not satisfy the specifications on the phase and gain margins, then find the frequency point where the phase angle of the open-loop transfer function is equal to $-180°$ plus the required phase margin. The required phase margin is the specified phase margin plus 5° to 12°. (The addition of 5° to 12° compensates for the phase lag of the lag compensator.) Choose this frequency as the new gain crossover frequency.

3. To prevent detrimental effects of phase lag due to the lag compensator, the pole and zero of the lag compensator must be located substantially lower than the new gain crossover frequency. Therefore, choose the corner frequency $\omega = 1/T$ (corresponding to the zero of the lag compensator) 1 octave to 1 decade below the new gain crossover frequency. (If the time constants of the lag compensator do not become too large, the corner frequency $\omega = 1/T$ may be chosen 1 decade below the new gain crossover frequency.)

   Notice that we choose the compensator pole and zero sufficiently small. Thus the phase lag occurs at the low-frequency region so that it will not affect the phase margin.

4. Determine the attenuation necessary to bring the magnitude curve down to 0 dB at the new gain crossover frequency. Noting that this attenuation is $-20 \log \beta$, determine the value of β. Then the other corner frequency (corresponding to the pole of the lag compensator) is determined from $\omega = 1/(\beta T)$.

5. Using the value of K determined in step 1 and that of β determined in step 4, calculate constant K_c from

$$K_c = \frac{K}{\beta}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 9–2 Consider the system shown in Figure 9–15. The open-loop transfer function is given by

$$G(s) = \frac{1}{s(s+1)(0.5s+1)}$$

It is desired to compensate the system so that the static velocity error constant K_v is 5 sec^{-1}, the phase margin is at least 40°, and the gain margin is at least 10 dB.

We shall use a lag compensator of the form

$$G_c(s) = K_c\beta\frac{Ts+1}{\beta Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\beta T}} \qquad (\beta > 1)$$

Define

$$K_c\beta = K$$

Define also

$$G_1(s) = KG(s) = \frac{K}{s(s+1)(0.5s+1)}$$

The first step in the design is to adjust the gain K to meet the required static velocity error constant. Thus,

$$K_v = \lim_{s\to 0} sG_c(s)G(s) = \lim_{s\to 0} s\frac{Ts+1}{\beta Ts+1}G_1(s) = \lim_{s\to 0} sG_1(s)$$

$$= \lim_{s\to 0}\frac{sK}{s(s+1)(0.5s+1)} = K = 5$$

or

$$K = 5$$

With $K = 5$, the compensated system satisfies the steady-state performance requirement.

We shall next plot the Bode diagram of

$$G_1(j\omega) = \frac{5}{j\omega(j\omega+1)(0.5j\omega+1)}$$

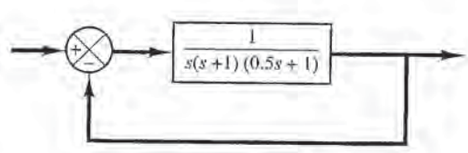


Figure 9–15
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

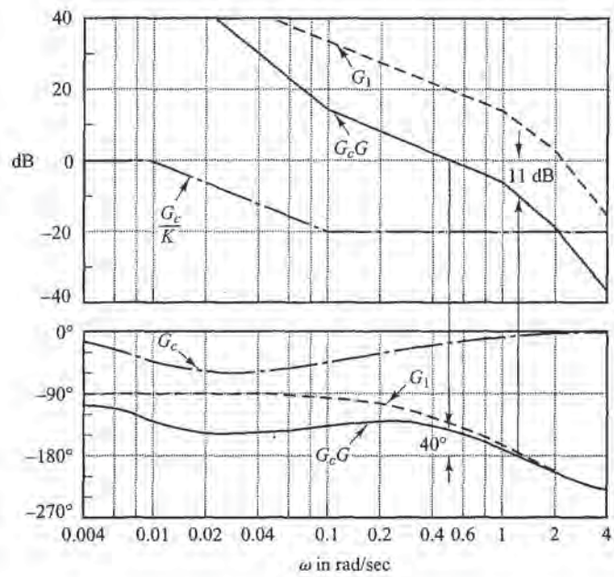


Figure 9–16
Bode diagrams for G_1 (gain-adjusted but uncompensated open-loop transfer function), G_c (compensator), and G_cG (compensated open-loop transfer function).

The magnitude curve and phase-angle curve of $G_1(j\omega)$ are shown in Figure 9–16. From this plot, the phase margin is found to be −20°, which means that the gain-adjusted but uncompensated system is unstable.

Noting that the addition of a lag compensator modifies the phase curve of the Bode diagram, we must allow 5° to 12° to the specified phase margin to compensate for the modification of the phase curve. Since the frequency corresponding to a phase margin of 40° is 0.7 rad/sec, the new gain crossover frequency (of the compensated system) must be chosen near this value. To avoid overly large time constants for the lag compensator, we shall choose the corner frequency $\omega = 1/T$ (which corresponds to the zero of the lag compensator) to be 0.1 rad/sec. Since this corner frequency is not too far below the new gain crossover frequency, the modification in the phase curve may not be small. Hence, we add about 12° to the given phase margin as an allowance to account for the lag angle introduced by the lag compensator. The required phase margin is now 52°. The phase angle of the uncompensated open-loop transfer function is −128° at about $\omega = 0.5$ rad/sec. So we choose the new gain crossover frequency to be 0.5 rad/sec. To bring the magnitude curve down to 0 dB at this new gain crossover frequency, the lag compensator must give the necessary attenuation, which in this case is −20 dB. Hence,

$$20 \log \frac{1}{\beta} = -20$$

or,

$$\beta = 10$$

The other corner frequency $\omega = 1(\beta T)$, which corresponds to the pole of the lag compensator, is then determined as

$$\frac{1}{\beta T} = 0.01 \text{ rad/sec}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Thus, the transfer function of the lag compensator is

$$G_c(s) = K_c(10)\frac{10s + 1}{100s + 1} = K_c\frac{s + \dfrac{1}{10}}{s + \dfrac{1}{100}}$$

Since the gain K was determined to be 5 and β was determined to be 10, we have

$$K_c = \frac{K}{\beta} = \frac{5}{10} = 0.5$$

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = \frac{5(10s + 1)}{s(100s + 1)(s + 1)(0.5s + 1)}$$

The magnitude and phase-angle curves of $G_c(j\omega)G(j\omega)$ are also shown in Figure 9–16.

The phase margin of the compensated system is about 40°, which is the required value. The gain margin is about 11 dB, which is quite acceptable. The static velocity error constant is 5 $\sec^{-1}$, as required. The compensated system, therefore, satisfies the requirements on both the steady state and the relative stability.

Note that the new gain crossover frequency is decreased from approximately 1 to 0.5 rad/sec. This means that the bandwidth of the system is reduced.

To further show the effects of lag compensation, the log-magnitude-versus-phase plots of the gain-adjusted but uncompensated system $G_1(j\omega)$ and of the compensated system $G_c(j\omega)G(j\omega)$ are shown in Figure 9–17. The plot of $G_1(j\omega)$ clearly shows that the gain-adjusted but uncompensated system is unstable. The addition of the lag compensator stabilizes the system. The plot of $G_c(j\omega)G(j\omega)$ is tangent to the $M = 3$ dB locus. Thus, the resonant peak value is 3 dB, or 1.4, and this peak occurs at $\omega = 0.5$ rad/sec.

Compensators designed by different methods or by different designers (even using the same approach) may look sufficiently different. Any of the well-designed systems, however, will give similar transient and steady-state performance. The best among many alternatives may be chosen from the economic consideration that the time constants of the lag compensator should not be too large.

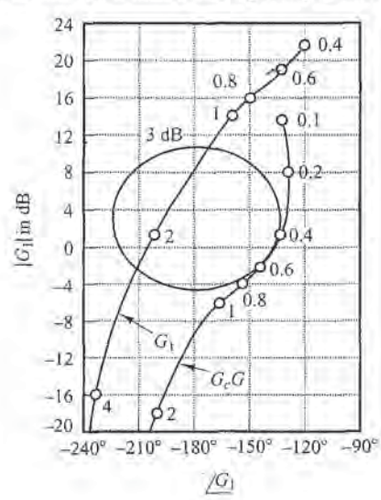


Figure 9–17
Log-magnitude-versus-phase plots of G_1 (gain-adjusted but uncompensated open-loop transfer function) and G_cG (compensated open-loop transfer function).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Finally, we shall examine the unit-step response and unit-ramp response of the compensated system and the original uncompensated system without gain adjustment. The closed-loop transfer functions of the compensated and uncompensated systems are

$$\frac{C(s)}{R(s)} = \frac{50s + 5}{50s^4 + 150.5s^3 + 101.5s^2 + 51s + 5}$$

and

$$\frac{C(s)}{R(s)} = \frac{1}{0.5s^3 + 1.5s^2 + s + 1}$$

respectively. MATLAB Program 9–2 will produce the unit-step and unit-ramp responses of the compensated and uncompensated systems. The resulting unit-step response curves and unit-ramp response curves are shown in Figures 9–18 and 9–19, respectively. From the response curves we find that the designed system satisfies the given specifications and is satisfactory.

MATLAB Program 9–2

```
%*****Unit-step response*****

num = [0 0 0 1];
den = [0.5 1.5 1 1];
numc = [0 0 0 50 5];
denc = [50 150.5 101.5 51 5];
t = 0:0.1:40;
[c1,x1,t] = step(num,den,t);
[c2,x2,t] = step(numc,denc,t);
plot(t,c1,'.',t,c2,'-')
grid
title('Unit-Step Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(12.7,1.27,'Compensated system')
text(12.2,0.7,'Uncompensated system')

%*****Unit-ramp response*****

num1 = [0 0 0 0 1];
den1 = [0.5 1.5 1 1 0];
num1c = [0 0 0 0 50 5];
den1c = [50 150.5 101.5 51 5 0];
t = 0:0.1:20;
[y1,z1,t] = step(num1,den1,t);
[y2,z2,t] = step(num1c,den1c,t);
plot(t,y1,'.',t,y2,'-',t,t,'--');
grid
title('Unit-Ramp Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(8.3,3,'Compensated system')
text(8.3,5,'Uncompensated system')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

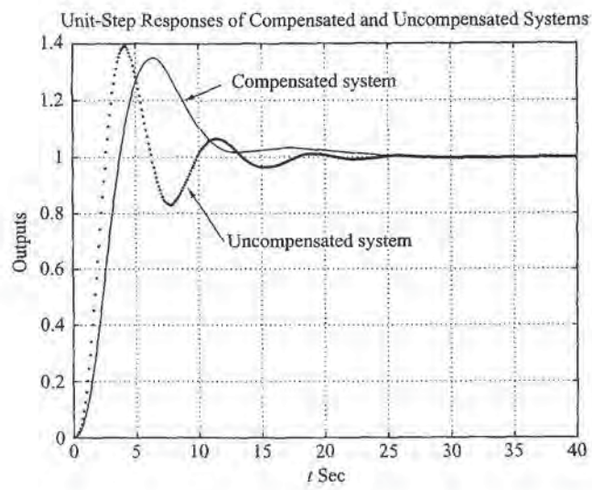


Figure 9–18
Unit-step response curves for the compensated and uncompensated systems (Example 9–2).

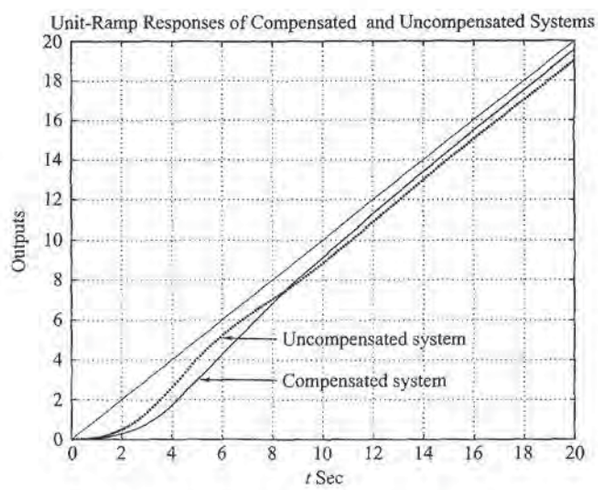


Figure 9–19
Unit-ramp response curves for the compensated and uncompensated systems (Example 9–2).

Note that the zero and poles of the designed closed-loop system are as follows:

Zero at $s = -0.1$

Poles at $s = -0.2859 \pm j0.5196, \quad s = -0.1228, \quad s = -2.3155$

The dominant closed-loop poles are very close to the $j\omega$ axis with the result that the response is slow. Also, a pair of the closed-loop pole at $s = -0.1228$ and the zero at $s = -0.1$ produces a slowly decreasing tail of small amplitude.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A Few Comments on Lag Compensation.

1. Lag compensators are essentially low-pass filters. Therefore, lag compensation permits a high gain at low frequencies (which improves the steady-state performance) and reduces gain in the higher critical range of frequencies so as to improve the phase margin. Note that in lag compensation we utilize the attenuation characteristic of the lag compensator at high frequencies rather than the phase-lag characteristic. (The phase-lag characteristic is of no use for compensation purposes.)

2. Suppose that the zero and pole of a lag compensator are located at $s = -z$ and $s = -p$, respectively. Then the exact location of the zero and pole is not critical provided that they are close to the origin and the ratio z/p is equal to the required multiplication factor of the static velocity error constant.

   It should be noted, however, that the zero and pole of the lag compensator should not be located unnecessarily close to the origin, because the lag compensator will create an additional closed-loop pole in the same region as the zero and pole of the lag compensator.

   The closed-loop pole located near the origin gives a very slowly decaying transient response, although its magnitude will become very small because the zero of the lag compensator will almost cancel the effect of this pole. However, the transient response (decay) due to this pole is so slow that the settling time will be adversely affected.

   It is also noted that in the system compensated by a lag compensator the transfer function between the plant disturbance and the system error may not involve a zero that is near this pole. Therefore, the transient response to the disturbance input may last very long.

3. The attenuation due to the lag compensator will shift the gain crossover frequency to a lower frequency point where the phase margin is acceptable. Thus, the lag compensator will reduce the bandwidth of the system and will result in slower transient response. [The phase angle curve of $G_c(j\omega)G(j\omega)$ is relatively unchanged near and above the new gain crossover frequency.]

4. Since the lag compensator tends to integrate the input signal, it acts approximately as a proportional-plus-integral controller. Because of this, a lag-compensated system tends to become less stable. To avoid this undesirable feature, the time constant T should be made sufficiently larger than the largest time constant of the system.

5. Conditional stability may occur when a system having saturation or limiting is compensated by use of a lag compensator. When the saturation or limiting takes place in the system, it reduces the effective loop gain. Then the system becomes less stable and unstable operation may even result, as shown in Figure 9–20. To avoid this, the system must be designed so that the effect of lag compensation becomes significant only when the amplitude of the input to the saturating element is small. (This can be done by means of minor feedback-loop compensation.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

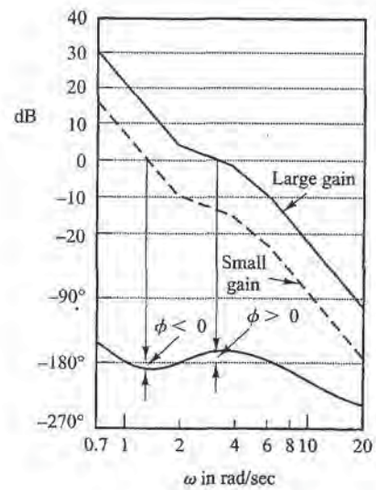


Figure 9–20
Bode diagram of a conditionally stable system.

9–4 LAG–LEAD COMPENSATION

We shall first examine the frequency-response characteristics of the lag–lead compensator. Then we present the lag–lead compensation technique based on the frequency-response approach.

Characteristic of Lag–Lead Compensator. Consider the lag–lead compensator given by

$$G_c(s) = K_c\left(\frac{s+\dfrac{1}{T_1}}{s+\dfrac{\gamma}{T_1}}\right)\left(\frac{s+\dfrac{1}{T_2}}{s+\dfrac{1}{\beta T_2}}\right) \tag{9–3}$$

where $\gamma > 1$ and $\beta > 1$. The term

$$\frac{s+\dfrac{1}{T_1}}{s+\dfrac{\gamma}{T_1}} = \frac{1}{\gamma}\left(\frac{T_1 s+1}{\dfrac{T_1}{\gamma}s+1}\right) \qquad (\gamma > 1)$$

produces the effect of the lead network, and the term

$$\frac{s+\dfrac{1}{T_2}}{s+\dfrac{1}{\beta T_2}} = \beta\left(\frac{T_2 s+1}{\beta T_2 s+1}\right) \qquad (\beta > 1)$$

produces the effect of the lag network.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

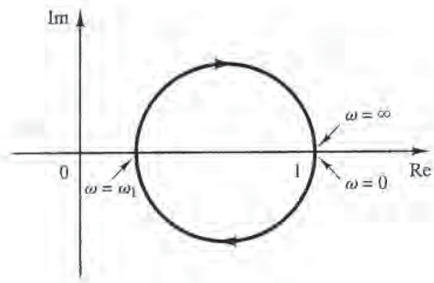


Figure 9–21
Polar plot of a lag–lead compensator given by Equation (9–3), with $K_c = 1$ and $\gamma = \beta$.

In designing a lag–lead compensator, we frequently chose $\gamma = \beta$. (This is not necessary. We can, of course, choose $\gamma \neq \beta$.) In what follows, we shall consider the case where $\gamma = \beta$. The polar plot of the lag–lead compensator with $K_c = 1$ and $\gamma = \beta$ becomes as shown in Figure 9–21. It can be seen that, for $0 < \omega < \omega_1$, the compensator acts as a lag compensator, while for $\omega_1 < \omega < \infty$ it acts as a lead compensator. The frequency ω_1 is the frequency at which the phase angle is zero. It is given by

$$\omega_1 = \frac{1}{\sqrt{T_1 T_2}}$$

(To derive this equation, see Problem **A–9–2**.)

Figure 9–22 shows the Bode diagram of a lag–lead compensator when $K_c = 1$, $\gamma = \beta = 10$, and $T_2 = 10T_1$. Notice that the magnitude curve has the value 0 dB at the low- and high-frequency regions.

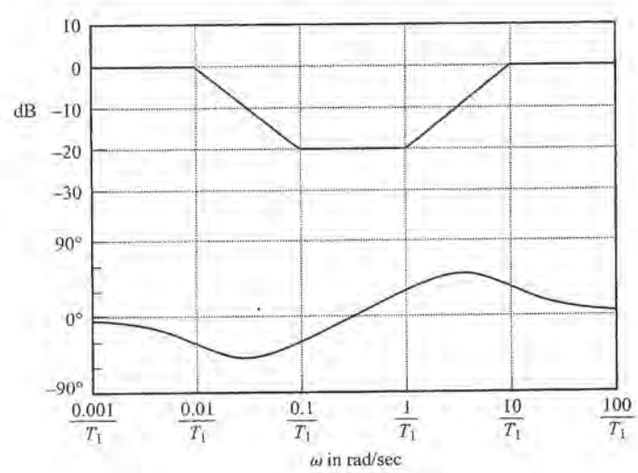


Figure 9–22
Bode diagram of a lag–lead compensator given by Equation (9–3) with $K_c = 1$, $\gamma = \beta = 10$, and $T_2 = 10T_1$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Lag–Lead Compensation Based on the Frequency-Response Approach. The design of a lag–lead compensator by the frequency-response approach is based on the combination of the design techniques discussed under lead compensation and lag compensation.

Let us assume that the lag–lead compensator is of the following form:

$$G_c(s) = K_c \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta} s + 1\right)(\beta T_2 s + 1)} = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \tag{9–4}$$

where $\beta > 1$. The phase lead portion of the lag–lead compensator (the portion involving T_1) alters the frequency-response curve by adding phase-lead angle and increasing the phase margin at the gain crossover frequency. The phase-lag portion (the portion involving T_2) provides attenuation near and above the gain crossover frequency and thereby allows an increase of gain at the low-frequency range to improve the steady-state performance.

We shall illustrate the details of the procedures for designing a lag–lead compensator by an example.

EXAMPLE 9–3 Consider the unity-feedback system whose open-loop transfer function is

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

It is desired that the static velocity error constant be 10 $\sec^{-1}$, the phase margin be 50°, and the gain margin be 10 dB or more.

Assume that we use the lag–lead compensator given by Equation (9–4). [Note that the phase-lead portion increases both the phase margin and the system bandwidth (which implies increasing the speed of response). The phase lag portion maintains the low-frequency gain.]

The open-loop transfer function of the compensated system is $G_c(s)G(s)$. Since the gain K of the plant is adjustable, let us assume that $K_c = 1$. Then, $\lim_{s \to 0} G_c(s) = 1$.

From the requirement on the static velocity error constant, we obtain

$$K_v = \lim_{s \to 0} sG_c(s)G(s) = \lim_{s \to 0} sG_c(s) \frac{K}{s(s + 1)(s + 2)} = \frac{K}{2} = 10$$

Hence,

$$K = 20$$

We shall next draw the Bode diagram of the uncompensated system with $K = 20$, as shown in Figure 9–23. The phase margin of the gain-adjusted but uncompensated system is found to be −32°, which indicates that the gain-adjusted but uncompensated system is unstable.

The next step in the design of a lag–lead compensator is to choose a new gain crossover frequency. From the phase-angle curve for $G(j\omega)$, we notice that $\angle G(j\omega) = -180°$ at $\omega = 1.5$ rad/sec. It is convenient to choose the new gain crossover frequency to be 1.5 rad/sec so that the phase-lead angle required at $\omega = 1.5$ rad/sec is about 50°, which is quite possible by use of a single lag–lead network.

Once we choose the gain crossover frequency to be 1.5 rad/sec, we can determine the corner frequency of the phase-lag portion of the lag–lead compensator. Let us choose the corner frequency $\omega = 1/T_2$ (which corresponds to the zero of the phase-lag portion of the compensator) to be 1 decade below the new gain crossover frequency, or at $\omega = 0.15$ rad/sec.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

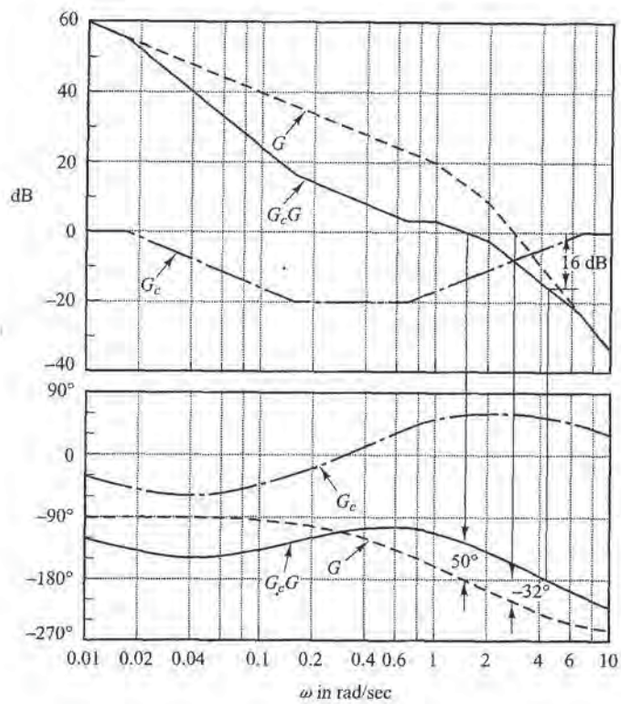


Figure 9–23
Bode diagrams for G (gain-adjusted but uncompensated open-loop transfer function), G_c (compensator), and G_cG (compensated open-loop transfer function).

Recall that for the lead compensator the maximum phase-lead angle ϕ_m is given by Equation (9–1), where α is $1/\beta$ in the present case. By substituting $\alpha = 1/\beta$ in Equation (9–1), we have

$$\sin\phi_m = \frac{1 - \dfrac{1}{\beta}}{1 + \dfrac{1}{\beta}} = \frac{\beta - 1}{\beta + 1}$$

Notice that $\beta = 10$ corresponds to $\phi_m = 54.9°$. Since we need a 50° phase margin, we may choose $\beta = 10$. (Note that we will be using several degrees less than the maximum angle, 54.9°.) Thus,

$$\beta = 10$$

Then the corner frequency $\omega = 1/\beta T_2$ (which corresponds to the pole of the phase-lag portion of the compensator) becomes $\omega = 0.015$ rad/sec. The transfer function of the phase-lag portion of the lag–lead compensator then becomes

$$\frac{s + 0.15}{s + 0.015} = 10\left(\frac{6.67s + 1}{66.7s + 1}\right)$$

The phase-lead portion can be determined as follows: Since the new gain crossover frequency is $\omega = 1.5$ rad/sec, from Figure 9–23, $G(j1.5)$ is found to be 13 dB. Hence, if the lag–lead compensator contributes −13 dB at $\omega = 1.5$ rad/sec, then the new gain crossover frequency is as desired. From this requirement, it is possible to draw a straight line of slope 20 dB/decade, passing through the point (1.5 rad/sec, −13 dB). The intersections of this line and the 0-dB line and −20-dB line determine the corner frequencies. Thus, the corner frequencies for the lead portion

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

are $\omega = 0.7$ rad/sec and $\omega = 7$ rad/sec. Thus, the transfer function of the lead portion of the lag–lead compensator becomes

$$\frac{s + 0.7}{s + 7} = \frac{1}{10}\left(\frac{1.43s + 1}{0.143s + 1}\right)$$

Combining the transfer functions of the lag and lead portions of the compensator, we obtain the transfer function of the lag–lead compensator. Since we chose $K_c = 1$, we have

$$G_c(s) = \left(\frac{s + 0.7}{s + 7}\right)\left(\frac{s + 0.15}{s + 0.015}\right) = \left(\frac{1.43s + 1}{0.143s + 1}\right)\left(\frac{6.67s + 1}{66.7s + 1}\right)$$

The magnitude and phase-angle curves of the lag–lead compensator just designed are shown in Figure 9–23. The open-loop transfer function of the compensated system is

$$\begin{aligned} G_c(s)G(s) &= \frac{(s + 0.7)(s + 0.15)20}{(s + 7)(s + 0.015)s(s + 1)(s + 2)} \\ &= \frac{10(1.43s + 1)(6.67s + 1)}{s(0.143s + 1)(66.7s + 1)(s + 1)(0.5s + 1)} \end{aligned} \tag{9–5}$$

The magnitude and phase-angle curves of the system of Equation (9–5) are also shown in Figure 9–23. The phase margin of the compensated system is 50°, the gain margin is 16 dB, and the static velocity error constant is 10 $\sec^{-1}$. All the requirements are therefore met, and the design has been completed.

Figure 9–24 shows the polar plots of $G(j\omega)$ (gain-adjusted but uncompensated open-loop transfer function) and $G_c(j\omega)G(j\omega)$ (compensated open-loop transfer function). The $G_c(j\omega)G(j\omega)$ locus is tangent to the $M = 1.2$ circle at about $\omega = 2$ rad/sec. Clearly, this indicates that the compensated system has satisfactory relative stability. The bandwidth of the compensated system is slightly larger than 2 rad/sec.

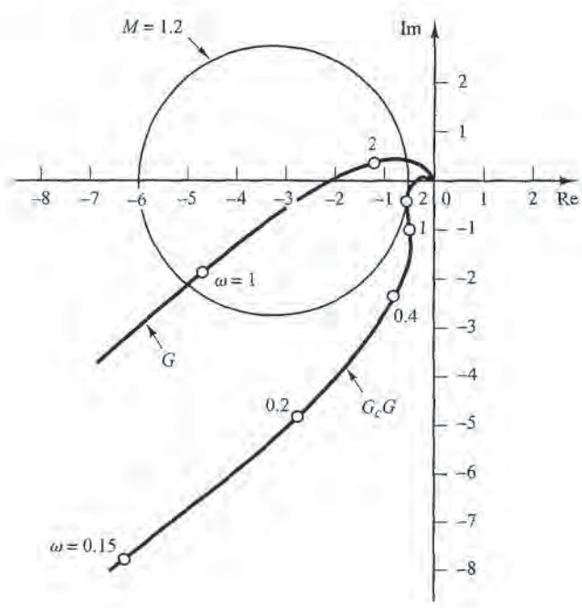


Figure 9–24
Polar plots of G (gain adjusted) and G_cG.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

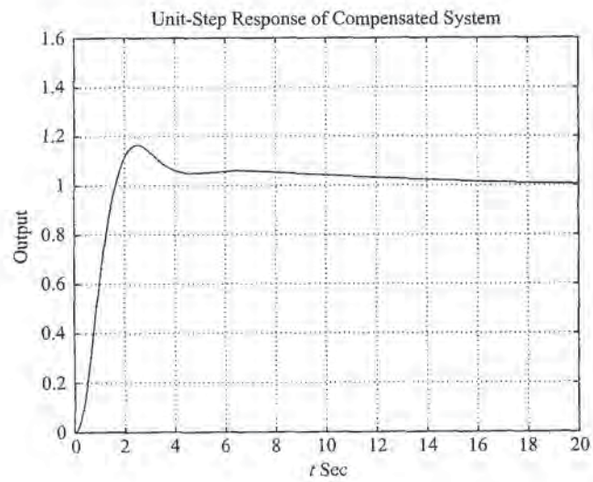


Figure 9–25
Unit-step response of the compensated system (Example 9–3).

In the following we shall examine the transient-response characteristics of the compensated system. (The gain-adjusted but uncompensated system is unstable.) The closed-loop transfer function of the compensated system is

$$\frac{C(s)}{R(s)} = \frac{95.381s^2 + 81s + 10}{4.7691s^5 + 47.7287s^4 + 110.3026s^3 + 163.724s^2 + 82s + 10}$$

The unit-step and unit-ramp response curves obtained with MATLAB are shown in Figures 9–25 and 9–26, respectively.

Note that the designed closed-loop control system has the following closed-loop zeros and poles:

$$\text{Zeros at } s = -0.1499, \qquad s = -0.6993$$

$$\text{Poles at } s = -0.8973 \pm j1.4439$$

$$s = -0.1785, \qquad s = -0.5425, \qquad s = -7.4923$$

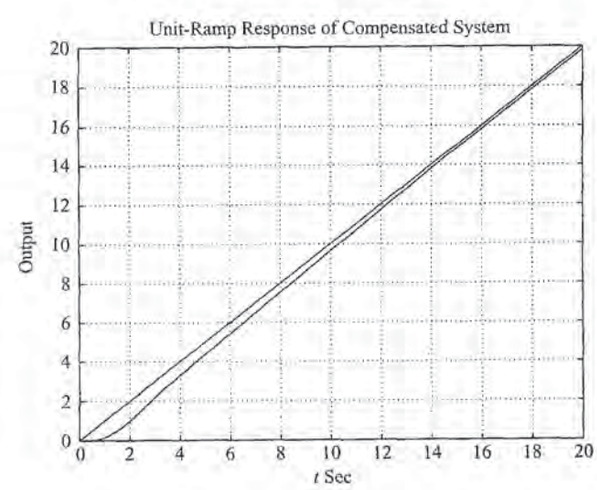


Figure 9–26
Unit-ramp response of the compensated system (Example 9–3).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The pole at $s = -0.1785$ and zero at $s = -0.1499$ are located very close to each other. Such a pair of pole and zero produces a long tail of small amplitude in the step response, as seen in Figure 9–25. Also, the pole at $s = -0.5425$ and zero at $s = -0.6993$ are located fairly close to each other. This pair adds amplitude to the long tail.

9–5 CONCLUDING COMMENTS

This chapter has presented detailed procedures for designing lead, lag, and lag–lead compensators by the use of simple examples. We have shown that the design of a compensator to satisfy the given specifications (in terms of the phase margin and gain margin) can be carried out in the Bode diagram in a simple and straightforward manner. It is noted that not every system can be compensated with a lead, lag, or lag–lead compensator. In some cases compensators with complex poles and zeros may be used. For systems that cannot be designed by use of the root-locus or frequency-response methods, the pole-placement method may be used. (See Chapter 12.) In a given design problem if both conventional design methods and the pole-placement method can be used, conventional methods (root-locus or frequency-response methods) usually result in a lower-order stable compensator. Note that a satisfactory design of a compensator for a complex system may require a creative application of all available design methods.

Comparison of Lead, Lag, and Lag–Lead Compensation.

1. Lead compensation achieves the desired result through the merits of its phase-lead contribution, whereas lag compensation accomplishes the result through the merits of its attenuation property at high frequencies. (In some design problems both lag compensation and lead compensation may satisfy the specifications.)
2. Lead compensation is commonly used for improving stability margins. Lead compensation yields a higher gain crossover frequency than is possible with lag compensation. The higher gain crossover frequency means a larger bandwidth. A large bandwidth means reduction in the settling time. The bandwidth of a system with lead compensation is always greater than that with lag compensation. Therefore, if a large bandwidth or fast response is desired, lead compensation should be employed. If, however, noise signals are present, then a large bandwidth may not be desirable, since it makes the system more susceptible to noise signals because of an increase in the high-frequency gain.
3. Lead compensation requires an additional increase in gain to offset the attenuation inherent in the lead network. This means that lead compensation will require a larger gain than that required by lag compensation. A larger gain, in most cases, implies larger space, greater weight, and higher cost.
4. The lead compensation may generate large signals in the system. Such large signals are not desirable because they will cause saturation in the system.
5. Lag compensation reduces the system gain at higher frequencies without reducing the system gain at lower frequencies. Since the system bandwidth is reduced, the system has a slower speed to respond. Because of the reduced high-frequency gain, the total system gain can be increased, and thereby low-frequency gain can be increased and the steady-state accuracy can be improved. Also, any high-frequency noises involved in the system can be attenuated.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

6. Lag compensation will introduce a pole-zero combination near the origin that will generate a long tail with small amplitude in the transient response.
7. If both fast responses and good static accuracy are desired, a lag–lead compensator may be employed. By use of the lag–lead compensator, the low-frequency gain can be increased (which means an improvement in steady-state accuracy), while at the same time the system bandwidth and stability margins can be increased.
8. Although a large number of practical compensation tasks can be accomplished with lead, lag, or lag–lead compensators, for complicated systems, simple compensation by use of these compensators may not yield satisfactory results. Then, different compensators having different pole–zero configurations must be employed.

Graphical Comparison. Figure 9–27(a) shows a unit-step response curve and unit-ramp response curve of an uncompensated system. Typical unit-step response and unit-ramp response curves for the compensated system using a lead, lag, and lag–lead compensator, respectively, are shown in Figures 9–27(b), (c), and (d). The system with a lead compensator exhibits the fastest response, while that with a lag compensator exhibits the slowest response, but with marked improvements in the unit-ramp response. The system with a lag–lead compensator will give a compromise; reasonable improvements in both the transient response and steady-state response can be expected. The response curves shown depict the nature of improvements that may be expected from using different types of compensators.

Feedback Compensation. A tachometer is one of the rate feedback devices. Another common rate feedback device is the rate gyro. Rate gyros are commonly used in aircraft autopilot systems.

Velocity feedback using a tachometer is very commonly used in positional servo systems. It is noted that, if the system is subjected to noise signals, velocity feedback may generate some difficulty if a particular velocity feedback scheme performs differentiation of the output signal. (The result is the accentuation of the noise effects.)

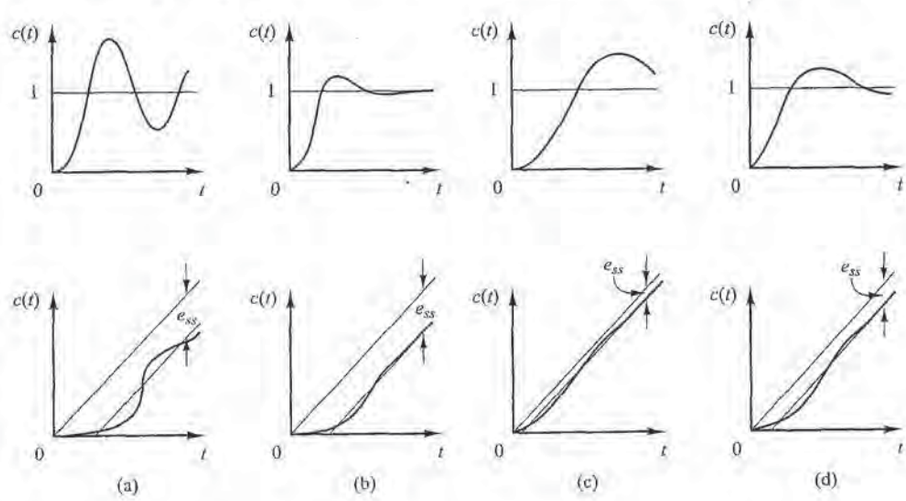


Figure 9–27
Unit-step response curves and unit-ramp response curves. (a) Uncompensated system; (b) lead compensated system; (c) lag compensated system; (d) lag–lead compensated system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Cancellation of Undesirable Poles. Since the transfer function of elements in cascade is the product of their individual transfer functions, it is possible to cancel some undesirable poles or zeros by placing a compensating element in cascade, with its poles and zeros being adjusted to cancel the undesirable poles or zeros of the original system. For example, a large time constant T_1 may be canceled by use of the lead network $(T_1 s + 1)/(T_2 s + 1)$ as follows:

$$\left(\frac{1}{T_1 s + 1}\right)\left(\frac{T_1 s + 1}{T_2 s + 1}\right) = \frac{1}{T_2 s + 1}$$

If T_2 is much smaller than T_1, we can effectively eliminate the large time constant T_1. Figure 9–28 shows the effect of canceling a large time constant in step transient response.

If an undesirable pole in the original system lies in the right-half s plane, this compensation scheme should not be used since, although mathematically it is possible to cancel the undesirable pole with an added zero, exact cancellation is physically impossible because of inaccuracies involved in the location of the poles and zeros. A pole in the right-half s plane not exactly canceled by the compensator zero will eventually lead to unstable operation, because the response will involve an exponential term that increases with time.

It is noted that if a left-half plane pole is almost canceled but not exactly canceled, as is almost always the case, the uncanceled pole-zero combination will cause the response to have a small amplitude but long-lasting transient-response component. If the cancellation is not exact but is reasonably good, then this component will be small.

It should be noted that the ideal control system is not the one that has a transfer function of unity. Physically, such a control system cannot be built since it cannot instantaneously transfer energy from the input to the output. In addition, since noise is almost always present in one form or another, a system with a unity transfer function is not desirable. A desired control system, in many practical cases, may have one set of dominant complex-conjugate closed-loop poles with a reasonable damping ratio and undamped natural frequency. The determination of the significant part of the closed-loop pole-zero configuration, such as the location of the dominant closed-loop poles, is based on the specifications that give the required system performance.

Cancellation of Undesirable Complex-Conjugate Poles. If the transfer function of a plant contains one or more pairs of complex-conjugate poles, then a lead, lag, or lag–lead compensator may not give satisfactory results. In such a case, a network that has two zeros and two poles may prove to be useful. If the zeros are chosen so as to cancel the undesirable complex-conjugate poles of the plant, then we can essentially replace the undesirable poles by acceptable poles. That is, if the undesirable complex-conjugate poles are in the left-half s plane and are in the form

$$\frac{1}{s^2 + 2\zeta_1\omega_1 s + \omega_1^2}$$

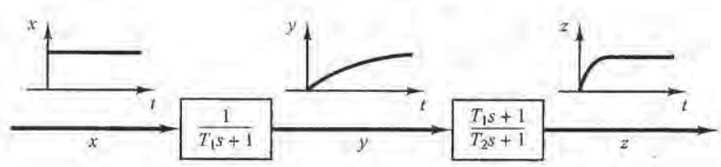
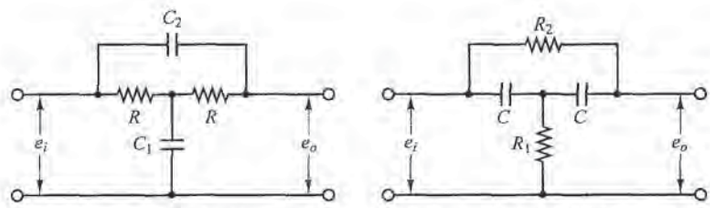


Figure 9–28
Step-response curves showing the effect of canceling a large time constant.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565



$$\frac{E_o(s)}{E_i(s)} = \frac{RC_1RC_2s^2 + 2RC_2s + 1}{RC_1RC_2s^2 + (RC_1 + 2RC_2)s + 1} \qquad \frac{E_o(s)}{E_i(s)} = \frac{R_1CR_2Cs^2 + 2R_1Cs + 1}{R_1CR_2Cs^2 + (R_2C + 2R_1C)s + 1}$$

(a) (b)

Figure 9–29
Bridged-T networks.

then the insertion of a compensating network having the transfer function

$$\frac{s^2 + 2\zeta_1\omega_1 s + \omega_1^2}{s^2 + 2\zeta_2\omega_2 s + \omega_2^2}$$

will result in an effective change of the undesirable complex-conjugate poles to acceptable poles. Note that even though the cancellation may not be exact the compensated system will exhibit better response characteristics. (As stated earlier, this approach cannot be used if the undesirable complex-conjugate poles are in the right-half s plane.)

Familiar networks consisting only of RC components whose transfer functions possess two zeros and two poles are the bridged-T networks. Examples of bridged-T networks and their transfer functions are shown in Figure 9–29. (The derivations of the transfer functions of the bridged-T networks were given in Problem **A–3–19**.)

Concluding Comments. In the design examples presented in this chapter, we have been primarily concerned only with the transfer functions of compensators. In actual design problems, we must choose the hardware. Thus, we must satisfy additional design constraints such as cost, size, weight, and reliability.

The system designed may meet the specifications under normal operating conditions but may deviate considerably from the specifications when environmental changes are considerable. Since the changes in the environment affect the gain and time constants of the system, it is necessary to provide automatic or manual means to adjust the gain to compensate for such environmental changes, for nonlinear effects that were not taken into account in the design, and also to compensate for manufacturing tolerances from unit to unit in the production of system components. (The effects of manufacturing tolerances are suppressed in a closed-loop system; therefore, the effects may not be critical in closed-loop operation but critical in open-loop operation.) In addition to this, the designer must remember that any system is subject to small variations due mainly to the normal deterioration of the system.

EXAMPLE PROBLEMS AND SOLUTIONS

A–9–1. Show that the lead network and lag network inserted in cascade in an open loop act as proportional-plus-derivative control (in the region of small ω) and proportional-plus-integral control (in the region of large ω), respectively.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

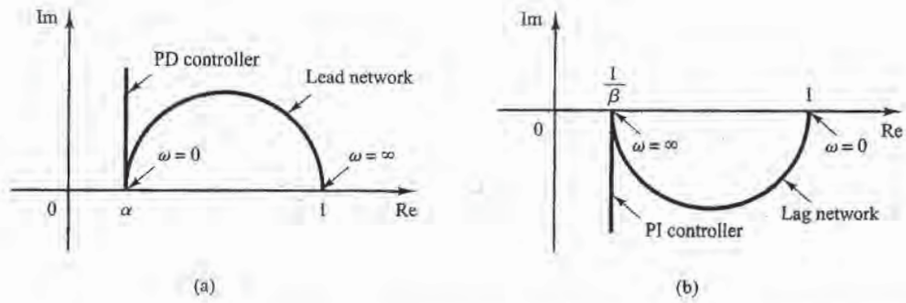


Figure 9–30
(a) Polar plots of a lead network and a proportional-plus-derivative controller; (b) polar plots of a lag network and a proportional-plus-integral controller.

Solution. In the region of small ω, the polar plot of the lead network is approximately the same as that of the proportional-plus-derivative controller. This is shown in Figure 9–30(a).

Similarly, in the region of large ω, the polar plot of the lag network approximates the proportional-plus-integral controller, as shown in Figure 9–30(b).

A–9–2. Consider a lag–lead compensator $G_c(s)$ defined by

$$G_c(s) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)}$$

Show that at frequency ω_1, where

$$\omega_1 = \frac{1}{\sqrt{T_1 T_2}}$$

the phase angle of $G_c(j\omega)$ becomes zero. (This compensator acts as a lag compensator for $0 < \omega < \omega_1$ and acts as a lead compensator for $\omega_1 < \omega < \infty$.)

Solution. The angle of $G_c(j\omega)$ is given by

$$\begin{aligned} \underline{/G_c(j\omega)} &= \underline{/j\omega + \frac{1}{T_1}} + \underline{/j\omega + \frac{1}{T_2}} - \underline{/j\omega + \frac{\beta}{T_1}} - \underline{/j\omega + \frac{1}{\beta T_2}} \\ &= \tan^{-1}\omega T_1 + \tan^{-1}\omega T_2 - \tan^{-1}\omega T_1/\beta - \tan^{-1}\omega T_2\beta \end{aligned}$$

At $\omega = \omega_1 = 1/\sqrt{T_1 T_2}$, we have

$$\underline{/G_c(j\omega_1)} = \tan^{-1}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\sqrt{\frac{T_2}{T_1}} - \tan^{-1}\frac{1}{\beta}\sqrt{\frac{T_1}{T_2}} - \tan^{-1}\beta\sqrt{\frac{T_2}{T_1}}$$

Since

$$\tan\left(\tan^{-1}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\sqrt{\frac{T_2}{T_1}}\right) = \frac{\sqrt{\frac{T_1}{T_2}} + \sqrt{\frac{T_2}{T_1}}}{1 - \sqrt{\frac{T_1}{T_2}}\sqrt{\frac{T_2}{T_1}}} = \infty$$

or

$$\tan^{-1}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\sqrt{\frac{T_2}{T_1}} = 90°$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

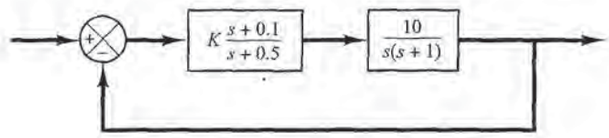


Figure 9–31
Control system.

and also

$$\tan^{-1}\frac{1}{\beta}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\beta\sqrt{\frac{T_2}{T_1}} = 90^\circ$$

we have

$$\underline{/G_c(j\omega_1)} = 0^\circ$$

Thus, the angle of $G_c(j\omega_1)$ becomes 0° at $\omega = \omega_1 = 1/\sqrt{T_1T_2}$.

A–9–3. Consider the control system shown in Figure 9–31. Determine the value of gain K such that the phase margin is 60°. What is the gain margin with this value of gain K?

Solution. The open-loop transfer function is

$$G(s) = K\frac{s+0.1}{s+0.5}\frac{10}{s(s+1)}$$
$$= \frac{K(10s+1)}{s^3+1.5s^2+0.5s}$$

Let us plot the Bode diagram of $G(s)$ when $K = 1$. MATLAB Program 9–3 may be used for this purpose. Figure 9–32 shows the Bode diagram produced by this program. From this diagram the required phase margin of 60° occurs at the frequency $\omega = 1.15$ rad/sec. The magnitude of $G(j\omega)$ at this frequency is found to be 14.5 dB. Then gain K must satisfy the following equation:

$$20\log K = -14.5 \text{ dB}$$

or

$$K = 0.188$$

MATLAB Program 9–3

```
num = [0 0 10 1];
den = [1 1.5 0.5 0];
bode(num,den)
title('Bode Diagram of G(s) = (10s + 1)/[s(s + 0.5)(s + 1)]')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

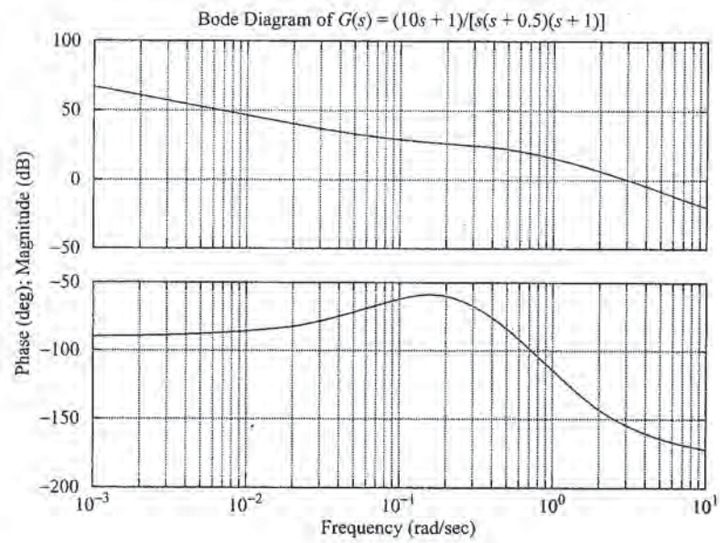


Figure 9–32
Bode diagram of
$$G(s) = \frac{10s + 1}{s(s + 0.5)(s + 1)}.$$

Thus, we have determined the value of gain K. Since the angle curve does not cross the $-180°$ line, the gain margin is $+\infty$ dB.

To verify the results, let us draw a Nyquist plot of G for the frequency range

```
w = 0.5:0.01:1.15
```

The end point of the locus ($\omega = 1.15$ rad/sec) will be on a unit circle in the Nyquist plane. To check the phase margin, it is convenient to draw the Nyquist plot on a polar diagram, using polar grids.

To draw the Nyquist plot on a polar diagram, first define a complex vector z by

$$z = re + i*im = re^{i\theta}$$

where r and θ (theta) are given by

```
r = abs(z)
theta = angle(z)
```

The abs means the square root of the sum of the real part squared and imaginary part squared, angle means $\tan^{-1}$ (imaginary part/real part).

If we use the command

```
polar(theta,r)
```

MATLAB will produce a plot in the polar coordinates. Subsequent use of the grid command draws polar grid lines and grid circles.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–4 produces the Nyquist plot of $G(j\omega)$, where ω is between 0.5 and 1.15 rad/sec. The resulting plot is shown in Figure 9–33. Notice that point $G(j1.15)$ lies on the unit circle, and the phase angle of this point is $-120°$. Hence, the phase margin is 60°. The fact that point $G(j1.15)$ is on the unit circle verifies that at $\omega = 1.15$ rad/sec the magnitude is equal to 1 or 0 dB. (Thus, $\omega = 1.15$ is the gain crossover frequency.) Thus, $K = 0.188$ gives the desired phase margin of 60°.

Note that in writing 'text' in the polar diagram we enter the text command as follows:

text(x,y,' ')

For example, to write 'Nyquist plot' starting at point $(0.1, -1.5)$, enter the command

text(0.1, −1.5,'Nyquist plot')

The text is written horizontally on the screen.

MATLAB Program 9–4

```
%*****Nyquist plot in rectangular coordinates*****

num = [0  0  1.88  0.188];
den = [1  1.5  0.5  0];
w = 0.5:0.01:1.15;
[re,im,w] = nyquist(num,den,w);

%*****Convert rectangular coordinates into polar coordinates
% by defining z, r, theta as follows*****

z = re + i*im;
r = abs(z);
theta = angle(z);

%*****To draw polar plot, enter command 'polar(theta,r)'*****

polar(theta,r)
text(-1.2,3,'Check of Phase Margin')

%*****We shall superimpose a unit circle to the polar plot*****

hold
Current plot held

k = 0:0.01:2*pi; re_k = sin(k); im_k = cos(k);
zz = re_k + i*im_k; theta_k = angle(zz); rr = abs(zz);
polar(theta_k,rr)

text(0.1,-1.45,'Nyquist plot')
text(-2,-0.3,'Phase margin').
text(-2,-0.6,'is 60 degrees')
text(1.4,-0.3,'Unit circle')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

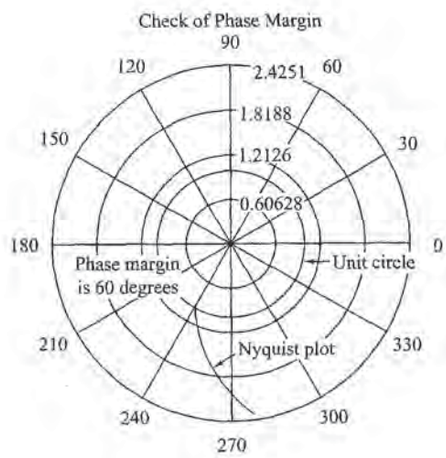


Figure 9–33
Nyquist plot of $G(j\omega)$ showing that the phase margin is 60°.

A–9–4. If the open-loop transfer function $G(s)$ involves lightly damped complex-conjugant poles, then more than one M locus may be tangent to the $G(j\omega)$ locus.

Consider the unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{9}{s(s + 0.5)(s^2 + 0.6s + 10)} \tag{9–6}$$

Draw the Bode diagram for this open-loop transfer function. Draw also the log-magnitude-versus-phase plot, and show that two M loci are tangent to the $G(j\omega)$ locus. Finally, plot the Bode diagram for the closed-loop transfer function.

Solution. Figure 9–34 shows the Bode diagram of $G(j\omega)$. Figure 9–35 shows the log-magnitude-versus-phase plot of $G(j\omega)$. It is seen that the $G(j\omega)$ locus is tangent to the $M = 8$-dB locus at $\omega = 0.97$ rad/sec, and it is tangent to the $M = -4$-dB locus at $\omega = 2.8$ rad/sec.

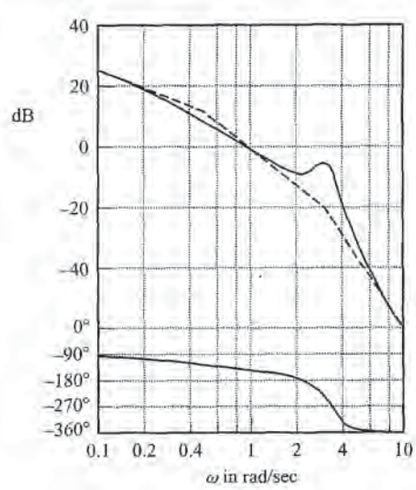


Figure 9–34
Bode diagram of $G(j\omega)$ given by Equation (9–6).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

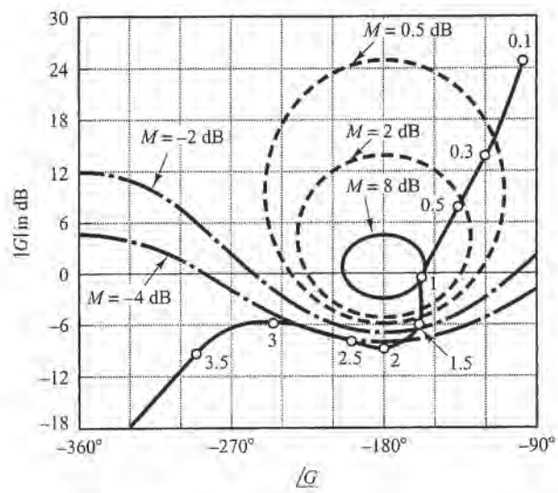


Figure 9–35
Log-magnitude-versus-phase plot of $G(j\omega)$ given by Equation (9–6).

Figure 9–36 shows the Bode diagram of the closed-loop transfer function. The magnitude curve of the closed-loop frequency response shows two resonant peaks. Note that such a case occurs when the closed-loop transfer function involves the product of two lightly damped second-order terms and the two corresponding resonant frequencies are sufficiently separated from each other. As a matter of fact, the closed-loop transfer function of this system can be written

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

$$= \frac{9}{(s^2 + 0.487s + 1)(s^2 + 0.613s + 9)}$$

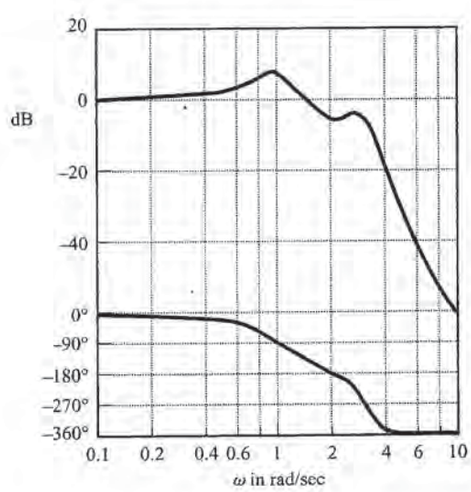


Figure 9–36
Bode diagram of $G(j\omega)/[1 + G(j\omega)]$, where $G(j\omega)$ is given by Equation (9–6).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Clearly, the closed-loop transfer function is a product of two lightly damped second-order terms (the damping ratios are 0.243 and 0.102), and the two resonant frequencies are sufficiently separated.

A–9–5. Consider a unity-feedback system whose feedforward transfer function is given by

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

It is desired to insert a series compensator so that the open-loop frequency-response curve is tangent to the $M = 3$-dB circle at $\omega = 3$ rad/sec. The system is subjected to high-frequency noises and sharp cutoff is desired. Design an appropriate series compensator.

Solution. To stabilize the system, we may insert a proportional-plus-derivative type of compensator or a lead compensator. Since sharp cutoff is required, we choose a lead compensator. Consider the following lead compensator:

$$G_c(s) = K_c \frac{Ts + 1}{\alpha Ts + 1} \qquad (\alpha < 1)$$

The compensated open-loop frequency-response curve must be tangent to the $M = 3$-dB locus. To minimize the additional gain K_c, we choose the tangent point to the 3-dB locus as shown in Figure 9–37. From Figure 9–37 we see that the lead compensator must provide about 45°. Then the necessary value of α is determined from

$$\sin 45° = \frac{1 - \alpha}{1 + \alpha}$$

or $\alpha = 0.172 \doteq \frac{1}{6}$. Let us choose $\alpha = \frac{1}{6}$. Since it is required that the open-loop frequency-response curve $G_c(j\omega)G(j\omega)$ be tangent to the $M = 3$-dB locus at $\omega = 3$ rad/sec, we obtain

$$\begin{aligned} 20 \log|G_c(j\omega)G(j\omega)|_{\omega=3} &= 20 \log|G_c(j3)| + 20 \log|G(j3)| \\ &= 20 \log|G_c(j3)| + 20 \log\left|\frac{1}{9}\right| = 3 \text{ dB} \end{aligned}$$

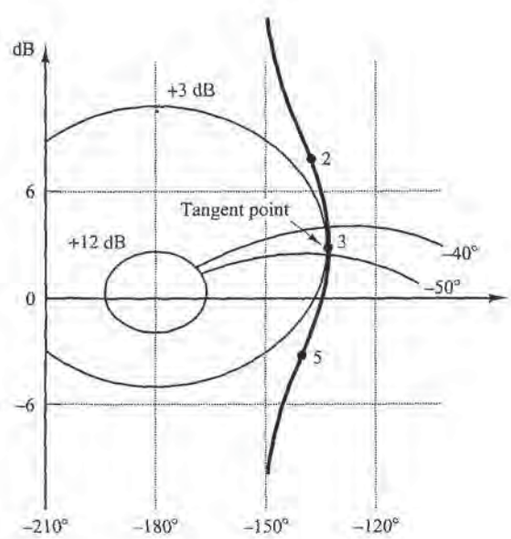


Figure 9–37
Nichols chart showing that the $G_c(j\omega)G(j\omega)$ locus is tangent to the $M = 3$-dB locus at $\omega = 3$ rad/sec.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

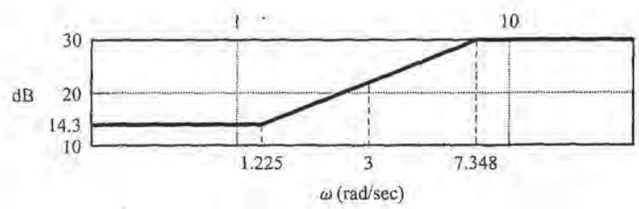


Figure 9–38
Bode diagram of the lead compensator designed in Problem A–9–5.

or

$$20 \log|G_c(j3)| = 22.085 \text{ dB}$$

The two time constants T and αT of the lead compensator can be determined as follows: Noting that

$$\sqrt{\frac{1}{T} \cdot \frac{1}{\alpha T}} = 3$$

we have

$$\frac{1}{T} = \frac{3}{\sqrt{6}} = 1.225, \qquad \frac{1}{\alpha T} = 3\sqrt{6} = 7.348$$

From the Bode diagram as shown in Figure 9–38, we find the gain K_c to be 14.3 dB or 5.19. Thus, the designed compensator is given by

$$G_c(s) = 5.19 \frac{0.816s + 1}{0.136s + 1}$$

A–9–6. Consider the system shown in Figure 9–39. Design a lead compensator such that the closed-loop system will have the phase margin of 50° and gain margin of not less than 10 dB. Assume that

$$G_c(s) = K_c\alpha\left(\frac{Ts + 1}{\alpha Ts + 1}\right) \qquad (0 < \alpha < 1)$$

It is desired that the bandwidth of the closed-loop system be 1 ~ 2 rad/sec. What are the values of M_r and ω_r of the compensated system?

Solution. Notice that

$$G_c(j\omega)G(j\omega) = K_c\alpha\left(\frac{Tj\omega + 1}{\alpha Tj\omega + 1}\right)\frac{0.2}{(j\omega)^2(0.2j\omega + 1)}$$

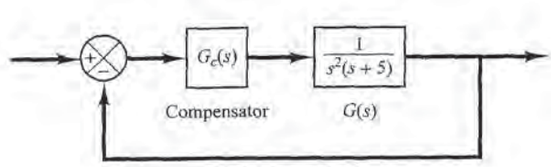


Figure 9–39
Closed-loop system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Since the bandwidth of the closed-loop system is close to the gain crossover frequency, we choose the gain crossover frequency to be 1 rad/sec. At $\omega = 1$, the phase angle of $G(j\omega)$ is $-191.31°$. Hence, the lead network needs to supply $50° + 11.31° = 61.31°$ at $\omega = 1$. Hence, α can be determined from

$$\sin\phi_m = \sin 61.31° = \frac{1-\alpha}{1+\alpha} = 0.8772$$

as follows:

$$\alpha = 0.06541$$

Noting that the maximum phase-lead angle ϕ_m occurs at the geometric mean of the two corner frequencies, we have

$$\omega_m = \sqrt{\frac{1}{T}\frac{1}{\alpha T}} = \frac{1}{\sqrt{\alpha}T} = \frac{1}{\sqrt{0.06541}T} = \frac{3.910}{T} = 1$$

Thus,

$$\frac{1}{T} = \frac{1}{3.910} = 0.2558$$

and

$$\frac{1}{\alpha T} = \frac{0.2558}{0.06541} = 3.910$$

Hence,

$$G_c(j\omega)G(j\omega) = 0.06541K_c\frac{3.910j\omega + 1}{0.2558j\omega + 1}\frac{0.2}{(j\omega)^2(0.2j\omega + 1)}$$

or

$$\frac{G_c(j\omega)G(j\omega)}{0.06541K_c} = \frac{3.910j\omega + 1}{0.2558j\omega + 1}\frac{0.2}{(j\omega)^2(0.2j\omega + 1)}$$

A Bode diagram for $G_c(j\omega)G(j\omega)/(0.06541\ K_c)$ is shown in Figure 9–40. By simple calculations (or from the Bode diagram), we find that the magnitude curve must be raised by 2.306 dB so that the magnitude equals 0 dB at $\omega = 1$ rad/sec. Hence, we set

$$20\log 0.06541\ K_c = 2.306$$

or

$$0.06541\ K_c = 1.3041$$

which yields

$$K_c = 19.94$$

The magnitude and phase curves of the compensated system show that the system has the phase margin of 50° and gain margin of 16 dB. Hence, the design specifications are satisfied.

Figure 9–41 shows the $G_c(j\omega)G(j\omega)$ locus superimposed on the Nichols chart. From this diagram, we find the bandwidth to be approximately 1.9 rad/sec. The values of M_r and ω_r are read from this diagram as follows:

$$M_r = 2.13\text{ dB} \qquad \omega_r = 0.58\text{ rad/sec}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

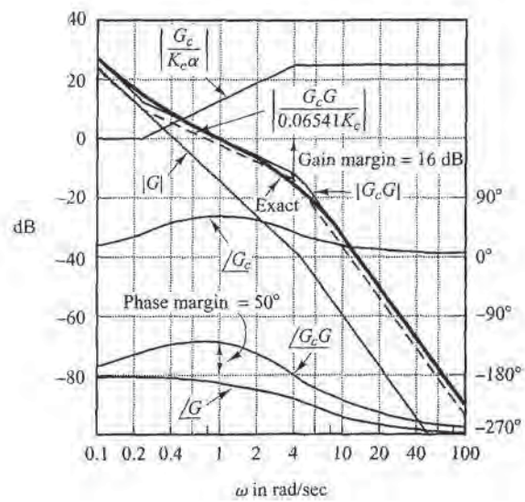


Figure 9–40
Bode diagram of the system shown in Figure 9–39.

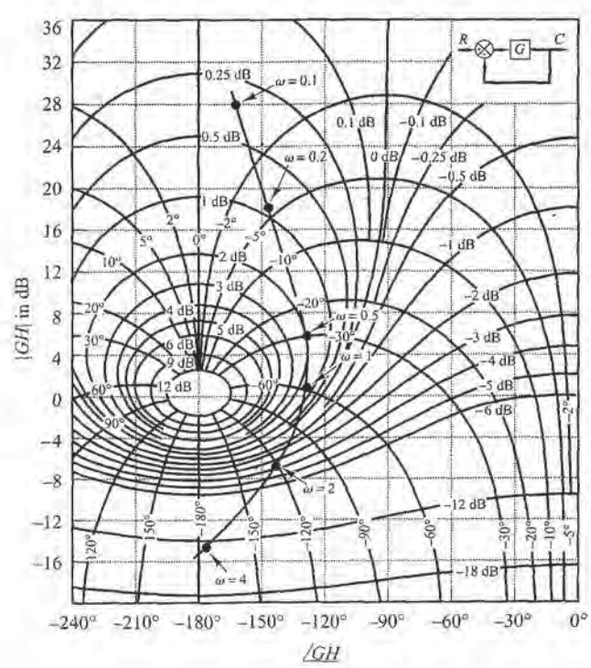


Figure 9–41
$G_c(j\omega)G(j\omega)$ locus superimposed on Nichols chart (Problem A–9–6).

A–9–7. Referring to Example 9–1, draw Nyquist plots of $G(j\omega)$, $G_1(j\omega)$, and $G_c(j\omega)G(j\omega)$ with MATLAB. (Compare the Nyquist plots obtained here with Figure 9–10.) Write a possible MATLAB program

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

for drawing the Nyquist plots in one diagram. Note that $G(j\omega)$, $G_1(j\omega)$, and $G_c(j\omega)G(j\omega)$ are given by

$$G(j\omega) = \frac{4}{j\omega(j\omega + 2)}$$

$$G_1(j\omega) = \frac{40}{j\omega(j\omega + 2)}$$

$$G_c(j\omega)G(j\omega) = 41.7\frac{j\omega + 4.41}{j\omega + 18.4}\frac{4}{j\omega(j\omega + 2)}$$

Solution. A possible MATLAB program for this problem is given in MATLAB Program 9–5. The resulting Nyquist plots are shown in Figure 9–42.

MATLAB Program 9–5

```
%*****Nyquist plots in polar coordinates*****

num1 = [0  0  4];
den1 = [1  2  0];
num2 = [0  0  40];
den2 = [1  2  0];
num3 = [0  0  166.8  735.588];
den3 = [1  20.4  36.8  0];
w = 0.2:0.1:10;
ww = 1.5:0.1:10;
[re1,im1,w] = nyquist(num1,den1,w);
z1 = re1 + i*im1;
r1 = abs(z1);
theta1 = angle(z1);
polar(theta1,r1,'o')
hold
Current plot held
[re2,im2,ww] = nyquist(num2,den2,ww);
z2 = re2 + i*im2;
r2 = abs(z2);
theta2 = angle(z2);
polar(theta2,r2,'o')
[re3,im3,ww] = nyquist(num3,den3,ww);
z3 = re3 + i*im3;
r3 = abs(z3);
theta3 = angle(z3);
polar(theta3,r3,'x')
text(-11,12.4,'Nyquist Plots of G(jw), G1(jw), and Gc(jw)G(jw)')
text(0.7,-8.8,'G(jw)')
text(-11.7,-8.5,'G1(jw)')
text(-6.6,-12.3,'Gc(jw)G(jw)')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

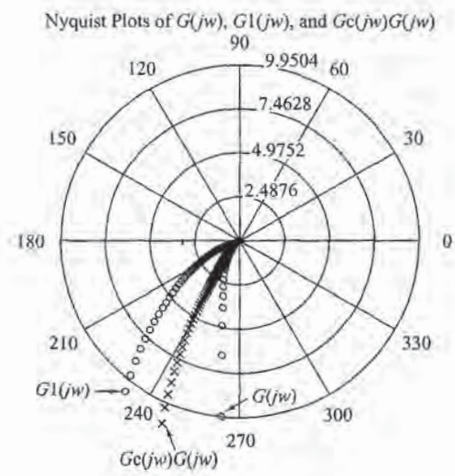


Figure 9–42
Nyquist plots of $G(j\omega)$, $G_1(j\omega)$, and $G_c(j\omega)G(j\omega)$.

A–9–8. Consider the system shown in Figure 9–43(a). Design a compensator such that the closed-loop system will satisfy the requirements that the static velocity error constant = 20 sec^{-1}, phase margin = 50°, and gain margin ≧ 10 dB.

Solution. To satisfy the requirements, we shall try a lead compensator $G_c(s)$ of the form

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1}$$

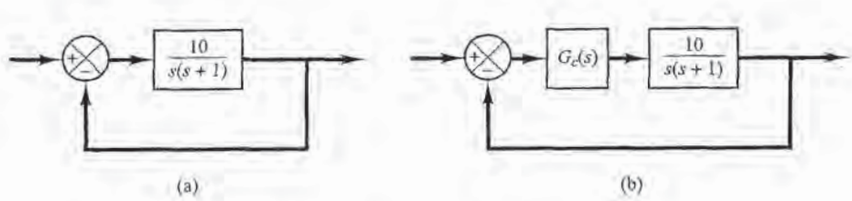
$$= K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}$$

(If the lead compensator does not work, then we shall employ a compensator of different form.) The compensated system is shown in Figure 9–43(b).

Define

$$G_1(s) = KG(s) = \frac{10K}{s(s+1)}$$

where $K = K_c\alpha$. The first step in the design is to adjust the gain K to meet the steady-state performance specification or to provide the required static velocity error constant. Since the static velocity error constant K_v is given as 20 sec^{-1}, we have



(a) (b)

Figure 9–43
(a) Control system; (b) compensated system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$
\begin{aligned}
K_v &= \lim_{s \to 0} sG_c(s)G(s) \\
&= \lim_{s \to 0} s \frac{Ts + 1}{\alpha Ts + 1} G_1(s) \\
&= \lim_{s \to 0} \frac{s10K}{s(s + 1)} \\
&= 10K = 20
\end{aligned}
$$

or

$$K = 2$$

With $K = 2$, the compensated system will satisfy the steady-state requirement.

We shall next plot the Bode diagram of

$$G_1(s) = \frac{20}{s(s + 1)}$$

MATLAB Program 9–6 produces the Bode diagram shown in Figure 9–44. From this plot, the phase margin is found to be 14°. The gain margin is $+\infty$ dB.

MATLAB Program 9–6

```
num = [0 0 20];
den = [1 1 0];
w = logspace(-1,2,100);
bode(num,den,w)
title('Bode Diagram of G1(s) = 20/[s(s + 1)]')
```

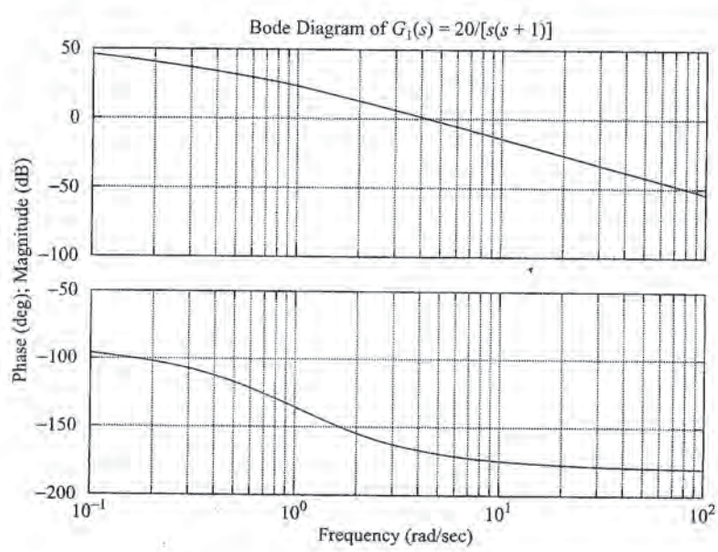


Figure 9–44
Bode diagram of $G_1(s)$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Since the specification calls for a phase margin of 50°, the additional phase lead necessary to satisfy the phase-margin requirement is 36°. A lead compensator can contribute this amount.

Noting that the addition of a lead compensator modifies the magnitude curve in the Bode diagram, we realize that the gain crossover frequency will be shifted to the right. We must offset the increased phase lag of $G_1(j\omega)$ due to this increase in the gain crossover frequency. Taking the shift of the gain crossover frequency into consideration, we may assume that ϕ_m, the maximum phase lead required, is approximately 41°. (This means that approximately 5° has been added to compensate for the shift in the gain crossover frequency.) Since

$$\sin\phi_m = \frac{1-\alpha}{1+\alpha}$$

$\phi_m = 41°$ corresponds to $\alpha = 0.2077$. Note that $\alpha = 0.21$ corresponds to $\phi_m = 40.76°$. Whether we choose $\phi_m = 41°$ or $\phi_m = 40.76°$ does not make much difference in the final solution. Hence, let us choose $\alpha = 0.21$.

Once the attenuation factor α has been determined on the basis of the required phase-lead angle, the next step is to determine the corner frequencies $\omega = 1/T$ and $\omega = 1/(\alpha T)$ of the lead compensator. Notice that the maximum phase-lead angle ϕ_m occurs at the geometric mean of the two corner frequencies, or $\omega = 1/(\sqrt{\alpha}T)$.

The amount of the modification in the magnitude curve at $\omega = 1/(\sqrt{\alpha}T)$ due to the inclusion of the term $(Ts+1)/(\alpha Ts+1)$ is

$$\left|\frac{1+j\omega T}{1+j\omega\alpha T}\right|_{\omega=\frac{1}{\sqrt{\alpha}T}} = \left|\frac{1+j\dfrac{1}{\sqrt{\alpha}}}{1+j\alpha\dfrac{1}{\sqrt{\alpha}}}\right| = \frac{1}{\sqrt{\alpha}}$$

Note that

$$\frac{1}{\sqrt{\alpha}} = \frac{1}{\sqrt{0.21}} = 6.7778 \text{ dB}$$

We need to find the frequency point where, when the lead compensator is added, the total magnitude becomes 0 dB.

From Figure 9–44 we see that the frequency point where the magnitude of $G_1(j\omega)$ is -6.7778 dB occurs between $\omega = 1$ and 10 rad/sec. Hence, we plot a new Bode diagram of $G_1(j\omega)$ in the frequency range between $\omega = 1$ and 10 to locate the exact point where $G_1(j\omega) = -6.7778$ dB. MATLAB Program 9–7 produces the Bode diagram in this frequency range, which is shown in Figure 9–45. From this diagram, we find the frequency point where $|G_1(j\omega)| = -6.7778$ dB

MATLAB Program 9–7

```
num = [0 0 20];
den = [1 1 0];
w = logspace(0,1,100);
bode(num,den,w)
title('Bode Diagram of G1(s) = 20/[s(s + 1)]')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

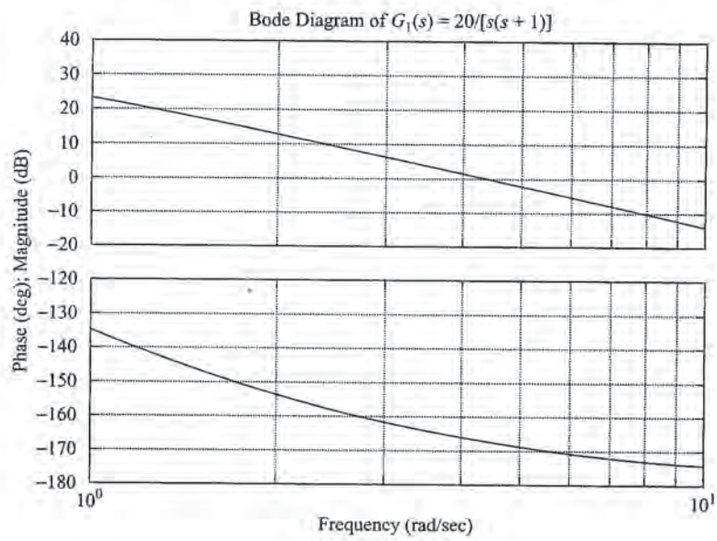


Figure 9–45
Bode diagram of $G_1(s)$.

occurs at $\omega = 6.5$ rad/sec. Let us select this frequency to be the new gain crossover frequency, or $\omega_c = 6.5$ rad/sec. Noting that this frequency corresponds to $1/(\sqrt{\alpha}T)$, or

$$\omega_c = \frac{1}{\sqrt{\alpha}T}$$

we obtain

$$\frac{1}{T} = \omega_c\sqrt{\alpha} = 6.5\sqrt{0.21} = 2.9787$$

and

$$\frac{1}{\alpha T} = \frac{\omega_c}{\sqrt{\alpha}} = \frac{6.5}{\sqrt{0.21}} = 14.1842$$

The lead compensator thus determined is

$$G_c(s) = K_c\frac{s + 2.9787}{s + 14.1842} = K_c\,\alpha\frac{0.3357s + 1}{0.07050s + 1}$$

where K_c is determined as

$$K_c = \frac{K}{\alpha} = \frac{2}{0.21} = 9.5238$$

Thus, the transfer function of the compensator becomes

$$G_c(s) = 9.5238\frac{s + 2.9787}{s + 14.1842} = 2\frac{0.3357s + 1}{0.07050s + 1}$$

MATLAB Program 9–8 produces the Bode diagram of this lead compensator, which is shown in Figure 9–46.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–8

```
numc = [9.5238  28.3685];
denc = [1  14.1842];
w = logspace(-1,3,100);
bode(numc,denc,w)
title('Bode Diagram of Gc(s) = 9.5238(s + 2.9787)/(s + 14.1842)')
```

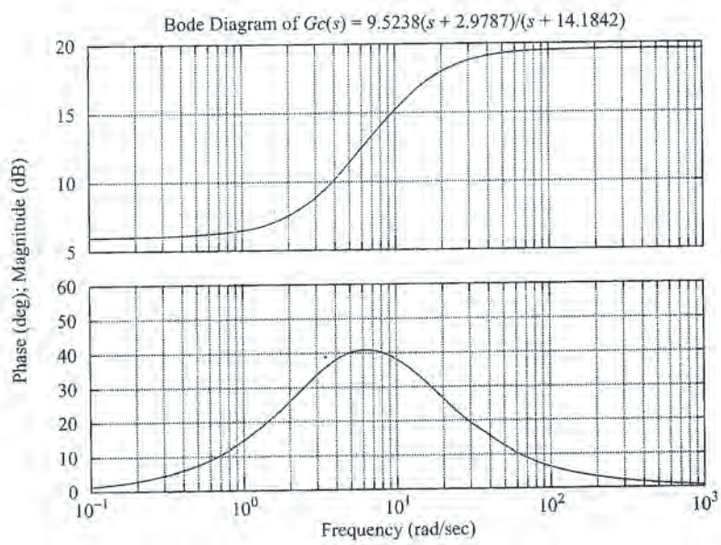


Figure 9–46
Bode diagram of $G_c(s)$.

The open-loop transfer function of the designed system is

$$G_c(s)G(s) = 9.5238\,\frac{s + 2.9787}{s + 14.1842}\,\frac{10}{s(s + 1)}$$

$$= \frac{95.238s + 283.6854}{s^3 + 15.1842s^2 + 14.1842s}$$

MATLAB Program 9–9 will produce the Bode diagram of $G_c(s)G(s)$, which is shown in Figure 9–47.

MATLAB Program 9–9

```
num = [0  0  95.238  283.6854];
den = [1  15.1842  14.1842  0];
w = logspace(−1,3,100);
bode(num,den,w)
title('Bode Diagram of Gc(s)G(s)')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

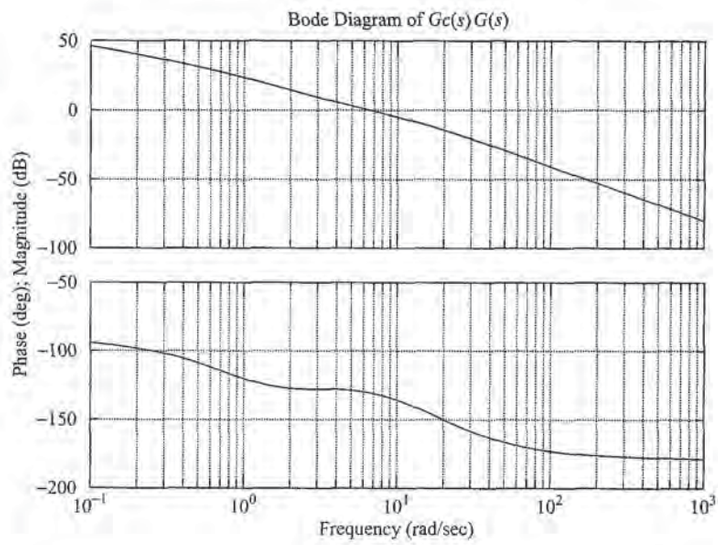


Figure 9–47
Bode diagram of $G_c(s)G(s)$.

From Figure 9–47 it is clearly seen that the phase margin is approximately 50° and the gain margin is $+\infty$ dB. Since the static velocity error constant K_v is 20 sec^{-1}, all the specifications are met. Before we conclude this problem, we need to check the transient-response characteristics.

Unit-Step Response: We shall compare the unit-step response of the compensated system with that of the original uncompensated system.

The closed-loop transfer function of the original uncompensated system is

$$\frac{C(s)}{R(s)} = \frac{10}{s^2 + s + 10}$$

The closed-loop transfer function of the compensated system is

$$\frac{C(s)}{R(s)} = \frac{95.238s + 283.6854}{s^3 + 15.1842s^2 + 109.4222s + 283.6854}$$

MATLAB Program 9–10 produces the unit-step responses of the uncompensated and compensated systems. The resulting response curves are shown in Figure 9–48. Clearly, the compensated system exhibits a satisfactory response. Note that the closed-loop zero and poles are located as follows:

$$\text{Zero at } s = -2.9787$$

$$\text{Poles at } s = -5.2270 \pm j5.7141, \quad s = -4.7303$$

Unit-Ramp Response: It is worthwhile to check the unit-ramp response of the compensated system. Since $K_v = 20$ sec^{-1}, the steady-state error following the unit-ramp input will be $1/K_v = 0.05$. The static velocity error constant of the uncompensated system is 10 sec^{-1}. Hence, the original uncompensated system will have twice as large a steady-state error in following the unit-ramp input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–10

```
%*****Unit-step responses*****
num1 = [0 0 10];
den1 = [1 1 10];
num2 = [0 0 95.238 283.6854];
den2 = [1 15.1842 109.4222 283.6854];
t = 0:0.01:6;
[c1,x1,t] = step(num1,den1,t);
[c2,x2,t] = step(num2,den2,t);
plot(t,c1,'.',t,c2,'-')
grid
title('Unit-Step Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(1.1,0.5,'Compensated system')
text(1.7,1.46,'Uncompensated system')
```

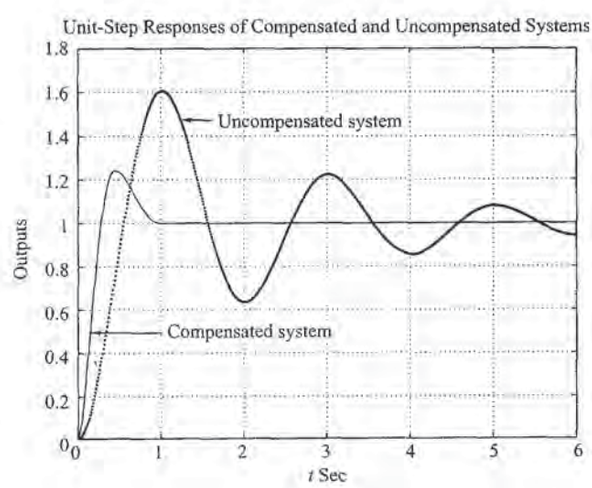


Figure 9–48
Unit-step responses of the compensated and uncompensated systems.

MATLAB Program 9–11 produces the unit-ramp response curves. [Note that the unit-ramp response is obtained as the unit-step response of $C(s)/sR(s)$.] The resulting curves are shown in Figure 9–49. The compensated system has a steady-state error equal to one-half that of the original uncompensated system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–11

```
%*****Unit-ramp responses*****
num1 = [0 0 0 10];
den1 = [1 1 10 0];
num2 = [0 0 0 95.238 283.6854];
den2 = [1 15.1842 109.4222 283.6854 0];
t = 0:0.01:3;
[c1,x1,t] = step(num1,den1,t);
[c2,x2,t] = step(num2,den2,t);
plot(t,c1,'.',t,c2,'-',t,t,'--')
grid
title('Unit-Ramp Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs')
text(0.1,1.3,'Compensated system')
text(1.2,0.65,'Uncompensated system')
```

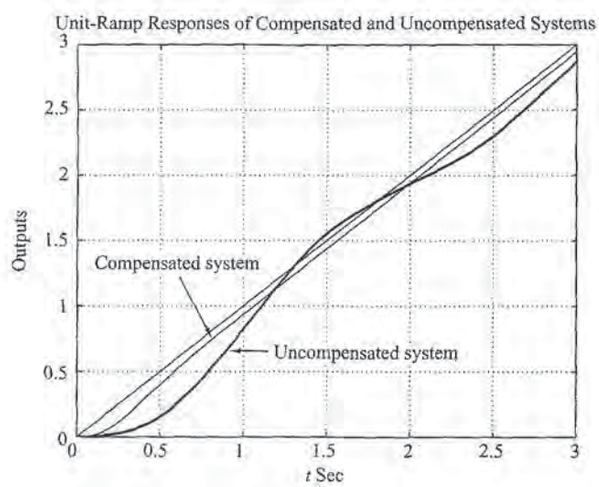


Figure 9–49
Unit-ramp responses of the compensated and uncompensated systems.

A–9–9. Consider the unity-feedback system whose open-loop transfer function is

$$G(s) = \frac{K}{s(s+1)(s+4)}$$

Design a lag–lead compensator $G_c(s)$ such that the static velocity error constant is 10 $\sec^{-1}$, the phase margin is 50°, and the gain margin is 10 dB or more.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Solution. We shall design a lag–lead compensator of the form

$$G_c(s) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)}$$

Then the open-loop transfer function of the compensated system is $G_c(s)G(s)$. Since the gain K of the plant is adjustable, let us assume that $K_c = 1$. Then $\lim_{s \to 0} G_c(s) = 1$. From the requirement on the static velocity error constant, we obtain

$$K_v = \lim_{s \to 0} sG_c(s)G(s) = \lim_{s \to 0} sG_c(s) \frac{K}{s(s+1)(s+4)}$$

$$= \frac{K}{4} = 10$$

Hence,

$$K = 40$$

We shall first plot a Bode diagram of the uncompensated system with $K = 40$. MATLAB Program 9–12 may be used to plot this Bode diagram. The diagram obtained is shown in Figure 9–50.

MATLAB Program 9–12

```
num = [0 0 0 40];
den = [1 5 4 0];
w = logspace(-1,1,100);
bode(num,den,w)
title('Bode Diagram of G(s) = 40/[s(s + 1)(s + 4)]')
```

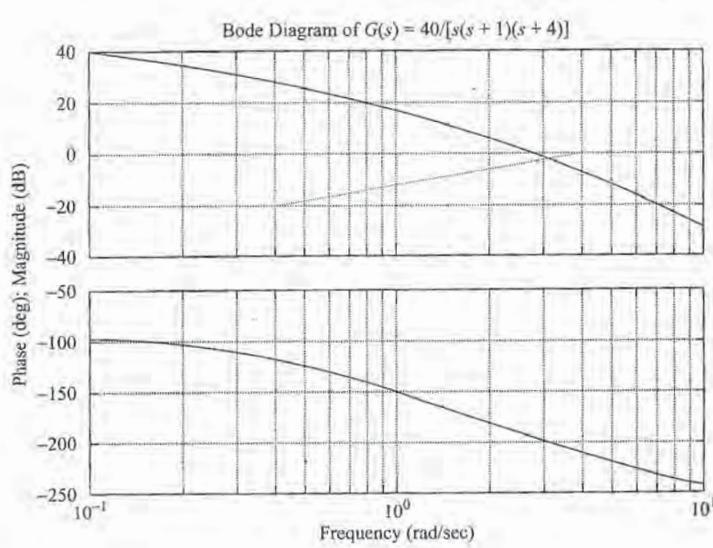


Figure 9–50
Bode diagram of $G(s) = 40/[s(s+1)(s+4)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

From Figure 9–50, the phase margin of the gain-adjusted but uncompensated system is found to be $-16°$, which indicates that this system is unstable. The next step in the design of a lag–lead compensator is to choose a new gain crossover frequency. From the phase-angle curve for $G(j\omega)$, we notice that the phase crossover frequency is $\omega = 2$ rad/sec. We may choose the new gain crossover frequency to be 2 rad/sec so that the phase-lead angle required at $\omega = 2$ rad/sec is about 50°. A single lag–lead compensator can provide this amount of phase-lead angle quite easily.

Once we choose the gain crossover frequency to be 2 rad/sec, we can determine the corner frequencies of the phase-lag portion of the lag–lead compensator. Let us choose the corner frequency $\omega = 1/T_2$ (which corresponds to the zero of the phase-lag portion of the compensator) to be 1 decade below the new gain crossover frequency, or at $\omega = 0.2$ rad/sec. For another corner frequency $\omega = 1/(\beta T_2)$, we need the value of β. The value of β can be determined from the consideration of the lead portion of the compensator, as shown next.

For the lead compensator, the maximum phase-lead angle ϕ_m is given by

$$\sin\phi_m = \frac{\beta - 1}{\beta + 1}$$

Notice that $\beta = 10$ corresponds to $\phi_m = 54.9°$. Since we need a 50° phase margin, we may choose $\beta = 10$. (Note that we will be using several degrees less than the maximum angle, 54.9°.) Thus,

$$\beta = 10$$

Then the corner frequency $\omega = 1/(\beta T_2)$ (which corresponds to the pole of the phase-lag portion of the compensator) becomes

$$\omega = 0.02$$

The transfer function of the phase-lag portion of the lag–lead compensator becomes

$$\frac{s + 0.2}{s + 0.02} = 10\left(\frac{5s + 1}{50s + 1}\right)$$

The phase-lead portion can be determined as follows: Since the new gain crossover frequency is $\omega = 2$ rad/sec, from Figure 9–50, $|G(j2)|$ is found to be 6 dB. Hence, if the lag–lead compensator contributes -6 dB at $\omega = 2$ rad/sec, then the new gain crossover frequency is as desired. From this requirement, it is possible to draw a straight line of slope 20 dB/decade passing through the point (2 rad/sec, -6 dB). (Such a line has been manually drawn on Figure 9–50.) The intersections of this line and the 0-dB line and -20-dB line determine the corner frequencies. From this consideration, the corner frequencies for the lead portion can be determined as $\omega = 0.4$ rad/sec and $\omega = 4$ rad/sec. Thus, the transfer function of the lead portion of the lag–lead compensator becomes

$$\frac{s + 0.4}{s + 4} = \frac{1}{10}\left(\frac{2.5s + 1}{0.25s + 1}\right)$$

Combining the transfer functions of the lag and lead portions of the compensator, we can obtain the transfer function $G_c(s)$ of the lag–lead compensator. Since we chose $K_c = 1$, we have

$$G_c(s) = \frac{s + 0.4}{s + 4}\,\frac{s + 0.2}{s + 0.02} = \frac{(2.5s + 1)(5s + 1)}{(0.25s + 1)(50s + 1)}$$

The Bode diagram of the lag–lead compensator $G_c(s)$ can be obtained by entering MATLAB Program 9–13 into the computer. The resulting plot is shown in Figure 9–51.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–13

```
numc = [1  0.6  0.08];
denc = [1  4.02  0.08];
bode(numc,denc)
title('Bode Diagram of Lag–Lead Compensator')
```

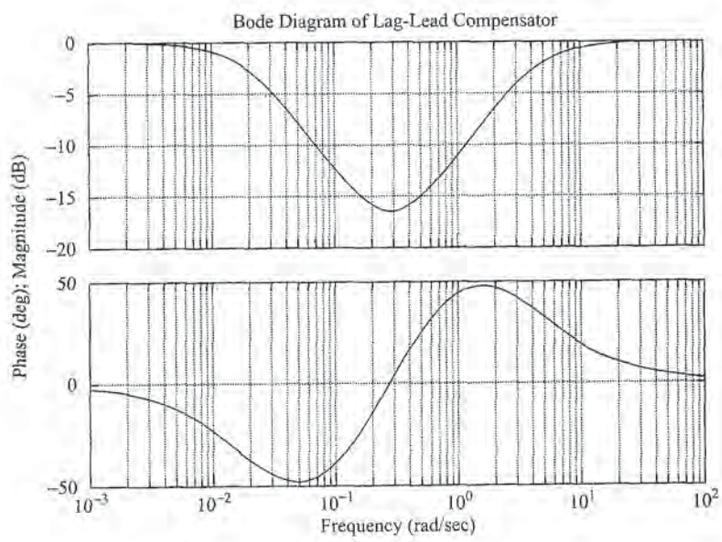


Figure 9–51
Bode diagram of the designed lag–lead compensator.

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = \frac{(s+0.4)(s+0.2)}{(s+4)(s+0.02)}\frac{40}{s(s+1)(s+4)}$$

$$= \frac{40s^2 + 24s + 3.2}{s^5 + 9.02s^4 + 24.18s^3 + 16.48s^2 + 0.32s}$$

Using MATLAB Program 9–14 the magnitude and phase-angle curves of the designed open-loop transfer function $G_c(s)G(s)$ can be obtained as shown in Figure 9–52. Note that the denominator polynomial den1 was obtained using the conv command, as follows:

```
a = [1  4.02  0.08];.
b = [1  5  4  0];
conv(a,b)

ans =

      1.0000  9.0200  24.1800  16.4800  0.320000  0
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 9–14

```
num1 = [0 0 0 40 24 3.2];
den1 = [1 9.02 24.18 16.48 0.32 0];
bode(num1,den1)
title('Bode Diagram of Gc(s)G(s)')
```

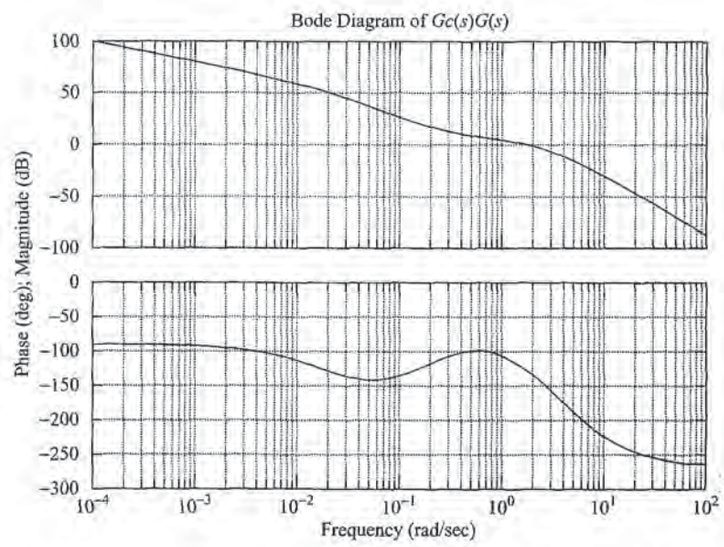


Figure 9–52
Bode diagram of the open-loop transfer function $G_c(s)G(s)$ of the compensated system.

Since the phase margin of the compensated system is 50°, the gain margin is 12 dB, and the static velocity error constant is 10 $\sec^{-1}$, all the requirements are met.

We shall next investigate the transient-response characteristics of the designed system.

Unit-Step Response: Noting that

$$G_c(s)G(s) = \frac{40(s + 0.4)(s + 0.2)}{(s + 4)(s + 0.02)s(s + 1)(s + 4)}$$

we have

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

$$= \frac{40(s + 0.4)(s + 0.2)}{(s + 4)(s + 0.02)s(s + 1)(s + 4) + 40(s + 0.4)(s + 0.2)}$$

To determine the denominator polynomial with MATLAB, we may proceed as follows: Define

$$a(s) = (s + 4)(s + 0.02) = s^2 + 4.02s + 0.08$$

$$b(s) = s(s + 1)(s + 4) = s^3 + 5s^2 + 4s$$

$$c(s) = 40(s + 0.4)(s + 0.2) = 40s^2 + 24s + 3.2$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Then we have

```
a = [1  4.02  0.08]
b = [1  5  4  0]
c = [40  24  3.2]
```

Using the following MATLAB program, we obtain the denominator polynomial.

```
a = [1  4.02  0.08];
b = [1  5  4  0];
c = [40  24  3.2];
p = [conv(a,b)] + [0  0  0  c]
p =
    1.0000  9.0200  24.1800  56.4800  24.3200  3.2000
```

MATLAB Program 9–15 is used to obtain the unit-step response of the compensated system. The resulting unit-step response curve is shown in Figure 9–53. (Note that the gain-adjusted but uncompensated system is unstable.)

MATLAB Program 9–15

```
%*****Unit-step response****

num = [0  0  0  40  24  3.2];
den = [1  9.02  24.18  56.48  24.32  3.2];
t = 0:0.2:40;
step(num,den,t)
grid
title('Unit-Step Response of Compensated System')
```

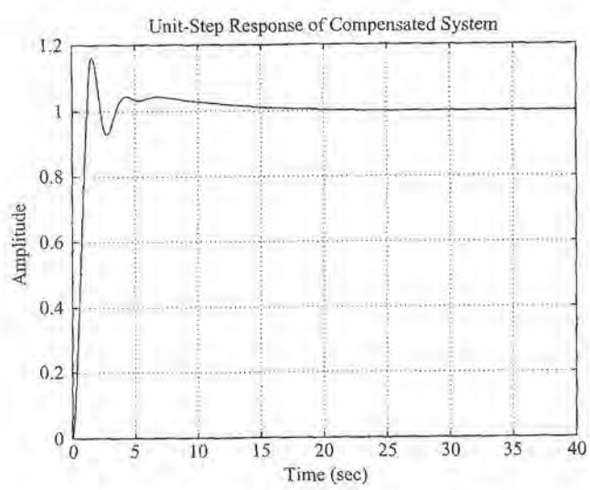


Figure 9–53
Unit-step response curve of the compensated system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Unit-Ramp Response: The unit-ramp response of this system may be obtained by entering MATLAB Program 9–16 into the computer. Here we converted the unit-ramp response of $G_cG/(1 + G_cG)$ into the unit-step response of $G_cG/[s(1 + G_cG)]$. The unit-ramp response curve obtained using this program is shown in Figure 9–54.

MATLAB Program 9–16

```
%*****Unit-ramp response*****

num = [0 0 0 0 40 24 3.2];
den = [1 9.02 24.18 56.48 24.32 3.2 0];
t = 0:0.05:20;
c = step(num,den,t);
plot(t,c,'-',t,t,'.')
grid
title('Unit-Ramp Response of Compensated System')
xlabel('Time (sec)')
ylabel('Unit-Ramp Input and Output c(t)')
```

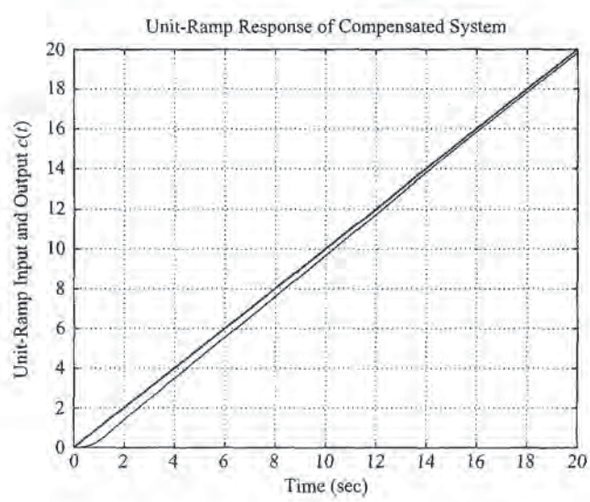


Figure 9–54
Unit-ramp response of the compensated system.

A–9–10. Consider the system shown in Figure 9–55. Design a compensator such that the static velocity error constant is 4 sec^{-1}, phase margin is 50°, and gain margin is 10 dB or more. Obtain the unit-step and unit-ramp response curves of the compensated system with MATLAB.

Solution. Since the plant does not have an integrator, it is necessary to add an integrator in the compensator. Let us choose the compensator to be

$$G_c(s) = \frac{K}{s}\hat{G}_c(s)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

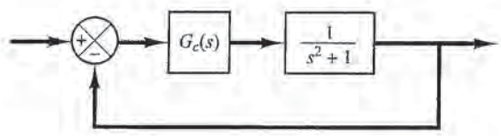


Figure 9–55
Control system.

where

$$\lim_{s \to 0} \hat{G}_c(s) = 1$$

[$\hat{G}_c(s)$ is to be determined later.] Since the static velocity error constant K_v is specified as 4 sec^{-1}, we have

$$K_v = \lim_{s \to 0} sG_c(s) \frac{1}{s^2 + 1} = \lim_{s \to 0} s \frac{K}{s} \hat{G}_c(s) \frac{1}{s^2 + 1}$$
$$= K = 4$$

Thus

$$G_c(s) = \frac{4}{s} \hat{G}_c(s)$$

Next, we plot a Bode diagram of

$$G(s) = \frac{4}{s(s^2 + 1)}$$

MATLAB Program 9–17 produces a Bode diagram of $G(s)$. The resulting Bode diagram is shown in Figure 9–56.

MATLAB Program 9–17

```
num = [0 0 0 4];
den = [1 0.00000000001 1 0];
w = logspace(-1,1,200);
bode(num,den,w)
title('Bode Diagram of 4/[s(s^2+1)]')
```

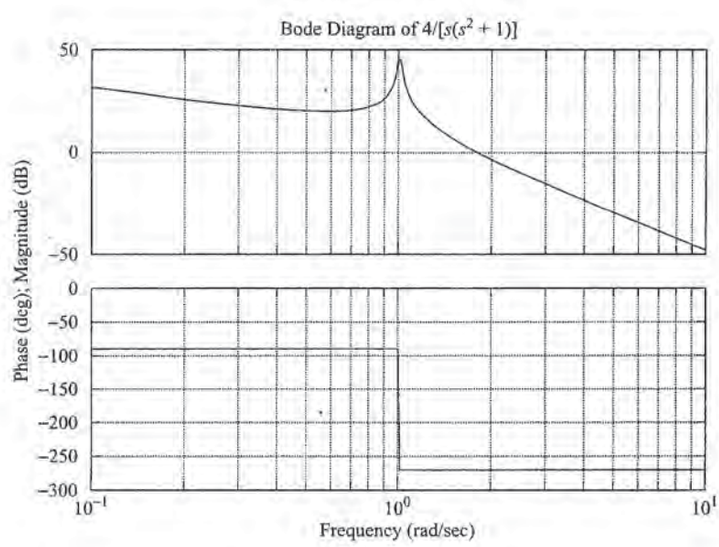


Figure 9–56
Bode diagram of $4/[s(s^2 + 1)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

We need the phase margin of 50° and gain margin of 10 dB or more. From the Bode diagram of Figure 9–56, we notice that the gain crossover frequency is approximately $\omega = 1.8$ rad/sec. Let us assume the gain crossover frequency of the compensated system to be somewhere between $\omega = 1$ and $\omega = 10$ rad/sec.

Let us choose $\hat{G}_c(s)$ to be

$$\hat{G}_c(s) = (as + 1)(bs + 1)$$

and choose $a = 5$. Then, $(as + 1)$ will contribute up to 90° phase lead in the high-frequency region. MATLAB Program 9–18 produces the Bode diagram of

$$\frac{4(5s + 1)}{s(s^2 + 1)}$$

The resulting Bode diagram is shown in Figure 9–57.

MATLAB Program 9–18

```
num = [0 0 20 4];
den = [1 0.00000000001 1 0];
w = logspace(-2,1,101);
bode(num,den,w)
title('Bode Diagram of G(s) = 4(5s+1)/[s(s^2+l)]')
```

Based on the Bode diagram of Figure 9–57, we choose the value of b. The term $(bs + 1)$ needs to give the phase margin of 50°. By simple MATLAB trials, we find $b = 0.25$ to give the phase margin of 50° and gain margin of $+\infty$ dB. Therefore, by choosing $b = 0.25$, we have

$$\hat{G}_c(s) = (5s + 1)(0.25s + 1)$$

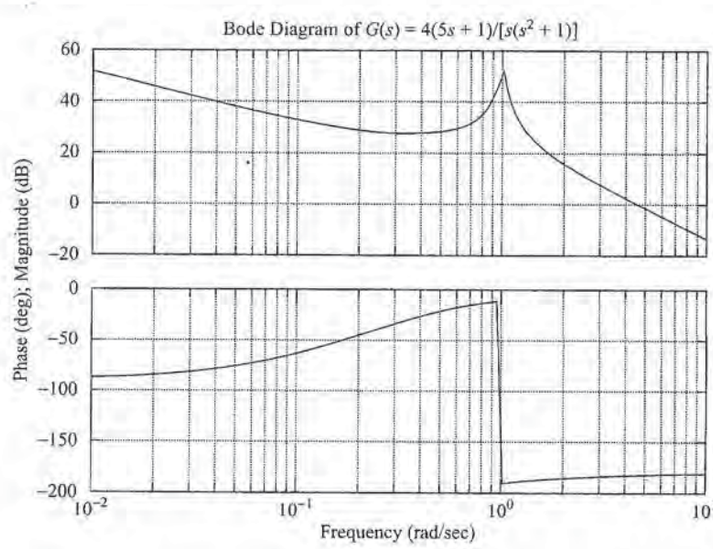


Figure 9–57
Bode diagram of $G(s) = 4(5s + 1)/[s(s^2 + 1)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

and the open-loop transfer function of the designed system becomes

$$\text{Open-loop transfer function} = \frac{4(5s+1)(0.25s+1)}{s}\frac{1}{s^2+1}$$

$$= \frac{5s^2+21s+4}{s^3+s}$$

Note that the controller designed here is a PID controller. MATLAB Program 9–19 produces the Bode diagram of the open-loop transfer function. The resulting Bode diagram is shown in Figure 9–58. From it we see that the static velocity error constant is 4 $\sec^{-1}$, the phase margin is 50°, and the gain margin is $+\infty$ dB. Therefore, the designed system satisfies all the requirements. Thus, the designed system is acceptable. (Note that there exist infinitely many systems that satisfy all the requirements. The present system is just one of them.)

MATLAB Program 9–19

```
num = [0 5 21 4];
den = [1 0 1 0];
w = logspace(-2,2,100);
bode(num,den,w)
title('Bode Diagram of 4(5s+1)(0.25s+1)/[s(s^2+1)]')
```

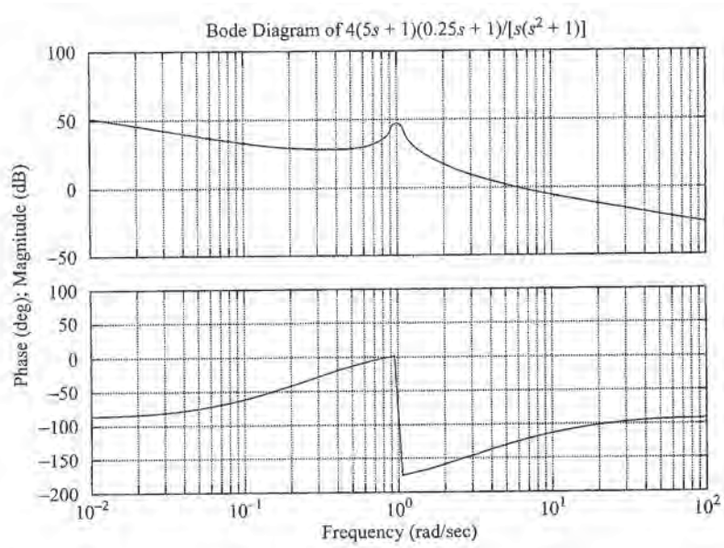


Figure 9–58
Bode diagram of $4(5s+1)(0.25s+1)/[s(s^2+1)]$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Next, we shall obtain the unit-step response and the unit-ramp response of the designed system. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{5s^2 + 21s + 4}{s^3 + 5s^2 + 22s + 4}$$

Note that the closed-loop zeros are located at

$$s = -4, \qquad s = -0.2$$

The closed-loop poles are located at

$$s = -2.4052 + j3.9119$$

$$s = -2.4052 - j3.9119$$

$$s = -0.1897$$

Notice that the complex-conjugate closed-loop poles have the damping ratio of 0.5237. MATLAB Program 9–20 produces the unit-step response and the unit-ramp response. The resulting unit-step response curve is shown in Figure 9–59 and the unit-ramp response curve in Figure 9–60. Notice that the closed-loop pole at $s = -0.1897$ and the zero at $s = -0.2$ produce a long tail of small amplitude in the unit-step response.

MATLAB Program 9–20

```
%***** Unit-step response *****

num = [0 5 21 4];
den = [1 5 22 4];
t = 0:0.01:14;
c = step(num,den,t);
plot(t,c)
grid
title('Unit-Step Response of Compensated System')
xlabel('t (sec)')
ylabel('Output c(t)')

%***** Unit-ramp response *****

num1 = [0 0 5 21 4];
den1 = [1 5 22 4 0];
t = 0:0.02:20;
c = step(num1,den1,t);
plot(t,c,'-',t,t,'--')
title('Unit-Ramp Response of Compensated System')
xlabel('t (sec)')
ylabel('Unit-Ramp Input and Output c(t)')
text(10.8,8,'Compensated System')
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

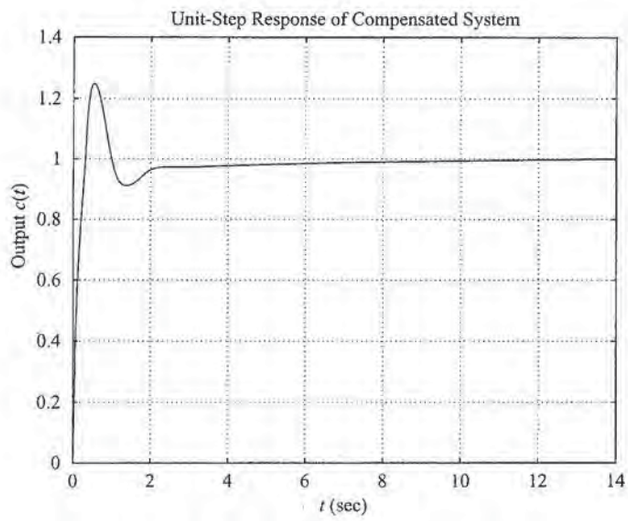


Figure 9–59
Unit-step response curve.

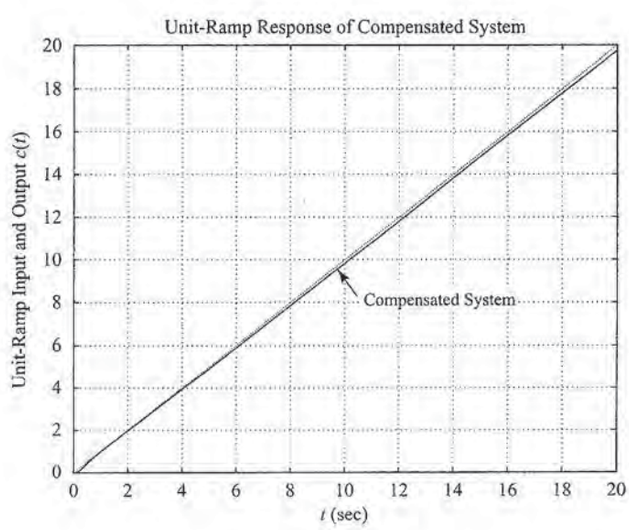


Figure 9–60
Unit-ramp input and the output curve.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

PROBLEMS

B–9–1. Draw Bode diagrams of the lead network and lag network shown in Figures 9–61(a) and (b), respectively. Assume that $R = 50\ k\Omega$ and $C = 10\mu F$.

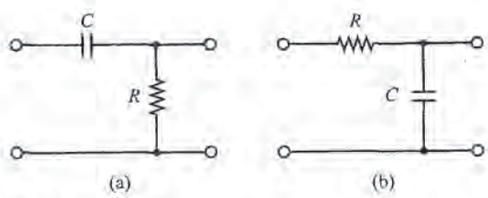


Figure 9–61
(a) Lead network; (b) lag network.

B–9–2. Draw Bode diagrams of the PI controller given by

$$G_c(s) = 5\left(1 + \frac{1}{2s}\right)$$

and the PD controller given by

$$G_c(s) = 5(1 + 0.5s)$$

B–9–3. Consider a PID controller given by

$$G_c(s) = 30.3215\,\frac{(s + 0.65)^2}{s}$$

Draw a Bode diagram of the controller.

B–9–4. Figure 9–62 shows a block diagram of a space vehicle attitude control system. Determine the proportional gain constant K_p and derivative time T_d such that the bandwidth of the closed-loop system is 0.4 to 0.5 rad/sec. (Note that the closed-loop bandwidth is close to the gain crossover frequency.) The system must have an adequate phase margin. Plot both the open-loop and closed-loop frequency response curves on Bode diagrams.

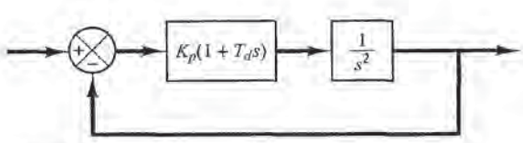


Figure 9–62
Block diagram of space vehicle attitude control system.

B–9–5. Referring to the closed-loop system shown in Figure 9–63, design a lead compensator $G_c(s)$ such that the phase margin is 45°, gain margin is not less than 8 dB, and the static velocity error constant K_v is 4.0 sec^{-1}. Plot unit-step and unit-ramp response curves of the compensated system with MATLAB.

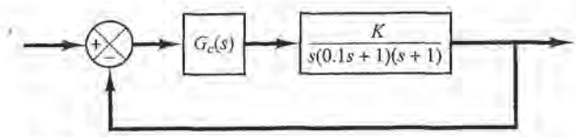


Figure 9–63
Closed-loop system.

B–9–6. Consider the system shown in Figure 9–64. Design a compensator such that the static velocity error constant K_v is 50 sec^{-1}, phase margin is 50°, and gain margin not less than 8 dB. Plot unit-step and unit-ramp response curves of the compensated and uncompensated systems with MATLAB.

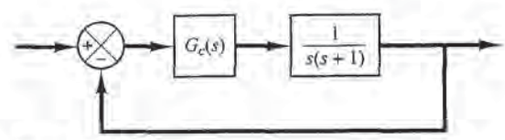


Figure 9–64
Control system.

B–9–7. Consider the system shown in Figure 9–65. Design a compensator such that the static velocity error constant is 4 sec^{-1}, phase margin is 50°, and gain margin is 10 dB or more. Plot unit-step and unit-ramp response curves of the compensated system with MATLAB. Also, draw a Nyquist plot of the compensated system with MATLAB.

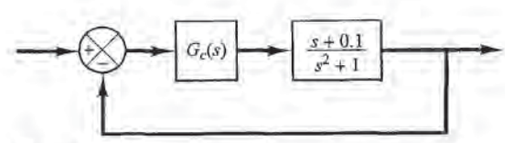


Figure 9–65
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–9–8. Consider the system shown in Figure 9–66. It is desired to design a compensator such that the static velocity error constant is 4 sec^{-1}, phase margin is 50°, and gain margin is 8 dB or more. Plot the unit-step and unit-ramp response curves of the compensated system with MATLAB.

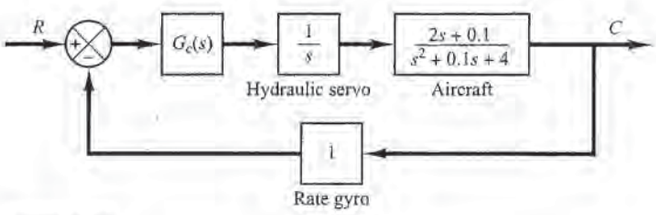


Figure 9–66
Control system.

B–9–9. Consider the system shown in Figure 9–67. Design a lag–lead compensator such that the static velocity error constant K_v is 20 sec^{-1}, phase margin is 60°, and gain margin is not less than 8 dB. Plot the unit-step and unit-ramp response curves of the compensated system with MATLAB.

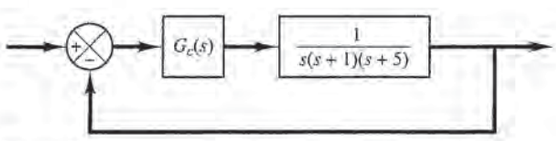


Figure 9–67
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

10

PID Controls and Two-Degrees-of-Freedom Control Systems

10-1 INTRODUCTION

In previous chapters, we occasionally discussed the basic PID control schemes. For example, we presented electronic, hydraulic, and pneumatic PID controllers. We also designed control systems where PID controllers were involved.

It is interesting to note that more than half of the industrial controllers in use today utilize PID or modified PID control schemes.

Because most PID controllers are adjusted on-site, many different types of tuning rules have been proposed in the literature. Using these tuning rules, delicate and fine tuning of PID controllers can be made on-site. Also, automatic tuning methods have been developed and some of the PID controllers may possess on-line automatic tuning capabilities. Modified forms of PID control, such as I-PD control and two-degrees-of-freedom PID control, are currently in use in industry. Many practical methods for bumpless switching (from manual operation to automatic operation) and gain scheduling are commercially available.

The usefulness of PID controls lies in their general applicability to most control systems. In particular, when the mathematical model of the plant is not known and therefore analytical design methods cannot be used, PID controls prove to be most useful. In the field of process control systems, it is well known that the basic and modified PID control schemes have proved their usefulness in providing satisfactory control, although in many given situations they may not provide optimal control.

In this chapter we first present the design of a PID controlled system. We next discuss modified PID controls such as PI-D control and I-PD control. Then we introduce

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

two-degrees-of-freedom control systems, which can satisfy conflicting requirements that single-degree-of-freedom control systems cannot.

In practical cases, there may be one requirement on the response to disturbance input and another requirement on the response to reference input. Often these two requirements conflict with each other and cannot be satisfied in the single-degree-of-freedom case. By increasing the degrees of freedom, we are able to satisfy both. In this chapter we present two-degrees-of-freedom control systems in detail.

By using PID controllers and a two-degrees-of-freedom configuration, we design control systems that satisfy two independent (or sometimes conflicting) requirements (such as the response characteristics to disturbance input and those to reference input). This chapter also presents a computational approach to the design of control systems. The approach can be used to design both single-degree-of-freedom and multiple-degrees-of-freedom control systems.

Finally, we present a very powerful computational approach with MATLAB to search optimal sets of parameter values to satisfy given transient response specifications (such as that the maximum overshoot in the response to the unit-step reference input be less than a specified value and the settling time be less than a specified value). This approach can be directly applied to the design of high-performance control systems.

Outline of the Chapter. Section 10–1 has presented introductory material for the chapter. Section 10–2 deals with tuning methods for the basic PID control, commonly known as Ziegler–Nichols tuning rules. Section 10–3 discusses a computational approach to search optimal sets of parameter values of PID controllers. Section 10–4 treats modified PID control schemes, such as PI-D control and I-PD control. Section 10–5 introduces two-degrees-of-freedom control schemes. Section 10–6 presents the zero-placement approach to improve response characteristics and to design high-performance control systems.

10–2 TUNING RULES FOR PID CONTROLLERS

PID Control of Plants. Figure 10–1 shows a PID control of a plant. If a mathematical model of the plant can be derived, then it is possible to apply various design techniques for determining parameters of the controller that will meet the transient and steady-state specifications of the closed-loop system. However, if the plant is so complicated that its mathematical model cannot be easily obtained, then an analytical approach to the design of a PID controller is not possible. Then we must resort to experimental approaches to the tuning of PID controllers.

The process of selecting the controller parameters to meet given performance specifications is known as controller tuning. Ziegler and Nichols suggested rules for tuning PID controllers (meaning to set values K_p, T_i, and T_d) based on experimental step responses or based on the value of K_p that results in marginal stability when only proportional control action is used. Ziegler–Nichols rules, which are briefly presented in

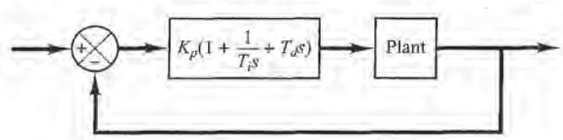


Figure 10–1 PID control of a plant.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

the following, are useful when mathematical models of plants are not known. (These rules can, of course, be applied to the design of systems with known mathematical models.) Such rules suggest a set of values of K_p, T_i, and T_d that will give a stable operation of the system. However, the resulting system may exhibit a large maximum overshoot in the step response, which is unacceptable. In such a case we need series of fine tunings until an acceptable result is obtained. In fact, the Ziegler–Nichols tuning rules give an educated guess for the parameter values and provide a starting point for fine tuning, rather than giving the final settings for K_p, T_i, and T_d in a single shot.

Ziegler–Nichols Rules for Tuning PID Controllers. Ziegler and Nichols proposed rules for determining values of the proportional gain K_p, integral time T_i, and derivative time T_d based on the transient response characteristics of a given plant. Such determination of the parameters of PID controllers or tuning of PID controllers can be made by engineers on-site by experiments on the plant. (Numerous tuning rules for PID controllers have been proposed since the Ziegler–Nichols proposal. They are available in the literature and from the manufacturers of such controllers.)

There are two methods called Ziegler–Nichols tuning rules: the first method and the second method. We shall give a brief presentation of these two methods.

First Method. In the first method, we obtain experimentally the response of the plant to a unit-step input, as shown in Figure 10–2. If the plant involves neither integrator(s) nor dominant complex-conjugate poles, then such a unit-step response curve may look S-shaped, as shown in Figure 10–3. This method applies if the response to a step input exhibits an S-shaped curve. Such step-response curves may be generated experimentally or from a dynamic simulation of the plant.

The S-shaped curve may be characterized by two constants, delay time L and time constant T. The delay time and time constant are determined by drawing a tangent line at the inflection point of the S-shaped curve and determining the intersections of the tangent line with the time axis and line $c(t) = K$, as shown in Figure 10–3. The transfer

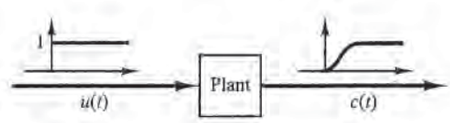


Figure 10–2
Unit-step response of a plant.

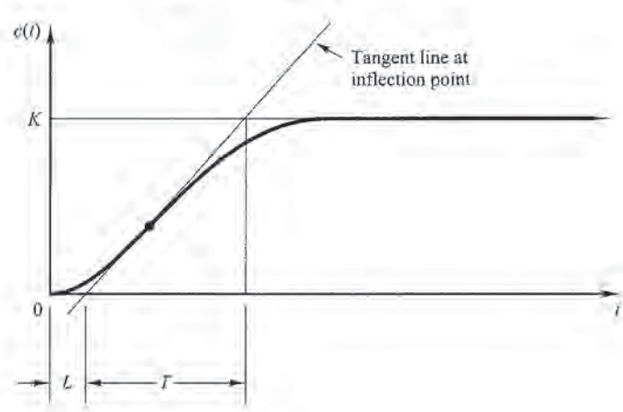


Figure 10–3
S-shaped response curve.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Table 10–1 Ziegler–Nichols Tuning Rule Based on Step Response of Plant (First Method)

Type of Controller	K_p	T_i	T_d
P	$\frac{T}{L}$	∞	0
PI	$0.9\frac{T}{L}$	$\frac{L}{0.3}$	0
PID	$1.2\frac{T}{L}$	$2L$	$0.5L$

function $C(s)/U(s)$ may then be approximated by a first-order system with a transport lag as follows:

$$\frac{C(s)}{U(s)} = \frac{Ke^{-Ls}}{Ts + 1}$$

Ziegler and Nichols suggested to set the values of K_p, T_i, and T_d according to the formula shown in Table 10–1.

Notice that the PID controller tuned by the first method of Ziegler–Nichols rules gives

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 1.2\frac{T}{L}\left(1 + \frac{1}{2Ls} + 0.5Ls\right) \\ &= 0.6T\frac{\left(s + \frac{1}{L}\right)^2}{s} \end{aligned}$$

Thus, the PID controller has a pole at the origin and double zeros at $s = -1/L$.

Second Method. In the second method, we first set $T_i = \infty$ and $T_d = 0$. Using the proportional control action only (see Figure 10–4), increase K_p from 0 to a critical value K_{cr} at which the output first exhibits sustained oscillations. (If the output does not exhibit sustained oscillations for whatever value K_p may take, then this method does not apply.) Thus, the critical gain K_{cr} and the corresponding period P_{cr} are experimentally

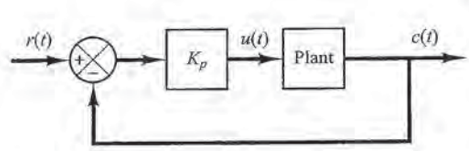


Figure 10–4 Closed-loop system with a proportional controller.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

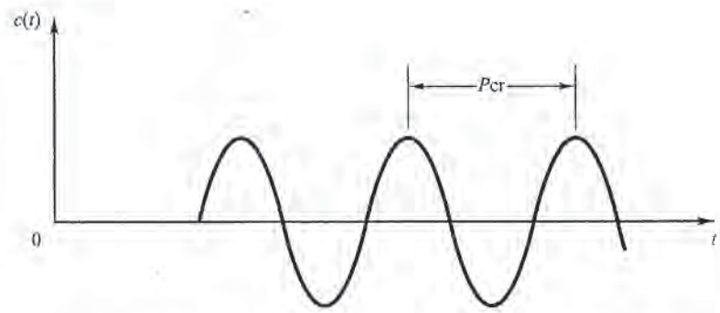


Figure 10–5
Sustained oscillation with period P_{cr}.

determined (see Figure 10–5). Ziegler and Nichols suggested that we set the values of the parameters K_p, T_i, and T_d according to the formula shown in Table 10–2.

Table 10–2 Ziegler–Nichols Tuning Rule Based on Critical Gain K_{cr} and Critical Period P_{cr} (Second Method)

Type of Controller	K_p	T_i	T_d
P	$0.5K_{cr}$	∞	0
PI	$0.45K_{cr}$	$\frac{1}{1.2}P_{cr}$	0
PID	$0.6K_{cr}$	$0.5P_{cr}$	$0.125P_{cr}$

Notice that the PID controller tuned by the second method of Ziegler–Nichols rules gives

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 0.6K_{cr}\left(1 + \frac{1}{0.5P_{cr}s} + 0.125P_{cr}s\right) \\ &= 0.075K_{cr}P_{cr}\frac{\left(s + \frac{4}{P_{cr}}\right)^2}{s} \end{aligned}$$

Thus, the PID controller has a pole at the origin and double zeros at $s = -4/P_{cr}$.

Note that if the system has a known mathematical model (such as the transfer function), then we can use the root-locus method to find the critical gain K_{cr} and the frequency of the sustained oscillations ω_{cr}, where $2\pi/\omega_{cr} = P_{cr}$. These values can be found from the crossing points of the root-locus branches with the $j\omega$ axis. (Obviously, if the root-locus branches do not cross the $j\omega$ axis, this method does not apply.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Comments. Ziegler–Nichols tuning rules (and other tuning rules presented in the literature) have been widely used to tune PID controllers in process control systems where the plant dynamics are not precisely known. Over many years, such tuning rules proved to be very useful. Ziegler–Nichols tuning rules can, of course, be applied to plants whose dynamics are known. (If the plant dynamics are known, many analytical and graphical approaches to the design of PID controllers are available, in addition to Ziegler–Nichols tuning rules.)

EXAMPLE 10–1 Consider the control system shown in Figure 10–6 in which a PID controller is used to control the system. The PID controller has the transfer function

$$G_c(s) = K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$$

Although many analytical methods are available for the design of a PID controller for the present system, let us apply a Ziegler–Nichols tuning rule for the determination of the values of parameters K_p, T_i, and T_d. Then obtain a unit-step response curve and check to see if the designed system exhibits approximately 25% maximum overshoot. If the maximum overshoot is excessive (40% or more), make a fine tuning and reduce the amount of the maximum overshoot to approximately 25% or less.

Since the plant has an integrator, we use the second method of Ziegler–Nichols tuning rules. By setting $T_i = \infty$ and $T_d = 0$, we obtain the closed-loop transfer function as follows:

$$\frac{C(s)}{R(s)} = \frac{K_p}{s(s+1)(s+5) + K_p}$$

The value of K_p that makes the system marginally stable so that sustained oscillation occurs can be obtained by use of Routh's stability criterion. Since the characteristic equation for the closed-loop system is

$$s^3 + 6s^2 + 5s + K_p = 0$$

the Routh array becomes as follows:

$$\begin{array}{ccc} s^3 & 1 & 5 \\ s^2 & 6 & K_p \\ s^1 & \dfrac{30 - K_p}{6} & \\ s^0 & K_p & \end{array}$$

Examining the coefficients of the first column of the Routh table, we find that sustained oscillation will occur if $K_p = 30$. Thus, the critical gain K_{cr} is

$$K_{cr} = 30$$

With gain K_p set equal to K_{cr} (= 30), the characteristic equation becomes

$$s^3 + 6s^2 + 5s + 30 = 0$$

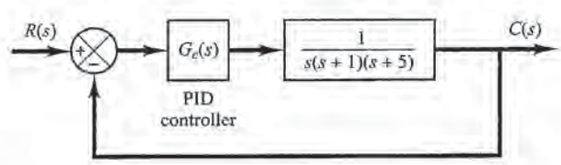


Figure 10–6 PID-controlled system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

To find the frequency of the sustained oscillation, we substitute $s = j\omega$ into this characteristic equation as follows:

$$(j\omega)^3 + 6(j\omega)^2 + 5(j\omega) + 30 = 0$$

or

$$6(5 - \omega^2) + j\omega(5 - \omega^2) = 0$$

from which we find the frequency of the sustained oscillation to be $\omega^2 = 5$ or $\omega = \sqrt{5}$. Hence, the period of sustained oscillation is

$$P_{cr} = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{5}} = 2.8099$$

Referring to Table 10–2, we determine K_p, T_i, and T_d as follows:

$$K_p = 0.6K_{cr} = 18$$

$$T_i = 0.5P_{cr} = 1.405$$

$$T_d = 0.125P_{cr} = 0.35124$$

The transfer function of the PID controller is thus

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 18\left(1 + \frac{1}{1.405s} + 0.35124s\right) \\ &= \frac{6.3223(s + 1.4235)^2}{s} \end{aligned}$$

The PID controller has a pole at the origin and double zero at $s = -1.4235$. A block diagram of the control system with the designed PID controller is shown in Figure 10–7.

Next, let us examine the unit-step response of the system. The closed-loop transfer function $C(s)/R(s)$ is given by

$$\frac{C(s)}{R(s)} = \frac{6.3223s^2 + 18s + 12.811}{s^4 + 6s^3 + 11.3223s^2 + 18s + 12.811}$$

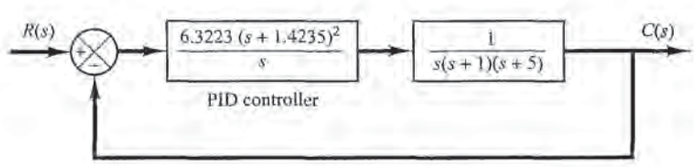


Figure 10–7
Block diagram of the system with PID controller designed by use of the Ziegler–Nichols tuning rule (second method).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The unit-step response of this system can be obtained easily with MATLAB. See MATLAB Program 10–1. The resulting unit-step response curve is shown in Figure 10–8. The maximum overshoot in the unit-step response is approximately 62%. The amount of maximum overshoot is excessive. It can be reduced by fine tuning the controller parameters. Such fine tuning can be made on the computer. We find that by keeping $K_p = 18$ and by moving the double zero of the PID controller to $s = -0.65$, that is, using the PID controller

$$G_c(s) = 18\left(1 + \frac{1}{3.077s} + 0.7692s\right) = 13.846\frac{(s + 0.65)^2}{s} \tag{10–1}$$

the maximum overshoot in the unit-step response can be reduced to approximately 18% (see Figure 10–9). If the proportional gain K_p is increased to 39.42, without changing the location of the double zero ($s = -0.65$), that is, using the PID controller

$$G_c(s) = 39.42\left(1 + \frac{1}{3.077s} + 0.7692s\right) = 30.322\frac{(s + 0.65)^2}{s} \tag{10–2}$$

then the speed of response is increased, but the maximum overshoot is also increased to approximately 28%, as shown in Figure 10–10. Since the maximum overshoot in this case is fairly close to 25% and the response is faster than the system with $G_c(s)$ given by Equation (10–1), we may

MATLAB Program 10–1

```
% --------- Unit-step response ---------
num = [0 0 6.3223 18 12.811];
den = [1 6 11.3223 18 12.811];
step(num,den)
grid
title('Unit-Step Response')
```

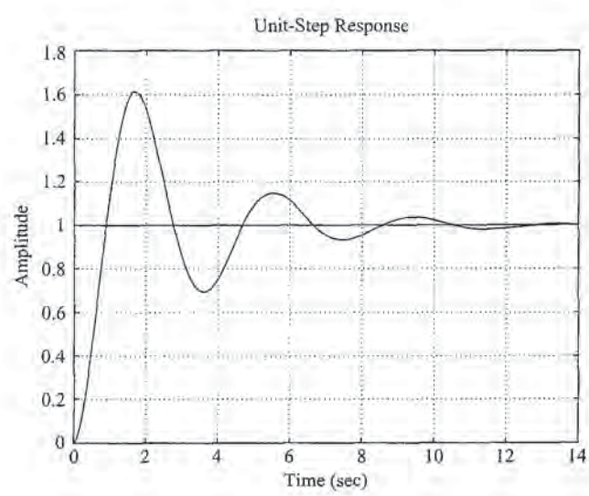


Figure 10–8
Unit-step response curve of PID-controlled system designed by use of the Ziegler–Nichols tuning rule (second method).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

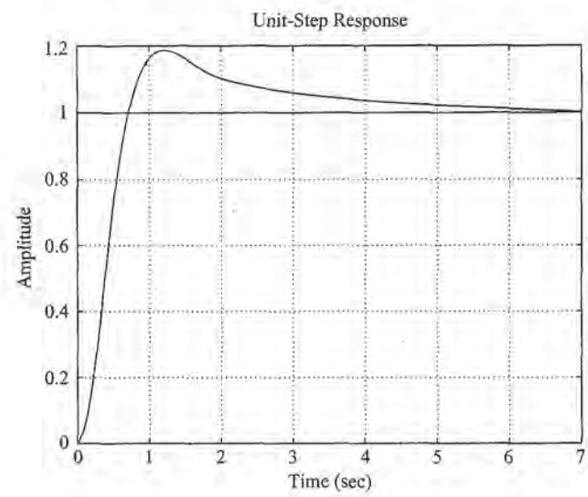


Figure 10–9
Unit-step response of the system shown in Figure 10–6 with PID controller having parameters $K_p = 18$, $T_i = 3.077$, and $T_d = 0.7692$.

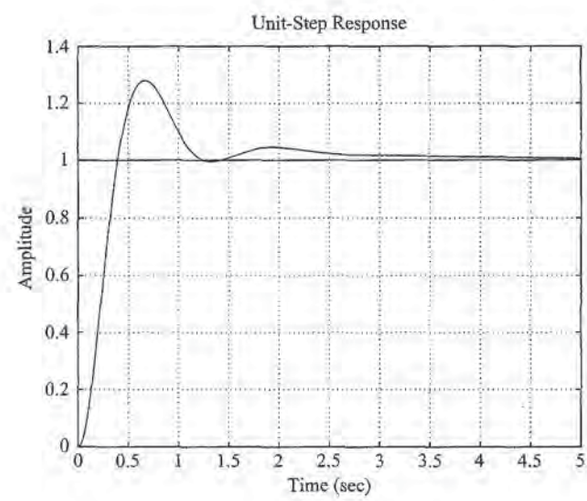


Figure 10–10
Unit-step response of the system shown in Figure 10–6 with PID controller having parameters $K_p = 39.42$, $T_i = 3.077$, and $T_d = 0.7692$.

consider $G_c(s)$ as given by Equation (10–2) as acceptable. Then the tuned values of K_p, T_i, and T_d become

$$K_p = 39.42, \qquad T_i = 3.077, \qquad T_d = 0.7692$$

It is interesting to observe that these values respectively are approximately twice the values suggested by the second method of the Ziegler–Nichols tuning rule. The important thing to note here is that the Ziegler–Nichols tuning rule has provided a starting point for fine tuning.

It is instructive to note that, for the case where the double zero is located at $s = -1.4235$, increasing the value of K_p increases the speed of response, but as far as the percentage maximum overshoot is concerned, varying gain K_p has very little effect. The reason for this may be seen

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

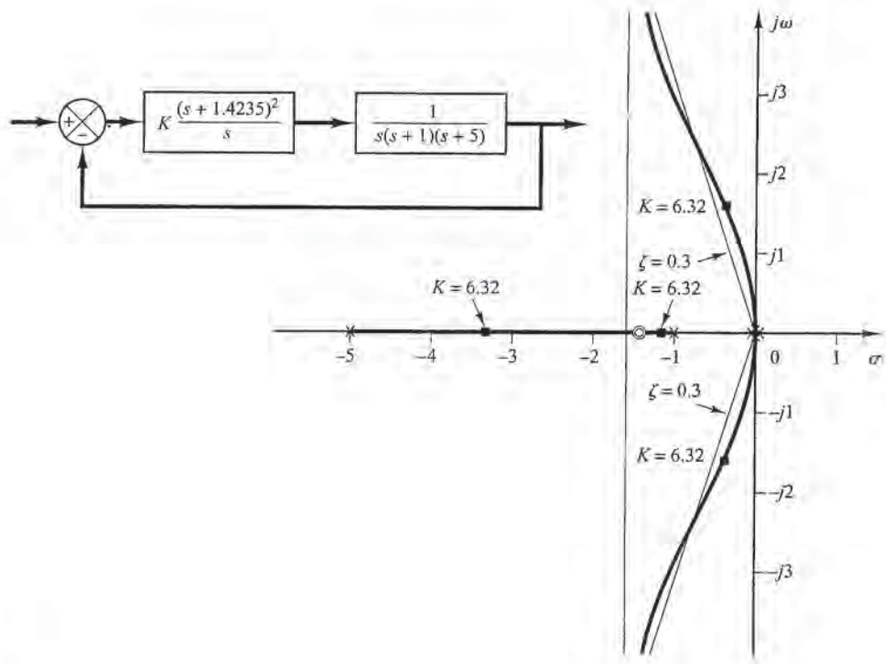


Figure 10–11
Root-locus diagram of system when PID controller has double zero at $s = -1.4235$.

from the root-locus analysis. Figure 10–11 shows the root-locus diagram for the system designed by use of the second method of Ziegler–Nichols tuning rules. Since the dominant branches of root loci are along the $\zeta = 0.3$ lines for a considerable range of K, varying the value of K (from 6 to 30) will not change the damping ratio of the dominant closed-loop poles very much. However, varying the location of the double zero has a significant effect on the maximum overshoot, because the damping ratio of the dominant closed-loop poles can be changed significantly. This can also be seen from the root-locus analysis. Figure 10–12 shows the root-locus diagram for the system where the PID controller has the double zero at $s = -0.65$. Notice the change of the root-locus configuration. This change in the configuration makes it possible to change the damping ratio of the dominant closed-loop poles.

In Figure 10–12, notice that, in the case where the system has gain $K = 30.322$, the closed-loop poles at $s = -2.35 \pm j4.82$ act as dominant poles. Two additional closed-loop poles are very near the double zero at $s = -0.65$, with the result that these closed-loop poles and the double zero almost cancel each other. The dominant pair of closed-loop poles indeed determines the nature of the response. On the other hand, when the system has $K = 13.846$, the closed-loop poles at $s = -2.35 \pm j2.62$ are not quite dominant because the two other closed-loop poles near the double zero at $s = -0.65$ have considerable effect on the response. The maximum overshoot in the step response in this case (18%) is much larger than the case where the system is of second order and having only dominant closed-loop poles. (In the latter case the maximum overshoot in the step response would be approximately 6%.)

It is possible to make a third, a fourth, and still further trials to obtain a better response. But this will take a lot of computations and time. If more trials are desired, it is desirable to use the computational approach presented in Section 10–3. Problem **A–10–11** solves this problem with the computational approach with MATLAB. It finds sets of parameter values that will yield the

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

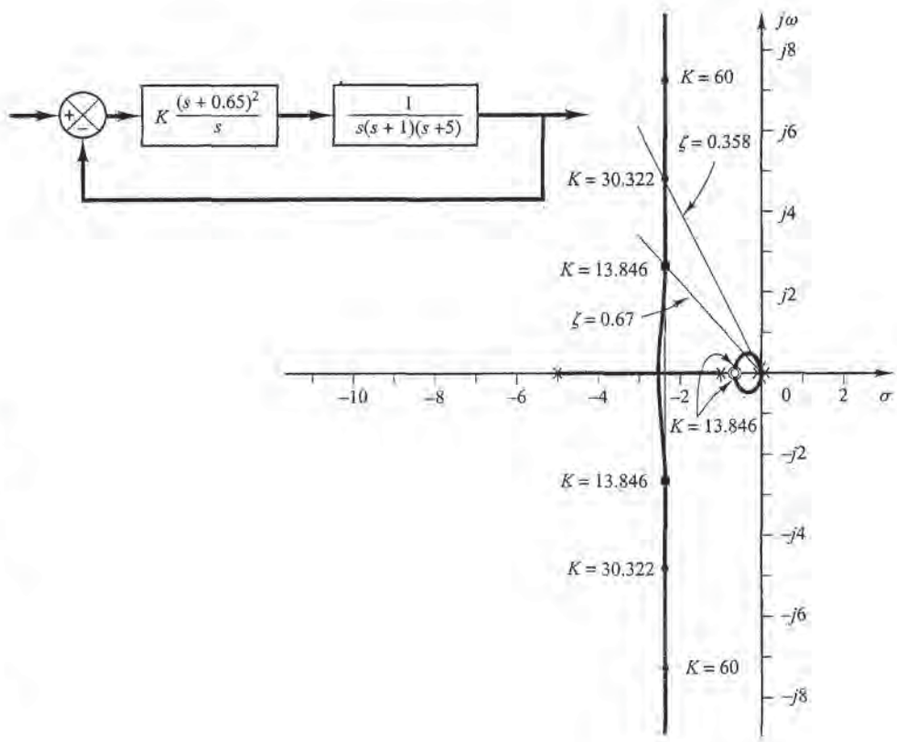


Figure 10–12
Root-locus diagram of system when PID controller has double zero at $s = -0.65$. $K = 13.846$ corresponds to $G_c(s)$ given by Equation (10–1) and $K = 30.322$ corresponds to $G_c(s)$ given by Equation (10–2).

maximum overshoot of 10% or less and the settling time of 3 sec or less. A solution to the present problem obtained in Problem **A–10–11** is that for the PID controller defined by

$$G_c(s) = K\frac{(s+a)^2}{s}$$

the values of K and a are

$$K = 29, \qquad a = 0.25$$

with the maximum overshoot equal to 9.52% and settling time equal to 1.78 sec. Another possible solution obtained there is that

$$K = 27, \qquad a = 0.2$$

with the 5.5% maximum overshoot and 2.89 sec of settling time. See Problem **A–10–11** for details.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

10–3 COMPUTATIONAL APPROACH TO OBTAIN OPTIMAL SETS OF PARAMETER VALUES

In this section we shall explore how to obtain an optimal set (or optimal sets) of parameter values to satisfy the transient response specification by use of MATLAB. We shall present two examples to illustrate the approach.

EXAMPLE 10–2 Consider the PID-controlled system shown in Figure 10–13. The PID controller is given by

$$G_c(s) = K\frac{(s+a)^2}{s}$$

It is desired to find a combination of K and a such that the closed-loop system is underdamped and the maximum overshoot in the unit-step response is less than 10%, but more than 5%, to avoid an overdamped or a close-to-overdamped response. (Other conditions can be included, such as that the settling time be less than a specified value and the rise time be less than a certain specified value.)

To solve this problem with MATLAB, it is necessary to write a program such that in the unit-step response it will find a combination of K and a which will satisfy the criterion that the maximum overshoot is less than 10%, but more than 5%. Assume that the region to search for K and a is bounded by

$$2 \le K \le 5 \qquad \text{and} \qquad 0.5 \le a \le 1.5$$

To avoid an overly large amount of computation in this problem, let us choose the step size to be reasonable—say, 0.2 for both K and a.

There may be more than one possible combination of K and a that will satisfy the given condition. In this problem, however, it is desired to obtain the first solution. (To find all possible sets of solutions in the search region, see Example 10–3.) We proceed by writing a MATLAB program such that the nested loops in the program begin with the highest values of K and a and step toward the lowest. It is also desired that the MATLAB program either automatically plot the unit-step response curve of the system using the first-found solution of K and a or else enable us to plot the unit-step response of the system with the first-found solution of K and a by use of the command 'plot'. (If no values of K and a are found, the MATLAB program should not plot a unit-step response curve.)

To solve this problem, one may write many different MATLAB programs. We shall present two possible programs to solve this problem, MATLAB Program 10–2 and MATLAB Program 10–3.

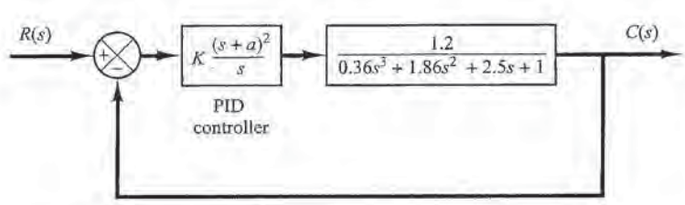


Figure 10–13 PID-controlled system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

In MATLAB Program 10–2, notice that we use two 'for' loops. We start the program with the outer loop to vary the 'K' values. Then we vary the 'a' values in the inner loop. We define the numerator and denominator of the closed-loop system and obtain the unit-step response of the system. We define the maximum overshoot m in the response (m = maximum value in the response y). If the condition "$m < 1.10$ and $m > 1.05$" is satisfied, we break the inner loop and terminate the process of varying the 'a' values. Then we break the outer loop to finish up the searching process. When MATLAB finishes computation for the first-found set of 'K' and 'a' that satisfies the condition, we enter the 'plot' command to get a plot of the response (see Figure 10–14). Note that, depending on the system and the ranges of search for K and a and the step size, it may take from several seconds to a few minutes for MATLAB to compute the desired set of the values. In the program, the statement

```
sol = [K; a; m]
```

displays the chosen set of values of K, a, and m.

MATLAB Program 10–2

```
t = 0:0.01:8;
for K = 5:-0.2:2; % Starts the outer loop to vary the K values
   for a = 1.5:-0.2:0.5; % Starts the inner loop to vary the a values
      num = [0  0  1.2*K  2.4*K*a  1.2*K*a^2];
      den = [0.36  1.86  2.5+1.2*K  1+2.4*K*a  1.2*K*a^2];
         y = step(num,den,t);
         m = max(y);
         if m < 1.1 & m > 1.05
         break; % Breaks the inner loop
      end
      end
   if m < 1.1 & m > 1.05
   break; % Breaks the outer loop
  end
  end
plot(t,y)
grid
title('Unit-Step Response')
xlabel('t Sec')
ylabel('Output')
KK = num2str(K); % String value of K to be printed on plot
aa = num2str(a); % String value of a to be printed on plot
text(4.25,0.54,'K = '), text(4.75,0.54,KK)
text(4.25,0.46,'a = '), text(4.75,0.46,aa)
sol = [K;a;m]

Sol =

4.2000
0.7000
1.0962
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

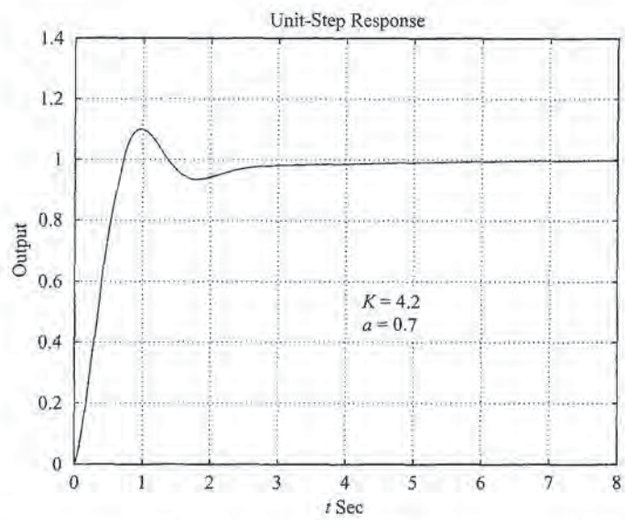


Figure 10–14
Unit-step response curve obtained by use of MATLAB Program 10–2.

MATLAB Program 10–3 is basically the same as MATLAB Program 10–2. We use two loops (the outer loop for varying 'K' values and the inner loop for varying 'a' values). In this program we define the PID controller by num1 and den1 and obtain the transfer function expression tf1 for the controller. Then we define the plant by num2 and den2 and obtain the transfer function expression tf2 for the plant. The open-loop transfer function for the system is obtained by multiplying tf1 and tf2. Using the 'feedback' command, we get the closed-loop transfer function 'sys'. Then we get the unit-step response of the closed-loop system. If the maximum overshoot m is less than 1.10, but more than 1.05, MATLAB will plot the unit-step response curve of the first-found solution of K and a. (If no sets of values of K and a satisfy the condition, this program will not plot a response curve.) This MATLAB program automatically displays the first-found solution of K, a, and m. Figure 10–15 is the unit-step response curve obtained by the program.

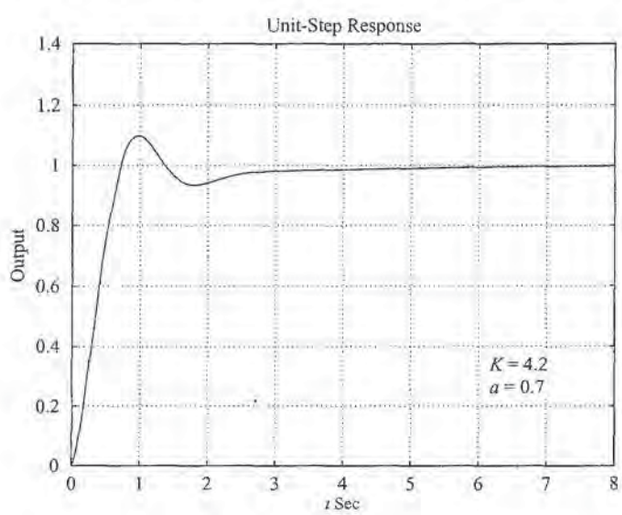


Figure 10–15
Unit-step response curve obtained by use of MATLAB Program 10–3.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

MATLAB Program 10–3

```
t = 0:0.01:8;
for K = 5:-0.2:2; % Starts the outer loop to vary the K values
   for a = 1.5:-0.2:0.5; % Starts the inner loop to vary the a values
         num1 = K*[1  2*a  a^2];
         den1 = [0  1  0];
         tf1 = tf(num1, den1);
         num2 = [0  0  0  1.2];
         den2 = [0.36  1.86  2.5  1];
         tf2 = tf(num2,den2);
         tf3 = tf1*tf2;
         sys = feedback(tf3,1);
         y = step(sys,t);
         m = max(y);
         if m < 1.1 & m > 1.05;
         plot(t,y);
         grid;
         title('Unit-Step Response')
         xlabel('t Sec')
         ylabel('Output')
         sol = [K;a;m]
         break; % Breaks the inner loop
      end
      end
   if m < 1.1 & m > 1.05;
   break; % Breaks the outer loop
end
end
sol =
      4.2000
      0.7000
      1.0962
text(6.2, 0.35,'K = '), text(6.65, 0.35,num2str(K))
text(6.2, 0.25,'a = '), text(6.65, 0.25,num2str(a))
```

EXAMPLE 10–3 In Example 10–2 we wrote MATLAB programs to find the first set of parameters to satisfy the given specifications. There may be more than one set of parameters that satisfy the specifications. In this example, we shall obtain all sets of parameters that satisfy the given specifications.

Consider the same system as in Example 10–2, except that the problem here is to find all sets of K and a that will satisfy the given specification that the maximum overshoot in the unit-step response be less than 10%. (This means that overdamped systems are included.) Assume the search region to be

$$2 \le K \le 3, \qquad 0.5 \le a \le 1.5$$

In the actual design process, the step size should be sufficiently small. In this example problem, however, we choose a fairly large step size to make the total number of search points reasonable. Thus, we choose the step size for both K and a to be 0.2.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

To solve this problem it is possible to write many different MATLAB programs. We present here one such program, MATLAB Program 10–4.

In this program the statement

```
solution(k,:) = [K(i)  a(j)  m]
```

will produce a table of K, a, m values. (In the present system there are 15 sets of K and a that will exhibit $m < 1.10$—that is, the maximum overshoot is less than 10%.)

To sort out the solution sets in the order of the magnitude of the maximum overshoot (starting from the smallest value of m and ending at the largest value of m in the table), we use the command

```
sortsolution = sortrows(solution,3)
```

To plot the unit-step response curve of the last set of the K and a values in the sorted table, we enter the commands

```
K = sortsolution (k,1)
a = sortsolution (k,2)
```

and use the step command. (The resulting unit-step response curve is shown in Figure 10–16.) To plot the unit-step response curve with the smallest overshoot that is greater than 0% found in the sorted table, enter the commands

```
K = sortsolution (11,1)
a = sortsolution (11,2)
```

and use the step command. (The resulting unit-step response curve is shown in Figure 10–17.) To plot the unit-step response curve of the system with any set shown in the sorted table, we specify the K and a values by entering an appropriate sortsolution command.

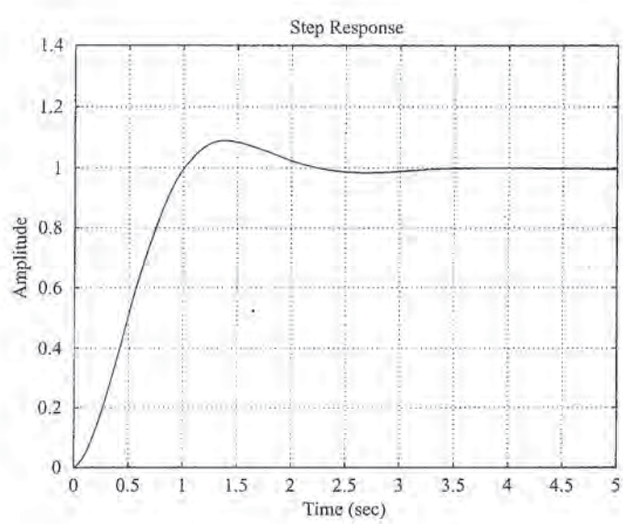


Figure 10–16
Unit-step response of the system with $K = 2.4$ and $a = 0.9$. (The maximum overshoot is 9.23%.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

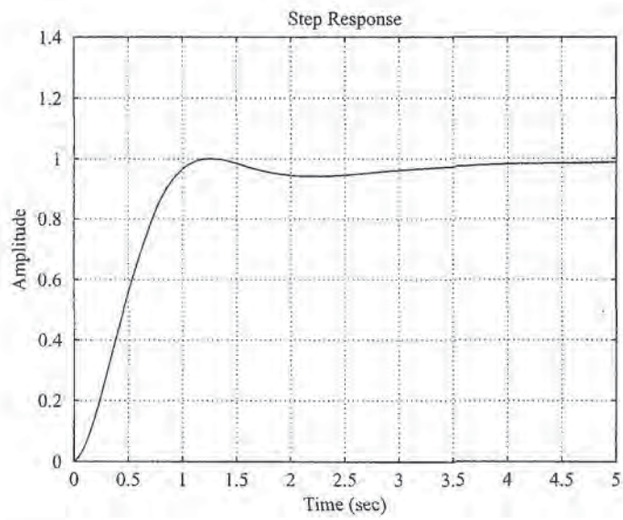


Figure 10–17
Unit-step response of the system with $K = 2.8$ and $a = 0.7$. (The maximum overshoot is 0.24%.)

Note that for a specification that the maximum overshoot be between 10% and 5%, there would be three sets of solutions:

$$K = 2.0000, \quad a = 0.9000, \quad m = 1.0614$$
$$K = 2.2000, \quad a = 0.9000, \quad m = 1.0772$$
$$K = 2.4000, \quad a = 0.9000, \quad m = 1.0923$$

Unit-step response curves for these three cases are shown in Figure 10–18. Notice that the system with a larger gain K has a smaller rise time and larger maximum overshoot. Which one of these three systems is best depends on the system's objective.

MATLAB Program 10–4

```
%'K' and 'a' values to test

K = [2.0  2.2  2.4  2.6  2.8  3.0];
a = [0.5  0.7  0.9  1.1  1.3  1.5];

% Evaluate closed-loop unit-step response at each 'K' and 'a' combination
% that will yield the maximum overshoot less than 10%

t = 0:0.01:5;
g = tf([0  0  0  1.2],[0.36  1.86  2.5  1]);
k = 0;
for i = 1:6;
   for j= 1:6;
      gc = tf(K(i)*[1  2*a(j)  a(j)^2], [0  1  0]); % controller
         G = gc*g/(1 + gc*g); % closed-loop transfer function
         y = step(G,t);
```

(*continues on next page*)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

```
        m = max(y);
        if m < 1.10
        k = k+1;
        solution(k,:) = [K(i)  a(j)  m];
        end
      end
    end
solution  % Print solution table

solution =

   2.0000  0.5000  0.9002
   2.0000  0.7000  0.9807
   2.0000  0.9000  1.0614
   2.2000  0.5000  0.9114
   2.2000  0.7000  0.9837
   2.2000  0.9000  1.0772
   2.4000  0.5000  0.9207
   2.4000  0.7000  0.9859
   2.4000  0.9000  1.0923
   2.6000  0.5000  0.9283
   2.6000  0.7000  0.9877
   2.8000  0.5000  0.9348
   2.8000  0.7000  1.0024
   3.0000  0.5000  0.9402
   3.0000  0.7000  1.0177
sortsolution = sortrows(solution,3)  % Print solution table sorted by
                                     % column 3

sortsolution =

   2.0000  0.5000  0.9002
   2.2000  0.5000  0.9114
   2.4000  0.5000  0.9207
   2.6000  0.5000  0.9283
   2.8000  0.5000  0.9348
   3.0000  0.5000  0.9402
   2.0000  0.7000  0.9807
   2.2000  0.7000  0.9837
   2.4000  0.7000  0.9859
   2.6000  0.7000  0.9877
   2.8000  0.7000  1.0024
   3.0000  0.7000  1.0177
   2.0000  0.9000  1.0614
   2.2000  0.9000  1.0772
   2.4000  0.9000  1.0923
```

(*continues on next page*)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

```
% Plot the response with the largest overshoot that is less than 10%

K = sortsolution(k,1)

K =

   2.4000

a = sortsolution(k,2)

a =

   0.9000

gc = tf(K*[1  2*a  a^2], [0  1  0]);
G = gc*g/(1 + gc*g);
step(G,t)
grid  % See Figure 10-16

% If you wish to plot the response with the smallest overshoot that is
% greater than 0%, then enter the following values of 'K' and 'a'

K = sortsolution(11,1)

K =

   2.8000

a = sortsolution(11,2)

a =

   0.7000

gc = tf(K*[1  2*a  a^2], [0  1  0]);
G = gc*g/(1 + gc*g);
step(G,t)
grid  % See Figure 10-17
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

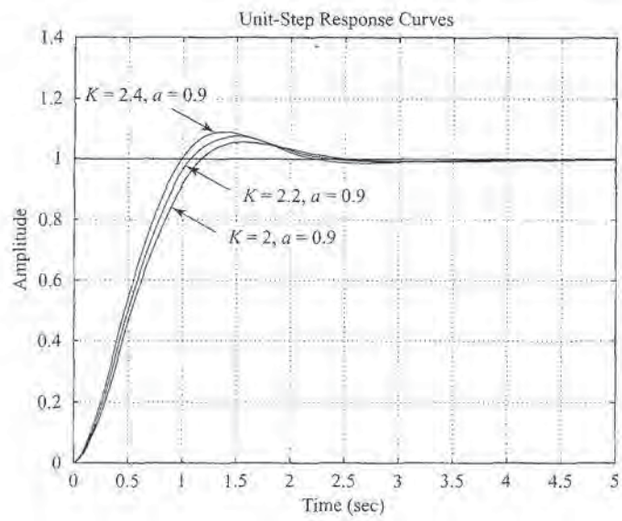


Figure 10–18
Unit-step response curves of system with $K = 2$, $a = 0.9$; $K = 2.2$, $a = 0.9$; and $K = 2.4$, $a = 0.9$.

10–4 MODIFICATIONS OF PID CONTROL SCHEMES

Consider the basic PID control system shown in Figure 10–19(a), where the system is subjected to disturbances and noises. Figure 10–19(b) is a modified block diagram of the same system. In the basic PID control system such as the one shown in Figure 10–19(b), if the reference input is a step function, then, because of the presence of the derivative term in the control action, the manipulated variable $u(t)$ will involve an impulse function (delta function). In an actual PID controller, instead of the pure derivative term $T_d s$, we employ

$$\frac{T_d s}{1 + \gamma T_d s}$$

where the value of γ is somewhere around 0.1. Therefore, when the reference input is a step function, the manipulated variable $u(t)$ will not involve an impulse function, but will involve a sharp pulse function. Such a phenomenon is called *set-point kick*.

PI-D Control. To avoid the set-point kick phenomenon, we may wish to operate the derivative action only in the feedback path so that differentiation occurs only on the feedback signal and not on the reference signal. The control scheme arranged in this way is called the PI-D control. Figure 10–20 shows a PI-D-controlled system.

From Figure 10–20, it can be seen that the manipulated signal $U(s)$ is given by

$$U(s) = K_p\left(1 + \frac{1}{T_i s}\right)R(s) - K_p\left(1 + \frac{1}{T_i s} + T_d s\right)B(s)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

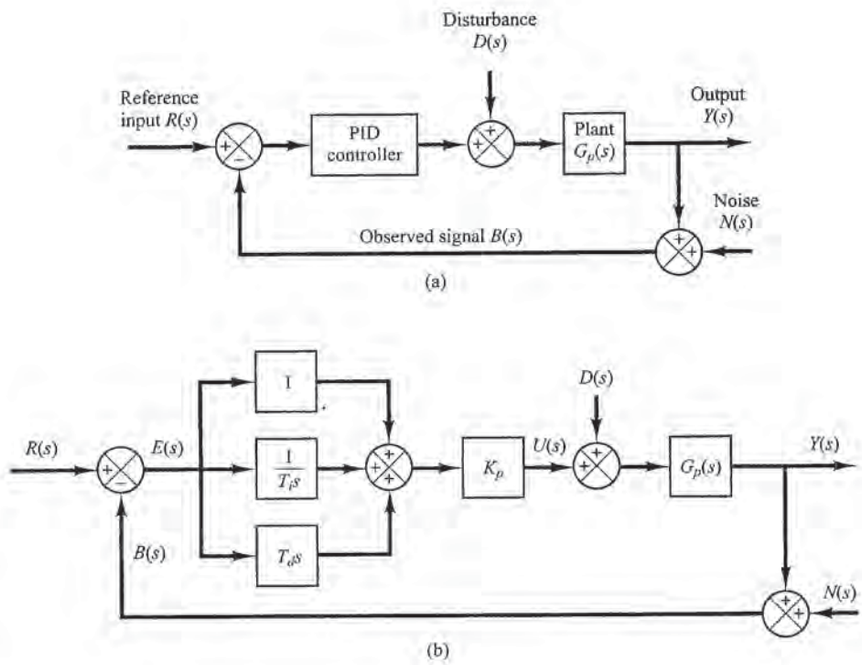


Figure 10–19
(a) PID-controlled system;
(b) equivalent block diagram.

Notice that in the absence of the disturbances and noises, the closed-loop transfer function of the basic PID control system [shown in Figure 10–19(b)] and the PI-D control system (shown in Figure 10–20) are given, respectively, by

$$\frac{Y(s)}{R(s)} = \left(1 + \frac{1}{T_i s} + T_d s\right) \frac{K_p G_p(s)}{1 + \left(1 + \frac{1}{T_i s} + T_d s\right) K_p G_p(s)}$$

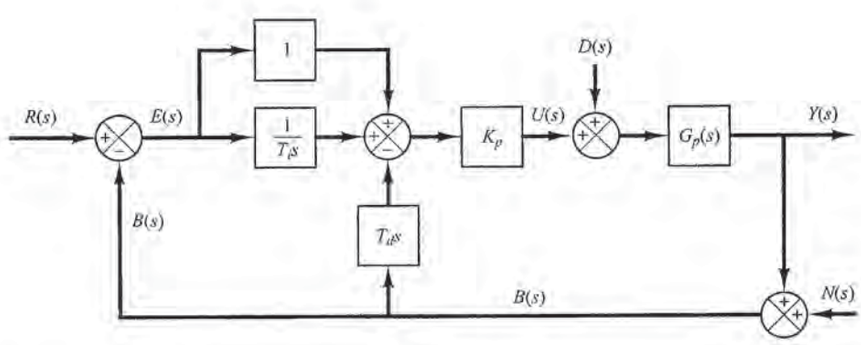


Figure 10–20
PI-D-controlled system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

and

$$\frac{Y(s)}{R(s)} = \left(1 + \frac{1}{T_i s}\right) \frac{K_p G_p(s)}{1 + \left(1 + \frac{1}{T_i s} + T_d s\right) K_p G_p(s)}$$

It is important to point out that in the absence of the reference input and noises, the closed-loop transfer function between the disturbance $D(s)$ and the output $Y(s)$ in either case is the same and is given by

$$\frac{Y(s)}{D(s)} = \frac{G_p(s)}{1 + K_p G_p(s)\left(1 + \frac{1}{T_i s} + T_d s\right)}$$

I-PD Control. Consider again the case where the reference input is a step function. Both PID control and PI-D control involve a step function in the manipulated signal. Such a step change in the manipulated signal may not be desirable in many occasions. Therefore, it may be advantageous to move the proportional action and derivative action to the feedback path so that these actions affect the feedback signal only. Figure 10–21 shows such a control scheme. It is called the I-PD control. The manipulated signal is given by

$$U(s) = K_p \frac{1}{T_i s} R(s) - K_p\left(1 + \frac{1}{T_i s} + T_d s\right) B(s)$$

Notice that the reference input $R(s)$ appears only in the integral control part. Thus, in I-PD control, it is imperative to have the integral control action for proper operation of the control system.

The closed-loop transfer function $Y(s)/R(s)$ in the absence of the disturbance input and noise input is given by

$$\frac{Y(s)}{R(s)} = \left(\frac{1}{T_i s}\right) \frac{K_p G_p(s)}{1 + K_p G_p(s)\left(1 + \frac{1}{T_i s} + T_d s\right)}$$

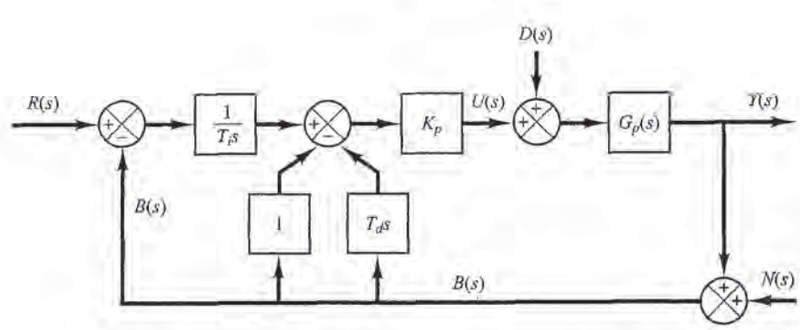


Figure 10–21
I-PD-controlled system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

It is noted that in the absence of the reference input and noise signals, the closed-loop transfer function between the disturbance input and the output is given by

$$\frac{Y(s)}{D(s)} = \frac{G_p(s)}{1 + K_pG_p(s)\left(1 + \dfrac{1}{T_is} + T_ds\right)}$$

This expression is the same as that for PID control or PI-D control.

Two-Degrees-of-Freedom PID Control. We have shown that PI-D control is obtained by moving the derivative control action to the feedback path, and I-PD control is obtained by moving the proportional control and derivative control actions to the feedback path. Instead of moving the entire derivative control action or proportional control action to the feedback path, it is possible to move only portions of these control actions to the feedback path, retaining the remaining portions in the feedforward path. In the literature, PI-PD control has been proposed. The characteristics of this control scheme lie between PID control and I-PD control. Similarly, PID-PD control can be considered. In these control schemes, we have a controller in the feedforward path and another controller in the feedback path. Such control schemes lead us to a more general two-degrees-of-freedom control scheme. We shall discuss details of such a two-degrees-of-freedom control scheme in subsequent sections of this chapter.

10–5 TWO-DEGREES-OF-FREEDOM CONTROL

Consider the system shown in Figure 10–22, where the system is subjected to the disturbance input $D(s)$ and noise input $N(s)$, in addition to the reference input $R(s)$. $G_c(s)$ is the transfer function of the controller and $G_p(s)$ is the transfer function of the plant. We assume that $G_p(s)$ is fixed and unalterable.

For this system, three closed-loop transfer functions $Y(s)/R(s) = G_{yr}$, $Y(s)/D(s) = G_{yd}$, and $Y(s)/N(s) = G_{yn}$ may be derived. They are

$$G_{yr} = \frac{Y(s)}{R(s)} = \frac{G_cG_p}{1 + G_cG_p}$$

$$G_{yd} = \frac{Y(s)}{D(s)} = \frac{G_p}{1 + G_cG_p}$$

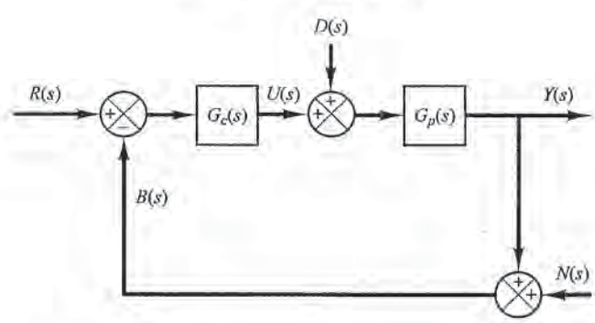


Figure 10–22
One-degree-of-freedom control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$G_{yn} = \frac{Y(s)}{N(s)} = -\frac{G_c G_p}{1 + G_c G_p}$$

[In deriving $Y(s)/R(s)$, we assumed $D(s) = 0$ and $N(s) = 0$. Similar comments apply to the derivations of $Y(s)/D(s)$ and $Y(s)/N(s)$.] The degrees of freedom of the control system refers to how many of these closed-loop transfer functions are independent. In the present case, we have

$$G_{yr} = \frac{G_p - G_{yd}}{G_p}$$

$$G_{yn} = \frac{G_{yd} - G_p}{G_p}$$

Among the three closed-loop transfer functions G_{yr}, G_{yn}, and G_{yd}, if one of them is given, the remaining two are fixed. This means that the system shown in Figure 10–22 is a one-degree-of-freedom control system.

Next consider the system shown in Figure 10–23, where $G_p(s)$ is the transfer function of the plant. For this system, closed-loop transfer functions G_{yr}, G_{yn}, and G_{yd} are given, respectively, by

$$G_{yr} = \frac{Y(s)}{R(s)} = \frac{G_{c1} G_p}{1 + (G_{c1} + G_{c2}) G_p}$$

$$G_{yd} = \frac{Y(s)}{D(s)} = \frac{G_p}{1 + (G_{c1} + G_{c2}) G_p}$$

$$G_{yn} = \frac{Y(s)}{N(s)} = -\frac{(G_{c1} + G_{c2}) G_p}{1 + (G_{c1} + G_{c2}) G_p}$$

Hence, we have

$$G_{yr} = G_{c1} G_{yd}$$

$$G_{yn} = \frac{G_{yd} - G_p}{G_p}$$

In this case, if G_{yd} is given, then G_{yn} is fixed, but G_{yr} is not fixed, because G_{c1} is independent of G_{yd}. Thus, two closed-loop transfer functions among three closed-loop

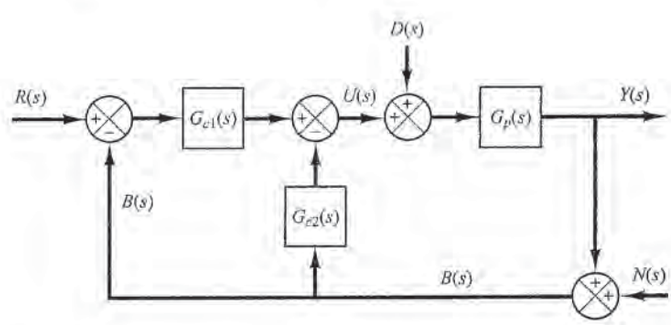


Figure 10–23
Two-degrees-of-freedom control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

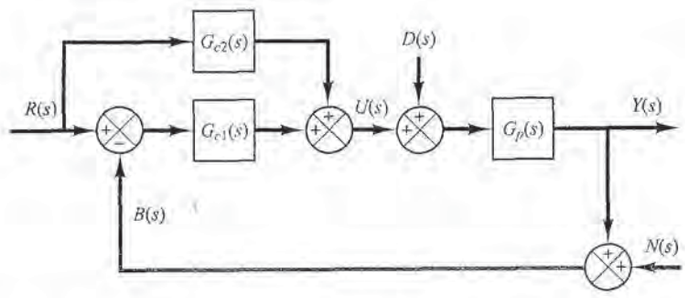


Figure 10–24
Two-degrees-of-freedom control system.

transfer functions G_{yr}, G_{yd}, and G_{yn} are independent. Hence, this system is a two-degrees-of-freedom control system.

Similarly, the system shown in Figure 10–24 is also a two-degrees-of-freedom control system, because for this system

$$G_{yr} = \frac{Y(s)}{R(s)} = \frac{G_{c1}G_p}{1 + G_{c1}G_p} + \frac{G_{c2}G_p}{1 + G_{c1}G_p}$$

$$G_{yd} = \frac{Y(s)}{D(s)} = \frac{G_p}{1 + G_{c1}G_p}$$

$$G_{yn} = \frac{Y(s)}{N(s)} = -\frac{G_{c1}G_p}{1 + G_{c1}G_p}$$

Hence,

$$G_{yr} = G_{c2}G_{yd} + \frac{G_p - G_{yd}}{G_p}$$

$$G_{yn} = \frac{G_{yd} - G_p}{G_p}$$

Clearly, if G_{yd} is given, then G_{yn} is fixed, but G_{yr} is not fixed, because G_{c2} is independent of G_{yd}.

It will be seen in Section 10–6 that, in such a two-degrees-of-freedom control system, both the closed-loop characteristics and the feedback characteristics can be adjusted independently to improve the system response performance.

10–6 ZERO-PLACEMENT APPROACH TO IMPROVE RESPONSE CHARACTERISTICS

We shall show here that by use of the zero-placement approach presented later in this section, we can achieve the following:

> The responses to the ramp reference input and acceleration reference input exhibit no steady-state errors.

In high-performance control systems it is always desired that the system output follow the changing input with minimum error. For step, ramp, and acceleration inputs, it is desired that the system output exhibit no steady-state error.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

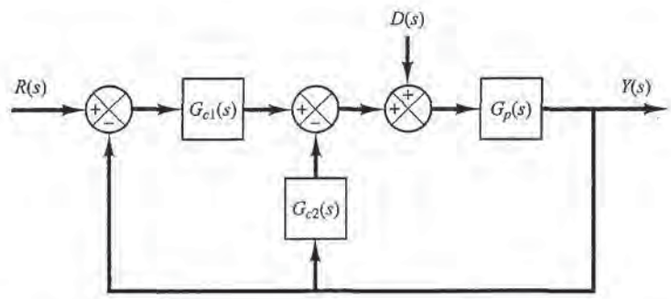


Figure 10–25
Two-degrees-of-freedom control system.

In what follows, we shall demonstrate how to design control systems that will exhibit no steady-state errors in following ramp and acceleration inputs and at the same time force the response to the step disturbance input to approach zero quickly.

Consider the two-degrees-of-freedom control system shown in Figure 10–25. Assume that the plant transfer function $G_p(s)$ is a minimum-phase transfer function and is given by

$$G_p(s) = K\frac{A(s)}{B(s)}$$

where

$$A(s) = (s + z_1)(s + z_2)\cdots(s + z_m)$$
$$B(s) = s^N(s + p_{N+1})(s + p_{N+2})\cdots(s + p_n)$$

where N may be 0, 1, 2 and $n \geq m$. Assume also that G_{c1} is a PID controller followed by a filter $1/A(s)$, or

$$G_{c1}(s) = \frac{\alpha_1 s + \beta_1 + \gamma_1 s^2}{s}\frac{1}{A(s)}$$

and G_{c2} is a PID, PI, PD, I, D, or P controller followed by a filter $1/A(s)$. That is

$$G_{c2}(s) = \frac{\alpha_2 s + \beta_2 + \gamma_2 s^2}{s}\frac{1}{A(s)}$$

where some of α_2, β_2, and γ_2 may be zero. Then it is possible to write $G_{c1} + G_{c2}$ as

$$G_{c1} + G_{c2} = \frac{\alpha s + \beta + \gamma s^2}{s}\frac{1}{A(s)} \tag{10–3}$$

where α, β, and γ are constants. Then

$$\frac{Y(s)}{D(s)} = \frac{G_p}{1 + (G_{c1} + G_{c2})G_p} = \frac{K\dfrac{A(s)}{B(s)}}{1 + \dfrac{\alpha s + \beta + \gamma s^2}{s}\dfrac{K}{B(s)}}$$
$$= \frac{sKA(s)}{sB(s) + (\alpha s + \beta + \gamma s^2)K}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Because of the presence of s in the numerator, the response $y(t)$ to a step disturbance input approaches zero as t approaches infinity, as shown below. Since

$$Y(s) = \frac{sKA(s)}{sB(s) + (\alpha s + \beta + \gamma s^2)K} D(s)$$

if the disturbance input is a step function of magnitude d, or

$$D(s) = \frac{d}{s}$$

and assuming the system is stable, then

$$\begin{aligned} y(\infty) &= \lim_{s \to 0} s \left[\frac{sKA(s)}{sB(s) + (\alpha s + \beta + \gamma s^2)K} \right] \frac{d}{s} \\ &= \lim_{s \to 0} \frac{sKA(0)d}{sB(0) + \beta K} \\ &= 0 \end{aligned}$$

The response $y(t)$ to a step disturbance input will have the general form shown in Figure 10–26.

Note that $Y(s)/R(s)$ and $Y(s)/D(s)$ are given by

$$\frac{Y(s)}{R(s)} = \frac{G_{c1}G_p}{1 + (G_{c1} + G_{c2})G_p}, \qquad \frac{Y(s)}{D(s)} = \frac{G_p}{1 + (G_{c1} + G_{c2})G_p}$$

Notice that the denominators of $Y(s)/R(s)$ and $Y(s)/D(s)$ are the same. (This is always true for any system. There is only one characteristic equation for the system.)

Before we choose the poles of $Y(s)/R(s)$, we need to place the zeros of $Y(s)/R(s)$.

Zero Placement. Consider the system

$$\frac{Y(s)}{R(s)} = \frac{p(s)}{s^{n+1} + a_n s^n + a_{n-1}s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0}$$

If we choose $p(s)$ as

$$p(s) = a_2 s^2 + a_1 s + a_0 = a_2(s + s_1)(s + s_2)$$

that is, choose the zeros $s = -s_1$ and $s = -s_2$ such that, together with a_2, the numerator polynomial $p(s)$ is equal to the sum of the last three terms of the denominator polynomial—then the system will exhibit no steady-state errors in response to the step input, ramp input, and acceleration input.

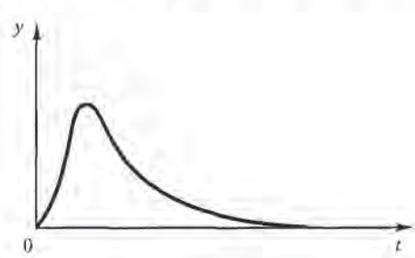


Figure 10–26
Typical response curve to a step disturbance input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Requirement Placed on System Response Characteristics. Suppose that it is desired that the maximum overshoot in the response to the unit-step reference input be between arbitrarily selected upper and lower limits—for example,

$$2\% < \text{maximum overshoot} < 10\%$$

where we choose the lower limit to be slightly above zero to avoid having overdamped systems. The smaller the upper limit, the harder it is to determine the coefficient a's. In some cases, no combination of the a's may exist to satisfy the specification, so we must allow a higher upper limit for the maximum overshoot. We use MATLAB to search at least one set of the a's to satisfy the specification. As a practical computational matter, instead of searching for the a's, we try to obtain acceptable closed-loop poles by searching a reasonable region in the left-half s plane for each closed-loop pole. Once we determine all closed-loop poles, then all coefficients $a_n, a_{n-1}, \ldots, a_1, a_0$ will be determined.

Determination of G_{c2}. Now that the coefficients of the transfer function $Y(s)/R(s)$ are all known and $Y(s)/R(s)$ is given by

$$\frac{Y(s)}{R(s)} = \frac{a_2 s^2 + a_1 s + a_0}{s^{n+1} + a_n s^n + a_{n-1} s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0} \tag{10–4}$$

we have

$$\begin{aligned}\frac{Y(s)}{R(s)} &= G_{c1}\frac{Y(s)}{D(s)} \\ &= \frac{G_{c1} sKA(s)}{sB(s) + (\alpha s + \beta + \gamma s^2)K} \\ &= \frac{G_{c1} sKA(s)}{s^{n+1} + a_n s^n + a_{n-1} s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0}\end{aligned}$$

Since G_{c1} is a PID controller and is given by

$$G_{c1} = \frac{\alpha_1 s + \beta_1 + \gamma_1 s^2}{s}\frac{1}{A(s)}$$

$Y(s)/R(s)$ can be written as

$$\frac{Y(s)}{R(s)} = \frac{K(\alpha_1 s + \beta_1 + \gamma_1 s^2)}{s^{n+1} + a_n s^n + a_{n-1} s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0}$$

Therefore, we choose

$$K\gamma_1 = a_2, \qquad K\alpha_1 = a_1, \qquad K\beta_1 = a_0$$

so that

$$G_{c1} = \frac{a_1 s + a_0 + a_2 s^2}{Ks}\frac{1}{A(s)} \tag{10–5}$$

The response of this system to the unit-step reference input can be made to exhibit the maximum overshoot between the chosen upper and lower limits, such as

$$2\% < \text{maximum overshoot} < 10\%$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The response of the system to the ramp reference input or acceleration reference input can be made to exhibit no steady-state error. The characteristic of the system of Equation (10–4) is that it generally exhibits a short settling time. If we wish to further shorten the settling time, then we need to allow a larger maximum overshoot—for example,

$$2\% < \text{maximum overshoot} < 20\%$$

The controller G_{c2} can now be determined from Equations (10–3) and (10–5). Since

$$G_{c1} + G_{c2} = \frac{\alpha s + \beta + \gamma s^2}{s} \frac{1}{A(s)}$$

we have

$$\begin{aligned} G_{c2} &= \left[\frac{\alpha s + \beta + \gamma s^2}{s} - \frac{a_1 s + a_0 + a_2 s^2}{Ks} \right] \frac{1}{A(s)} \\ &= \frac{(K\alpha - a_1)s + (K\beta - a_0) + (K\gamma - a_2)s^2}{Ks} \frac{1}{A(s)} \end{aligned} \tag{10–6}$$

The two controllers G_{c1} and G_{c2} are given by Equations (10–5) and (10–6), respectively.

EXAMPLE 10–4 Consider the two-degrees-of-freedom control system shown in Figure 10–27. The plant transfer function $G_p(s)$ is given by

$$G_p(s) = \frac{10}{s(s+1)}$$

Design controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that the maximum overshoot in the response to the unit-step reference input be less than 19%, but more than 2%, and the settling time be less than 1 sec. It is desired that the steady-state errors in following the ramp reference input and acceleration reference input be zero. The response to the unit-step disturbance input should have a small amplitude and settle to zero quickly.

To design suitable controllers $G_{c1}(s)$ and $G_{c2}(s)$, first note that

$$\frac{Y(s)}{D(s)} = \frac{G_p}{1 + G_p(G_{c1} + G_{c2})}$$

To simplify the notation, let us define

$$G_c = G_{c1} + G_{c2}$$

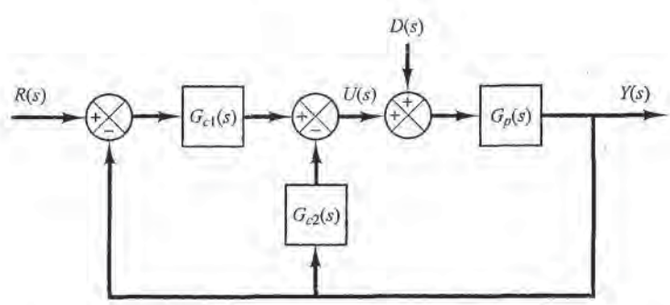


Figure 10–27 Two-degrees-of-freedom control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Then

$$\frac{Y(s)}{D(s)} = \frac{G_p}{1 + G_pG_c} = \frac{\dfrac{10}{s(s+1)}}{1 + \dfrac{10}{s(s+1)}G_c}$$

$$= \frac{10}{s(s+1) + 10G_c}$$

Second, note that

$$\frac{Y(s)}{R(s)} = \frac{G_pG_{c1}}{1 + G_pG_c} = \frac{10G_{c1}}{s(s+1) + 10G_c}$$

Notice that the characteristic equation for $Y(s)/D(s)$ and the one for $Y(s)/R(s)$ are identical. (The system always has only one characteristic equation.)

We may be tempted to choose a zero of $G_c(s)$ at $s = -1$ to cancel a pole at $s = -1$ of the plant $G_p(s)$. However, the canceled pole $s = -1$ becomes a closed-loop pole of the entire system, as seen below. If we define $G_c(s)$ as a PID controller such that

$$G_c(s) = \frac{K(s+1)(s+\beta)}{s} \tag{10–7}$$

then

$$\frac{Y(s)}{D(s)} = \frac{10}{s(s+1) + \dfrac{10K(s+1)(s+\beta)}{s}}$$

$$= \frac{10s}{(s+1)[s^2 + 10K(s+\beta)]}$$

The closed-loop pole at $s = -1$ is a slow-response pole, and if this closed-loop pole is included in the system, the settling time will not be less than 1 sec. Therefore, we should not choose $G_c(s)$ as given by Equation (10–7).

The design of controllers $G_{c1}(s)$ and $G_{c2}(s)$ consists of two steps.

Design Step 1: We design $G_c(s)$ to satisfy the requirements on the response to the step-disturbance input $D(s)$. In this design stage, we assume that the reference input is zero.

Suppose that we assume that $G_c(s)$ is a PID controller of the form

$$G_c(s) = \frac{K(s+\alpha)(s+\beta)}{s}$$

Then the closed-loop transfer function $Y(s)/D(s)$ becomes

$$\frac{Y(s)}{D(s)} = \frac{10}{s(s+1) + 10G_c}$$

$$= \frac{10}{s(s+1) + \dfrac{10K(s+\alpha)(s+\beta)}{s}}$$

$$= \frac{10s}{s^2(s+1) + 10K(s+\alpha)(s+\beta)}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Note that the presence of "s" in the numerator of $Y(s)/D(s)$ assures that the steady-state response to the step disturbance input is zero.

Let us assume that the desired dominant closed-loop poles are complex conjugates and are given by

$$s = -a \pm jb$$

and the remaining closed-loop pole is real and is located at

$$s = -c$$

Note that in this problem there are three requirements. The first requirement is that the response to the step disturbance input damp out quickly. The second requirement is that the maximum overshoot in the response to the unit-step reference input be between 19% and 2% and the settling time be less than 1 sec. The third requirement is that the steady-state errors in the responses to both the ramp and acceleration reference inputs be zero.

A set (or sets) of reasonable values of a, b, and c must be searched using a computational approach. To satisfy the first requirement, we choose the search region for a, b, and c to be

$$2 \le a \le 6, \qquad 2 \le b \le 6, \qquad 6 \le c \le 12$$

This region is shown in Figure 10–28. If the dominant closed-loop poles $s = -a \pm jb$ are located anywhere in the shaded region, the response to a step disturbance input will damp out quickly. (The first requirement will be met.)

Notice that the denominator of $Y(s)/D(s)$ can be written as

$$\begin{aligned}
&s^2(s+1) + 10K(s+\alpha)(s+\beta) \\
&= s^3 + (1+10K)s^2 + 10K(\alpha+\beta)s + 10K\alpha\beta \\
&= (s+a+jb)(s+a-jb)(s+c) \\
&= s^3 + (2a+c)s^2 + (a^2+b^2+2ac)s + (a^2+b^2)c
\end{aligned}$$

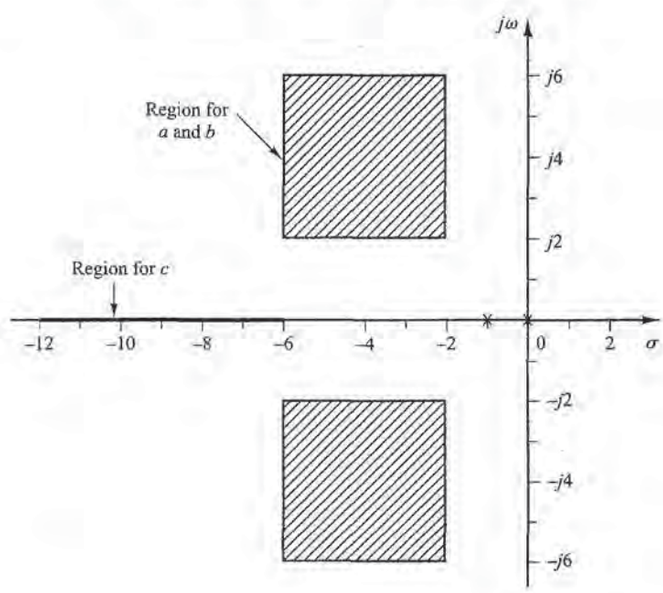


Figure 10–28
Search regions for a, b, and c.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Since the denominators of $Y(s)/D(s)$ and $Y(s)/R(s)$ are the same, the denominator of $Y(s)/D(s)$ determines also the response characteristics for the reference input. To satisfy the third requirement, we refer to the zero-placement method and choose the closed-loop transfer function $Y(s)/R(s)$ to be of the following form:

$$\frac{Y(s)}{R(s)} = \frac{(2a + c)s^2 + (a^2 + b^2 + 2ac)s + (a^2 + b^2)c}{s^3 + (2a + c)s^2 + (a^2 + b^2 + 2ac)s + (a^2 + b^2)c}$$

in which case the third requirement is automatically satisfied.

Our problem then becomes a search of a set or sets of desired closed-loop poles in terms of a, b, and c in the specified region, such that the system will satisfy the requirement on the response to the unit-step reference input that the maximum overshoot be between 19% and 2% and the settling time be less than 1 sec. (If an acceptable set cannot be found in the search region, we need to widen the region.)

In the computational search, we need to assume a reasonable step size. In this problem, we assume it to be 0.2.

MATLAB Program 10–5 produces a table of sets of acceptable values of a, b, and c. Using this program, we find that the requirement on the response to the unit-step reference input is met by any of the 23 sets shown in the table in MATLAB Program 10–5. Note that the last row in the table corresponds to the last search point. This point does not satisfy the requirement and thus it should simply be ignored. (In the program written, the last search point produces the last row in the table whether or not it satisfies the requirement.)

MATLAB Program 10–5

```
t = 0:0.01:4;
k = 0;
for i = 1:21;
   a(i) = 6.2-i*0.2;
   for j = 1:21;
      b(j) = 6.2-j*0.2;
      for h = 1:31;
         c(h) = 12.2-h*0.2;
      num = [0  2*a(i)+c(h)  a(i)^2+b(j)^2+2*a(i)*c(h)  (a(i)^2+b(j)^2)*c(h)];
      den = [1  2*a(i)+c(h)  a(i)^2+b(j)^2+2*a(i)*c(h)  (a(i)^2+b(j)^2)*c(h)];
            y = step(num,den,t);
            m = max(y);
            s = 401; while y(s) > 0.98 & y(s) < 1.02;
            s = s-1; end;
            ts = (s-1)*0.01;
         if m < 1.19 & m > 1.02 & ts < 1.0;
         k = k+1;
         table(k,:) = [a(i)  b(j)  c(h)  m  ts];
         end
      end
   end
 end
```

(*continues on next page*)

0725

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

```
table(k,:) = [a(i)  b(j)  c(h)  m  ts]

table =

    4.2000  2.0000  12.0000  1.1896  0.8500
    4.0000  2.0000  12.0000  1.1881  0.8700
    4.0000  2.0000  11.8000  1.1890  0.8900
    4.0000  2.0000  11.6000  1.1899  0.9000
    3.8000  2.2000  12.0000  1.1883  0.9300
    3.8000  2.2000  11.8000  1.1894  0.9400
    3.8000  2.0000  12.0000  1.1861  0.8900
    3.8000  2.0000  11.8000  1.1872  0.9100
    3.8000  2.0000  11.6000  1.1882  0.9300
    3.8000  2.0000  11.4000  1.1892  0.9400
    3.6000  2.4000  12.0000  1.1893  0.9900
    3.6000  2.2000  12.0000  1.1867  0.9600
    3.6000  2.2000  11.8000  1.1876  0.9800
    3.6000  2.2000  11.6000  1.1886  0.9900
    3.6000  2.0000  12.0000  1.1842  0.9200
    3.6000  2.0000  11.8000  1.1852  0.9400
    3.6000  2.0000  11.6000  1.1861  0.9500
    3.6000  2.0000  11.4000  1.1872  0.9700
    3.6000  2.0000  11.2000  1.1883  0.9800
    3.4000  2.0000  12.0000  1.1820  0.9400
    3.4000  2.0000  11.8000  1.1831  0.9600
    3.4000  2.0000  11.6000  1.1842  0.9800
    3.2000  2.0000  12.0000  1.1797  0.9600
    2.0000  2.0000   6.0000  1.2163  1.8900
```

As noted above, 23 sets of variables a, b, and c satisfy the requirement. Unit-step response curves of the system with any of the 23 sets are about the same. The unit-step response curve with

$$a = 4.2, \qquad b = 2, \qquad c = 12$$

is shown in Figure 10–29(a). The maximum overshoot is 18.96% and the settling time is 0.85 sec. Using these values of a, b, and c, the desired closed-loop poles are located at

$$s = -4.2 \pm j2, \qquad s = -12$$

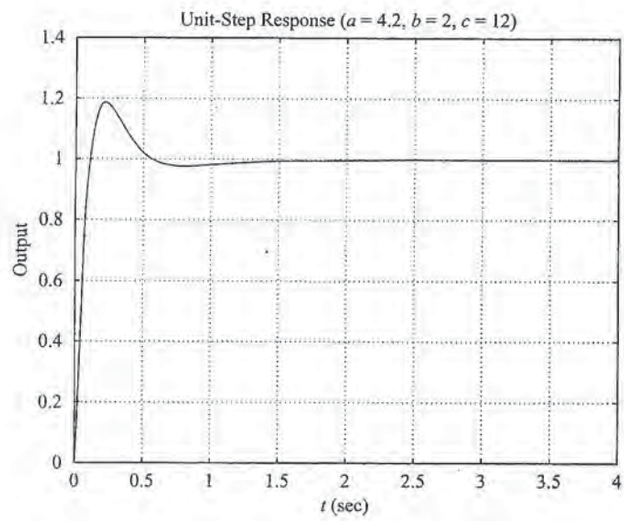
Using these closed-loop poles, the denominator of $Y(s)/D(s)$ becomes

$$s^2(s + 1) + 10K(s + \alpha)(s + \beta) = (s + 4.2 + j2)(s + 4.2 - j2)(s + 12)$$

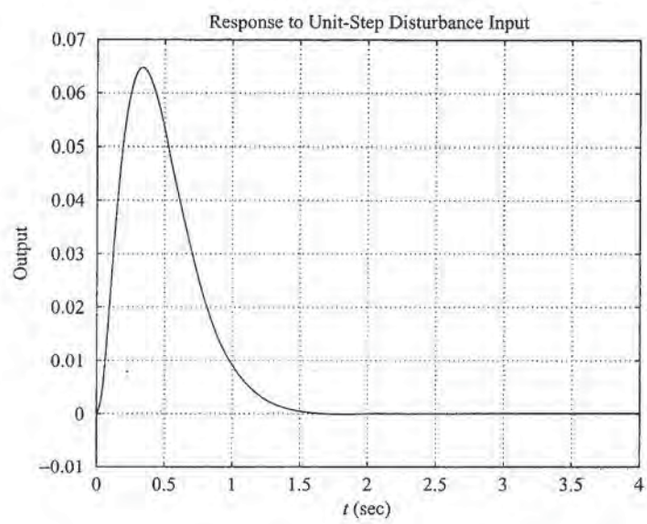
or

$$s^3 + (1 + 10K)s^2 + 10K(\alpha + \beta)s + 10K\alpha\beta = s^3 + 20.4s^2 + 122.44s + 259.68$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565



(a)



(b)

Figure 10–29
(a) Response to unit-step reference input ($a = 4.2, b = 2, c = 12$); (b) response to unit-step disturbance input ($a = 4.2, b = 2, c = 12$).

By equating the coefficients of equal powers of s on both sides of this last equation, we obtain

$$1 + 10K = 20.4$$

$$10K(\alpha + \beta) = 122.44$$

$$10K\alpha\beta = 259.68$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Hence

$$K = 1.94, \qquad \alpha + \beta = \frac{122.44}{19.4}, \qquad \alpha\beta = \frac{259.68}{19.4}$$

Then $G_c(s)$ can be written as

$$\begin{aligned} G_c(s) &= K\frac{(s+\alpha)(s+\beta)}{s} \\ &= \frac{K[s^2 + (\alpha+\beta)s + \alpha\beta]}{s} \\ &= \frac{1.94s^2 + 12.244s + 25.968}{s} \end{aligned}$$

The closed-loop transfer function $Y(s)/D(s)$ becomes

$$\begin{aligned} \frac{Y(s)}{D(s)} &= \frac{10}{s(s+1) + 10G_c} \\ &= \frac{10}{s(s+1) + 10\dfrac{1.94s^2 + 12.244s + 25.968}{s}} \\ &= \frac{10s}{s^3 + 20.4s^2 + 122.44s + 259.68} \end{aligned}$$

Using this expression, the response $y(t)$ to a unit-step disturbance input can be obtained as shown in Figure 10–29(b).

Figure 10–30(a) shows the response of the system to the unit-step reference input when a, b, and c are chosen as

$$a = 3.2, \qquad b = 2, \qquad c = 12$$

Figure 10–30(b) shows the response of this system when it is subjected to a unit-step disturbance input. Comparing Figures 10–29(a) and Figure 10–30(a), we find that they are about the same. However, comparing Figures 10–29(b) and 10–30(b), we find the former to be a little bit better than the latter. Comparing the responses of systems with each set in the table, we conclude the first set of values ($a = 4.2$, $b = 2$, $c = 12$) to be one of the best. Therefore, as the solution to this problem, we choose

$$a = 4.2, \qquad b = 2, \qquad c = 12$$

Design Step 2: Next, we determine G_{c1}. Since $Y(s)/R(s)$ can be given by

$$\begin{aligned} \frac{Y(s)}{R(s)} &= \frac{G_pG_{c1}}{1 + G_pG_c} \\ &= \frac{\dfrac{10}{s(s+1)}G_{c1}}{1 + \dfrac{10}{s(s+1)}\dfrac{1.94s^2 + 12.244s + 25.968}{s}} \\ &= \frac{10sG_{c1}}{s^3 + 20.4s^2 + 122.44s + 259.68} \end{aligned}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

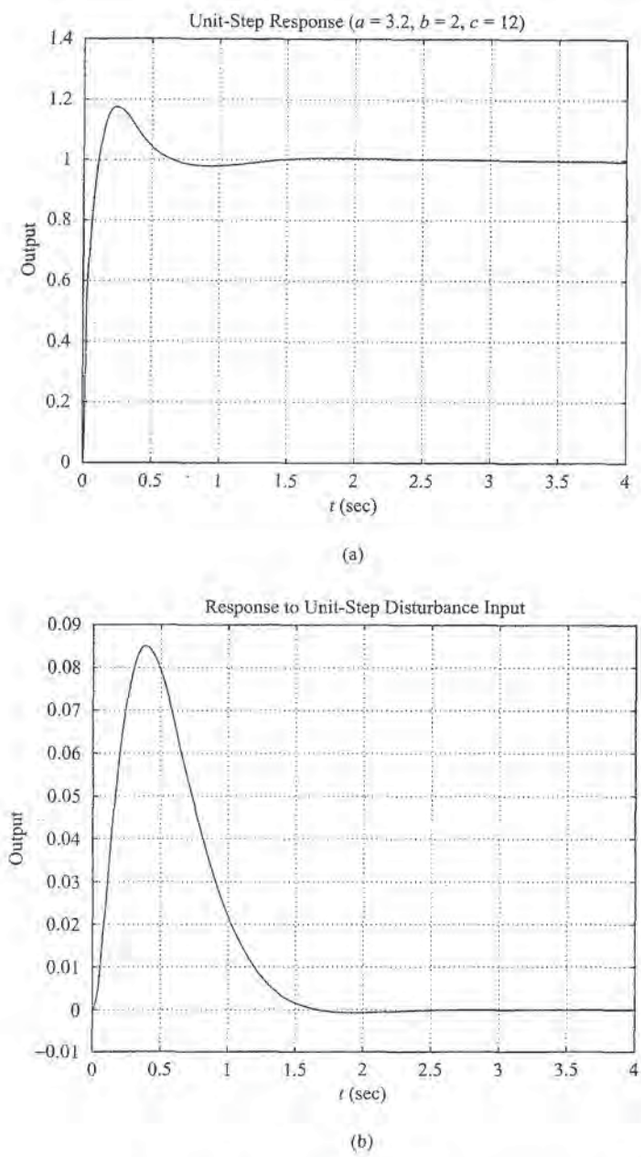


Figure 10–30
(a) Response to unit-step reference input ($a = 3.2, b = 2, c = 12$); (b) response to unit-step disturbance input ($a = 3.2, b = 2, c = 12$).

our problem becomes that of designing $G_{c1}(s)$ to satisfy the requirements on the responses to the step, ramp, and acceleration inputs.

Since the numerator involves "s", $G_{c1}(s)$ must include an integrator to cancel this "s". [Although we want "s" in the numerator of the closed-loop transfer function $Y(s)/D(s)$ to obtain zero steady-state error to the step disturbance input, we do not want to have "s" in the numerator of the closed-loop transfer function $Y(s)/R(s)$.] To eliminate the offset in the response to the step reference input and eliminate the steady-state errors in following the ramp reference input

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

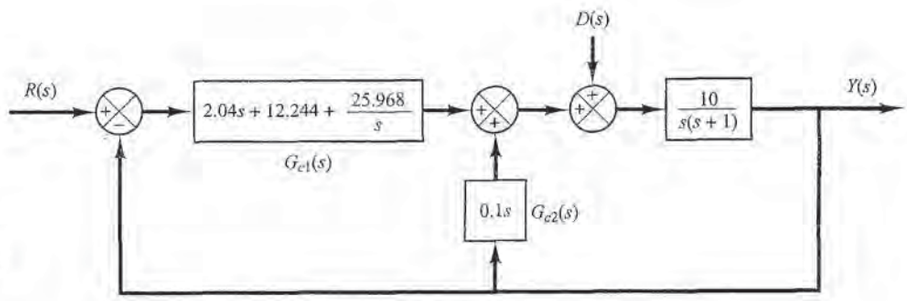


Figure 10–31
Block diagram of the designed system.

and acceleration reference input, the numerator of $Y(s)/R(s)$ must be equal to the last three terms of the denominator, as mentioned earlier. That is,

$$10sG_{c1}(s) = 20.4s^2 + 122.44s + 259.68$$

or

$$G_{c1}(s) = 2.04s + 12.244 + \frac{25.968}{s}$$

Thus, $G_{c1}(s)$ is a PID controller. Since $G_c(s)$ is given as

$$G_c(s) = G_{c1}(s) + G_{c2}(s) = \frac{1.94s^2 + 12.244s + 25.968}{s}$$

we obtain

$$\begin{aligned} G_{c2}(s) &= G_c(s) - G_{c1}(s) \\ &= \left(1.94s + 12.244 + \frac{25.968}{s}\right) - \left(2.04s + 12.244 + \frac{25.968}{s}\right) \\ &= -0.1s \end{aligned}$$

Thus, $G_{c2}(s)$ is a derivative controller. A block diagram of the designed system is shown in Figure 10–31.

The closed-loop transfer function $Y(s)/R(s)$ now becomes

$$\frac{Y(s)}{R(s)} = \frac{20.4s^2 + 122.44s + 259.68}{s^3 + 20.4s^2 + 122.44s + 259.68}$$

The response to the unit-ramp reference input and that to the unit-acceleration reference input are shown in Figures 10–32(a) and (b), respectively. The steady-state errors in following the ramp input and acceleration input are zero. Thus, all the requirements of the problem are satisfied. Hence, the designed controllers $G_{c1}(s)$ and $G_{c2}(s)$ are acceptable.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

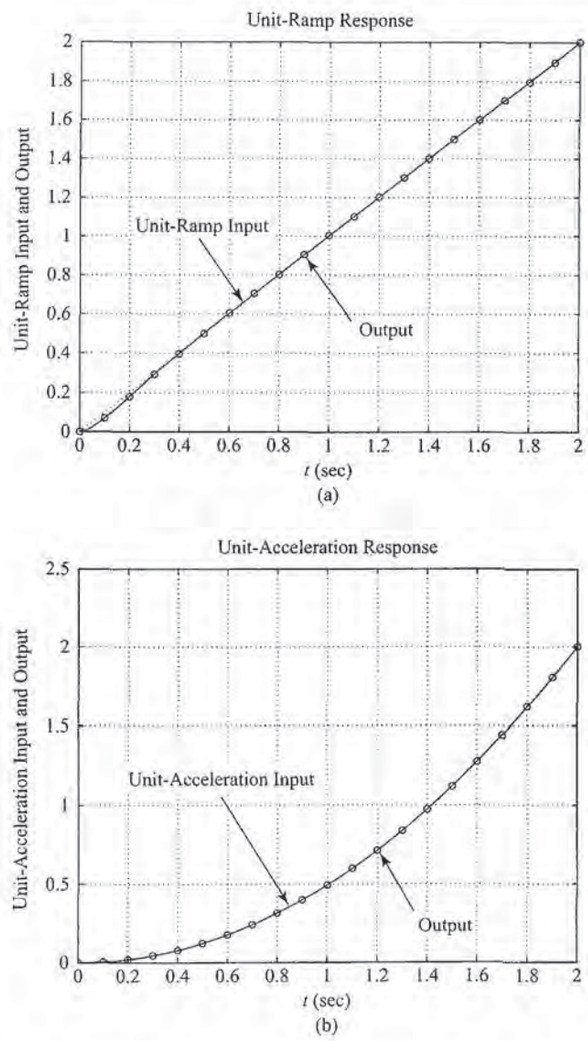


Figure 10–32
(a) Response to unit-ramp reference input; (b) response to unit-acceleration reference input.

EXAMPLE 10–5 Consider the control system shown in Figure 10–33. This is a two-degrees-of-freedom system. In the design problem considered here, we assume that the noise input $N(s)$ is zero. Assume that the plant transfer function $G_p(s)$ is given by

$$G_p(s) = \frac{5}{(s+1)(s+5)}$$

Assume also that the controller $G_{c1}(s)$ is of PID type. That is,

$$G_{c1}(s) = K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

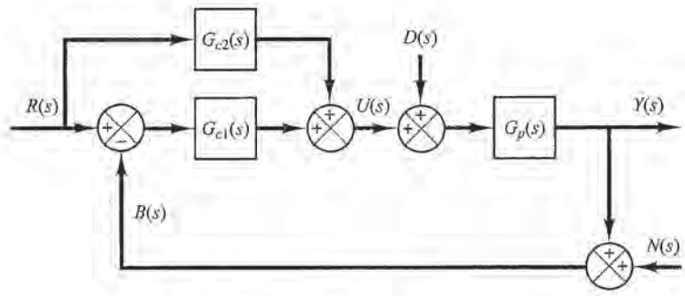


Figure 10–33
Two-degrees-of-freedom control system.

The controller $G_{c2}(s)$ is of P or PD type. [If $G_{c2}(s)$ involves integral control action, then this will introduce a ramp component in the input signal, which is not desirable. Therefore, $G_{c2}(s)$ should not include the integral control action.] Thus, we assume that

$$G_{c2}(s) = \hat{K}_p(1 + \hat{T}_d s)$$

where $\hat{T}_d$ may be zero.

Let us design controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that the responses to the step-disturbance input and the step-reference input are of "desirable characteristics" in the sense that

1. The response to the step-disturbance input will have a small peak and eventually approach zero. (That is, there will be no steady-state error.)
2. The response to the step reference input will exhibit less than 25% overshoot with a settling time less than 2 sec. The steady-state errors to the ramp reference input and acceleration reference input should be zero.

The design of this two-degrees-of-freedom control system may be carried out by following the steps **1** and **2** below.

1. Determine $G_{c1}(s)$ so that the response to the step-disturbance input is of desirable characteristics.
2. Design $G_{c2}(s)$ so that the responses to the reference inputs are of desirable characteristics without changing the response to the step disturbance considered in step **1**.

Design of $G_{c1}(s)$: First, note that we assumed the noise input $N(s)$ to be zero. To obtain the response to the step-disturbance input, we assume that the reference input is zero. Then the block diagram which relates $Y(s)$ and $D(s)$ can be drawn as shown in Figure 10–34. The transfer function $Y(s)/D(s)$ is given by

$$\frac{Y(s)}{D(s)} = \frac{G_p}{1 + G_{c1}G_p}$$

where

$$G_{c1}(s) = K_p\left(1 + \frac{1}{T_i s} + T_d s\right)$$

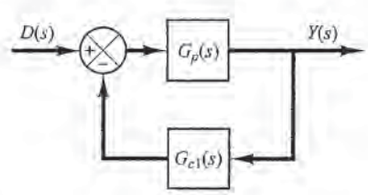


Figure 10–34
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

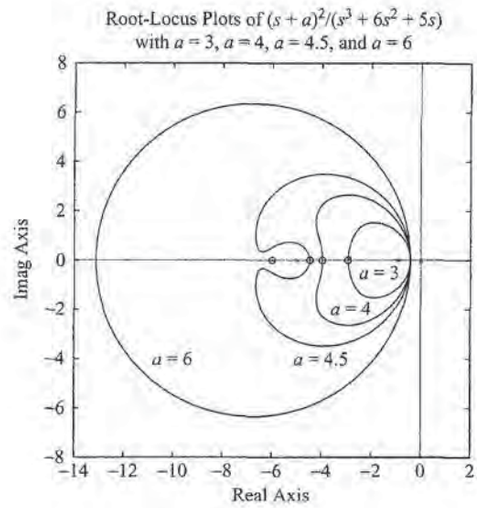


Figure 10–35
Root-locus plots of $5K(s + a)^2/[s(s + 1)(s + 5)]$ when $a = 3$, $a = 4$, $a = 4.5$, and $a = 6$.

This controller involves one pole at the origin and two zeros. If we assume that the two zeros are located at the same place (a double zero), then $G_{c1}(s)$ can be written as

$$G_{c1}(s) = K\frac{(s + a)^2}{s}$$

Then the characteristic equation for the system becomes

$$1 + G_{c1}(s)G_p(s) = 1 + \frac{K(s + a)^2}{s}\frac{5}{(s + 1)(s + 5)} = 0$$

or

$$s(s + 1)(s + 5) + 5K(s + a)^2 = 0$$

which can be rewritten as

$$s^3 + (6 + 5K)s^2 + (5 + 10Ka)s + 5Ka^2 = 0 \tag{10–8}$$

If we place the double zero between $s = -3$ and $s = -6$, then the root-locus plot of $G_{c1}(s)G_p(s)$ may look like the one shown in Figure 10–35. The speed of response should be fast, but not faster than necessary, because faster response generally implies larger or more expensive components. Therefore, we may choose the dominant closed-loop poles at

$$s = -3 \pm j2$$

(Note that this choice is not unique. There are infinitely many possible closed-loop poles that we may choose from.)

Since the system is of third order, there are three closed-loop poles. The third one is located on the negative real axis to the left of point $s = -5$.

Let us substitute $s = -3 + j2$ into Equation (10–8).

$$(-3 + j2)^3 + (6 + 5K)(-3 + j2)^2 + (5 + 10Ka)(-3 + j2) + 5Ka^2 = 0$$

which can be simplified to

$$24 + 25K - 30Ka + 5Ka^2 + j(-16 - 60K + 20Ka) = 0$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

By equating the real part and imaginary part to zero, respectively, we obtain

$$24 + 25K - 30Ka + 5Ka^2 = 0 \tag{10–9}$$

$$-16 - 60K + 20Ka = 0 \tag{10–10}$$

From Equation (10–10), we have

$$K = \frac{4}{5a - 15} \tag{10–11}$$

Substituting Equation (10–11) into Equation (10–9), we get

$$a^2 = 13$$

or $a = 3.6056$ or -3.6056. Notice that the values of K become

$$K = 1.3210 \qquad \text{for } a = 3.6056$$

$$K = -0.1211 \qquad \text{for } a = -3.6056$$

Since $G_{c1}(s)$ is in the feedforward path, the gain K should be positive. Hence, we choose

$$K = 1.3210, \qquad a = 3.6056$$

Then $G_{c1}(s)$ can be given by

$$\begin{aligned} G_{c1}(s) &= K\frac{(s + a)^2}{s} \\ &= 1.3210\frac{(s + 3.6056)^2}{s} \\ &= \frac{1.3210s^2 + 9.5260s + 17.1735}{s} \end{aligned}$$

To determine K_p, T_i, and T_d, we proceed as follows:

$$\begin{aligned} G_{c1}(s) &= \frac{1.3210(s^2 + 7.2112s + 13)}{s} \\ &= 9.5260\left(1 + \frac{1}{0.5547s} + 0.1387s\right) \end{aligned} \tag{10–12}$$

Thus,

$$K_p = 9.5260, \qquad T_i = 0.5547, \qquad T_d = 0.1387$$

To check the response to a unit-step disturbance input, we obtain the closed-loop transfer function $Y(s)/D(s)$.

$$\begin{aligned} \frac{Y(s)}{D(s)} &= \frac{G_p}{1 + G_{c1}G_p} \\ &= \frac{5s}{s(s + 1)(s + 5) + 5K(s + a)^2} \\ &= \frac{5s}{s^3 + 12.605s^2 + 52.63s + 85.8673} \end{aligned}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

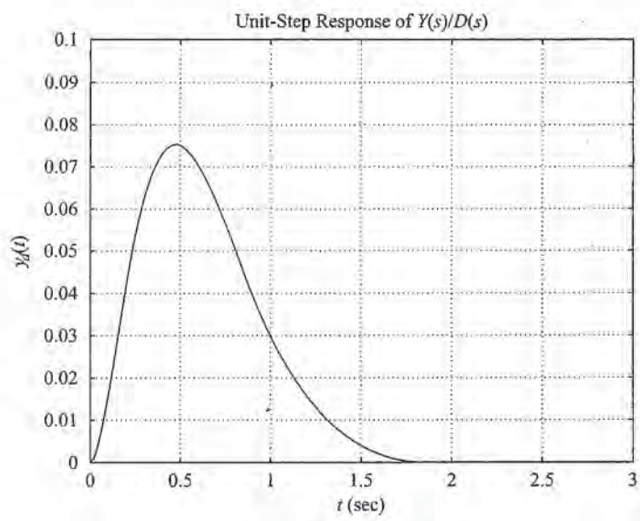


Figure 10–36
Response to unit-step disturbance input.

The response to the unit-step disturbance input is shown in Figure 10–36. The response curve seems good and acceptable. Note that the closed-loop poles are located at $s = -3 \pm j2$ and $s = -6.6051$. The complex-conjugate closed-loop poles act as dominant closed-loop poles.

Design of $G_{c2}(s)$: We now design $G_{c2}(s)$ to obtain the desired responses to the reference inputs. The closed-loop transfer function $Y(s)/R(s)$ can be given by

$$\begin{aligned}\frac{Y(s)}{R(s)} &= \frac{(G_{c1} + G_{c2})G_p}{1 + G_{c1}G_p} \\ &= \frac{\left[\dfrac{1.321s^2 + 9.526s + 17.1735}{s} + \hat{K}_p(1 + \hat{T}_d s)\right]\dfrac{5}{(s+1)(s+5)}}{1 + \dfrac{1.321s^2 + 9.526s + 17.1735}{s}\dfrac{5}{(s+1)(s+5)}} \\ &= \frac{(6.6051 + 5\hat{K}_p\hat{T}_d)s^2 + (47.63 + 5\hat{K}_p)s + 85.8673}{s^3 + 12.6051s^2 + 52.63s + 85.8673}\end{aligned}$$

Zero placement. We place two zeros together with the dc gain constant such that the numerator is the same as the sum of the last three terms of the denominator. That is,

$$(6.6051 + 5\hat{K}_p\hat{T}_d)s^2 + (47.63 + 5\hat{K}_p)s + 85.8673 = 12.6051s^2 + 52.63s + 85.8673$$

By equating the coefficients of s^2 terms and s terms on both sides of this last equation,

$$6.6051 + 5\hat{K}_p\hat{T}_d = 12.6051$$
$$47.63 + 5\hat{K}_p = 52.63$$

from which we get

$$\hat{K}_p = 1, \qquad \hat{T}_d = 1.2$$

Therefore,

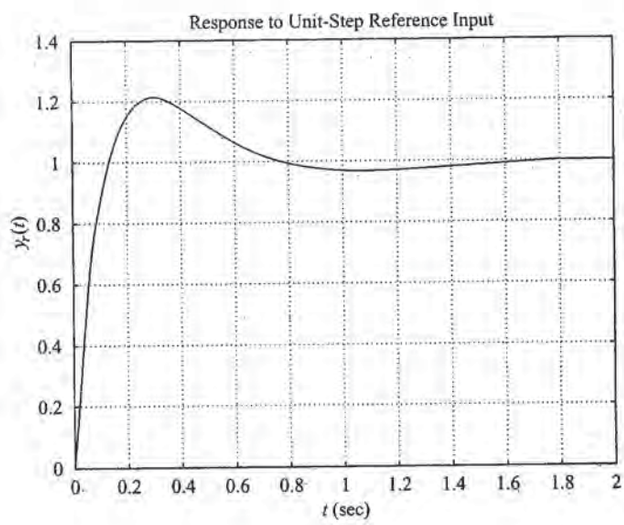
$$G_{c2}(s) = 1 + 1.2s \tag{10–13}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

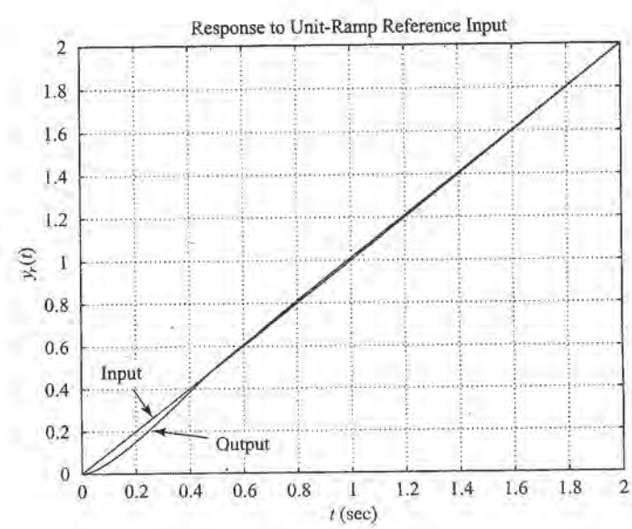
With this controller $G_{c2}(s)$, the closed-loop transfer function $Y(s)/R(s)$ becomes

$$\frac{Y(s)}{R(s)} = \frac{12.6051s^2 + 52.63s + 85.8673}{s^3 + 12.6051s^2 + 52.63s + 85.8673}$$

The response to the unit-step reference input becomes as shown in Figure 10–37(a).



(a)



(b)

Figure 10–37
(a) Response to unit-step reference input; (b) response to unit-ramp reference input; (c) response to unit-acceleration reference input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

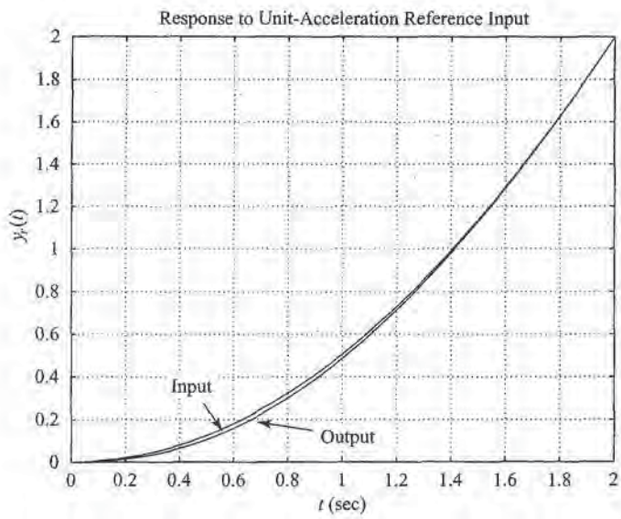


Figure 10–37
(continued)

(c)

The response exhibits the maximum overshoot of 21% and the settling time is approximately 1.6 sec. Figures 10–37(b) and (c) show the ramp response and acceleration response. The steady-state errors in both responses are zero. The response to the step disturbance was satisfactory. Thus, the designed controllers $G_{c1}(s)$ and $G_{c2}(s)$ given by Equations (10–12) and (10–13), respectively, are satisfactory.

If the response characteristics to the unit-step reference input are not satisfactory, we need to change the location of the dominant closed-loop poles and repeat the design process. The dominant closed-loop poles should lie in a certain region in the left-half s plane (such as $2 \le a \le 6$, $2 \le b \le 6, 6 \le c \le 12$). If the computational search is desired, write a computer program (similar to MATLAB Program 10–5) and execute the search process. Then a desired set or sets of values of a, b, and c may be found such that the system response to the unit-step reference input satisfies all requirements on maximum overshoot and settling time.

EXAMPLE PROBLEMS AND SOLUTIONS

A–10–1. Describe briefly the dynamic characteristics of the PI controller, PD controller, and PID controller.

Solution. The PI controller is characterized by the transfer function

$$G_c(s) = K_p\left(1 + \frac{1}{T_i s}\right)$$

The PI controller is a lag compensator. It possesses a zero at $s = -1/T_i$ and a pole at $s = 0$. Thus, the characteristic of the PI controller is infinite gain at zero frequency. This improves the steady-state characteristics. However, inclusion of the PI control action in the system increases the type number of the compensated system by 1, and this causes the compensated system to be less

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

stable or even makes the system unstable. Therefore, the values of K_p and T_i must be chosen carefully to ensure a proper transient response. By properly designing the PI controller, it is possible to make the transient response to a step input exhibit relatively small or no overshoot. The speed of response, however, becomes much slower. This is because the PI controller, being a low-pass filter, attenuates the high-frequency components of the signal.

The PD controller is a simplified version of the lead compensator. The PD controller has the transfer function $G_c(s)$, where

$$G_c(s) = K_p(1 + T_d s)$$

The value of K_p is usually determined to satisfy the steady-state requirement. The corner frequency $1/T_d$ is chosen such that the phase lead occurs in the neighborhood of the gain crossover frequency. Although the phase margin can be increased, the magnitude of the compensator continues to increase for the frequency region $1/T_d < \omega$. (Thus, the PD controller is a high-pass filter.) Such a continued increase of the magnitude is undesirable, since it amplifies high-frequency noises that may be present in the system. Lead compensation can provide a sufficient phase lead, while the increase of the magnitude for the high-frequency region is very much smaller than that for PD control. Therefore, lead compensation is preferred over PD control.

Because the transfer function of the PD controller involves one zero, but no pole, it is not possible to electrically realize it by passive *RLC* elements only. Realization of the PD controller using op amps, resistors, and capacitors is possible, but because the PD controller is a high-pass filter, as mentioned earlier, the differentiation process involved may cause serious noise problems in some cases. There is, however, no problem if the PD controller is realized by use of the hydraulic or pneumatic elements.

The PD control, as in the case of the lead compensator, improves the transient-response characteristics, improves system stability, and increases the system bandwidth, which implies fast rise time.

The PID controller is a combination of the PI and PD controllers. It is a lag–lead compensator. Note that the PI control action and PD control action occur in different frequency regions. The PI control action occurs at the low-frequency region and PD control action occurs at the high-frequency region. The PID control may be used when the system requires improvements in both transient and steady-state performances.

A–10–2. Show that the transfer function $U(s)/E(s)$ of the PID controller shown in Figure 10–38 is

$$\frac{U(s)}{E(s)} = K_0 \frac{T_1 + T_2}{T_1}\left[1 + \frac{1}{(T_1 + T_2)s} + \frac{T_1 T_2 s}{T_1 + T_2}\right]$$

Assume that the gain K is very large compared with unity, or $K \gg 1$.

Solution.

$$\frac{U(s)}{E(s)} = \frac{K}{1 + K\left(\frac{1}{K_0}\frac{T_1 s}{1 + T_1 s}\frac{1}{1 + T_2 s}\right)}$$

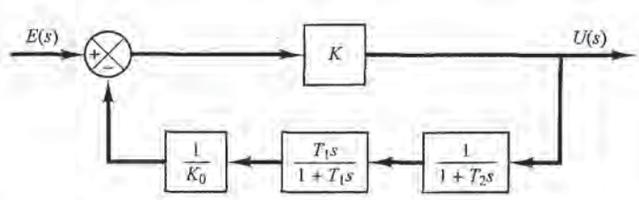


Figure 10–38
PID controller.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$
\begin{aligned}
&\doteq \frac{K}{K\left(\dfrac{1}{K_0}\dfrac{T_1 s}{1+T_1 s}\dfrac{1}{1+T_2 s}\right)} \\
&= \frac{K_0(1+T_1 s)(1+T_2 s)}{T_1 s} \\
&= K_0\left(1+\frac{1}{T_1 s}\right)(1+T_2 s) \\
&= K_0\left(1+\frac{1}{T_1 s}+T_2 s+\frac{T_2}{T_1}\right) \\
&= K_0\frac{T_1+T_2}{T_1}\left[1+\frac{1}{(T_1+T_2)s}+\frac{T_1 T_2 s}{T_1+T_2}\right]
\end{aligned}
$$

A–10–3. Consider the electronic circuit involving two operational amplifiers shown in Figure 10–39. This is a modified PID controller in that the transfer function involves an integrator and a first-order lag term. Obtain the transfer function of this PID controller.

Solution. Since

$$
Z_1 = \frac{1}{\dfrac{1}{R_1}+C_1 s} + R_3 = \frac{R_1+R_3+R_1R_3C_1 s}{1+R_1C_1 s}
$$

and

$$
Z_2 = R_2 + \frac{1}{C_2 s}
$$

we have

$$
\frac{E(s)}{E_i(s)} = -\frac{Z_2}{Z_1} = -\frac{(R_2C_2+1)(R_1C_1 s+1)}{C_2 s(R_1+R_3+R_1R_3C_1 s)}
$$

Also,

$$
\frac{E_o(s)}{E(s)} = -\frac{R_5}{R_4}
$$

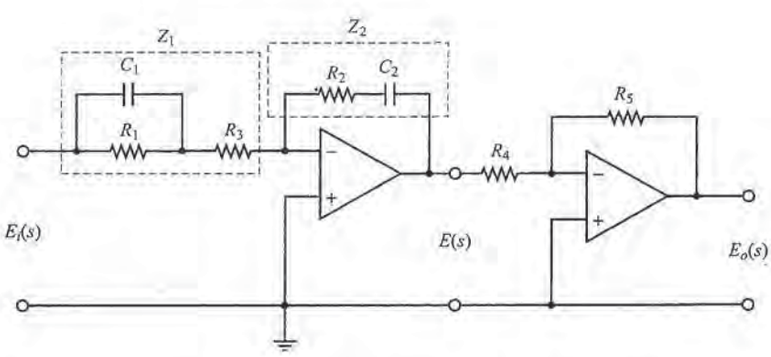


Figure 10–39
Modified PID controller.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

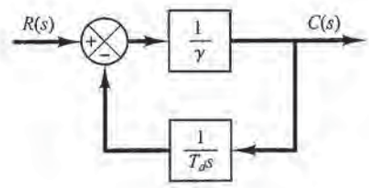


Figure 10–40
Approximate differentiator.

Consequently,

$$\frac{E_o(s)}{E_i(s)} = \frac{E_o(s)}{E(s)}\frac{E(s)}{E_i(s)} = \frac{R_5}{R_4(R_1 + R_3)C_2}\frac{(R_1C_1s + 1)(R_2C_2s + 1)}{s\left(\dfrac{R_1R_3}{R_1 + R_3}C_1s + 1\right)}$$

$$= \frac{R_5R_2}{R_4R_3}\frac{\left(s + \dfrac{1}{R_1C_1}\right)\left(s + \dfrac{1}{R_2C_2}\right)}{s\left(s + \dfrac{R_1 + R_3}{R_1R_3C_1}\right)}$$

Notice that R_1C_1 and R_2C_2 determine the locations of the zeros of the controller, while R_1, R_3, and C_1 affect the location of the pole on the negative real axis. R_5/R_4 adjusts the gain of the controller.

A–10–4. In practice, it is impossible to realize the true differentiator. Hence, we always have to approximate the true differentiator T_ds by something like

$$\frac{T_ds}{1 + \gamma T_ds}$$

One way to realize such an approximate differentiator is to utilize an integrator in the feedback path. Show that the closed-loop transfer function of the system shown in Figure 10–40 is given by the preceding expression. (In the commercially available differentiator, the value of γ may be set as 0.1).

Solution. The closed-loop transfer function of the system shown in Figure 10–40 is

$$\frac{C(s)}{R(s)} = \frac{\dfrac{1}{\gamma}}{1 + \dfrac{1}{\gamma T_ds}} = \frac{T_ds}{1 + \gamma T_ds}$$

Note that such a differentiator with first-order delay reduces the bandwidth of the closed-loop control system and reduces the detrimental effect of noise signals.

A–10–5. Consider the system shown in Figure 10–41. This is a PID control of a second-order plant $G(s)$. Assume that disturbances $D(s)$ enter the system as shown in the diagram. It is assumed that the reference input $R(s)$ is normally held constant, and the response characteristics to disturbances are a very important consideration in this system.

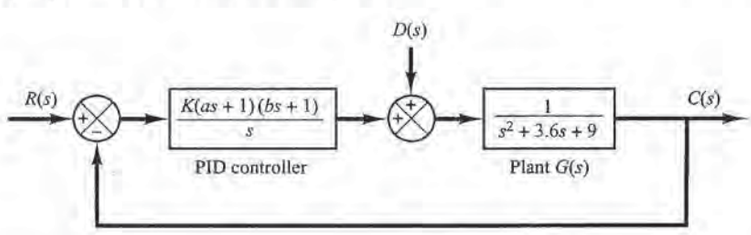


Figure 10–41
PID-controlled system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Design a control system such that the response to any step disturbance will be damped out quickly (in 2 to 3 sec in terms of the 2% settling time). Choose the configuration of the closed-loop poles such that there is a pair of dominant closed-loop poles. Then obtain the response to the unit-step disturbance input. Also, obtain the response to the unit-step reference input.

Solution. The PID controller has the transfer function

$$G_c(s) = \frac{K(as + 1)(bs + 1)}{s}$$

For the disturbance input in the absence of the reference input, the closed-loop transfer function becomes

$$\begin{aligned}\frac{C_d(s)}{D(s)} &= \frac{s}{s(s^2 + 3.6s + 9) + K(as + 1)(bs + 1)} \\ &= \frac{s}{s^3 + (3.6 + Kab)s^2 + (9 + Ka + Kb)s + K} \qquad (10\text{–}14)\end{aligned}$$

The specification requires that the response to the unit-step disturbance be such that the settling time be 2 to 3 sec and the system have a reasonable damping. We may interpret the specification as $\zeta = 0.5$ and $\omega_n = 4$ rad/sec for the dominant closed-loop poles. We may choose the third pole at $s = -10$ so that the effect of this real pole on the response is small. Then the desired characteristic equation can be written as

$$(s + 10)(s^2 + 2 \times 0.5 \times 4s + 4^2) = (s + 10)(s^2 + 4s + 16) = s^3 + 14s^2 + 56s + 160$$

The characteristic equation for the system given by Equation (10–14) is

$$s^3 + (3.6 + Kab)s^2 + (9 + Ka + Kb)s + K = 0$$

Hence, we require

$$\begin{aligned}3.6 + Kab &= 14 \\ 9 + Ka + Kb &= 56 \\ K &= 160\end{aligned}$$

which yields

$$ab = 0.065, \qquad a + b = 0.29375$$

The PID controller now becomes

$$\begin{aligned}G_c(s) &= \frac{K[abs^2 + (a + b)s + 1]}{s} \\ &= \frac{160(0.065s^2 + 0.29375s + 1)}{s} \\ &= \frac{10.4(s^2 + 4.5192s + 15.385)}{s}\end{aligned}$$

With this PID controller, the response to the disturbance is given by

$$\begin{aligned}C_d(s) &= \frac{s}{s^3 + 14s^2 + 56s + 160} D(s) \\ &= \frac{s}{(s + 10)(s^2 + 4s + 16)} D(s)\end{aligned}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Clearly, for a unit-step disturbance input, the steady-state output is zero, since

$$\lim_{t\to\infty} c_d(t) = \lim_{s\to 0} sC_d(s) = \lim_{s\to 0} \frac{s^2}{(s+10)(s^2+4s+16)}\frac{1}{s} = 0$$

The response to a unit-step disturbance input can be obtained easily with MATLAB. MATLAB Program 10–6 produces a response curve as shown in Figure 10–42(a). From the response curve, we see that the settling time is approximately 2.7 sec. The response damps out quickly. Therefore, the system designed here is acceptable.

MATLAB Program 10–6

```
% ***** Response to unit-step disturbance input *****

numd = [0  0  1  0];
dend = [1  14  56  160];
t = 0:0.01:5;
[c1,x1,t] = step(numd,dend,t);
plot(t,c1)
grid
title('Response to Unit-Step Disturbance Input')
xlabel('t Sec')
ylabel('Output to Disturbance Input')

% ***** Response to unit-step reference input *****

numr = [0  10.4  47  160];
denr = [1  14  56  160];
[c2,x2,t] = step(numr,denr,t);
plot(t,c2)
grid
title('Response to Unit-Step Reference Input')
xlabel('t Sec')
ylabel('Output to Reference Input')
```

For the reference input $r(t)$, the closed-loop transfer function is

$$\frac{C_r(s)}{R(s)} = \frac{10.4(s^2+4.5192s+15.385)}{s^3+14s^2+56s+160}$$

$$= \frac{10.4s^2+47s+160}{s^3+14s^2+56s+160}$$

The response to a unit-step reference input can also be obtained by use of MATLAB Program 10–6. The resulting response curve is shown in Figure 10–42(b). The response curve shows that the maximum overshoot is 7.3% and the settling time is 1.2 sec. The system has quite acceptable response characteristics.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

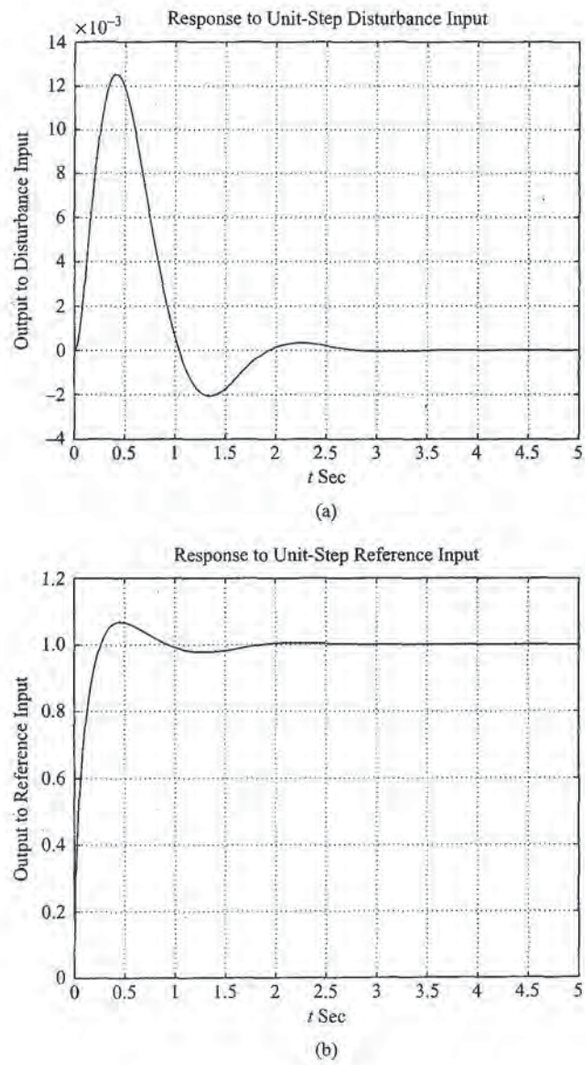


Figure 10–42
(a) Response to unit-step disturbance input; (b) response to unit-step reference input.

A–10–6. Consider the system shown in Figure 10–43. It is desired to design a PID controller $G_c(s)$ such that the dominant closed-loop poles are located at $s = -1 \pm j\sqrt{3}$. For the PID controller, choose $a = 1$ and then determine the values of K and b. Sketch the root-locus diagram for the designed system.

Solution. Since

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

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

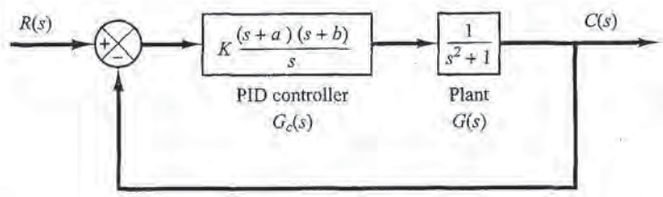


Figure 10–43
PID-controlled system.

the sum of the angles at $s = -1 + j\sqrt{3}$, one of the desired closed-loop poles, from the zero at $s = -1$ and poles at $s = 0$, $s = j$, and $s = -j$ is

$$90° - 143.794° - 120° - 110.104° = -283.898°$$

Hence the zero at $s = -b$ must contribute 103.898°. This requires that the zero be located at

$$b = 0.5714$$

The gain constant K can be determined from the magnitude condition.

$$\left| K \frac{(s+1)(s+0.5714)}{s} \frac{1}{s^2+1} \right|_{s=-1+j\sqrt{3}} = 1$$

or

$$K = 2.3333$$

Then the compensator can be written as follows:

$$G_c(s) = 2.3333 \frac{(s+1)(s+0.5714)}{s}$$

The open-loop transfer function becomes

$$G_c(s)G(s) = \frac{2.3333(s+1)(s+0.5714)}{s} \frac{1}{s^2+1}$$

From this equation a root-locus plot for the compensated system can be drawn. Figure 10–44 is a root-locus plot.

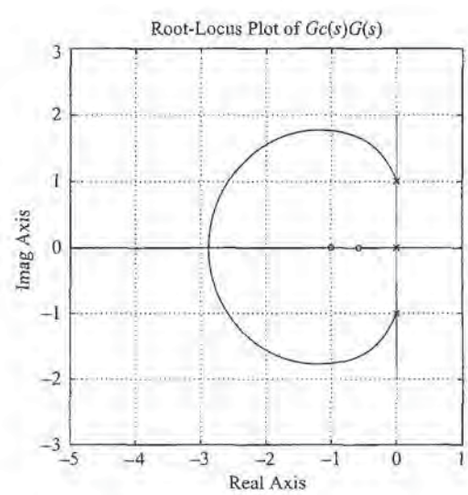


Figure 10–44
Root-locus plot of the compensated system (Problem **A–10–6**).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

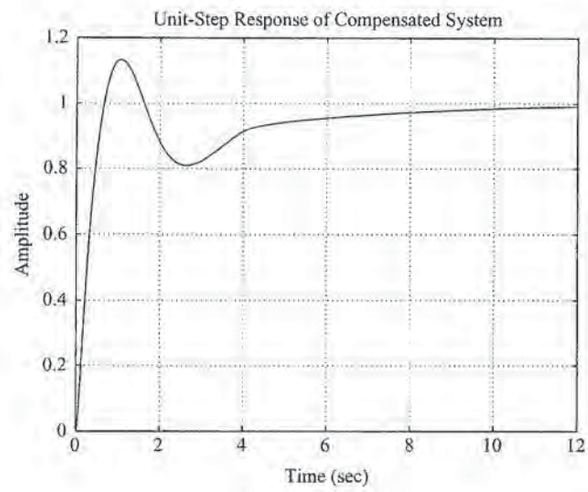


Figure 10–45
Unit-step response of the compensated system (Problem **A–10–6**).

The closed-loop transfer function is given by

$$\frac{C(s)}{R(s)} = \frac{2.3333(s+1)(s+0.5714)}{s^3 + s + 2.3333(s+1)(s+0.5714)}$$

The closed-loop poles are located at $s = -1 \pm j\sqrt{3}$ and $s = -0.3333$. A unit-step response curve is shown in Figure 10–45. The closed-loop pole at $s = -0.3333$ and a zero at $s = -0.5714$ produce a long tail of small amplitude.

A–10–7. Show that the I-PD-controlled system shown in Figure 10–46(a) is equivalent to the PID-controlled system with input filter shown in Figure 10–46(b).

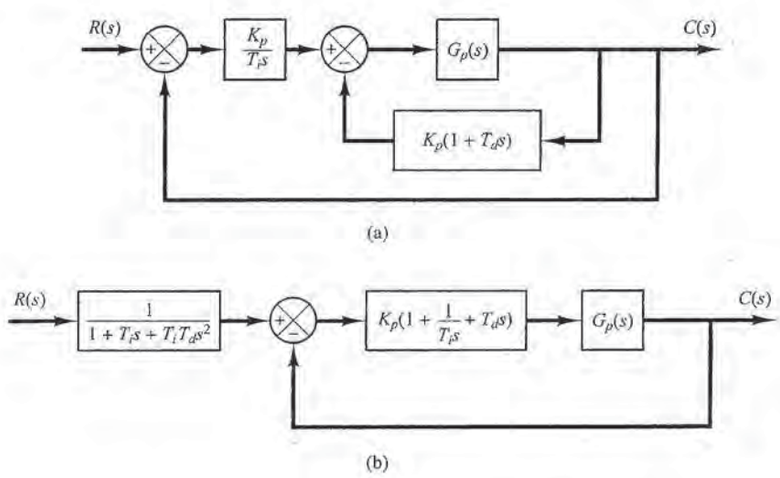


Figure 10–46
(a) I-PD-controlled system;
(b) PID-controlled system with input filter.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Solution. The closed-loop transfer function $C(s)/R(s)$ of the I-PD-controlled system is

$$\frac{C(s)}{R(s)} = \frac{\dfrac{K_p}{T_i s} G_p(s)}{1 + K_p\left(1 + \dfrac{1}{T_i s} + T_d s\right) G_p(s)}$$

The closed-loop transfer function $C(s)/R(s)$ of the PID-controlled system with input filter shown in Figure 10–46(b) is

$$\frac{C(s)}{R(s)} = \frac{1}{1 + T_i s + T_i T_d s^2} \frac{K_p\left(1 + \dfrac{1}{T_i s} + T_d s\right) G_p(s)}{1 + K_p\left(1 + \dfrac{1}{T_i s} + T_d s\right) G_p(s)}$$

$$= \frac{\dfrac{K_p}{T_i s} G_p(s)}{1 + K_p\left(1 + \dfrac{1}{T_i s} + T_d s\right) G_p(s)}$$

The closed-loop transfer functions of both systems are the same. Thus, the two systems are equivalent.

A–10–8. The basic idea of the I-PD control is to avoid large control signals (which will cause saturation phenomenon) within the system. By bringing the proportional and derivative control actions to the feedback path, it is possible to choose larger values for K_p and T_d than those possible by the PID control scheme.

Compare, qualitatively, the responses of the PID-controlled system and I-PD-controlled system to the disturbance input and to the reference input.

Solution. Consider first the response of the I-PD-controlled system to the disturbance input. Since, in the I-PD control of a plant, it is possible to select larger values for K_p and T_d than those of the PID-controlled case, the I-PD-controlled system will attenuate the effect of disturbance faster than the PID-controlled case.

Next, consider the response of the I-PD-controlled system to a reference input. Since the I-PD-controlled system is equivalent to the PID-controlled system with input filter (refer to Problem **A–10–7**), the PID-controlled system will have faster responses than the corresponding I-PD-controlled system, provided a saturation phenomenon does not occur in the PID-controlled system.

A–10–9. In some cases it is desirable to provide an input filter as shown in Figure 10–47(a). Notice that the input filter $G_f(s)$ is outside the loop. Therefore, it does not affect the stability of the closed-loop portion of the system. An advantage of having the input filter is that the zeros of the closed-loop transfer function can be modified (canceled or replaced by other zeros) so that the closed-loop response is acceptable.

Show that the configuration in Figure 10–47(a) can be modified to that shown in Figure 10–47(b), where $G_d(s) = [G_f(s) - 1]G_c(s)$. The compensation structure shown in Figure 10–47(b) is sometimes called command compensation.

Solution. For the system of Figure 10–47(a), we have

$$\frac{C(s)}{R(s)} = G_f(s) \frac{G_c(s)G_p(s)}{1 + G_c(s)G_p(s)} \tag{10–15}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

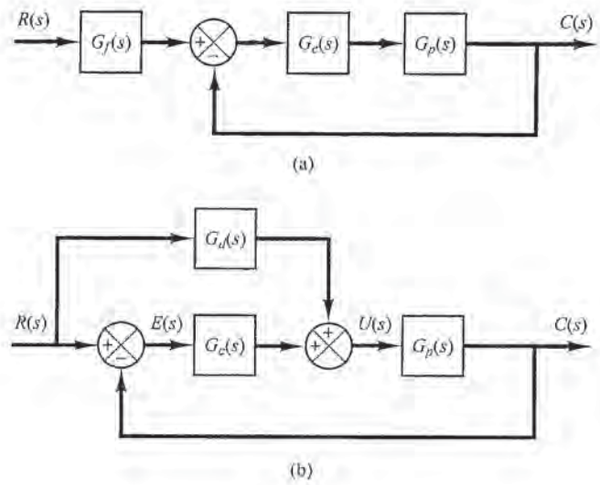


Figure 10–47
(a) Block diagram of control system with input filter;
(b) modified block diagram.

For the system of Figure 10–47(b), we have

$$U(s) = G_d(s)R(s) + G_c(s)E(s)$$

$$E(s) = R(s) - C(s)$$

$$C(s) = G_p(s)U(s)$$

Thus

$$C(s) = G_p(s)\{G_d(s)R(s) + G_c(s)[R(s) - C(s)]\}$$

or

$$\frac{C(s)}{R(s)} = \frac{[G_d(s) + G_c(s)]G_p(s)}{1 + G_c(s)G_p(s)} \tag{10–16}$$

By substituting $G_d(s) = [G_f(s) - 1]G_c(s)$ into Equation (10–16), we obtain

$$\frac{C(s)}{R(s)} = \frac{[G_f(s)G_c(s) - G_c(s) + G_c(s)]G_p(s)}{1 + G_c(s)G_p(s)}$$

$$= G_f(s)\frac{G_c(s)G_p(s)}{1 + G_c(s)G_p(s)}$$

which is the same as Equation (10–15). Hence, we have shown that the systems shown in Figures 10–47(a) and (b) are equivalent.

It is noted that the system shown in Figure 10–47(b) has a feedforward controller $G_d(s)$. In such a case, $G_d(s)$ does not affect the stability of the closed-loop portion of the system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–10–10. A closed-loop system has the characteristic that the closed-loop transfer function is nearly equal to the inverse of the feedback transfer function whenever the open-loop gain is much greater than unity.

The open-loop characteristic may be modified by adding an internal feedback loop with a characteristic equal to the inverse of the desired open-loop characteristic. Suppose that a unity-feedback system has the open-loop transfer function

$$G(s) = \frac{K}{(T_1 s + 1)(T_2 s + 1)}$$

Determine the transfer function $H(s)$ of the element in the internal feedback loop so that the inner loop becomes ineffective at both low and high frequencies.

Solution. Figure 10–48(a) shows the original system. Figure 10–48(b) shows the addition of the internal feedback loop around $G(s)$. Since

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

if the gain around the inner loop is large compared with unity, then $G(s)H(s)/[1 + G(s)H(s)]$ is approximately equal to unity, and the transfer function $C(s)/E(s)$ is approximately equal to $1/H(s)$.

On the other hand, if the gain $|G(s)H(s)|$ is much less than unity, the inner loop becomes ineffective and $C(s)/E(s)$ becomes approximately equal to $G(s)$.

To make the inner loop ineffective at both the low- and high-frequency ranges, we require that

$$|G(j\omega)H(j\omega)| \ll 1, \quad \text{for } \omega \ll 1 \text{ and } \omega \gg 1$$

Since, in this problem,

$$G(j\omega) = \frac{K}{(1 + j\omega T_1)(1 + j\omega T_2)}$$

the requirement can be satisfied if $H(s)$ is chosen to be

$$H(s) = ks$$

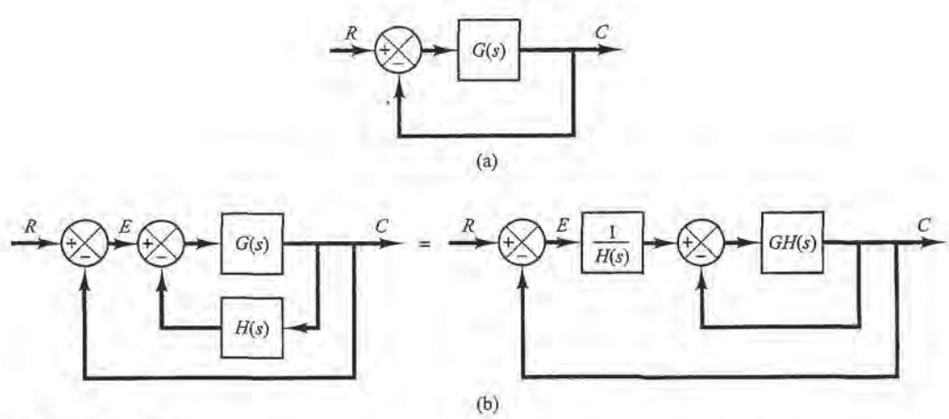


Figure 10–48
(a) Control system; (b) addition of the internal feedback loop to modify the closed-loop characteristic.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

because

$$\lim_{\omega \to 0} G(j\omega)H(j\omega) = \lim_{\omega \to 0} \frac{Kkj\omega}{(1 + j\omega T_1)(1 + j\omega T_2)} = 0$$

$$\lim_{\omega \to \infty} G(j\omega)H(j\omega) = \lim_{\omega \to \infty} \frac{Kkj\omega}{(1 + j\omega T_1)(1 + j\omega T_2)} = 0$$

Thus, with $H(s) = ks$ (velocity feedback), the inner loop becomes ineffective at both the low- and high-frequency regions. It becomes effective only in the intermediate-frequency region.

A–10–11. Consider the control system shown in Figure 10–49. This is the same system as that considered in Example 10–1. In that example we designed a PID controller $G_c(s)$, starting with the second method of the Ziegler–Nichols tuning rule. Here we design a PID controller using the computational approach with MATLAB. We shall determine the values of K and a of the PID controller

$$G_c(s) = K\frac{(s + a)^2}{s}$$

such that the unit-step response will exhibit the maximum overshoot between 10% and 2% ($1.02 \leq$ maximum output ≤ 1.10) and the settling time will be less than 3 sec. The search region is

$$2 \leq K \leq 50, \qquad 0.05 \leq a \leq 2$$

Let us choose the step size for K to be 1 and that for a to be 0.05.

Write a MATLAB program to find the first set of variables K and a that will satisfy the given specifications. Also, write a MATLAB program to find all possible sets of variables K and a that will satisfy the given specifications. Plot the unit-step response curves of the designed system with the chosen sets of variables K and a.

Solution. The transfer function of the plant is

$$G_p(s) = \frac{1}{s^3 + 6s^2 + 5s}$$

The closed-loop transfer function $C(s)/R(s)$ is given by

$$\frac{C(s)}{R(s)} = \frac{Ks^2 + 2Kas + Ka^2}{s^4 + 6s^3 + (5 + K)s^2 + 2Kas + Ka^2}$$

A possible MATLAB program that will produce the first set of variables K and a that will satisfy the given specifications is given in MATLAB Program 10–7. In this program we use two 'for' loops. The specification for the settling time is interpreted by the following four lines:

```
s = 501; while y(s) > 0.98 and y(s) < 1.02;
s = s - 1; end;
ts = (s - 1) * 0.01
ts < 3.0
```

Note that for $t = 0{:}0.01{:}5$, we have 501 computing time points. $s = 501$ corresponds to the last computing time point.

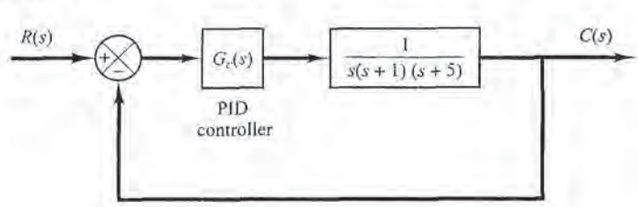


Figure 10–49
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The solution obtained by this program is

$$K = 32, \qquad a = 0.2$$

with the maximum overshoot equal to 9.69% and the settling time equal to 2.64 sec. The resulting unit-step response curve is shown in Figure 10–50.

MATLAB Program 10–7

```
t = 0:0.01:5;
for K = 50:-1:2;
   for a = 2:-0.05:0.05;
      num = [0  0  K  2*K*a  K*a^2];
      den = [1  6  5+K  2*K*a  K*a^2];
         y = step(num,den,t);
         m = max(y);
         s = 501; while y(s) > 0.98 & y(s) < 1.02;
         s = s-1; end;
         ts = (s-1)*0.01;
      if m < 1.10 & m > 1.02 & ts < 3.0
      break;
      end
      end
   if m < 1.10 & m > 1.02 & ts < 3.0
   break
   end
   end
plot(t,y)
grid
title('Unit-Step Response')
xlabel('t sec')
ylabel('Output')
solution = [K;a;m;ts]

solution =

32.0000
 0.2000
 1.0969
 2.6400
```

Next, we shall consider the case where we want to find all sets of variables that will satisfy the given specifications. A possible MATLAB program for this purpose is given in MATLAB Program 10–8. Note that in the table shown in the program, the last row should be ignored. (These are the last K and a values for searching purposes.)

From the sorttable, it seems that

$$K = 29,\ a = 0.25 \text{ (max overshoot} = 9.52\%, \text{settling time} = 1.78 \text{ sec)}$$

and

$$K = 27,\ a = 0.2 \text{ (max overshoot} = 5.5\%, \text{settling time} = 2.89 \text{ sec)}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

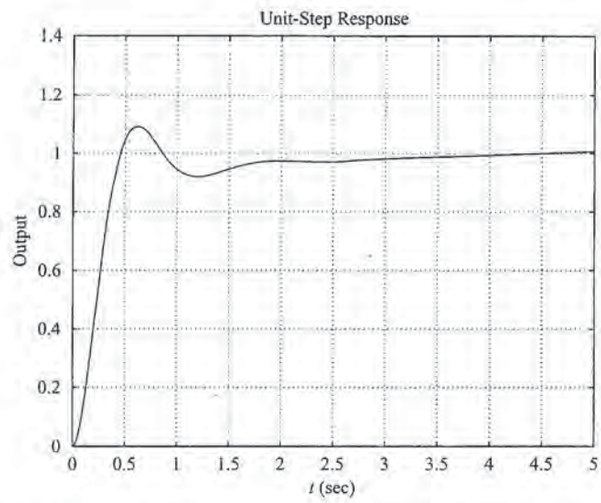


Figure 10–50
Unit-step response curve.

are two of the best choices. The unit-step response curves for these two cases are shown in Figure 10–51. From these curves, we might conclude that the best choice depends on the system objective. If a small maximum overshoot is desired, $K = 27$, $a = 0.2$ will be the best choice. If the shorter settling time is more important than a small maximum overshoot, then $K = 29$, $a = 0.25$ will be the best choice.

MATLAB Program 10–8

```
t = 0:0.01:5;
k = 0;
for i = 1:49;
   K(i) = 51-i*1;
      for j = 1:40;
      a(j) = 2.05-j*0.05;
      num = [0  0  K(i)  2*K(i)*a(j)  K(i)*a(j)*a(j)];
      den = [1  6  5+K(i)  2*K(i)*a(j)  K(i)*a(j)*a(j)];
         y = step(num,den,t);
         m = max(y);
         s = 501; while y(s) > 0.98 & y(s) < 1.02;
         s = s-1; end;
         ts = (s-1)*0.01;
      if m < 1.10 & m > 1.02 & ts < 3.0
      k = k+1;
      table(k,:) = [K(i)  a(j)  m  ts];
      end
   end
 end
table(k,:) = [K(i)  a(j)  m  ts]

table =
```

(*continues on next page*)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

```
   32.0000  0.2000  1.0969  2.6400
   31.0000  0.2000  1.0890  2.6900
   30.0000  0.2000  1.0809  2.7300
   29.0000  0.2500  1.0952  1.7800
   29.0000  0.2000  1.0726  2.7800
   28.0000  0.2000  1.0639  2.8300
   27.0000  0.2000  1.0550  2.8900
    2.0000  0.0500  0.3781  5.0000

sorttable = sortrows(table,3)

sorttable =

    2.0000  0.0500  0.3781  5.0000
   27.0000  0.2000  1.0550  2.8900
   28.0000  0.2000  1.0639  2.8300
   29.0000  0.2000  1.0726  2.7800
   30.0000  0.2000  1.0809  2.7300
   31.0000  0.2000  1.0890  2.6900
   29.0000  0.2500  1.0952  1.7800
   32.0000  0.2000  1.0969  2.6400

K = sorttable(7,1)

K =

   29

a = sorttable(7,2)

a=

   0.2500

num = [0  0  K  2*K*a  K*a^2];
den = [1  6  5+K  2*K*a  K*a^2];
y = step(num,den,t);
plot(t,y)
grid
hold
Current plot held
K = sorttable(2,1)

K=

   27

a = sorttable(2,2)

a=

   0.2000
```

(*continues on next page*)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

```
num = [0 0 K 2*K*a K*a^2];
den = [1 6 5+K 2*K*a K*a^2];
y = step(num,den,t);
plot(t,y)
title('Unit-Step Response Curves')
xlabel('t (sec)')
ylabel('Output')
text(1.22,1.22,'K = 29, a = 0.25')
text(1.22,0.72,'K = 27, a = 0.2')
```

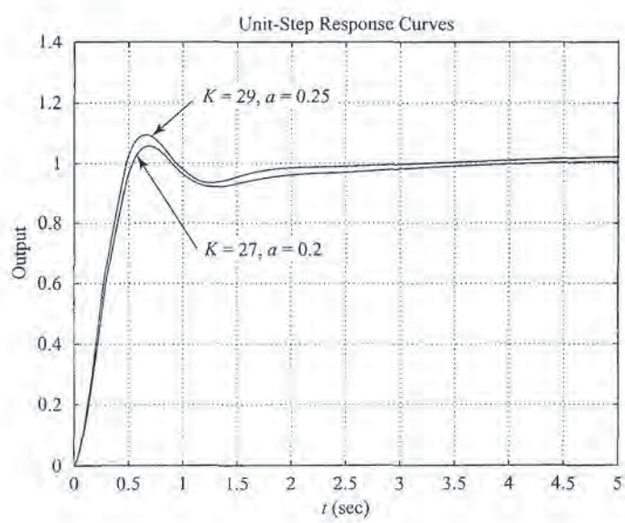


Figure 10–51
Unit-step response curves.

A–10–12. Consider the two-degrees-of-freedom control system shown in Figure 10–52. The plant $G_p(s)$ is given by

$$G_p(s) = \frac{100}{s(s+1)}$$

Assuming that the noise input $N(s)$ is zero, design controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that the designed system satisfies the following:

1. The response to the step disturbance input has a small amplitude and settles to zero quickly (on the order of 1 sec to 2 sec).
2. The response to the unit-step reference input has a maximum overshoot of 25% or less, and the settling time is 1 sec or less.
3. The steady-state errors in following ramp reference input and acceleration reference input are zero.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

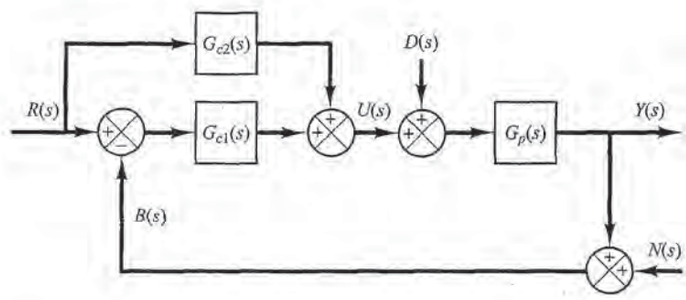


Figure 10–52
Two-degrees-of-freedom control system.

Solution. The closed-loop transfer functions for the disturbance input and reference input are given, respectively, by

$$\frac{Y(s)}{D(s)} = \frac{G_p(s)}{1 + G_{c1}(s)G_p(s)}$$

$$\frac{Y(s)}{R(s)} = \frac{[G_{c1}(s) + G_{c2}(s)]G_p(s)}{1 + G_{c1}(s)G_p(s)}$$

Let us assume that $G_{c1}(s)$ is a PID controller and has the following form:

$$G_{c1}(s) = \frac{K(s + a)^2}{s}$$

The characteristic equation for the system is

$$1 + G_{c1}(s)G_p(s) = 1 + \frac{K(s + a)^2}{s}\frac{100}{s(s + 1)}$$

Notice that the open-loop poles are located at $s = 0$ (a double pole) and $s = -1$. The zeros are located at $s = -a$ (a double zero).

In what follows, we shall use the root-locus approach to determine the values of a and K. Let us choose the dominant closed-loop poles at $s = -5 \pm j5$. Then, the angle deficiency at the desired closed-loop pole at $s = -5 + j5$ is

$$-135° - 135° - 128.66° + 180° = -218.66°$$

The double zero at $s = -a$ must contribute 218.66°. (Each zero must contribute 109.33°.) By a simple calculation, we find

$$a = -3.2460$$

The controller $G_{c1}(s)$ is then determined as

$$G_{c1}(s) = \frac{K(s + 3.2460)^2}{s}$$

The constant K must be determined by use of the magnitude condition. This condition is

$$|G_{c1}(s)G_p(s)|_{s=-5+j5} = 1$$

Since

$$G_{c1}(s)G_p(s) = \frac{K(s + 3.2460)^2}{s}\frac{100}{s(s + 1)}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

we obtain

$$K = \left| \frac{s^2(s+1)}{100(s+3.2460)^2} \right|_{s=-5+j5}$$
$$= 0.11403$$

The controller $G_{c1}(s)$ thus becomes

$$G_{c1}(s) = \frac{0.11403(s+3.2460)^2}{s}$$
$$= \frac{0.11403s^2 + 0.7403s + 1.20148}{s}$$
$$= 0.7403 + \frac{1.20148}{s} + 0.11403s \tag{10–17}$$

Then, the closed-loop transfer function $Y(s)/D(s)$ is obtained as follows:

$$\frac{Y(s)}{D(s)} = \frac{G_p(s)}{1 + G_{c1}(s)G_p(s)}$$
$$= \frac{\dfrac{100}{s(s+1)}}{1 + \dfrac{0.11403(s+3.2460)^2}{s}\dfrac{100}{s(s+1)}}$$
$$= \frac{100s}{s^3 + 12.403s^2 + 74.0284s + 120.148}$$

The response curve when $D(s)$ is a unit-step disturbance is shown in Figure 10–53.

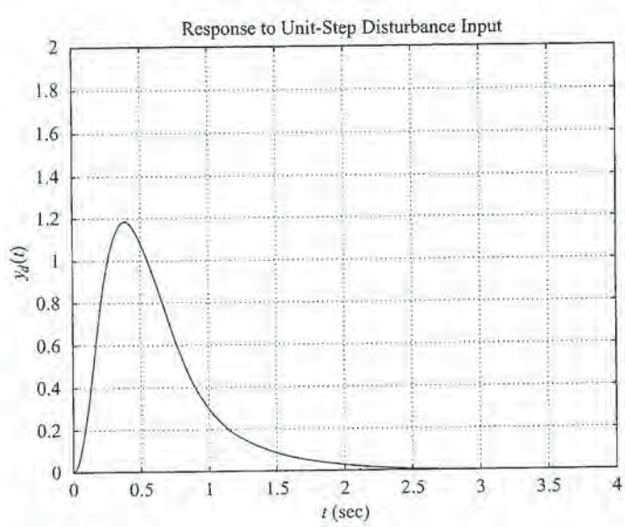


Figure 10–53
Response to unit-step disturbance input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Next, we consider the responses to reference inputs. The closed-loop transfer function $Y(s)/R(s)$ is

$$\frac{Y(s)}{R(s)} = \frac{[G_{c1}(s) + G_{c2}(s)]G_p(s)}{1 + G_{c1}(s)G_p(s)}$$

Let us define

$$G_{c1}(s) + G_{c2}(s) = G_c(s)$$

Then

$$\frac{Y(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1 + G_{c1}(s)G_p(s)}$$
$$= \frac{100sG_c(s)}{s^3 + 12.403s^2 + 74.0284s + 120.148}$$

To satisfy the requirements on the responses to the ramp reference input and acceleration reference input, we use the zero-placement approach. That is, we choose the numerator of $Y(s)/R(s)$ to be the sum of the last three terms of the denominator, or

$$100sG_c(s) = 12.403s^2 + 74.0284s + 120.148$$

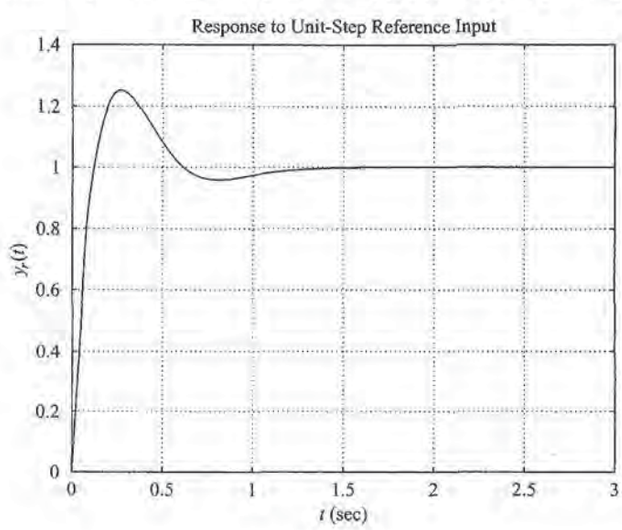
from which we get

$$G_c(s) = \frac{0.12403s^2 + 0.740284s + 1.20148}{s}$$
$$= 0.7403 + \frac{1.20148}{s} + 0.12403s \tag{10–18}$$

Hence, the closed-loop transfer function $Y(s)/R(s)$ becomes as

$$\frac{Y(s)}{R(s)} = \frac{12.403s^2 + 74.0284s + 120.148}{s^3 + 12.403s^2 + 74.0284s + 120.148}$$

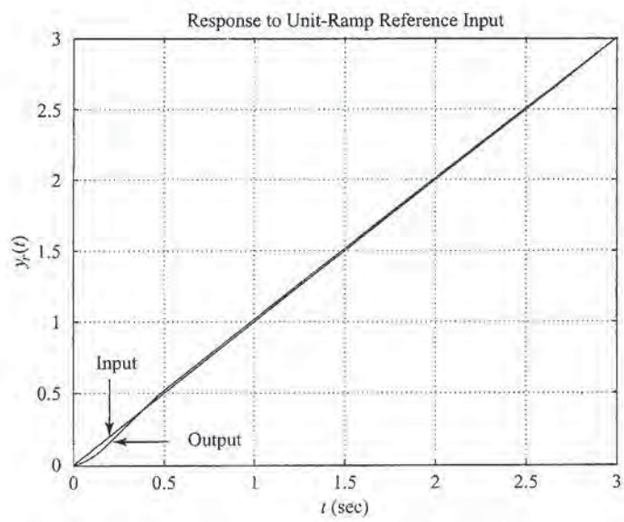
The response curves to the unit-step reference input, unit-ramp reference input, and unit-acceleration reference input are shown in Figures 10–54(a), (b), and (c), respectively. The maximum



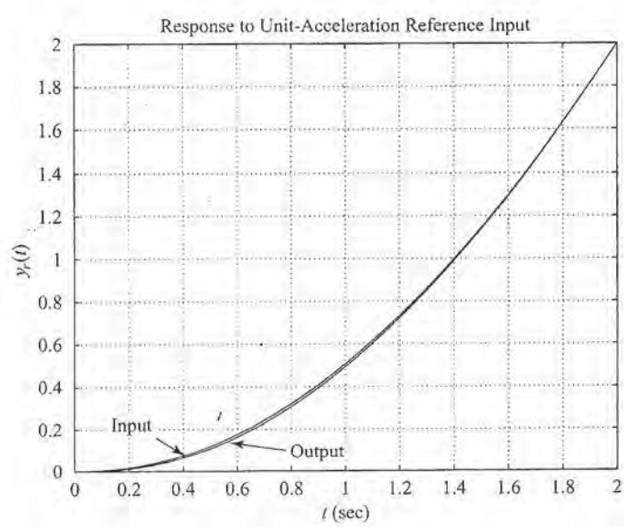
(a)

Figure 10–54
(a) Response to unit-step reference input; (b) response to unit-ramp reference input; (c) response to unit-acceleration reference input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565



(b)



(c)

Figure 10–54
(continued)

overshoot in the unit-step response is approximately 25% and the settling time is approximately 1.2 sec. The steady-state errors in the ramp response and acceleration response are zero. Therefore, the designed controller $G_c(s)$ given by Equation (10–18) is satisfactory.

Finally, we determine $G_{c2}(s)$. Noting that

$$G_{c2}(s) = G_c(s) - G_{c1}(s)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

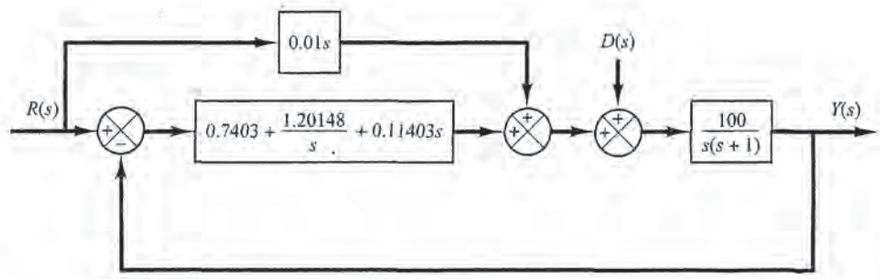


Figure 10–55
Block diagram of the designed system.

and from Equation (10–17)

$$G_{c1}(s) = 0.7403 + \frac{1.20148}{s} + 0.11403s$$

we obtain

$$\begin{aligned} G_{c2}(s) &= \left(0.7403 + \frac{1.20148}{s} + 0.12403s\right) \\ &\quad - \left(0.7403 + \frac{1.20148}{s} + 0.11403s\right) \\ &= 0.01s \end{aligned} \tag{10–19}$$

Equations (10–17) and (10–19) give the transfer functions of the controllers $G_{c1}(s)$ and $G_{c2}(s)$, respectively. The block diagram of the designed system is shown in Figure 10–55.

Note that if the maximum overshoot were much higher than 25% and/or the settling time were much larger than 1.2 sec, then we might assume a search region (such as $3 \le a \le 6$, $3 \le b \le 6$, and $6 \le c \le 12$) and use the computational method presented in Example 10–4 to find a set or sets of variables that would give the desired response to the unit-step reference input.

PROBLEMS

B–10–1. Consider the electronic PID controller shown in Figure 10–56. Determine the values of R_1, R_2, R_3, R_4, C_1, and C_2 of the controller such that the transfer function $G_c(s) = E_o(s)/E_i(s)$ is

$$\begin{aligned} G_c(s) &= 39.42\left(1 + \frac{1}{3.077s} + 0.7692s\right) \\ &= 30.3215\,\frac{(s + 0.65)^2}{s} \end{aligned}$$

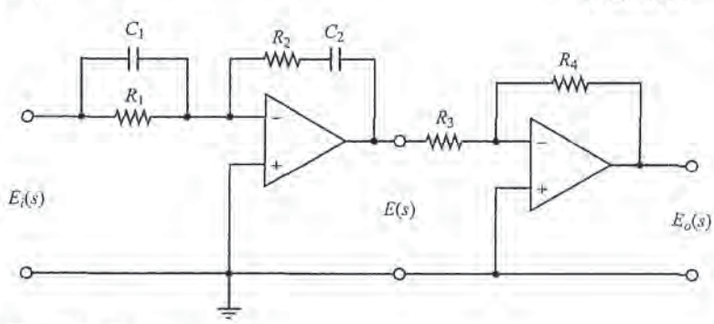


Figure 10–56
Electronic PID controller.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–10–2. Consider the system shown in Figure 10–57. Assume that disturbances $D(s)$ enter the system as shown in the diagram. Determine parameters K, a, and b such that the response to the unit-step disturbance input and the response to the unit-step reference input satisfy the following specifications: The response to the step disturbance input should attenuate rapidly with no steady-state error, and the response to the step reference input exhibits a maximum overshoot of 20% or less and a settling time of 2 sec.

B–10–3. Show that the PID-controlled system shown in Figure 10–58(a) is equivalent to the I-PD-controlled system with feedforward control shown in Figure 10–58(b).

B–10–4. Consider the systems shown in Figures 10–59(a) and (b). The system shown in Figure 10–59(a) is the system designed in Example 10–1. The response to the unit-step reference input in the absence of the disturbance input is shown in Figure 10–10. The system shown in Figure 10–59(b) is the I-PD-controlled system using the same K_p, T_i, and T_d as the system shown in Figure 10–59(a).

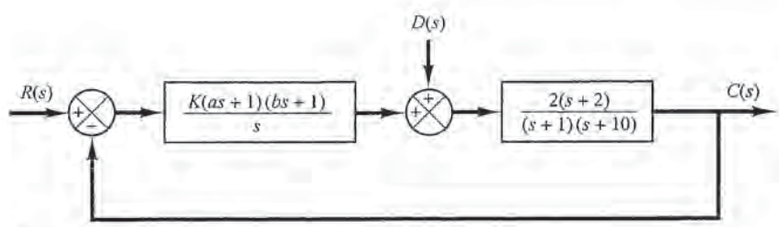


Figure 10–57
Control system.

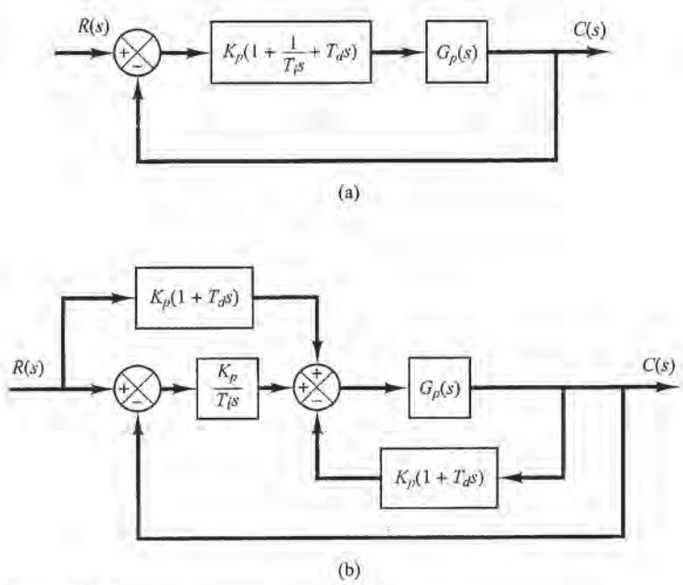


Figure 10–58
(a) PID-controlled system; (b) I-PD-controlled system with feedforward control.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Obtain the response of the I-PD-controlled system to the unit-step reference input with MATLAB. Compare the unit-step response curves of the two systems.

B–10–5. Referring to Problem B–10–4, obtain the response of the PID-controlled system shown in Figure 10–59(a) to the unit-step disturbance input.

Show that for the disturbance input, the responses of the PID-controlled system shown in Figure 10–59(a) and of the I-PD-controlled system shown in Figure 10–59(b) are exactly the same. [When considering $D(s)$ to be the input, assume that the reference input $R(s)$ is zero, and vice versa.] Also, compare the closed-loop transfer function $C(s)/R(s)$ of both systems.

B–10–6. Consider the system shown in Figure 10–60. This system is subjected to three input signals: the reference input, disturbance input, and noise input. Show that the characteristic equation of this system is the same regardless of which input signal is chosen as input.

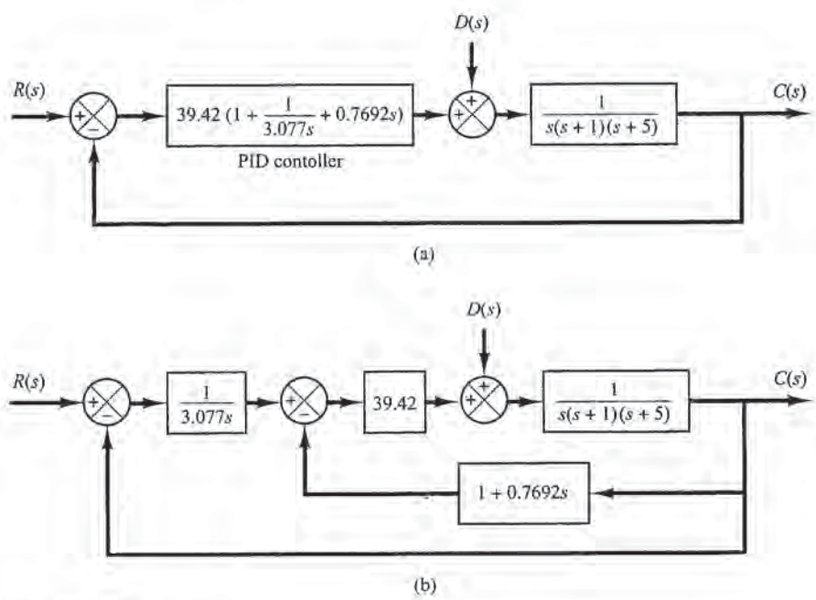


Figure 10–59
(a) PID-controlled system; (b) I-PD-controlled system.

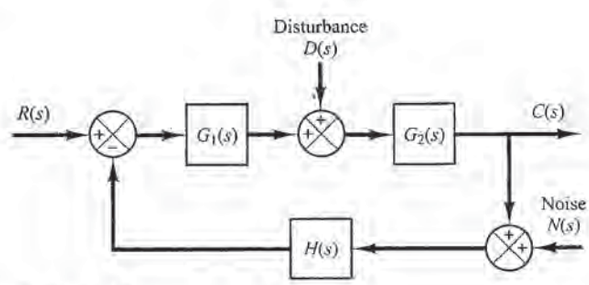


Figure 10–60
Control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–10–7. Consider the system shown in Figure 10–61. Obtain the closed-loop transfer function $C(s)/R(s)$ for the reference input and the closed-loop transfer function $C(s)/D(s)$ for the disturbance input. When considering $R(s)$ as the input, assume that $D(s)$ is zero, and vice versa.

B–10–8. Consider the system shown in Figure 10–62(a), where K is an adjustable gain and $G(s)$ and $H(s)$ are fixed components. The closed-loop transfer function for the disturbance is

$$\frac{C(s)}{D(s)} = \frac{1}{1 + KG(s)H(s)}$$

To minimize the effect of disturbances, the adjustable gain K should be chosen as large as possible.

Is this true for the system in Figure 10–62(b), too?

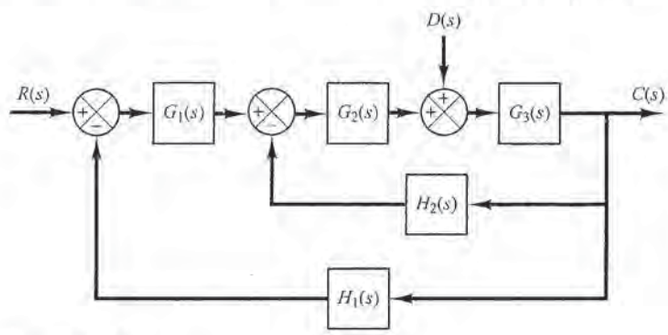


Figure 10–61
Control system.

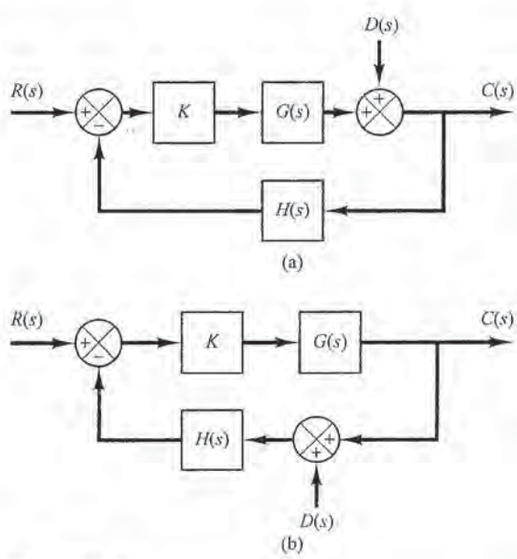


Figure 10–62
(a) Control system with disturbance entering in the feedforward path; (b) control system with disturbance entering in the feedback path.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–10–9. Show that the control systems shown in Figures 10–63(a), (b), and (c) are two-degrees-of-freedom systems. In the diagrams, G_{c1} and G_{c2} are controllers and G_p is the plant.

B–10–10. Show that the control system shown in Figure 10–64 is a three-degrees-of freedom system. The transfer functions G_{c1}, G_{c2}, and G_{c3} are controllers. The plant consists of transfer functions G_1 and G_2.

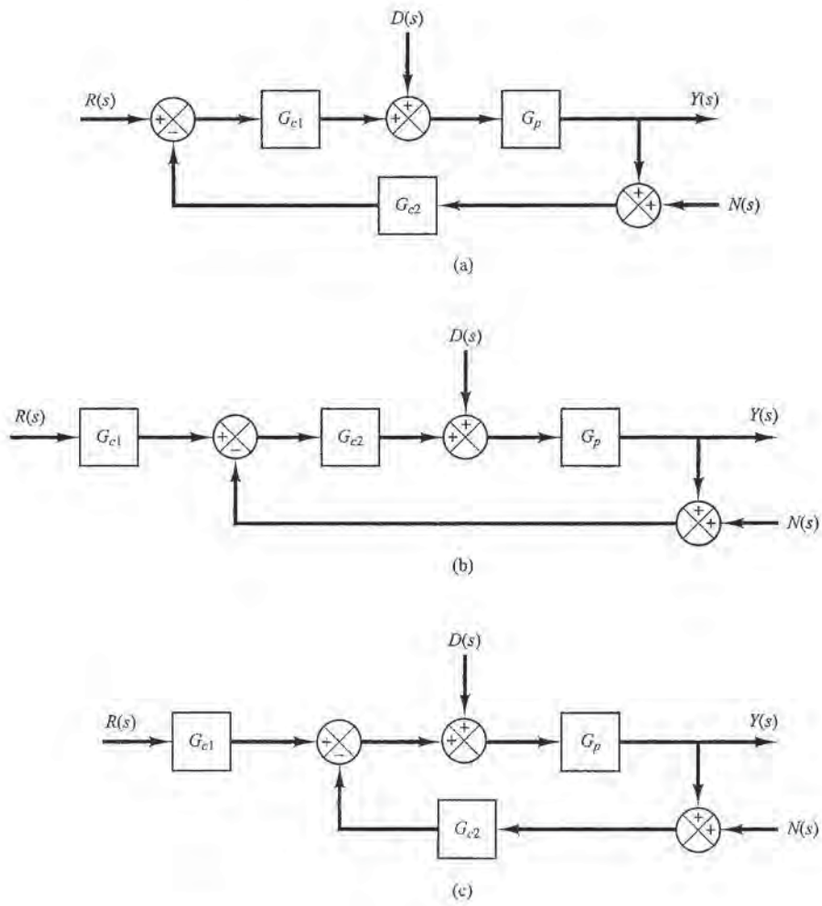


Figure 10–63
(a), (b), (c) Two degrees-of-freedom systems.

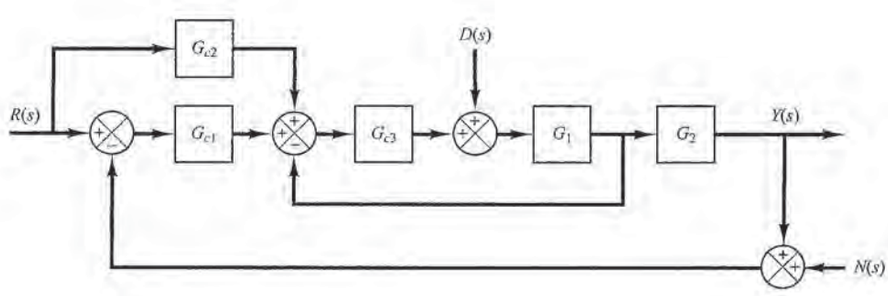


Figure 10–64
Three-degrees-of-freedom system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–10–11. Consider the control system shown in Figure 10–65. Assume that the PID controller is given by

$$G_c(s) = K\frac{(s+a)^2}{s}$$

It is desired that the unit-step response of the system exhibit the maximum overshoot of less than 10%, but more than 2% (to avoid an almost overdamped system), and the settling time be less than 2 sec.

Using the computational approach presented in Section 10–3, write a MATLAB program to determine the values of K and a that will satisfy the given specifications. Choose the search region to be

$$1 \le K \le 4, \qquad 0.4 \le a \le 4$$

Choose the step size for K and a to be 0.05. Write the program such that the nested loops start with the highest values of K and a and step toward the lowest.

Using the first-found solution, plot the unit-step response curve.

B–10–12. Consider the same control system as treated in Problem B–10–11 (Figure 10–65). The PID controller is given by

$$G_c(s) = K\frac{(s+a)^2}{s}$$

It is desired to determine the values of K and a such that the unit-step response of the system exhibits the maximum overshoot of less than 8%, but more than 3%, and the settling time is less than 2 sec. Choose the search region to be

$$2 \le K \le 4, \qquad 0.5 \le a \le 3$$

Choose the step size for K and a to be 0.05.

First, write a MATLAB program such that the nested loops in the program start with the highest values of K and a and step toward the lowest and the computation stop when a successful set of K and a is found for the first time.

Next, write a MATLAB program that will find all possible sets of K and a that will satisfy the given specifications.

Among multiple sets of K and a that satisfy the given specifications, determine the best choice. Then, plot the unit-step response curves of the system with the best choice of K and a.

B–10–13. Consider the two-degrees-of-freedom control system shown in Figure 10–66. The plant $G_p(s)$ is given by

$$G_p(s) = \frac{3(s+5)}{s(s+1)(s^2+4s+13)}$$

Design controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that the response to the unit-step disturbance input should have small amplitude and settle to zero quickly (in approximately 2 sec). The response to the unit-step reference input should be such that the maximum overshoot is 25% (or less) and the settling time is 2 sec. Also, the steady-state errors in the response to the ramp and acceleration reference inputs should be zero.

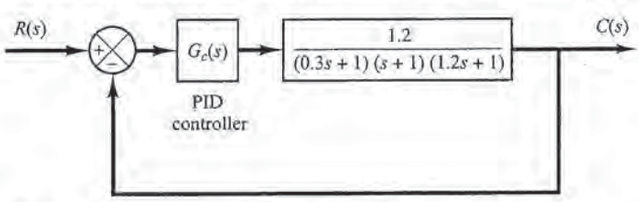


Figure 10–65
Control system.

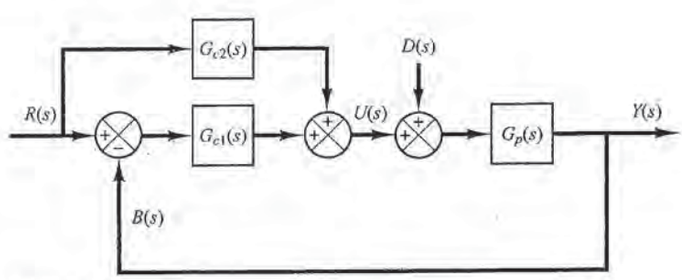


Figure 10–66
Two-degrees-of-freedom control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

B–10–14. Consider the system shown in Figure 10–67. The plant $G_p(s)$ is given by

$$G_p(s) = \frac{2(s+1)}{s(s+3)(s+5)}$$

Determine the controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that, for the step disturbance input, the response shows a small amplitude and approaches zero quickly (in a matter of 1 to 2 sec). For the response to the unit-step reference input, it is desired that the maximum overshoot be 20% or less and the settling time 1 sec or less. For the ramp reference input and acceleration reference input, the steady-state errors should be zero.

B–10–15. Consider the two-degrees-of-freedom control system shown in Figure 10–68. Design controllers $G_{c1}(s)$ and $G_{c2}(s)$ such that the response to the step disturbance input shows a small amplitude and settles to zero quickly (in 1 to 2 sec) and the response to the step reference input exhibits 25% or less maximum overshoot and the settling time be less than 1 sec. The steady-state error in following the ramp reference input or acceleration reference input should be zero.

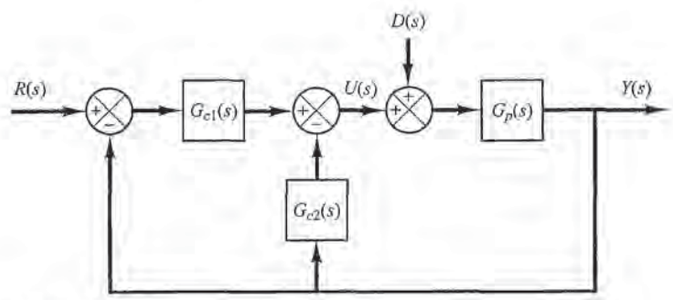


Figure 10–67
Two-degrees-of-freedom control system.

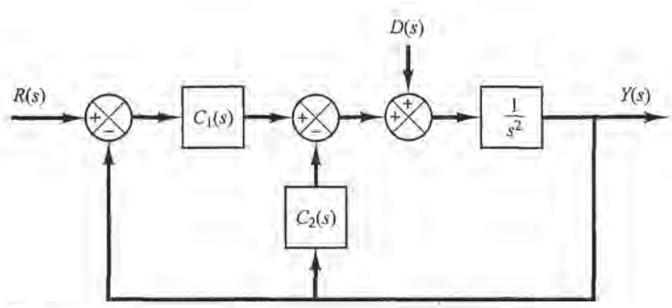


Figure 10–68
Two-degrees-of-freedom control system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

11

Analysis of Control Systems in State Space

11-1 INTRODUCTION*

A modern complex system may have many inputs and many outputs, and these may be interrelated in a complicated manner. To analyze such a system, it is essential to reduce the complexity of the mathematical expressions, as well as to resort to computers for most of the tedious computations necessary in the analysis. The state-space approach to system analysis is best suited from this viewpoint.

While conventional control theory is based on the input–output relationship, or transfer function, modern control theory is based on the description of system equations in terms of n first-order differential equations, which may be combined into a first-order vector-matrix differential equation. The use of vector-matrix notation greatly simplifies the mathematical representation of systems of equations. The increase in the number of state variables, the number of inputs, or the number of outputs does not increase the complexity of the equations. In fact, the analysis of complicated multiple-input–multiple-output systems can be carried out by the procedures that are only slightly more complicated than those required for the analysis of systems of first-order scalar differential equations.

This chapter and the next deal with the state-space analysis and design of control systems. Basic materials of state-space analysis, including the state-space representation of

*It is noted that in this book an asterisk used as a superscript of matrix, such as $\mathbf{A}^*$, implies that it is a **conjugate transpose** of matrix $\mathbf{A}$. The conjugate transpose is the conjugate of the transpose of a matrix. For a real matrix (a matrix whose elements are all real), the conjugate transpose $\mathbf{A}^*$ is the same as the transpose $\mathbf{A}^T$.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

systems, controllability, and observability are presented in this chapter. Basic design methods based on state-feedback control are given in Chapter 12.

Outline of the Chapter. Section 11–1 has presented an introduction to state-space analysis of control systems. Section 11–2 deals with the state-space representation of transfer-function systems. Here we present various canonical forms of state-space equations. Section 11–3 discusses the transformation of system models (such as from transfer function to state-space models, and vice versa) with MATLAB. Section 11–4 presents the solution of time-invariant state equations. Section 11–5 gives some useful results in vector-matrix analysis that are necessary in studying the state-space analysis of control systems. Section 11–6 discusses the controllability of control systems and Section 11–7 treats the observability of control systems.

11–2 STATE-SPACE REPRESENTATIONS OF TRANSFER-FUNCTION SYSTEMS

Many techniques are available for obtaining state-space representations of transfer-function systems. In Chapter 3 we presented a few such methods. This section presents state-space representations in the controllable, observable, diagonal, or Jordan canonical form. (Methods for obtaining such state-space representations from transfer functions are discussed in Problems **A–11–1** through **A–11–4**.)

State-Space Representation in Canonical Forms. Consider a system defined by

$$\overset{(n)}{y} + a_1 \overset{(n-1)}{y} + \cdots + a_{n-1}\dot{y} + a_n y = b_0 \overset{(n)}{u} + b_1 \overset{(n-1)}{u} + \cdots + b_{n-1}\dot{u} + b_n u \tag{11–1}$$

where u is the input and y is the output. This equation can also be written as

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \tag{11–2}$$

In what follows we shall present state-space representations of the system defined by Equation (11–1) or (11–2) in controllable canonical form, observable canonical form, and diagonal (or Jordan) canonical form.

Controllable Canonical Form. The following state-space representation is called a controllable canonical form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} u \tag{11–3}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$y = \left[b_n - a_n b_0 \;\vdots\; b_{n-1} - a_{n-1} b_0 \;\vdots\; \cdots \;\vdots\; b_1 - a_1 b_0\right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u \qquad (11\text{–}4)$$

The controllable canonical form is important in discussing the pole-placement approach to the control systems design.

Observable Canonical Form. The following state-space representation is called an observable canonical form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} b_n - a_n b_0 \\ b_{n-1} - a_{n-1} b_0 \\ \cdot \\ \cdot \\ \cdot \\ b_1 - a_1 b_0 \end{bmatrix} u \qquad (11\text{–}5)$$

$$y = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + b_0 u \qquad (11\text{–}6)$$

Note that the $n \times n$ state matrix of the state equation given by Equation (11–5) is the transpose of that of the state equation defined by Equation (11–3).

Diagonal Canonical Form. Consider the transfer function system defined by Equation (11–2). Here we consider the case where the denominator polynomial involves only distinct roots. For the distinct-roots case, Equation (11–2) can be written as

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{(s + p_1)(s + p_2)\cdots(s + p_n)}$$

$$= b_0 + \frac{c_1}{s + p_1} + \frac{c_2}{s + p_2} + \cdots + \frac{c_n}{s + p_n} \qquad (11\text{–}7)$$

The diagonal canonical form of the state-space representation of this system is given by

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} -p_1 & & & & 0 \\ & -p_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & -p_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} u \tag{11-8}$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u \tag{11-9}$$

Jordan Canonical Form. Next we shall consider the case where the denominator polynomial of Equation (11–2) involves multiple roots. For this case, the preceding diagonal canonical form must be modified into the Jordan canonical form. Suppose, for example, that the p_i's are different from one another, except that the first three p_i's are equal, or $p_1 = p_2 = p_3$. Then the factored form of $Y(s)/U(s)$ becomes

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1} s + b_n}{(s + p_1)^3 (s + p_4)(s + p_5) \cdots (s + p_n)}$$

The partial-fraction expansion of this last equation becomes

$$\frac{Y(s)}{U(s)} = b_0 + \frac{c_1}{(s + p_1)^3} + \frac{c_2}{(s + p_1)^2} + \frac{c_3}{s + p_1} + \frac{c_4}{s + p_4} + \cdots + \frac{c_n}{s + p_n}$$

A state-space representation of this system in the Jordan canonical form is given by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \left[\begin{array}{ccc|ccc} -p_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -p_1 & 1 & \vdots & & \vdots \\ 0 & 0 & -p_1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & -p_4 & & 0 \\ \cdot & & \cdot & & \cdot & \\ \cdot & & \cdot & & \cdot & \\ \cdot & & \cdot & & & \cdot \\ 0 & \cdots & 0 & 0 & & -p_n \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} u \tag{11-10}$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u \tag{11-11}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11-1 Consider the system given by

$$\frac{Y(s)}{U(s)} = \frac{s+3}{s^2+3s+2}$$

Obtain state-space representations in the controllable canonical form, observable canonical form, and diagonal canonical form.

Controllable Canonical Form:

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

$$y(t) = \begin{bmatrix} 3 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

Observable Canonical Form:

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

$$y(t) = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

Diagonal Canonical Form:

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

$$y(t) = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}$$

Eigenvalues of an $n \times n$ Matrix A. The eigenvalues of an $n \times n$ matrix $\mathbf{A}$ are the roots of the characteristic equation

$$|\lambda \mathbf{I} - \mathbf{A}| = 0$$

The eigenvalues are also called the characteristic roots.

Consider, for example, the following matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}$$

The characteristic equation is

$$\begin{aligned} |\lambda \mathbf{I} - \mathbf{A}| &= \begin{vmatrix} \lambda & -1 & 0 \\ 0 & \lambda & -1 \\ 6 & 11 & \lambda + 6 \end{vmatrix} \\ &= \lambda^3 + 6\lambda^2 + 11\lambda + 6 \\ &= (\lambda + 1)(\lambda + 2)(\lambda + 3) = 0 \end{aligned}$$

The eigenvalues of $\mathbf{A}$ are the roots of the characteristic equation, or -1, -2, and -3.

Diagonalization of $n \times n$ Matrix. Note that if an $n \times n$ matrix $\mathbf{A}$ with distinct eigenvalues is given by

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \tag{11–12}$$

the transformation $\mathbf{x} = \mathbf{P}\mathbf{z}$, where

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_n^2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{bmatrix}$$

$$\lambda_1, \lambda_2, \ldots, \lambda_n = n \text{ distinct eigenvalues of } \mathbf{A}$$

will transform $\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ into the diagonal matrix, or

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{bmatrix} \lambda_1 & & & & 0 \\ & \lambda_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & \lambda_n \end{bmatrix}$$

If the matrix $\mathbf{A}$ defined by Equation (11–12) involves multiple eigenvalues, then diagonalization is impossible. For example, if the 3×3 matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}$$

has the eigenvalues $\lambda_1, \lambda_1, \lambda_3$, then the transformation $\mathbf{x} = \mathbf{S}\mathbf{z}$, where

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 1 \\ \lambda_1 & 1 & \lambda_3 \\ \lambda_1^2 & 2\lambda_1 & \lambda_3^2 \end{bmatrix}$$

will yield

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

This is in the Jordan canonical form.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11–2 Consider the following state-space representation of a system.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} u \tag{11–13}$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{11–14}$$

Equations (11–13) and (11–14) can be put in a standard form as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{11–15}$$

$$y = \mathbf{C}\mathbf{x} \tag{11–16}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

The eigenvalues of matrix $\mathbf{A}$ are

$$\lambda_1 = -1, \qquad \lambda_2 = -2, \qquad \lambda_3 = -3$$

Thus, three eigenvalues are distinct. If we define a set of new state variables z_1, z_2, and z_3 by the transformation

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix}$$

or

$$\mathbf{x} = \mathbf{P}\mathbf{z} \tag{11–17}$$

where

$$\mathbf{P} = \begin{bmatrix} 1 & 1 & 1 \\ \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1^2 & \lambda_2^2 & \lambda_3^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix} \tag{11–18}$$

then, by substituting Equation (11–17) into Equation (11–15), we obtain

$$\mathbf{P}\dot{\mathbf{z}} = \mathbf{A}\mathbf{P}\mathbf{z} + \mathbf{B}u$$

By premultiplying both sides of this last equation by $\mathbf{P}^{-1}$, we get

$$\dot{\mathbf{z}} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{z} + \mathbf{P}^{-1}\mathbf{B}u \tag{11–19}$$

or

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} 3 & 2.5 & 0.5 \\ -3 & -4 & -1 \\ 1 & 1.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 3 & 2.5 & 0.5 \\ -3 & -4 & -1 \\ 1 & 1.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 6 \end{bmatrix} u$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Simplifying gives

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix} u \tag{11–20}$$

Equation (11–20) is also a state equation that describes the same system as defined by Equation (11–13).

The output equation, Equation (11–16), is modified to

$$y = \mathbf{CPz}$$

or

$$\begin{aligned} y &= [1 \quad 0 \quad 0] \begin{bmatrix} 1 & 1 & 1 \\ -1 & -2 & -3 \\ 1 & 4 & 9 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \\ &= [1 \quad 1 \quad 1] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \end{aligned} \tag{11–21}$$

Notice that the transformation matrix $\mathbf{P}$, defined by Equation (11–18), modifies the coefficient matrix of $\mathbf{z}$ into the diagonal matrix. As is clearly seen from Equation (11–20), the three scalar state equations are uncoupled. Notice also that the diagonal elements of the matrix $\mathbf{P}^{-1}\mathbf{AP}$ in Equation (11–19) are identical with the three eigenvalues of $\mathbf{A}$. It is very important to note that the eigenvalues of $\mathbf{A}$ and those of $\mathbf{P}^{-1}\mathbf{AP}$ are identical. We shall prove this for a general case in what follows.

Invariance of Eigenvalues. To prove the invariance of the eigenvalues under a linear transformation, we must show that the characteristic polynomials $|\lambda\mathbf{I} - \mathbf{A}|$ and $|\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{AP}|$ are identical.

Since the determinant of a product is the product of the determinants, we obtain

$$\begin{aligned} |\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{AP}| &= |\lambda\mathbf{P}^{-1}\mathbf{P} - \mathbf{P}^{-1}\mathbf{AP}| \\ &= |\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}| \\ &= |\mathbf{P}^{-1}||\lambda\mathbf{I} - \mathbf{A}||\mathbf{P}| \\ &= |\mathbf{P}^{-1}||\mathbf{P}||\lambda\mathbf{I} - \mathbf{A}| \end{aligned}$$

Noting that the product of the determinants $|\mathbf{P}^{-1}|$ and $|\mathbf{P}|$ is the determinant of the product $|\mathbf{P}^{-1}\mathbf{P}|$, we obtain

$$\begin{aligned} |\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{AP}| &= |\mathbf{P}^{-1}\mathbf{P}||\lambda\mathbf{I} - \mathbf{A}| \\ &= |\lambda\mathbf{I} - \mathbf{A}| \end{aligned}$$

Thus, we have proved that the eigenvalues of $\mathbf{A}$ are invariant under a linear transformation.

Nonuniqueness of a Set of State Variables. It has been stated that a set of state variables is not unique for a given system. Suppose that $x_1, x_2, \ldots, x_n$ are a set of state variables. Then we may take as another set of state variables any set of functions

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$\hat{x}_1 = X_1(x_1, x_2, \ldots, x_n)$$
$$\hat{x}_2 = X_2(x_1, x_2, \ldots, x_n)$$
$$\vdots$$
$$\hat{x}_n = X_n(x_1, x_2, \ldots, x_n)$$

provided that, for every set of values $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n$, there corresponds a unique set of values $x_1, x_2, \ldots, x_n$, and vice versa. Thus, if $\mathbf{x}$ is a state vector, then $\hat{\mathbf{x}}$, where

$$\hat{\mathbf{x}} = \mathbf{P}\mathbf{x}$$

is also a state vector, provided the matrix $\mathbf{P}$ is nonsingular. Different state vectors convey the same information about the system behavior.

11-3 TRANSFORMATION OF SYSTEM MODELS WITH MATLAB

In this section we shall consider the transformation of the system model from transfer function to state space, and vice versa. We shall begin our discussion with the transformation from transfer function to state space.

Let us write the closed-loop transfer function as

$$\frac{Y(s)}{U(s)} = \frac{\text{numerator polynomial in } s}{\text{denominator polynomial in } s} = \frac{\text{num}}{\text{den}}$$

Once we have this transfer-function expression, the MATLAB command

```
[A, B, C, D] = tf2ss(num,den)
```

will give a state-space representation. It is important to note that the state-space representation for any system is not unique. There are many (indeed, infinitely many) state-space representations for the same system. The MATLAB command gives one possible such state-space representation.

State-Space Formulation of Transfer-Function Systems. Consider the transfer-function system

$$\frac{Y(s)}{U(s)} = \frac{10s + 10}{s^3 + 6s^2 + 5s + 10} \tag{11-22}$$

There are many (again, infinitely many) possible state-space representations for this system. One possible state-space representation is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -10 & -5 & -6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 10 \\ -50 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [0]u$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Another possible state-space representation (among infinitely many alternatives) is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -6 & -5 & -10 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u \tag{11–23}$$

$$y = \begin{bmatrix} 0 & 10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [0]u \tag{11–24}$$

MATLAB transforms the transfer function given by Equation (11–22) into the state-space representation given by Equations (11–23) and (11–24). For the example system considered here, MATLAB Program 11–1 will produce matrices **A**, **B**, **C**, and D.

Transformation from State Space to Transfer Function. To obtain the transfer function from state-space equations, use the following command:

```
[num,den] = ss2tf(A,B,C,D,iu)
```

iu must be specified for systems with more than one input. For example, if the system has three inputs ($u1$, $u2$, $u3$), then iu must be either 1, 2, or 3, where 1 implies $u1$, 2 implies $u2$, and 3 implies $u3$.

If the system has only one input, then either

```
[num,den] = ss2tf(A,B,C,D)
```

or

```
[num,den] = ss2tf(A,B,C,D,1)
```

may be used. (See Example 11–3 and MATLAB Program 11–2.)

MATLAB Program 11–1

```
num = [0 0 10 10];
den = [1 6 5 10];
[A,B,C,D] = tf2ss(num,den)

A =

    -6   -5   -10
     1    0     0
     0    1     0

B =

     1
     0
     0

C =

     0   10   10

D =

     0
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

For the case where the system has multiple inputs and multiple outputs, see Example 11–4.

EXAMPLE 11–3 Obtain the transfer function of the system defined by the following state-space equations:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -5.008 & -25.1026 & -5.03247 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 25.04 \\ -121.005 \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

MATLAB Program 11–2 will produce the transfer function for the given system. The transfer function obtained is given by

$$\frac{Y(s)}{U(s)} = \frac{25.04s + 5.008}{s^3 + 5.0325s^2 + 25.1026s + 5.008}$$

MATLAB Program 11–2

```
A = [0 1 0;0 0 1;-5.008 -25.1026 -5.03247];
B = [0;25.04; -121.005];
C = [1 0 0];
D = [0];
[num,den] = ss2tf(A,B,C,D)

num =

        0   -0.0000   25.0400   5.0080

den =

   1.0000   5.0325   25.1026   5.0080

% ***** The same result can be obtained by entering the following command *****


[num,den] = ss2tf(A,B,C,D,1)

num =

        0   -0.0000   25.0400   5.0080

den =

   1.0000   5.0325   25.1026   5.0080
```


PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11–4 Consider a system with multiple inputs and multiple outputs. When the system has more than one output, the command

```
[NUM,den] = ss2tf(A,B,C,D,iu)
```

produces transfer functions for all outputs to each input. (The numerator coefficients are returned to matrix NUM with as many rows as there are outputs.)

Consider the system defined by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -25 & -4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

This system involves two inputs and two outputs. Four transfer functions are involved: $Y_1(s)/U_1(s)$, $Y_2(s)/U_1(s)$, $Y_1(s)/U_2(s)$, and $Y_2(s)/U_2(s)$. (When considering input u_1, we assume that input u_2 is zero and vice versa.) See the output of MATLAB Program 11–3.

MATLAB Program 11–3

```
A = [0 1;-25 -4];
B = [1 1;0 1];
C = [1 0;0 1];
D = [0 0;0 0];
[NUM,den] = ss2tf(A,B,C,D,1)

NUM =

    0    1    4
    0    0  -25

den =

    1  4  25

[NUM,den] = ss2tf(A,B,C,D,2)

NUM =

    0  1.0000    5.0000
    0  1.0000  -25.0000

den =

    1    4    25
```

This is the MATLAB representation of the following four transfer functions:

$$\frac{Y_1(s)}{U_1(s)} = \frac{s+4}{s^2+4s+25}, \qquad \frac{Y_2(s)}{U_1(s)} = \frac{-25}{s^2+4s+25}$$

$$\frac{Y_1(s)}{U_2(s)} = \frac{s+5}{s^2+4s+25}, \qquad \frac{Y_2(s)}{U_2(s)} = \frac{s-25}{s^2+4s+25}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

11–4 SOLVING THE TIME-INVARIANT STATE EQUATION

In this section, we shall obtain the general solution of the linear time-invariant state equation. We shall first consider the homogeneous case and then the nonhomogeneous case.

Solution of Homogeneous State Equations. Before we solve vector-matrix differential equations, let us review the solution of the scalar differential equation

$$\dot{x} = ax \tag{11–25}$$

In solving this equation, we may assume a solution $x(t)$ of the form

$$x(t) = b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k + \cdots \tag{11–26}$$

By substituting this assumed solution into Equation (11–25), we obtain

$$\begin{aligned} &b_1 + 2b_2 t + 3b_3 t^2 + \cdots + kb_k t^{k-1} + \cdots \\ &= a(b_0 + b_1 t + b_2 t^2 + \cdots + b_k t^k + \cdots) \end{aligned} \tag{11–27}$$

If the assumed solution is to be the true solution, Equation (11–27) must hold for any t. Hence, equating the coefficients of the equal powers of t, we obtain

$$\begin{aligned} b_1 &= ab_0 \\ b_2 &= \frac{1}{2} ab_1 = \frac{1}{2} a^2 b_0 \\ b_3 &= \frac{1}{3} ab_2 = \frac{1}{3 \times 2} a^3 b_0 \\ &\vdots \\ b_k &= \frac{1}{k!} a^k b_0 \end{aligned}$$

The value of b_0 is determined by substituting $t = 0$ into Equation (11–26), or

$$x(0) = b_0$$

Hence, the solution $x(t)$ can be written as

$$\begin{aligned} x(t) &= \left(1 + at + \frac{1}{2!} a^2 t^2 + \cdots + \frac{1}{k!} a^k t^k + \cdots\right) x(0) \\ &= e^{at} x(0) \end{aligned}$$

We shall now solve the vector-matrix differential equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{11–28}$$

where $\mathbf{x} = n$-vector
$\mathbf{A} = n \times n$ constant matrix

By analogy with the scalar case, we assume that the solution is in the form of a vector power series in t, or

$$\mathbf{x}(t) = \mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{b}_2 t^2 + \cdots + \mathbf{b}_k t^k + \cdots \tag{11–29}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

By substituting this assumed solution into Equation (11–28), we obtain

$$\begin{aligned}&\mathbf{b}_1 + 2\mathbf{b}_2 t + 3\mathbf{b}_3 t^2 + \cdots + k\mathbf{b}_k t^{k-1} + \cdots \\ &= \mathbf{A}(\mathbf{b}_0 + \mathbf{b}_1 t + \mathbf{b}_2 t^2 + \cdots + \mathbf{b}_k t^k + \cdots)\end{aligned} \tag{11–30}$$

If the assumed solution is to be the true solution, Equation (11–30) must hold for all t. Thus, by equating the coefficients of like powers of t on both sides of Equation (11–30), we obtain

$$\begin{aligned}\mathbf{b}_1 &= \mathbf{A}\mathbf{b}_0 \\ \mathbf{b}_2 &= \frac{1}{2}\mathbf{A}\mathbf{b}_1 = \frac{1}{2}\mathbf{A}^2\mathbf{b}_0 \\ \mathbf{b}_3 &= \frac{1}{3}\mathbf{A}\mathbf{b}_2 = \frac{1}{3 \times 2}\mathbf{A}^3\mathbf{b}_0 \\ &\vdots \\ \mathbf{b}_k &= \frac{1}{k!}\mathbf{A}^k\mathbf{b}_0\end{aligned}$$

By substituting $t = 0$ into Equation (11–29), we obtain

$$\mathbf{x}(0) = \mathbf{b}_0$$

Thus, the solution $\mathbf{x}(t)$ can be written as

$$\mathbf{x}(t) = \left(\mathbf{I} + \mathbf{A}t + \frac{1}{2!}\mathbf{A}^2 t^2 + \cdots + \frac{1}{k!}\mathbf{A}^k t^k + \cdots\right)\mathbf{x}(0)$$

The expression in the parentheses on the right-hand side of this last equation is an $n \times n$ matrix. Because of its similarity to the infinite power series for a scalar exponential, we call it the matrix exponential and write

$$\mathbf{I} + \mathbf{A}t + \frac{1}{2!}\mathbf{A}^2 t^2 + \cdots + \frac{1}{k!}\mathbf{A}^k t^k + \cdots = e^{\mathbf{A}t}$$

In terms of the matrix exponential, the solution of Equation (11–28) can be written as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) \tag{11–31}$$

Since the matrix exponential is very important in the state-space analysis of linear systems, we shall next examine its properties.

Matrix Exponential. It can be proved that the matrix exponential of an $n \times n$ matrix $\mathbf{A}$,

$$e^{\mathbf{A}t} = \sum_{k=0}^{\infty} \frac{\mathbf{A}^k t^k}{k!}$$

converges absolutely for all finite t. (Hence, computer calculations for evaluating the elements of $e^{\mathbf{A}t}$ by using the series expansion can be easily carried out.)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Because of the convergence of the infinite series $\sum_{k=0}^{\infty} \mathbf{A}^k t^k / k!$, the series can be differentiated term by term to give

$$\begin{aligned}
\frac{d}{dt} e^{\mathbf{A}t} &= \mathbf{A} + \mathbf{A}^2 t + \frac{\mathbf{A}^3 t^2}{2!} + \cdots + \frac{\mathbf{A}^k t^{k-1}}{(k-1)!} + \cdots \\
&= \mathbf{A}\left[\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \cdots + \frac{\mathbf{A}^{k-1} t^{k-1}}{(k-1)!} + \cdots\right] = \mathbf{A}e^{\mathbf{A}t} \\
&= \left[\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \cdots + \frac{\mathbf{A}^{k-1} t^{k-1}}{(k-1)!} + \cdots\right]\mathbf{A} = e^{\mathbf{A}t}\mathbf{A}
\end{aligned}$$

The matrix exponential has the property that

$$e^{\mathbf{A}(t+s)} = e^{\mathbf{A}t} e^{\mathbf{A}s}$$

This can be proved as follows:

$$\begin{aligned}
e^{\mathbf{A}t} e^{\mathbf{A}s} &= \left(\sum_{k=0}^{\infty} \frac{\mathbf{A}^k t^k}{k!}\right)\left(\sum_{k=0}^{\infty} \frac{\mathbf{A}^k s^k}{k!}\right) \\
&= \sum_{k=0}^{\infty} \mathbf{A}^k \left(\sum_{i=0}^{\infty} \frac{t^i s^{k-i}}{i!\,(k-i)!}\right) \\
&= \sum_{k=0}^{\infty} \mathbf{A}^k \frac{(t+s)^k}{k!} \\
&= e^{\mathbf{A}(t+s)}
\end{aligned}$$

In particular, if $s = -t$, then

$$e^{\mathbf{A}t} e^{-\mathbf{A}t} = e^{-\mathbf{A}t} e^{\mathbf{A}t} = e^{\mathbf{A}(t-t)} = \mathbf{I}$$

Thus, the inverse of $e^{\mathbf{A}t}$ is $e^{-\mathbf{A}t}$. Since the inverse of $e^{\mathbf{A}t}$ always exists, $e^{\mathbf{A}t}$ is nonsingular.

It is very important to remember that

$$\begin{aligned}
e^{(\mathbf{A}+\mathbf{B})t} &= e^{\mathbf{A}t} e^{\mathbf{B}t}, \qquad \text{if } \mathbf{AB} = \mathbf{BA} \\
e^{(\mathbf{A}+\mathbf{B})t} &\neq e^{\mathbf{A}t} e^{\mathbf{B}t}, \qquad \text{if } \mathbf{AB} \neq \mathbf{BA}
\end{aligned}$$

To prove this, note that

$$\begin{aligned}
e^{(\mathbf{A}+\mathbf{B})t} &= \mathbf{I} + (\mathbf{A}+\mathbf{B})t + \frac{(\mathbf{A}+\mathbf{B})^2}{2!} t^2 + \frac{(\mathbf{A}+\mathbf{B})^3}{3!} t^3 + \cdots \\
e^{\mathbf{A}t} e^{\mathbf{B}t} &= \left(\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \cdots\right)\left(\mathbf{I} + \mathbf{B}t + \frac{\mathbf{B}^2 t^2}{2!} + \frac{\mathbf{B}^3 t^3}{3!} + \cdots\right) \\
&= \mathbf{I} + (\mathbf{A}+\mathbf{B})t + \frac{\mathbf{A}^2 t^2}{2!} + \mathbf{AB}t^2 + \frac{\mathbf{B}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} \\
&\quad + \frac{\mathbf{A}^2\mathbf{B}t^3}{2!} + \frac{\mathbf{AB}^2 t^3}{2!} + \frac{\mathbf{B}^3 t^3}{3!} + \cdots
\end{aligned}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Hence,

$$e^{(\mathbf{A}+\mathbf{B})t} - e^{\mathbf{A}t}e^{\mathbf{B}t} = \frac{\mathbf{BA} - \mathbf{AB}}{2!}t^2$$
$$+ \frac{\mathbf{BA}^2 + \mathbf{ABA} + \mathbf{B}^2\mathbf{A} + \mathbf{BAB} - 2\mathbf{A}^2\mathbf{B} - 2\mathbf{AB}^2}{3!}t^3 + \cdots$$

The difference between $e^{(\mathbf{A}+\mathbf{B})t}$ and $e^{\mathbf{A}t}e^{\mathbf{B}t}$ vanishes if $\mathbf{A}$ and $\mathbf{B}$ commute.

Laplace Transform Approach to the Solution of Homogeneous State Equations. Let us first consider the scalar case:

$$\dot{x} = ax \tag{11–32}$$

Taking the Laplace transform of Equation (11–32), we obtain

$$sX(s) - x(0) = aX(s) \tag{11–33}$$

where $X(s) = \mathscr{L}[x]$. Solving Equation (11–33) for $X(s)$ gives

$$X(s) = \frac{x(0)}{s-a} = (s-a)^{-1}x(0)$$

The inverse Laplace transform of this last equation gives the solution

$$x(t) = e^{at}x(0)$$

The foregoing approach to the solution of the homogeneous scalar differential equation can be extended to the homogeneous state equation:

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \tag{11–34}$$

Taking the Laplace transform of both sides of Equation (11–34), we obtain

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s)$$

where $\mathbf{X}(s) = \mathscr{L}[\mathbf{x}]$. Hence,

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0)$$

Premultiplying both sides of this last equation by $(s\mathbf{I} - \mathbf{A})^{-1}$, we obtain

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0)$$

The inverse Laplace transform of $\mathbf{X}(s)$ gives the solution $\mathbf{x}(t)$. Thus,

$$\mathbf{x}(t) = \mathscr{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}]\mathbf{x}(0) \tag{11–35}$$

Note that

$$(s\mathbf{I} - \mathbf{A})^{-1} = \frac{\mathbf{I}}{s} + \frac{\mathbf{A}}{s^2} + \frac{\mathbf{A}^2}{s^3} + \cdots$$

Hence, the inverse Laplace transform of $(s\mathbf{I} - \mathbf{A})^{-1}$ gives

$$\mathscr{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}] = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2t^2}{2!} + \frac{\mathbf{A}^3t^3}{3!} + \cdots = e^{\mathbf{A}t} \tag{11–36}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

(The inverse Laplace transform of a matrix is the matrix consisting of the inverse Laplace transforms of all elements.) From Equations (11–35) and (11–36), the solution of Equation (11–34) is obtained as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0)$$

The importance of Equation (11–36) lies in the fact that it provides a convenient means for finding the closed solution for the matrix exponential.

State-Transition Matrix. We can write the solution of the homogeneous state equation

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{11–37}$$

as

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\mathbf{x}(0) \tag{11–38}$$

where $\mathbf{\Phi}(t)$ is an $n \times n$ matrix and is the unique solution of

$$\dot{\mathbf{\Phi}}(t) = \mathbf{A}\mathbf{\Phi}(t), \qquad \mathbf{\Phi}(0) = \mathbf{I}$$

To verify this, note that

$$\mathbf{x}(0) = \mathbf{\Phi}(0)\mathbf{x}(0) = \mathbf{x}(0)$$

and

$$\dot{\mathbf{x}}(t) = \dot{\mathbf{\Phi}}(t)\mathbf{x}(0) = \mathbf{A}\mathbf{\Phi}(t)\mathbf{x}(0) = \mathbf{A}\mathbf{x}(t)$$

We thus confirm that Equation (11–38) is the solution of Equation (11–37).

From Equations (11–31), (11–35), and (11–38), we obtain

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

Note that

$$\mathbf{\Phi}^{-1}(t) = e^{-\mathbf{A}t} = \mathbf{\Phi}(-t)$$

From Equation (11–38), we see that the solution of Equation (11–37) is simply a transformation of the initial condition. Hence, the unique matrix $\mathbf{\Phi}(t)$ is called the state-transition matrix. The state-transition matrix contains all the information about the free motions of the system defined by Equation (11–37).

If the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the matrix $\mathbf{A}$ are distinct, than $\mathbf{\Phi}(t)$ will contain the n exponentials

$$e^{\lambda_1 t}, e^{\lambda_2 t}, \ldots, e^{\lambda_n t}$$

In particular, if the matrix $\mathbf{A}$ is diagonal, then

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \begin{bmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_n t} \end{bmatrix} \qquad (\mathbf{A}\text{: diagonal})$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

If there is a multiplicity in the eigenvalues, for example, if the eigenvalues of $\mathbf{A}$ are

$$\lambda_1, \lambda_1, \lambda_1, \lambda_4, \lambda_5, \ldots, \lambda_n,$$

then $\mathbf{\Phi}(t)$ will contain, in addition to the exponentials $e^{\lambda_1 t}, e^{\lambda_4 t}, e^{\lambda_5 t}, \ldots, e^{\lambda_n t}$, terms like $te^{\lambda_1 t}$ and $t^2 e^{\lambda_1 t}$.

Properties of State-Transition Matrices. We shall now summarize the important properties of the state-transition matrix $\mathbf{\Phi}(t)$. For the time-invariant system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

for which

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t}$$

we have the following:

1. $\mathbf{\Phi}(0) = e^{\mathbf{A}0} = \mathbf{I}$
2. $\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \left(e^{-\mathbf{A}t}\right)^{-1} = \left[\mathbf{\Phi}(-t)\right]^{-1}$ or $\mathbf{\Phi}^{-1}(t) = \mathbf{\Phi}(-t)$
3. $\mathbf{\Phi}(t_1 + t_2) = e^{\mathbf{A}(t_1+t_2)} = e^{\mathbf{A}t_1}e^{\mathbf{A}t_2} = \mathbf{\Phi}(t_1)\mathbf{\Phi}(t_2) = \mathbf{\Phi}(t_2)\mathbf{\Phi}(t_1)$
4. $\left[\mathbf{\Phi}(t)\right]^n = \mathbf{\Phi}(nt)$
5. $\mathbf{\Phi}(t_2 - t_1)\mathbf{\Phi}(t_1 - t_0) = \mathbf{\Phi}(t_2 - t_0) = \mathbf{\Phi}(t_1 - t_0)\mathbf{\Phi}(t_2 - t_1)$

EXAMPLE 11–5 Obtain the state-transition matrix $\mathbf{\Phi}(t)$ of the following system:

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

Obtain also the inverse of the state-transition matrix, $\mathbf{\Phi}^{-1}(t)$.

For this system,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$$

The state-transition matrix $\mathbf{\Phi}(t)$ is given by

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

Since

$$s\mathbf{I} - \mathbf{A} = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} = \begin{bmatrix} s & -1 \\ 2 & s+3 \end{bmatrix}$$

the inverse of $(s\mathbf{I} - \mathbf{A})$ is given by

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

$$= \begin{bmatrix} \dfrac{s+3}{(s+1)(s+2)} & \dfrac{1}{(s+1)(s+2)} \\ \dfrac{-2}{(s+1)(s+2)} & \dfrac{s}{(s+1)(s+2)} \end{bmatrix}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Hence,

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}]$$
$$= \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix}$$

Noting that $\mathbf{\Phi}^{-1}(t) = \mathbf{\Phi}(-t)$, we obtain the inverse of the state-transition matrix as follows:

$$\mathbf{\Phi}^{-1}(t) = e^{-\mathbf{A}t} = \begin{bmatrix} 2e^{t} - e^{2t} & e^{t} - e^{2t} \\ -2e^{t} + 2e^{2t} & -e^{t} + 2e^{2t} \end{bmatrix}$$

Solution of Nonhomogeneous State Equations. We shall begin by considering the scalar case

$$\dot{x} = ax + bu \tag{11–39}$$

Let us rewrite Equation (11–39) as

$$\dot{x} - ax = bu$$

Multiplying both sides of this equation by e^{-at}, we obtain

$$e^{-at}[\dot{x}(t) - ax(t)] = \frac{d}{dt}[e^{-at}x(t)] = e^{-at}bu(t)$$

Integrating this equation between 0 and t gives

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

or

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

The first term on the right-hand side is the response to the initial condition and the second term is the response to the input $u(t)$.

Let us now consider the nonhomogeneous state equation described by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{11–40}$$

where $\mathbf{x}$ = n-vector
$\mathbf{u}$ = r-vector
$\mathbf{A}$ = $n \times n$ constant matrix
$\mathbf{B}$ = $n \times r$ constant matrix

By writing Equation (11–40) as

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

and premultiplying both sides of this equation by $e^{-\mathbf{A}t}$, we obtain

$$e^{-\mathbf{A}t}[\dot{\mathbf{x}}(t) - \mathbf{A}\mathbf{x}(t)] = \frac{d}{dt}[e^{-\mathbf{A}t}\mathbf{x}(t)] = e^{-\mathbf{A}t}\mathbf{B}\mathbf{u}(t)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Integrating the preceding equation between 0 and t gives

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

or

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{11–41}$$

Equation (11–41) can also be written as

$$\mathbf{x}(t) = \mathbf{\Phi}(t)\mathbf{x}(0) + \int_0^t \mathbf{\Phi}(t-\tau)\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{11–42}$$

where $\mathbf{\Phi}(t) = e^{\mathbf{A}t}$. Equation (11–41) or (11–42) is the solution of Equation (11–40). The solution $\mathbf{x}(t)$ is clearly the sum of a term consisting of the transition of the initial state and a term arising from the input vector.

Laplace Transform Approach to the Solution of Nonhomogeneous State Equations. The solution of the nonhomogeneous state equation

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

can also be obtained by the Laplace transform approach. The Laplace transform of this last equation yields

$$s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s) + \mathbf{B}\mathbf{U}(s)$$

or

$$(s\mathbf{I} - \mathbf{A})\mathbf{X}(s) = \mathbf{x}(0) + \mathbf{B}\mathbf{U}(s)$$

Premultiplying both sides of this last equation by $(s\mathbf{I} - \mathbf{A})^{-1}$, we obtain

$$\mathbf{X}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{x}(0) + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{U}(s)$$

Using the relationship given by Equation (11–36) gives

$$\mathbf{X}(s) = \mathscr{L}[e^{\mathbf{A}t}]\mathbf{x}(0) + \mathscr{L}[e^{\mathbf{A}t}]\mathbf{B}\mathbf{U}(s)$$

The inverse Laplace transform of this last equation can be obtained by use of the convolution integral as follows:

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

Solution in Terms of $\mathbf{x}(t_0)$. Thus far we have assumed the initial time to be zero. If, however, the initial time is given by t_0 instead of 0, then the solution to Equation (11–40) must be modified to

$$\mathbf{x}(t) = e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}\mathbf{u}(\tau)\,d\tau \tag{11–43}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11–6 Obtain the time response of the following system:

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

where $u(t)$ is the unit-step function occurring at $t = 0$, or

$$u(t) = 1(t)$$

For this system,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The state-transition matrix $\mathbf{\Phi}(t) = e^{\mathbf{A}t}$ was obtained in Example 11–5 as

$$\mathbf{\Phi}(t) = e^{\mathbf{A}t} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix}$$

The response to the unit-step input is then obtained as

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t \begin{bmatrix} 2e^{-(t-\tau)} - e^{-2(t-\tau)} & e^{-(t-\tau)} - e^{-2(t-\tau)} \\ -2e^{-(t-\tau)} + 2e^{-2(t-\tau)} & -e^{-(t-\tau)} + 2e^{-2(t-\tau)} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} [1]\, d\tau$$

or

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{bmatrix} \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} + \begin{bmatrix} \frac{1}{2} - e^{-t} + \frac{1}{2}e^{-2t} \\ e^{-t} - e^{-2t} \end{bmatrix}$$

If the initial state is zero, or $\mathbf{x}(0) = \mathbf{0}$, then $\mathbf{x}(t)$ can be simplified to

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} - e^{-t} + \frac{1}{2}e^{-2t} \\ e^{-t} - e^{-2t} \end{bmatrix}$$

11–5 SOME USEFUL RESULTS IN VECTOR-MATRIX ANALYSIS

In this section we present some useful results in vector-matrix analysis that we use in Section 11–6. Specifically, we present the Cayley–Hamilton theorem, the minimal polynomial, Sylvester's interpolation method for calculating $e^{\mathbf{A}t}$, and the linear independence of vectors.

Cayley–Hamilton Theorem. The Cayley–Hamilton theorem is very useful in proving theorems involving matrix equations or solving problems involving matrix equations.

Consider an $n \times n$ matrix $\mathbf{A}$ and its characteristic equation:

$$|\lambda\mathbf{I} - \mathbf{A}| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = 0$$

The Cayley–Hamilton theorem states that the matrix $\mathbf{A}$ satisfies its own characteristic equation, or that

$$\mathbf{A}^n + a_1\mathbf{A}^{n-1} + \cdots + a_{n-1}\mathbf{A} + a_n\mathbf{I} = \mathbf{0} \tag{11–44}$$

To prove this theorem, note that $\operatorname{adj}(\lambda\mathbf{I} - \mathbf{A})$ is a polynomial in λ of degree $n - 1$. That is,

$$\operatorname{adj}(\lambda\mathbf{I} - \mathbf{A}) = \mathbf{B}_1\lambda^{n-1} + \mathbf{B}_2\lambda^{n-2} + \cdots + \mathbf{B}_{n-1}\lambda + \mathbf{B}_n$$

0785

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

where $\mathbf{B}_1 = \mathbf{I}$. Since

$$(\lambda\mathbf{I} - \mathbf{A})\,\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A}) = [\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})](\lambda\mathbf{I} - \mathbf{A}) = |\lambda\mathbf{I} - \mathbf{A}|\mathbf{I}$$

we obtain

$$\begin{aligned}|\lambda\mathbf{I} - \mathbf{A}|\mathbf{I} &= \mathbf{I}\lambda^n + a_1\mathbf{I}\lambda^{n-1} + \cdots + a_{n-1}\mathbf{I}\lambda + a_n\mathbf{I}\\ &= (\lambda\mathbf{I} - \mathbf{A})(\mathbf{B}_1\lambda^{n-1} + \mathbf{B}_2\lambda^{n-2} + \cdots + \mathbf{B}_{n-1}\lambda + \mathbf{B}_n)\\ &= (\mathbf{B}_1\lambda^{n-1} + \mathbf{B}_2\lambda^{n-2} + \cdots + \mathbf{B}_{n-1}\lambda + \mathbf{B}_n)(\lambda\mathbf{I} - \mathbf{A})\end{aligned}$$

From this equation, we see that $\mathbf{A}$ and $\mathbf{B}_i$ $(i = 1, 2, \ldots, n)$ commute. Hence, the product of $(\lambda\mathbf{I} - \mathbf{A})$ and $\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})$ becomes zero if either of these is zero. If $\mathbf{A}$ is substituted for λ in this last equation, then clearly $\lambda\mathbf{I} - \mathbf{A}$ becomes zero. Hence, we obtain

$$\mathbf{A}^n + a_1\mathbf{A}^{n-1} + \cdots + a_{n-1}\mathbf{A} + a_n\mathbf{I} = \mathbf{0}$$

This proves the Cayley–Hamilton theorem, or Equation (11–44).

Minimal Polynomial. Referring to the Cayley–Hamilton theorem, every $n \times n$ matrix $\mathbf{A}$ satisfies its own characteristic equation. The characteristic equation is not, however, necessarily the scalar equation of least degree that $\mathbf{A}$ satisfies. The least-degree polynomial having $\mathbf{A}$ as a root is called the *minimal polynomial.* That is, the minimal polynomial of an $n \times n$ matrix $\mathbf{A}$ is defined as the polynomial $\phi(\lambda)$ of least degree,

$$\phi(\lambda) = \lambda^m + a_1\lambda^{m-1} + \cdots + a_{m-1}\lambda + a_m, \qquad m \le n$$

such that $\phi(\mathbf{A}) = \mathbf{0}$, or

$$\phi(\mathbf{A}) = \mathbf{A}^m + a_1\mathbf{A}^{m-1} + \cdots + a_{m-1}\mathbf{A} + a_m\mathbf{I} = \mathbf{0}$$

The minimal polynomial plays an important role in the computation of polynomials in an $n \times n$ matrix.

Let us suppose that $d(\lambda)$, a polynomial in λ, is the greatest common divisor of all the elements of $\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})$. We can show that if the coefficient of the highest-degree term in λ of $d(\lambda)$ is chosen as 1, then the minimal polynomial $\phi(\lambda)$ is given by

$$\phi(\lambda) = \frac{|\lambda\mathbf{I} - \mathbf{A}|}{d(\lambda)} \tag{11–45}$$

[See Problem **A–11–8** for the derivation of Equation (11–45).]

It is noted that the minimal polynomial $\phi(\lambda)$ of an $n \times n$ matrix $\mathbf{A}$ can be determined by the following procedure:

1. Form $\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})$ and write the elements of $\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})$ as factored polynomials in λ.
2. Determine $d(\lambda)$ as the greatest common divisor of all the elements of $\mathrm{adj}(\lambda\mathbf{I} - \mathbf{A})$. Choose the coefficient of the highest-degree term in λ of $d(\lambda)$ to be 1. If there is no common divisor, $d(\lambda) = 1$.
3. The minimal polynomial $\phi(\lambda)$ is then given as $|\lambda\mathbf{I} - \mathbf{A}|$ divided by $d(\lambda)$.

Matrix Exponential $e^{\mathbf{A}t}$. In solving control engineering problems, it often becomes necessary to compute $e^{\mathbf{A}t}$. If matrix $\mathbf{A}$ is given with all elements in numerical values, MATLAB provides a simple way to compute $e^{\mathbf{A}T}$, where T is a constant.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Aside from computational methods, several analytical methods are available for the computation of $e^{\mathbf{A}t}$. We shall present three methods here.

Computation of $e^{\mathbf{A}t}$: Method 1. If matrix $\mathbf{A}$ can be transformed into a diagonal form, then $e^{\mathbf{A}t}$ can be given by

$$e^{\mathbf{A}t} = \mathbf{P}e^{\mathbf{D}t}\mathbf{P}^{-1} = \mathbf{P}\begin{bmatrix} e^{\lambda_1 t} & & & & 0 \\ & e^{\lambda_2 t} & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & e^{\lambda_n t} \end{bmatrix}\mathbf{P}^{-1} \tag{11–46}$$

where $\mathbf{P}$ is a diagonalizing matrix for $\mathbf{A}$. [For the derivation of Equation (11–46), see Problem **A–11–11**.]

If matrix $\mathbf{A}$ can be transformed into a Jordan canonical form, then $e^{\mathbf{A}t}$ can be given by

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

As an example, consider the following matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$$

The characteristic equation is

$$|\lambda\mathbf{I} - \mathbf{A}| = \lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3 = 0$$

Thus, matrix $\mathbf{A}$ has a multiple eigenvalue of order 3 at $\lambda = 1$. It can be shown that matrix $\mathbf{A}$ has a multiple eigenvector of order 3. The transformation matrix that will transform matrix $\mathbf{A}$ into a Jordan canonical form can be given by

$$\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

The inverse of matrix $\mathbf{S}$ is

$$\mathbf{S}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}$$

Then it can be seen that

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix}\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

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

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Noting that

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

we find

$$\begin{aligned} e^{\mathbf{A}t} &= \mathbf{S}e^{\mathbf{J}t}\mathbf{S}^{-1} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} e^t & te^t & \frac{1}{2}t^2e^t \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} e^t - te^t + \frac{1}{2}t^2e^t & te^t - t^2e^t & \frac{1}{2}t^2e^t \\ \frac{1}{2}t^2e^t & e^t - te^t - t^2e^t & te^t + \frac{1}{2}t^2e^t \\ te^t + \frac{1}{2}t^2e^t & -3te^t - t^2e^t & e^t + 2te^t + \frac{1}{2}t^2e^t \end{bmatrix} \end{aligned}$$

Computation of $e^{\mathbf{A}t}$: Method 2. The second method of computing $e^{\mathbf{A}t}$ uses the Laplace transform approach. Referring to Equation (11–36), $e^{\mathbf{A}t}$ can be given as follows:

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

Thus, to obtain $e^{\mathbf{A}t}$, first invert the matrix $(s\mathbf{I} - \mathbf{A})$. This results in a matrix whose elements are rational functions of s. Then take the inverse Laplace transform of each element of the matrix.

EXAMPLE 11–7 Consider the following matrix $\mathbf{A}$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix}$$

Compute $e^{\mathbf{A}t}$ by use of the two analytical methods presented previously.

Method 1. The eigenvalues of $\mathbf{A}$ are 0 and -2 $(\lambda_1 = 0, \lambda_2 = -2)$. A necessary transformation matrix $\mathbf{P}$ may be obtained as

$$\mathbf{P} = \begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix}$$

Then, from Equation (11–46), $e^{\mathbf{A}t}$ is obtained as follows:

$$e^{\mathbf{A}t} = \begin{bmatrix} 1 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} e^0 & 0 \\ 0 & e^{-2t} \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{2}(1 - e^{-2t}) \\ 0 & e^{-2t} \end{bmatrix}$$

Method 2. Since

$$s\mathbf{I} - \mathbf{A} = \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} = \begin{bmatrix} s & -1 \\ 0 & s+2 \end{bmatrix}$$

we obtain

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

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Hence,

$$e^{\mathbf{A}t} = \mathcal{L}^{-1}[(s\mathbf{I} - \mathbf{A})^{-1}] = \begin{bmatrix} 1 & \frac{1}{2}(1 - e^{-2t}) \\ 0 & e^{-2t} \end{bmatrix}$$

Computation of $e^{\mathbf{A}t}$: Method 3. The third method is based on Sylvester's interpolation method. (For Sylvester's interpolation formula, see Problem **A–11–12**.) We shall first consider the case where the roots of the minimal polynomial $\phi(\lambda)$ of $\mathbf{A}$ are distinct. Then we shall deal with the case of multiple roots.

Case 1: Minimal Polynomial of $\mathbf{A}$ *Involves Only Distinct Roots.* We shall assume that the degree of the minimal polynomial of $\mathbf{A}$ is m. By using Sylvester's interpolation formula, it can be shown that $e^{\mathbf{A}t}$ can be obtained by solving the following determinant equation:

$$\begin{vmatrix} 1 & \lambda_1 & \lambda_1^2 & \cdots & \lambda_1^{m-1} & e^{\lambda_1 t} \\ 1 & \lambda_2 & \lambda_2^2 & \cdots & \lambda_2^{m-1} & e^{\lambda_2 t} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ 1 & \lambda_m & \lambda_m^2 & \cdots & \lambda_m^{m-1} & e^{\lambda_m t} \\ \mathbf{I} & \mathbf{A} & \mathbf{A}^2 & \cdots & \mathbf{A}^{m-1} & e^{\mathbf{A}t} \end{vmatrix} = \mathbf{0} \tag{11–47}$$

By solving Equation (11–47) for $e^{\mathbf{A}t}$, $e^{\mathbf{A}t}$ can be obtained in terms of the $\mathbf{A}^k$ $(k = 0, 1, 2, \ldots, m - 1)$ and the $e^{\lambda_i t}$ $(i = 1, 2, 3, \ldots, m)$. [Equation (11–47) may be expanded, for example, about the last column.]

Notice that solving Equation (11–47) for $e^{\mathbf{A}t}$ is the same as writing

$$e^{\mathbf{A}t} = \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} + \alpha_2(t)\mathbf{A}^2 + \cdots + \alpha_{m-1}(t)\mathbf{A}^{m-1} \tag{11–48}$$

and determining the $\alpha_k(t)$ $(k = 0, 1, 2, \ldots, m - 1)$ by solving the following set of m equations for the $\alpha_k(t)$:

$$\begin{aligned} \alpha_0(t) + \alpha_1(t)\lambda_1 + \alpha_2(t)\lambda_1^2 + \cdots + \alpha_{m-1}(t)\lambda_1^{m-1} &= e^{\lambda_1 t} \\ \alpha_0(t) + \alpha_1(t)\lambda_2 + \alpha_2(t)\lambda_2^2 + \cdots + \alpha_{m-1}(t)\lambda_2^{m-1} &= e^{\lambda_2 t} \\ &\vdots \\ \alpha_0(t) + \alpha_1(t)\lambda_m + \alpha_2(t)\lambda_m^2 + \cdots + \alpha_{m-1}(t)\lambda_m^{m-1} &= e^{\lambda_m t} \end{aligned}$$

If $\mathbf{A}$ is an $n \times n$ matrix and has distinct eigenvalues, then the number of $\alpha_k(t)$'s to be determined is $m = n$. If $\mathbf{A}$ involves multiple eigenvalues, but its minimal polynomial has only simple roots, however, then the number m of $\alpha_k(t)$'s to be determined is less than n.

Case 2: Minimal Polynomial of $\mathbf{A}$ *Involves Multiple Roots.* As an example, consider the case where the minimal polynomial of $\mathbf{A}$ involves three equal roots $(\lambda_1 = \lambda_2 = \lambda_3)$ and has other roots $(\lambda_4, \lambda_5, \ldots, \lambda_m)$ that are all distinct. By applying Sylvester's interpolation formula, it can be shown that $e^{\mathbf{A}t}$ can be obtained from the following determinant equation:

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$
\begin{vmatrix}
0 & 0 & 1 & 3\lambda_1 & \cdots & \dfrac{(m-1)(m-2)}{2}\lambda_1^{m-3} & \dfrac{t^2}{2}e^{\lambda_1 t} \\
0 & 1 & 2\lambda_1 & 3\lambda_1^2 & \cdots & (m-1)\lambda_1^{m-2} & te^{\lambda_1 t} \\
1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 & \cdots & \lambda_1^{m-1} & e^{\lambda_1 t} \\
1 & \lambda_4 & \lambda_4^2 & \lambda_4^3 & \cdots & \lambda_4^{m-1} & e^{\lambda_4 t} \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\
1 & \lambda_m & \lambda_m^2 & \lambda_m^3 & \cdots & \lambda_m^{m-1} & e^{\lambda_m t} \\
\mathbf{I} & \mathbf{A} & \mathbf{A}^2 & \mathbf{A}^3 & \cdots & \mathbf{A}^{m-1} & e^{\mathbf{A}t}
\end{vmatrix} = \mathbf{0} \qquad (11\text{–}49)
$$

Equation (11–49) can be solved for $e^{\mathbf{A}t}$ by expanding it about the last column.

It is noted that, just as in case 1, solving Equation (11–49) for $e^{\mathbf{A}t}$ is the same as writing

$$
e^{\mathbf{A}t} = \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} + \alpha_2(t)\mathbf{A}^2 + \cdots + \alpha_{m-1}(t)\mathbf{A}^{m-1} \qquad (11\text{–}50)
$$

and determining the $\alpha_k(t)$'s ($k = 0, 1, 2, \ldots, m-1$) from

$$
\begin{aligned}
\alpha_2(t) + 3\alpha_3(t)\lambda_1 + \cdots + \frac{(m-1)(m-2)}{2}\alpha_{m-1}(t)\lambda_1^{m-3} &= \frac{t^2}{2}e^{\lambda_1 t} \\
\alpha_1(t) + 2\alpha_2(t)\lambda_1 + 3\alpha_3(t)\lambda_1^2 + \cdots + (m-1)\alpha_{m-1}(t)\lambda_1^{m-2} &= te^{\lambda_1 t} \\
\alpha_0(t) + \alpha_1(t)\lambda_1 + \alpha_2(t)\lambda_1^2 + \cdots + \alpha_{m-1}(t)\lambda_1^{m-1} &= e^{\lambda_1 t} \\
\alpha_0(t) + \alpha_1(t)\lambda_4 + \alpha_2(t)\lambda_4^2 + \cdots + \alpha_{m-1}(t)\lambda_4^{m-1} &= e^{\lambda_4 t} \\
&\vdots \\
\alpha_0(t) + \alpha_1(t)\lambda_m + \alpha_2(t)\lambda_m^2 + \cdots + \alpha_{m-1}(t)\lambda_m^{m-1} &= e^{\lambda_m t}
\end{aligned}
$$

The extension to other cases where, for example, there are two or more sets of multiple roots will be apparent. Note that if the minimal polynomial of $\mathbf{A}$ is not found, it is possible to substitute the characteristic polynomial for the minimal polynomial. The number of computations may, of course, be increased.

EXAMPLE 11–8 Consider the matrix

$$
\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix}
$$

Compute $e^{\mathbf{A}t}$ using Sylvester's interpolation formula.

From Equation (11–47), we get

$$
\begin{vmatrix} 1 & \lambda_1 & e^{\lambda_1 t} \\ 1 & \lambda_2 & e^{\lambda_2 t} \\ \mathbf{I} & \mathbf{A} & e^{\mathbf{A}t} \end{vmatrix} = \mathbf{0}
$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Substituting 0 for λ_1 and -2 for λ_2 in this last equation, we obtain

$$\begin{vmatrix} 1 & 0 & 1 \\ 1 & -2 & e^{-2t} \\ \mathbf{I} & \mathbf{A} & e^{\mathbf{A}t} \end{vmatrix} = \mathbf{0}$$

Expanding the determinant, we obtain

$$-2e^{\mathbf{A}t} + \mathbf{A} + 2\mathbf{I} - \mathbf{A}e^{-2t} = \mathbf{0}$$

or

$$\begin{aligned} e^{\mathbf{A}t} &= \tfrac{1}{2}(\mathbf{A} + 2\mathbf{I} - \mathbf{A}e^{-2t}) \\ &= \frac{1}{2}\left\{\begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} + \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & -2 \end{bmatrix} e^{-2t}\right\} \\ &= \begin{bmatrix} 1 & \frac{1}{2}(1 - e^{-2t}) \\ 0 & e^{-2t} \end{bmatrix} \end{aligned}$$

An alternative approach is to use Equation (11–48). We first determine $\alpha_0(t)$ and $\alpha_1(t)$ from

$$\alpha_0(t) + \alpha_1(t)\lambda_1 = e^{\lambda_1 t}$$

$$\alpha_0(t) + \alpha_1(t)\lambda_2 = e^{\lambda_2 t}$$

Since $\lambda_1 = 0$ and $\lambda_2 = -2$, the last two equations become

$$\alpha_0(t) = 1$$

$$\alpha_0(t) - 2\alpha_1(t) = e^{-2t}$$

Solving for $\alpha_0(t)$ and $\alpha_1(t)$ gives

$$\alpha_0(t) = 1, \qquad \alpha_1(t) = \frac{1}{2}(1 - e^{-2t})$$

Then $e^{\mathbf{A}t}$ can be written as

$$e^{\mathbf{A}t} = \alpha_0(t)\mathbf{I} + \alpha_1(t)\mathbf{A} = \mathbf{I} + \frac{1}{2}(1 - e^{-2t})\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2}(1 - e^{-2t}) \\ 0 & e^{-2t} \end{bmatrix}$$

Linear Independence of Vectors. The vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be linearly independent if

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n = \mathbf{0}$$

where $c_1, c_2, \ldots, c_n$ are constants, implies that

$$c_1 = c_2 = \cdots = c_n = 0$$

Conversely, the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be linearly dependent if and only if $\mathbf{x}_i$ can be expressed as a linear combination of $\mathbf{x}_j$ $(j = 1, 2, \ldots, n;\ j \neq i)$, or

$$\mathbf{x}_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} c_j\mathbf{x}_j$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

for some set of constants c_j. This means that if $\mathbf{x}_i$ can be expressed as a linear combination of the other vectors in the set, it is linearly dependent on them or it is not an independent member of the set.

EXAMPLE 11–9 The vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 2 \\ 4 \end{bmatrix}$$

are linearly dependent since

$$\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_3 = \mathbf{0}$$

The vectors

$$\mathbf{y}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{y}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{y}_3 = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

are linearly independent since

$$c_1\mathbf{y}_1 + c_2\mathbf{y}_2 + c_3\mathbf{y}_3 = \mathbf{0}$$

implies that

$$c_1 = c_2 = c_3 = 0$$

Note that if an $n \times n$ matrix is nonsingular (that is, the matrix is of rank n or the determinant is nonzero) then n column (or row) vectors are linearly independent. If the $n \times n$ matrix is singular (that is, the rank of the matrix is less than n or the determinant is zero), then n column (or row) vectors are linearly dependent. To demonstrate this, notice that

$$[\mathbf{x}_1 \mid \mathbf{x}_2 \mid \mathbf{x}_3] = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 2 \\ 3 & 1 & 4 \end{bmatrix} = \text{singular}$$

$$[\mathbf{y}_1 \mid \mathbf{y}_2 \mid \mathbf{y}_3] = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 2 \\ 3 & 1 & 2 \end{bmatrix} = \text{nonsingular}$$

11–6 CONTROLLABILITY

Controllability and Observability. A system is said to be controllable at time t_0 if it is possible by means of an unconstrained control vector to transfer the system from any initial state $\mathbf{x}(t_0)$ to any other state in a finite interval of time.

A system is said to be observable at time t_0 if, with the system in state $\mathbf{x}(t_0)$, it is possible to determine this state from the observation of the output over a finite time interval.

The concepts of controllability and observability were introduced by Kalman. They play an important role in the design of control systems in state space. In fact, the conditions of controllability and observability may govern the existence of a complete solution to the control system design problem. The solution to this problem may not

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

exist if the system considered is not controllable. Although most physical systems are controllable and observable, corresponding mathematical models may not possess the property of controllability and observability. Then it is necessary to know the conditions under which a system is controllable and observable. This section deals with controllability and the next section discusses observability.

In what follows, we shall first derive the condition for complete state controllability. Then we derive alternative forms of the condition for complete state controllability followed by discussions of complete output controllability. Finally, we present the concept of stabilizability.

Complete State Controllability of Continuous-Time Systems. Consider the continuous-time system.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}u \tag{11–51}$$

where $\mathbf{x}$ = state vector (n-vector)
u = control signal (scalar)
$\mathbf{A} = n \times n$ matrix
$\mathbf{B} = n \times 1$ matrix

The system described by Equation (11–51) is said to be state controllable at $t = t_0$ if it is possible to construct an unconstrained control signal that will transfer an initial state to any final state in a finite time interval $t_0 \le t \le t_1$. If every state is controllable, then the system is said to be completely state controllable.

We shall now derive the condition for complete state controllability. Without loss of generality, we can assume that the final state is the origin of the state space and that the initial time is zero, or $t_0 = 0$.

The solution of Equation (11–51) is

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

Applying the definition of complete state controllability just given, we have

$$\mathbf{x}(t_1) = \mathbf{0} = e^{\mathbf{A}t_1}\mathbf{x}(0) + \int_0^{t_1} e^{\mathbf{A}(t_1-\tau)}\mathbf{B}u(\tau)\,d\tau$$

or

$$\mathbf{x}(0) = -\int_0^{t_1} e^{-\mathbf{A}\tau}\mathbf{B}u(\tau)\,d\tau \tag{11–52}$$

Referring to Equation (11–48) or (11–50), $e^{-\mathbf{A}\tau}$ can be written

$$e^{-\mathbf{A}\tau} = \sum_{k=0}^{n-1} \alpha_k(\tau)\mathbf{A}^k \tag{11–53}$$

Substituting Equation (11–53) into Equation (11–52) gives

$$\mathbf{x}(0) = -\sum_{k=0}^{n-1} \mathbf{A}^k\mathbf{B}\int_0^{t_1} \alpha_k(\tau)u(\tau)\,d\tau \tag{11–54}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Let us put

$$\int_0^{t_1} \alpha_k(\tau)u(\tau)\,d\tau = \beta_k$$

Then Equation (11–54) becomes

$$\mathbf{x}(0) = -\sum_{k=0}^{n-1} \mathbf{A}^k\mathbf{B}\beta_k$$

$$= -\left[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}\right]\begin{bmatrix} \beta_0 \\ \hdashline \beta_1 \\ \hdashline \cdot \\ \cdot \\ \cdot \\ \hdashline \beta_{n-1} \end{bmatrix} \tag{11–55}$$

If the system is completely state controllable, then, given any initial state $\mathbf{x}(0)$, Equation (11–55) must be satisfied. This requires that the rank of the $n \times n$ matrix

$$\left[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}\right]$$

be n.

From this analysis, we can state the condition for complete state controllability as follows: The system given by Equation (11–51) is completely state controllable if and only if the vectors $\mathbf{B}, \mathbf{AB}, \ldots, \mathbf{A}^{n-1}\mathbf{B}$ are linearly independent, or the $n \times n$ matrix

$$\left[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}\right]$$

is of rank n.

The result just obtained can be extended to the case where the control vector $\mathbf{u}$ is r-dimensional. If the system is described by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

where $\mathbf{u}$ is an r-vector, then it can be proved that the condition for complete state controllability is that the $n \times nr$ matrix

$$\left[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}\right]$$

be of rank n, or contain n linearly independent column vectors. The matrix

$$\left[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}\right]$$

is commonly called the *controllability matrix*.

EXAMPLE 11–10 Consider the system given by

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

Since

$$\left[\mathbf{B} \mid \mathbf{AB}\right] = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \text{singular}$$

the system is not completely state controllable.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11–11 Consider the system given by

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

For this case,

$$[\mathbf{B} \mid \mathbf{AB}] = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} = \text{nonsingular}$$

The system is therefore completely state controllable.

Alternative Form of the Condition for Complete State Controllability. Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \tag{11–56}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{u}$ = control vector (r-vector)
$\mathbf{A} = n \times n$ matrix
$\mathbf{B} = n \times r$ matrix

If the eigenvectors of $\mathbf{A}$ are distinct, then it is possible to find a transformation matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{AP} = \mathbf{D} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

Note that if the eigenvalues of $\mathbf{A}$ are distinct, then the eigenvectors of $\mathbf{A}$ are distinct; however, the converse is not true. For example, an $n \times n$ real symmetric matrix having multiple eigenvalues has n distinct eigenvectors. Note also that each column of the $\mathbf{P}$ matrix is an eigenvector of $\mathbf{A}$ associated with λ_i ($i = 1, 2, \ldots, n$).

Let us define

$$\mathbf{x} = \mathbf{Pz} \tag{11–57}$$

Substituting Equation (11–57) into Equation (11–56), we obtain

$$\dot{\mathbf{z}} = \mathbf{P}^{-1}\mathbf{APz} + \mathbf{P}^{-1}\mathbf{Bu} \tag{11–58}$$

By defining

$$\mathbf{P}^{-1}\mathbf{B} = \mathbf{F} = (f_{ij})$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

we can rewrite Equation (11–58) as

$$\begin{aligned}
\dot{z}_1 &= \lambda_1 z_1 + f_{11}u_1 + f_{12}u_2 + \cdots + f_{1r}u_r \\
\dot{z}_2 &= \lambda_2 z_2 + f_{21}u_1 + f_{22}u_2 + \cdots + f_{2r}u_r \\
&\;\;\vdots \\
\dot{z}_n &= \lambda_n z_n + f_{n1}u_1 + f_{n2}u_2 + \cdots + f_{nr}u_r
\end{aligned}$$

If the elements of any one row of the $n \times r$ matrix $\mathbf{F}$ are all zero, then the corresponding state variable cannot be controlled by any of the u_i. Hence, the condition of complete state controllability is that if the eigenvectors of $\mathbf{A}$ are distinct, then the system is completely state controllable if and only if no row of $\mathbf{P}^{-1}\mathbf{B}$ has all zero elements. It is important to note that, to apply this condition for complete state controllability, we must put the matrix $\mathbf{P}^{-1}\mathbf{AP}$ in Equation (11–58) in diagonal form.

If the $\mathbf{A}$ matrix in Equation (11–56) does not possess distinct eigenvectors, then diagonalization is impossible. In such a case, we may transform $\mathbf{A}$ into a Jordan canonical form. If, for example, $\mathbf{A}$ has eigenvalues $\lambda_1, \lambda_1, \lambda_1, \lambda_4, \lambda_4, \lambda_6, \ldots, \lambda_n$ and has $n - 3$ distinct eigenvectors, then the Jordan canonical form of $\mathbf{A}$ is

$$\mathbf{J} = \left[\begin{array}{ccc|cc|c|ccc}
\lambda_1 & 1 & 0 & & & & & & 0 \\
0 & \lambda_1 & 1 & & & & & & \\
0 & 0 & \lambda_1 & & & & & & \\
\hline
 & & & \lambda_4 & 1 & & & & \\
 & & & 0 & \lambda_4 & & & & \\
\hline
 & & & & & \lambda_6 & & & \\
 & & & & & & \ddots & & \\
 & & & & & & & \ddots & \\
0 & & & & & & & & \lambda_n
\end{array}\right]$$

The square submatrices on the main diagonal are called *Jordan blocks.*

Suppose that we can find a transformation matrix $\mathbf{S}$ such that

$$\mathbf{S}^{-1}\mathbf{AS} = \mathbf{J}$$

If we define a new state vector $\mathbf{z}$ by

$$\mathbf{x} = \mathbf{Sz} \tag{11–59}$$

then substitution of Equation (11–59) into Equation (11–56) yields

$$\begin{aligned}
\dot{\mathbf{z}} &= \mathbf{S}^{-1}\mathbf{ASz} + \mathbf{S}^{-1}\mathbf{Bu} \\
&= \mathbf{Jz} + \mathbf{S}^{-1}\mathbf{Bu} \qquad (11\text{–}60)
\end{aligned}$$

The condition for complete state controllability of the system of Equation (11–56) may then be stated as follows: The system is completely state controllable if and only if (1)

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

no two Jordan blocks in $\mathbf{J}$ of Equation (11–60) are associated with the same eigenvalues, (2) the elements of any row of $\mathbf{S}^{-1}\mathbf{B}$ that correspond to the last row of each Jordan block are not all zero, and (3) the elements of each row of $\mathbf{S}^{-1}\mathbf{B}$ that correspond to distinct eigenvalues are not all zero.

EXAMPLE 11–12 The following systems are completely state controllable:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 2 \\ 5 \end{bmatrix} u$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 4 \\ 3 \end{bmatrix} u$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \left[\begin{array}{ccc|cc} -2 & 1 & 0 & & 0 \\ 0 & -2 & 1 & & \\ 0 & 0 & -2 & & \\ \hline & & & -5 & 1 \\ 0 & & & 0 & -5 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 3 & 0 \\ 0 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

The following systems are not completely state controllable:

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

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 4 & 2 \\ 0 & 0 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \left[\begin{array}{ccc|cc} -2 & 1 & 0 & & 0 \\ 0 & -2 & 1 & & \\ 0 & 0 & -2 & & \\ \hline & & & -5 & 1 \\ 0 & & & 0 & -5 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} + \begin{bmatrix} 4 \\ 2 \\ 1 \\ 3 \\ 0 \end{bmatrix} u$$

Condition for Complete State Controllability in the s Plane. The condition for complete state controllability can be stated in terms of transfer functions or transfer matrices.

It can be proved that a necessary and sufficient condition for complete state controllability is that no cancellation occur in the transfer function or transfer matrix. If cancellation occurs, the system cannot be controlled in the direction of the canceled mode.

EXAMPLE 11–13 Consider the following transfer function:

$$\frac{X(s)}{U(s)} = \frac{s + 2.5}{(s + 2.5)(s - 1)}$$

Clearly, cancellation of the factor $(s + 2.5)$ occurs in the numerator and denominator of this transfer function. (Thus one degree of freedom is lost.) Because of this cancellation, this system is not completely state controllable.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The same conclusion can be obtained by writing this transfer function in the form of a state equation. A state-space representation is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2.5 & -1.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u$$

Since

$$[\mathbf{B} \mid \mathbf{AB}] = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

the rank of the matrix $[\mathbf{B} \mid \mathbf{AB}]$ is 1. Therefore, we arrive at the same conclusion: The system is not completely state controllable.

Output Controllability. In the practical design of a control system, we may want to control the output rather than the state of the system. Complete state controllability is neither necessary nor sufficient for controlling the output of the system. For this reason, it is desirable to define separately complete output controllability.

Consider the system described by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu} \tag{11–61}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du} \tag{11–62}$$

where $\mathbf{x}$ = state vector (n-vector)

$\mathbf{u}$ = control vector (r-vector)

$\mathbf{y}$ = output vector (m-vector)

$\mathbf{A} = n \times n$ matrix

$\mathbf{B} = n \times r$ matrix

$\mathbf{C} = m \times n$ matrix

$\mathbf{D} = m \times r$ matrix

The system described by Equations (11–61) and (11–62) is said to be completely output controllable if it is possible to construct an unconstrained control vector $\mathbf{u}(t)$ that will transfer any given initial output $\mathbf{y}(t_0)$ to any final output $\mathbf{y}(t_1)$ in a finite time interval $t_0 \le t \le t_1$.

It can be proved that the condition for complete output controllability is as follows: The system described by Equations (11–61) and (11–62) is completely output controllable if and only if the $m \times (n+1)r$ matrix

$$[\mathbf{CB} \mid \mathbf{CAB} \mid \mathbf{CA}^2\mathbf{B} \mid \cdots \mid \mathbf{CA}^{n-1}\mathbf{B} \mid \mathbf{D}]$$

is of rank m. (For a proof, see Problem **A–11–16**.) Note that the presence of the $\mathbf{Du}$ term in Equation (11–62) always helps to establish output controllability.

Uncontrollable System. An uncontrollable system has a subsystem that is physically disconnected from the input.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Stabilizability. For a partially controllable system, if the uncontrollable modes are stable and the unstable modes are controllable, the system is said to be stabilizable. For example, the system defined by

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

is not state controllable. The stable mode that corresponds to the eigenvalue of -1 is not controllable. The unstable mode that corresponds to the eigenvalue of 1 is controllable. Such a system can be made stable by the use of a suitable feedback. Thus this system is stabilizable.

11–7 OBSERVABILITY

In this section we discuss the observability of linear systems. Consider the unforced system described by the following equations:

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{11–63}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{11–64}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{y}$ = output vector (m-vector)
$\mathbf{A} = n \times n$ matrix
$\mathbf{C} = m \times n$ matrix

The system is said to be completely observable if every state $\mathbf{x}(t_0)$ can be determined from the observation of $\mathbf{y}(t)$ over a finite time interval, $t_0 \le t \le t_1$. The system is, therefore, completely observable if every transition of the state eventually affects every element of the output vector. The concept of observability is useful in solving the problem of reconstructing unmeasurable state variables from measurable variables in the minimum possible length of time. In this section we treat only linear, time-invariant systems. Therefore, without loss of generality, we can assume that $t_0 = 0$.

The concept of observability is very important because, in practice, the difficulty encountered with state feedback control is that some of the state variables are not accessible for direct measurement, with the result that it becomes necessary to estimate the unmeasurable state variables in order to construct the control signals. It will be shown in Section 12–5 that such estimates of state variables are possible if and only if the system is completely observable.

In discussing observability conditions, we consider the unforced system as given by Equations (11–63) and (11–64). The reason for this is as follows: If the system is described by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$$

then

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

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

and $\mathbf{y}(t)$ is

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

Since the matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are known and $\mathbf{u}(t)$ is also known, the last two terms on the right-hand side of this last equation are known quantities. Therefore, they may be subtracted from the observed value of $\mathbf{y}(t)$. Hence, for investigating a necessary and sufficient condition for complete observability, it suffices to consider the system described by Equations (11–63) and (11–64).

Complete Observability of Continuous-Time Systems. Consider the system described by Equations (11–63) and (11–64). The output vector $\mathbf{y}(t)$ is

$$\mathbf{y}(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{x}(0)$$

Referring to Equation (11–48) or (11–50), we have

$$e^{\mathbf{A}t} = \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{A}^k$$

Hence, we obtain

$$\mathbf{y}(t) = \sum_{k=0}^{n-1} \alpha_k(t)\mathbf{C}\mathbf{A}^k\mathbf{x}(0)$$

or

$$\mathbf{y}(t) = \alpha_0(t)\mathbf{C}\mathbf{x}(0) + \alpha_1(t)\mathbf{C}\mathbf{A}\mathbf{x}(0) + \cdots + \alpha_{n-1}(t)\mathbf{C}\mathbf{A}^{n-1}\mathbf{x}(0) \qquad (11\text{–}65)$$

If the system is completely observable, then, given the output $\mathbf{y}(t)$ over a time interval $0 \le t \le t_1$, $\mathbf{x}(0)$ is uniquely determined from Equation (11–65). It can be shown that this requires the rank of the $nm \times n$ matrix

$$\begin{bmatrix} \mathbf{C} \\ \hdashline \mathbf{C}\mathbf{A} \\ \hdashline \cdot \\ \cdot \\ \cdot \\ \hdashline \mathbf{C}\mathbf{A}^{n-1} \end{bmatrix}$$

to be n. (See Problem **A–11–19** for the derivation of this condition.)

From this analysis, we can state the condition for complete observability as follows: The system described by Equations (11–63) and (11–64) is completely observable if and only if the $n \times nm$ matrix

$$[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^* \mid \cdots \mid (\mathbf{A}^*)^{n-1}\mathbf{C}^*]$$

is of rank n or has n linearly independent column vectors. This matrix is called the *observability matrix.*

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

EXAMPLE 11–14 Consider the system described by

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

$$y = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Is this system controllable and observable?

Since the rank of the matrix

$$[\mathbf{B} \mid \mathbf{AB}] = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix}$$

is 2, the system is completely state controllable.

For output controllability, let us find the rank of the matrix $[\mathbf{CB} \mid \mathbf{CAB}]$. Since

$$[\mathbf{CB} \mid \mathbf{CAB}] = [0 \quad 1]$$

the rank of this matrix is 1. Hence, the system is completely output controllable.

To test the observability condition, examine the rank of $[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^*]$. Since

$$[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^*] = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

the rank of $[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^*]$ is 2. Hence, the system is completely observable.

Conditions for Complete Observability in the s Plane. The conditions for complete observability can also be stated in terms of transfer functions or transfer matrices. The necessary and sufficient conditions for complete observability is that no cancellation occur in the transfer function or transfer matrix. If cancellation occurs, the canceled mode cannot be observed in the output.

EXAMPLE 11–15 Show that the following system is not completely observable.

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{B}u$$
$$y = \mathbf{Cx}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -6 & -11 & -6 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{C} = [4 \quad 5 \quad 1]$$

Note that the control function u does not affect the complete observability of the system. To examine complete observability, we may simply set $u = 0$. For this system, we have

$$[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^* \mid (\mathbf{A}^*)^2\mathbf{C}^*] = \begin{bmatrix} 4 & -6 & 6 \\ 5 & -7 & 5 \\ 1 & -1 & -1 \end{bmatrix}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Note that

$$\begin{vmatrix} 4 & -6 & 6 \\ 5 & -7 & 5 \\ 1 & -1 & -1 \end{vmatrix} = 0$$

Hence, the rank of the matrix $[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^* \mid (\mathbf{A}^*)^2\mathbf{C}^*]$ is less than 3. Therefore, the system is not completely observable.

In fact, in this system, cancellation occurs in the transfer function of the system. The transfer function between $X_1(s)$ and $U(s)$ is

$$\frac{X_1(s)}{U(s)} = \frac{1}{(s+1)(s+2)(s+3)}$$

and the transfer function between $Y(s)$ and $X_1(s)$ is

$$\frac{Y(s)}{X_1(s)} = (s+1)(s+4)$$

Therefore, the transfer function between the output $Y(s)$ and the input $U(s)$ is

$$\frac{Y(s)}{U(s)} = \frac{(s+1)(s+4)}{(s+1)(s+2)(s+3)}$$

Clearly, the two factors $(s+1)$ cancel each other. This means that there are nonzero initial states $\mathbf{x}(0)$, which cannot be determined from the measurement of $y(t)$.

Comments. The transfer function has no cancellation if and only if the system is completely state controllable and completely observable. This means that the canceled transfer function does not carry along all the information characterizing the dynamic system.

Alternative Form of the Condition for Complete Observability. Consider the system described by Equations (11–63) and (11–64), rewritten

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} \tag{11–66}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x} \tag{11–67}$$

Suppose that the transformation matrix $\mathbf{P}$ transforms $\mathbf{A}$ into a diagonal matrix, or

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

where $\mathbf{D}$ is a diagonal matrix. Let us define

$$\mathbf{x} = \mathbf{P}\mathbf{z}$$

Then Equations (11–66) and (11–67) can be written

$$\dot{\mathbf{z}} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}\mathbf{z} = \mathbf{D}\mathbf{z}$$

$$\mathbf{y} = \mathbf{C}\mathbf{P}\mathbf{z}$$

Hence,

$$\mathbf{y}(t) = \mathbf{C}\mathbf{P}e^{\mathbf{D}t}\mathbf{z}(0)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

or

$$\mathbf{y}(t) = \mathbf{CP}\begin{bmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_n t} \end{bmatrix}\mathbf{z}(0) = \mathbf{CP}\begin{bmatrix} e^{\lambda_1 t}z_1(0) \\ e^{\lambda_2 t}z_2(0) \\ \vdots \\ e^{\lambda_n t}z_n(0) \end{bmatrix}$$

The system is completely observable if none of the columns of the $m \times n$ matrix **CP** consists of all zero elements. This is because, if the ith column of **CP** consists of all zero elements, then the state variable $z_i(0)$ will not appear in the output equation and therefore cannot be determined from observation of $\mathbf{y}(t)$. Thus, $\mathbf{x}(0)$, which is related to $\mathbf{z}(0)$ by the nonsingular matrix **P**, cannot be determined. (Remember that this test applies only if the matrix $\mathbf{P}^{-1}\mathbf{AP}$ is in diagonal form.)

If the matrix **A** cannot be transformed into a diagonal matrix, then by use of a suitable transformation matrix **S**, we can transform **A** into a Jordan canonical form, or

$$\mathbf{S}^{-1}\mathbf{AS} = \mathbf{J}$$

where **J** is in the Jordan canonical form.

Let us define

$$\mathbf{x} = \mathbf{Sz}$$

Then Equations (11–66) and (11–67) can be written

$$\dot{\mathbf{z}} = \mathbf{S}^{-1}\mathbf{ASz} = \mathbf{Jz}$$
$$\mathbf{y} = \mathbf{CSz}$$

Hence,

$$\mathbf{y}(t) = \mathbf{CS}e^{\mathbf{J}t}\mathbf{z}(0)$$

The system is completely observable if (1) no two Jordan blocks in **J** are associated with the same eigenvalues, (2) no columns of **CS** that correspond to the first row of each Jordan block consist of zero elements, and (3) no columns of **CS** that correspond to distinct eigenvalues consist of zero elements.

To clarify condition (2), in Example 11–16 we have encircled by dashed lines the columns of **CS** that correspond to the first row of each Jordan block.

EXAMPLE 11–16 The following systems are completely observable.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 4 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \left[\begin{array}{ccc:cc} 2 & 1 & 0 & & 0 \\ 0 & 2 & 1 & & \\ 0 & 0 & 2 & & \\ \hdashline & & & -3 & 1 \\ 0 & & & 0 & -3 \end{array}\right]\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The following systems are not completely observable.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \qquad y = \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 \\ 0 & 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \dot{x}_5 \end{bmatrix} = \left[\begin{array}{ccc|cc} 2 & 1 & 0 & & 0 \\ 0 & 2 & 1 & & \\ 0 & 0 & 2 & & \\ \hline & & & -3 & 1 \\ 0 & & & 0 & -3 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}, \qquad \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

Principle of Duality. We shall now discuss the relationship between controllability and observability. We shall introduce the principle of duality, due to Kalman, to clarify apparent analogies between controllability and observability.

Consider the system S_1 described by

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$$
$$\mathbf{y} = \mathbf{C}\mathbf{x}$$

where $\mathbf{x}$ = state vector (n-vector)
$\mathbf{u}$ = control vector (r-vector)
$\mathbf{y}$ = output vector (m-vector)
$\mathbf{A} = n \times n$ matrix
$\mathbf{B} = n \times r$ matrix
$\mathbf{C} = m \times n$ matrix

and the dual system S_2 defined by

$$\dot{\mathbf{z}} = \mathbf{A}^*\mathbf{z} + \mathbf{C}^*\mathbf{v}$$
$$\mathbf{n} = \mathbf{B}^*\mathbf{z}$$

where $\mathbf{z}$ = state vector (n-vector)
$\mathbf{v}$ = control vector (m-vector)
$\mathbf{n}$ = output vector (r-vector)
$\mathbf{A}^*$ = conjugate transpose of $\mathbf{A}$
$\mathbf{B}^*$ = conjugate transpose of $\mathbf{B}$
$\mathbf{C}^*$ = conjugate transpose of $\mathbf{C}$

The principle of duality states that the system S_1 is completely state controllable (observable) if and only if system S_2 is completely observable (state controllable).

To verify this principle, let us write down the necessary and sufficient conditions for complete state controllability and complete observability for systems S_1 and S_2.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

For system S_1:

1. A necessary and sufficient condition for complete state controllability is that the rank of the $n \times nr$ matrix

$$[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}]$$

be n.

2. A necessary and sufficient condition for complete observability is that the rank of the $n \times nm$ matrix

$$[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^* \mid \cdots \mid (\mathbf{A}^*)^{n-1}\mathbf{C}^*]$$

be n.

For system S_2:

1. A necessary and sufficient condition for complete state controllability is that the rank of the $n \times nm$ matrix

$$[\mathbf{C}^* \mid \mathbf{A}^*\mathbf{C}^* \mid \cdots \mid (\mathbf{A}^*)^{n-1}\mathbf{C}^*]$$

be n.

2. A necessary and sufficient condition for complete observability is that the rank of the $n \times nr$ matrix

$$[\mathbf{B} \mid \mathbf{AB} \mid \cdots \mid \mathbf{A}^{n-1}\mathbf{B}]$$

be n.

By comparing these conditions, the truth of this principle is apparent. By use of this principle, the observability of a given system can be checked by testing the state controllability of its dual.

Detectability. For a partially observable system, if the unobservable modes are stable and the observable modes are unstable, the system is said to be detectable. Note that the concept of detectability is dual to the concept of stabilizability.

EXAMPLE PROBLEMS AND SOLUTIONS

A–11–1. Consider the transfer function system defined by Equation (11–2), rewritten

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \tag{11–68}$$

Derive the following controllable canonical form of the state-space representation for this transfer function system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_1 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_{n-1} \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 1 \end{bmatrix} u \tag{11–69}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

$$y = \left[b_n - a_n b_0 \;\vdots\; b_{n-1} - a_{n-1} b_0 \;\vdots\; \cdots \;\vdots\; b_1 - a_1 b_0\right] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + b_0 u \tag{11–70}$$

Solution. Equation (11–68) can be written as

$$\frac{Y(s)}{U(s)} = b_0 + \frac{(b_1 - a_1 b_0)s^{n-1} + \cdots + (b_{n-1} - a_{n-1} b_0)s + (b_n - a_n b_0)}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n}$$

which can be modified to

$$Y(s) = b_0 U(s) + \hat{Y}(s) \tag{11–71}$$

where

$$\hat{Y}(s) = \frac{(b_1 - a_1 b_0)s^{n-1} + \cdots + (b_{n-1} - a_{n-1} b_0)s + (b_n - a_n b_0)}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n} U(s)$$

Let us rewrite this last equation in the following form:

$$\frac{\hat{Y}(s)}{(b_1 - a_1 b_0)s^{n-1} + \cdots + (b_{n-1} - a_{n-1} b_0)s + (b_n - a_n b_0)} = \frac{U(s)}{s^n + a_1 s^{n-1} + \cdots + a_{n-1} s + a_n} = Q(s)$$

From this last equation, the following two equations may be obtained:

$$s^n Q(s) = -a_1 s^{n-1} Q(s) - \cdots - a_{n-1} s Q(s) - a_n Q(s) + U(s) \tag{11–72}$$

$$\hat{Y}(s) = (b_1 - a_1 b_0)s^{n-1} Q(s) + \cdots + (b_{n-1} - a_{n-1} b_0)s Q(s) + (b_n - a_n b_0)Q(s) \tag{11–73}$$

Now define state variables as follows:

$$\begin{aligned} X_1(s) &= Q(s) \\ X_2(s) &= sQ(s) \\ &\cdot \\ &\cdot \\ &\cdot \\ X_{n-1}(s) &= s^{n-2} Q(s) \\ X_n(s) &= s^{n-1} Q(s) \end{aligned}$$

Then, clearly,

$$\begin{aligned} sX_1(s) &= X_2(s) \\ sX_2(s) &= X_3(s) \\ &\cdot \\ &\cdot \\ &\cdot \\ sX_{n-1}(s) &= X_n(s) \end{aligned}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

which may be rewritten as

$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= x_3 \\ &\cdot \\ &\cdot \\ &\cdot \\ \dot{x}_{n-1} &= x_n \end{aligned} \tag{11–74}$$

Noting that $s^n Q(s) = sX_n(s)$, we can rewrite Equation (11–72) as

$$sX_n(s) = -a_1 X_n(s) - \cdots - a_{n-1}X_2(s) - a_n X_1(s) + U(s)$$

or

$$\dot{x}_n = -a_n x_1 - a_{n-1}x_2 - \cdots - a_1 x_n + u \tag{11–75}$$

Also, from Equations (11–71) and (11–73), we obtain

$$\begin{aligned} Y(s) &= b_0 U(s) + (b_1 - a_1 b_0)s^{n-1}Q(s) + \cdots + (b_{n-1} - a_{n-1}b_0)sQ(s) \\ &\quad + (b_n - a_n b_0)Q(s) \\ &= b_0 U(s) + (b_1 - a_1 b_0)X_n(s) + \cdots + (b_{n-1} - a_{n-1}b_0)X_2(s) \\ &\quad + (b_n - a_n b_0)X_1(s) \end{aligned}$$

The inverse Laplace transform of this output equation becomes

$$y = (b_n - a_n b_0)x_1 + (b_{n-1} - a_{n-1}b_0)x_2 + \cdots + (b_1 - a_1 b_0)x_n + b_0 u \tag{11–76}$$

Combining Equations (11–74) and (11–75) into one vector–matrix differential equation, we obtain Equation (11–69). Equation (11–76) can be rewritten as given by Equation (11–70). Equations (11–69) and (11–70) are said to be in the controllable canonical form. Figure 11–1 shows the block diagram representation of the system defined by Equations (11–69) and (11–70).

A–11–2. Consider the following transfer function system:

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{s^n + a_1 s^{n-1} + \cdots + a_{n-1}s + a_n} \tag{11–77}$$

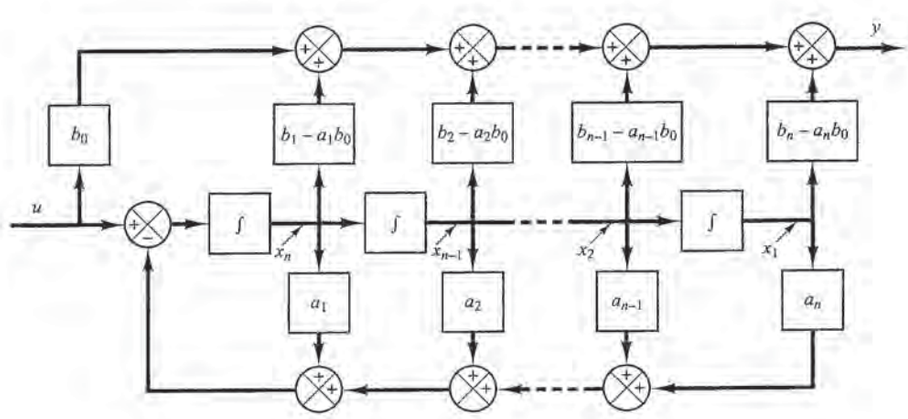


Figure 11–1 Block diagram representation of the system defined by Equations (11–69) and (11–70) (controllable canonical form)

0807

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Derive the following observable canonical form of the state-space representation for this transfer function system:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ 0 & 0 & \cdots & 1 & -a_1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} b_n - a_n b_0 \\ b_{n-1} - a_{n-1} b_0 \\ \cdot \\ \cdot \\ \cdot \\ b_1 - a_1 b_0 \end{bmatrix} u \tag{11–78}$$

$$y = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_{n-1} \\ x_n \end{bmatrix} + b_0 u \tag{11–79}$$

Solution. Equation (11–77) can be modified into the following form:

$$s^n[Y(s) - b_0U(s)] + s^{n-1}[a_1Y(s) - b_1U(s)] + \cdots$$
$$+ s[a_{n-1}Y(s) - b_{n-1}U(s)] + a_nY(s) - b_nU(s) = 0$$

By dividing the entire equation by s^n and rearranging, we obtain

$$Y(s) = b_0U(s) + \frac{1}{s}[b_1U(s) - a_1Y(s)] + \cdots$$
$$+ \frac{1}{s^{n-1}}[b_{n-1}U(s) - a_{n-1}Y(s)] + \frac{1}{s^n}[b_nU(s) - a_nY(s)] \tag{11–80}$$

Now define state variables as follows:

$$\begin{aligned} X_n(s) &= \frac{1}{s}[b_1U(s) - a_1Y(s) + X_{n-1}(s)] \\ X_{n-1}(s) &= \frac{1}{s}[b_2U(s) - a_2Y(s) + X_{n-2}(s)] \\ &\cdot \\ &\cdot \\ &\cdot \\ X_2(s) &= \frac{1}{s}[b_{n-1}U(s) - a_{n-1}Y(s) + X_1(s)] \\ X_1(s) &= \frac{1}{s}[b_nU(s) - a_nY(s)] \end{aligned} \tag{11–81}$$

Then Equation (11–80) can be written as

$$Y(s) = b_0U(s) + X_n(s) \tag{11–82}$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

By substituting Equation (11–82) into Equation (11–81) and multiplying both sides of the equations by s, we obtain

$$sX_n(s) = X_{n-1}(s) - a_1X_n(s) + (b_1 - a_1b_0)U(s)$$
$$sX_{n-1}(s) = X_{n-2}(s) - a_2X_n(s) + (b_2 - a_2b_0)U(s)$$
$$\vdots$$
$$sX_2(s) = X_1(s) - a_{n-1}X_n(s) + (b_{n-1} - a_{n-1}b_0)U(s)$$
$$sX_1(s) = -a_nX_n(s) + (b_n - a_nb_0)U(s)$$

Taking the inverse Laplace transforms of the preceding n equations and writing them in the reverse order, we get

$$\dot{x}_1 = -a_nx_n + (b_n - a_nb_0)u$$
$$\dot{x}_2 = x_1 - a_{n-1}x_n + (b_{n-1} - a_{n-1}b_0)u$$
$$\vdots$$
$$\dot{x}_{n-1} = x_{n-2} - a_2x_n + (b_2 - a_2b_0)u$$
$$\dot{x}_n = x_{n-1} - a_1x_n + (b_1 - a_1b_0)u$$

Also, the inverse Laplace transform of Equation (11–82) gives

$$y = x_n + b_0u$$

Rewriting the state and output equations in the standard vector-matrix forms gives Equations (11–78) and (11–79). Figure 11–2 shows a block diagram representation of the system defined by Equations (11–78) and (11–79).

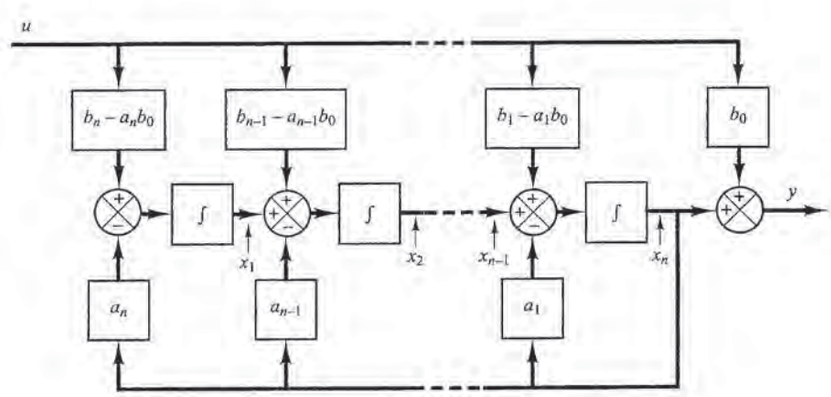


Figure 11–2
Block diagram representation of the system defined by Equations (11–78) and (11–79) (observable canonical form).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

A–11–3. Consider the transfer function system defined by

$$\frac{Y(s)}{U(s)} = \frac{b_0 s^n + b_1 s^{n-1} + \cdots + b_{n-1}s + b_n}{(s+p_1)(s+p_2)\cdots(s+p_n)}$$

$$= b_0 + \frac{c_1}{s+p_1} + \frac{c_2}{s+p_2} + \cdots + \frac{c_n}{s+p_n} \tag{11–83}$$

where $p_i \neq p_j$. Derive the state-space representation of this system in the following diagonal canonical form:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} -p_1 & & & 0 \\ & -p_2 & & \\ & & \ddots & \\ 0 & & & -p_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} u \tag{11–84}$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b_0 u \tag{11–85}$$

Solution. Equation (11–83) may be written as

$$Y(s) = b_0 U(s) + \frac{c_1}{s+p_1}U(s) + \frac{c_2}{s+p_2}U(s) + \cdots + \frac{c_n}{s+p_n}U(s) \tag{11–86}$$

Define the state variables as follows:

$$X_1(s) = \frac{1}{s+p_1}U(s)$$
$$X_2(s) = \frac{1}{s+p_2}U(s)$$
$$\vdots$$
$$X_n(s) = \frac{1}{s+p_n}U(s)$$

which may be rewritten as

$$sX_1(s) = -p_1 X_1(s) + U(s)$$
$$sX_2(s) = -p_2 X_2(s) + U(s)$$
$$\vdots$$
$$sX_n(s) = -p_n X_n(s) + U(s)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The inverse Laplace transforms of these equations give

$$\begin{aligned}\dot{x}_1 &= -p_1x_1 + u\\ \dot{x}_2 &= -p_2x_2 + u\\ &\vdots\\ \dot{x}_n &= -p_nx_n + u\end{aligned} \tag{11–87}$$

These n equations make up a state equation.

In terms of the state variables $X_1(s), X_2(s), \ldots, X_n(s)$, Equation (11–86) can be written as

$$Y(s) = b_0U(s) + c_1X_1(s) + c_2X_2(s) + \cdots + c_nX_n(s)$$

The inverse Laplace transform of this last equation is

$$y = c_1x_1 + c_2x_2 + \cdots + c_nx_n + b_0u \tag{11–88}$$

which is the output equation.

Equation (11–87) can be put in the vector-matrix equation as given by Equation (11–84). Equation (11–88) can be put in the form of Equation (11–85).

Figure 11–3 shows a block diagram representation of the system defined by Equations (11–84) and (11–85).

It is noted that if we choose the state variables as

$$\begin{aligned}\hat{X}_1(s) &= \frac{c_1}{s+p_1}U(s)\\ \hat{X}_2(s) &= \frac{c_2}{s+p_2}U(s)\\ &\vdots\\ \hat{X}_n(s) &= \frac{c_n}{s+p_n}U(s)\end{aligned}$$

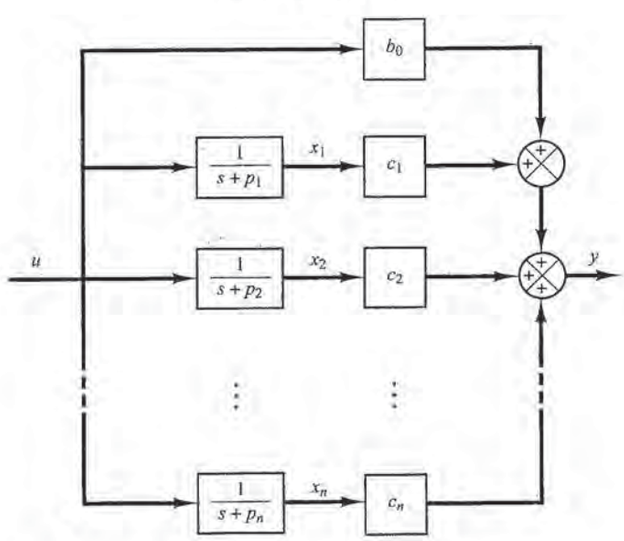


Figure 11–3
Block diagram representation of the system defined by Equations (11–84) and (11–85) (diagonal canonical form).

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

then we get a slightly different state-space representation. This choice of state variables gives

$$
\begin{aligned}
s\hat{X}_1(s) &= -p_1\hat{X}_1(s) + c_1U(s) \\
s\hat{X}_2(s) &= -p_2\hat{X}_2(s) + c_2U(s) \\
&\vdots \\
s\hat{X}_n(s) &= -p_n\hat{X}_n(s) + c_nU(s)
\end{aligned}
$$

from which we obtain

$$
\begin{aligned}
\dot{\hat{x}}_1 &= -p_1\hat{x}_1 + c_1u \\
\dot{\hat{x}}_2 &= -p_2\hat{x}_2 + c_2u \\
&\vdots \\
\dot{\hat{x}}_n &= -p_n\hat{x}_n + c_nu
\end{aligned}
\tag{11–89}
$$

Referring to Equation (11–86), the output equation becomes

$$Y(s) = b_0U(s) + \hat{X}_1(s) + \hat{X}_2(s) + \cdots + \hat{X}_n(s)$$

from which we get

$$y = \hat{x}_1 + \hat{x}_2 + \cdots + \hat{x}_n + b_0u \tag{11–90}$$

Equations (11–89) and (11–90) give the following state-space representation for the system:

$$
\begin{bmatrix} \dot{\hat{x}}_1 \\ \dot{\hat{x}}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{\hat{x}}_n \end{bmatrix}
=
\begin{bmatrix} -p_1 & & & & 0 \\ & -p_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ 0 & & & & -p_n \end{bmatrix}
\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{x}_n \end{bmatrix}
+
\begin{bmatrix} c_1 \\ c_2 \\ \cdot \\ \cdot \\ \cdot \\ c_n \end{bmatrix} u
$$

$$
y = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}
\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \hat{x}_n \end{bmatrix} + b_0u
$$

A–11–4. Consider the system defined by

$$\frac{Y(s)}{U(s)} = \frac{b_0s^n + b_1s^{n-1} + \cdots + b_{n-1}s + b_n}{(s + p_1)^3(s + p_4)(s + p_5)\cdots(s + p_n)} \tag{11–91}$$

where the system involves a triple pole at $s = -p_1$. (We assume that, except for the first three p_i's being equal, the p_i's are different from one another.) Obtain the Jordan canonical form of the state-space representation for this system.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

Solution. The partial-fraction expansion of Equation (11–91) becomes

$$\frac{Y(s)}{U(s)} = b_0 + \frac{c_1}{(s+p_1)^3} + \frac{c_2}{(s+p_1)^2} + \frac{c_3}{s+p_1} + \frac{c_4}{s+p_4} + \cdots + \frac{c_n}{s+p_n}$$

which may be written as

$$Y(s) = b_0U(s) + \frac{c_1}{(s+p_1)^3}U(s) + \frac{c_2}{(s+p_1)^2}U(s) + \frac{c_3}{s+p_1}U(s) + \frac{c_4}{s+p_4}U(s) + \cdots + \frac{c_n}{s+p_n}U(s) \tag{11–92}$$

Define

$$X_1(s) = \frac{1}{(s+p_1)^3}U(s)$$

$$X_2(s) = \frac{1}{(s+p_1)^2}U(s)$$

$$X_3(s) = \frac{1}{s+p_1}U(s)$$

$$X_4(s) = \frac{1}{s+p_4}U(s)$$

$$\vdots$$

$$X_n(s) = \frac{1}{s+p_n}U(s)$$

Notice that the following relationships exist among $X_1(s)$, $X_2(s)$, and $X_3(s)$:

$$\frac{X_1(s)}{X_2(s)} = \frac{1}{s+p_1}$$

$$\frac{X_2(s)}{X_3(s)} = \frac{1}{s+p_1}$$

Then, from the preceding definition of the state variables and the preceding relationships, we obtain

$$sX_1(s) = -p_1X_1(s) + X_2(s)$$

$$sX_2(s) = -p_1X_2(s) + X_3(s)$$

$$sX_3(s) = -p_1X_3(s) + U(s)$$

$$sX_4(s) = -p_4X_4(s) + U(s)$$

$$\vdots$$

$$sX_n(s) = -p_nX_n(s) + U(s)$$

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565

The inverse Laplace transforms of the preceding n equations give

$$\begin{aligned}
\dot{x}_1 &= -p_1 x_1 + x_2 \\
\dot{x}_2 &= -p_1 x_2 + x_3 \\
\dot{x}_3 &= -p_1 x_3 + u \\
\dot{x}_4 &= -p_4 x_4 + u \\
&\vdots \\
\dot{x}_n &= -p_n x_n + u
\end{aligned}$$

The output equation, Equation (11–92), can be rewritten as

$$Y(s) = b_0 U(s) + c_1 X_1(s) + c_2 X_2(s) + c_3 X_3(s) + c_4 X_4(s) + \cdots + c_n X_n(s)$$

The inverse Laplace transform of this output equation is

$$y = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 + \cdots + c_n x_n + b_0 u$$

Thus, the state-space representation of the system for the case when the denominator polynomial involves a triple root $-p_1$ can be given as follows:

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \left[\begin{array}{ccc|ccc} -p_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -p_1 & 1 & \vdots & & \vdots \\ 0 & 0 & -p_1 & 0 & \cdots & 0 \\ \hline 0 & \cdots & 0 & -p_4 & & \\ \vdots & & \vdots & & \ddots & \\ \vdots & & \vdots & & & \ddots \\ 0 & \cdots & 0 & 0 & & -p_n \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} u \tag{11–93}$$

$$y = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + b_0 u \tag{11–94}$$

The state-space representation in the form given by Equations (11–93) and (11–94) is said to be in the Jordan canonical form. Figure 11–4 shows a block-diagram representation of the system given by Equations (11–93) and (11–94).

A–11–5. Consider the transfer function system

$$\frac{Y(s)}{U(s)} = \frac{25.04s + 5.008}{s^3 + 5.03247s^2 + 25.1026s + 5.008}$$

Obtain a state-space representation of this system with MATLAB.

PETITIONER'S EXHIBIT 1002
US PATENT NO. 6,581,565