6

Root-Locus Analysis

6-1 INTRODUCTION

The basic characteristic of the transient response of a closed-loop system is closely related to the location of the closed-loop poles. If the system has a variable loop gain, then the location of the closed-loop poles depends on the value of the loop gain chosen. It is important, therefore, that the designer know how the closed-loop poles move in the s plane as the loop gain is varied.

From the design viewpoint, in some systems simple gain adjustment may move the closed-loop poles to desired locations. Then the design problem may become the selection of an appropriate gain value. If the gain adjustment alone does not yield a desired result, addition of a compensator to the system will become necessary. (This subject is discussed in detail in Chapter 7.)

The closed-loop poles are the roots of the characteristic equation. Finding the roots of the characteristic equation of degree higher than 3 is laborious and will need computer solution. (MATLAB provides a simple solution to this problem.) However, just finding the roots of the characteristic equation may be of limited value, because as the gain of the open-loop transfer function varies the characteristic equation changes and the computations must be repeated.

A simple method for finding the roots of the characteristic equation has been developed by W. R. Evans and used extensively in control engineering. This method, called the *root-locus method,* is one in which the roots of the characteristic equation are plotted for all values of a system parameter. The roots corresponding to a particular value of this parameter can then be located on the resulting graph. Note that the parameter is usually the gain, but any other variable of the open-loop transfer function may be used. Unless otherwise stated, we shall assume that the gain of the open-loop transfer function is the parameter to be varied through all values, from zero to infinity.

By using the root-locus method the designer can predict the effects on the location of the closed-loop poles of varying the gain value or adding open-loop poles and/or open-loop zeros. Therefore, it is desired that the designer have a good understanding of the method for generating the root loci of the closed-loop system, both by hand and by use of a computer software like MATLAB.

Root-locus method. The basic idea behind the root-locus method is that the values of s that make the transfer function around the loop equal -1 must satisfy the characteristic equation of the system.

The locus of roots of the characteristic equation of the closed-loop system as the gain is varied from zero to infinity gives the method its name. Such a plot clearly shows the contributions of each open-loop pole or zero to the locations of the closed-loop poles.

In designing a linear control system, we find that the root-locus method proves quite useful since it indicates the manner in which the open-loop poles and zeros should be modified so that the response meets system performance specifications. This method is particularly suited to obtaining approximate results very quickly.

Some control systems may involve more than one parameter to be adjusted. The root-locus diagram for a system having multiple parameters may be constructed by varying one parameter at a time. In this chapter we include the discussion of the root loci for a system having two parameters. The root loci for such a case is called the *root contour.*

The root-locus method is a very powerful graphical technique for investigating the effects of the variation of a system parameter on the location of the closed-loop poles. In most cases, the system parameter is the loop gain K, although the parameter can be any other variable of the system. If the designer follows the general rules for constructing the root loci, sketching the root loci of a given system may become a simple matter.

Because generating the root loci by use of MATLAB is a very simple matter, one may think sketching the root loci by hand is a waste of time and effort. However, experience in sketching the root loci by hand is invaluable for interpreting computer-generated root loci, as well as for getting a rough idea of the root loci very quickly.

By using the root-locus method, it is possible to determine the value of the loop gain K that will make the damping ratio of the dominant closed-loop poles as prescribed. If the location of an open-loop pole or zero is a system variable, then the root-locus method suggests the way to choose the location of an open-loop pole or zero. (See Example 6–8 and Problems A-6-12 through A-6-14.) More on the control system design based on the root-locus method will be given in Chapter 7.

Outline of the chapter. This chapter introduces the basic concept of the root-locus method and presents useful rules for graphically constructing the root loci, as well as the generation of root loci with MATLAB.

The outline of the chapter is as follows: Section 6–1 has presented an introduction to the root-locus method. Section 6–2 details the concepts underlying the root-locus method and presents the general procedure for sketching root loci using illustrative examples. Section 6–3 summarizes general rules for constructing root loci, and Section 6–4 discusses generating root-locus plots with MATLAB. Section 6–5 treats special cases: the first case occurs when the variable K does not appear as a multiplicative factor and the second case when the closed-loop system has positive feedback. Section 6–6 analyzes closed-loop systems by use of the root-locus method. Section 6–7 extends the root-locus method to treat closed-loop systems with transport lag. Finally, Section 6–8 discusses root-contour plots.

6-2 ROOT-LOCUS PLOTS

Angle and magnitude conditions. Consider the system shown in Figure 6–1. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)H(s)} \tag{6-1}$$

The characteristic equation for this closed-loop system is obtained by setting the denominator of the right-hand side of Equation (6–1) equal to zero. That is,

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

or

$$G(s)H(s) = -1 \tag{6-2}$$

Here we assume that $G(s)H(s)$ is a ratio of polynomials in s. [Later in Section 6–7 we extend the analysis to the case when $G(s)H(s)$ involves the transport lag e^{-Ts}.] Since $G(s)H(s)$ is a complex quantity, Equation (6–2) can be split into two equations by equating the angles and magnitudes of both sides, respectively, to obtain the following:

Angle condition:

$$\underline{/G(s)H(s)} = \pm 180°(2k + 1) \qquad (k = 0, 1, 2, \ldots) \tag{6-3}$$

Magnitude condition:

$$|G(s)H(s)| = 1 \tag{6-4}$$

The values of s that fulfill both the angle and magnitude conditions are the roots of the characteristic equation, or the closed-loop poles. A plot of the points in the complex plane satisfying the angle condition alone is the root locus. The roots of the characteristic equation (the closed-loop poles) corresponding to a given value of the gain can be determined from the magnitude condition. The details of applying the angle and magnitude conditions to obtain the closed-loop poles are presented later in this section.

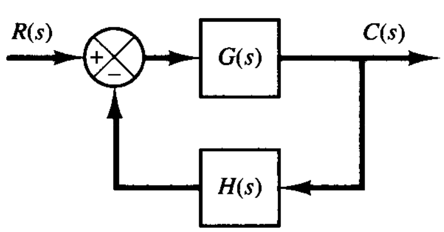


Figure 6–1
Control system.

In many cases, $G(s)H(s)$ involves a gain parameter K, and the characteristic equation may be written as

$$1 + \frac{K(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} = 0 \tag{6-5}$$

Then the root loci for the system are the loci of the closed-loop poles as the gain K is varied from zero to infinity.

Note that to begin sketching the root loci of a system by the root-locus method we must know the location of the poles and zeros of $G(s)H(s)$. Remember that the angles of the complex quantities originating from the open-loop poles and open-loop zeros to the test point s are measured in the counterclockwise direction. For example, if $G(s)H(s)$ is given by

$$G(s)H(s) = \frac{K(s + z_1)}{(s + p_1)(s + p_2)(s + p_3)(s + p_4)}$$

where $-p_2$ and $-p_3$ are complex-conjugate poles, then the angle of $G(s)H(s)$ is

$$\underline{/G(s)H(s)} = \phi_1 - \theta_1 - \theta_2 - \theta_3 - \theta_4$$

where ϕ_1, θ_1, θ_2, θ_3, and θ_4 are measured counterclockwise as shown in Figures 6–2(a) and (b). The magnitude of $G(s)H(s)$ for this system is

$$|G(s)H(s)| = \frac{KB_1}{A_1A_2A_3A_4}$$

where A_1, A_2, A_3, A_4, and B_1 are the magnitudes of the complex quantities $s + p_1, s + p_2, s + p_3, s + p_4$, and $s + z_1$, respectively, as shown in Figure 6–2(a).

Note that, because the open-loop complex-conjugate poles and complex-conjugate zeros, if any, are always located symmetrically about the real axis, the root loci are always

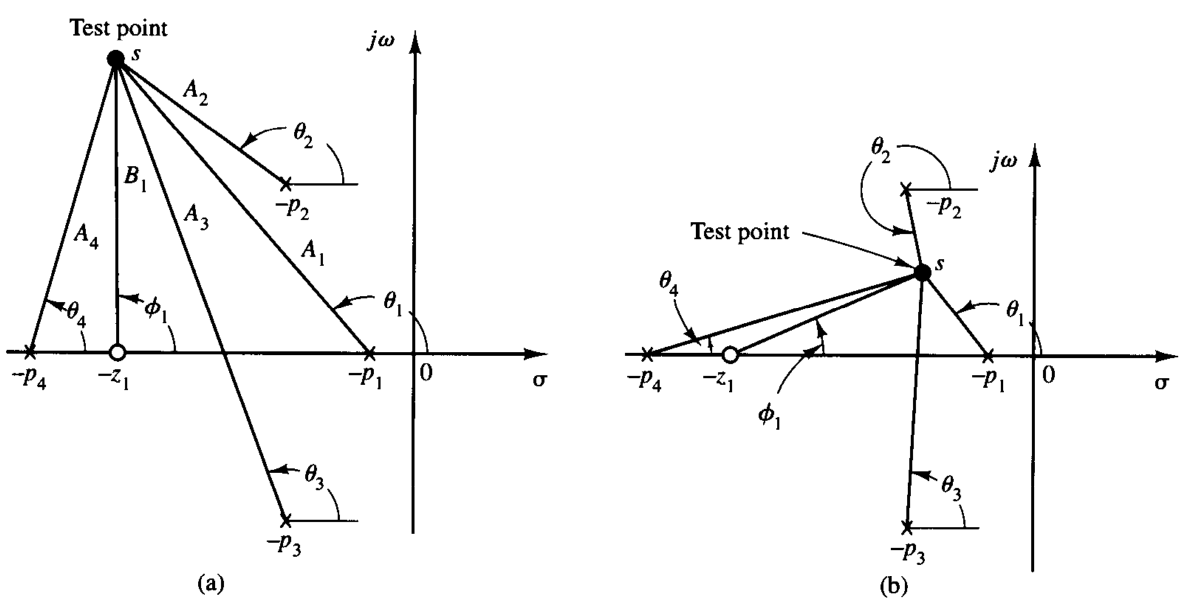


Figure 6–2
(a) and (b) Diagrams showing angle measurements from open-loop poles and open-loop zero to test point s.

symmetrical with respect to this axis. Therefore, we only need to construct the upper half of the root loci and draw the mirror image of the upper half in the lower-half s plane.

Illustrative examples. In what follows, two illustrative examples for constructing root-locus plots will be presented. Although computer approaches to the construction of the root loci are easily available, here we shall use graphical computation, combined with inspection, to determine the root loci upon which the roots of the characteristic equation of the closed-loop system must lie. Such a graphical approach will enhance understanding of how the closed-loop poles move in the complex plane as the open-loop poles and zeros are moved. Although we employ only simple systems for illustrative purposes, the procedure for finding the root loci is no more complicated for higher-order systems.

The first step in the procedure for constructing a root-locus plot is to seek out the loci of possible roots using the angle condition. Then, if necessary, the loci can be scaled, or graduated, in gain using the magnitude condition.

Because graphical measurements of angles and magnitudes are involved in the analysis, we find it necessary to use the same divisions on the abscissa as on the ordinate axis when sketching the root locus on graph paper.

EXAMPLE 6–1 Consider the system shown in Figure 6–3. (We assume that the value of gain K is nonnegative.) For this system,

$$G(s) = \frac{K}{s(s+1)(s+2)}, \qquad H(s) = 1$$

Let us sketch the root-locus plot and then determine the value of K such that the damping ratio ζ of a pair of dominant complex-conjugate closed-loop poles is 0.5.

For the given system, the angle condition becomes

$$\begin{aligned} \angle G(s) &= \angle \frac{K}{s(s+1)(s+2)} \\ &= -\angle s - \angle s+1 - \angle s+2 \\ &= \pm 180^\circ(2k+1) \qquad (k = 0, 1, 2, \ldots) \end{aligned}$$

The magnitude condition is

$$|G(s)| = \left| \frac{K}{s(s+1)(s+2)} \right| = 1$$

A typical procedure for sketching the root-locus plot is as follows:

1. *Determine the root loci on the real axis.* The first step in constructing a root-locus plot is to locate the open-loop poles, $s = 0$, $s = -1$, and $s = -2$, in the complex plane. (There are no

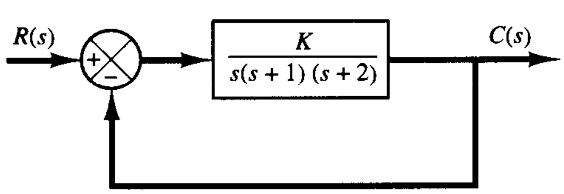


Figure 6–3
Control system.

open-loop zeros in this system.) The locations of the open-loop poles are indicated by crosses. (The locations of the open-loop zeros in this book will be indicated by small circles.) Note that the starting points of the root loci (the points corresponding to $K = 0$) are open-loop poles. The number of individual root loci for this system is three, which is the same as the number of open-loop poles.

To determine the root loci on the real axis, we select a test point, s. If the test point is on the positive real axis, then

$$\angle s = \angle s + 1 = \angle s + 2 = 0°$$

This shows that the angle condition cannot be satisfied. Hence, there is no root locus on the positive real axis. Next, select a test point on the negative real axis between 0 and -1. Then

$$\angle s = 180°, \qquad \angle s + 1 = \angle s + 2 = 0°$$

Thus

$$-\angle s - \angle s + 1 - \angle s + 2 = -180°$$

and the angle condition is satisfied. Therefore, the portion of the negative real axis between 0 and -1 forms a portion of the root locus. If a test point is selected between -1 and -2, then

$$\angle s = \angle s + 1 = 180°, \qquad \angle s + 2 = 0°$$

and

$$-\angle s - \angle s + 1 - \angle s + 2 = -360°$$

It can be seen that the angle condition is not satisfied. Therefore, the negative real axis from -1 to -2 is not a part of the root locus. Similarly, if a test point is located on the negative real axis from -2 to $-\infty$, the angle condition is satisfied. Thus, root loci exist on the negative real axis between 0 and -1 and between -2 and $-\infty$.

2. *Determine the asymptotes of the root loci.* The asymptotes of the root loci as s approaches infinity can be determined as follows: If a test point s is selected very far from the origin, then

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

and the angle condition becomes

$$-3\angle s = \pm 180°(2k+1) \qquad (k = 0, 1, 2, \ldots)$$

or

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k+1)}{3} \qquad (k = 0, 1, 2, \ldots)$$

Since the angle repeats itself as k is varied, the distinct angles for the asymptotes are determined as 60°, $-60°$, and 180°. Thus, there are three asymptotes. The one having the angle of 180° is the negative real axis.

Before we can draw these asymptotes in the complex plane, we must find the point where they intersect the real axis. Since

$$G(s) = \frac{K}{s(s+1)(s+2)} \tag{6–6}$$

if a test point is located very far from the origin, then $G(s)$ may be written as

$$G(s) = \frac{K}{s^3 + 3s^2 + \cdots} \tag{6-7}$$

Since the characteristic equation is

$$G(s) = -1$$

referring to Equation (6–7) the characteristic equation may be written as

$$s^3 + 3s^2 + \cdots = -K$$

For a large value of s, this last equation may be approximated by

$$(s + 1)^3 = 0$$

If the abscissa of the intersection of the asymptotes and the real axis is denoted by $s = -\sigma_a$, then

$$\sigma_a = -1$$

and the point of origin of the asymptotes is $(-1, 0)$. The asymptotes are almost part of the root loci in regions very far from the origin.

3. *Determine the breakaway point.* To plot root loci accurately, we must find the breakaway point, where the root-locus branches originating from the poles at 0 and -1 break away (as K is increased) from the real axis and move into the complex plane. The breakaway point corresponds to a point in the s plane where multiple roots of the characteristic equation occur.

A simple method for finding the breakaway point is available. We shall present this method in the following: Let us write the characteristic equation as

$$f(s) = B(s) + KA(s) = 0 \tag{6-8}$$

where $A(s)$ and $B(s)$ do not contain K. Note that $f(s) = 0$ has multiple roots at points where

$$\frac{df(s)}{ds} = 0$$

This can be seen as follows: Suppose that $f(s)$ has multiple roots of order r. Then $f(s)$ may be written as

$$f(s) = (s - s_1)^r(s - s_2) \cdots (s - s_n)$$

If we differentiate this equation with respect to s and set $s = s_1$, then we get

$$\left.\frac{df(s)}{ds}\right|_{s=s_1} = 0 \tag{6-9}$$

This means that multiple roots of $f(s)$ will satisfy Equation (6–9). From Equation (6–8), we obtain

$$\frac{df(s)}{ds} = B'(s) + KA'(s) = 0 \tag{6-10}$$

where

$$A'(s) = \frac{dA(s)}{ds}, \qquad B'(s) = \frac{dB(s)}{ds}$$

The particular value of K that will yield multiple roots of the characteristic equation is obtained from Equation (6–10) as

$$K = -\frac{B'(s)}{A'(s)}$$

If we substitute this value of K into Equation (6–8), we get

$$f(s) = B(s) - \frac{B'(s)}{A'(s)} A(s) = 0$$

or

$$B(s)A'(s) - B'(s)A(s) = 0 \tag{6–11}$$

If Equation (6–11) is solved for s, the points where multiple roots occur can be obtained. On the other hand, from Equation (6–8) we obtain

$$K = -\frac{B(s)}{A(s)}$$

and

$$\frac{dK}{ds} = -\frac{B'(s)A(s) - B(s)A'(s)}{A^2(s)}$$

If dK/ds is set equal to zero, we get the same equation as Equation (6–11). Therefore, the breakaway points can be simply determined from the roots of

$$\frac{dK}{ds} = 0$$

It should be noted that not all the solutions of Equation (6–11) or of $dK/ds = 0$ correspond to actual breakaway points. If a point at which $df(s)/ds = 0$ is on a root locus, it is an actual breakaway or break-in point. Stated differently, if at a point at which $df(s)/ds = 0$ the value of K takes a real positive value then that point is an actual breakaway or break-in point.

For the present example, the characteristic equation $G(s) + 1 = 0$ is given by

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

or

$$K = -(s^3 + 3s^2 + 2s)$$

By setting $dK/ds = 0$, we obtain

$$\frac{dK}{ds} = -(3s^2 + 6s + 2) = 0$$

or

$$s = -0.4226, \qquad s = -1.5774$$

Since the breakaway point must lie on a root locus between 0 and -1, it is clear that $s = -0.4226$ corresponds to the actual breakaway point. Point $s = -1.5774$ is not on the root locus. Hence, this point is not an actual breakaway or break-in point. In fact, evaluation of the values of K corresponding to $s = -0.4226$ and $s = -1.5774$ yields

$$K = 0.3849, \qquad \text{for } s = -0.4226$$

$$K = -0.3849, \qquad \text{for } s = -1.5774$$

4. *Determine the points where the root loci cross the imaginary axis.* These points can be found by use of Routh's stability criterion as follows: Since the characteristic equation for the present system is

$$s^3 + 3s^2 + 2s + K = 0$$

the Routh array becomes

$$\begin{array}{lcc} s^3 & 1 & 2 \\ s^2 & 3 & K \\ s^1 & \dfrac{6-K}{3} & \\ s^0 & K & \end{array}$$

The value of K that makes the s^1 term in the first column equal zero is $K = 6$. The crossing points on the imaginary axis can then be found by solving the auxiliary equation obtained from the s^2 row; that is,

$$3s^2 + K = 3s^2 + 6 = 0$$

which yields

$$s = \pm j\sqrt{2}$$

The frequencies at the crossing points on the imaginary axis are thus $\omega = \pm\sqrt{2}$. The gain value corresponding to the crossing points is $K = 6$.

An alternative approach is to let $s = j\omega$ in the characteristic equation, equate both the real part and the imaginary part to zero, and then solve for ω and K. For the present system, the characteristic equation, with $s = j\omega$, is

$$(j\omega)^3 + 3(j\omega)^2 + 2(j\omega) + K = 0$$

or

$$(K - 3\omega^2) + j(2\omega - \omega^3) = 0$$

Equating both the real and imaginary parts of this last equation to zero, we obtain

$$K - 3\omega^2 = 0, \qquad 2\omega - \omega^3 = 0$$

from which

$$\omega = \pm\sqrt{2}, \qquad K = 6 \qquad \text{or} \qquad \omega = 0, \qquad K = 0$$

Thus, root loci cross the imaginary axis at $\omega = \pm\sqrt{2}$, and the value of K at the crossing points is 6. Also, a root-locus branch on the real axis touches the imaginary axis at $\omega = 0$.

5. *Choose a test point in the broad neighborhood of the $j\omega$ axis and the origin,* as shown in Figure 6–4, and apply the angle condition. If a test point is on the root loci, then the sum of the

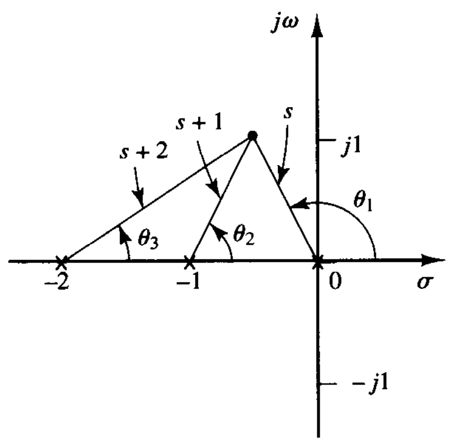


Figure 6–4
Construction of root locus.

three angles, $\theta_1 + \theta_2 + \theta_3$, must be 180°. If the test point does not satisfy the angle condition, select another test point until it satisfies the condition. (The sum of the angles at the test point will indicate which direction the test point should be moved.) Continue this process and locate a sufficient number of points satisfying the angle condition.

6. *Draw the root loci,* based on the information obtained in the foregoing steps, as shown in Figure 6–5.

7. *Determine a pair of dominant complex-conjugate closed-loop poles such that the damping ratio ζ is 0.5.* Closed-loop poles with $\zeta = 0.5$ lie on lines passing through the origin and making the angles $\pm\cos^{-1}\zeta = \pm\cos^{-1} 0.5 = \pm 60°$ with the negative real axis. From Figure 6–5, such closed-loop poles having $\zeta = 0.5$ are obtained as follows:

$$s_1 = -0.3337 + j0.5780, \qquad s_2 = -0.3337 - j0.5780$$

The value of K that yields such poles is found from the magnitude condition as follows:

$$\begin{aligned} K &= \left|s(s+1)(s+2)\right|_{s=-0.3337+j0.5780} \\ &= 1.0383 \end{aligned}$$

Using this value of K, the third pole is found at $s = -2.3326$.

Note that, from step 4, it can be seen that for $K = 6$ the dominant closed-loop poles lie on the imaginary axis at $s = \pm j\sqrt{2}$. With this value of K, the system will exhibit sustained oscillations. For $K > 6$, the dominant closed-loop poles lie in the right-half s plane, resulting in an unstable system.

Finally, note that, if necessary, the root loci can be easily graduated in terms of K by use of the magnitude condition. We simply pick out a point on a root locus, measure the magnitudes of the three complex quantities s, $s + 1$, and $s + 2$, and multiply these magnitudes; the product is equal to the gain value K at that point, or

$$|s| \cdot |s+1| \cdot |s+2| = K$$

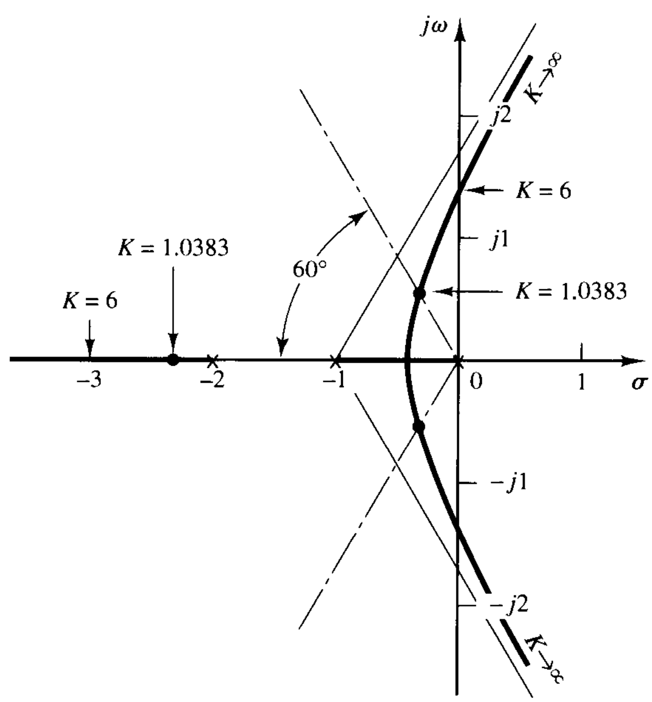


Figure 6–5
Root-locus plot.

EXAMPLE 6–2 In this example, we shall sketch the root-locus plot of a system with complex-conjugate open-loop poles. Consider the system shown in Figure 6–6. For this system,

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

It is seen that $G(s)$ has a pair of complex-conjugate poles at

$$s = -1 + j\sqrt{2}, \qquad s = -1 - j\sqrt{2}$$

A typical procedure for sketching the root-locus plot is as follows:

1. *Determine the root loci on the real axis.* For any test point s on the real axis, the sum of the angular contributions of the complex-conjugate poles is 360°, as shown in Figure 6–7. Thus the net effect of the complex-conjugate poles is zero on the real axis. The location of the root locus on the real axis is determined from the open-loop zero on the negative real axis. A simple test reveals that a section of the negative real axis, that between -2 and $-\infty$, is a part of the root locus. It is noted that, since this locus lies between two zeros (at $s = -2$ and $s = -\infty$), it is actually a part of two root loci, each of which starts from one of the two complex-conjugate poles. In other words, two root loci break in the part of the negative real axis between -2 and $-\infty$.

Since there are two open-loop poles and one zero, there is one asymptote, which coincides with the negative real axis.

2. *Determine the angle of departure from the complex-conjugate open-loop poles.* The presence of a pair of complex-conjugate open-loop poles requires determination of the angle of departure from these poles. Knowledge of this angle is important since the root locus near a complex pole yields information as to whether the locus originating from the complex pole migrates toward the real axis or extends toward the asymptote.

Referring to Figure 6–8, if we choose a test point and move it in the very vicinity of the complex open-loop pole at $s = -p_1$, we find that the sum of the angular contributions from the pole at $s = -p_2$ and zero at $s = -z_1$ to the test point can be considered remaining the same. If the test

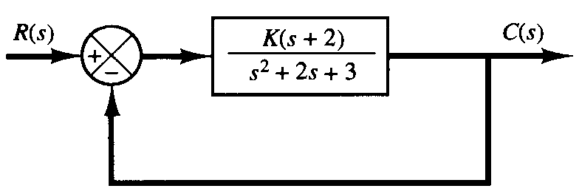


Figure 6–6
Control system.

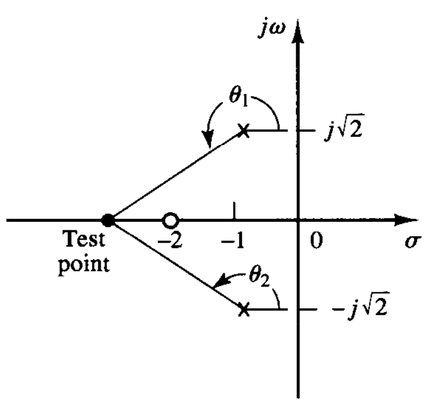


Figure 6–7
Determination of the root locus on the real axis.

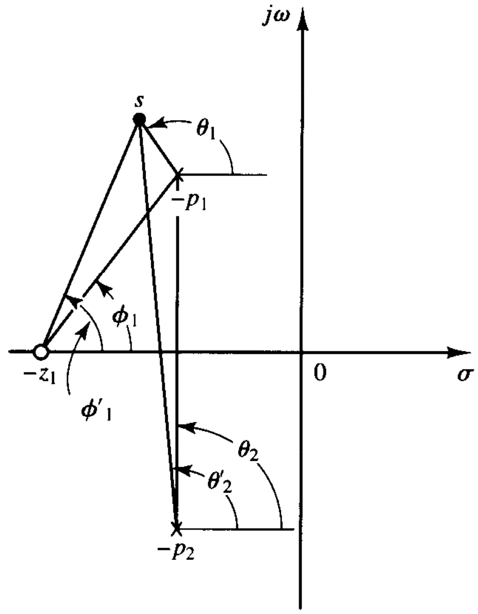


Figure 6–8
Determination of the angle of departure.

point is to be on the root locus, then the sum of ϕ'_1, $-\theta_1$, and $-\theta'_2$ must be $\pm 180°(2k + 1)$, where $k = 0, 1, 2, \ldots$. Thus, in the example,

$$\phi'_1 - (\theta_1 + \theta'_2) = \pm 180°(2k + 1)$$

or

$$\theta_1 = 180° - \theta'_2 + \phi'_1 = 180° - \theta_2 + \phi_1$$

The angle of departure is then

$$\theta_1 = 180° - \theta_2 + \phi_1 = 180° - 90° + 55° = 145°$$

Since the root locus is symmetric about the real axis, the angle of departure from the pole at $s = -p_2$ is $-145°$.

3. *Determine the break-in point.* A break-in point exists where a pair of root-locus branches coalesce as K is increased. For this problem, the break-in point can be found as follows: Since

$$K = -\frac{s^2 + 2s + 3}{s + 2}$$

we have

$$\frac{dK}{ds} = -\frac{(2s + 2)(s + 2) - (s^2 + 2s + 3)}{(s + 2)^2} = 0$$

which gives

$$s^2 + 4s + 1 = 0$$

or

$$s = -3.7320 \qquad \text{or} \qquad s = -0.2680$$

Notice that point $s = -3.7320$ is on the root locus. Hence this point is an actual break-in point. (Note that at point $s = -3.7320$ the corresponding gain value is $K = 5.4641$.) Since point $s = -0.2680$ is not on the root locus, it cannot be a break-in point. (For point $s = -0.2680$, the corresponding gain value is $K = -1.4641$.)

4. *Sketch a root-locus plot,* based on the information obtained in the foregoing steps. To determine accurate root loci, several points must be found by trial and error between the break-in point and the complex open-loop poles. (To facilitate sketching the root-locus plot, we should find the direction in which the test point should be moved by mentally summing up the changes on the angles of the poles and zeros.) Figure 6–9 shows a complete root-locus plot for the system considered.

The value of the gain K at any point on root locus can be found by applying the magnitude condition. For example, the value of K at which the complex-conjugate closed-loop poles have the damping ratio $\zeta = 0.7$ can be found by locating the roots, as shown in Figure 6–9, and computing the value of K as follows:

$$K = \left|\frac{(s + 1 - j\sqrt{2})(s + 1 + j\sqrt{2})}{s + 2}\right|_{s=-1.67+j1.70} = 1.34$$

It is noted that in this system the root locus in the complex plane is a part of a circle. Such a circular root locus will not occur in most systems. Circular root loci may occur in systems that involve two poles and one zero, two poles and two zeros, or one pole and two zeros. Even in such systems, whether circular root loci occur depends on the locations of poles and zeros involved.

To show the occurrence of a circular root locus in the present system, we need to derive the equation for the root locus. For the present system, the angle condition is

$$\underline{/s + 2} - \underline{/s + 1 - j\sqrt{2}} - \underline{/s + 1 + j\sqrt{2}} = \pm 180°(2k + 1)$$

If $s = \sigma + j\omega$ is substituted into this last equation, we obtain

$$\underline{/\sigma + 2 + j\omega} - \underline{/\sigma + 1 + j\omega - j\sqrt{2}} - \underline{/\sigma + 1 + j\omega + j\sqrt{2}} = \pm 180°(2k + 1)$$

which can be written as

$$\tan^{-1}\left(\frac{\omega}{\sigma + 2}\right) - \tan^{-1}\left(\frac{\omega - \sqrt{2}}{\sigma + 1}\right) - \tan^{-1}\left(\frac{\omega + \sqrt{2}}{\sigma + 1}\right) = \pm 180°(2k + 1)$$

or

$$\tan^{-1}\left(\frac{\omega - \sqrt{2}}{\sigma + 1}\right) + \tan^{-1}\left(\frac{\omega + \sqrt{2}}{\sigma + 1}\right) = \tan^{-1}\left(\frac{\omega}{\sigma + 2}\right) \pm 180°(2k + 1)$$

Taking tangents of both sides of this last equation using the relationship

$$\tan(x \pm y) = \frac{\tan x \pm \tan y}{1 \mp \tan x \tan y} \tag{6–12}$$

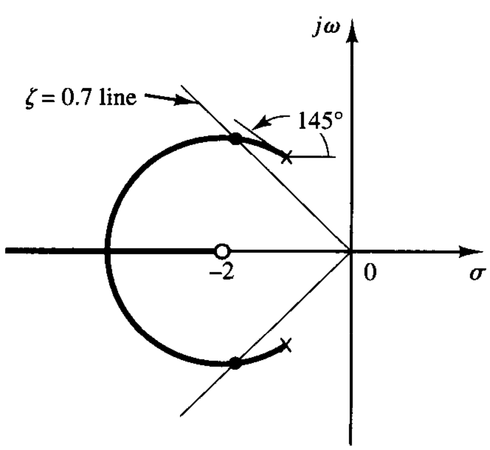


Figure 6–9
Root-locus plot.

we obtain

$$\tan\left[\tan^{-1}\left(\frac{\omega-\sqrt{2}}{\sigma+1}\right)+\tan^{-1}\left(\frac{\omega+\sqrt{2}}{\sigma+1}\right)\right]=\tan\left[\tan^{-1}\left(\frac{\omega}{\sigma+2}\right)\pm 180^\circ(2k+1)\right]$$

or

$$\frac{\dfrac{\omega-\sqrt{2}}{\sigma+1}+\dfrac{\omega+\sqrt{2}}{\sigma+1}}{1-\left(\dfrac{\omega-\sqrt{2}}{\sigma+1}\right)\left(\dfrac{\omega+\sqrt{2}}{\sigma+1}\right)}=\frac{\dfrac{\omega}{\sigma+2}\pm 0}{1\mp\dfrac{\omega}{\sigma+2}\times 0}$$

which can be simplified to

$$\frac{2\omega(\sigma+1)}{(\sigma+1)^2-(\omega^2-2)}=\frac{\omega}{\sigma+2}$$

or

$$\omega[(\sigma+2)^2+\omega^2-3]=0$$

This last equation is equivalent to

$$\omega=0 \qquad \text{or} \qquad (\sigma+2)^2+\omega^2=(\sqrt{3})^2$$

These two equations are the equations for the root loci for the present system. Notice that the first equation, $\omega = 0$, is the equation for the real axis. The real axis from $s = -2$ to $s = -\infty$ corresponds to a root locus for $K \geq 0$. The remaining part of the real axis corresponds to a root locus when K is negative. (In the present system, K is nonnegative.) The second equation for the root locus is an equation of a circle with center at $\sigma = -2, \omega = 0$ and the radius equal to $\sqrt{3}$. That part of the circle to the left of the complex-conjugate poles corresponds to a root locus for $K \geq 0$. The remaining part of the circle corresponds to a root locus when K is negative.

It is important to note that easily interpretable equations for the root locus can be derived for simple systems only. For complicated systems having many poles and zeros, any attempt to derive equations for the root loci is discouraged. Such derived equations are very complicated and their configuration in the complex plane is difficult to visualize.

6–3 SUMMARY OF GENERAL RULES FOR CONSTRUCTING ROOT LOCI

For a complicated system with many open-loop poles and zeros, constructing a root-locus plot may seem complicated, but actually it is not difficult if the rules for constructing the root loci are applied. By locating particular points and asymptotes and by computing angles of departure from complex poles and angles of arrival at complex zeros, we can construct the general form of the root loci without difficulty.

Some of the rules for constructing root loci were given in Section 6–2. The purpose of this section is to summarize the general rules for constructing root loci of the system shown in Figure 6–10. While the root-locus method is essentially based on a trial-and-error technique, the number of trials required can be greatly reduced if we use these rules.

General rules for constructing root loci. We shall now summarize the general rules and procedure for constructing the root loci of the system shown in Figure 6–10.

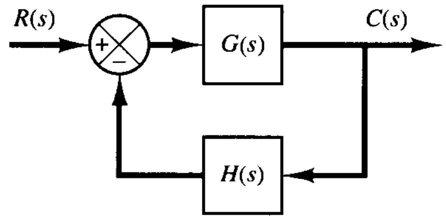


Figure 6–10
Control system.

First, obtain the characteristic equation

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

Then rearrange this equation so that the parameter of interest appears as the multiplying factor in the form

$$1 + \frac{K(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} = 0 \tag{6–13}$$

In the present discussions, we assume that the parameter of interest is the gain K, where $K > 0$. (If $K < 0$, which corresponds to the positive-feedback case, the angle condition must be modified. See Section 6–5.) Note, however, that the method is still applicable to systems with parameters of interest other than gain.

1. *Locate the poles and zeros of $G(s)H(s)$ on the s plane. The root-locus branches start from open-loop poles and terminate at zeros (finite zeros or zeros at infinity).* From the factored form of the open-loop transfer function, locate the open-loop poles and zeros in the s plane. [Note that the open-loop zeros are the zeros of $G(s)H(s)$, while the closed-loop zeros consist of the zeros of $G(s)$ and the poles of $H(s)$.]

Note that the root loci are symmetrical about the real axis of the s plane, because the complex poles and complex zeros occur only in conjugate pairs.

Find the starting points and terminating points of the root loci and find also the number of separate root loci. The points on the root loci corresponding to $K = 0$ are open-loop poles. This can be seen from the magnitude condition by letting K approach zero, or

$$\lim_{K \to 0} \left| \frac{(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} \right| = \lim_{K \to 0} \frac{1}{K} = \infty$$

This last equation implies that as K is decreased the value of s must approach one of the open-loop poles. Each root locus thus originates at a pole of the open-loop transfer function $G(s)H(s)$. As K is increased to infinity, each root-locus approaches either a zero of the open-loop transfer function or infinity in the complex plane. This can be seen as follows: If we let K approach infinity in the magnitude condition, then

$$\lim_{K \to \infty} \left| \frac{(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} \right| = \lim_{K \to \infty} \frac{1}{K} = 0$$

Hence the value of s must approach one of the finite open-loop zeros or an open-loop zero at infinity. [If the zeros at infinity are included in the count, $G(s)H(s)$ has the same number of zeros as poles.]

A root-locus plot will have just as many branches as there are roots of the characteristic equation. Since the number of open-loop poles generally exceeds that of zeros, the number of branches equals that of poles. If the number of closed-loop poles is the same as the number of open-loop poles, then the number of individual root-locus branches terminating at finite open-loop zeros is equal to the number m of the open-loop zeros. The remaining $n - m$ branches terminate at infinity ($n - m$ implicit zeros at infinity) along asymptotes.

If we include poles and zeros at infinity, the number of open-loop poles is equal to that of open-loop zeros. Hence we can always state that the root loci start at the poles of $G(s)H(s)$ and end at the zeros of $G(s)H(s)$, as K increases from zero to infinity, where the poles and zeros include both those in the finite s plane and those at infinity.

2. *Determine the root loci on the real axis.* Root loci on the real axis are determined by open-loop poles and zeros lying on it. The complex-conjugate poles and zeros of the open-loop transfer function have no effect on the location of the root loci on the real axis because the angle contribution of a pair of complex-conjugate poles or zeros is 360° on the real axis. Each portion of the root locus on the real axis extends over a range from a pole or zero to another pole or zero. In constructing the root loci on the real axis, choose a test point on it. If the total number of real poles and real zeros to the right of this test point is odd, then this point lies on a root locus. The root locus and its complement form alternate segments along the real axis.

3. *Determine the asymptotes of root loci.* If the test point s is located far from the origin, then the angle of each complex quantity may be considered the same. One open-loop zero and one open-loop pole then cancel the effects of the other. Therefore, the root loci for very large values of s must be asymptotic to straight lines whose angles (slopes) are given by

$$\text{Angles of asymptotes} = \frac{\pm 180^\circ(2k + 1)}{n - m} \qquad (k = 0, 1, 2, \ldots)$$

where n = number of finite poles of $G(s)H(s)$
m = number of finite zeros of $G(s)H(s)$

Here, $k = 0$ corresponds to the asymptotes with the smallest angle with the real axis. Although k assumes an infinite number of values, as k is increased, the angle repeats itself, and the number of distinct asymptotes is $n - m$.

All the asymptotes intersect on the real axis. The point at which they do so is obtained as follows: If both the numerator and denominator of the open-loop transfer function are expanded, the result is

$$G(s)H(s) = \frac{K[s^m + (z_1 + z_2 + \cdots + z_m)s^{m-1} + \cdots + z_1 z_2 \cdots z_m]}{s^n + (p_1 + p_2 + \cdots + p_n)s^{n-1} + \cdots + p_1 p_2 \cdots p_n}$$

If a test point is located very far from the origin, then by dividing the denominator by the numerator $G(s)H(s)$ may be written as

$$G(s)H(s) = \frac{K}{s^{n-m} + [(p_1 + p_2 + \cdots + p_n) - (z_1 + z_2 + \cdots + z_m)]s^{n-m-1} + \cdots}$$

Since the characteristic equation is

$$G(s)H(s) = -1$$

it may be written

$$s^{n-m} + [(p_1 + p_2 + \cdots + p_n) - (z_1 + z_2 + \cdots + z_m)]s^{n-m-1} + \cdots = -K \tag{6–14}$$

For a large value of s, Equation (6–14) may be approximated by

$$\left[s + \frac{(p_1 + p_2 + \cdots + p_n) - (z_1 + z_2 + \cdots + z_m)}{n - m}\right]^{n-m} = 0$$

If the abscissa of the intersection of the asymptotes and the real axis is denoted by $s = \sigma_a$, then

$$\sigma_a = -\frac{(p_1 + p_2 + \cdots + p_n) - (z_1 + z_2 + \cdots + z_m)}{n - m} \tag{6–15}$$

or

$$\sigma_a = \frac{(\text{sum of poles}) - (\text{sum of zeros})}{n - m} \tag{6–16}$$

Because all the complex poles and zeros occur in conjugate pairs, σ_a is always a real quantity. Once the intersection of the asymptotes and the real axis is found, the asymptotes can be readily drawn in the complex plane.

It is important to note that the asymptotes show the behavior of the root loci for $|s| \gg 1$. A root locus branch may lie on one side of the corresponding asymptote or may cross the corresponding asymptote from one side to the other side.

4. *Find the breakaway and break-in points.* Because of the conjugate symmetry of the root loci, the breakaway points and break-in points either lie on the real axis or occur in complex-conjugate pairs.

If a root locus lies between two adjacent open-loop poles on the real axis, then there exists at least one breakaway point between the two poles. Similarly, if the root locus lies between two adjacent zeros (one zero may be located at $-\infty$) on the real axis, then there always exists at least one break-in point between the two zeros. If the root locus lies between an open-loop pole and a zero (finite or infinite) on the real axis, then there may exist no breakaway or break-in points or there may exist both breakaway and break-in points.

Suppose that the characteristic equation is given by

$$B(s) + KA(s) = 0$$

The breakaway points and break-in points correspond to multiple roots of the characteristic equation. Hence, the breakaway and break-in points can be determined from the roots of

$$\frac{dK}{ds} = -\frac{B'(s)A(s) - B(s)A'(s)}{A^2(s)} = 0 \tag{6–17}$$

where the prime indicates differentiation with respect to s. It is important to note that the breakaway points and break-in points must be the roots of Equation (6–17), but not all roots of Equation (6–17) are breakaway or break-in points. If a real root of Equation (6–17) lies on the root locus portion of the real axis, then it is an actual breakaway or break-in point. If a real root of Equation (6–17) is not on the root locus portion of the real axis, then this root corresponds to neither a breakaway point nor a break-in point. If two roots $s = s_1$ and $s = -s_1$ of Equation (6–17) are a complex-conjugate pair and if it is not certain whether they are on root loci, then it is necessary to check the corresponding K value. If the value of K corresponding to a root $s = s_1$ of $dK/ds = 0$ is positive, point $s = s_1$ is an actual breakaway or break-in point. (Since K is assumed to be nonnegative, if the value of K thus obtained is negative, then point $s = s_1$ is neither a breakaway nor break-in point.)

5. *Determine the angle of departure (angle of arrival) of the root locus from a complex pole (at a complex zero).* To sketch the root loci with reasonable accuracy, we must find the directions of the root loci near the complex poles and zeros. If a test point is chosen and moved in the very vicinity of a complex pole (or complex zero), the sum of the angular contributions from all other poles and zeros can be considered remaining the same. Therefore, the angle of departure (or angle of arrival) of the root locus from a complex pole (or at a complex zero) can be found by subtracting from 180° the sum of all the angles of vectors from all other poles and zeros to the complex pole (or complex zero) in question, with appropriate signs included.

Angle of departure from a complex pole = 180°
− (sum of the angles of vectors to a complex pole in question from other poles)
+ (sum of the angles of vectors to a complex pole in question from zeros)

Angle of arrival at a complex zero = 180°
− (sum of the angles of vectors to a complex zero in question from other zeros)
+ (sum of the angles of vectors to a complex zero in question from poles)

The angle of departure is shown in Figure 6–11.

6. *Find the points where the root loci may cross the imaginary axis.* The points where the root loci intersect the $j\omega$ axis can be found easily by (a) use of Routh's stability criterion or (b) letting $s = j\omega$ in the characteristic equation, equating both the real part and the imaginary part to zero, and solving for ω and K. The values of ω thus found give the frequencies at which root loci cross the imaginary axis. The K value corresponding to each crossing frequency gives the gain at the crossing point.

7. *Taking a series of test points in the broad neighborhood of the origin of the s plane, sketch the root loci.* Determine the root loci in the broad neighborhood of the $j\omega$ axis and the origin. The most important part of the root loci is on neither the real axis nor the asymptotes but the part in the broad neighborhood of the $j\omega$ axis and the origin. The

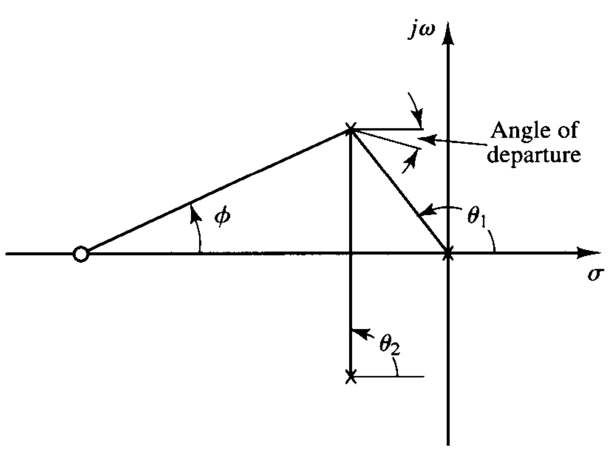


Figure 6–11
Construction of the root locus. [Angle of departure = 180° − $(\theta_1 + \theta_2) + \phi$.]

shape of the root loci in this important region in the s plane must be obtained with sufficient accuracy.

8. *Determine closed-loop poles.* A particular point on each root-locus branch will be a closed-loop pole if the value of K at that point satisfies the magnitude condition. Conversely, the magnitude condition enables us to determine the value of the gain K at any specific root location on the locus. (If necessary, the root loci may be graduated in terms of K. The root loci are continuous with K.)

The value of K corresponding to any point s on a root locus can be obtained using the magnitude condition, or

$$K = \frac{\text{product of lengths between point } s \text{ to poles}}{\text{product of lengths between point } s \text{ to zeros}}$$

This value can be evaluated either graphically or analytically.

If the gain K of the open-loop transfer function is given in the problem, then by applying the magnitude condition we can find the correct locations of the closed-loop poles for a given K on each branch of the root loci by a trial-and-error approach or by use of MATLAB, which will be presented in Section 6–4.

Comments on the root-locus plots. It is noted that the characteristic equation of the system whose open-loop transfer function is

$$G(s)H(s) = \frac{K(s^m + b_1 s^{m-1} + \cdots + b_m)}{s^n + a_1 s^{n-1} + \cdots + a_n} \qquad (n \geq m)$$

is an nth-degree algebraic equation in s. If the order of the numerator of $G(s)H(s)$ is lower than that of the denominator by two or more (which means that there are two or more zeros at infinity), then the coefficient a_1 is the negative sum of the roots of the equation and is independent of K. In such a case, if some of the roots move on the locus toward the left as K is increased, then the other roots must move toward the right as K is increased. This information is helpful in finding the general shape of the root loci.

It is also noted that a slight change in the pole–zero configuration may cause significant changes in the root-locus configurations. Figure 6–12 demonstrates the fact that a slight change in the location of a zero or pole will make the root-locus configuration look quite different.

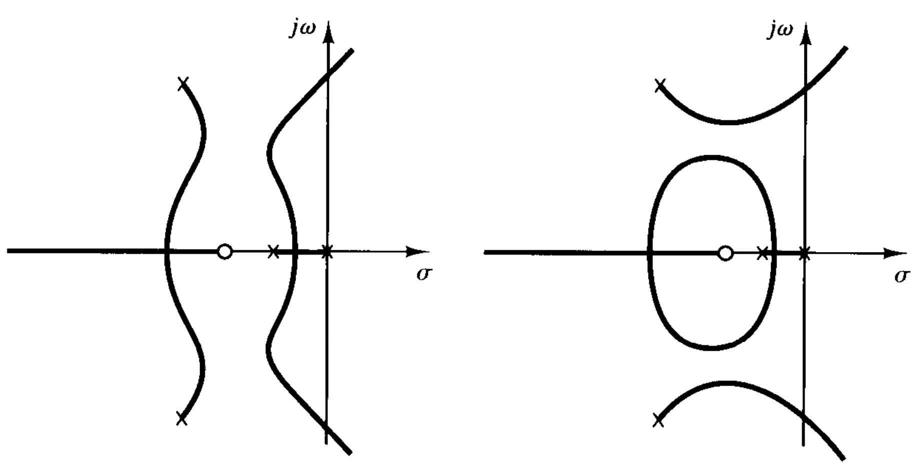


Figure 6–12
Root-locus plot.

Cancellation of poles $G(s)$ with zeros of $H(s)$. It is important to note that if the denominator of $G(s)$ and the numerator of $H(s)$ involve common factors then the corresponding open-loop poles and zeros will cancel each other, reducing the degree of the characteristic equation by one or more. For example, consider the system shown in Figure 6–13(a). (This system has velocity feedback.) By modifying the block diagram of Figure 6–13(a) to that shown in Figure 6–13(b), it is clearly seen that $G(s)$ and $H(s)$ have a common factor $s + 1$. The closed-loop transfer function $C(s)/R(s)$ is

$$\frac{C(s)}{R(s)} = \frac{K}{s(s + 1)(s + 2) + K(s + 1)}$$

The characteristic equation is

$$[s(s + 2) + K](s + 1) = 0$$

Because of the cancellation of the terms $(s + 1)$ appearing in $G(s)$ and $H(s)$, however, we have

$$\begin{aligned} 1 + G(s)H(s) &= 1 + \frac{K(s + 1)}{s(s + 1)(s + 2)} \\ &= \frac{s(s + 2) + K}{s(s + 2)} \end{aligned}$$

The reduced characteristic equation is

$$s(s + 2) + K = 0$$

The root-locus plot of $G(s)H(s)$ does not show all the roots of the characteristic equation, only the roots of the reduced equation.

To obtain the complete set of closed-loop poles, we must add the canceled pole of $G(s)H(s)$ to those closed-loop poles obtained from the root-locus plot of $G(s)H(s)$. The

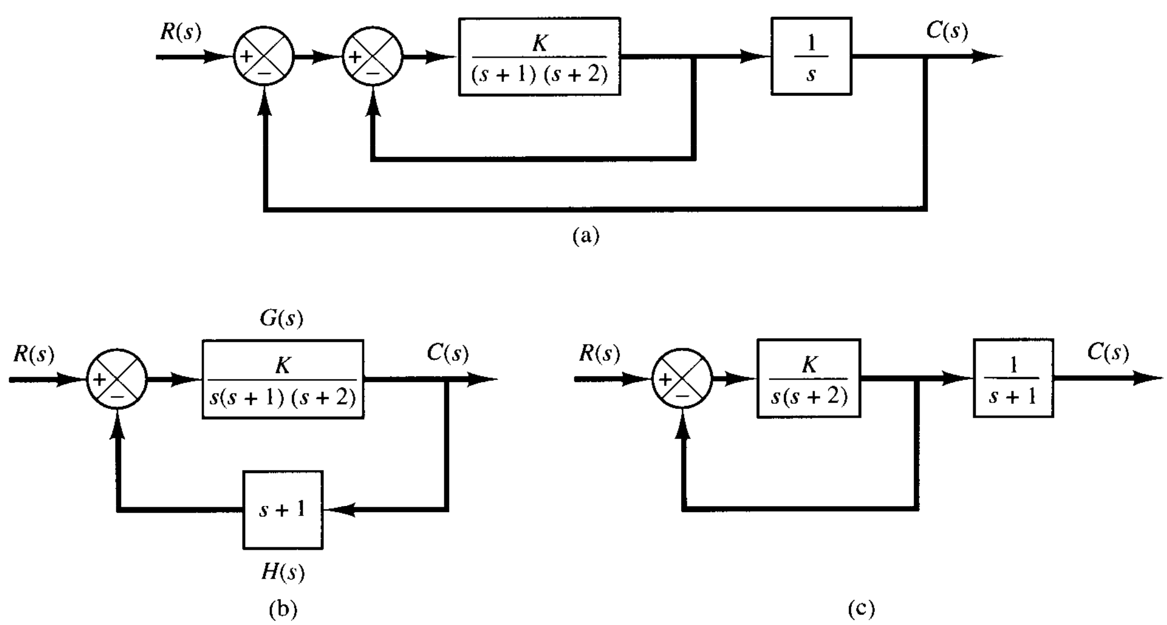
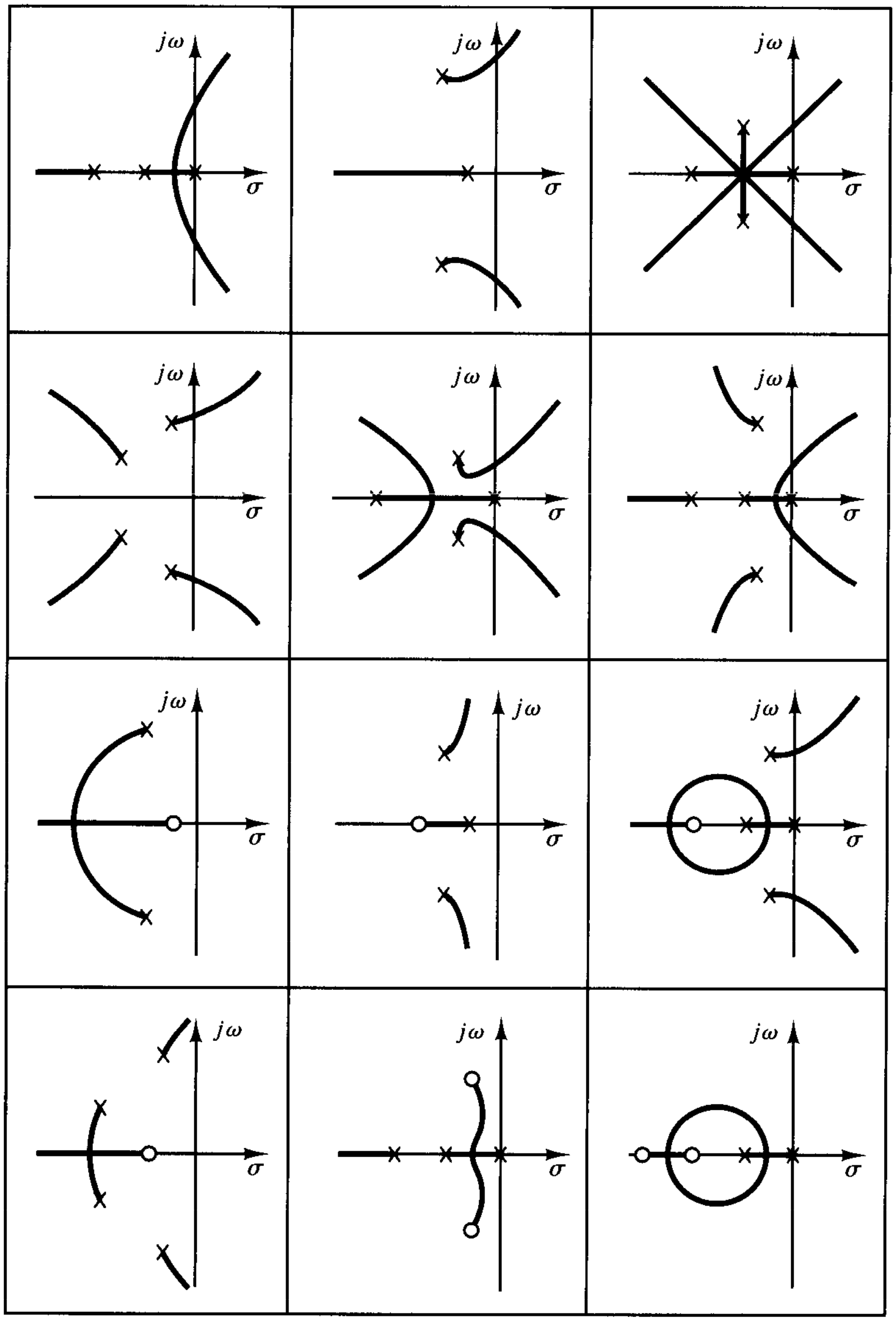


Figure 6–13
(a) Control system with velocity feedback; (b) and (c) modified block diagrams.

important thing to remember is that the canceled pole of $G(s)H(s)$ is a closed-loop pole of the system, as seen from Figure 6–13(c).

Typical pole–zero configurations and corresponding root loci. In concluding this section, we show several open-loop pole–zero configurations and their corresponding root loci in Table 6–1. The pattern of the root loci depends only on the relative separation of the open-loop poles and zeros. If the number of open-loop poles

Table 6–1 Open-Loop Pole–Zero Configurations and the Corresponding Root Loci



exceeds the number of finite zeros by three or more, there is a value of the gain K beyond which root loci enter the right-half s plane, and thus the system can become unstable. A stable system must have all its closed-loop poles in the left-half s plane.

Note that once we have some experience with the method we can easily evaluate the changes in the root loci due to the changes in the number and location of the open-loop poles and zeros by visualizing the root-locus plots resulting from various pole–zero configurations.

Summary. From the preceding discussions, it should be clear that it is possible to sketch a reasonably accurate root-locus diagram for a given system by following simple rules. (The reader is suggested to study various root-locus diagrams shown in the solved problems at the end of the chapter.) At preliminary design stages, we may not need the precise locations of the closed-loop poles. Often their approximate locations are all that is needed to make an estimate of system performance. Thus, it is important that the designer have the capability of quickly sketching the root loci for a given system.

6-4 ROOT-LOCUS PLOTS WITH MATLAB

In this section we present the MATLAB approach to the generation of root-locus plots.

Plotting root loci with MATLAB. In plotting root loci with MATLAB we deal with the system equation given in the form of Equation (6–13), which may be written as

$$1 + K\frac{\text{num}}{\text{den}} = 0$$

where num is the numerator polynomial and den is the denominator polynomial. That is,

$$\begin{aligned}
\text{num} &= (s + z_1)(s + z_2)\cdots(s + z_m) \\
&= s^m + (z_1 + z_2 + \cdots + z_m)s^{m-1} + \cdots + z_1 z_2 \cdots z_m \\
\text{den} &= (s + p_1)(s + p_2)\cdots(s + p_n) \\
&= s^n + (p_1 + p_2 + \cdots + p_n)s^{n-1} + \cdots + p_1 p_2 \cdots p_n
\end{aligned}$$

Note that both vectors num and den must be written in descending powers of s.

A MATLAB command commonly used for plotting root loci is

```
rlocus(num,den)
```

Using this command, the root-locus plot is drawn on the screen. The gain vector K is automatically determined. The command rlocus works for both continuous- and discrete-time systems.

For the systems defined in state space, rlocus(A,B,C,D) plots the root locus of the system with the gain vector automatically determined.

Note that commands

```
rlocus(num,den,K)        and        rlocus(A,B,C,D,K)
```

use the user-supplied gain vector K. (The vector K contains all the gain values for which the closed-loop poles are to be computed.)

If invoked with left-hand arguments

```
[r,K] = rlocus(num,den)
[r,K] = rlocus(num,den,K)
[r,K] = rlocus(A,B,C,D)
[r,K] = rlocus(A,B,C,D,K)
```

the screen will show the matrix r and gain vector K. (r has length K rows and length den − 1 columns containing the complex root locations. Each row of the matrix corresponds to a gain from vector K.) The plot command

```
plot(r,' ')
```

plots the root loci.

If it is desired to plot the root loci with marks 'o' or 'x', it is necessary to use the following command:

```
r = rlocus(num,den)
plot(r,'o')        or        plot(r,'x')
```

Plotting root loci using marks 'o' or 'x' is instructive, since each calculated closed-loop pole is graphically shown; in some portion of the root loci those marks are densely placed and in another portion of the root loci they are sparsely placed. MATLAB supplies its own set of gain values used to calculate a root-locus plot. It does so by an internal adaptive step-size routine. Also, MATLAB uses the automatic axis-scaling feature of the *plot* command.

Finally, note that, since the gain vector is automatically determined, root-locus plots of

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

$$G(s)H(s) = \frac{10K(s+1)}{s(s+2)(s+3)}$$

$$G(s)H(s) = \frac{200K(s+1)}{s(s+2)(s+3)}$$

are all the same. The num and den set of the system is the same for all three systems. The num and den are

```
num = [0   0   1   1]
den = [1   5   6   0]
```

EXAMPLE 6–3 Consider the control system shown in Figure 6–14. To plot the root-locus diagram with MATLAB, it is necessary to find the numerator and denominator polynomials of the open loop.

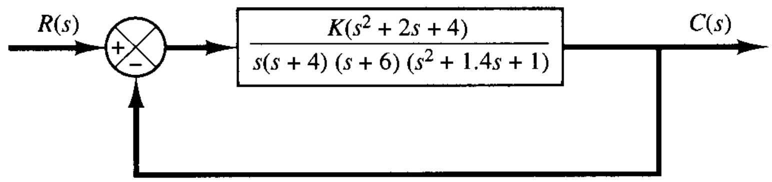


Figure 6–14
Control system.

For this problem, the numerator is already given as a polynomial in s. However, the denominator is given as a product of first- and second-order terms, with the result that we must multiply these terms to get a polynomial in s. The multiplication of these terms can be done easily by use of the *convolution command,* as shown next.

Define

$$a = s(s + 4) = s^2 + 4s \quad : \quad \mathrm{a = [1 \quad 4 \quad 0]}$$
$$b = s + 6 \quad : \quad \mathrm{b = [1 \quad 6]}$$
$$c = s^2 + 1.4s + 1 \quad : \quad \mathrm{c = [1 \quad 1.4 \quad 1]}$$

Then use the following command:

```
d = conv(a,b);      e = conv(c,d)
```

[Note that conv(a,b) gives the product of two polynomials a and b.] See the following computer output:

```
a = [1   4   0];
b = [1   6];
c = [1   1.4   1];
d = conv(a,b)

d =

    1   10   24   0

e = conv(c,d)

e =

    1.0000   11.4000   39.0000   43.6000   24.0000   0
```

The denominator polynomial is thus found to be

```
den = [1   11.4   39   43.6   24   0]
```

To find the open-loop zeros of the given transfer function, we may use the following roots command:

```
p = [1   2   4]
r = roots(p)
```

The command and the computer output are shown next.

```
p = [1   2   4];
r = roots(p)

r =

   -1.0000 + 1.7321i
   -1.0000 - 1.7321i
```

Similarly, to find the complex-conjugate open-loop poles (the roots of $s^2 + 1.4s + 1 = 0$), we may enter the roots commands as follows:

```
q = roots(c)

q =

    -0.7000 + 0.7141i
    -0.7000 - 0.7141i
```

Thus the system has the following open-loop zeros and open-loop poles;

Open-loop zeros:	$s = -1 + j1.7321,$		$s = -1 - j1.7321$
Open-loop poles:	$s = -0.7 + j0.7141,$		$s = -0.7 - j0.7141$
	$s = 0,$	$s = -4,$	$s = -6$

MATLAB Program 6–1 will plot the root-locus diagram for this system. The plot is shown in Figure 6–15.

MATLAB Program 6–1

```
% ---------- Root-locus plot ----------

num = [0   0   0   1   2   4];
den = [1   11.4   39   43.6   24   0];
rlocus(num,den)

Warning: Divide by zero
v = [-10   10   -10   10]; axis(v)
grid
title('Root-Locus Plot of G(s) = K(s^2 + 2s + 4)/[s(s + 4)(s + 6)(s^2 + 1.4s + 1)]')
```

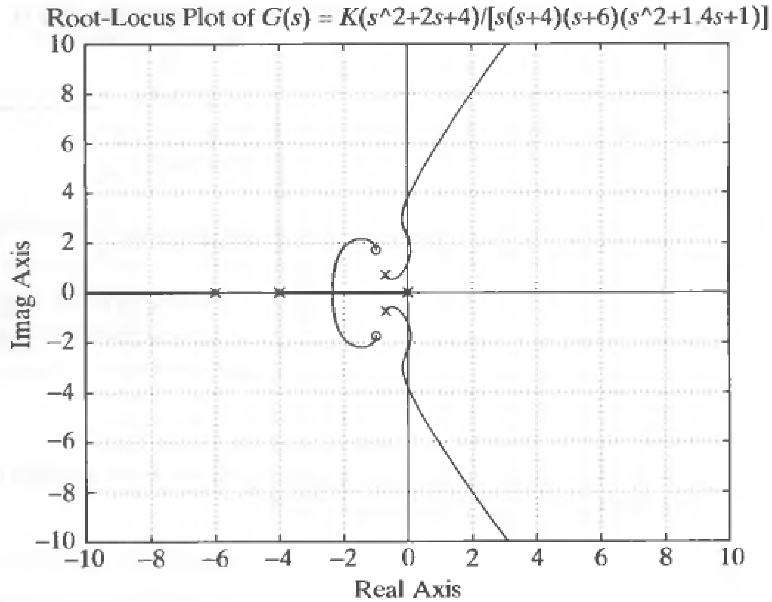


Figure 6–15
Root-locus plot.

EXAMPLE 6–4 Consider the system shown in Figure 6–16, where the open-loop transfer function $G(s)H(s)$ is

$$G(s)H(s) = \frac{K(s + 0.2)}{s^2(s + 3.6)}$$

The open-loop zero is at $s = -0.2$, and open-loop poles are at $s = 0, s = 0$, and $s = -3.6$.

MATLAB Program 6–2 generates a root-locus plot. The resulting root-locus plot is shown in Figure 6–17.

MATLAB Program 6–2

```
% ---------- Root-locus plot ----------

num = [0   0   1   0.2];
den = [1   3.6   0   0];
rlocus(num,den)
v = [-4   2   -4   4]; axis(v)
grid
title('Root-Locus Plot of G(s) = K(s + 0.2)/[s^2(s + 3.6)]')
```

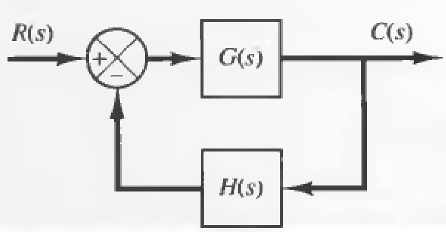


Figure 6–16
Control system.

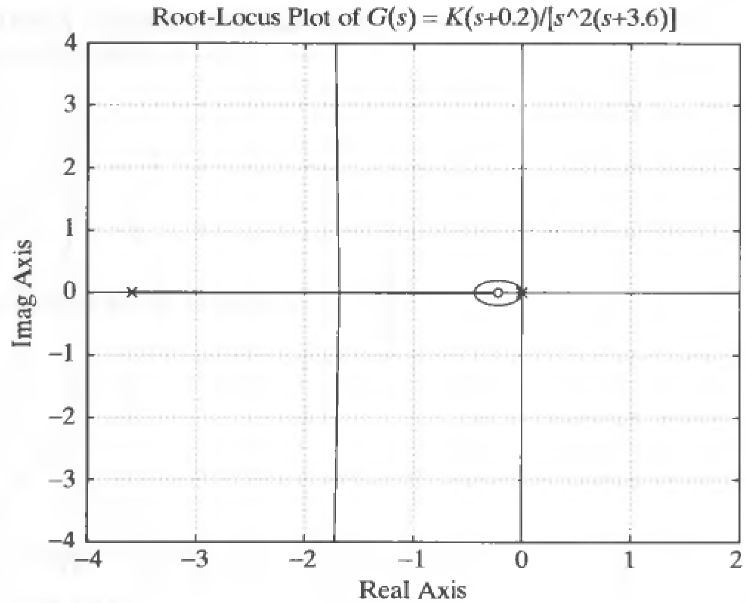


Figure 6–17
Root-locus plot.

EXAMPLE 6–5 Consider the system shown in Figure 6–18. Plot root loci with a square aspect ratio so that a line with slope 1 is a true 45° line.

To set the plot region on the screen to be square, enter the command axis('square'). With this command, a line with slope 1 is at a true 45°, not skewed by the irregular shape of the screen. (It is important to note that a hard-copy plot may or may not be of a square region depending on a printer.)

MATLAB Program 6–3 produces a root-locus plot in a square region. The resulting plot is shown in Figure 6–19.

MATLAB Program 6–3

```
% ---------- Root-locus plot ----------

num = [0   0   0   1   1];
den = [1   3   12   -16   0];
rlocus(num,den)
v = [-6   6   -6   6]; axis(v);axis('square')
grid
title('Root-Locus Plot of G(s) = K(s + 1)/[s(s - 1)(s^2 + 4s + 16)]')
```

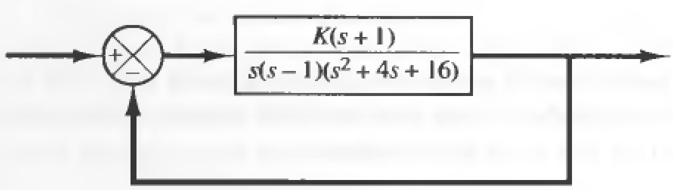


Figure 6–18
Control system.

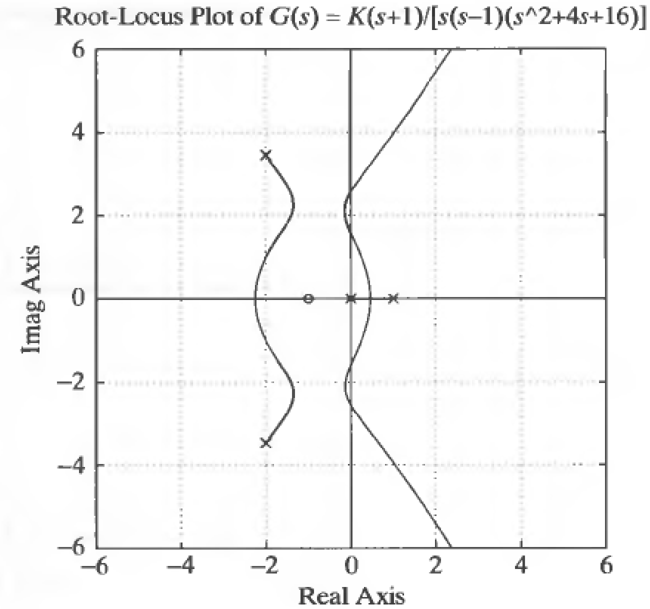


Figure 6–19
Root-locus plot.

EXAMPLE 6–6 Consider the system whose open-loop transfer function $G(s)H(s)$ is

$$G(s)H(s) = \frac{K}{s(s + 0.5)(s^2 + 0.6s + 10)}$$

$$= \frac{K}{s^4 + 1.1s^3 + 10.3s^2 + 5s}$$

There are no open-loop zeros. Open-loop poles are located at $s = -0.3 + j3.1480$, $s = -0.3 - j3.1480$, $s = -0.5$, and $s = 0$.

Entering MATLAB Program 6–4 into the computer, we obtain the root-locus plot shown in Figure 6–20.

MATLAB Program 6–4

```
% ---------- Root-locus plot ----------

num = [0   0   0   0   1];
den = [1   1.1   10.3   5   0];
rlocus(num,den)
grid
title('Root-Locus Plot of G(s) = K/[s(s + 0.5)(s^2 + 0.6s + 10)]')
```

Notice that in the regions near $x = -0.3$, $y = 2.3$ and $x = -0.3$, $y = -2.3$ two loci approach each other. We may wonder if these two branches should touch or not. To explore this situation, we may plot the root loci using the command

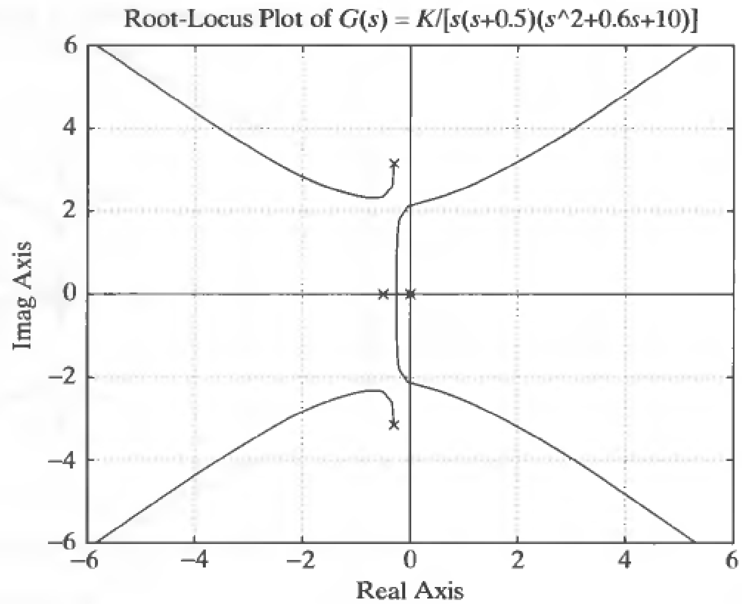


Figure 6–20
Root-locus plot.

```
r = rlocus(num,den)
plot(r,'o')
```

as shown in MATLAB Program 6–5. Figure 6–21 shows the resulting plot.

MATLAB Program 6–5

```
% ---------- Root-locus plot ----------

num = [0   0   0   0   1];
den = [1   1.1   10.3   5   0];
r = rlocus(num,den);
plot(r,'or')
v = [-6   6   -6   6]; axis(v)
grid
title('Root-Locus Plot of G(s) = K/[s(s + 0.5)(s^2 + 0.6s + 10)]')
xlabel('Real Axis')
ylabel('Imag Axis')

% ***** Note that the command 'plot(r,'or')' gives small circles
% in the screen plot in red color *****
```

Since there are no computed points near $(-0.3, 2.3)$ and $(-0.3, -2.3)$, it is necessary to adjust steps in gain K. By a trial and error approach, we find the particular region of interest to be $20 \le K \le 30$. By entering MATLAB Program 6–6, we obtain the root-locus plot shown in Figure 6–22. From this plot, it is clear that the two branches that approach in the upper half-plane (or in the lower half-plane) do not touch.

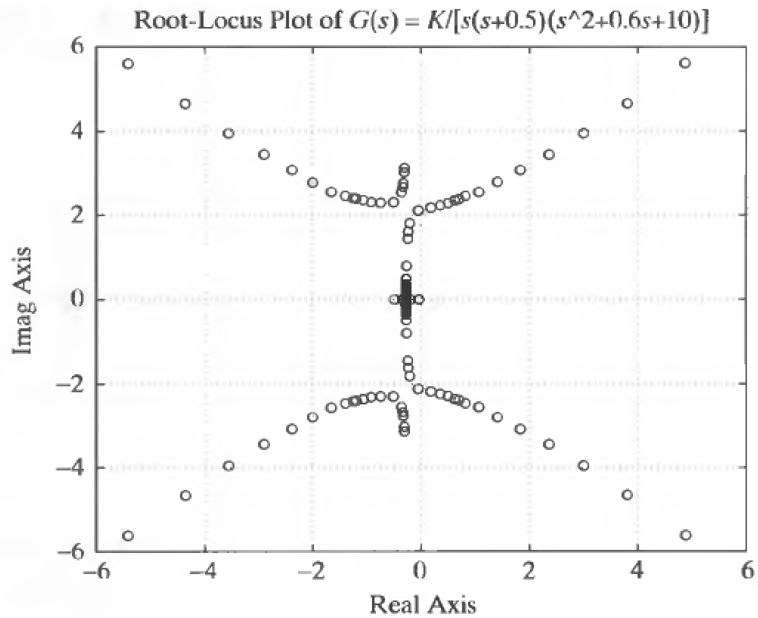


Figure 6–21
Root-locus plot.

MATLAB Program 6–6

```
% ---------- Root-locus plot ----------

num = [0   0   0   0   1];
den = [1   1.1   10.3   5   0];
K1 = 0:0.2:20;
K2 = 20:0.1:30;
K3 = 30:5:1000;
K = [K1   K2   K3];
r = rlocus(num,den,K);
plot(r,'ob')
v = [-4   4   -4   4]; axis(v)
grid
title('Root-Locus Plot of G(s) = K/[s(s + 0.5)(s^2 + 0.6s + 10)]')
xlabel('Real Axis')
ylabel('Imag Axis')

% ***** Note that the command 'plot(r,'ob')' gives small circles
% in the screen plot in blue color *****
```

EXAMPLE 6–7 Consider the system shown in Figure 6–23. The system equations are

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

$$y = \mathbf{C}\mathbf{x} + Du$$

$$u = r - y$$

In this example problem we shall obtain the root-locus diagram of the system defined in state space. Let us assume, for example, that matrices **A**, **B**, **C**, and D are given by

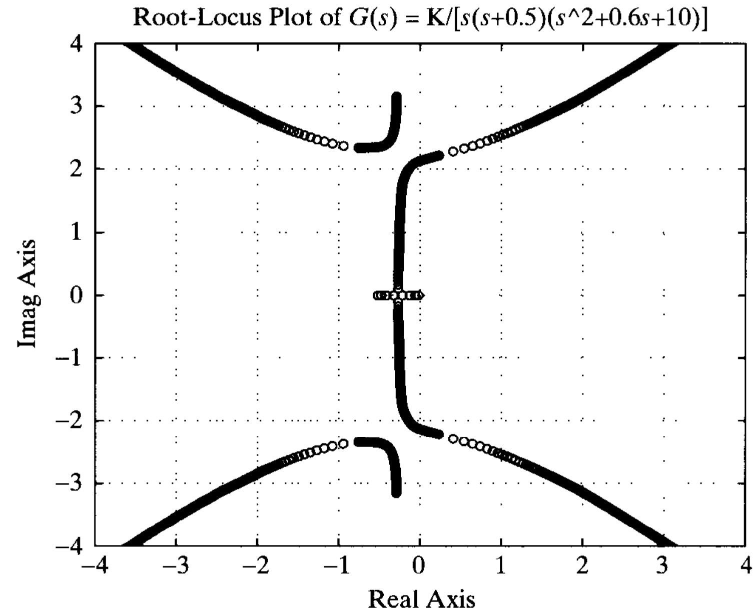


Figure 6–22
Root-locus plot.

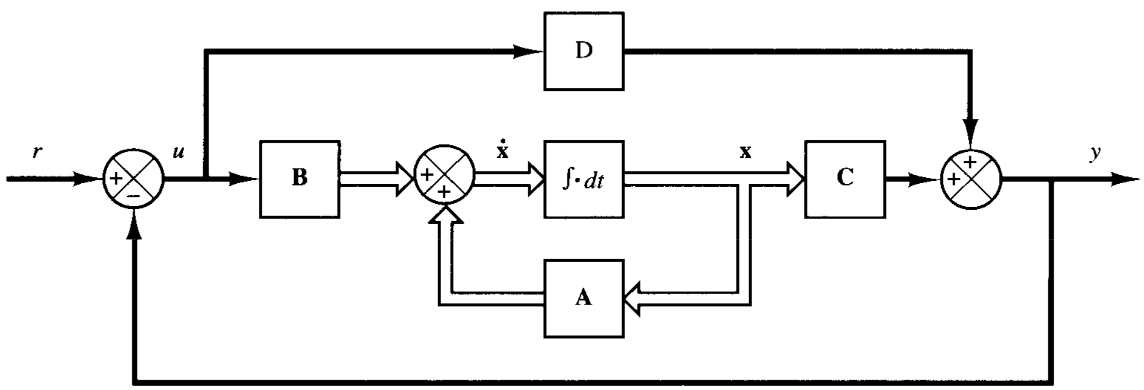


Figure 6–23
Closed-loop control system.

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -160 & -56 & -14 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ -14 \end{bmatrix} \tag{6–18}$$

$$\mathbf{C} = [1 \quad 0 \quad 0], \qquad D = [0]$$

The root-locus plot for this system can be obtained with MATLAB by use of the following command:

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

This command will produce the same root-locus plot as can be obtained by use of the rlocus (num,den) command, where num and den are obtained from

```
[num,den] = ss2tf(A,B,C,D)
```

as follows:

```
num = [0   0   1   0]
den = [1   14   56   160]
```

MATLAB Program 6–7 gives a program that will generate the root-locus plot as shown in Figure 6–24.

MATLAB Program 6–7

```
% ---------- Root-locus plot ----------

% ***** Root-locus plot of system defined in state space *****

% ***** Enter matrices A, B, C, and D *****

A = [0  1  0;0  0  1;-160  -56  -14];
B = [0;1;-14];
C = [1  0  0];
D = [0];

% ***** Enter rlocus(A,B,C,D) command in the computer *****

rlocus(A,B,C,D)
grid
title('Root-Locus Plot of System defined in State Space')
```

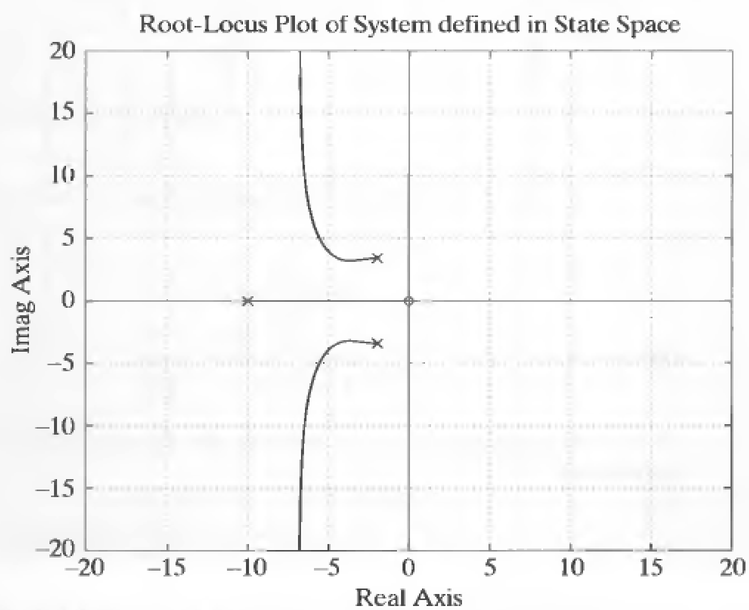


Figure 6–24 Root-locus plot of system defined in state space, where **A**, **B**, **C**, and D are as given by Equation (6–18).

6–5 SPECIAL CASES

In this section we shall consider two special cases. One is the case in which the gain K does not appear as a multiplicative factor, and in the other the closed-loop system is a positive-feedback system, rather than a negative-feedback system.

Constructing root loci when a variable parameter does not appear as a multiplying factor. In some cases the variable parameter K may not appear as a multiplying factor of $G(s)H(s)$. In such cases it may be possible to rewrite the characteristic equation such that the variable parameter K appears as a multiplying factor of $G(s)H(s)$. Example 6–8 illustrates how to proceed in such a case.

EXAMPLE 6–8 Consider the system in Figure 6–25. Draw a root-locus diagram. Then determine the value of k such that the damping ratio of the dominant closed-loop poles is 0.4.

Here the system involves velocity feedback. The open-loop transfer function is

$$\text{Open-loop transfer function} = \frac{20}{s(s+1)(s+4)+20ks}$$

Notice that the adjustable variable k does not appear as a multiplying factor. The characteristic equation for the system is

$$s^3 + 5s^2 + 4s + 20 + 20ks = 0 \tag{6–19}$$

Define

$$20k = K$$

Then Equation (6–19) becomes

$$s^3 + 5s^2 + 4s + Ks + 20 = 0 \tag{6–20}$$

Dividing both sides of Equation (6–20) by the sum of the terms that do not contain K, we get

$$1 + \frac{Ks}{s^3 + 5s^2 + 4s + 20} = 0$$

or

$$1 + \frac{Ks}{(s+j2)(s-j2)(s+5)} = 0 \tag{6–21}$$

Equation (6–21) is now of the form of Equation (6–5).

We shall now sketch the root loci of the system given by Equation (6–21). Notice that the open-loop poles are located at $s = j2, s = -j2, s = -5$, and the open-loop zero is located at $s = 0$. The root locus exists on the real axis between 0 and -5. Since

$$\lim_{s\to\infty} \frac{Ks}{(s+j2)(s-j2)(s+5)} = \lim_{s\to\infty} \frac{K}{s^2}$$

we have

$$\text{Angle of asymptote} = \frac{\pm 180°(2k+1)}{2} = \pm 90°$$

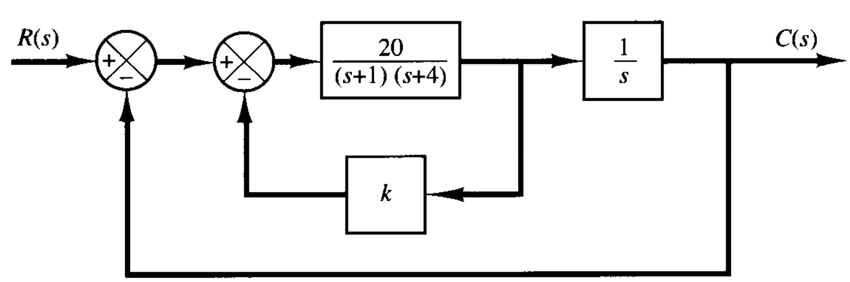


Figure 6–25 Control system.

The intersection of the asymptotes with the real axis can be found from

$$\lim_{s\to\infty} \frac{Ks}{s^3 + 5s^2 + 4s + 20} = \lim_{s\to\infty} \frac{K}{s^2 + 5s + \cdots} = \lim_{s\to\infty} \frac{K}{(s + 2.5)^2}$$

as

$$\sigma_a = -2.5$$

The angle of departure (angle θ) from the pole at $s = j2$ is obtained as follows:

$$\theta = 180° - 90° - 21.8° + 90° = 158.2°$$

Thus, the angle of departure from the pole $s = j2$ is 158.2°. Figure 6–26 shows a root-locus plot for the system.

Note that the closed-loop poles with $\zeta = 0.4$ must lie on straight lines passing through the origin and making the angles $\pm 66.42°$ with the negative real axis. In the present case, there are two intersections of the root-locus branch in the upper half s plane and the straight line of angle 66.42°. Thus, two values of K will give the damping ratio ζ of the closed-loop poles equal to 0.4. At point P, the value of K is

$$K = \left| \frac{(s + j2)(s - j2)(s + 5)}{s} \right|_{s=-1.0490+j2.4065} = 8.9801$$

Hence

$$k = \frac{K}{20} = 0.4490 \qquad \text{at point } P$$

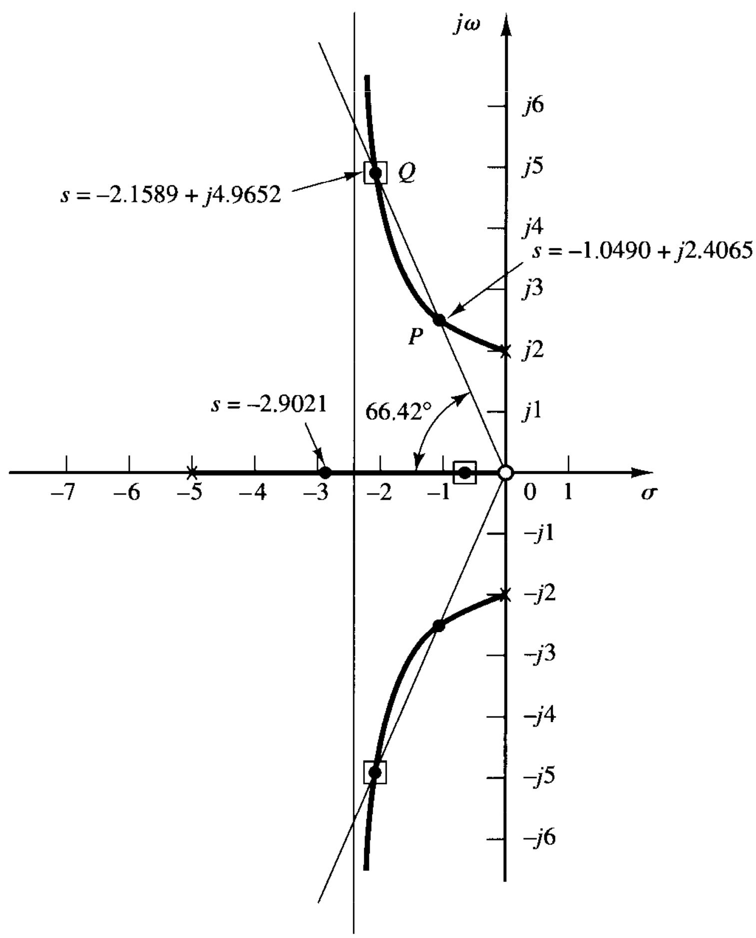


Figure 6–26 Root-locus plot for the system shown in Figure 6–25.

At point Q, the value of K is

$$K = \left| \frac{(s + j2)(s - j2)(s + 5)}{s} \right|_{s=-2.1589+j4.9652} = 28.260$$

Hence

$$k = \frac{K}{20} = 1.4130 \qquad \text{at point } Q$$

Thus, we have two solutions for this problem. For $k = 0.4490$, the three closed-loop poles are located at

$$s = -1.0490 + j2.4065, \qquad s = -1.0490 - j2.4065, \qquad s = -2.9021$$

For $k = 1.4130$, the three closed-loop poles are located at

$$s = -2.1589 + j4.9652, \qquad s = -2.1589 - j4.9652, \qquad s = -0.6823$$

It is important to point out that the zero at the origin is the open-loop zero, but not the closed-loop zero. This is evident, because the original system shown in Figure 6–25 does not have a closed-loop zero, since

$$\frac{C(s)}{R(s)} = \frac{20}{s(s + 1)(s + 4) + 20(1 + ks)}$$

The open-loop zero at $s = 0$ was introduced in the process of modifying the characteristic equation such that the adjustable variable $K = 20k$ was to appear as a multiplying factor.

We have obtained two different values of k to satisfy the requirement that the damping ratio of the dominant closed-loop poles be equal to 0.4. The closed-loop transfer function with $k = 0.4490$ is given by

$$\frac{C(s)}{R(s)} = \frac{20}{s^3 + 5s^2 + 12.98s + 20}$$

$$= \frac{20}{(s + 1.0490 + j2.4065)(s + 1.0490 - j2.4065)(s + 2.9021)}$$

The closed-loop transfer function with $k = 1.4130$ is given by

$$\frac{C(s)}{R(s)} = \frac{20}{s^3 + 5s^2 + 32.26s + 20}$$

$$= \frac{20}{(s + 2.1589 + j4.9652)(s + 2.1589 - j4.9652)(s + 0.6823)}$$

Notice that the system with $k = 0.4490$ has a pair of dominant complex-conjugate closed-loop poles, while in the system with $k = 1.4130$ the real closed-loop pole at $s = -0.6823$ is dominant, and the complex-conjugate closed-loop poles are not dominant. In this case, the response characteristic is primarily determined by the real closed-loop pole.

Let us compare the unit-step responses of both systems. MATLAB Program 6–8 may be used for plotting the unit-step response curves in one diagram. The resulting unit-step response curves [$c_1(t)$ for $k = 0.4490$ and $c_2(t)$ for $k = 1.4130$] are shown in Figure 6–27.

From Figure 6–27 we notice that the response of the system with $k = 0.4490$ is oscillatory. (The effect of the closed-loop pole at $s = -2.9021$ on the unit-step response is small.) For the system with $k = 1.4130$, the oscillations due to the closed-loop poles at $s = -2.1589 \pm j4.9652$ damp out much faster than purely exponential response due to the closed-loop pole at $s = -0.6823$.

MATLAB Program 6–8

```
% ---------- Unit-step response ----------

% ***** Enter numerators and denominators of system with
% k = 0.4490 and k = 1.4130, respectively. *****

num1 = [0   0   0   20];
den1 = [1   5   12.98   20];
num2 = [0   0   0   20];
den2 = [1   5   32.26   20];
t = 0:0.1:10;
[c1,x1,t] = step(num1,den1,t);
[c2,x2,t] = step(num2,den2,t);
plot(t,c1,t,c2)
text(2.5,1.12,'k = 0.4490')
text(3.7,0.85,'k = 1.4130')
grid
title('Unit-Step Responses of Two Systems')
xlabel('t Sec')
ylabel('Outputs c1 and c2')
```

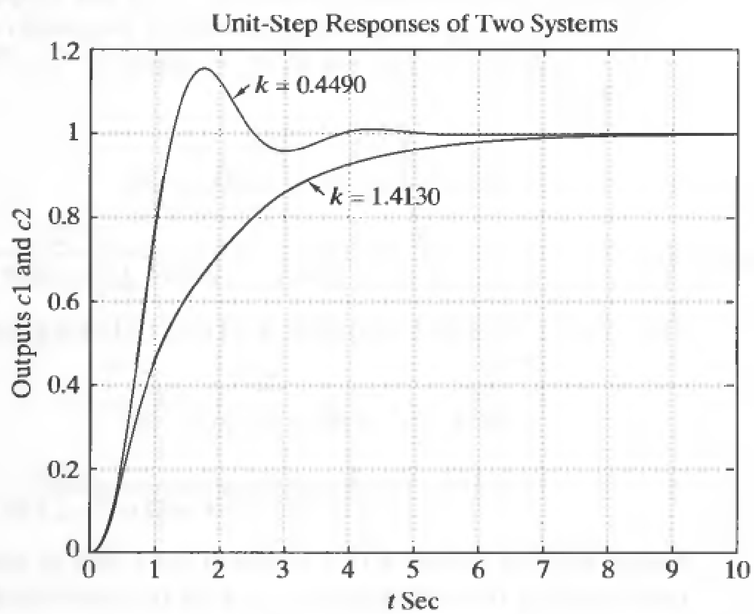


Figure 6–27 Unit-step response curves for the system shown in Figure 6–25 when the damping ratio ζ of the dominant closed-loop poles is set equal to 0.4. (Two possible values of k give the damping ratio ζ equal to 0.4.)

The system with $k = 0.4490$ (which exhibits a faster response with relatively small overshoot) has a much better response characteristic than the system with $k = 1.4130$ (which exhibits a slow overdamped response). Therefore, we should choose $k = 0.4490$ for the present system.

Root loci for positive-feedback systems.* In a complex control system, there may be a positive-feedback inner loop as shown in Figure 6–28. Such a loop is usu-

* Reference W–5.

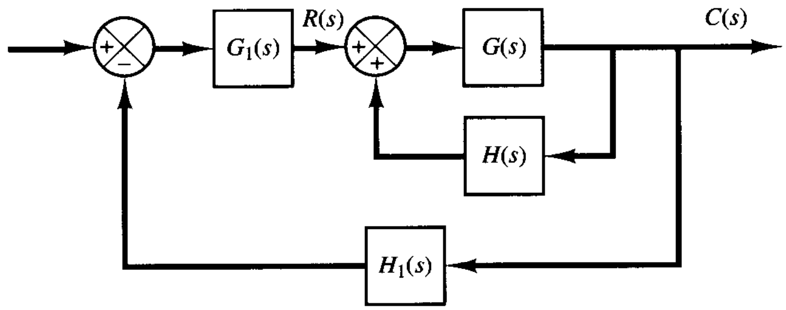


Figure 6–28
Control system.

ally stabilized by the outer loop. In what follows, we shall be concerned only with the positive-feedback inner loop. The closed-loop transfer function of the inner loop is

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

The characteristic equation is

$$1 - G(s)H(s) = 0 \tag{6–22}$$

This equation can be solved in a manner similar to the development of the root-locus method in Section 6–2. The angle condition, however, must be altered.

Equation (6–22) can be rewritten as

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

which is equivalent to the following two equations:

$$\underline{/G(s)H(s)} = 0° \pm k360° \qquad (k = 0, 1, 2, \ldots)$$

$$|G(s)H(s)| = 1$$

The total sum of all angles from the open-loop poles and zeros must be equal to $0° \pm k360°$. Thus the root locus follows a 0° locus in contrast to the 180° locus considered previously. The magnitude condition remains unaltered.

To illustrate the root-locus plot for the positive feedback system, we shall use the following transfer functions $G(s)$ and $H(s)$ as an example.

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

The gain K is assumed to be positive.

The general rules for constructing root loci given in Section 6–3 must be modified in the following way:

Rule 2 is modified as follows: If the total number of real poles and real zeros to the right of a test point on the real axis is even, then this test point lies on the root locus.

Rule 3 is modified as follows:

$$\text{Angles of asymptotes} = \frac{\pm k360°}{n - m} \qquad (k = 0, 1, 2, \ldots)$$

where n = number of finite poles of $G(s)H(s)$
m = number of finite zeros of $G(s)H(s)$

Rule 5 is modified as follows: When calculating the angle of departure (or angle of arrival) from a complex open-loop pole (or at a complex zero), subtract from 0° the sum of all angles of the vectors from all the other poles and zeros to the complex pole (or complex zero) in question, with appropriate signs included.

Other rules for constructing the root-locus plot remain the same. We shall now apply the modified rules to construct the root-locus plot.

1. Plot the open-loop poles ($s = -1 + j$, $s = -1 - j$, $s = -3$) and zero ($s = -2$) in the complex plane. As K is increased from 0 to ∞, the closed-loop poles start at the open-loop poles and terminate at the open-loop zeros (finite or infinite), just as in the case of negative-feedback systems.
2. Determine the root loci on the real axis. Root loci exist on the real axis between -2 and $+\infty$ and between -3 and $-\infty$.
3. Determine the asymptotes of the root loci. For the present system,

$$\text{Angle of asymptote} = \frac{\pm k360°}{3 - 1} = \pm 180°$$

This simply means that asymptotes are on the real axis.
4. Determine the breakaway and break-in points. Since the characteristic equation is

$$(s + 3)(s^2 + 2s + 2) - K(s + 2) = 0$$

we obtain

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

By differentiating K with respect to s, we obtain

$$\frac{dK}{ds} = \frac{2s^3 + 11s^2 + 20s + 10}{(s + 2)^2}$$

Note that

$$\begin{aligned} 2s^3 + 11s^2 + 20s + 10 &= 2(s + 0.8)(s^2 + 4.7s + 6.24) \\ &= 2(s + 0.8)(s + 2.35 + j0.77)(s + 2.35 - j0.77) \end{aligned}$$

Point $s = -0.8$ is on the root locus. Since this point lies between two zeros (a finite zero and an infinite zero), it is an actual break-in point. Points $s = -2.35 \pm j0.77$ do not satisfy the angle condition and, therefore, they are neither breakaway nor break-in points.
5. Find the angle of departure of the root locus from a complex pole. For the complex pole at $s = -1 + j$, the angle of departure θ is

$$\theta = 0° - 27° - 90° + 45°$$

or

$$\theta = -72°$$

(The angle of departure from the complex pole at $s = -1 - j$ is 72°.)

6. Choose a test point in the broad neighborhood of the $j\omega$ axis and the origin and apply the angle condition. Locate a sufficient number of points that satisfy the angle condition.

Figure 6–29 shows the root loci for the given positive-feedback system. The root loci are shown with dashed lines and curve.

Note that if

$$K > \left. \frac{(s+3)(s^2+2s+2)}{s+2} \right|_{s=0} = 3$$

one real root enters the right-half s plane. Hence, for values of K greater than 3, the system becomes unstable. (For $K > 3$, the system must be stabilized with an outer loop.)

Note that the closed-loop transfer function for the positive-feedback system is given by

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 - G(s)H(s)}$$
$$= \frac{K(s+2)}{(s+3)(s^2+2s+2) - K(s+2)}$$

To compare this root-locus plot with that of the corresponding negative-feedback system, we show in Figure 6–30 the root loci for the negative-feedback system whose closed-loop transfer function is

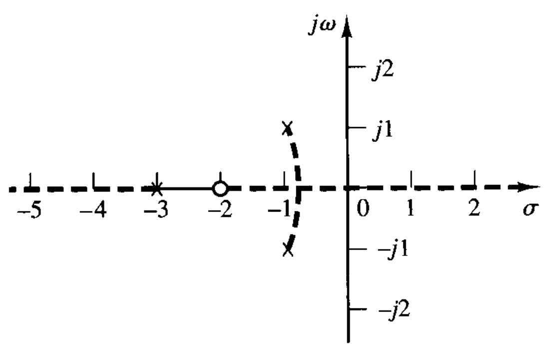


Figure 6–29
Root-locus plot for the positive-feedback system with $G(s) = K(s+2)/[(s+3)(s^2+2s+2)]$, $H(s) = 1$.

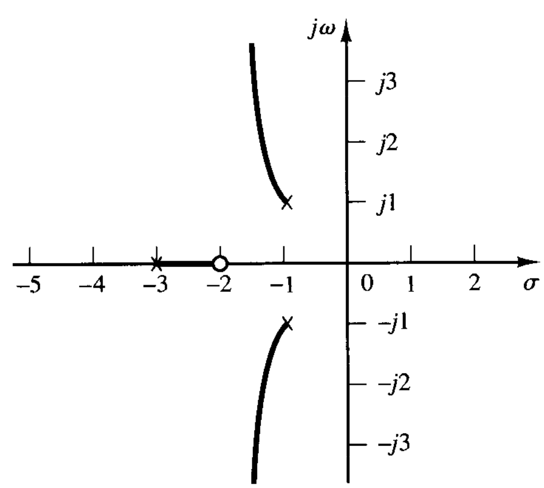
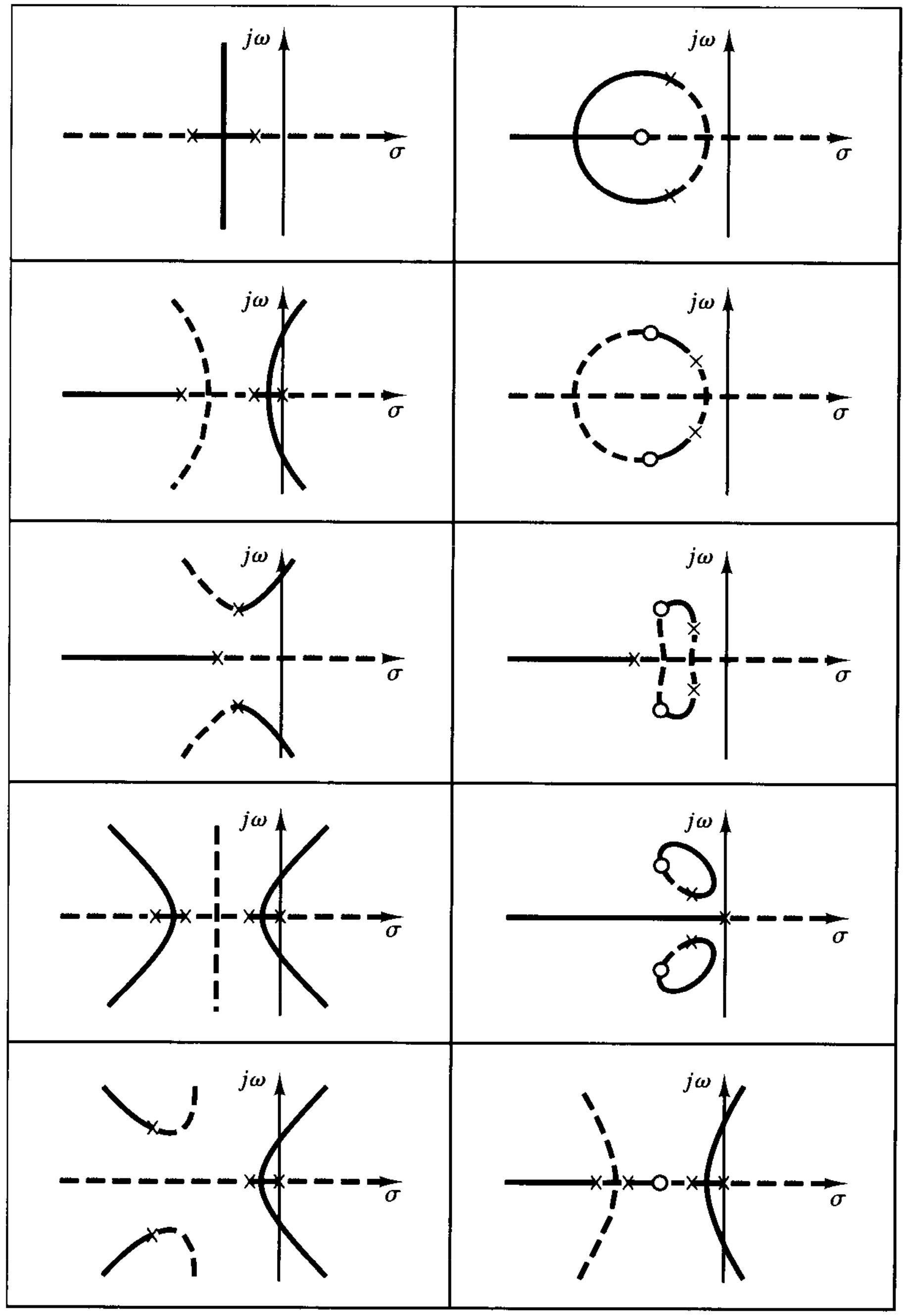


Figure 6–30
Root-locus plot for the negative-feedback system with $G(s) = K(s+2)/[(s+3)(s^2+2s+2)]$, $H(s) = 1$.

$$\frac{C(s)}{R(s)} = \frac{K(s+2)}{(s+3)(s^2+2s+2)+K(s+2)}$$

Table 6–2 shows various root-locus plots of negative-feedback and positive-feedback systems. The closed-loop transfer functions are given by

Table 6–2 Root-Locus Plots of Negative-Feedback and Positive-Feedback Systems



Heavy lines and curves correspond to negative-feedback systems; dashed lines and curves correspond to positive-feedback systems.

$$\frac{C}{R} = \frac{G}{1 + GH}, \qquad \text{for negative-feedback systems}$$

$$\frac{C}{R} = \frac{G}{1 - GH}, \qquad \text{for positive-feedback systems}$$

where GH is the open-loop transfer function. In Table 6–2, the root loci for negative-feedback systems are drawn with heavy lines and curves and those for positive-feedback systems are drawn with dashed lines and curves.

6–6 ROOT-LOCUS ANALYSIS OF CONTROL SYSTEMS

In this section we shall first discuss orthogonality of the root loci and constant-gain loci for the closed-loop systems. Next, we discuss conditionally stable systems. Finally, we analyze nonminimum-phase systems.

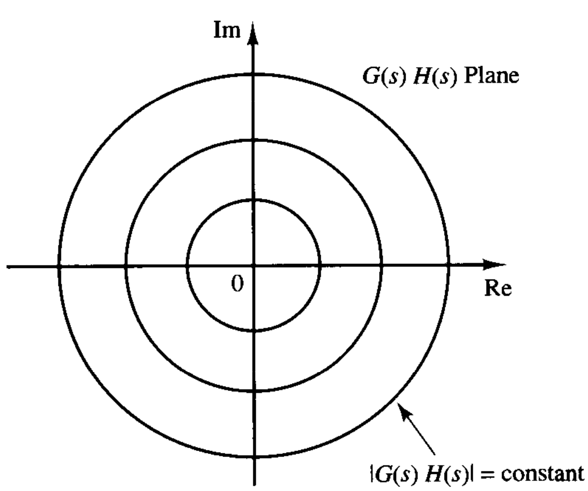
Orthogonality of root loci and constant-gain loci. Consider the system whose open-loop transfer function is $G(s)H(s)$. In the $G(s)H(s)$ plane, the loci of $|G(s)H(s)| =$ constant are circles centered at the origin, and the loci corresponding to $\underline{/G(s)H(s)} = \pm 180°(2k + 1)$ $(k = 0, 1, 2, \ldots)$ lie on the negative real axis of the $G(s)H(s)$ plane, as shown in Figure 6–31. [Note that the complex plane employed here is not the s plane, but the $G(s)H(s)$ plane.]

The root loci and constant-gain loci in the s plane are conformal mappings of the loci of $\underline{/G(s)H(s)} = \pm 180°(2k + 1)$ and of $|G(s)H(s)| =$ constant in the $G(s)H(s)$ plane.

Since the constant-phase and constant-gain loci in the $G(s)H(s)$ plane are orthogonal, the root loci and constant-gain loci in the s plane are orthogonal. Figure 6–32(a) shows the root loci and constant-gain loci for the following system:

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

Notice that since the pole–zero configuration is symmetrical about the real axis the constant-gain loci are also symmetrical about the real axis.



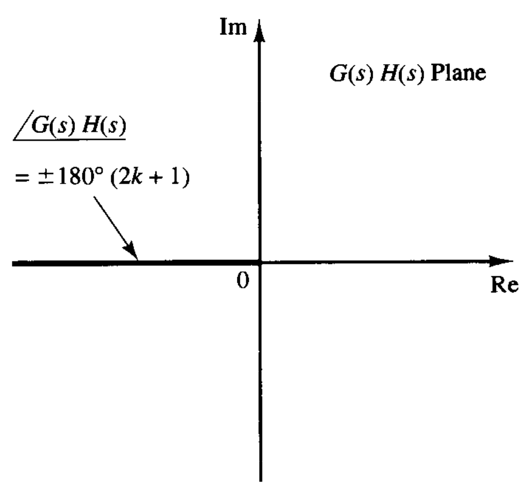


Figure 6–31 Plots of constant-gain and constant-phase loci in the $G(s)H(s)$ plane.

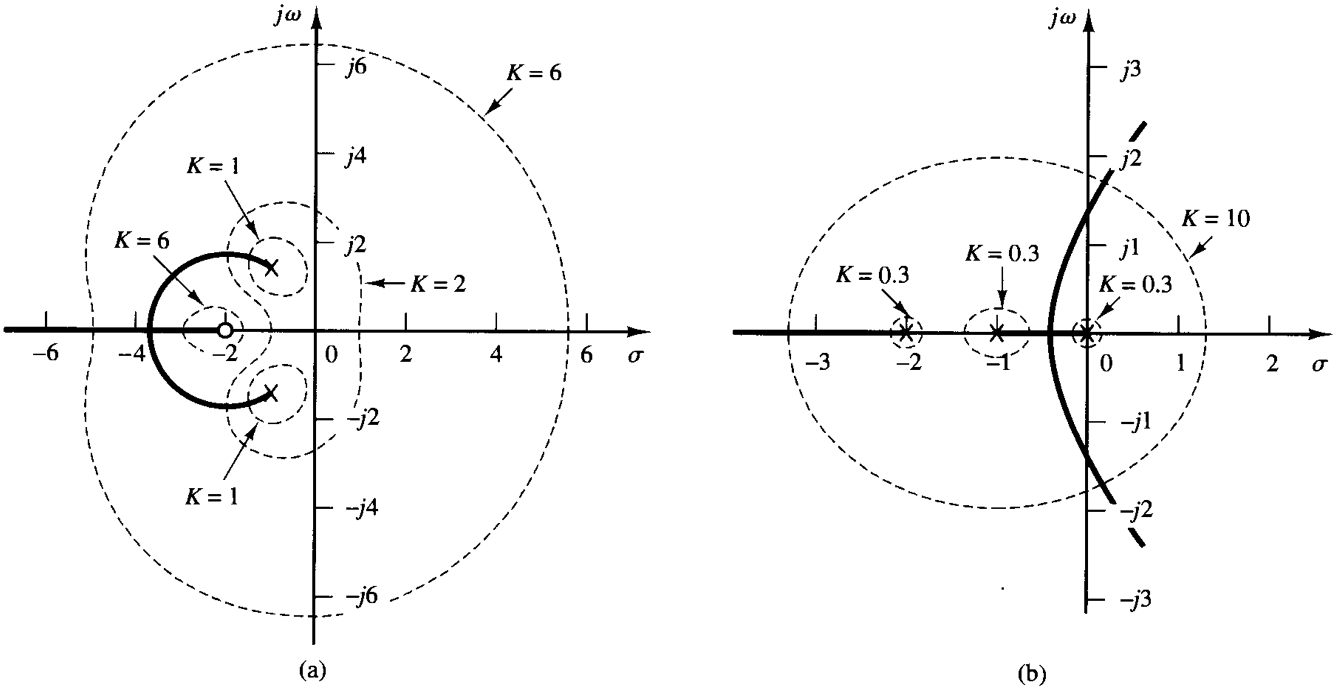


Figure 6–32
Plots of root loci and constant-gain loci. (a) System with $G(s) = K(s + 2)/(s^2 + 2s + 3)$, $H(s) = 1$; (b) system with $G(s) = K/[s(s + 1)(s + 2)]$, $H(s) = 1$.

Figure 6–32(b) shows the root loci and constant-gain loci for the system:

$$G(s) = \frac{K}{s(s + 1)(s + 2)}, \qquad H(s) = 1$$

Notice that since the configuration of the poles in the s plane is symmetrical about the real axis and the line parallel to the imaginary axis passing through point ($\sigma = -1, \omega = 0$), the constant-gain loci are symmetrical about the $\omega = 0$ line (real axis) and the $\sigma = -1$ line.

Conditionally stable systems. Consider the system shown in Figure 6–33(a). The root loci for this system can be plotted by applying the general rules and procedure for constructing root loci. A root-locus plot for this system is shown in Figure 6–33(b). It can be seen that this system is stable only for limited ranges of the value of K; that is, $0 < K < 14$ and $64 < K < 195$. The system becomes unstable for $14 < K < 64$ and $195 < K$. If K assumes a value corresponding to unstable operation, the system may break down or may become nonlinear due to a saturation nonlinearity that may exist. Such a system is called *conditionally stable.*

In practice, conditionally stable systems are not desirable. Conditional stability is dangerous but does occur in certain systems, in particular, a system that has an unstable feedforward path. Such a feedforward path may occur if the system has a minor loop. It is advisable to avoid such conditional stability since, if the gain drops beyond the critical value for some reason, the system becomes unstable. Note that the addition of a proper compensating network will eliminate conditional stability. [An addition of a zero will cause the root loci to bend to the left. (See Section 7–2). Hence conditional stability may be eliminated by adding proper compensation.]

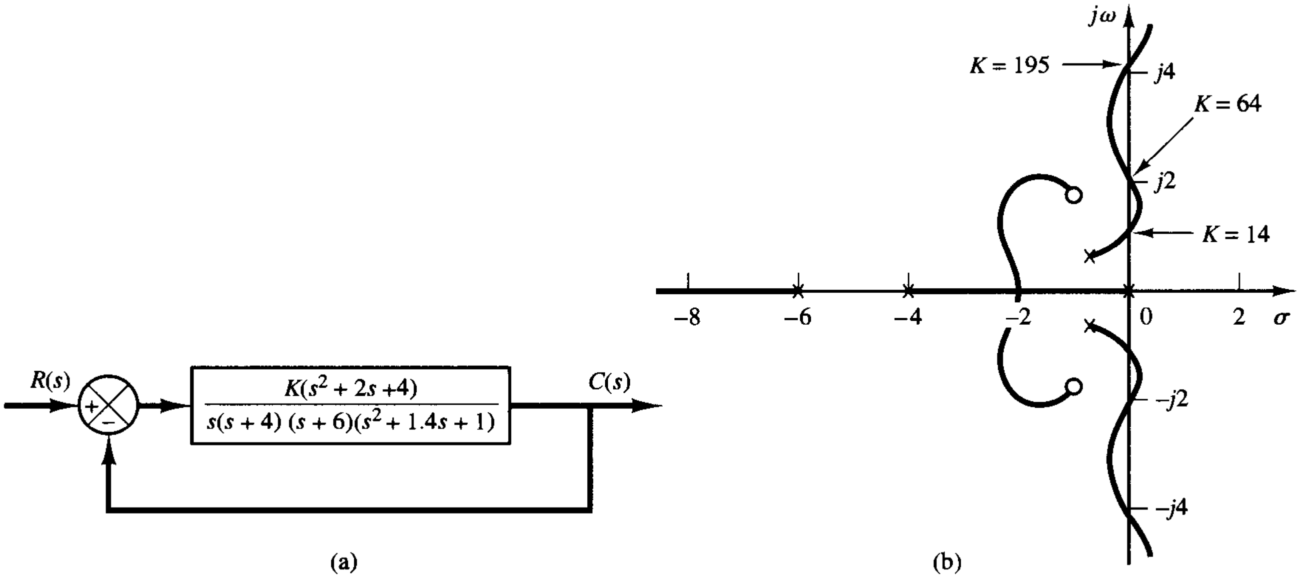


Figure 6–33
(a) Conditionally stable system; (b) root-locus plot.

Nonminimum-phase systems. If all the poles and zeros of a system lie in the left-half s plane, then the system is called *minimum phase.* If a system has at least one pole or zero in the right-half s plane, then the system is called *nonminimum phase.* The term nonminimum phase comes from the phase shift characteristics of such a system when subjected to sinusoidal inputs.

Consider the system shown in Figure 6–34(a). For this system

$$G(s) = \frac{K(1 - T_a s)}{s(Ts + 1)} \qquad (T_a > 0), \qquad H(s) = 1$$

This is a nonminimum-phase system since there is one zero in the right-half s plane. For this system, the angle condition becomes

$$\angle G(s) = \angle -\frac{K(T_a s - 1)}{s(Ts + 1)}$$

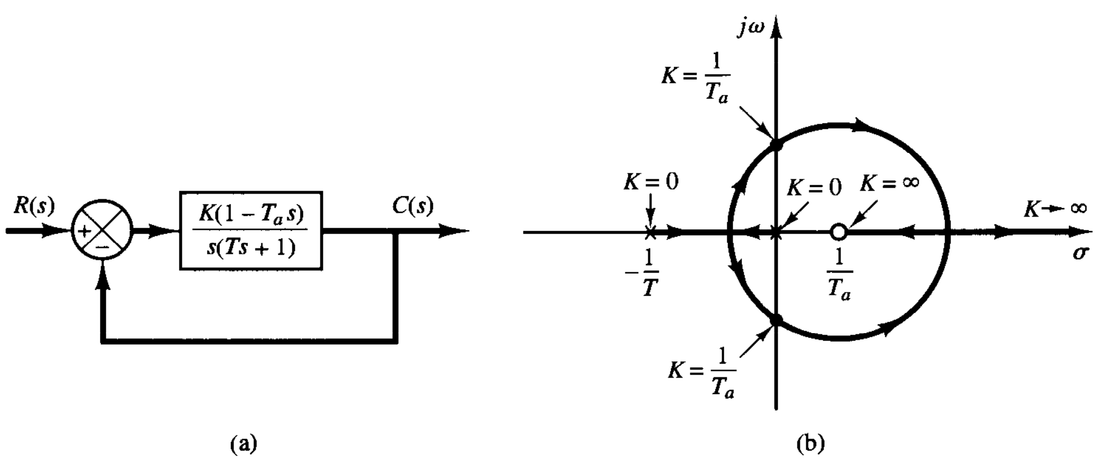


Figure 6–34
(a) Nonminimum phase system; (b) root-locus plot.

$$= \underline{\Big/ \frac{K(T_a s - 1)}{s(Ts + 1)}} + 180°$$

$$= \pm 180°(2k + 1) \qquad (k = 0, 1, 2, \ldots)$$

or

$$\underline{\Big/ \frac{K(T_a s - 1)}{s(Ts + 1)}} = 0° \tag{6–23}$$

The root loci can be obtained from Equation (6–23). Figure 6–34(b) shows a root-locus plot for this system. From the diagram, we see that the system is stable if the gain K is less than $1/T_a$.

6–7 ROOT LOCI FOR SYSTEMS WITH TRANSPORT LAG

Figure 6–35 shows a thermal system in which hot air is circulated to keep the temperature of a chamber constant. In this system, the measuring element is placed downstream a distance L ft from the furnace, the air velocity is v ft/sec, and $T = L/v$ sec would elapse before any change in the furnace temperature is sensed by the thermometer. Such a delay in measuring, delay in controller action, or delay in actuator operation, and the like, is called *transport lag* or *dead time*. Dead time is present in most process control systems.

The input $x(t)$ and the output $y(t)$ of a transport lag or dead time element are related by

$$y(t) = x(t - T)$$

where T is dead time. The transfer function of transport lag or dead time is given by

$$\text{Transfer function of transport lag or dead time} = \frac{\mathscr{L}[x(t - T)1(t - T)]}{\mathscr{L}[x(t)1(t)]}$$

$$= \frac{X(s)e^{-Ts}}{X(s)} = e^{-Ts}$$

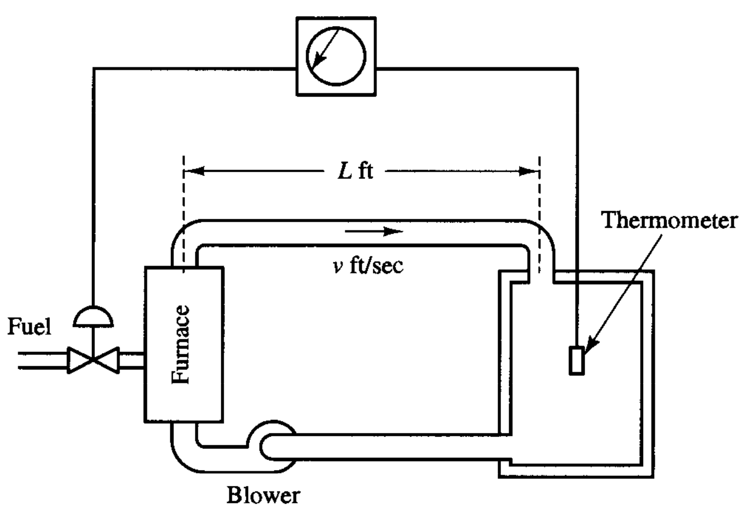


Figure 6–35
Thermal system.

Suppose that the feedforward transfer function of this thermal system can be approximated by

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

as shown in Figure 6–36. Let us construct a root-locus plot for this system. The characteristic equation for this closed-loop system is

$$1 + \frac{Ke^{-Ts}}{s + 1} = 0 \qquad (6\text{–}24)$$

It is noted that for systems with transport lag the rules of construction presented earlier need to be modified. For example, the number of the root-locus branches is infinite, since the characteristic equation has an infinite number of roots. The number of asymptotes is infinite. They are all parallel to the real axis of the s plane.

From Equation (6–24), we obtain

$$\frac{Ke^{-Ts}}{s + 1} = -1$$

Thus, the angle condition becomes

$$\angle \frac{Ke^{-Ts}}{s + 1} = \angle e^{-Ts} - \angle s + 1 = \pm 180°(2k + 1) \qquad (k = 0, 1, 2, \ldots) \qquad (6\text{–}25)$$

To find the angle of e^{-Ts}, substitute $s = \sigma + j\omega$. Then we obtain

$$e^{-Ts} = e^{-T\sigma - j\omega T}$$

Since $e^{-T\sigma}$ is a real quantity, the angle of $e^{-T\sigma}$ is zero. Hence

$$\begin{aligned} \angle e^{-Ts} = \angle e^{-j\omega T} &= \angle \cos \omega T - j \sin \omega T \\ &= -\omega T \qquad \text{(radians)} \\ &= -57.3\omega T \qquad \text{(degrees)} \end{aligned}$$

The angle condition, Equation (6–25), then becomes

$$-57.3\omega T - \angle s + 1 = \pm 180°(2k + 1)$$

Since T is a given constant, the angle of e^{-Ts} is a function of ω only.

We shall next determine the angle contribution due to e^{-Ts}. For $k = 0$, the angle condition may be written

$$\angle s + 1 = \pm 180° - 57.3°\omega T \qquad (6\text{–}26)$$

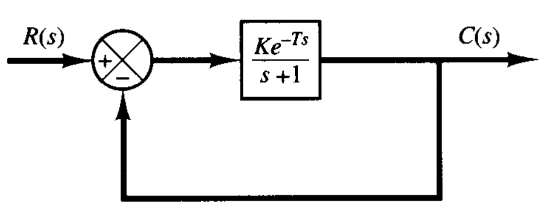


Figure 6–36
Block diagram of the system shown in Figure 6–35.

Since the angle contribution of e^{-Ts} is zero for $\omega = 0$, the real axis from -1 to $-\infty$ forms a part of the root loci. Now assume a value ω_1 for ω and compute $57.3°\omega_1 T$. At point -1 on the negative real axis, draw a line that makes an angle of $180° - 57.3°\omega_1 T$ with the real axis. Find the intersection of this line and the horizontal line $\omega = \omega_1$. This intersection, point P in Figure 6–37(a), is a point satisfying Equation (6–26) and hence is on a root locus. Continuing the same process, we obtain the root-locus plot as shown in Figure 6–37(b).

Note that as s approaches minus infinity, the open-loop transfer function

$$\frac{Ke^{-Ts}}{s+1}$$

approaches minus infinity since

$$\begin{aligned}\lim_{s=-\infty} \frac{Ke^{-Ts}}{s+1} &= \left.\frac{\dfrac{d}{ds}(Ke^{-Ts})}{\dfrac{d}{ds}(s+1)}\right|_{s=-\infty} \\ &= -KTe^{-Ts}\Big|_{s=-\infty} \\ &= -\infty\end{aligned}$$

Therefore, $s = -\infty$ is a pole of the open-loop transfer function. Thus, root loci start from $s = -1$ or $s = -\infty$ and terminate at $s = \infty$, as K increases from zero to infinity. Since the right-hand side of the angle condition given by Equation (6–25) has an infinite number of values, there are an infinite number of root loci, as the value of k ($k = 0, 1, 2, \ldots$) goes from zero to infinity. For example, if $k = 1$, the angle condition becomes

$$\begin{aligned}\underline{/s+1} &= \pm 540° - 57.3°\omega T \qquad \text{(degrees)} \\ &= \pm 3\pi - \omega T \qquad \text{(radians)}\end{aligned}$$

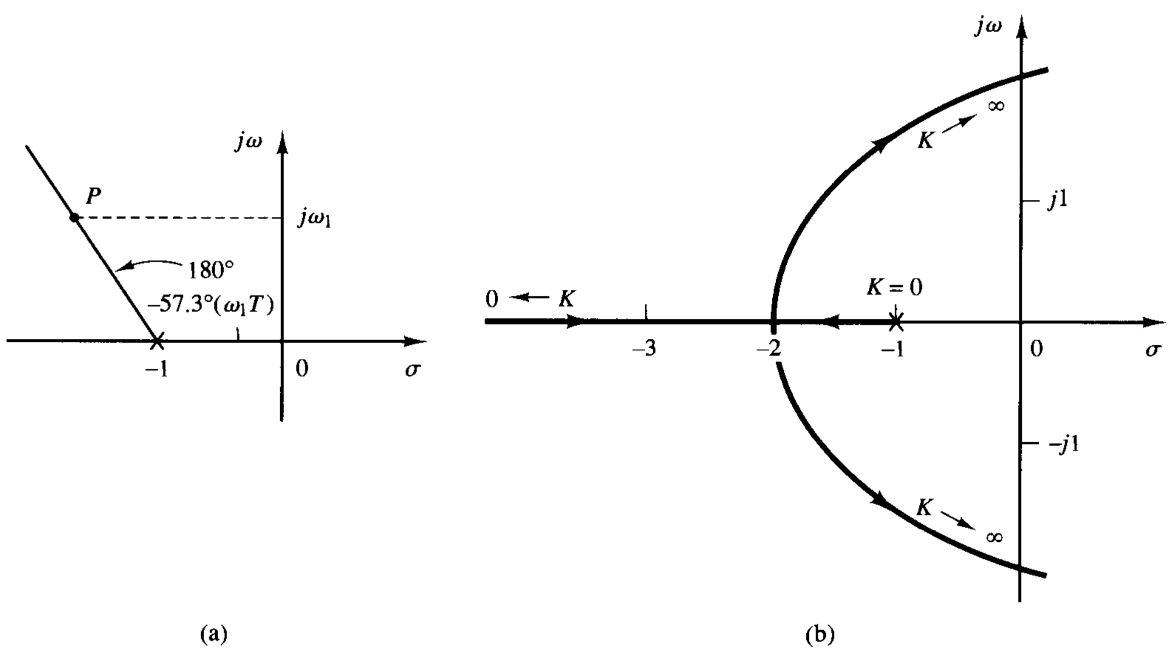


Figure 6–37
(a) Construction of the root locus; (b) root-locus plot.

The construction of the root loci for $k = 1$ is the same as that for $k = 0$. A plot of root loci for $k = 0, 1$, and 2 when $T = 1$ sec is shown in Figure 6–38.

The magnitude condition states that

$$\left|\frac{Ke^{-Ts}}{s + 1}\right| = 1$$

Since the magnitude of e^{-Ts} is equal to that of $e^{-T\sigma}$ or

$$\left|e^{-Ts}\right| = \left|e^{-T\sigma}\right| \cdot \left|e^{-j\omega T}\right| = e^{-T\sigma}$$

the magnitude condition becomes

$$\left|s + 1\right| = Ke^{-T\sigma}$$

The root loci shown in Figure 6–38 are graduated in terms of K when $T = 1$ sec.

Although there are an infinite number of root-locus branches, the primary branch that lies between $-j\pi$ and $j\pi$ is most important. Referring to Figure 6–38, the critical value of K at the primary branch is equal to 2, while the critical values of K at other branches are much higher (8, 14, . . .). Therefore, the critical value $K = 2$ on the primary branch is most significant from the stability viewpoint. The transient response of the system is determined by the roots located closest to the $j\omega$ axis and lie on the primary branch. In summary, the root-locus branch corresponding to $k = 0$ is the dominant one; other branches corresponding to $k = 1, 2, 3, \ldots$ are not so important and may be neglected.

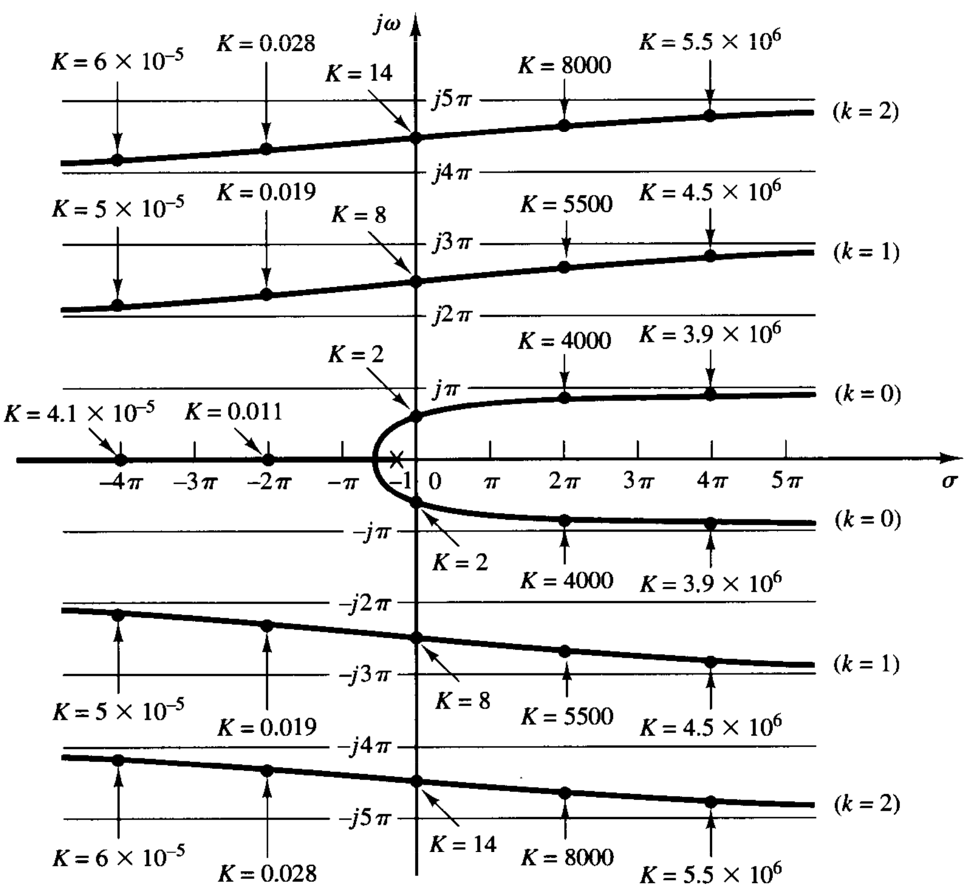


Figure 6–38 Root-locus plot for the system shown in Figure 6–36 ($T = 1$ sec).

This example illustrates the fact that dead time can cause instability even in the first-order system because the root loci enter the right-half s plane for large values of K. Therefore, although the gain K of the first-order system can be set at a high value in the absence of dead time, it cannot be set too high if dead time is present. (For the system considered here, the value of gain K must be considerably less than 2 for a satisfactory operation.)

Approximation of transport lag or dead time. If the dead time T is very small, then e^{-Ts} is frequently approximated by

$$e^{-Ts} \doteqdot 1 - Ts$$

or

$$e^{-Ts} \doteqdot \frac{1}{Ts + 1}$$

Such approximations are good if the dead time is very small and, in addition, the input time function $f(t)$ to the dead-time element is smooth and continuous. [This means that the second- and higher-order derivatives of $f(t)$ are small.]

A more elaborate expression to approximate e^{-Ts} is available and is

$$e^{-Ts} = \frac{1 - \dfrac{Ts}{2} + \dfrac{(Ts)^2}{8} - \dfrac{(Ts)^3}{48} + \cdots}{1 + \dfrac{Ts}{2} + \dfrac{(Ts)^2}{8} + \dfrac{(Ts)^3}{48} + \cdots}$$

If only the first two terms in the numerator and denominator are taken, then

$$e^{-Ts} \doteqdot \frac{1 - \dfrac{Ts}{2}}{1 + \dfrac{Ts}{2}} = \frac{2 - Ts}{2 + Ts}$$

This approximation is also used frequently.

6–8 ROOT-CONTOUR PLOTS

Effects of parameter variations on closed-loop poles. In many design problems, the effects on the closed-loop poles of the variations of parameters other than the gain K need to be investigated. Such effects can be easily investigated by the root-locus method. When two (or more) parameters are varied, the corresponding root loci are called *root contours*.

We shall use an example to illustrate the construction of the root contours when two parameters are varied, respectively, from zero to infinity.

Consider a servo system having tachometer feedback as shown in Figure 6–39(a). By eliminating the minor loop, the block diagram can be simplified [Figure 6–39(b)]. By defining

$$a = b + KK_h$$

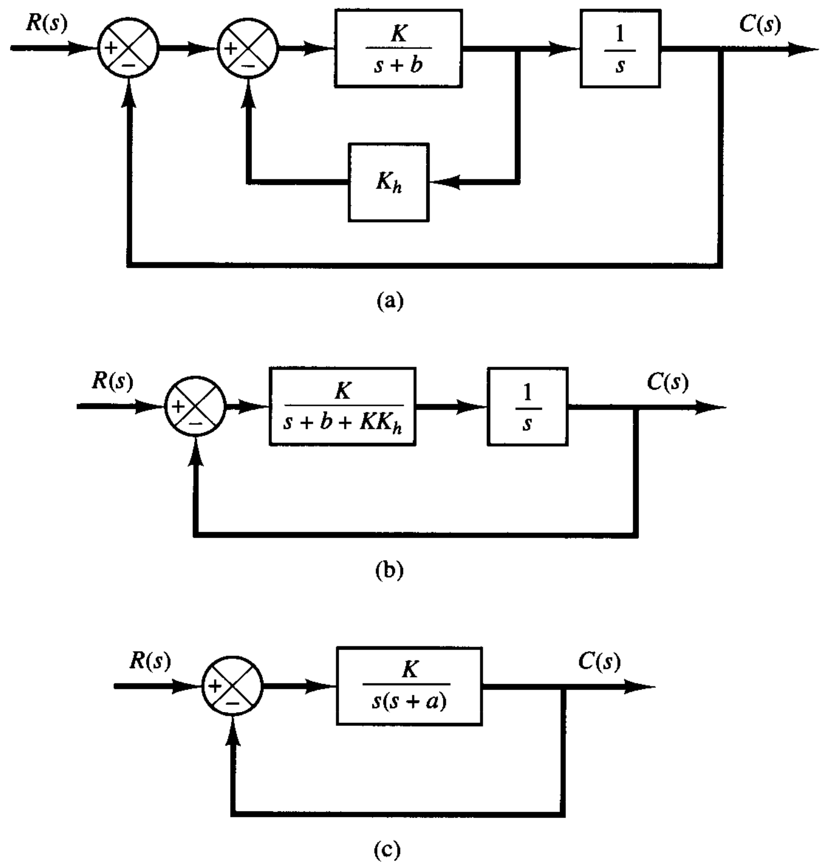


Figure 6–39
(a) Servo system with tachometer feedback; (b), (c) simplified block diagrams ($a = b + KK_h$).

this block diagram can be modified to that shown in Figure 6–39(c). This system involves two variables, parameter a and gain K.

In what follows we shall investigate the effect of varying the parameter a as well as the gain K. The closed-loop transfer function of this system becomes

$$\frac{C(s)}{R(s)} = \frac{K}{s^2 + as + K}$$

The characteristic equation is

$$s^2 + as + K = 0 \tag{6–27}$$

which may be rewritten

$$1 + \frac{as}{s^2 + K} = 0$$

or

$$\frac{as}{s^2 + K} = -1 \tag{6–28}$$

In Equation (6–28), the parameter a is written as a multiplying factor. For a given value of K, the effect of a on the closed-loop poles can be investigated from Equation (6–28). The root contours for this system can be constructed by following the usual procedure for constructing root loci.

We shall now construct the root contours as K and a vary, respectively, from zero to infinity. The root contours start from the poles (at $s = \pm j\sqrt{K}$) and terminate at the zeros (at $s = 0$ and infinity).

We shall first construct the locus of roots when $a = 0$. This can be done easily as follows: Substitute $a = 0$ into Equation (6–27). Then

$$s^2 + K = 0$$

or

$$\frac{K}{s^2} = -1 \tag{6–29}$$

The open-loop poles are thus a double pole at the origin. The root-locus plot of Equation (6–29) is shown in Figure 6–40(a).

To construct the root contours, let us assume that K is a constant; for example, $K = 4$. Then Equation (6–28) becomes

$$\frac{as}{s^2 + 4} = -1 \tag{6–30}$$

The open-loop poles are $s = \pm j2$. The finite open-loop zero is at the origin. The root-locus plot corresponding to Equation (6–30) is shown in Figure 6–40(b). For different values of K, Equation (6–30) yields similar root loci.

The root contour, the diagram showing the root loci corresponding to $0 \le K \le \infty$, $0 \le a \le \infty$, can be plotted as in Figure 6–40(c). Clearly, the root contours start at the

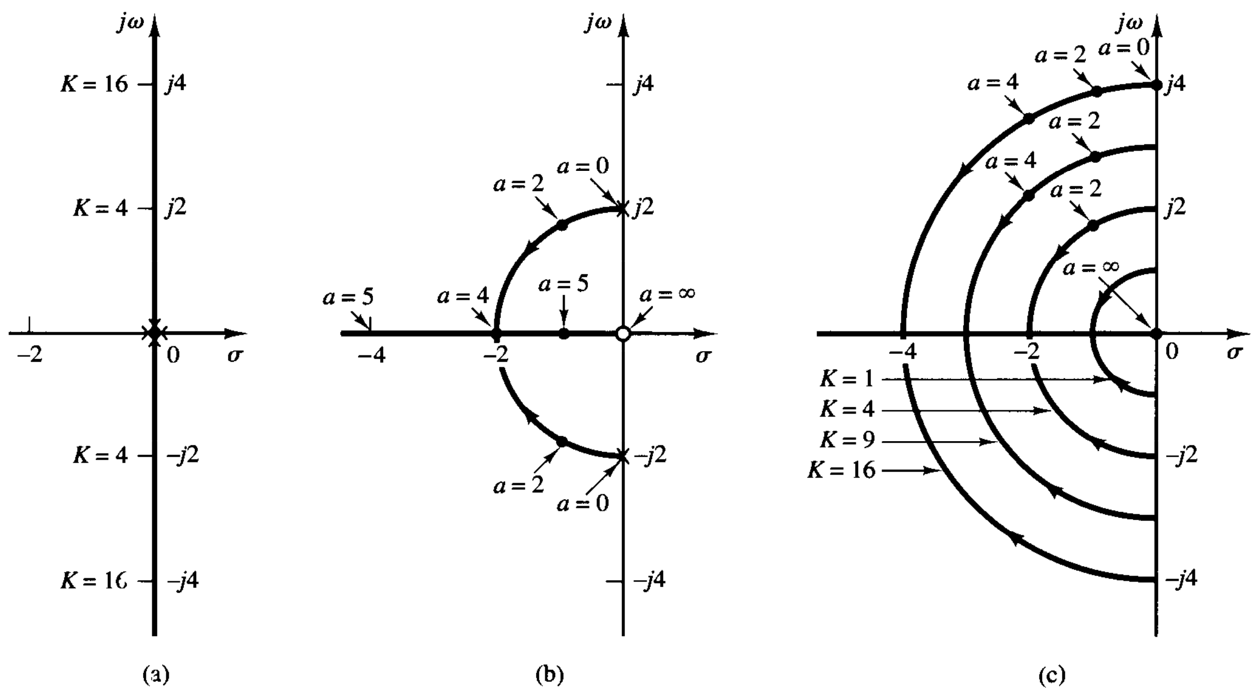


Figure 6–40
(a) Root-locus plot for the system shown in Figure 6–39(c) ($a = 0, 0 \le K \le \infty$); (b) root locus plot ($0 \le a \le \infty, K = 4$); (c) root-contour plot.

poles of and end at the zeros of the transfer function $as/(s^2 + K)$. The arrowheads on the root contours indicate the direction of increase in the value of a.

The root contours show the effects of the variations of system parameters on the closed-loop poles. From the root-contour plot shown in Figure 6–40(c), we see that, for $0 < K < \infty$, $0 < a < \infty$, the closed-loop poles lie in the left-half s plane and the system is stable.

Note that if the value of K is fixed, say $K = 4$, then the root contours become simply the root loci, as shown in Figure 6–40(b).

We have illustrated a method for constructing root contours when the gain K and parameter a are varied, respectively, from zero to infinity. Basically, we assign one parameter a constant value at a time and vary the other parameter from 0 to ∞ and sketch the root loci. Then we change the value of the first parameter and repeat sketching the root loci. By repeating this process we can sketch the root contour.

A MATLAB program to generate the root-contour plot is given in MATLAB Program 6–9. The resulting plot is shown in Figure 6–41.

MATLAB Program 6–9

```
% ---------- Root-contour plot ----------

% ***** Plot root contour of the system shown in Figure 6–39(c),
% where a and K are variables *****

% ***** In Equation (6–28), as/(s^2 + K) = –1, assume K = 1, 4, 9,
% 16, . . . and plot root loci as a varies from zero to infinity *****

% ***** Enter the numerator and denominators *****

num = [0   1   0];
den1 = [1   0   1];
den2 = [1   0   4];
den3 = [1   0   9];
den4 = [1   0   16];

% ***** Enter rlocus(num,den) command *****

rlocus(num,den1)
hold
Current plot held
rlocus(num,den2)
rlocus(num,den3)
rlocus(num,den4)
v = [–5   2   –5   5]; axis(v); axis('square');
grid
title('Root-Contour Plot')

% ***** Remove hold on graphics *****

hold
Current plot released
```

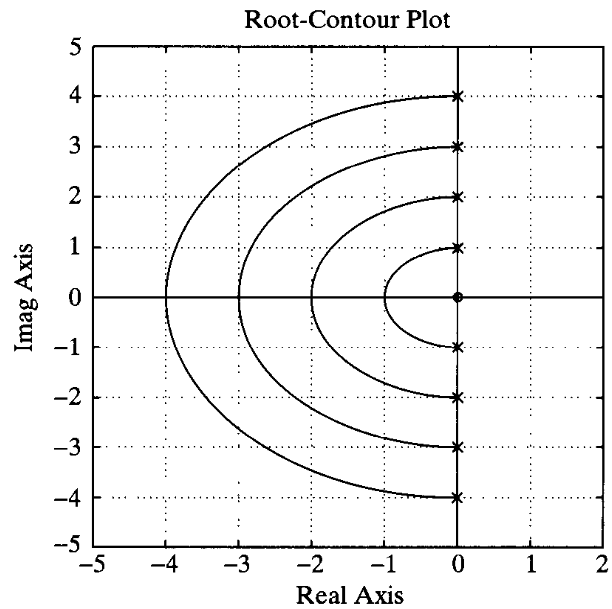


Figure 6–41
Root-contour plot generated with MATLAB.

EXAMPLE PROBLEMS AND SOLUTIONS

A-6-1. Sketch the root loci for the system shown in Figure 6–42(a). (The gain K is assumed to be positive.) Observe that for small or large values of K the system is overdamped and for medium values of K it is underdamped.

Solution. The procedure for plotting the root loci is as follows:

1. Locate the open-loop poles and zeros on the complex plane. Root loci exist on the negative real axis between 0 and -1 and between -2 and -3.

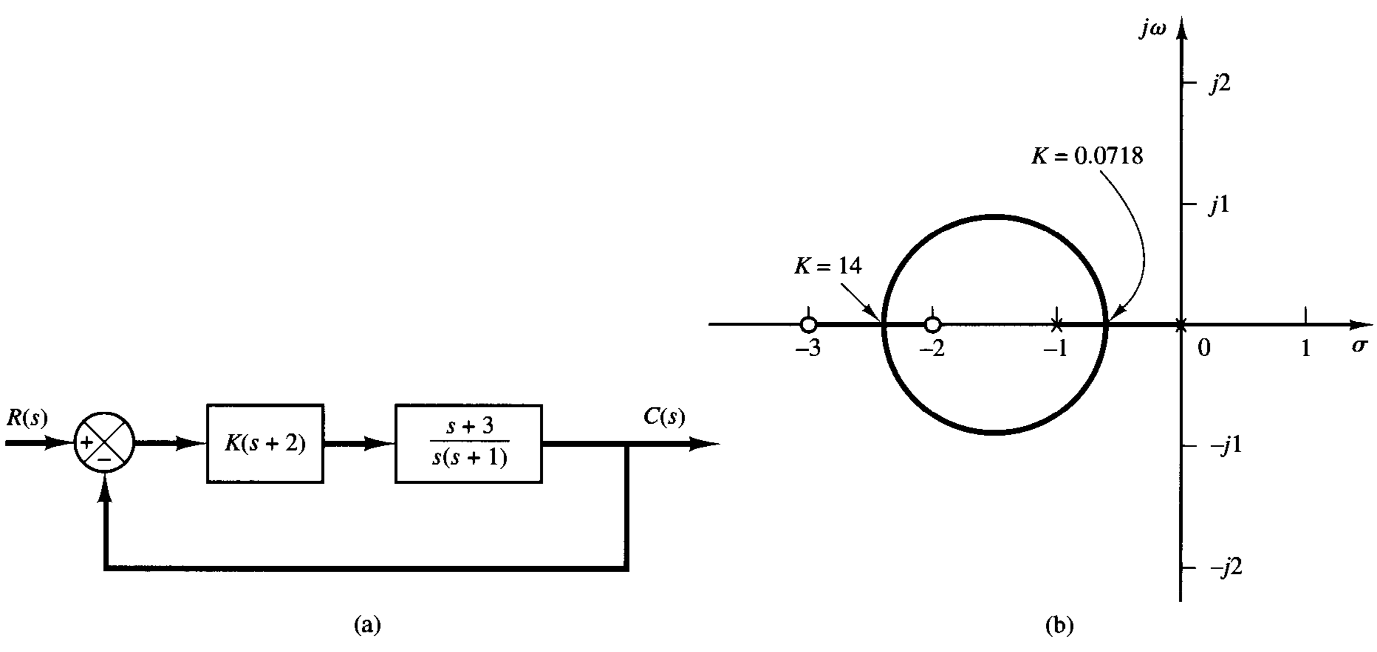


Figure 6–42
(a) Control system; (b) root-locus plot.

2. The number of open-loop poles and that of finite zeros are the same. This means that there are no asymptotes in the complex region of the s plane.

3. Determine the breakaway and break-in points. The characteristic equation for the system is

$$1 + \frac{K(s + 2)(s + 3)}{s(s + 1)} = 0$$

or

$$K = -\frac{s(s + 1)}{(s + 2)(s + 3)}$$

The breakaway and break-in points are determined from

$$\begin{aligned}\frac{dK}{ds} &= -\frac{(2s + 1)(s + 2)(s + 3) - s(s + 1)(2s + 5)}{[(s + 2)(s + 3)]^2} \\ &= -\frac{4(s + 0.634)(s + 2.366)}{[(s + 2)(s + 3)]^2} \\ &= 0\end{aligned}$$

as follows:

$$s = -0.634, \qquad s = -2.366$$

Notice that both points are on root loci. Therefore, they are actual breakaway or break-in points. At point $s = -0.634$, the value of K is

$$K = -\frac{(-0.634)(0.366)}{(1.366)(2.366)} = 0.0718$$

Similarly, at $s = -2.366$,

$$K = -\frac{(-2.366)(-1.366)}{(-0.366)(0.634)} = 14$$

(Because point $s = -0.634$ lies between two poles, it is a breakaway point and because point $s = -2.366$ lies between two zeros, it is a break-in point.)

4. Determine a sufficient number of points that satisfy the angle condition. (It can be found that the root locus is a circle with center at -1.5 that passes through the breakaway and break-in points.) The root-locus plot for this system is shown in Figure 6–42(b).

Note that this system is stable for any positive value of K since all the root loci lie in the left-half s plane.

Small values of K $(0 < K < 0.0718)$ correspond to an overdamped system. Medium values of K $(0.0718 < K < 14)$ correspond to an underdamped system. Finally, large values of K $(14 < K)$ correspond to an overdamped system. With a large value of K, the steady state can be reached in much shorter time than with a small value of K.

The value of K should be adjusted so that system performance is optimum according to a given performance index.

A-6-2. A simplified form of the open-loop transfer function of an airplane with an autopilot in the longitudinal mode is

$$G(s)H(s) = \frac{K(s + a)}{s(s - b)(s^2 + 2\zeta\omega_n s + \omega_n^2)}, \qquad a > 0, \qquad b > 0$$

Such a system involving an open-loop pole in the right-half s plane may be conditionally stable. Sketch the root loci when $a = b = 1$, $\zeta = 0.5$, and $\omega_n = 4$. Find the range of gain K for stability.

Solution. The open-loop transfer function for the system is

$$G(s)H(s) = \frac{K(s + 1)}{s(s - 1)(s^2 + 4s + 16)}$$

To sketch the root loci, we follow this procedure:

1. Locate the open-loop poles and zero in the complex plane. Root loci exist on the real axis between 1 and 0 and between -1 and $-\infty$.

2. Determine the asymptotes of the root loci. There are three asymptotes whose angles can be determined as

$$\text{Angles of asymptotes} = \frac{180°(2k + 1)}{4 - 1} = 60°, -60°, 180°$$

Referring to Equation (6–15), the abscissa of the intersection of the asymptotes and the real axis is

$$\sigma_a = -\frac{(0 - 1 + 2 + j2\sqrt{3} + 2 - j2\sqrt{3}) - 1}{4 - 1} = -\frac{2}{3}$$

3. Determine the breakaway and break-in points. Since the characteristic equation is

$$1 + \frac{K(s + 1)}{s(s - 1)(s^2 + 4s + 16)} = 0$$

we obtain

$$K = -\frac{s(s - 1)(s^2 + 4s + 16)}{s + 1}$$

By differentiating K with respect to s, we get

$$\frac{dK}{ds} = -\frac{3s^4 + 10s^3 + 21s^2 + 24s - 16}{(s + 1)^2}$$

The numerator can be factored as follows:

$$3s^4 + 10s^3 + 21s^2 + 24s - 16$$
$$= 3(s + 0.76 + j2.16)(s + 0.76 - j2.16)(s + 2.26)(s - 0.45)$$

Points $s = 0.45$ and $s = -2.26$ are on root loci on the real axis. Hence, these points are actual breakaway and break-in points, respectively. Points $s = -0.76 \pm j2.16$ do not satisfy the angle condition. Hence, they are neither breakaway nor break-in points.

4. Using Routh's stability criterion, determine the value of K at which the root loci cross the imaginary axis. Since the characteristic equation is

$$s^4 + 3s^3 + 12s^2 + (K - 16)s + K = 0$$

the Routh array becomes

$$
\begin{array}{llll}
s^4 & 1 & 12 & K \\
s^3 & 3 & K-16 & 0 \\
s^2 & \dfrac{52-K}{3} & K & 0 \\
s^1 & \dfrac{-K^2+59K-832}{52-K} & 0 & \\
s^0 & K & &
\end{array}
$$

The values of K that make the s^1 term in the first column equal zero are $K = 35.7$ and $K = 23.3$.

The crossing points on the imaginary axis can be found by solving the auxiliary equation obtained from the s^2 row, that is, by solving the following equation for s:

$$\frac{52-K}{3}s^2 + K = 0$$

The results are

$$s = \pm j2.56, \qquad \text{for } K = 35.7$$

$$s = \pm j1.56, \qquad \text{for } K = 23.3$$

The crossing points on the imaginary axis are thus $s = \pm j2.56$ and $s = \pm j1.56$.

5. Find the angles of departure of the root loci from the complex poles. For the open-loop pole at $s = -2 + j2\sqrt{3}$, the angle of departure θ is

$$\theta = 180° - 120° - 130.5° - 90° + 106°$$

or

$$\theta = -54.5°$$

(The angle of departure from the open-loop pole at $s = -2 - j2\sqrt{3}$ is 54.5°.)

6. Choose a test point in the broad neighborhood of the $j\omega$ axis and the origin and apply the angle condition. If the test point does not satisfy the angle condition, select another test point until it does. Continue the same process and locate a sufficient number of points that satisfy the angle condition.

Figure 6–43 shows the root loci for this system. From step 4 the system is stable for $23.3 < K < 35.7$. Otherwise, it is unstable.

A-6-3. Sketch the root loci of the control system shown in Figure 6–44(a).

Solution. The open-loop poles are located at $s = 0$, $s = -3 + j4$, and $s = -3 - j4$. A root locus branch exists on the real axis between the origin and $-\infty$. There are three asymptotes for the root loci. The angles of asymptotes are

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k+1)}{3} = 60°, -60°, 180°$$

Referring to Equation (6–15), the intersection of the asymptotes and the real axis is obtained as

$$\sigma_a = -\frac{0+3+3}{3} = -2$$

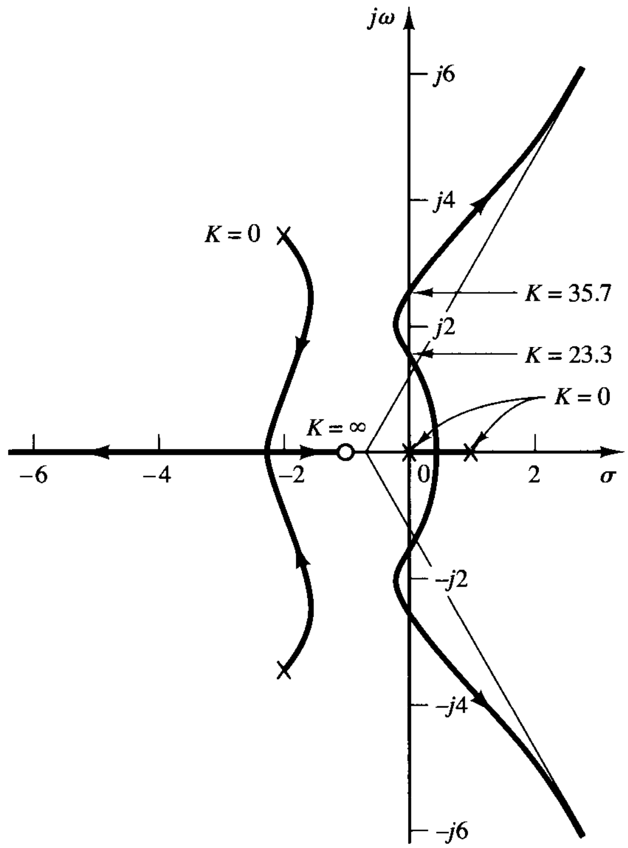


Figure 6–43
Root-locus plot.

Next we check the breakaway and break-in points. For this system we have

$$K = -s(s^2 + 6s + 25)$$

Now we set

$$\frac{dK}{ds} = -(3s^2 + 12s + 25) = 0$$

which yields

$$s = -2 + j2.0817, \qquad s = -2 - j2.0817$$

Notice that at points $s = -2 \pm j2.0817$ the angle condition is not satisfied. Hence, they are neither breakaway nor break-in points. In fact, if we calculate the value of K, we obtain

$$K = -s(s^2 + 6s + 25)\Big|_{s=-2\pm j2.0817} = 34 \pm j18.04$$

(To be an actual breakaway or break-in point, the corresponding value of K must be real and positive.)

The angle of departure from the complex pole in the upper half s plane is

$$\theta = 180^\circ - 126.87^\circ - 90^\circ$$

or

$$\theta = -36.87^\circ$$

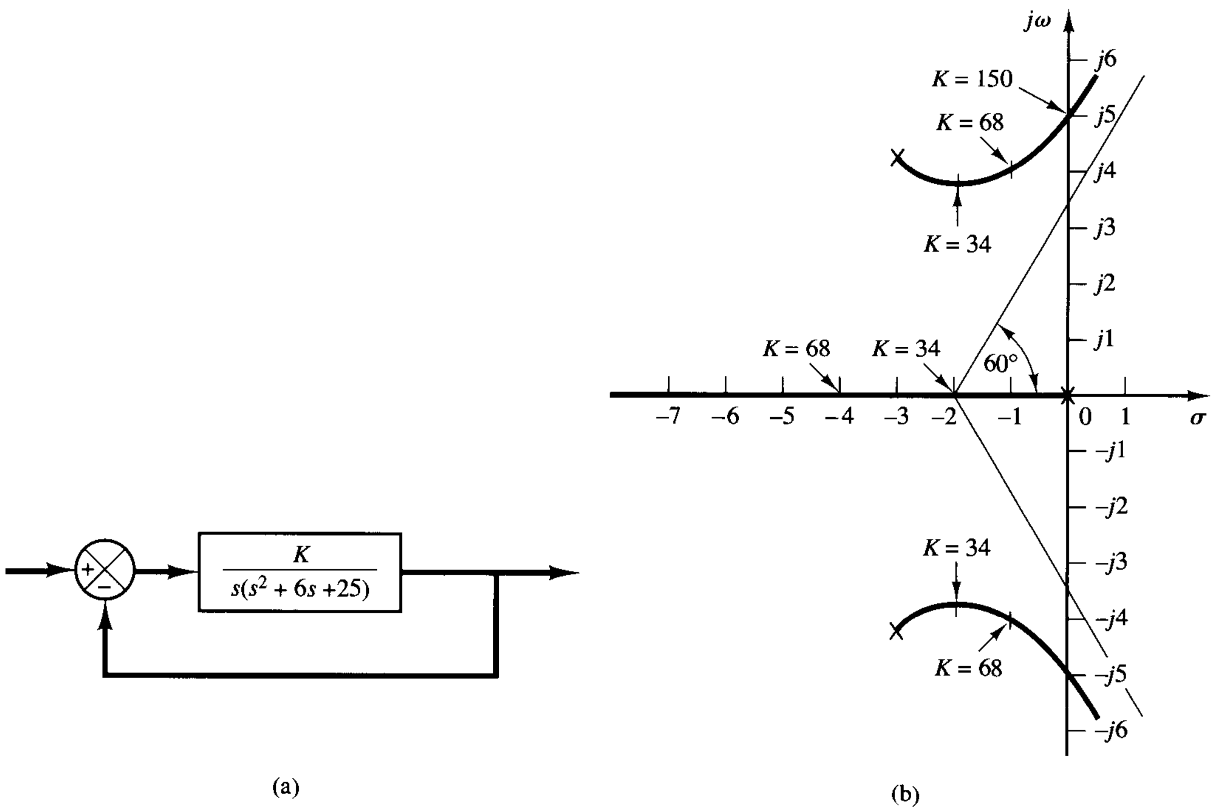


Figure 6–44
(a) Control system; (b) root-locus plot.

The points where root-locus branches cross the imaginary axis may be found by substituting $s = j\omega$ into the characteristic equation and solving the equation for ω and K as follows: Noting that the characteristic equation is

$$s^3 + 6s^2 + 25s + K = 0$$

we have

$$(j\omega)^3 + 6(j\omega)^2 + 25(j\omega) + K = (-6\omega^2 + K) + j\omega(25 - \omega^2) = 0$$

which yields

$$\omega = \pm 5, \qquad K = 150 \qquad \text{or} \qquad \omega = 0, \qquad K = 0$$

Root-locus branches cross the imaginary axis at $\omega = 5$ and $\omega = -5$. The value of gain K at the crossing points is 150. Also, the root-locus branch on the real axis touches the imaginary axis at $\omega = 0$. Figure 6–44(b) shows a root-locus plot for the system.

It is noted that if the order of the numerator of $G(s)H(s)$ is lower than that of the denominator by two or more, and if some of the closed-loop poles move on the root locus toward the right as gain K is increased, then other closed-loop poles must move toward the left as gain K is increased. This fact can be seen clearly in this problem. If the gain K is increased from $K = 34$ to $K = 68$, the complex-conjugate closed-loop poles are moved from $s = -2 + j3.65$ to $s = -1 + j4$; the third pole is moved from $s = -2$ (which corresponds to $K = 34$) to $s = -4$ (which corresponds to $K = 68$). Thus, the movements of two complex-conjugate closed-loop poles to the right by one unit cause the remaining closed-loop pole (real pole in this case) to move to the left by two units.

A-6-4. Consider the system shown in Figure 6–45(a). Sketch the root loci for the system. Observe that for small or large values of K the system is underdamped and for medium values of K it is overdamped.

Solution. A root locus exists on the real axis between the origin and $-\infty$. The angles of asymptotes of the root-locus branches are obtained as

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k+1)}{3} = 60°, -60°, -180°$$

The intersection of the asymptotes and the real axis is located on the real axis at

$$\sigma_a = -\frac{0+2+2}{3} = -1.3333$$

The breakaway and break-in points are found from $dK/ds = 0$. Since the characteristic equation is

$$s^3 + 4s^2 + 5s + K = 0$$

we have

$$K = -(s^3 + 4s^2 + 5s)$$

Now we set

$$\frac{dK}{ds} = -(3s^2 + 8s + 5) = 0$$

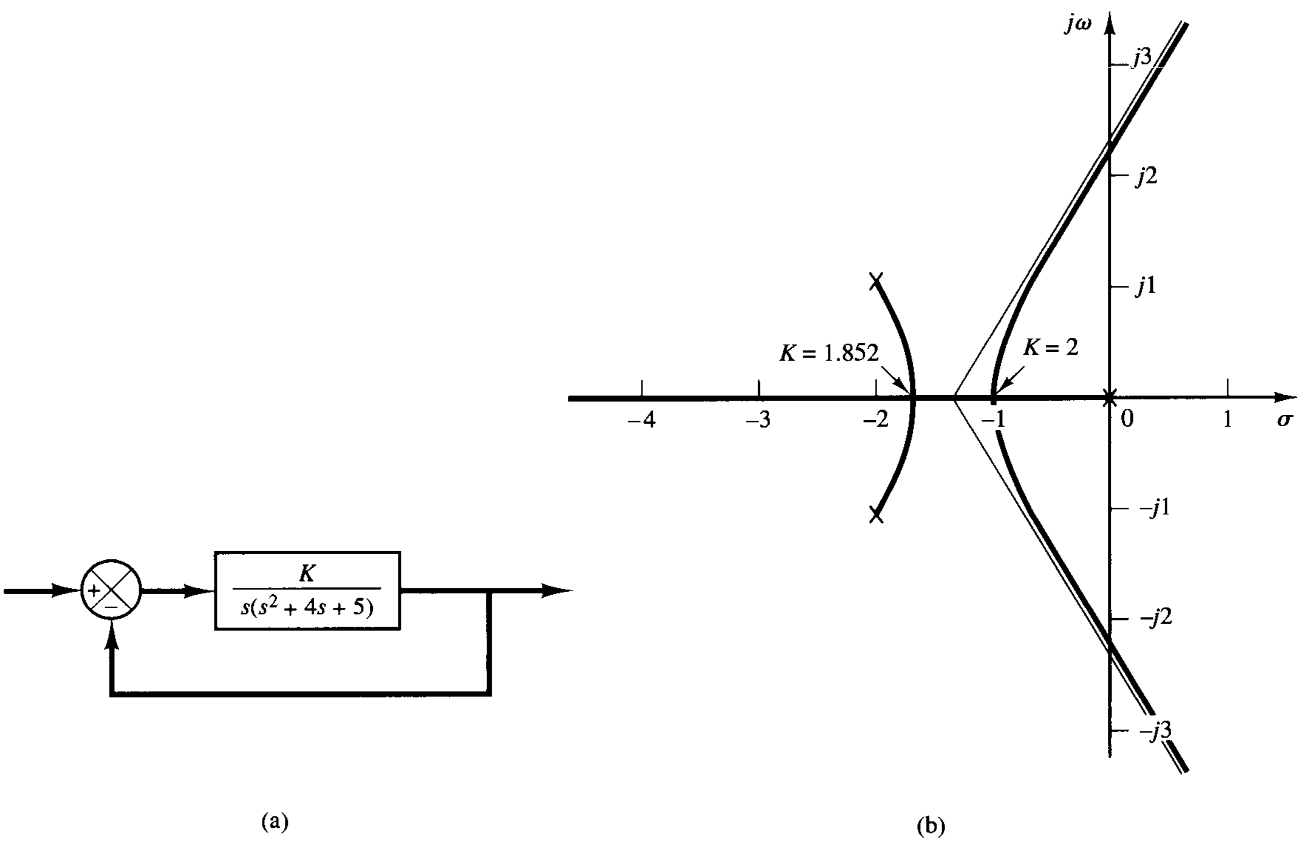


Figure 6–45
(a) Control system; (b) root-locus plot.

which yields

$$s = -1, \qquad s = -1.6667$$

Since these points are on root loci, they are actual breakaway or break-in points. (At point $s = -1$, the value of K is 2, and at point $s = -1.6667$, the value of K is 1.852.)

The angle of departure from a complex pole in the upper half s plane is obtained from

$$\theta = 180° - 153.43° - 90°$$

or

$$\theta = -63.43°$$

The root-locus branch from the complex pole in the upper half s plane breaks into the real axis at $s = -1.6667$.

Next we determine the points where root-locus branches cross the imaginary axis. By substituting $s = j\omega$ into the characteristic equation, we have

$$(j\omega)^3 + 4(j\omega)^2 + 5(j\omega) + K = 0$$

or

$$(K - 4\omega^2) + j\omega(5 - \omega^2) = 0$$

from which we obtain

$$\omega = \pm\sqrt{5}, \qquad K = 20 \quad \text{or} \quad \omega = 0, \qquad K = 0$$

Root-locus branches cross the imaginary axis at $\omega = \sqrt{5}$ and $\omega = -\sqrt{5}$. The root-locus branch on the real axis touches the $j\omega$ axis at $\omega = 0$. A sketch of the root loci for the system is shown in Figure 6–45(b).

Note that since this system is of third order there are three closed-loop poles. The nature of the system response to a given input depends on the locations of the closed-loop poles.

For $0 < K < 1.852$, there are a set of complex-conjugate closed-loop poles and a real closed-loop pole. For $1.852 \le K \le 2$, there are three real closed-loop poles. For example, the closed-loop poles are located at

$$s = -1.667, \qquad s = -1.667, \qquad s = -0.667, \qquad \text{for } K = 1.852$$

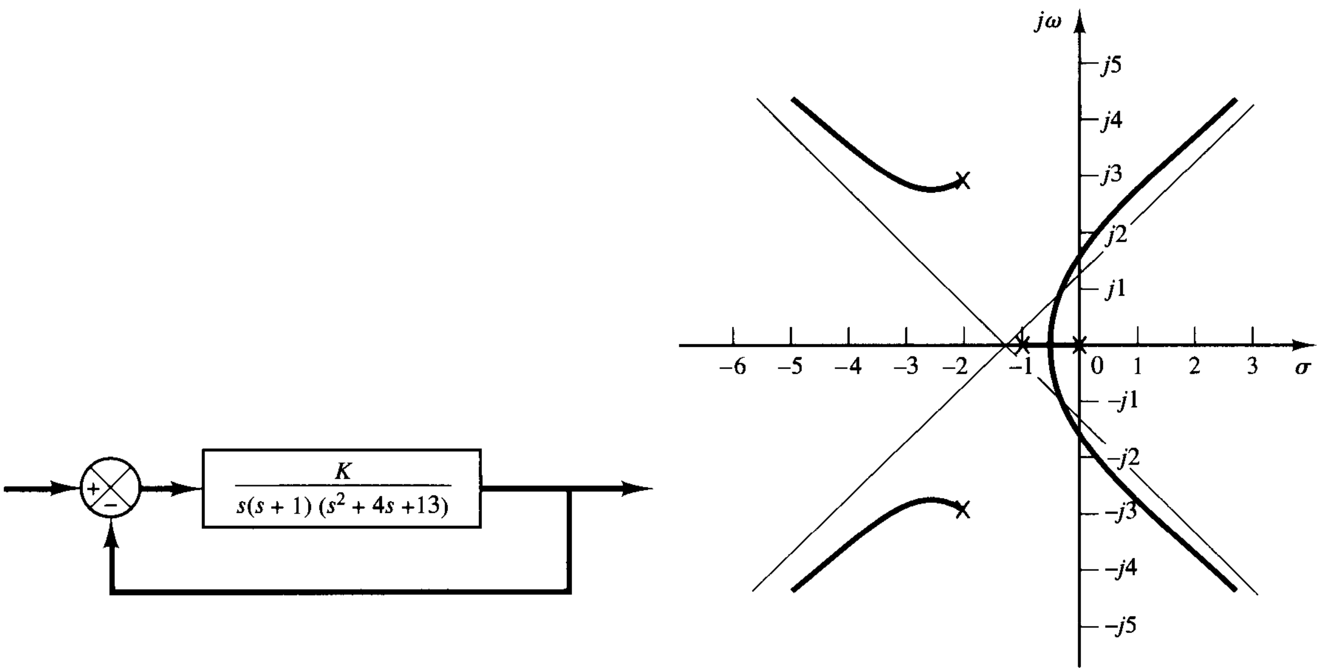
$$s = -1, \qquad s = -1, \qquad s = -2, \qquad \text{for } K = 2$$

For $2 < K$, there are a set of complex-conjugate closed-loop poles and a real closed-loop pole. Thus, small values of K $(0 < K < 1.852)$ correspond to an underdamped system. (Since the real closed-loop pole dominates, only a small ripple may show up in the transient response.) Medium values of K $(1.852 \le K \le 2)$ correspond to an overdamped system. Large values of K $(2 < K)$ correspond to an underdamped system. With a large value of K, the system responds much faster than with a smaller value of K.

A-6-5. Sketch the root loci for the system shown in Figure 6–46(a).

Solution. The open-loop poles are located at $s = 0$, $s = -1$, $s = -2 + j3$, and $s = -2 - j3$. A root locus exists on the real axis between points $s = 0$ and $s = -1$. The asymptotes are found as follows:

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k + 1)}{4} = 45°, -45°, 135°, -135°$$



(a) (b)

Figure 6–46
(a) Control system; (b) root-locus plot.

The intersection of the asymptotes and the real axis is found from

$$\sigma_a = -\frac{0 + 1 + 2 + 2}{4} = -1.25$$

The breakaway and break-in points are found from $dK/ds = 0$. Noting that

$$K = -s(s + 1)(s^2 + 4s + 13) = -(s^4 + 5s^3 + 17s^2 + 13s)$$

we have

$$\frac{dK}{ds} = -(4s^3 + 15s^2 + 34s + 13) = 0$$

from which we get

$$s = -0.467, \qquad s = -1.642 + j2.067, \qquad s = -1.642 - j2.067$$

The point $s = -0.467$ is on a root locus. Therefore, it is an actual breakaway point. The gain values K corresponding to points $s = -1.642 \pm j2.067$ are complex quantities. Since the gain values are not real positive, these points are neither breakaway nor break-in points.

The angle of departure from the complex pole in the upper half s plane is

$$\theta = 180° - 123.69° - 108.44° - 90°$$

or

$$\theta = -142.13°$$

Next we shall find the points where root loci may cross the $j\omega$ axis. Since the characteristic equation is

$$s^4 + 5s^3 + 17s^2 + 13s + K = 0$$

by substituting $s = j\omega$ into it we obtain

$$(j\omega)^4 + 5(j\omega)^3 + 17(j\omega)^2 + 13(j\omega) + K = 0$$

or

$$(K + \omega^4 - 17\omega^2) + j\omega(13 - 5\omega^2) = 0$$

from which we obtain

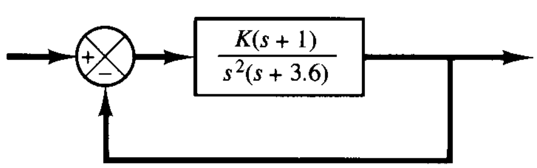
$$\omega = \pm 1.6125, \quad K = 37.44 \quad \text{or} \quad \omega = 0, \quad K = 0$$

The root-locus branches that extend to the right-half s plane cross the imaginary axis at $\omega = \pm 1.6125$. Also, the root-locus branch on the real axis touches the imaginary axis at $\omega = 0$. Figure 6–46(b) shows a sketch of the root loci for the system. Notice that each root-locus branch that extends to the right half s plane crosses its own asymptote.

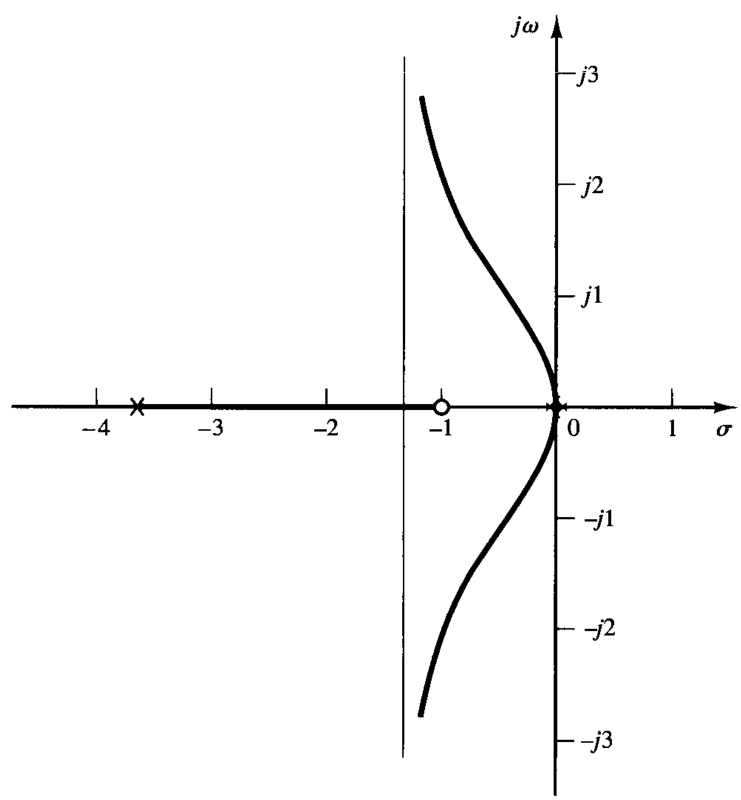
A-6-6. Sketch the root loci for the system shown in Figure 6–47(a).

Solution. A root locus exists on the real axis between points $s = -1$ and $s = -3.6$. The asymptotes can be determined as follows:

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k + 1)}{3 - 1} = 90°, -90°$$



(a)



(b)

Figure 6–47
(a) Control system; (b) root-locus plot.

The intersection of the asymptotes and the real axis is found from

$$\sigma_a = -\frac{0 + 0 + 3.6 - 1}{3 - 1} = -1.3$$

Since the characteristic equation is

$$s^3 + 3.6s^2 + K(s + 1) = 0$$

we have

$$K = -\frac{s^3 + 3.6s^2}{s + 1}$$

The breakaway and break-in points are found from

$$\frac{dK}{ds} = -\frac{(3s^2 + 7.2s)(s + 1) - (s^3 + 3.6s^2)}{(s + 1)^2} = 0$$

or

$$s^3 + 3.3s^2 + 3.6s = 0$$

from which we get

$$s = 0, \qquad s = -1.65 + j0.9367, \qquad s = -1.65 - j0.9367$$

Point $s = 0$ corresponds to the actual breakaway point. But points $s = -1.65 \pm j0.9367$ are neither breakaway nor break-in points, because the corresponding gain values K become complex quantities.

To check the points where root-locus branches may cross the imaginary axis, substitute $s = j\omega$ into the characteristic equation.

$$(j\omega)^3 + 3.6(j\omega)^2 + Kj\omega + K = 0$$

or

$$(K - 3.6\omega^2) + j\omega(K - \omega^2) = 0$$

Notice that this equation can be satisfied only if $\omega = 0$, $K = 0$. Because of the presence of a double pole at the origin, the root locus is tangent to the $j\omega$ axis at $\omega = 0$. The root-locus branches do not cross the $j\omega$ axis. Figure 6–47(b) is a sketch of the root loci for this system.

A-6-7. Sketch the root loci for the system shown in Figure 6–48(a).

Solution. A root locus exists on the real axis between point $s = -0.4$ and $s = -3.6$. The asymptotes can be found as follows:

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k + 1)}{3 - 1} = 90°, -90°$$

The intersection of the asymptotes and the real axis is obtained from

$$\sigma_a = -\frac{0 + 0 + 3.6 - 0.4}{3 - 1} = -1.6$$

Next we shall find the breakaway points. Since the characteristic equation is

$$s^3 + 3.6s^2 + Ks + 0.4K = 0$$

we have

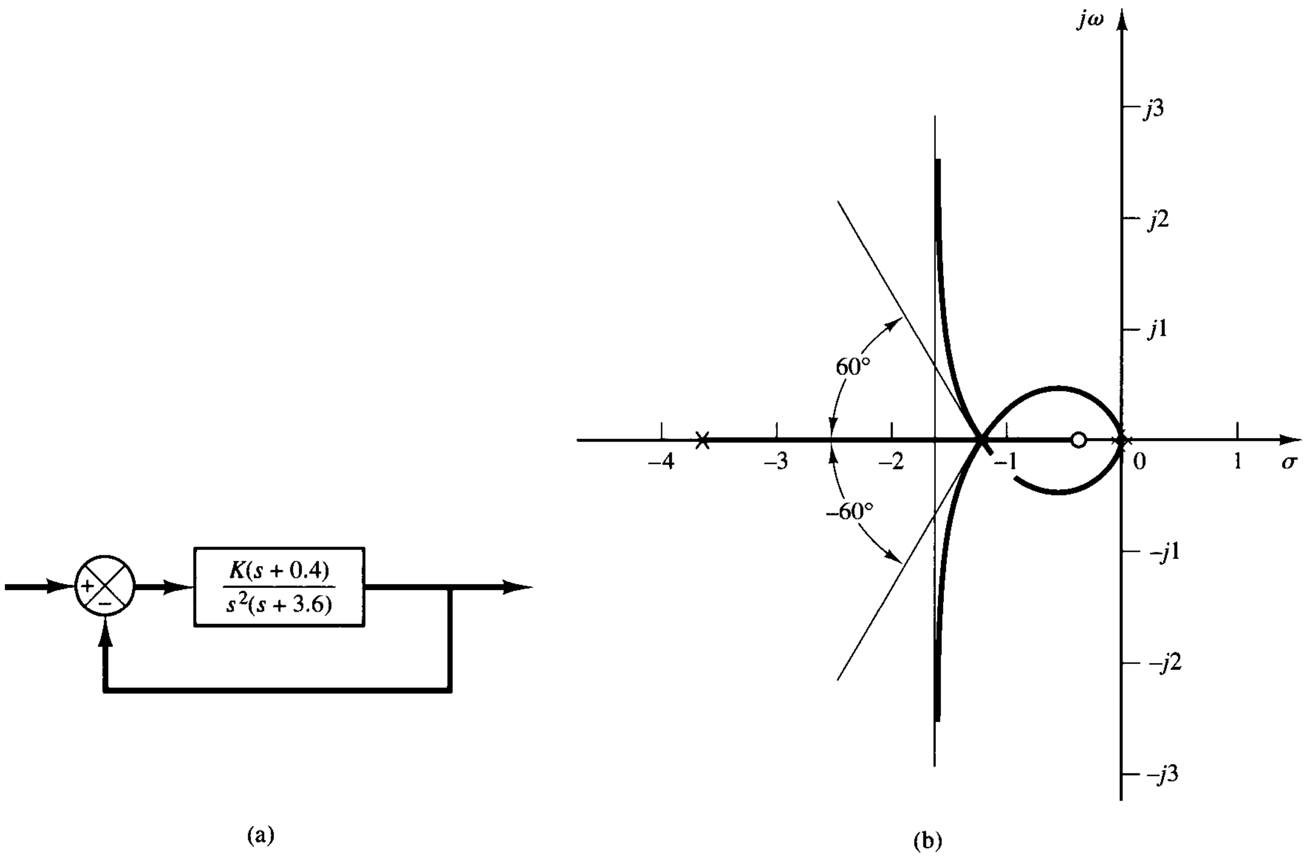


Figure 6–48
(a) Control system; (b) root-locus plot.

$$K = -\frac{s^3 + 3.6s^2}{s + 0.4}$$

The breakaway and break-in points are found from

$$\frac{dK}{ds} = -\frac{(3s^2 + 7.2s)(s + 0.4) - (s^3 + 3.6s^2)}{(s + 0.4)^2} = 0$$

from which we get

$$s^3 + 2.4s^2 + 1.44s = 0$$

or

$$s(s + 1.2)^2 = 0$$

Thus, the breakaway or break-in points are at $s = 0$ and $s = -1.2$. Note that $s = -1.2$ is a double root. When a double root occurs in $dK/ds = 0$ at point $s = -1.2$, $d^2K/(ds^2) = 0$ at this point. The value of gain K at point $s = -1.2$ is

$$K = -\frac{s^3 + 3.6s^2}{s + 0.4}\bigg|_{s=-1.2} = 4.32$$

This means that with $K = 4.32$ the characteristic equation has a triple root at point $s = -1.2$. This can be easily verified as follows:

$$s^3 + 3.6s^2 + 4.32s + 1.728 = (s + 1.2)^3 = 0$$

Hence, three root-locus branches meet at point $s = -1.2$. The angles of departures at point $s = -1.2$ of the root locus branches that approach the asymptotes are $\pm 180°/3$, that is, 60° and −60°. (See Problem A-6-8.)

Finally, we shall examine if root-locus branches cross the imaginary axis. By substituting $s = j\omega$ into the characteristic equation, we have

$$(j\omega)^3 + 3.6(j\omega)^2 + K(j\omega) + 0.4K = 0$$

or

$$(0.4K - 3.6\omega^2) + j\omega(K - \omega^2) = 0$$

This equation can be satisfied only if $\omega = 0, K = 0$. At point $\omega = 0$, the root locus is tangent to the $j\omega$ axis because of the presence of a double pole at the origin. There are no points that root-locus branches cross the imaginary axis.

A sketch of the root loci for this system is shown in Figure 6–48(b).

A-6-8. Referring to Problem A-6-7, obtain the equations for the root-locus branches for the system shown in Figure 6–48(a). Show that the root-locus branches cross the real axis at the breakaway point at angles ±60°.

Solution. The equations for the root-locus branches can be obtained from the angle condition

$$\angle \frac{K(s + 0.4)}{s^2(s + 3.6)} = \pm 180°(2k + 1)$$

which can be rewritten as

$$\angle s + 0.4 - 2\angle s - \angle s + 3.6 = \pm 180°(2k + 1)$$

By substituting $s = \sigma + j\omega$, we obtain

$$\angle \sigma + j\omega + 0.4 - 2\angle \sigma + j\omega - \angle \sigma + j\omega + 3.6 = \pm 180°(2k + 1)$$

or

$$\tan^{-1}\left(\frac{\omega}{\sigma + 0.4}\right) - 2\tan^{-1}\left(\frac{\omega}{\sigma}\right) - \tan^{-1}\left(\frac{\omega}{\sigma + 3.6}\right) = \pm 180°(2k + 1)$$

By rearranging, we have

$$\tan^{-1}\left(\frac{\omega}{\sigma + 0.4}\right) - \tan^{-1}\left(\frac{\omega}{\sigma}\right) = \tan^{-1}\left(\frac{\omega}{\sigma}\right) + \tan^{-1}\left(\frac{\omega}{\sigma + 3.6}\right) \pm 180°(2k + 1)$$

Taking tangents of both sides of this last equation, and noting that

$$\tan\left[\tan^{-1}\left(\frac{\omega}{\sigma + 3.6}\right) \pm 180°(2k + 1)\right] = \frac{\omega}{\sigma + 3.6}$$

we obtain

$$\frac{\dfrac{\omega}{\sigma + 0.4} - \dfrac{\omega}{\sigma}}{1 + \dfrac{\omega}{\sigma + 0.4}\dfrac{\omega}{\sigma}} = \frac{\dfrac{\omega}{\sigma} + \dfrac{\omega}{\sigma + 3.6}}{1 - \dfrac{\omega}{\sigma}\dfrac{\omega}{\sigma + 3.6}}$$

which can be simplified to

$$\frac{\omega\sigma - \omega(\sigma + 0.4)}{(\sigma + 0.4)\sigma + \omega^2} = \frac{\omega(\sigma + 3.6) + \omega\sigma}{\sigma(\sigma + 3.6) - \omega^2}$$

or

$$\omega(\sigma^3 + 2.4\sigma^2 + 1.44\sigma + 1.6\omega^2 + \sigma\omega^2) = 0$$

which can be further simplified to

$$\omega[\sigma(\sigma + 1.2)^2 + (\sigma + 1.6)\omega^2] = 0$$

For $\sigma \neq -1.6$, we may write this last equation as

$$\omega\left[\omega - (\sigma + 1.2)\sqrt{\frac{-\sigma}{\sigma + 1.6}}\right]\left[\omega + (\sigma + 1.2)\sqrt{\frac{-\sigma}{\sigma + 1.6}}\right] = 0$$

which gives the equations for the root-locus as follows:

$$\omega = 0$$

$$\omega = (\sigma + 1.2)\sqrt{\frac{-\sigma}{\sigma + 1.6}}$$

$$\omega = -(\sigma + 1.2)\sqrt{\frac{-\sigma}{\sigma + 1.6}}$$

The equation $\omega = 0$ represents the real axis. The root locus for $0 \leq K \leq \infty$ is between points $s = -0.4$ and $s = -3.6$. (The real axis other than this line segment and the origin $s = 0$ corresponds to the root locus for $-\infty \leq K < 0$.)

The equations

$$\omega = \pm(\sigma + 1.2)\sqrt{\frac{-\sigma}{\sigma + 1.6}} \tag{6-31}$$

represent the complex branches for $0 \leq K \leq \infty$. These two branches lie between $\sigma = -1.6$ and $\sigma = 0$. [See Figure 6–48(b).] The slopes of the complex root-locus branches at the breakaway point ($\sigma = -1.2$) can be found by evaluating $d\omega/d\sigma$ of Equation (6–31) at point $\sigma = -1.2$.

$$\left.\frac{d\omega}{d\sigma}\right|_{\sigma=-1.2} = \pm\left.\sqrt{\frac{-\sigma}{\sigma + 1.6}}\right|_{\sigma=-1.2} = \pm\sqrt{\frac{1.2}{0.4}} = \pm\sqrt{3}$$

Since $\tan^{-1}\sqrt{3} = 60°$, the root-locus branches intersect the real axis with angles $\pm 60°$.

A-6-9. Consider the system shown in Figure 6–49, which has an unstable feedforward transfer function. Sketch the root-locus plot and locate the closed-loop poles. Show that, although the closed-loop poles lie on the negative real axis and the system is not oscillatory, the unit-step response curve will exhibit overshoot.

Solution. The root-locus plot for this system is shown in Figure 6–50. The closed-loop poles are located at $s = -2$ and $s = -5$.

The closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{10(s + 1)}{s^2 + 7s + 10}$$

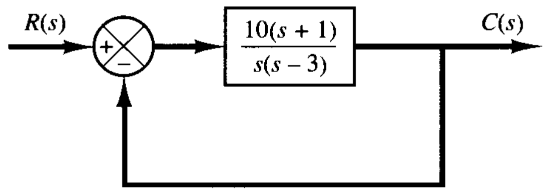


Figure 6–49
Control system.

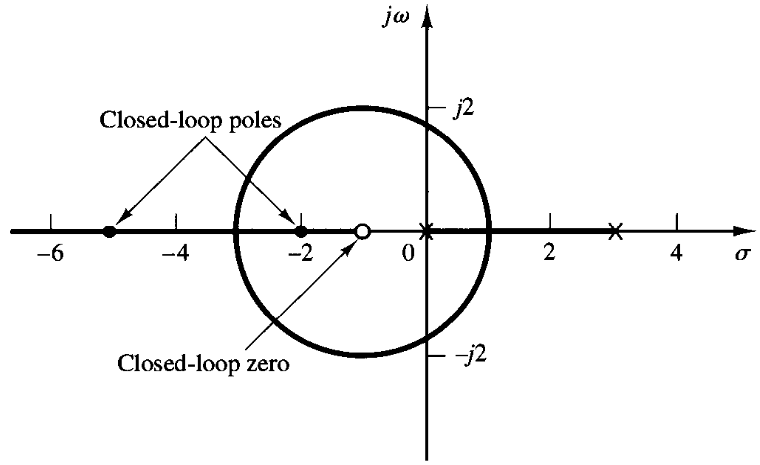


Figure 6–50
Root-locus plot for the system shown in Figure 6–49.

The unit-step response of this system is

$$C(s) = \frac{10(s+1)}{s(s+2)(s+5)}$$

The inverse Laplace transform of $C(s)$ gives

$$c(t) = 1 + 1.666e^{-2t} - 2.666e^{-5t}, \qquad \text{for } t \geq 0$$

The unit-step response curve is shown in Figure 6–51. Although the system is not oscillatory, the unit-step response curve exhibits overshoot. (This is due to the presence of a zero at $s = -1$.)

A-6-10. Sketch the root loci of the control system shown in Figure 6–52(a). Determine the range of gain K for stability.

Solution. Open-loop poles are located at $s = 1, s = -2 + j\sqrt{3}$, and $s = -2 - j\sqrt{3}$. A root locus exists on the real axis between points $s = 1$ and $s = -\infty$. The asymptotes of the root-locus branches are found as follows:

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k+1)}{3} = 60°, -60°, 180°$$

The intersection of the asymptotes and the real axis is obtained as

$$\sigma_a = -\frac{-1+2+2}{3} = -1$$

The breakaway and break-in points can be located from $dK/ds = 0$. Since

$$K = -(s-1)(s^2+4s+7) = -(s^3+3s^2+3s-7)$$

we have

$$\frac{dK}{ds} = -(3s^2+6s+3) = 0$$

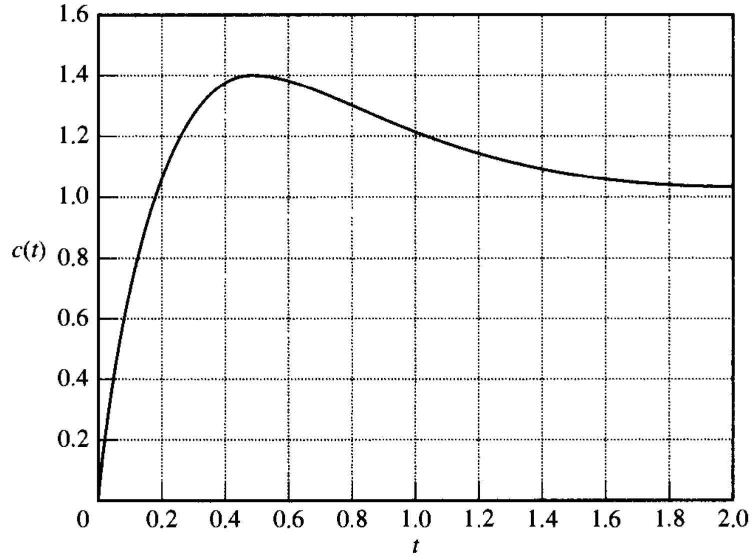
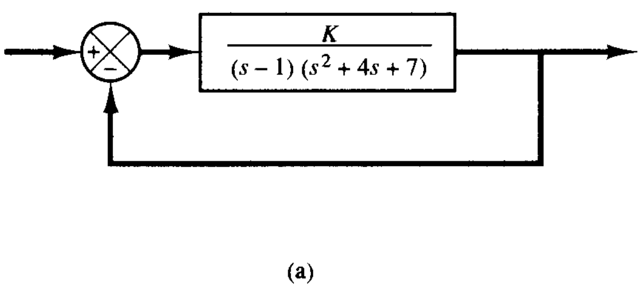


Figure 6–51
Unit-step response curve for the system shown in Figure 6–49.



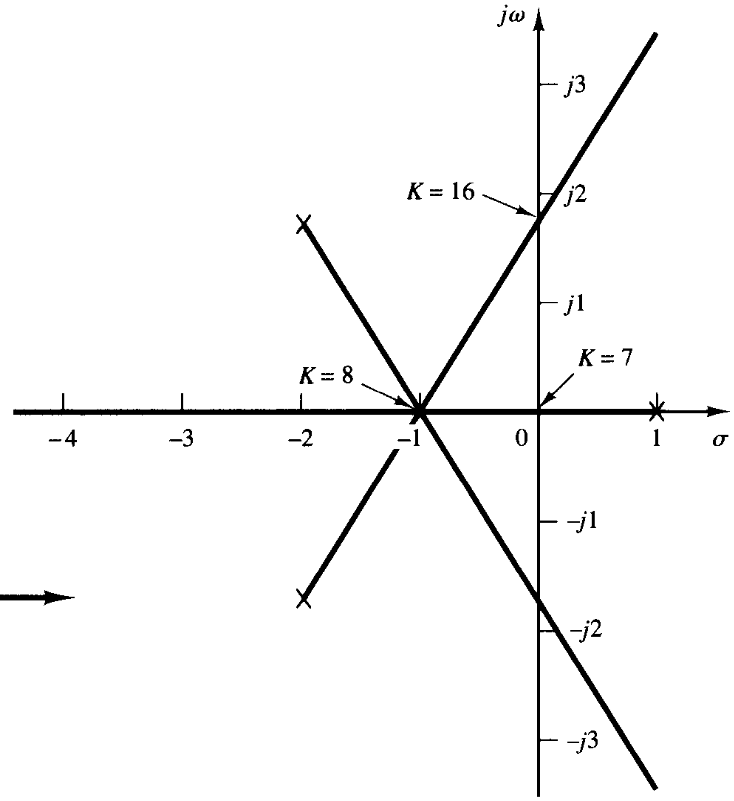


Figure 6–52
(a) Control system; (b) root-locus plot.

which yields

$$(s + 1)^2 = 0$$

Thus the equation $dK/ds = 0$ has a double root at $s = -1$. The breakaway point is located at $s = -1$. Three root locus branches meet at this breakaway point. The angles of departure of the branches at the breakaway point are $\pm 180°/3$, that is, $60°$ and $-60°$.

We shall next determine the points where root-locus branches may cross the imaginary axis. Noting that the characteristic equation is

$$(s - 1)(s^2 + 4s + 7) + K = 0$$

or

$$s^3 + 3s^2 + 3s - 7 + K = 0$$

we substitute $s = j\omega$ into it and obtain

$$(j\omega)^3 + 3(j\omega)^2 + 3(j\omega) - 7 + K = 0$$

By rewriting this last equation, we have

$$(K - 7 - 3\omega^2) + j\omega(3 - \omega^2) = 0$$

This equation is satisfied when

$$\omega = \pm\sqrt{3}, \qquad K = 7 + 3\omega^2 = 16 \qquad \text{or} \qquad \omega = 0, \qquad K = 7$$

The root-locus branches cross the imaginary axis at $\omega = \pm\sqrt{3}$ (where $K = 16$) and $\omega = 0$ (where $K = 7$). Since the value of gain K at the origin is 7, the range of gain value K for stability is

$$7 < K < 16$$

Figure 6–52(b) shows a sketch of the root loci for the system. Notice that all branches consist of parts of straight lines.

The fact that the root-locus branches consist of straight lines can be verified as follows: Since the angle condition is

$$\angle \frac{K}{(s - 1)(s + 2 + j\sqrt{3})(s + 2 - j\sqrt{3})} = \pm 180°(2k + 1)$$

we have

$$-\angle s - 1 - \angle s + 2 + j\sqrt{3} - \angle s + 2 - j\sqrt{3} = \pm 180°(2k + 1)$$

By substituting $s = \sigma + j\omega$ into this last equation,

$$\angle \sigma - 1 + j\omega + \angle \sigma + 2 + j\omega + j\sqrt{3} + \angle \sigma + 2 + j\omega - j\sqrt{3} = \pm 180°(2k + 1)$$

or

$$\angle \sigma + 2 + j(\omega + \sqrt{3}) + \angle \sigma + 2 + j(\omega - \sqrt{3}) = -\angle \sigma - 1 + j\omega \pm 180°(2k + 1)$$

which can be rewritten as

$$\tan^{-1}\left(\frac{\omega + \sqrt{3}}{\sigma + 2}\right) + \tan^{-1}\left(\frac{\omega - \sqrt{3}}{\sigma + 2}\right) = -\tan^{-1}\left(\frac{\omega}{\sigma - 1}\right) \pm 180°(2k + 1)$$

Taking tangents of both sides of this last equation, we obtain

$$\frac{\dfrac{\omega + \sqrt{3}}{\sigma + 2} + \dfrac{\omega - \sqrt{3}}{\sigma + 2}}{1 - \left(\dfrac{\omega + \sqrt{3}}{\sigma + 2}\right)\left(\dfrac{\omega - \sqrt{3}}{\sigma + 2}\right)} = -\frac{\omega}{\sigma - 1}$$

or

$$\frac{2\omega(\sigma + 2)}{\sigma^2 + 4\sigma + 4 - \omega^2 + 3} = -\frac{\omega}{\sigma - 1}$$

which can be simplified to

$$2\omega(\sigma + 2)(\sigma - 1) = -\omega(\sigma^2 + 4\sigma + 7 - \omega^2)$$

or

$$\omega(3\sigma^2 + 6\sigma + 3 - \omega^2) = 0$$

Further simplification of this last equation yields

$$\omega\left(\sigma + 1 + \frac{1}{\sqrt{3}}\omega\right)\left(\sigma + 1 - \frac{1}{\sqrt{3}}\right) = 0$$

which defines three lines:

$$\omega = 0, \qquad \sigma + 1 + \frac{1}{\sqrt{3}}\omega = 0, \qquad \sigma + 1 - \frac{1}{\sqrt{3}}\omega = 0$$

Thus the root-locus branches consist of three lines. Note that the root loci for $K > 0$ consist of portions of the straight lines as shown in Figure 6–52(b). (Note that each straight line starts from an open-loop pole and extends to infinity in the direction of 180°, 60°, or −60° measured from the real axis.) The remaining portion of each straight line corresponds to $K < 0$.

A-6-11. Consider the system shown in Figure 6–53(a). Sketch the root loci.

Solution. The open-loop zeros of the system are located at $s = \pm j$. The open-loop poles are located at $s = 0$ and $s = -2$. This system involves two poles and two zeros. Hence, there is a possibility that a circular root-locus branch exists. In fact, such a circular root locus exists in this case, as shown in the following. The angle condition is

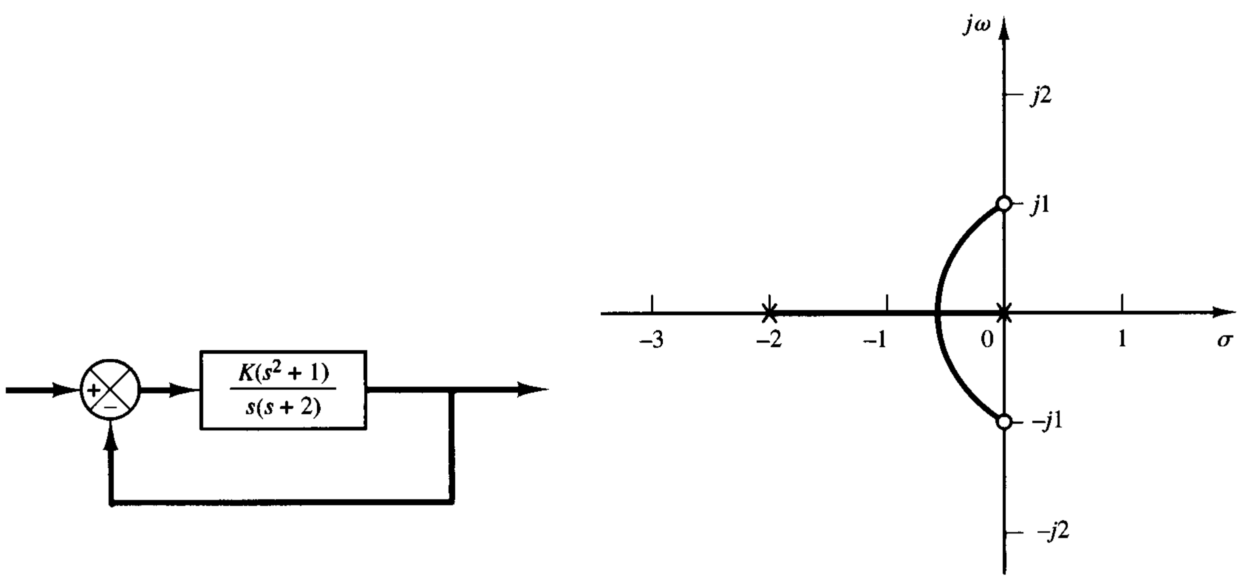


Figure 6–53
(a) Control system; (b) root-locus plot.

$$\underline{\Big/ \frac{K(s+j)(s-j)}{s(s+2)}} = \pm 180°(2k+1)$$

or

$$\underline{/s+j} + \underline{/s-j} - \underline{/s} - \underline{/s+2} = \pm 180°(2k+1)$$

By substituting $s = \sigma + j\omega$ into this last equation, we obtain

$$\underline{/\sigma + j\omega + j} + \underline{/\sigma + j\omega - j} = \underline{/\sigma + j\omega} + \underline{/\sigma + 2 + j\omega} \pm 180°(2k+1)$$

or

$$\tan^{-1}\left(\frac{\omega+1}{\sigma}\right) + \tan^{-1}\left(\frac{\omega-1}{\sigma}\right) = \tan^{-1}\left(\frac{\omega}{\sigma}\right) + \tan^{-1}\left(\frac{\omega}{\sigma+2}\right) \pm 180°(2k+1)$$

Taking tangents of both sides of this equation and noting that

$$\tan\left[\tan^{-1}\left(\frac{\omega}{\sigma+2}\right) \pm 180°\right] = \frac{\omega}{\sigma+2}$$

we obtain

$$\frac{\dfrac{\omega+1}{\sigma} + \dfrac{\omega-1}{\sigma}}{1 - \dfrac{\omega+1}{\sigma}\dfrac{\omega-1}{\sigma}} = \frac{\dfrac{\omega}{\sigma} + \dfrac{\omega}{\sigma+2}}{1 - \dfrac{\omega}{\sigma}\dfrac{\omega}{\sigma+2}}$$

or

$$\omega\left[\left(\sigma - \frac{1}{2}\right)^2 + \omega^2 - \frac{5}{4}\right] = 0$$

which is equivalent to

$$\omega = 0 \qquad \text{or} \qquad \left(\sigma - \frac{1}{2}\right)^2 + \omega^2 = \frac{5}{4}$$

These two equations are equations for the root loci. The first equation corresponds to the root locus on the real axis. (The segment between $s = 0$ and $s = -2$ corresponds to the root locus for $0 \leq K < \infty$. The remaining parts of the real axis correspond to the root locus for $K < 0$.) The second equation is an equation for a circle. Thus, there exists a circular root locus with center at $\sigma = \frac{1}{2}$, $\omega = 0$ and the radius equal to $\sqrt{5}/2$. The root loci are sketched in Figure 6–53(b). [That part of the circular locus to the left of the imaginary zeros corresponds to $K > 0$. The portion of the circular locus not shown in Figure 6–53(b) corresponds to $K < 0$.]

A-6-12. Consider the system shown in Figure 6–54. Determine the value of a such that the damping ratio ζ of the dominant closed-loop poles is 0.5.

Solution. In this system the characteristic equation is

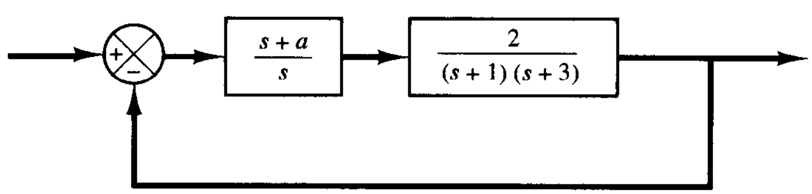


Figure 6–54 Control system.

$$1 + \frac{2(s + a)}{s(s + 1)(s + 3)} = 0$$

Notice that the variable a is not a multiplying factor. Hence, we need to rewrite the characteristic equation

$$s(s + 1)(s + 3) + 2s + 2a = 0$$

as follows:

$$1 + \frac{2a}{s^3 + 4s^2 + 5s} = 0$$

Define

$$2a = K$$

Then we get the characteristic equation in the form

$$1 + \frac{K}{s(s^2 + 4s + 5)} = 0 \tag{6–32}$$

In Problem A-6-4 we constructed the root-locus diagram for the system defined by Equation (6–32). Hence, the solution to this problem is available in Problem A-6-4. Referring to Figure 6–45(b), the closed-loop poles having the damping ratio $\zeta = 0.5$ can be located at $s = -0.63 \pm j1.09$. The value of K at point $s = -0.63 + j1.09$ may be found as 4.32. Hence, the value of a in this problem is obtained as follows:

$$a = \frac{K}{2} = 2.16$$

A-6-13. Consider the system shown in Figure 6–55(a). Determine the value of a such that the damping ratio ζ of the dominant closed poles is 0.5.

Solution. The characteristic equation is

$$1 + \frac{10(s + a)}{s(s + 1)(s + 8)} = 0$$

The variable a is not a multiplying factor. Hence, we need to modify the characteristic equation. Since the characteristic equation can be written as

$$s^3 + 9s^2 + 18s + 10a = 0$$

we rewrite this equation such that a appears as a multiplying factor as follows:

$$1 + \frac{10a}{s(s^2 + 9s + 18)} = 0$$

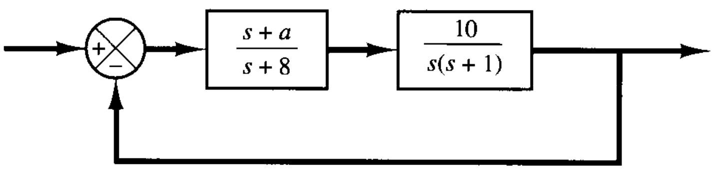
Define

$$10a = K$$

Then the characteristic equation becomes

$$1 + \frac{K}{s(s^2 + 9s + 18)} = 0$$

Notice that the characteristic equation is in a suitable form for the construction of the root loci.



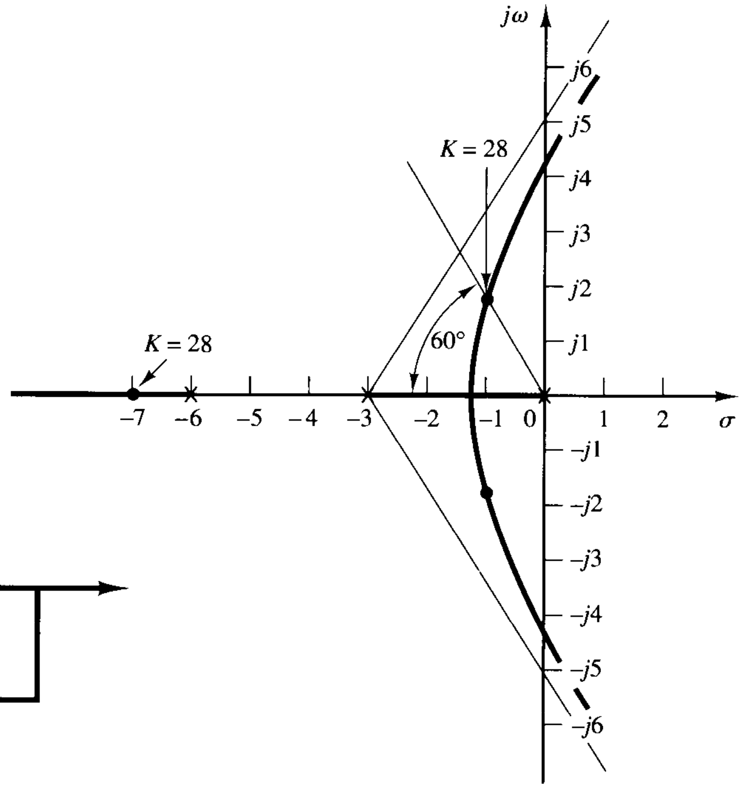


Figure 6–55
(a) Control system; (b) root-locus plot, where $K = 10a$.

This system involves three poles and no zero. The three poles are at $s = 0, s = -3$, and $s = -6$. A root-locus branch exists on the real axis between points $s = 0$ and $s = -3$. Also, another branch exists between points $s = -6$ and $s = -\infty$.

The asymptotes for the root loci are found as follows:

$$\text{Angles of asymptotes} = \frac{\pm 180°(2k + 1)}{3} = 60°, -60°, 180°$$

The intersection of the asymptotes and the real axis is obtained from

$$\sigma_a = -\frac{0 + 3 + 6}{3} = -3$$

The breakaway and break-in points can be determined from $dK/ds = 0$, where

$$K = -(s^3 + 9s^2 + 18s)$$

Now we set

$$\frac{dK}{ds} = -(3s^2 + 18s + 18) = 0$$

which yields

$$s^2 + 6s + 6 = 0$$

or

$$s = -1.268, \qquad s = -4.732$$

Point $s = -1.268$ is on a root-locus branch. Hence, point $s = -1.268$ is an actual breakaway point. But point $s = -4.732$ is not on the root locus and therefore is neither a breakaway nor break-in point.

Next we shall find points where root-locus branches cross the imaginary axis. We substitute $s = j\omega$ in the characteristic equation, which is

$$s^3 + 9s^2 + 18s + K = 0$$

as follows:

$$(j\omega)^3 + 9(j\omega)^2 + 18(j\omega) + K = 0$$

or

$$(K - 9\omega^2) + j\omega(18 - \omega^2) = 0$$

from which we get

$$\omega = \pm 3\sqrt{2}, \qquad K = 9\omega^2 = 162 \qquad \text{or} \qquad \omega = 0, \qquad K = 0$$

The crossing points are at $\omega = \pm 3\sqrt{2}$ and the corresponding value of gain K is 162. Also, a root-locus branch touches the imaginary axis at $\omega = 0$. Figure 6–55(b) shows a sketch of the root loci for the system.

Since the damping ratio of the dominant closed-loop poles is specified as 0.5, the desired closed-loop pole in the upper-half s plane is located at the intersection of the root-locus branch in the upper-half s plane and a straight line having an angle of 60° with the negative real axis. The desired dominant closed-loop poles are located at

$$s = -1 + j1.732, \qquad s = -1 - j1.732$$

At these points, the value of gain K is 28. Hence,

$$a = \frac{K}{10} = 2.8$$

Since the system involves two or more poles than zeros (in fact, three poles and no zero), the third pole can be located on the negative real axis from the fact that the sum of the three closed-loop poles is -9. Hence, the third pole is found to be at

$$s = -9 - (-1 + j1.732) - (-1 - j1.732)$$

or

$$s = -7$$

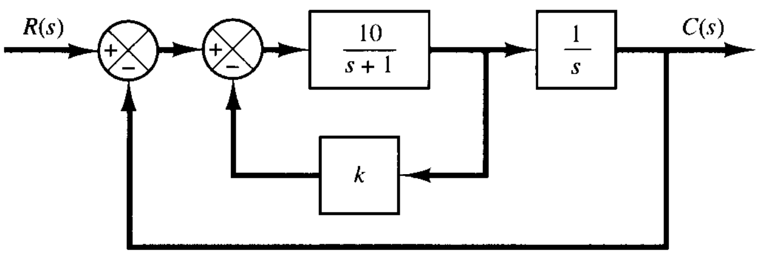
A-6-14. Consider the system shown in Figure 6–56(a). Sketch the root loci of the system as the velocity feedback gain k varies from zero to infinity. Determine the value of k such that the closed-loop poles have the damping ratio ζ of 0.7.

Solution. The open-loop transfer function is

$$\text{Open-loop transfer function} = \frac{10}{(s + 1 + 10k)s + 10}$$

Since k is not a multiplying factor, we modify the equation such that k appears as a multiplying factor. Since the characteristic equation is

$$s^2 + s + 10ks + 10 = 0$$



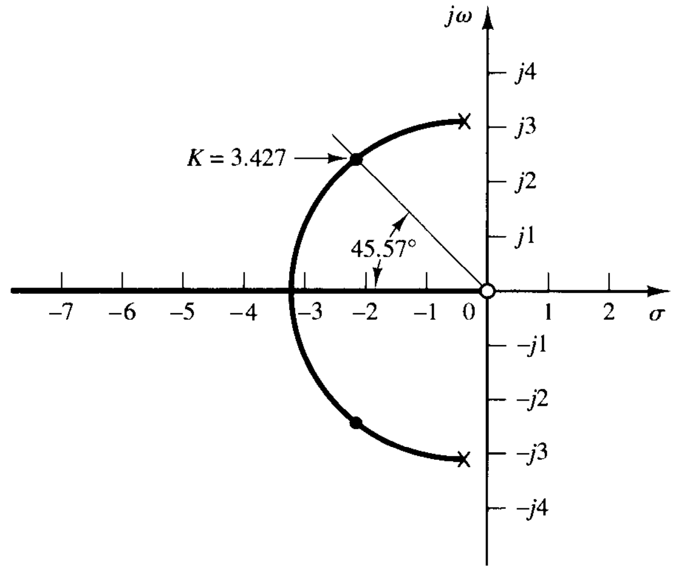


Figure 6–56
(a) Control system; (b) root-locus plot, where $K = 10k$.

we rewrite this equation as follows:

$$1 + \frac{10ks}{s^2 + s + 10} = 0 \tag{6–33}$$

Define

$$10k = K$$

Then Equation (6–33) becomes

$$1 + \frac{Ks}{s^2 + s + 10} = 0$$

Notice that the system has a zero at $s = 0$ and two poles at $s = -0.5 \pm j3.1225$. Since this system involves two poles and one zero, there is a possibility that a circular root locus exists. In fact, this system has a circular root locus, as will be shown. Since the angle condition is

$$\angle \frac{Ks}{s^2 + s + 10} = \pm 180°(2k + 1)$$

we have

$$\angle s - \angle s + 0.5 + j3.1225 - \angle s + 0.5 - j3.1225 = \pm 180°(2k + 1)$$

By substituting $s = \sigma + j\omega$ into this last equation and rearranging, we obtain

$$\angle \sigma + 0.5 + j(\omega + 3.1225) + \angle \sigma + 0.5 + j(\omega - 3.1225) = \angle \sigma + j\omega \pm 180°(2k + 1)$$

which can be rewritten as

$$\tan^{-1}\left(\frac{\omega + 3.1225}{\sigma + 0.5}\right) + \tan^{-1}\left(\frac{\omega - 3.1225}{\sigma + 0.5}\right) = \tan^{-1}\left(\frac{\omega}{\sigma}\right) \pm 180°(2k + 1)$$

Taking tangents of both sides of this last equation, we obtain

$$\frac{\dfrac{\omega + 3.1225}{\sigma + 0.5} + \dfrac{\omega - 3.1225}{\sigma + 0.5}}{1 - \left(\dfrac{\omega + 3.1225}{\sigma + 0.5}\right)\left(\dfrac{\omega - 3.1225}{\sigma + 0.5}\right)} = \frac{\omega}{\sigma}$$

which can be simplified to

$$\frac{2\omega(\sigma + 0.5)}{(\sigma + 0.5)^2 - (\omega^2 - 3.1225^2)} = \frac{\omega}{\sigma}$$

or

$$\omega(\sigma^2 - 10 + \omega^2) = 0$$

which yields

$$\omega = 0 \qquad \text{or} \qquad \sigma^2 + \omega^2 = 10$$

Notice that $\omega = 0$ corresponds to the real axis. The negative real axis (between $s = 0$ and $s = -\infty$) corresponds to $K \geq 0$, and the positive real axis corresponds to $K < 0$. The equation

$$\sigma^2 + \omega^2 = 10$$

is an equation of a circle with center at $\sigma = 0$, $\omega = 0$ with the radius equal to $\sqrt{10}$. A portion of this circle that lies to the left of the complex poles corresponds to the root locus for $K > 0$. The portion of the circle which lies to the right of the complex poles corresponds to the root locus for $K < 0$. Hence, this portion is not a root locus for the present system, where $K > 0$. Figure 6–56(b) shows a sketch of the root loci.

Since we require $\zeta = 0.7$ for the closed-loop poles, we find the intersection of the circular root locus and a line having an angle of 45.57° (note that cos 45.57° = 0.7) with the negative real axis. The intersection is at $s = -2.214 + j2.258$. The gain K corresponding to this point is 3.427. Hence, the desired value of the velocity feedback gain k is

$$k = \frac{K}{10} = 0.3427$$

A-6-15. Consider the control system shown in Figure 6–57. Plot root loci with MATLAB.

Solution. MATLAB Program 6–10 generates a root-locus plot as shown in Figure 6–58. The root loci must be symmetric about the real axis. However, Figure 6–58 shows otherwise.

MATLAB supplies its own set of gain values that are used to calculate a root-locus plot. It does so by an internal adaptive step-size routine. However, in certain systems, very small changes in the gain cause drastic changes in root locations within a certain range of gains. Thus, MATLAB takes too big a jump in its gain values when calculating the roots, and root locations change by a relatively large amount. When plotting, MATLAB connects these points and causes a strange looking graph at the location of sensitive gains. Such erroneous root-locus plots typically occur when the loci approach a double pole (or triple or higher pole), since the locus is very sensitive to small gain changes.

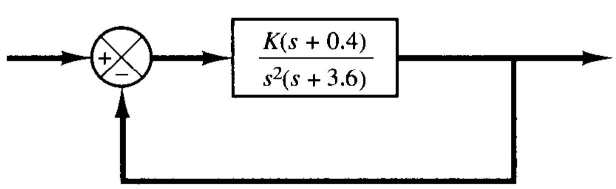


Figure 6–57 Control system.

MATLAB Program 6–10

```
% ---------- Root-locus plot ----------

num = [0   0   1   0.4];
den = [1   3.6   0   0];
rlocus(num,den);
v = [-5   1   -3   3]; axis(v)
grid
title('Root-Locus Plot of G(s) = K(s + 0.4)/[s^2(s + 3.6)]')
```

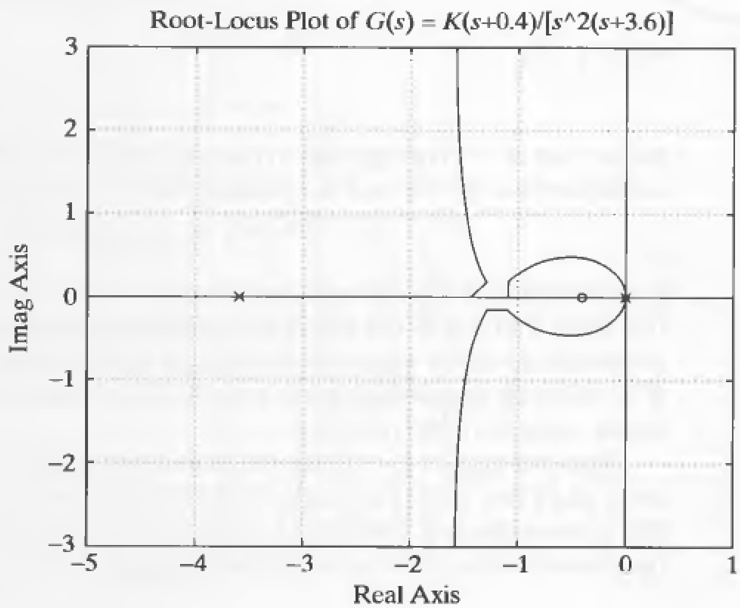


Figure 6–58
Root-locus plot.

In the problem considered here, the critical region of gain K is between 4.2 and 4.4. Thus we need to set the step size small enough in this region. We may divide the region for K as follows:

```
K1 = [0:0.2:4.2];
K2 = [4.2:0.002:4.4];
K3 = [4.4:0.2:10];
K4 = [10:5:200];
K = [K1   K2   K3   K4];
```

Entering MATLAB Program 6–11 into the computer, we obtain the plot as shown in Figure 6–59. If we change the plot command plot(r,'o') in MATLAB Program 6–11 to plot(r,'-'), we obtain Figure 6–60. Figures 6–59 and 6–60 respectively show satisfactory root-locus plots.

A-6-16. Consider the system whose open-loop transfer function $G(s)H(s)$ is given by

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

Using MATLAB, plot root loci and their asymptotes.

```
MATLAB Program 6–11

% ---------- Root-locus plot ----------

num = [0   0   1   0.4];
den = [1   3.6   0   0];
K1 = [0:0.2:4.2];
K2 = [4.2:0.002:4.4];
K3 = [4.4:0.02:10];
K4 = [10:5:200];
K = [K1   K2   K3   K4];
r = rlocus(num,den,K);
plot(r,'o')
v = [-5   1   -5   5]; axis(v)
grid
title('Root-Locus Plot of G(s) = K(s + 0.4)/[s^2(s + 3.6)]')
xlabel('Real Axis')
ylabel('Imag Axis')
```

Solution. We shall plot the root loci and asymptotes on one diagram. Since the open-loop transfer function is given by

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

the equation for the asymptotes may be obtained as follows: Noting that

$$\lim_{s\to\infty} \frac{K}{s^3 + 3s^2 + 2s} \doteq \lim_{s\to\infty} \frac{K}{s^3 + 3s^2 + 3s + 1} = \frac{K}{(s+1)^3}$$

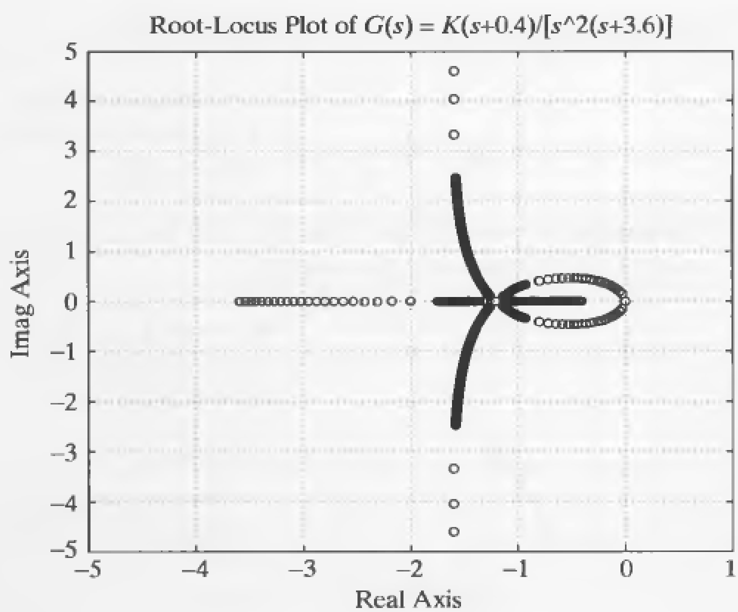


Figure 6–59
Root-locus plot.

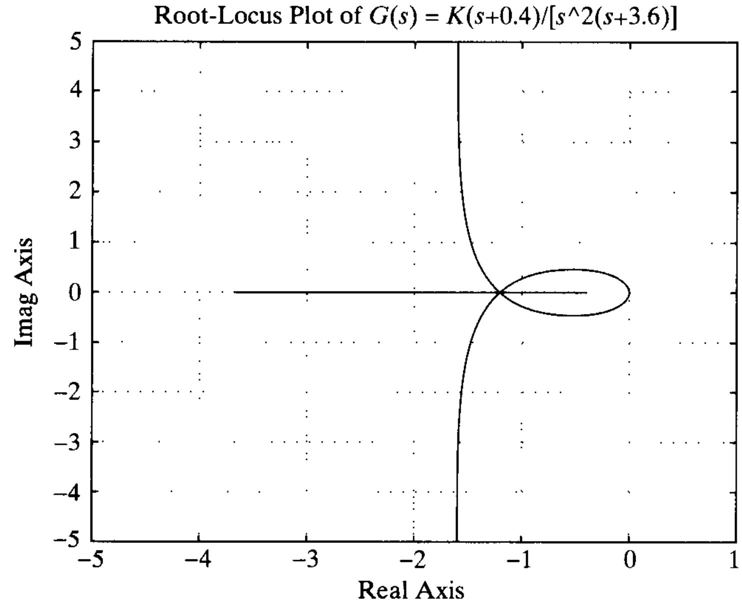


Figure 6–60
Root-locus plot.

the equation for the asymptotes may be given by

$$G_a(s)H_a(s) = \frac{K}{(s+1)^3}$$

Hence, for the system we have

```
num = [0  0  0  1]
den = [1  3  2  0]
```

and for the asymptotes,

```
numa = [0  0  0  1]
dena = [1  3  3  1]
```

In using the following root-locus and plot commands

```
r = rlocus(num,den)
a = rlocus(numa,dena)
plot([r   a])
```

the number of rows of r and that of a must be the same. To ensure this, we include the gain constant K in the commands. For example,

```
K1 = 0:0.1:0.3;
K2 = 0.3:0.005:0.5:
K3 = 0.5:0.5:10;
K4 = 10:5:100;
K = [K1   K2   K3   K4]
r = rlocus(num,den,K)
a = rlocus(numa,dena,K)
y = [r      a]
plot(y,' ')
```

Including gain K in rlocus command ensures that the r matrix and a matrix have the same number of rows. MATLAB Program 6–12 will generate a plot of root loci and their asymptotes. See Figure 6–61.

MATLAB Program 6–12

```
% ---------- Root-Locus Plots ----------

num = [0   0   0   1];
den = [1   3   2   0];
numa = [0   0   0   1];
dena = [1   3   3   1];
K1 = 0:0.1:0.3;
K2 = 0.3:0.005:0.5;
K3 = 0.5:0.5:10;
K4 = 10:5:100;
K = [K1   K2   K3   K4];
r = rlocus(num,den,K);
a = rlocus(numa,dena,K);
y = [r   a];
plot(y,'-')
v = [-4   4   -4   4]; axis(v)
grid
title('Root-Locus Plot of G(s) = K/[s(s + 1)(s + 2)] and Asymptotes')
xlabel('Real Axis')
ylabel('Imag Axis')

% ***** Manually draw open-loop poles in the hard copy *****
```

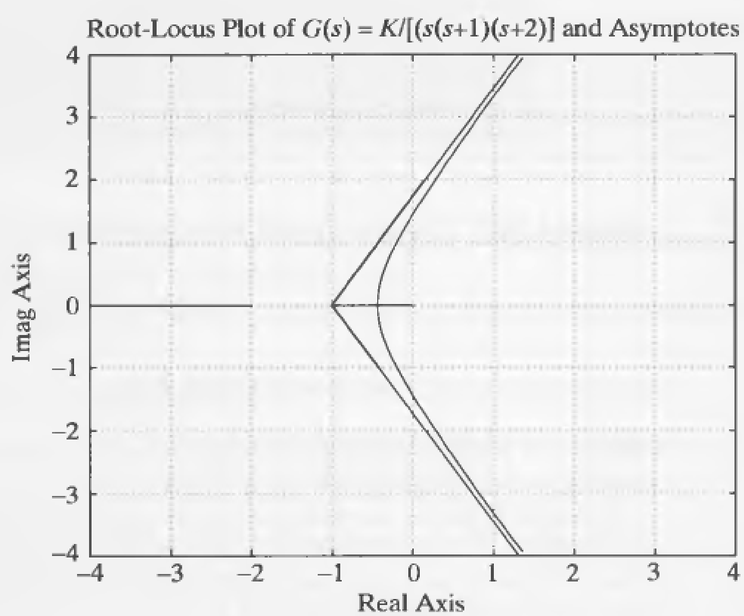


Figure 6–61
Root-locus plot.

Drawing two or more plots in one diagram can be accomplished by using the hold command. MATLAB Program 6–13 uses the hold command. The resulting root-locus plot is shown in Figure 6–62.

MATLAB Program 6–13

```
% ---------- Root-Locus Plots ----------

num = [0   0   0   1];
den = [1   3   2   0];
numa = [0   0   0   1];
dena = [1   3   3   1];
K1 = 0:0.1:0.3;
K2 = 0.3:0.005:0.5;
K3 = 0.5:0.5:10;
K4 = 10:5:100;
K = [K1   K2   K3   K4];
r = rlocus(num,den,K);
a = rlocus(numa,dena,K);
plot(r,'o')
hold
Current plot held
plot(a,'-')
v = [-4   4   -4   4]; axis(v)
grid
title('Root-Locus Plot of G(s) = K/[s(s + 1)(s + 2)] and Asymptotes')
xlabel('Real Axis')
ylabel('Imag Axis')

% ***** Manually draw open-loop poles in the hard copy *****

% ***** Remove hold on graphics *****

hold
Current plot released
```

A-6-17. Consider a unity-feedback system with the following feedforward transfer function $G(s)$:

$$G(s) = \frac{K(s+2)^2}{(s^2+4)(s+5)^2}$$

Plot root loci for the system with MATLAB.

Solution. A MATLAB program to plot the root loci is given as MATLAB Program 6–14. The resulting root-locus plot is shown in Figure 6–63.

Notice that this is a special case where no root locus exists on the real axis. This means that for any value of $K > 0$ the closed-loop poles of the system are two sets of complex-conjugate poles. (No real closed-loop poles exist.) Since no closed-loop poles exist in the right-half s plane, the system is stable for all values of $K > 0$.

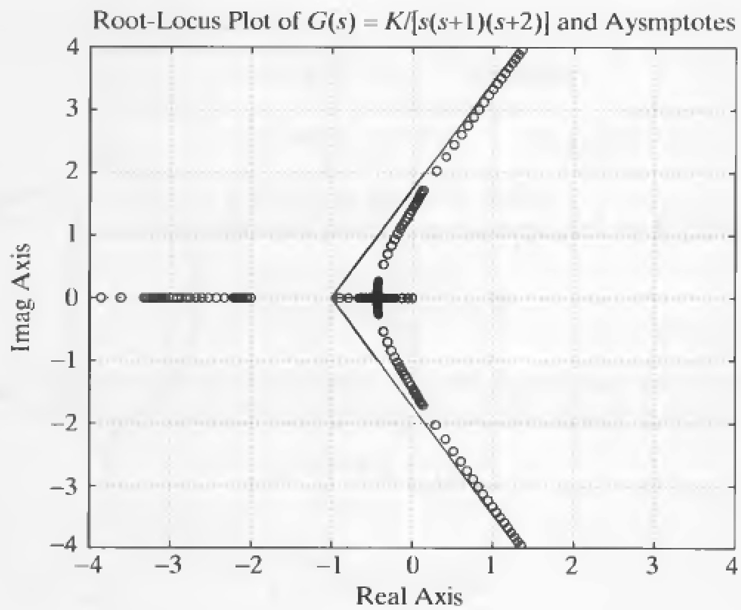


Figure 6–62
Root-locus plot.

MATLAB Program 6–14

```
% ---------- Root-Locus Plot ----------

num = [0   0   1   4   4];
den = [1   10   29   40   100];
r = rlocus(num,den);
plot(r,'o')
hold
Current plot held
plot(r,'-')
v = [-8   4   -6   6]; axis(v); axis('square')
grid
title('Root-Locus Plot of G(s) = (s + 2)^2/[(s^2 + 4)(s + 5)^2]')
xlabel('Real Axis')
ylabel('Imag Axis')
```

A-6-18. Consider the system with transport lag shown in Figure 6–64(a). Sketch the root loci and find the two pairs of closed-loop poles nearest the $j\omega$ axis.

Using only the dominant closed-loop poles, obtain the unit-step response and sketch the response curve.

Solution. The characteristic equation is

$$\frac{2e^{-0.3s}}{s+1} + 1 = 0$$

which is equivalent to the following angle and magnitude conditions:

$$\angle \frac{2e^{-0.3s}}{s+1} = \pm 180°(2k+1)$$

Root-Locus Plot of $G(s) = (s+2)^\wedge 2/[(s^\wedge 2+4)(s+5)^\wedge 2]$

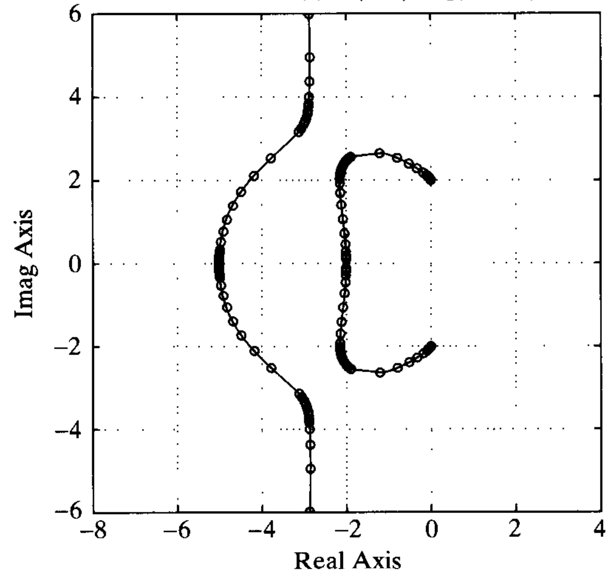


Figure 6–63
Root-locus plot.

$$\left| \frac{2e^{-0.3s}}{s+1} \right| = 1$$

The angle condition reduces to

$$\angle s + 1 = \pm\pi(2k + 1) - 0.3\omega \qquad \text{(radians)}$$

For $k = 0$,

$$\angle s + 1 = \pm\pi - 0.3\omega \qquad \text{(radians)}$$

$$= \pm 180° - 17.2°\omega \qquad \text{(degrees)}$$

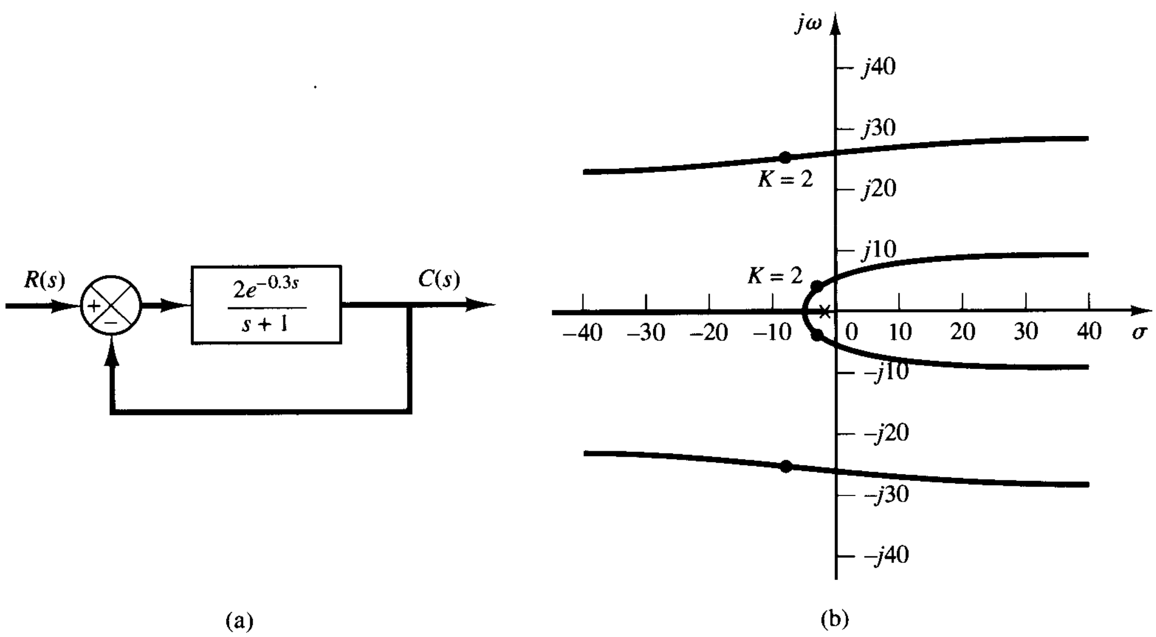


Figure 6–64
(a) Control system with transport lag;
(b) root-locus plot.

For $k = 1$,

$$\underline{/s + 1} = \pm 3\pi - 0.3\omega \qquad \text{(radians)}$$

$$= \pm 540° - 17.2°\omega \qquad \text{(degrees)}$$

The root-locus plot for this system is shown in Figure 6–64(b).

Let us set $s = \sigma + j\omega$ in the magnitude condition and replace 2 by K. Then we obtain

$$\frac{\sqrt{(1 + \sigma)^2 + \omega^2}}{e^{-0.3\sigma}} = K$$

By evaluating K at different points on the root loci, the points may be found for which $K = 2$. These points are closed-loop poles. The dominant pair of closed-loop poles is

$$s = -2.5 \pm j3.9$$

The next pair of closed-loop poles is

$$s = -8.6 \pm j25.1$$

Using only the pair of dominant closed-loop poles, the closed-loop transfer function may be approximated as follows: Noting that

$$\frac{C(s)}{R(s)} = \frac{2e^{-0.3s}}{1 + s + 2e^{-0.3s}}$$

$$= \frac{2e^{-0.3s}}{1 + s + 2\left(1 - 0.3s + \dfrac{0.09s^2}{2} + \cdots\right)}$$

$$= \frac{2e^{-0.3s}}{3 + 0.4s + 0.09s^2 + \cdots}$$

and

$$(s + 2.5 + j3.9)(s + 2.5 - j3.9) = s^2 + 5s + 21.46$$

we may approximate $C(s)/R(s)$ by

$$\frac{C(s)}{R(s)} = \frac{\frac{2}{3}(21.46)e^{-0.3s}}{s^2 + 5s + 21.46}$$

or

$$\frac{C(s)}{R(s)} = \frac{14.31e^{-0.3s}}{(s + 2.5)^2 + 3.9^2}$$

For a unit-step input,

$$C(s) = \frac{14.31e^{-0.3s}}{[(s + 2.5)^2 + 3.9^2]s}$$

Note that

$$\frac{14.31}{[(s + 2.5)^2 + 3.9^2]s} = \frac{\frac{2}{3}}{s} + \frac{-\frac{2}{3}s - \frac{10}{3}}{(s + 2.5)^2 + 3.9^2}$$

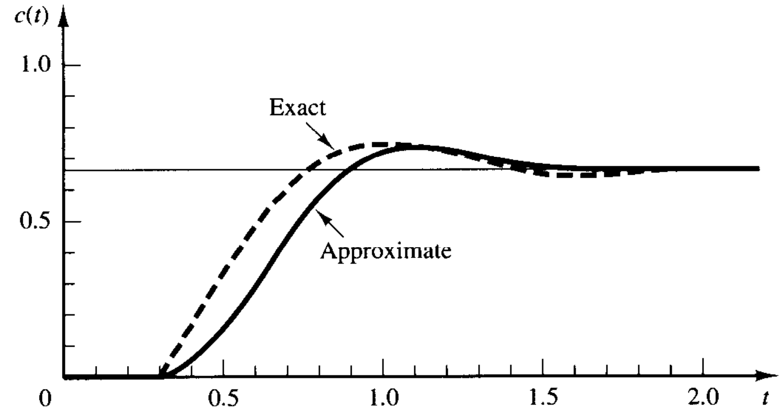


Figure 6–65 Unit-step response curves for the system shown in Figure 6–64(a).

Hence,

$$C(s) = \left(\frac{\frac{2}{3}}{s}\right)e^{-0.3s} + \left[\frac{-\frac{2}{3}s - \frac{10}{3}}{(s + 2.5)^2 + 3.9^2}\right]e^{-0.3s}$$

The inverse Laplace transform of $C(s)$ gives

$$c(t) = \tfrac{2}{3}[1 - e^{-2.5(t-0.3)} \cos 3.9(t - 0.3) - 0.641e^{-2.5(t-0.3)} \sin 3.9(t - 0.3)]1(t - 0.3)$$

where $1(t - 0.3)$ is the unit-step function occurring at $t = 0.3$.

Figure 6–65 shows the approximate response curve thus obtained, together with the exact unit-step response curve obtained by a computer simulation. Note that in this system a fairly good approximation can be obtained by use of only the dominant closed-loop poles.

PROBLEMS

B-6-1. Plot the root loci for the closed-loop control system with

$$G(s) = \frac{K}{s(s + 1)(s^2 + 4s + 5)}, \qquad H(s) = 1$$

B-6-2. Plot the root loci for a closed-loop control system with

$$G(s) = \frac{K(s + 9)}{s(s^2 + 4s + 11)}, \qquad H(s) = 1$$

Locate the closed-loop poles on the root loci such that the dominant closed-loop poles have a damping ratio equal to 0.5. Determine the corresponding value of gain K.

B-6-3. Plot the root loci for the system with

$$G(s) = \frac{K}{s(s + 0.5)(s^2 + 0.6s + 10)}, \qquad H(s) = 1$$

B-6-4. Plot the root loci for a system with

$$G(s) = \frac{K}{(s^2 + 2s + 2)(s^2 + 2s + 5)}, \qquad H(s) = 1$$

Determine the exact points where the root loci cross the $j\omega$ axis.

B-6-5. Show that the root loci for a control system with

$$G(s) = \frac{K(s^2 + 6s + 10)}{s^2 + 2s + 10}, \qquad H(s) = 1$$

are arcs of the circle centered at the origin with radius equal to $\sqrt{10}$.

B-6-6. Plot the root loci for a closed-loop control system with

$$G(s) = \frac{K(s + 0.2)}{s^2(s + 3.6)}, \qquad H(s) = 1$$

B-6-7. Plot the root loci for a closed-loop control system with

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

B-6-8. Plot the root loci for the system shown in Figure 6–66. Determine the range of gain K for stability.

B-6-9. Consider a unity-feedback control system with the following feedforward transfer function:

$$G(s) = \frac{K}{s(s^2 + 4s + 8)}$$

Plot the root loci for the system. If the value of gain K is set equal to 2, where are the closed-loop poles located?

B-6-10. Consider the system shown in Figure 6–67. Determine the values of the gain K and the velocity feedback coefficient K_h so that the closed-loop poles are at $s = -1 \pm j\sqrt{3}$. Then, using the determined value of K_h, plot the root loci.

B-6-11. Consider the system shown in Figure 6–68. The system involves velocity feedback. Determine the value of gain K such that the dominant closed-loop poles have a damping ratio of 0.5. Using the gain K thus determined, obtain the unit-step response of the system.

B-6-12. Consider the system whose open-loop transfer function $G(s)H(s)$ is given by

$$G(s)H(s) = \frac{K}{(s^2 + 2s + 2)(s^2 + 2s + 5)}$$
$$= \frac{K}{s^4 + 4s^3 + 11s^2 + 14s + 10}$$

Plot a root-locus diagram with MATLAB.

B-6-13. Consider the system whose open-loop transfer function is given by

$$G(s)H(s) = \frac{K(s - 0.6667)}{s^4 + 3.3401s^3 + 7.0325s^2}$$

Show that the equation for the asymptotes is given by

$$G_a(s)H_a(s) = \frac{K}{s^3 + 4.0068s^2 + 5.3515s + 2.3825}$$

Using MATLAB, plot the root loci and asymptotes for the system.

B-6-14. Consider the unity-feedback system whose feedforward transfer function is

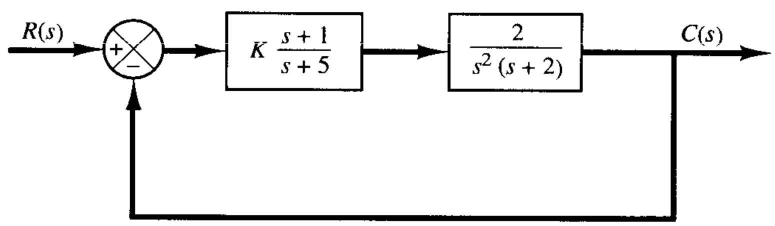


Figure 6–66
Control system.

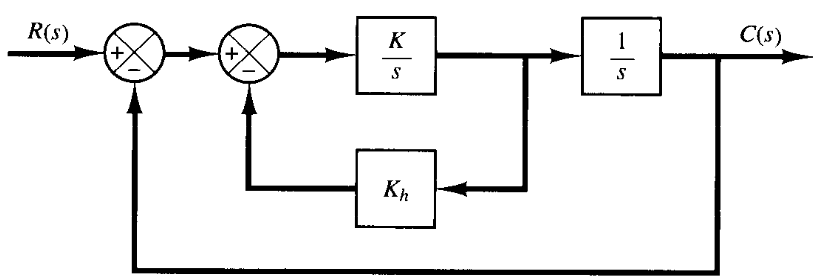


Figure 6–67
Control system.

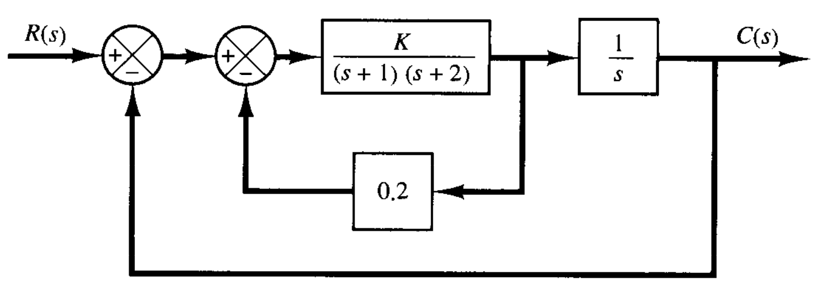


Figure 6–68
Control system.

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

The constant-gain locus for the system for a given value of K is defined by the following equation:

$$\left| \frac{K}{s(s+1)} \right| = 1$$

Show that the constant-gain loci for $0 \le K \le \infty$ may be given by

$$[\sigma(\sigma + 1) + \omega^2]^2 + \omega^2 = K^2$$

Sketch the constant-gain loci for $K = 1, 2, 5, 10$, and 20 on the s plane.

B-6-15. Consider the system shown in Figure 6–69. Plot the root loci. Locate the closed-loop poles when the gain K is set equal to 2.

B-6-16. Consider the system shown in Figure 6–70. Plot the root loci as a varies from 0 to ∞. Determine the value of a such that the damping ratio of the dominant closed-loop poles is 0.5.

B-6-17. Consider the system shown in Figure 6–71. Plot the root loci as the value of k varies from 0 to ∞. What value of k will give the damping ratio of the dominant closed-loop poles equal to 0.5? Find the static velocity error constant with this value of k.

B-6-18. Plot the root loci for the system shown in Figure 6–72. Show that the system may become unstable for large values of K.

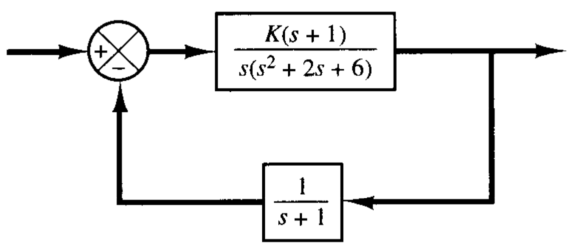


Figure 6–69
Control system.

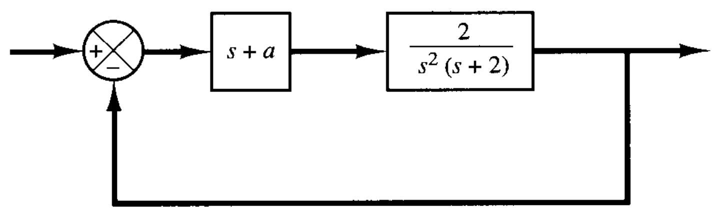


Figure 6–70
Control system.

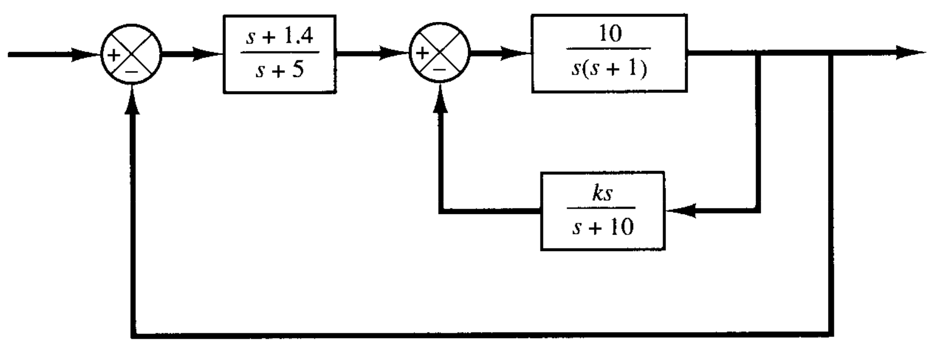


Figure 6–71
Control system.

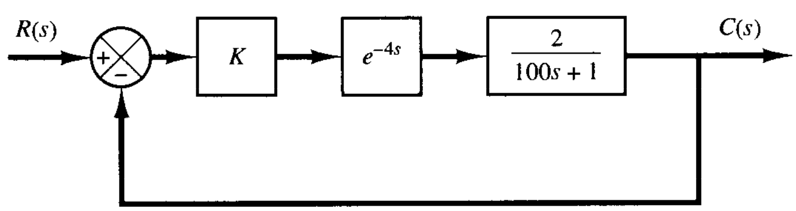


Figure 6–72
Control system.

B-6-19. Plot the root contours for the system shown in Figure 6–73 when the gain K and parameter a vary, respectively, from zero to infinity.

B-6-20. Consider the system shown in Figure 6–74. Assuming that the value of gain K varies from 0 to ∞, plot the root loci when $K_h = 0.5$. Then sketch the root contours for $0 \le K < \infty$ and $0 \le K_h < \infty$. Locate the closed-loop poles on the root contour when $K = 10$ and $K_h = 0.5$.

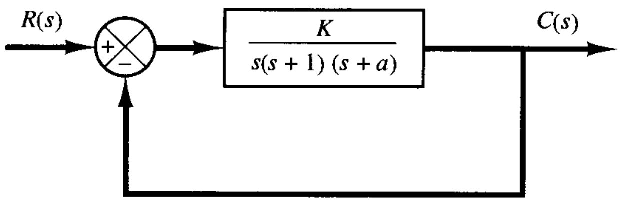


Figure 6–73
Control system.

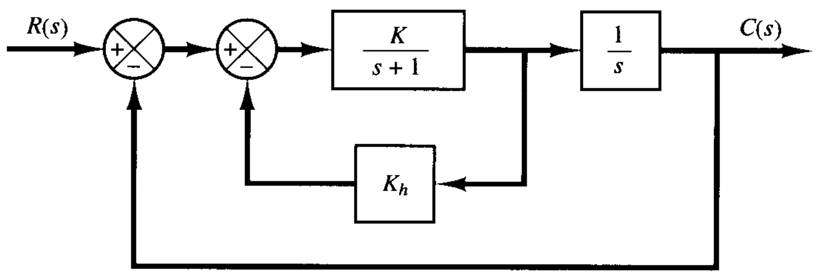


Figure 6–74
Control system.

7

Control Systems Design by the Root-Locus Method

7–1 INTRODUCTION

The primary objective of this chapter is to present procedures for the design and compensation of single-input–single-output linear time-invariant control systems. Compensation is the modification of the system dynamics to satisfy the given specifications. The approach to the control system design and compensation used in this chapter is the root-locus approach. (The frequency-response approach and the state-space approach to the control systems design and compensation will be presented in Chapter 9 and Chapter 11, respectively.)

Performance specifications. Control systems are designed to perform specific tasks. The requirements imposed on the control system are usually spelled out as performance specifications. They generally relate to accuracy, relative stability, and speed of response.

For routine design problems, the performance specifications may be given in terms of precise numerical values. In other cases, they may be given partially in terms of precise numerical values and partially in terms of qualitative statements. In the latter case, the specifications may have to be modified during the course of design since the given specifications may never be satisfied (because of conflicting requirements) or may lead to a very expensive system.

Generally, the performance specifications should not be more stringent than necessary to perform the given task. If the accuracy at steady-state operation is of prime importance in a given control system, then we should not require unnecessarily rigid performance specifications on the transient response since such specifications will require expensive components. Remember that the most important part of control system design is to state the performance specifications precisely so that they will yield an optimal control system for the given purpose.

System compensation. Setting the gain is the first step in adjusting the system for satisfactory performance. In many practical cases, however, the adjustment of the gain alone may not provide sufficient alteration of the system behavior to meet the given specifications. As is frequently the case, increasing the gain value will improve the steady-state behavior but will result in poor stability or even instability. It is then necessary to redesign the system (by modifying the structure or by incorporating additional devices or components) to alter the overall behavior so that the system will behave as desired. Such a redesign or addition of a suitable device is called *compensation*. A device inserted into the system for the purpose of satisfying the specifications is called a *compensator*. The compensator compensates for deficit performance of the original system.

Series compensation and feedback (or parallel) compensation. Figures 7–1 (a) and (b) show compensation schemes commonly used for feedback control systems. Figure 7–1(a) shows the configuration where the compensator $G_c(s)$ is placed in series with the plant. This scheme is called *series compensation*.

An alternative to series compensation is to feed back the signal(s) from some element(s) and place a compensator in the resulting inner feedback path, as shown in Figure 7–1(b). Such compensation is called *feedback compensation* or *parallel compensation*.

In compensating control systems, we see that the problem usually boils down to a suitable design of a series or feedback compensator. The choice between series compensation and feedback compensation depends on the nature of the signals in the

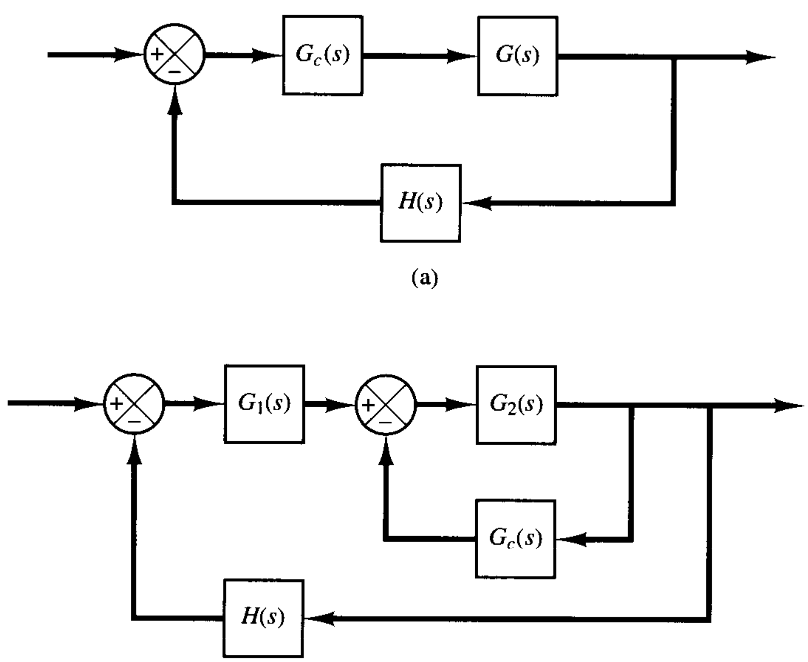


Figure 7–1
(a) Series compensation; (b) feedback or parallel compensation.

system, the power levels at various points, available components, the designer's experience, economic considerations, and so on.

In general, series compensation may be simpler than feedback compensation; however, series compensation frequently requires additional amplifiers to increase the gain and/or to provide isolation. (To avoid power dissipation, the series compensator is inserted at the lowest energy point in the feedforward path.) Note that, in general, the number of components required in feedback compensation will be less than the number of components in series compensation, provided a suitable signal is available, because the energy transfer is from a higher power level to a lower level. (This means that additional amplifiers may not be necessary.)

In discussing compensators, we frequently use such terminologies as *lead network, lag network,* and *lag–lead network.* As stated in Section 5–9, if a sinusoidal input e_i is applied to the input of a network and the steady-state output e_o (which is also sinusoidal) has a phase lead, then the network is called a lead network. (The amount of phase lead angle is a function of the input frequency.) If the steady-state output e_o has a phase lag, then the network is called a lag network. In a lag–lead network, both phase lag and phase lead occur in the output but in different frequency regions; phase lag occurs in the low-frequency region and phase lead occurs in the high-frequency region. A compensator having a characteristic of a lead network, lag network, or lag–lead network is called a lead compensator, lag compensator, or lag–lead compensator.

Compensators. If a compensator is needed to meet the performance specifications, the designer must realize a physical device that has the prescribed transfer function of the compensator.

Numerous physical devices have been used for such purposes. In fact, many noble and useful ideas for physically constructing compensators may be found in the literature.

Among the many kinds of compensators, widely employed compensators are the lead compensators, lag compensators, lag–lead compensators, and velocity-feedback (tachometer) compensators. In this chapter we shall limit our discussions mostly to these types. Lead, lag, and lag–lead compensators may be electronic devices (such as circuits using operational amplifiers) or *RC* networks (electrical, mechanical, pneumatic, hydraulic, or combinations thereof) and amplifiers.

In the actual design of a control system, whether to use an electronic, pneumatic, or hydraulic compensator is a matter that must be decided partially based on the nature of the controlled plant. For example, if the controlled plant involves flammable fluid, then we have to choose pneumatic components (both a compensator and an actuator) to avoid the possibility of sparks. If, however, no fire hazard exists, then electronic compensators are most commonly used. (In fact, we often transform nonelectrical signals into electrical signals because of the simplicity of transmission, increased accuracy, increased reliability, ease of compensation, and the like.)

Design procedures. In the trial-and-error approach to system design, we set up a mathematical model of the control system and adjust the parameters of a compensator. The most time-consuming part of such work is the checking of the system performance by analysis with each adjustment of the parameters. The designer should make use of a digital computer to avoid much of the numerical drudgery necessary for this checking.

Once a satisfactory mathematical model has been obtained, the designer must construct a prototype and test the open-loop system. If absolute stability of the closed loop is assured, the designer closes the loop and tests the performance of the resulting closed-loop system. Because of the neglected loading effects among the components, nonlinearities, distributed parameters, and so on, which were not taken into consideration in the original design work, the actual performance of the prototype system will probably differ from the theoretical predictions. Thus the first design may not satisfy all the requirements on performance. By trial and error, the designer must make changes in the prototype until the system meets the specifications. In doing this, he or she must analyze each trial, and the results of the analysis must be incorporated into the next trial. The designer must see that the final system meets the performance specifications and, at the same time, is reliable and economical.

It is noted that in designing control systems by the root-locus or frequency-response methods the final result is not unique, because the best or optimal solution may not be precisely defined if the time-domain specifications or frequency-domain specifications are given.

Outline of the chapter. Section 7–1 has presented an introduction to the compensation of control systems. Section 7–2 discusses preliminary considerations for the root-locus approach to the control systems design. Section 7–3 treats details of the lead compensation techniques based on the root-locus method. Section 7–4 deals with the lag compensation techniques by the root-locus method. Section 7–5 presents lag–lead compensation techniques. Detailed discussions of the design of lag–lead compensators are presented.

7–2 PRELIMINARY DESIGN CONSIDERATIONS

In building a control system, we know that proper modification of the plant dynamics may be a simple way to meet the performance specifications. This, however, may not be possible in many practical situations because the plant may be fixed and may not be modified. Then we must adjust parameters other than those in the fixed plant. In this book, we assume that the plant is given and unalterable.

The design problems, therefore, become those of improving system performance by insertion of a compensator. Compensation of a control system is reduced to the design of a filter whose characteristics tend to compensate for the undesirable and unalterable characteristics of the plant. Our discussions are limited to continuous-time compensators.

In Section 7–3 through 7–5, we shall specifically consider the design of lead compensators, lag compensators, and lag–lead compensators. In such design problems, we place a compensator in series with the unalterable transfer function $G(s)$ to obtain desirable behavior. The main problem then involves the judicious choice of the pole(s) and zero(s) of the compensator $G_c(s)$ to alter the root locus (or frequency response) so that the performance specifications will be met.

Root–locus approach to control system design. The root-locus method is a graphical method for determining the locations of all closed-loop poles from knowledge of the locations of the open-loop poles and zeros as some parameter (usually the gain) is varied from zero to infinity. The method yields a clear indication of the effects of parameter adjustment.

In practice, the root-locus plot of a system may indicate that the desired performance cannot be achieved just by the adjustment of gain. In fact, in some cases, the system may not be stable for all values of gain. Then it is necessary to reshape the root loci to meet the performance specifications.

In designing a control system, if other than a gain adjustment is required, we must modify the original root loci by inserting a suitable compensator. Once the effects on the root locus of the addition of poles and/or zeros are fully understood, we can readily determine the locations of the pole(s) and zero(s) of the compensator that will reshape the root locus as desired. In essence, in the design by the root-locus method, the root loci of the system are reshaped through the use of a compensator so that a pair of dominant closed-loop poles can be placed at the desired location. (Often, the damping ratio and undamped natural frequency of a pair of dominant closed-loop poles are specified.)

Effects of the addition of poles. The addition of a pole to the open-loop transfer function has the effect of pulling the root locus to the right, tending to lower the system's relative stability and to slow down the settling of the response. (Remember that the addition of integral control adds a pole at the origin, thus making the system less stable.) Figure 7–2 shows examples of root loci illustrating the effects of the addition of a pole to a single-pole system and the addition of two poles to a single-pole system.

Effects of the addition of zeros. The addition of a zero to the open-loop transfer function has the effect of pulling the root locus to the left, tending to make the system more stable and to speed up the settling of the response. (Physically, the addition of a zero in the feedforward transfer function means the addition of derivative control to the system. The effect of such control is to introduce a degree of anticipation into the system and speed up the transient response.) Figure 7–3(a) shows the root loci for a system that is stable for small gain but unstable for large gain. Figures 7–3(b), (c), and (d) show root-locus plots for the system when a zero is added to the open-loop transfer

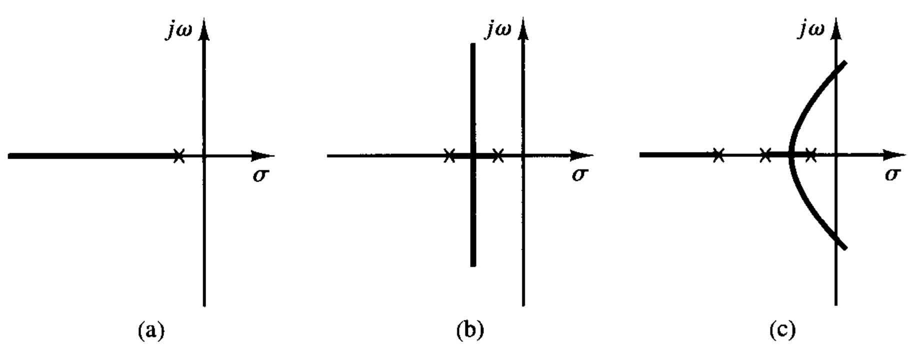


Figure 7–2
(a) Root-locus plot of a single-pole system; (b) root-locus plot of a two-pole system; (c) root-locus plot of a three-pole system.

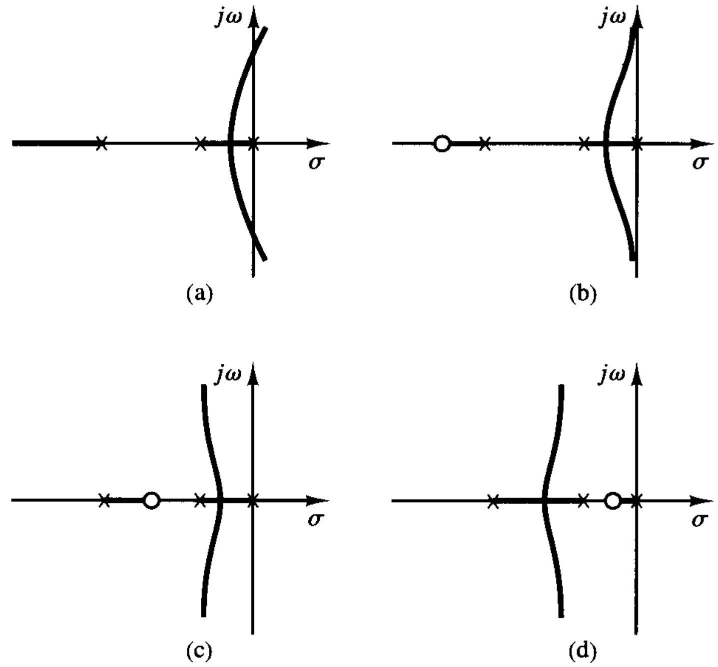


Figure 7–3
(a) Root-locus plot of a three-pole system; (b), (c), and (d) root-locus plots showing effects of addition of a zero to the three-pole system.

function. Notice that when a zero is added to the system of Figure 7–3(a) it becomes stable for all values of gain.

7-3 LEAD COMPENSATION

Lead compensators. There are many ways to realize continuous-time (or analog) lead compensators, such as electronic networks using operational amplifiers, electrical *RC* networks, and mechanical spring–dashpot systems. Compensators using operational amplifiers are frequently used in practice. (Refer to Chapter 5 for networks using operational amplifiers.)

Figure 7–4 shows an electronic circuit using operational amplifiers. The transfer function for this circuit was obtained in Chapter 5 as follows:

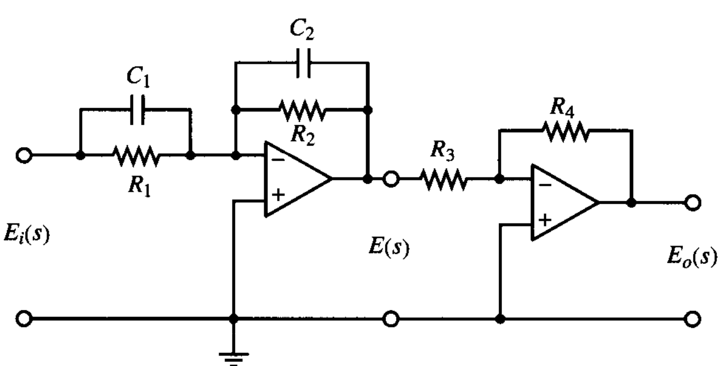


Figure 7–4
Electronic circuit that is a lead network if $R_1C_1 > R_2C_2$ and a lag network if $R_1C_1 < R_2C_2$.

$$\frac{E_o(s)}{E_i(s)} = \frac{R_2R_4}{R_1R_3}\frac{R_1C_1s + 1}{R_2C_2s + 1} = \frac{R_4C_1}{R_3C_2}\frac{s + \frac{1}{R_1C_1}}{s + \frac{1}{R_2C_2}}$$

$$= K_c\alpha\frac{Ts + 1}{\alpha Ts + 1} = K_c\frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}} \tag{7–1}$$

where

$$T = R_1C_1, \qquad \alpha T = R_2C_2, \qquad K_c = \frac{R_4C_1}{R_3C_2}$$

Notice that

$$K_c\alpha = \frac{R_4C_1}{R_3C_2}\frac{R_2C_2}{R_1C_1} = \frac{R_2R_4}{R_1R_3}, \qquad \alpha = \frac{R_2C_2}{R_1C_1}$$

This network has a dc gain of $K_c\alpha = R_2R_4/(R_1R_3)$.

From Equation (7–1), we see that this network is a lead network if $R_1C_1 > R_2C_2$, or $\alpha < 1$. It is a lag network if $R_1C_1 < R_2C_2$. The pole–zero configurations of this network when $R_1C_1 > R_2C_2$ and $R_1C_1 < R_2C_2$ are shown in Figure 7–5(a) and (b), respectively.

Lead compensation techniques based on the root-locus approach. The root-locus approach to design is very powerful when the specifications are given in terms of time-domain quantities, such as the damping ratio and undamped natural frequency of the desired dominant closed-loop poles, maximum overshoot, rise time, and settling time.

Consider a design problem in which the original system either is unstable for all values of gain or is stable but has undesirable transient-response characteristics. In such a

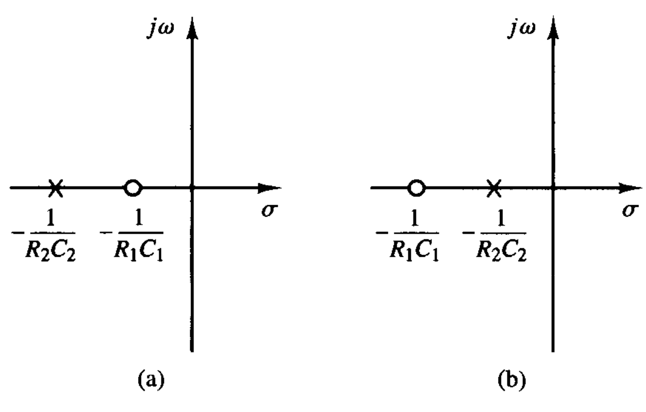


Figure 7–5
Pole–zero configurations: (a) lead network; (b) lag network.

case, the reshaping of the root locus is necessary in the broad neighborhood of the $j\omega$ axis and the origin in order that the dominant closed-loop poles be at desired locations in the complex plane. This problem may be solved by inserting an appropriate lead compensator in cascade with the feedforward transfer function.

The procedures for designing a lead compensator for the system shown in Figure 7–6 by the root-locus method may be stated as follows:

1. From the performance specifications, determine the desired location for the dominant closed-loop poles.

2. By drawing the root-locus plot, ascertain whether or not the gain adjustment alone can yield the desired closed-loop poles. If not, calculate the angle deficiency ϕ. This angle must be contributed by the lead compensator if the new root locus is to pass through the desired locations for the dominant closed-loop poles.

3. Assume the lead compensator $G_c(s)$ to be

$$G_c(s) = K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}}, \qquad (0<\alpha<1)$$

where α and T are determined from the angle deficiency. K_c is determined from the requirement of the open-loop gain.

4. If static error constants are not specified, determine the location of the pole and zero of the lead compensator so that the lead compensator will contribute the necessary angle ϕ. If no other requirements are imposed on the system, try to make the value of α as large as possible. A larger value of α generally results in a larger value of K_v, which is desirable. (If a particular static error constant is specified, it is generally simpler to use the frequency-response approach.)

5. Determine the open-loop gain of the compensated system from the magnitude condition.

Once a compensator has been designed, check to see whether all performance specifications have been met. If the compensated system does not meet the performance specifications, then repeat the design procedure by adjusting the compensator pole and zero until all such specifications are met. If a large static error constant is required, cascade a lag network or alter the lead compensator to a lag–lead compensator.

Note that if the selected dominant closed-loop poles are not really dominant, it will be necessary to modify the location of the pair of such selected dominant closed-loop

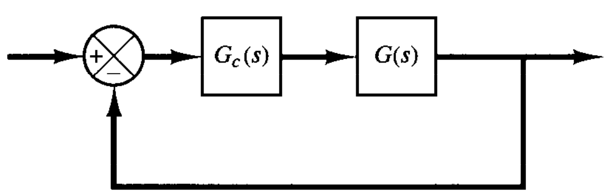


Figure 7–6
Control system.

poles. (The closed-loop poles other than dominant ones modify the response obtained from the dominant closed-loop poles alone. The amount of modification depends on the location of these remaining closed-loop poles.) Also, the closed-loop zeros affect the response if they are located near the origin.

EXAMPLE 7–1 Consider the system shown in Figure 7–7(a). The feedforward transfer function is

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

The root-locus plot for this system is shown in Figure 7–7(b). The closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{4}{s^2 + 2s + 4}$$

$$= \frac{4}{(s + 1 + j\sqrt{3})(s + 1 - j\sqrt{3})}$$

The closed-loop poles are located at

$$s = -1 \pm j\sqrt{3}$$

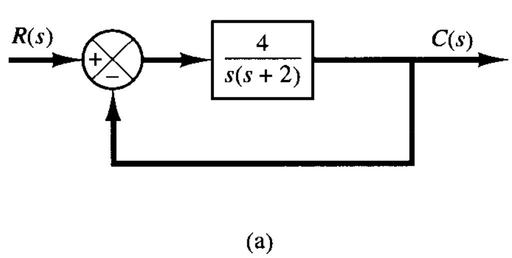
The damping ratio of the closed-loop poles is 0.5. The undamped natural frequency of the closed-loop poles is 2 rad/sec. The static velocity error constant is 2 sec^{-1}.

It is desired to modify the closed-loop poles so that an undamped natural frequency $\omega_n = 4$ rad/sec is obtained, without changing the value of the damping ratio, $\zeta = 0.5$.

Recall that in the complex plane the damping ratio ζ of a pair of complex conjugate poles can be expressed in terms of the angle θ, which is measured from the $j\omega$ axis, as shown in Figure 7–8(a), with

$$\zeta = \sin\theta$$

In other words, lines of constant damping ratio ζ are radial lines passing through the origin as shown in Figure 7–8(b). For example, a damping ratio of 0.5 requires that the complex poles lie on the lines drawn through the origin making angles of $\pm 60°$ with the negative real axis. (If the real part of a pair of complex poles is positive, which means that the system is unstable, the cor-



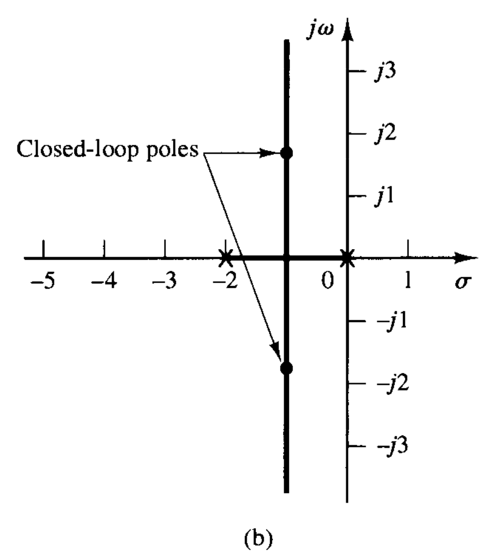


Figure 7–7
(a) Control system;
(b) root-locus plot.

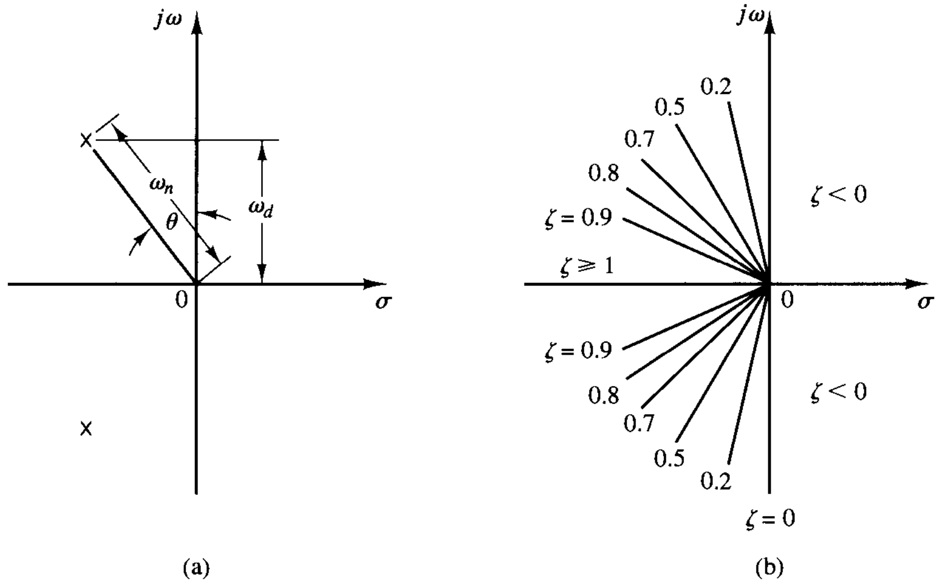


Figure 7–8
(a) Complex poles; (b) lines of constant damping ratio ζ.

responding ζ is negative.) The damping ratio determines the angular location of the poles, while the distance of the pole from the origin is determined by the undamped natural frequency ω_n.

In the present example, the desired locations of the closed-loop poles are

$$s = -2 \pm j2\sqrt{3}$$

In some cases, after the root loci of the original system have been obtained, the dominant closed-loop poles may be moved to the desired location by simple gain adjustment. This is, however, not the case for the present system. Therefore, we shall insert a lead compensator in the feedforward path.

A general procedure for determining the lead compensator is as follows: First, find the sum of the angles at the desired location of one of the dominant closed-loop poles with the open-loop poles and zeros of the original system, and determine the necessary angle ϕ to be added so that the total sum of the angles is equal to $\pm 180°(2k + 1)$. The lead compensator must contribute this angle ϕ. (If the angle ϕ is quite large, then two or more lead networks may be needed rather than a single one.)

If the original system has the open-loop transfer function $G(s)$, then the compensated system will have the open-loop transfer function

$$G_c(s)G(s) = \left(K_c \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\alpha T}} \right) G(s)$$

where

$$G_c(s) = K_c \alpha \frac{Ts + 1}{\alpha Ts + 1} = K_c \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\alpha T}}, \qquad (0 < \alpha < 1)$$

Notice that there are many possible values for T and α that will yield the necessary angle contribution at the desired closed-loop poles.

The next step is to determine the locations of the zero and pole of the lead compensator. There are many possibilities for the choice of such locations. (See the comments at the end of this example problem.) In what follows, we shall introduce a procedure to obtain the largest possible value for α. (Note that a larger value of α will produce a larger value of K_v. In most cases, the larger the K_v is, the better the system performance.) First, draw a horizontal line passing through point P, the desired location for one of the dominant closed-loop poles. This is shown as line PA in Figure 7–9. Draw also a line connecting point P and the origin. Bisect the angle between the lines PA and PO, as shown in Figure 7–9. Draw two lines PC and PD that make angles $\pm\phi/2$ with the bisector PB. The intersections of PC and PD with the negative real axis give the necessary location for the pole and zero of the lead network. The compensator thus designed will make point P a point on the root locus of the compensated system. The open-loop gain is determined by use of the magnitude condition.

In the present system, the angle of $G(s)$ at the desired closed-loop pole is

$$\left. \angle \frac{4}{s(s+2)} \right|_{s=-2+j2\sqrt{3}} = -210^\circ$$

Thus, if we need to force the root locus to go through the desired closed-loop pole, the lead compensator must contribute $\phi = 30^\circ$ at this point. By following the foregoing design procedure, we determine the zero and pole of the lead compensator, as shown in Figure 7–10, to be

$$\text{Zero at } s = -2.9, \qquad \text{Pole at } s = -5.4$$

or

$$T = \frac{1}{2.9} = 0.345, \qquad \alpha T = \frac{1}{5.4} = 0.185$$

Thus $\alpha = 0.537$. The open-loop transfer function of the compensated system becomes

$$G_c(s)G(s) = K_c \frac{s+2.9}{s+5.4} \frac{4}{s(s+2)} = \frac{K(s+2.9)}{s(s+2)(s+5.4)}$$

where $K = 4K_c$. The root-locus plot for the compensated system is shown in Figure 7–10. The gain K is evaluated from the magnitude condition as follows: Referring to the root-locus plot for the compensated system shown in Figure 7–10, the gain K is evaluated from the magnitude condition as

$$\left| \frac{K(s+2.9)}{s(s+2)(s+5.4)} \right|_{s=-2+j2\sqrt{3}} = 1$$

or

$$K = 18.7$$

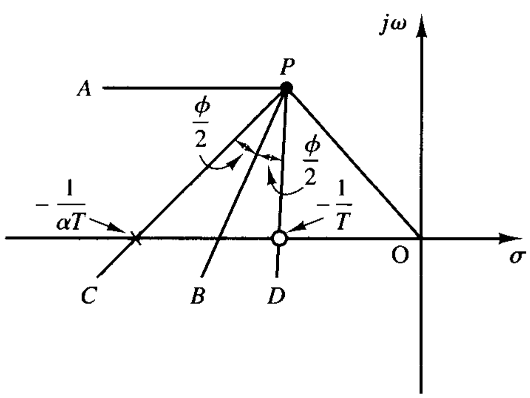


Figure 7–9
Determination of the pole and zero of a lead network.

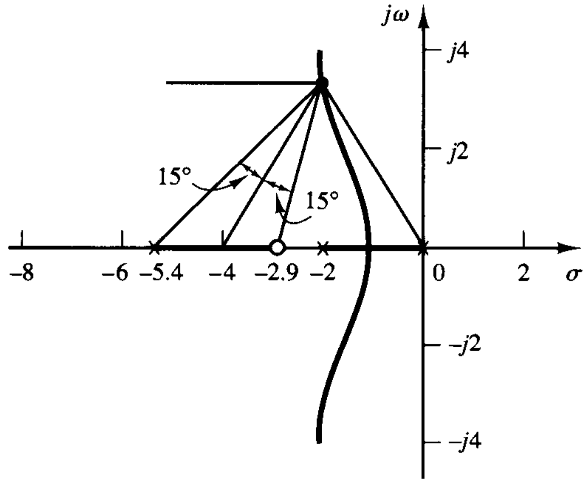


Figure 7–10
Root-locus plot of the compensated system.

It follows that

$$G_c(s)G(s) = \frac{18.7(s + 2.9)}{s(s + 2)(s + 5.4)}$$

The constant K_c of the lead compensator is

$$K_c = \frac{18.7}{4} = 4.68$$

Hence, $K_c\alpha = 2.51$. The lead compensator, therefore, has the transfer function

$$G_s(s) = 2.51 \frac{0.345s + 1}{0.185s + 1} = 4.68 \frac{s + 2.9}{s + 5.4}$$

If the electronic circuit using operational amplifiers as shown in Figure 7–4 is used as the lead compensator just designed, then the parameter values of the lead compensator are determined from

$$\frac{E_o(s)}{E_i(s)} = \frac{R_2 R_4}{R_1 R_3} \frac{R_1 C_1 s + 1}{R_2 C_2 s + 1} = 2.51 \frac{0.345s + 1}{0.185s + 1}$$

as shown in Figure 7–11, where we have arbitrarily chosen $C_1 = C_2 = 10\ \mu\text{F}$ and $R_3 = 10\ \text{k}\Omega$.

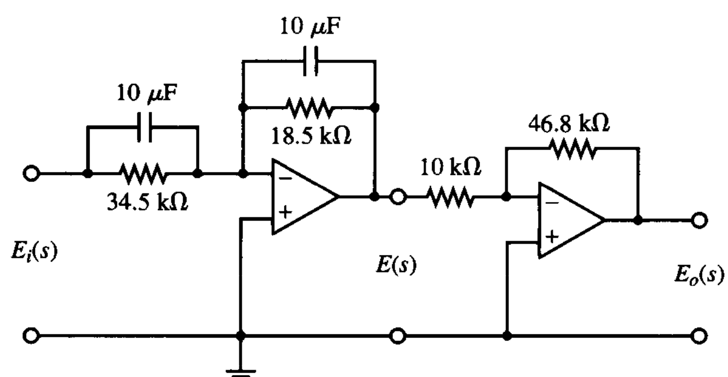


Figure 7–11
Lead compensator.

The static velocity error constant K_v is obtained from the expression

$$K_v = \lim_{s \to 0} sG_c(s)G(s)$$

$$= \lim_{s \to 0} \frac{s18.7(s + 2.9)}{s(s + 2)(s + 5.4)}$$

$$= 5.02 \text{ sec}^{-1}$$

Note that the third closed-loop pole of the designed system is found by dividing the characteristic equation by the known factors as follows:

$$s(s + 2)(s + 5.4) + 18.7(s + 2.9) = (s + 2 + j2\sqrt{3})(s + 2 - j2\sqrt{3})(s + 3.4)$$

The foregoing compensation method enables us to place the dominant closed-loop poles at the desired points in the complex plane. The third pole at $s = 3.4$ is close to the added zero at $s = -2.9$. Therefore, the effect of this pole on the transient response is relatively small. Since no restriction has been imposed on the nondominant pole and no specification has been given concerning the value of the static velocity error coefficient, we conclude that the present design is satisfactory.

Comments. We may place the zero of the compensator at $s = -2$ and pole at $s = -4$ so that the angle contribution of the lead compensator is 30°. (In this case the zero of the lead compensator will cancel a pole of the plant, resulting in the second-order system, rather than the third-order system as we designed.) It can be seen that the K_v value in this case is 4 sec^{-1}. Other combinations can be selected that will yield 30° phase lead. (For different combinations of a zero and pole of the compensator that contribute 30°, the value of α will be different and the value of K_v will also be different.) Although a certain change in the value of K_v can be made by altering the pole–zero location of the lead compensator, if a large increase in the value of K_v is desired, then we must alter the lead compensator to a lag–lead compensator. (See Section 7–5 for lag–lead compensation.)

Comparison of step responses of the compensated and uncompensated systems. In what follows we shall examine the unit-step responses of the compensated and uncompensated systems with MATLAB.

The closed-loop transfer function of the compensated system is

$$\frac{C(s)}{R(s)} = \frac{18.7(s + 2.9)}{s(s + 2)(s + 5.4) + 18.7(s + 2.9)}$$

$$= \frac{18.7s + 54.23}{s^3 + 7.4s^2 + 29.5s + 54.23}$$

Hence,

```
numc = [0   0     18.7   54.23]
denc = [1   7.4   29.5   54.23]
```

For the uncompensated system, the closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{4}{s^2 + 2s + 4}$$

Hence,

$$\text{num} = [0 \quad 0 \quad 4]$$
$$\text{den} = [1 \quad 2 \quad 4]$$

MATLAB Program 7–1 produces the unit-step response curves for the two systems. The resulting plot is shown in Figure 7–12.

MATLAB Program 7–1

```
% ---------- Unit-step response ----------

% ***** Unit-step responses of compensated and uncompensated
% systems *****

numc = [0   0   18.7   54.23];
denc = [1   7.4   29.5   54.23];
num = [0   0   4];
den = [1   2   4];
t = 0:0.05:5;
[c1,x1,t] = step(numc,denc,t);
[c2,x2,t] = step(num,den,t);
plot(t,c1,t,c1,'o',t,c2,t,c2,'x')
grid
title('Unit-Step Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs c1 and c2')
text(0.6,1.32,'Compensated system')
text(1.3,0.68,'Uncompensated system')
```

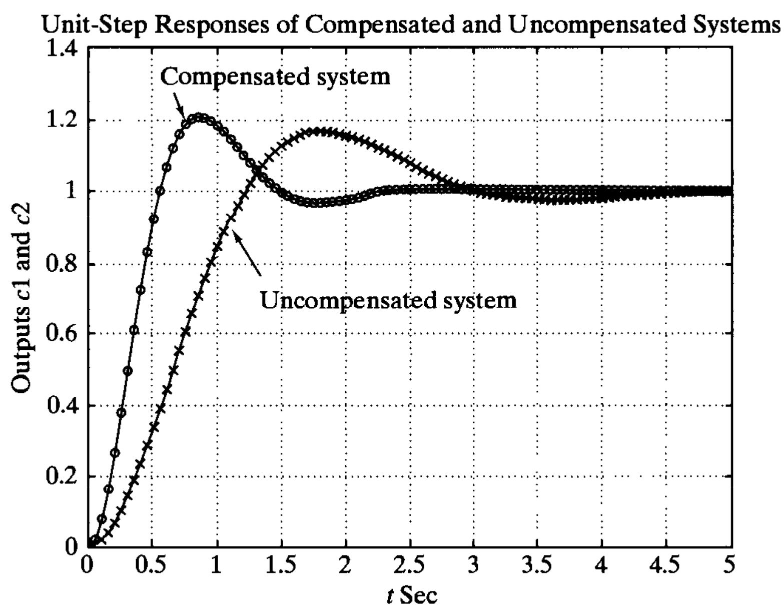


Figure 7–12 Unit-step responses of compensated and uncompensated systems.

Electronic lag compensator using operational amplifiers. The configuration of the electronic lag compensator using operational amplifiers is the same as that for the lead compensator shown in Figure 7–4. If we choose $R_2C_2 > R_1C_1$ in the circuit shown in Figure 7–4, it becomes a lag compensator. Referring to Figure 7–4, the transfer function of the lag compensator is given by

$$\frac{E_o(s)}{E_i(s)} = \hat{K}_c\beta\frac{Ts + 1}{\beta Ts + 1} = \hat{K}_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}}$$

where

$$T = R_1C_1, \qquad \beta T = R_2C_2, \qquad \beta = \frac{R_2C_2}{R_1C_1} > 1, \qquad \hat{K}_c = \frac{R_4C_1}{R_3C_2}$$

Note that we use β instead of α in the above expressions. [In the lead compensator we used α to indicate the ratio $R_2C_2/(R_1C_1)$, which was less than 1, or $0 < \alpha < 1$.] In this chapter we always assume that $0 < \alpha < 1$ and $\beta > 1$.

Lag compensation techniques based on the root–locus approach. Consider the problem of finding a suitable compensation network for the case where the system exhibits satisfactory transient-response characteristics but unsatisfactory steady-state characteristics. Compensation in this case essentially consists of increasing the open-loop gain without appreciably changing the transient-response characteristics. This means that the root locus in the neighborhood of the dominant closed-loop poles should not be changed appreciably, but the open-loop gain should be increased as much as needed. This can be accomplished if a lag compensator is put in cascade with the given feedforward transfer function.

To avoid an appreciable change in the root loci, the angle contribution of the lag network should be limited to a small amount, say 5°. To assure this, we place the pole and zero of the lag network relatively close together and near the origin of the s plane. Then the closed-loop poles of the compensated system will be shifted only slightly from their original locations. Hence, the transient-response characteristics will be changed only slightly.

Consider a lag compensator $G_c(s)$, where

$$G_c(s) = \hat{K}_c\beta\frac{Ts + 1}{\beta Ts + 1} = \hat{K}_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}} \tag{7–2}$$

If we place the zero and pole of the lag compensator very close to each other, then at $s = s_1$, where s_1 is one of the dominant closed-loop poles, the magnitudes $s_1 + (1/T)$ and $s_1 + [1/(\beta T)]$ are almost equal, or

$$|G_c(s_1)| = \left| \hat{K}_c \frac{s_1 + \dfrac{1}{T}}{s_1 + \dfrac{1}{\beta T}} \right| \doteq \hat{K}_c$$

This implies that if gain $\hat{K}_c$ of the lag compensator is set equal to 1 then the transient-response characteristics will not be altered. (This means that the overall gain of the open-loop transfer function can be increased by a factor of β where $\beta > 1$.) If the pole and zero are placed very close to the origin, then the value of β can be made large. (A large value of β may be used, provided physical realization of the lag compensator is possible.) It is noted that the value of T must be large, but its exact value is not critical. However, it should not be too large in order to avoid difficulties in realizing the phase lag compensator by physical components.

An increase in the gain means an increase in the static error constants. If the open-loop transfer function of the uncompensated system is $G(s)$, then the static velocity error constant K_v of the uncompensated system is

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

If the compensator is chosen as given by Equation (7–2), then for the compensated system with the open-loop transfer function $G_c(s)G(s)$ the static velocity error constant $\hat{K}_v$ becomes

$$\begin{aligned} \hat{K}_v &= \lim_{s \to 0} sG_c(s)G(s) \\ &= \lim_{s \to 0} G_c(s)K_v \\ &= \hat{K}_c \beta K_v \end{aligned}$$

Thus if the compensator is given by Equation (7–2), then the static velocity error constant is increased by a factor of $\hat{K}_c\beta$, where $\hat{K}_c$ is approximately unity.

Design procedures for lag compensation by the root-locus method. The procedure for designing lag compensators for the system shown in Figure 7–13 by the root-locus method may be stated as follows (we assume that the uncompensated system meets the transient-response specifications by simple gain adjustment; if this is not the case, refer to Section 7–5):

1. Draw the root-locus plot for the uncompensated system whose open-loop transfer function is $G(s)$. Based on the transient-response specifications, locate the dominant closed-loop poles on the root locus.

2. Assume the transfer function of the lag compensator to be

$$G_c(s) = \hat{K}_c \beta \frac{Ts + 1}{\beta Ts + 1} = \hat{K}_c \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}}$$

Then the open-loop transfer function of the compensated system becomes $G_c(s)G(s)$.

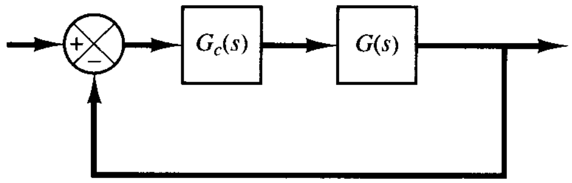


Figure 7–13
Control system.

3. Evaluate the particular static error constant specified in the problem.

4. Determine the amount of increase in the static error constant necessary to satisfy the specifications.

5. Determine the pole and zero of the lag compensator that produce the necessary increase in the particular static error constant without appreciably altering the original root loci. (Note that the ratio of the value of gain required in the specifications and the gain found in the uncompensated system is the required ratio between the distance of the zero from the origin and that of the pole from the origin.)

6. Draw a new root-locus plot for the compensated system. Locate the desired dominant closed-loop poles on the root locus. (If the angle contribution of the lag network is very small, that is, a few degrees, then the original and new root loci are almost identical. Otherwise, there will be a slight discrepancy between them. Then locate, on the new root locus, the desired dominant closed-loop poles based on the transient-response specifications.)

7. Adjust gain $\hat{K}_c$ of the compensator from the magnitude condition so that the dominant closed-loop poles lie at the desired location.

EXAMPLE 7–2 Consider the system shown in Figure 7–14(a). The feedforward transfer function is

$$G(s) = \frac{1.06}{s(s+1)(s+2)}$$

The root-locus plot for the system is shown in Figure 7–14(b). The closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{1.06}{s(s+1)(s+2) + 1.06}$$

$$= \frac{1.06}{(s + 0.3307 - j0.5864)(s + 0.3307 + j0.5864)(s + 2.3386)}$$

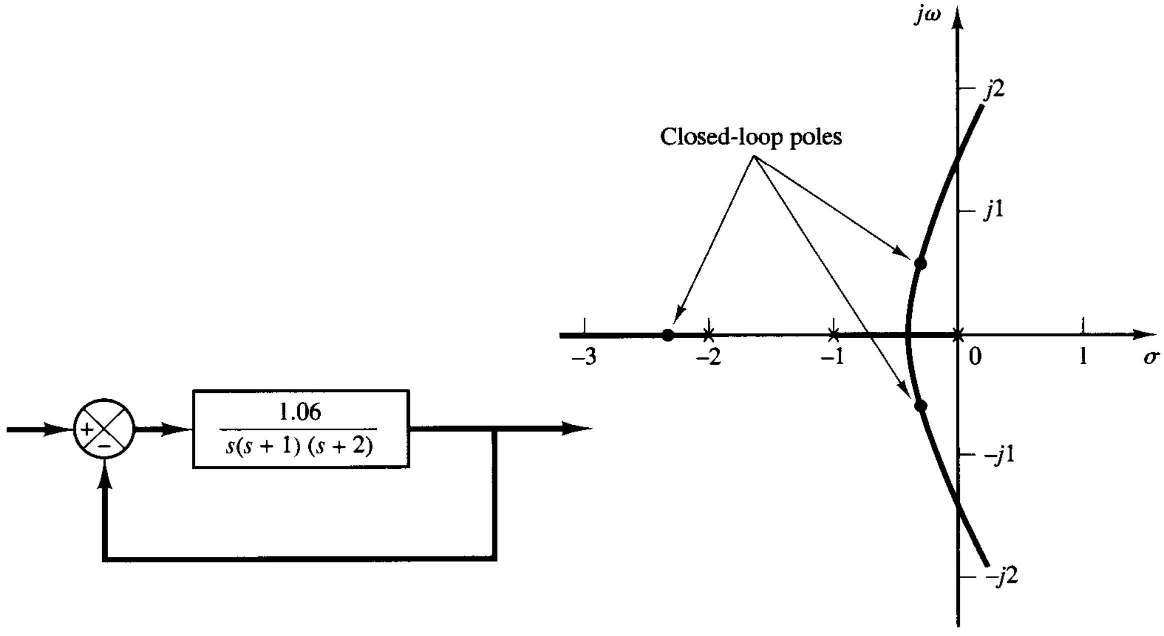
The dominant closed-loop poles are

$$s = -0.3307 \pm j0.5864$$

The damping ratio of the dominant closed-loop poles is $\zeta = 0.491$. The undamped natural frequency of the dominant closed-loop poles is 0.673 rad/sec. The static velocity error constant is 0.53 $\sec^{-1}$.

It is desired to increase the static velocity error constant K_v to about 5 $\sec^{-1}$ without appreciably changing the location of the dominant closed-loop poles.

To meet this specification, let us insert a lag compensator as given by Equation (7–2) in cascade with the given feedforward transfer function. To increase the static velocity error constant



(a)   (b)

Figure 7–14
(a) Control system;
(b) root-locus plot.

by a factor of about 10, let us choose $\beta = 10$ and place the zero and pole of the lag compensator at $s = -0.05$ and $s = -0.005$, respectively. The transfer function of the lag compensator becomes

$$G_c(s) = \hat{K}_c \frac{s + 0.05}{s + 0.005}$$

The angle contribution of this lag network near a dominant closed-loop pole is about 4°. Because this angle contribution is not very small, there is a small change in the new root locus near the desired dominant closed-loop poles.

The open-loop transfer function of the compensated system then becomes

$$G_c(s)G(s) = \hat{K}_c \frac{s + 0.05}{s + 0.005} \frac{1.06}{s(s + 1)(s + 2)}$$
$$= \frac{K(s + 0.05)}{s(s + 0.005)(s + 1)(s + 2)}$$

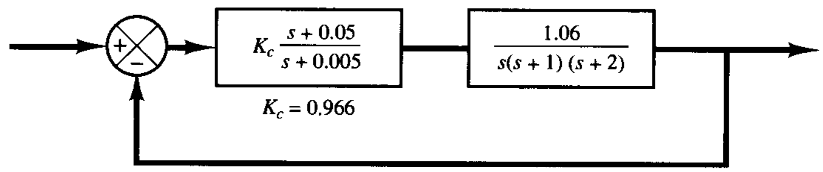
where

$$K = 1.06\hat{K}_c$$

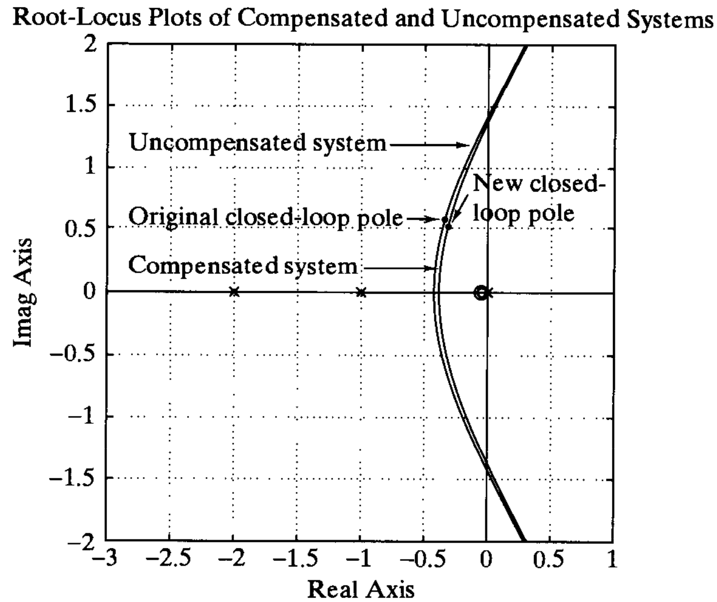
The block diagram of the compensated system is shown in Figure 7–15(a). The root-locus plot for the compensated system near the dominant closed-loop poles is shown in Figure 7–15(b), together with the original root-locus plot. Figure 7–15(c) shows the root-locus plot of the compensated system near the origin. The MATLAB program to generate the root-locus plots shown in Figures 7–15(b) and (c) is given in MATLAB Program 7–2.

If the damping ratio of the new dominant closed-loop poles is kept the same, then the poles are obtained from the new root-locus plot as follows:

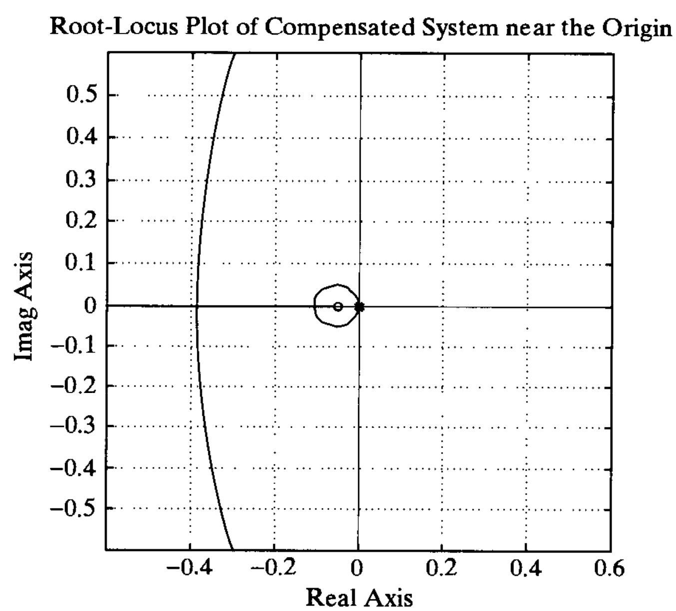
$$s_1 = -0.31 + j0.55, \qquad s_2 = -0.31 - j0.55$$



(a)



(b)



(c)

Figure 7–15
(a) Compensated system; (b) root-locus plots of the compensated system and the uncompensated system; (c) root-locus plot of compensated system near the origin.

MATLAB Program 7–2

```
% ---------- Root Locus Plot ---------

% ***** Root locus plots of the compensated system and
% uncompensated system *****

% ***** Enter the numerators and denominators of the
% compensated and uncompensated systems *****

numc = [0   0   0   1   0.05];
denc = [1   3.005   2.015   0.01   0];
num = [0   0   0   1.06];
den = [1   3   2   0];

% ***** Enter rlocus command. Plot the root loci of both
% systems *****

rlocus(numc,denc)
hold
Current plot held
rlocus(num,den)
v = [-3   1   -2   2]; axis(v); axis('square')
grid
text(-2,8,0.2,'Compensated system')
text(-2.8,1.2,'Uncompensated system')
text(-2.8,0.58,'Original closed-loop pole')
text(-0.1,0.85,'New closed-')
text(-0.1,0.62,'loop pole')
title('Root-Locus Plots of Compensated and Uncompensated Systems')

hold
Current plot released

% ***** Plot root loci of the compensated system near the origin *****

rlocus(numc,denc)
v = [-0.6   0.6   -0.6   0.6]; axis(v); axis('square')
grid
title('Root-Locus Plot of Compensated System near the Origin')
```

The open-loop gain K is

$$K = \left| \frac{s(s + 0.005)(s + 1)(s + 2)}{s + 0.05} \right|_{s=-0.31+j0.55}$$
$$= 1.0235$$

Then the lag compensator gain $\hat{K}_c$ is determined as

$$\hat{K}_c = \frac{K}{1.06} = \frac{1.0235}{1.06} = 0.9656$$

Thus the transfer function of the lag compensator designed is

$$G_c(s) = 0.9656 \frac{s + 0.05}{s + 0.005} = 9.656 \frac{20s + 1}{200s + 1}$$

Then the compensated system has the following open-loop transfer function:

$$G_1(s) = \frac{1.0235(s + 0.05)}{s(s + 0.005)(s + 1)(s + 2)}$$
$$= \frac{5.12(20s + 1)}{s(200s + 1)(s + 1)(0.5s + 1)}$$

The static velocity error constant K_v is

$$K_v = \lim_{s \to 0} sG_1(s) = 5.12 \text{ sec}^{-1}$$

In the compensated system, the static velocity error constant has increased to 5.12 sec^{-1}, or 5.12/0.53 = 9.66 times the original value. (The steady-state error with ramp inputs has decreased to about 10% of that of the original system.) We have essentially accomplished the design objective of increasing the static velocity error constant to about 5 sec^{-1}.

Note that, since the pole and zero of the lag compensator are placed close together and are located very near the origin, their effect on the shape of the original root loci has been small. Except for the presence of a small closed root locus near the origin, the root loci of the compensated and the uncompensated systems are very similar to each other. However, the value of the static velocity error constant of the compensated system is 9.66 times greater than that of the uncompensated system.

The two other closed-loop poles for the compensated system are found as follows:

$$s_3 = -2.326, \qquad s_4 = -0.0549$$

The addition of the lag compensator increases the order of the system from 3 to 4, adding one additional closed-loop pole close to the zero of the lag compensator. (The added closed-loop pole at $s = -0.0549$ is close to the zero at $s = -0.05$.) Such a pair of a zero and pole creates a long tail of small amplitude in the transient response, as we will see later in the unit-step response. Since the pole at $s = -2.326$ is very far from the $j\omega$ axis compared with the dominant closed-loop poles, the effect of this pole on the transient response is also small. Therefore, we may consider the closed-loop poles at $s = -0.31 \pm j0.55$ to be the dominant closed-loop poles.

The undamped natural frequency of the dominant closed-loop poles of the compensated system is 0.631 rad/sec. This value is about 6% less than the original value, 0.673 rad/sec. This implies that the transient response of the compensated system is slower than that of the original system. The response will take a longer time to settle down. The maximum overshoot in the step response will increase in the compensated system. If such adverse effects can be tolerated, the lag compensation as discussed here presents a satisfactory solution to the given design problem.

Next, we shall compare the unit-ramp responses of the compensated system against the uncompensated system and verify that the steady-state performance is much better in the compensated system than the uncompensated system.

To obtain the unit-ramp response with MATLAB, we use the step command for the system $C(s)/[sR(s)]$. Since $C(s)/[sR(s)]$ for the compensated system is

$$\frac{C(s)}{sR(s)} = \frac{1.0235(s + 0.05)}{s[s(s + 0.005)(s + 1)(s + 2) + 1.0235(s + 0.05)]}$$
$$= \frac{1.0235s + 0.0512}{s^5 + 3.005s^4 + 2.015s^3 + 1.0335s^2 + 0.0512s}$$

we have

```
numc = [0   0   0   0   1.0235   0.0512]
denc = [1   3.005   2.015   1.0335   0.0512   0]
```

Also, $C(s)/[sR(s)]$ for the uncompensated system is

$$\frac{C(s)}{sR(s)} = \frac{1.06}{s[s(s+1)(s+2)+1.06]}$$
$$= \frac{1.06}{s^4 + 3s^3 + 2s^2 + 1.06s}$$

Hence,

```
num = [0   0   0   0   1.06]
den = [1   3   2   1.06   0]
```

MATLAB Program 7–3 produces the plot of the unit-ramp response curves. Figure 7–16 shows the result. Clearly, the compensated system shows much smaller steady-state error (one-tenth of the original steady-state error) in following the unit-ramp input.

MATLAB Program 7–4 gives the unit-step response curves of the compensated and uncompensated systems. The unit-step response curves are shown in Figure 7–17. Notice that the

MATLAB Program 7–3

```
% ---------- Unit ramp response ----------

% ***** Unit-ramp responses of compensated system and
% uncompensated system *****

% ***** Unit-ramp response will be obtained as the unit-step
% response of C(s)/[sR(s)] *****

% ***** Enter the numerators and denominators of C1(s)/[sR(s)]
% and C2(s)/[sR(s)], where C1(s) and C2(s) are Laplace
% transforms of the outputs of the compensated and un-
% compensated systems, respectively. *****

numc = [0   0   0   0   1.0235   0.0512];
denc = [1   3.005   2.015   1.0335   0.0512   0];
num = [0   0   0   0   1.06];
den = [1   3   2   1.06   0];

% ***** Specify the time range (such as t = 0:0.1:50) and enter
% step command and plot command. *****

t = 0:0.1:50;
[c1,x1,t] = step(numc,denc,t);
[c2,x2,t] = step(num,den,t);
plot(t,c1,'-',t,c2,'.',t,t,'--')
grid
text(2.2,27,'Compensated system');
text(26,21.3,'Uncompensated system')
title('Unit-Ramp Responses of Compensated and Uncompensated Systems')
xlabel('t Sec');
ylabel('Outputs c1 and c2')
```

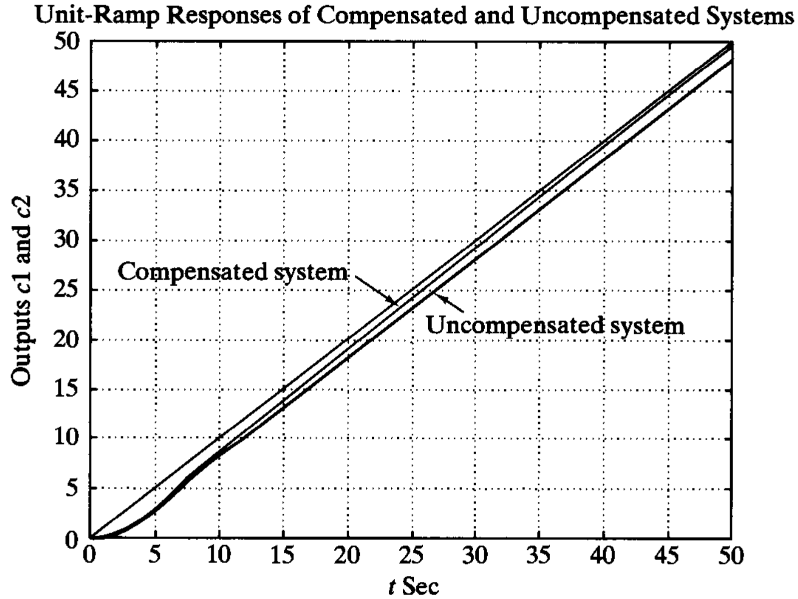


Figure 7–16
Unit-ramp responses of compensated and uncompensated systems.

MATLAB Program 7–4

```
% ---------- Unit-step response ----------

% ***** Unit-step responses of compensated system and
% uncompensated system *****

% ***** Enter the numerators and denominators of the
% compensated and uncompensated systems *****

numc = [0   0   0   1.0235   0.0512];
denc = [1   3.005   2.015   1.0335   0.0512];
num = [0   0   0   1.06];
den = [1   3   2   1.06];

% ***** Specify the time range (such as t = 0:0.1:40) and enter
% step command and plot command. *****

t = 0:0.1:40;
[c1,x1,t] = step(numc,denc,t);
[c2,x2,t] = step(num,den,t);
plot(t,c1,'-',t,c2,'.')
grid
text(13,1.12,'Compensated system')
text(13.6,0.88,'Uncompensated system')
title('Unit-Step Responses of Compensated and Uncompensated Systems')
xlabel('t Sec')
ylabel('Outputs c1 and c2')
```

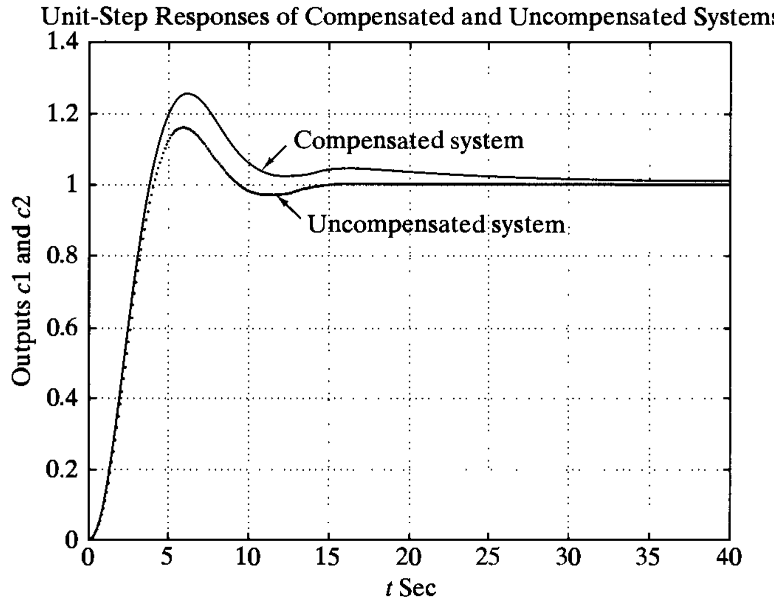


Figure 7–17 Unit-step responses of compensated and uncompensated systems.

lag-compensated system exhibits a larger maximum overshoot and slower response than the original uncompensated system. Notice that a pair of the pole at $s = -0.0549$ and zero at $s = -0.05$ generates a long tail of small amplitude in the transient response. If a larger maximum overshoot and a slower response are not desired, we need to use a lag–lead compensator as presented in Section 7–5.

7–5 LAG–LEAD COMPENSATION

Lead compensation basically speeds up the response and increases the stability of the system. Lag compensation improves the steady-state accuracy of the system, but reduces the speed of the response.

If improvements in both transient response and steady-state response are desired, then both a lead compensator and a lag compensator may be used simultaneously. Rather than introducing both a lead compensator and a lag compensator as separate elements, however, it is economical to use a single lag–lead compensator.

Lag-lead compensation combines the advantages of lag and lead compensations. Since the lag–lead compensator possesses two poles and two zeros, such a compensation increases the order of the system by 2, unless cancellation of pole(s) and zero(s) occurs in the compensated system.

Electronic lag–lead compensator using operational amplifiers. Figure 7–18 shows an electronic lag–lead compensator using operational amplifiers. The transfer function for this compensator may be obtained as follows: The complex impedance Z_1 is given by

$$\frac{1}{Z_1} = \frac{1}{R_1 + \dfrac{1}{C_1 s}} + \frac{1}{R_3}$$

or

$$Z_1 = \frac{(R_1 C_1 s + 1)R_3}{(R_1 + R_3)C_1 s + 1}$$

Similarly, complex impedance Z_2 is given by

$$Z_2 = \frac{(R_2 C_2 s + 1)R_4}{(R_2 + R_4)C_2 s + 1}$$

Hence, we have

$$\frac{E(s)}{E_i(s)} = -\frac{Z_2}{Z_1} = -\frac{R_4}{R_3}\frac{(R_1 + R_3)C_1 s + 1}{R_1 C_1 s + 1} \cdot \frac{R_2 C_2 s + 1}{(R_2 + R_4)C_2 s + 1}$$

The sign inverter has the transfer function

$$\frac{E_o(s)}{E(s)} = -\frac{R_6}{R_5}$$

Thus the transfer function of the compensator shown in Figure 7–18 is

$$\frac{E_o(s)}{E_i(s)} = \frac{E_o(s)}{E(s)}\frac{E(s)}{E_i(s)} = \frac{R_4 R_6}{R_3 R_5}\left[\frac{(R_1 + R_3)C_1 s + 1}{R_1 C_1 s + 1}\right]\left[\frac{R_2 C_2 s + 1}{(R_2 + R_4)C_2 s + 1}\right] \tag{7–3}$$

Let us define

$$T_1 = (R_1 + R_3)C_1, \qquad \frac{T_1}{\gamma} = R_1 C_1, \qquad T_2 = R_2 C_2, \qquad \beta T_2 = (R_2 + R_4)C_2$$

Then Equation (7–3) becomes

$$\frac{E_o(s)}{E_i(s)} = K_c \frac{\beta}{\gamma}\left(\frac{T_1 s + 1}{\frac{T_1}{\gamma}s + 1}\right)\left(\frac{T_2 s + 1}{\beta T_2 s + 1}\right) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\gamma}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \tag{7–4}$$

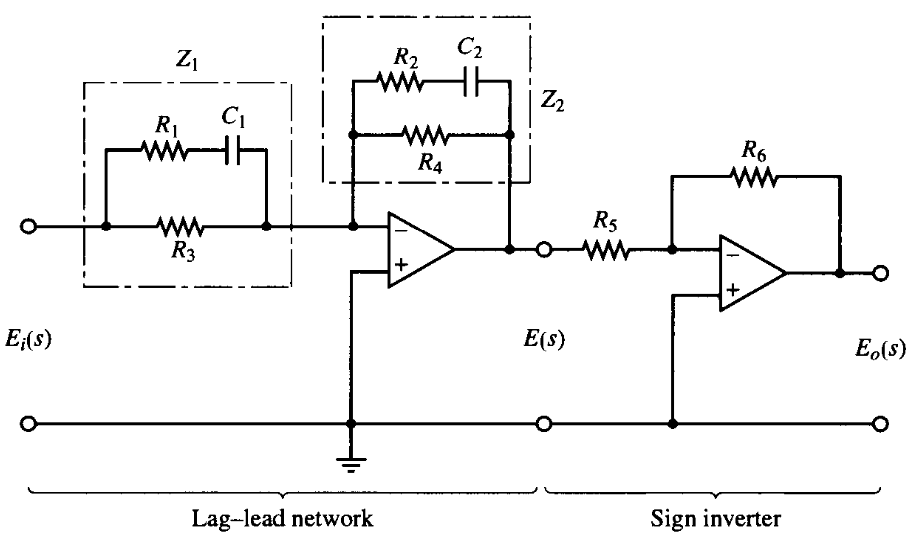


Figure 7–18
Lag-lead compensator.

where

$$\gamma = \frac{R_1 + R_3}{R_1} > 1, \qquad \beta = \frac{R_2 + R_4}{R_2} > 1, \qquad K_c = \frac{R_2 R_4 R_6}{R_1 R_3 R_5} \frac{R_1 + R_3}{R_2 + R_4}$$

Note that β is often chosen to be equal to γ.

Lag–lead compensation techniques based on the root-locus approach. Consider the system shown in Figure 7–19. Assume that we use the lag–lead compensator:

$$G_c(s) = K_c \frac{\beta}{\gamma} \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\gamma} s + 1\right)(\beta T_2 s + 1)} = K_c \left(\frac{s + \frac{1}{T_1}}{s + \frac{\gamma}{T_1}} \right) \left(\frac{s + \frac{1}{T_2}}{s + \frac{1}{\beta T_2}} \right) \tag{7–5}$$

where $\beta > 1$ and $\gamma > 1$. (Consider K_c to belong to the lead portion of the lag–lead compensator.)

In designing lag–lead compensators, we consider two cases where $\gamma \neq \beta$ and $\gamma = \beta$.

Case 1. $\gamma \neq \beta$. In this case, the design process is a combination of the design of the lead compensator and that of the lag compensator. The design procedure for the lag–lead compensator follows:

1. From the given performance specifications, determine the desired location for the dominant closed-loop poles.
2. Using the uncompensated open-loop transfer function $G(s)$, determine the angle deficiency ϕ if the dominant closed-loop poles are to be at the desired location. The phase-lead portion of the lag–lead compensator must contribute this angle ϕ.
3. Assuming that we later choose T_2 sufficiently large so that the magnitude of the lag portion

$$\left| \frac{s_1 + \frac{1}{T_2}}{s_1 + \frac{1}{\beta T_2}} \right|$$

is approximately unity, where $s = s_1$ is one of the dominant closed-loop poles, choose the values of T_1 and γ from the requirement that

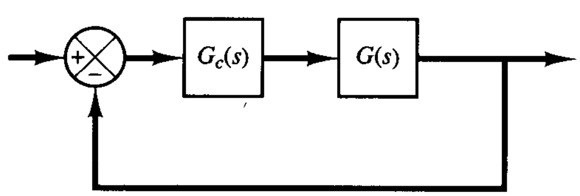


Figure 7–19
Control system.

$$\angle \frac{s_1 + \dfrac{1}{T_1}}{s_1 + \dfrac{\gamma}{T_1}} = \phi$$

The choice of T_1 and γ is not unique. (Infinitely many sets of T_1 and γ are possible.) Then determine the value of K_c from the magnitude condition:

$$\left| K_c \frac{s_1 + \dfrac{1}{T_1}}{s_1 + \dfrac{\gamma}{T_1}} G(s_1) \right| = 1$$

4. If the static velocity error constant K_v is specified, determine the value of β to satisfy the requirement for K_v. The static velocity error constant K_v is given by

$$\begin{aligned} K_v &= \lim_{s\to 0} sG_c(s)G(s) \\ &= \lim_{s\to 0} sK_c \left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\gamma}{T_1}} \right) \left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}} \right) G(s) \\ &= \lim_{s\to 0} sK_c \frac{\beta}{\gamma} G(s) \end{aligned}$$

where K_c and γ are aleady determined in step 3. Hence, given the value of K_v, the value of β can be determined from this last equation. Then, using the value of β thus determined, choose the value of T_2 such that

$$\left| \frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{1}{\beta T_2}} \right| \doteq 1$$

$$-5^\circ < \angle \frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{1}{\beta T_2}} < 0^\circ$$

(The preceding design procedure is illustrated in Example 7–3.)

Case 2. $\gamma = \beta$. If $\gamma = \beta$ is required in Equation (7–5), then the preceeding design procedure for the lag–lead compensator may be modified as follows:

1. From the given performance specifications, determine the desired location for the dominant closed-loop poles.

2. The lag–lead compensator given by Equation (7–5) is modified to

$$G_c(s) = K_c \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta}s + 1\right)(\beta T_2 s + 1)} = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \tag{7–6}$$

where $\beta > 1$. The open-loop transfer function of the compensated system is $G_c(s)G(s)$. If the static velocity error constant K_v is specified, determine the value of constant K_c from the following equation:

$$K_v = \lim_{s \to 0} sG_c(s)G(s)$$
$$= \lim_{s \to 0} sK_cG(s)$$

3. To have the dominant closed-loop poles at the desired location, calculate the angle contribution ϕ needed from the phase lead portion of the lag–lead compensator.

4. For the lag–lead compensator, we later choose T_2 sufficiently large so that

$$\left| \frac{s_1 + \frac{1}{T_2}}{s_1 + \frac{1}{\beta T_2}} \right|$$

is approximately unity, where $s = s_1$ is one of the dominant closed-loop poles. Determine the values of T_1 and β from the magnitude and angle conditions:

$$\left| K_c \left(\frac{s_1 + \frac{1}{T_1}}{s_1 + \frac{\beta}{T_1}} \right) G(s_1) \right| = 1$$

$$\angle \frac{s_1 + \frac{1}{T_1}}{s_1 + \frac{\beta}{T_1}} = \phi$$

5. Using the value of β just determined, choose T_2 so that

$$\left| \frac{s_1 + \frac{1}{T_2}}{s_1 + \frac{1}{\beta T_2}} \right| \doteqdot 1$$

$$-5^\circ < \angle \frac{s_1 + \frac{1}{T_2}}{S_1 + \frac{1}{\beta T_2}} < 0^\circ$$

The value of βT_2, the largest time constant of the lag–lead compensator, should not be too large to be physically realized. (An example of the design of the lag–lead compensator when $\gamma = \beta$ is given in Example 7–4).

EXAMPLE 7–3 Consider the control system shown in Figure 7–20. The feedforward transfer function is

$$G(s) = \frac{4}{s(s + 0.5)}$$

This system has closed-loop poles at

$$s = -0.2500 \pm j1.9843$$

The damping ratio is 0.125, the undamped natural frequency is 2 rad/sec, and the static velocity error constant is 8 sec^{-1}.

It is desired to make the damping ratio of the dominant closed-loop poles equal to 0.5 and to increase the undamped natural frequency to 5 rad/sec and the static velocity error constant to 80 sec^{-1}. Design an appropriate compensator to meet all the performance specifications.

Let us assume that we use a lag–lead compensator having the transfer function

$$G_c(s) = K_c\left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\gamma}{T_1}}\right)\left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}}\right) \qquad (\gamma > 1, \beta > 1)$$

where γ is not equal to β. Then the compensated system will have the transfer function

$$G_c(s)G(s) = K_c\left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\gamma}{T_1}}\right)\left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}}\right)G(s)$$

From the performance specifications, the dominant closed-loop poles must be at

$$s = -2.50 \pm j4.33$$

Since

$$\left.\angle\frac{4}{s(s + 0.5)}\right|_{s=-2.50+j4.33} = -235°$$

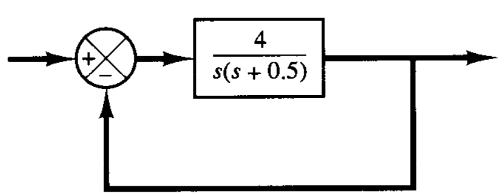


Figure 7–20
Control system.

the phase lead portion of the lag–lead compensator must contribute 55° so that the root locus passes through the desired location of the dominant closed-loop poles.

To design the phase lead portion of the compensator, we first determine the location of the zero and pole that will give 55° contribution. There are many possible choices, but we shall here choose the zero at $s = -0.5$ so that this zero will cancel the pole at $s = -0.5$ of the plant. Once the zero is chosen, the pole can be located such that the angle contribution is 55°. By simple calculation or graphical analysis, the pole must be located at $s = -5.021$. Thus, the phase lead portion of the lag–lead compensator becomes

$$K_c \frac{s + \dfrac{1}{T_1}}{s + \dfrac{\gamma}{T_1}} = K_c \frac{s + 0.5}{s + 5.021}$$

Thus

$$T_1 = 2, \qquad \gamma = \frac{5.021}{0.5} = 10.04$$

Next we determine the value of K_c from the magnitude condition:

$$\left| K_c \frac{s + 0.5}{s + 5.021} \frac{4}{s(s + 0.5)} \right|_{s=-2.5+j4.33} = 1$$

Hence,

$$K_c = \left| \frac{(s + 5.021)s}{4} \right|_{s=-2.5+j4.33} = 6.26$$

The phase lag portion of the compensator can be designed as follows: First the value of β is determined to satisfy the requirement on the static velocity error constant:

$$K_v = \lim_{s \to 0} sG_c(s)G(s) = \lim_{s \to 0} sK_c \frac{\beta}{\gamma} G(s)$$

$$= \lim_{s \to 0} s(6.26) \frac{\beta}{10.04} \frac{4}{s(s + 0.5)} = 4.988\beta = 80$$

Hence, β is determined as

$$\beta = 16.04$$

Finally, we choose the value of T_2 large enough so that

$$\left| \frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{16.04T_2}} \right|_{s=-2.5+j4.33} \doteqdot 1$$

and

$$-5^\circ < \left/ \frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{16.04T_2}} \right._{s=-2.5+j4.33} < 0^\circ$$

Since $T_2 \doteqdot 5$ (or any number greater than 5) satisfies the above two requirements, we may choose

$$T_2 = 5$$

Now the transfer function of the designed lag–lead compensator is given by

$$G_c(s) = (6.26)\left(\frac{s + \frac{1}{2}}{s + \frac{10.04}{2}}\right)\left(\frac{s + \frac{1}{5}}{s + \frac{1}{16.04 \times 5}}\right)$$

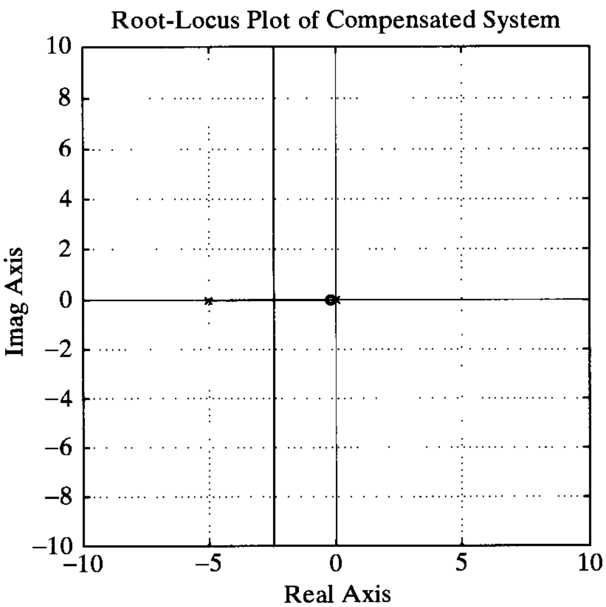
$$= 6.26\left(\frac{s + 0.5}{s + 5.02}\right)\left(\frac{s + 0.2}{s + 0.01247}\right)$$

$$= \frac{10(2s + 1)(5s + 1)}{(0.1992s + 1)(80.19s + 1)}$$

The compensated system will have the open-loop transfer function

$$G_c(s)G(s) = \frac{25.04(s + 0.2)}{s(s + 5.02)(s + 0.01247)}$$

Because of the cancellation of the $(s + 0.5)$ terms, the compensated system is a third-order system. (Mathematically, this cancellation is exact, but practically such cancellation will not be exact because some approximations are usually involved in deriving the mathematical model of the system and, as a result, the time constants are not precise.) The root-locus plot of the compensated system is shown in Figure 7–21(a). An enlarged view of the root-locus plot near the origin is shown in Figure 7–21(b). Because the angle contribution of the phase lag portion of the lag–lead compensator is quite small, there is only a small change in the location of the dominant closed-loop poles from the desired location, $s = -2.5 \pm j4.33$. In fact, the new closed-loop poles are located



(a)

Figure 7–21
(a) Root-locus plot of the compensated system; (b) root-locus plot near the origin.

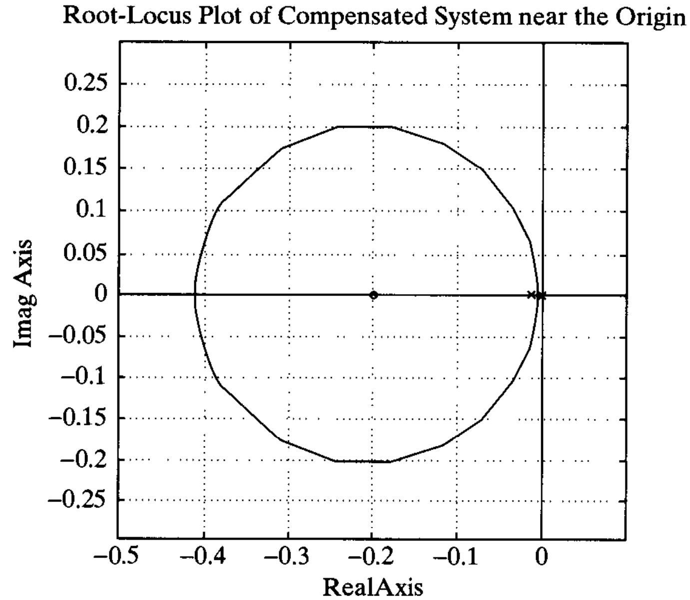
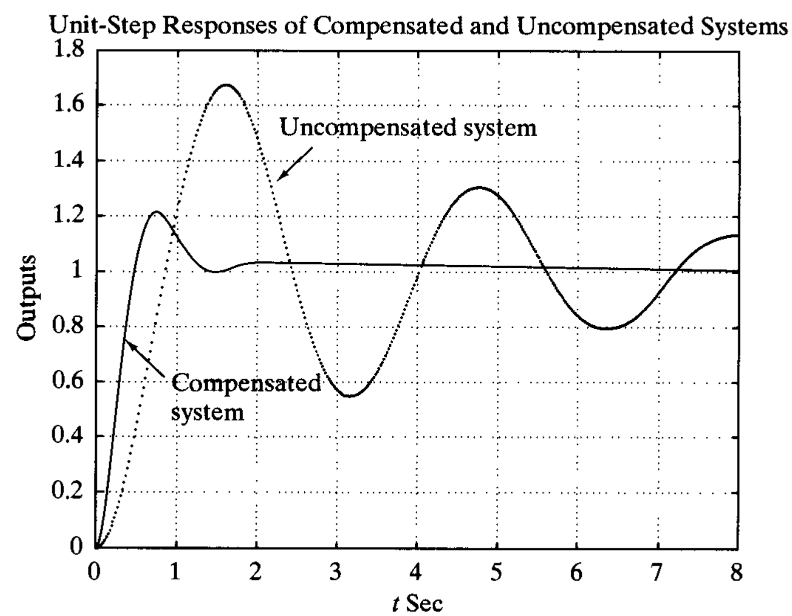


Figure 7–21
(Continued)

(b)

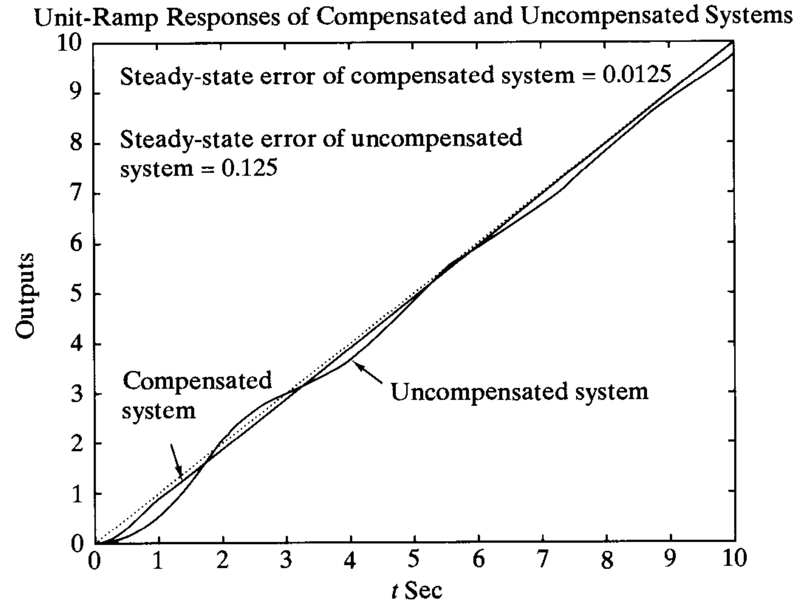
at $s = -2.4123 \pm j4.2756$. (The new damping ratio is $\zeta = 0.491$.) Therefore, the compensated system meets all the required performance specifications. The third closed-loop pole of the compensated system is located at $s = -0.2078$. Since this closed-loop pole is very close to the zero at $s = -0.2$, the effect of this pole on the response is small. (Note that, in general, if a pole and a zero lie close to each other on the negative real axis near the origin, then such a pole-zero combination will yield a long tail of small amplitude in the transient response.)

The unit-step response curves and unit-ramp response curves before and after compensation are shown in Figure 7–22.



(a)

Figure 7–22
Transient response curves for the compensated system and the uncompensated system. (a) Unit-step response curves; (b) unit-ramp response curves.



(b)

Figure 7–22 *(Continued)*

EXAMPLE 7–4 Consider the control system of Example 7–3. Suppose that we use a lag–lead compensator of the form given by Equation (7–6), or

$$G_c(s) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \qquad (\beta > 1)$$

Assuming the specifications are the same as those given in Example 7–3, design a compensator $G_c(s)$.

The desired locations for the dominant closed-loop poles are at

$$s = -2.50 \pm j4.33$$

The open-loop transfer function of the compensated system is

$$G_c(s)G(s) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \cdot \frac{4}{s(s + 0.5)}$$

Since the requirement on the static velocity error constant K_v is 80 sec^{-1}, we have

$$K_v = \lim_{s \to 0} sG_c(s)G(s) = \lim_{s \to 0} K_c \frac{4}{0.5} = 8K_c = 80$$

Thus

$$K_c = 10$$

The time constant T_1 and the value of β are determined from

$$\left| \frac{s + \frac{1}{T_1}}{s + \frac{\beta}{T_1}} \right| \left| \frac{40}{s(s + 0.5)} \right|_{s=-2.5+j4.33} = \left| \frac{s + \frac{1}{T_1}}{s + \frac{\beta}{T_1}} \right| \frac{8}{4.77} = 1$$

$$\left. \angle \frac{s + \dfrac{1}{T_1}}{s + \dfrac{\beta}{T_1}} \right|_{s = -2.5 + j4.33} = 55°$$

Referring to Figure 7–23, we can easily locate points A and B such that

$$\angle APB = 55°, \qquad \frac{\overline{PA}}{\overline{PB}} = \frac{4.77}{8}$$

(Use a graphical approach or a trigonometric approach.) The result is

$$\overline{AO} = 2.38, \qquad \overline{BO} = 8.34$$

or

$$T_1 = \frac{1}{2.38} = 0.420, \qquad \beta = 8.34T_1 = 3.503$$

The phase lead portion of the lag–lead network thus becomes

$$10\left(\frac{s + 2.38}{s + 8.34}\right)$$

For the phase lag portion, we may choose

$$T_2 = 10$$

Then

$$\frac{1}{\beta T_2} = \frac{1}{3.503 \times 10} = 0.0285$$

Thus, the lag–lead compensator becomes

$$G_c(s) = (10)\left(\frac{s + 2.38}{s + 8.34}\right)\left(\frac{s + 0.1}{s + 0.0285}\right)$$

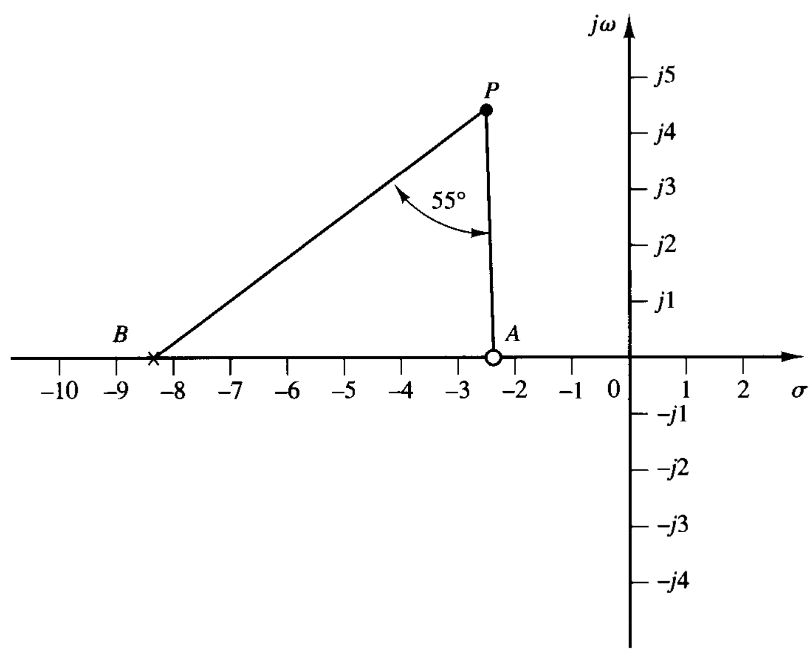


Figure 7–23 Determination of the desired pole-zero location.

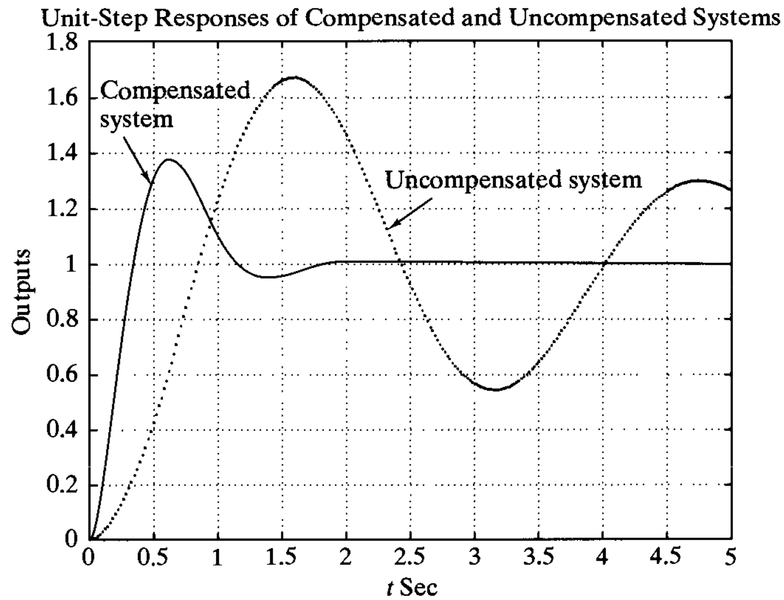
The compensated system will have the open-loop transfer function

$$G_c(s)G(s) = \frac{40(s + 2.38)(s + 0.1)}{(s + 8.34)(s + 0.0285)s(s + 0.5)}$$

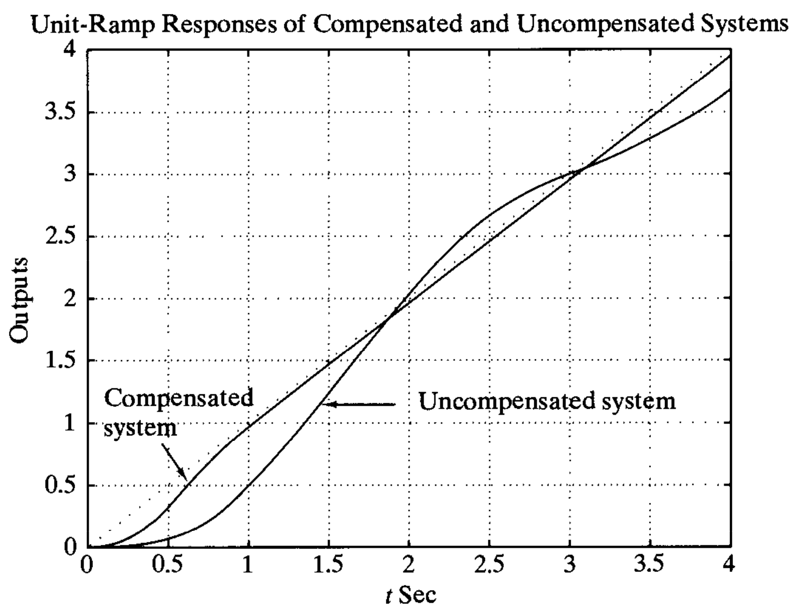
No cancellation occurs in this case, and the compensated system is of fourth order. Because the angle contribution of the phase lag portion of the lag–lead network is quite small, the dominant closed-loop poles are located very near the desired location. In fact, the dominant closed-loop poles are located at $s = -2.4539 \pm j4.3099$. The two other closed-loop poles are located at

$$s = -0.1003, \qquad s = -3.8604$$

Since the closed-loop pole at $s = -0.1003$ is very close to a zero at $s = -0.1$, they almost cancel each other. Thus, the effect of this closed-loop pole is very small. The remaining closed-loop pole



(a)



(a)

Figure 7–24
(a) Unit-step response curves for the compensated and uncompensated systems; (b) unit-ramp response curves for both systems.

($s = -3.8604$) does not quite cancel the zero at $s = -2.4$. The effect of this zero is to cause a larger overshoot in the step response than a similar system without such a zero. The unit-step response curves of the compensated and uncompensated systems are shown in Figure 7–24(a). The unit-ramp response curves for both systems are depicted in Figure 7–24(b).

EXAMPLE PROBLEMS AND SOLUTIONS

A-7-1. Obtain the transfer function of the mechanical system shown in Figure 7–25. Assume that the displacement x_i is the input and displacement x_o is the output of the system.

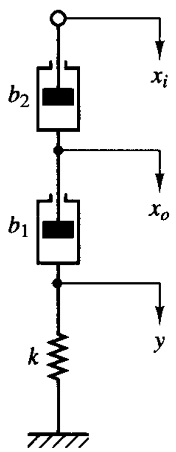


Figure 7–25 Mechanical system.

Solution. From the diagram we obtain the following equations of motion:

$$b_2(\dot{x}_i - \dot{x}_o) = b_1(\dot{x}_o - \dot{y})$$

$$b_1(\dot{x}_o - \dot{y}) = ky$$

Taking the Laplace transforms of these two equations, assuming zero initial conditions, and then eliminating $Y(s)$, we obtain

$$\frac{X_o(s)}{X_i(s)} = \frac{b_2}{b_1 + b_2} \frac{\frac{b_1}{k}s + 1}{\frac{b_2}{b_1 + b_2}\frac{b_1}{k}s + 1}$$

This is the transfer function between $X_o(s)$ and $X_i(s)$. By defining

$$\frac{b_1}{k} = T, \qquad \frac{b_2}{b_1 + b_2} = \alpha < 1$$

we obtain

$$\frac{X_o(s)}{X_i(s)} = \alpha\frac{Ts + 1}{\alpha Ts + 1} = \frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}}$$

This mechanical system is a mechanical lead network.

A-7-2. Obtain the transfer function of the mechanical system shown in Figure 7–26. Assume that the displacement x_i is the input and displacement x_o is the output.

Solntion. The equations of motion for this system are

$$b_2(\dot{x}_i - \dot{x}_o) + k_2(x_i - x_o) = b_1(\dot{x}_o - \dot{y})$$

$$b_1(\dot{x}_o - \dot{y}) = ky$$

By taking the Laplace transforms of these two equations, assuming zero initial conditions, we obtain

$$b_2[sX_i(s) - sX_o(s)] + k_2[X_i(s) - X_o(s)] = b_1[sX_o(s) - sY(s)]$$

$$b_1[sX_o(s) - sY(s)] = k_1Y(s)$$

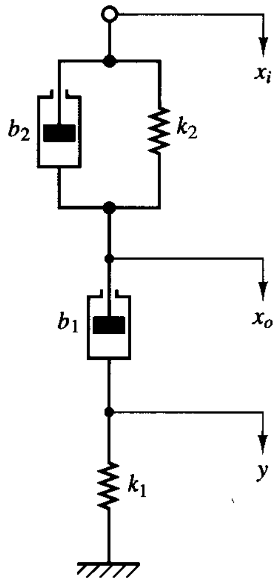


Figure 7–26
Mechanical system.

If we eliminate $Y(s)$ from the last two equations, the transfer function $X_o(s)/X_i(s)$ can be obtained as

$$\frac{X_o(s)}{X_i(s)} = \frac{\left(\frac{b_1}{k_1}s + 1\right)\left(\frac{b_2}{k_2}s + 1\right)}{\left(\frac{b_1}{k_1}s + 1\right)\left(\frac{b_2}{k_2}s + 1\right) + \frac{b_1}{k_2}s}$$

Define

$$T_1 = \frac{b_1}{k_1}, \qquad T_2 = \frac{b_2}{k_2}, \qquad \frac{b_1}{k_1} + \frac{b_2}{k_2} + \frac{b_1}{k_2} = \frac{T_1}{\beta} + \beta T_2 \qquad (\beta > 1)$$

Then $X_o(s)/X_i(s)$ can be simplified as

$$\frac{X_o(s)}{X_i(s)} = \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta}s + 1\right)(\beta T_2 s + 1)} = \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)}$$

From this transfer function we see that this mechanical system is a mechanical lag–lead network.

A-7-3. Consider the electrical network shown in Figure 7–27. Derive the transfer function of the network. (As usual in the derivation of the transfer function of any four-terminal network, we

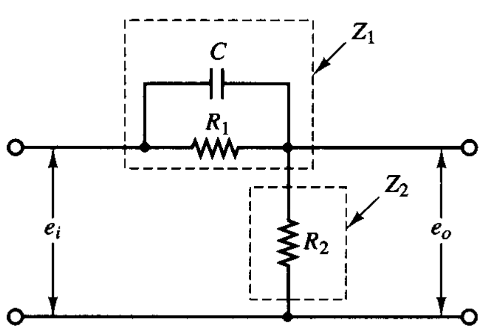


Figure 7–27
Electrical network.

assume that the source impedance that the network sees is zero and that the output load impedance is infinite.)

Solution. Using the symbols defined in Figure 7–27, we find that the complex impedances Z_1 and Z_2 are

$$Z_1 = \frac{R_1}{R_1Cs + 1}, \qquad Z_2 = R_2$$

The transfer function between the output $E_o(s)$ and the input $E_i(s)$ is

$$\frac{E_o(s)}{E_i(s)} = \frac{Z_2}{Z_1 + Z_2} = \frac{R_2}{R_1 + R_2}\frac{R_1Cs + 1}{\dfrac{R_1R_2}{R_1 + R_2}Cs + 1}$$

Define

$$R_1C = T, \qquad \frac{R_2}{R_1 + R_2} = \alpha < 1$$

Then the transfer function becomes

$$\frac{E_o(s)}{E_i(s)} = \alpha\frac{Ts + 1}{\alpha Ts + 1} = \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\alpha T}}$$

Since α is less than 1, this network is a lead network.

A-7-4. Obtain the transfer function of the network shown in Figure 7–28.

Solution. The complex impedances Z_1 and Z_2 are

$$Z_1 = \frac{R_1}{R_1C_1s + 1}, \qquad Z_2 = R_2 + \frac{1}{C_2s}$$

The transfer function between $E_o(s)$ and $E_i(s)$ is

$$\frac{E_o(s)}{E_i(s)} = \frac{Z_2}{Z_1 + Z_2} = \frac{(R_1C_1s + 1)(R_2C_2s + 1)}{(R_1C_1s + 1)(R_2C_2s + 1) + R_1C_2s}$$

The denominator of this transfer function can be factored into two real terms. Let us define

$$R_1C_1 = T_1, \qquad R_2C_2 = T_2, \qquad R_1C_1 + R_2C_2 + R_1C_2 = \frac{T_1}{\beta} + \beta T_2 \qquad (\beta > 1)$$

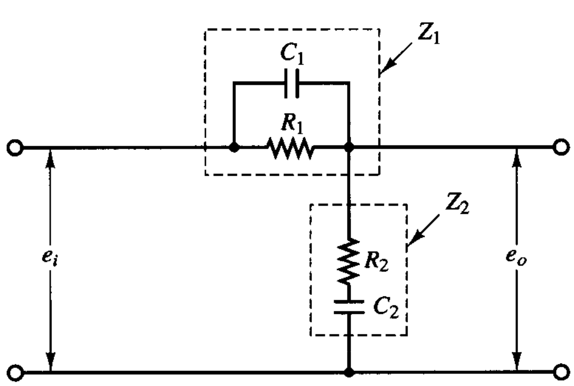


Figure 7–28
Electrical network.

Then $E_o(s)/E_i(s)$ can be simplified to

$$\frac{E_o(s)}{E_i(s)} = \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta} s + 1\right)\left(\beta T_2 s + 1\right)} = \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)}$$

This is a lag–lead network.

A-7-5. A control system with

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

is unstable for all positive values of gain K.

Plot the root loci of the system. By using this plot, show that this system can be stabilized by adding a zero on the negative real axis or by modifying $G(s)$ to $G_1(s)$, where

$$G_1(s) = \frac{K(s+a)}{s^2(s+1)} \qquad (0 \le a < 1)$$

Solution. A root-locus plot for the system with

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

is shown in Figure 7–29(a). Since two branches lie in the right half-plane, the system is unstable for any value of $K > 0$.

Addition of a zero to the transfer function $G(s)$ bends the right half-plane branches to the left and brings all root-locus branches to the left half-plane, as shown in the root-locus plot in Figure 7–29(b). Thus, the system with

$$G_1 = \frac{K(s+a)}{s^2(s+1)}, \qquad H(s) = 1 \qquad (0 \le a < 1)$$

is stable for all $K > 0$.

A-7-6. Consider a system with an unstable plant as shown in Figure 7–30a. Using the root-locus approach, design a proportional-plus-derivative controller (that is, determine the values of K_p and T_d) such that the damping ratio ζ of the closed-loop system is 0.7 and the undamped natural frequency ω_n is 0.5 rad/sec.

Solntion. Note that the open-loop transfer function involves two poles at $s = 1.085$ and $s = -1.085$ and one zero at $s = -1/T_d$, which is unknown at this point.

Since the desired closed-loop poles must have $\omega_n = 0.5$ rad/sec and $\zeta = 0.7$, they must be located at

$$s = 0.5 \underline{/180^\circ \pm 45.573^\circ}$$

($\zeta = 0.7$ corresponds to a line having an angle of 45.573° with the negative real axis.) Hence, the desired closed-loop poles are at

$$s = -0.35 \pm j0.357$$

The open-loop poles and the desired closed-loop pole in the upper half-plane are located in the diagram shown in Figure 7–30b. The angle deficiency at point $s = -0.35 + j0.357$ is

$$-166.026^\circ - 25.913^\circ + 180^\circ = -11.938^\circ$$

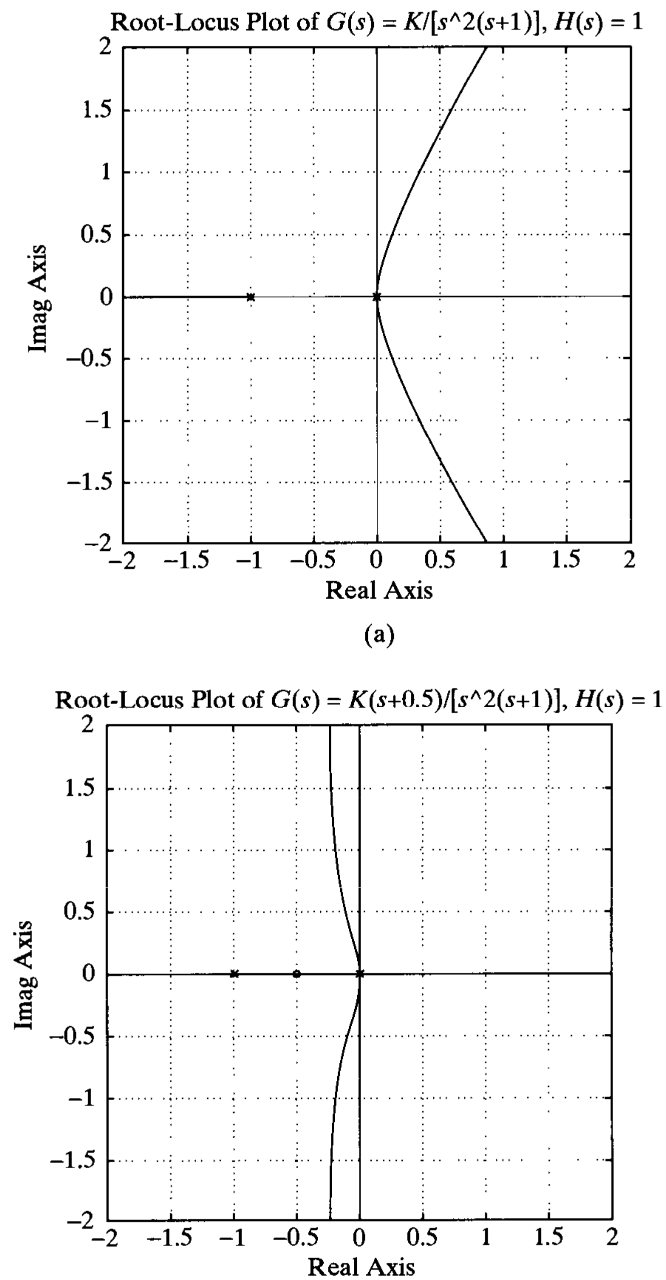


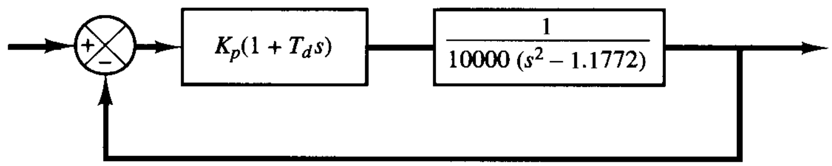
Figure 7–29
(a) Root-locus plot of the system with $G(s) = K/[s^2(s+1)]$ and $H(s) = 1$;
(b) root-locus plot of the system with $G_1(s) = K(s+a)/[s^2(s+1)]$ and $H(s) = 1$, where $a = 0.5$.

This means that the zero at $s = -1/T_d$ must contribute 11.938°, which, in turn, determines the location of the zero as follows:

$$s = -\frac{1}{T_d} = -2.039$$

Hence, we have

$$K_p(1 + T_d s) = K_p T_d\left(\frac{1}{T_d} + s\right) = K_p T_d(s + 2.039) \tag{7–7}$$



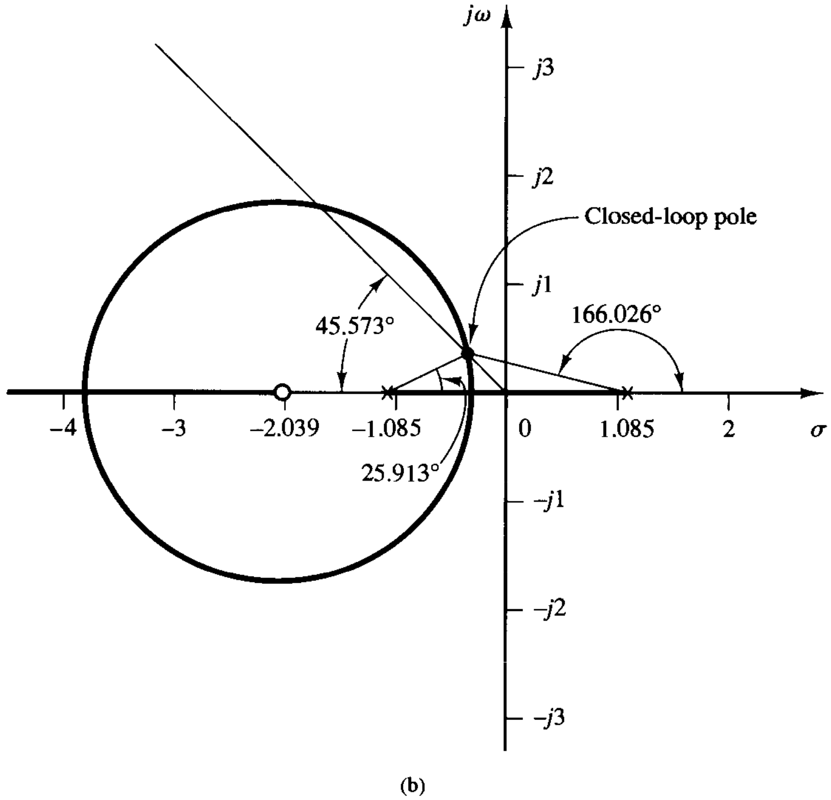


Figure 7–30
(a) PD control of an unstable plant; (b) root-locus diagram for the system.

The value of T_d is

$$T_d = \frac{1}{2.039} = 0.4904$$

The value of gain K_p can be determined from the magnitude condition as follows:

$$\left| K_p T_d \frac{s + 2.039}{10000(s^2 - 1.1772)} \right|_{s=-0.35+j0.357} = 1$$

or

$$K_p T_d = 6999.5$$

Hence,

$$K_p = \frac{6999.5}{0.4904} = 14{,}273$$

By substituting the numerical values of T_d and K_p into Equation (7–7), we obtain

$$K_p(1 + T_d s) = 14{,}273(1 + 0.4904s) = 6999.5(s + 2.039)$$

which gives the transfer function of the desired proportional-plus-derivative controller.

A-7-7. Consider the control system shown in Figure 7–31. Design a lag compensator $G_c(s)$ such that the static velocity error constant K_v is 50 $\sec^{-1}$ without appreciably changing the location of the original closed-loop poles, which are at $s = -2 \pm j\sqrt{6}$.

Solution. Assume that the transfer function of the lag compensator is

$$G_c(s) = \hat{K}_c \frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}} \qquad (\beta > 1)$$

Since K_v is specified as 50 $\sec^{-1}$, we have

$$K_v = \lim_{s \to 0} sG_c(s) \frac{10}{s(s+4)} = \hat{K}_c \beta 2.5 = 50$$

Thus

$$\hat{K}_c \beta = 20$$

Now choose $\hat{K}_c = 1$. Then

$$\beta = 20$$

Choose $T = 10$. Then the lag compensator can be given by

$$G_c(s) = \frac{s + 0.1}{s + 0.005}$$

The angle contribution of the lag compensator at the closed-loop pole $s = -2 + j\sqrt{6}$ is

$$\left. \underline{/G_c(s)} \right|_{s=-2+j\sqrt{6}} = \tan^{-1}\frac{\sqrt{6}}{-1.9} - \tan^{-1}\frac{\sqrt{6}}{-1.995}$$
$$= -1.3616°$$

which is small. Thus the change in the location of the dominant closed-loop poles is very small.

The open-loop transfer function of the system becomes

$$G_c(s)G(s) = \frac{s + 0.1}{s + 0.005} \frac{10}{s(s+4)}$$

The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{10s + 1}{s^3 + 4.005s^2 + 10.02s + 1}$$

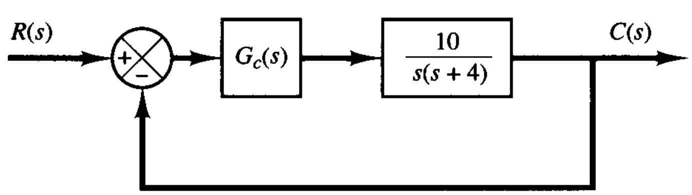


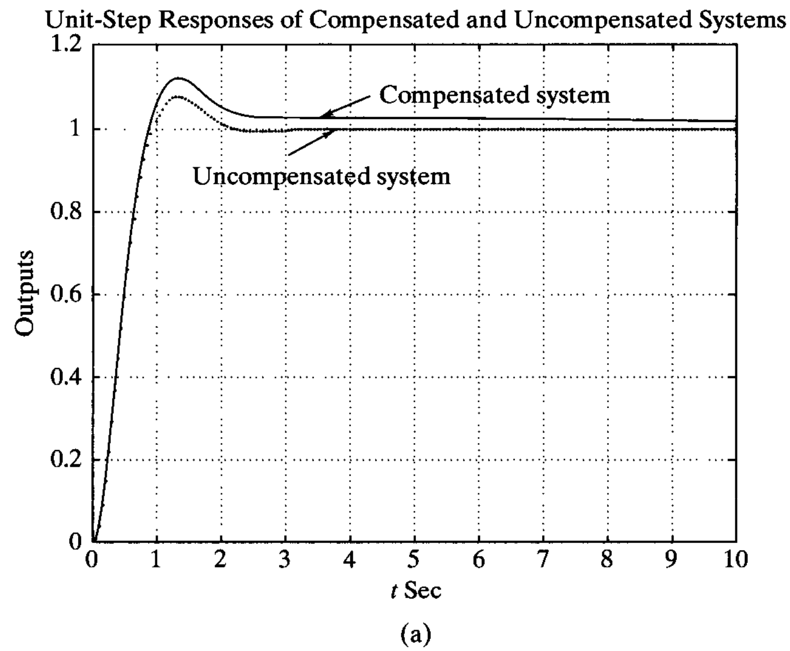
Figure 7–31
Control system.

To compare the transient-response characteristics before and after the compensation, the unit-step and unit-ramp responses of the compensated and uncompensated systems are shown in Figures 7–32(a) and (b), respectively. The steady-state error in the unit-ramp response is shown in Figure 7–32(c).

A-7-8. Consider a unity-feedback control system whose feedforward transfer function is given by

$$G(s) = \frac{10}{s(s+2)(s+8)}$$

Design a compensator such that the dominant closed-loop poles are located at $s = -2 \pm j2\sqrt{3}$ and the static velocity error constant K_v is equal to 80 sec^{-1}.



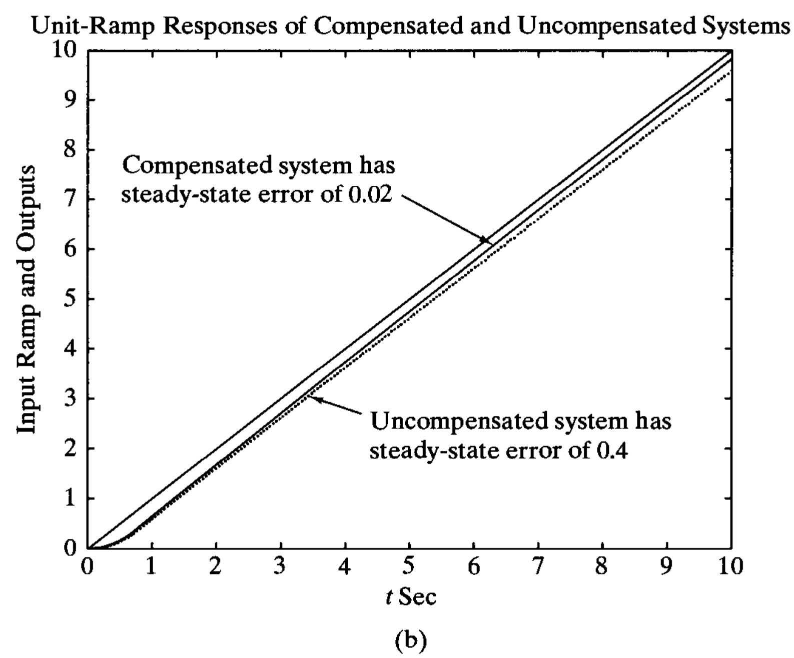
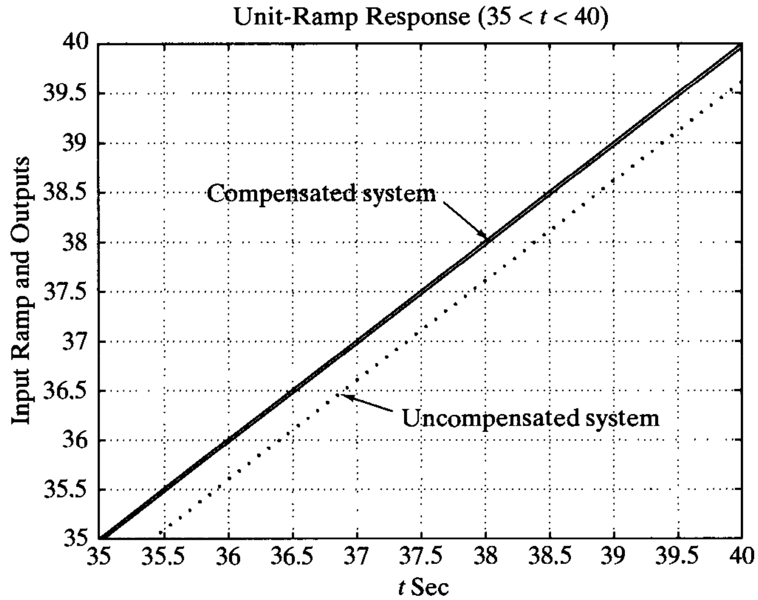


Figure 7–32
(a) Unit-step responses of the compensated and uncompensated systems; (b) unit-ramp responses of both systems; (c) unit-ramp responses showing steady-state errors.



(c)

Figure 7–32 *(Continued)*

Solution. The static velocity error constant of the uncompensated system is $K_v = \frac{10}{16} = 0.625$. Since $K_v = 80$ is required, we need to increase the open-loop gain by 128. (This implies that we need a lag compensator.) The root-locus plot of the uncompensated system reveals that it is not possible to bring the dominant closed-loop poles to $-2 \pm j2\sqrt{3}$ by just a gain adjustment alone. See Figure 7–33. (This means that we also need a lead compensator.) Therefore, we shall employ a lag–lead compensator.

Let us assume that the transfer function of the lag–lead compensator to be

$$G_c(s) = K_c \left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\beta}{T_1}} \right) \left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}} \right) \qquad (\alpha = \beta)$$

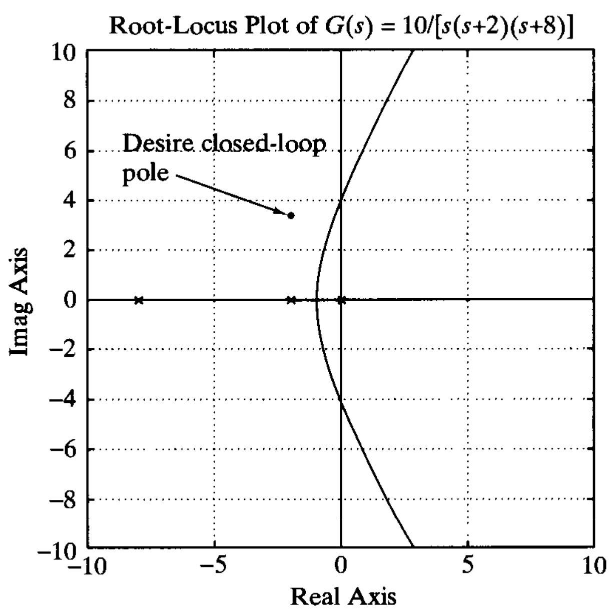


Figure 7–33 Root-locus plot of $G(s) = 10/[s(s + 2)(s + 8)]$.

where $K_c = 128$. This is because

$$K_v = \lim_{s \to 0} sG_c(s)G(s) = \lim_{s \to 0} sK_cG(s) = K_c \frac{10}{16} = 80$$

and we obtain $K_c = 128$. The angle deficiency at the desired closed-loop pole $s = -2 + j2\sqrt{3}$ is

$$\text{Angle deficiency} = 120° + 90° + 30° - 180° = 60°$$

The lead portion of the lag–lead compensator must contribute this angle. To choose T_1 we may use the graphical method presented in Section 7–5.

The lead portion must satisfy the following conditions:

$$\left| 128 \left(\frac{s_1 + \frac{1}{T_1}}{s_1 + \frac{\beta}{T_1}} \right) G(s_1) \right|_{s_1 = -2 + j2\sqrt{3}} = 1$$

and

$$\left. \angle \frac{s_1 + \frac{1}{T_1}}{s_1 + \frac{\beta}{T_1}} \right|_{s_1 = -2 + j2\sqrt{3}} = 60°$$

The first condition can be simplified as

$$\left| \frac{s_1 + \frac{1}{T_1}}{s_1 + \frac{\beta}{T_1}} \right|_{s_1 = -2 + j2\sqrt{3}} = \frac{1}{13.3333}$$

By using the same approach as used in Section 7–5, the zero ($s = 1/T_1$) and pole ($s = \beta/T_1$) can be determined as follows:

$$\frac{1}{T_1} = 3.70, \qquad \frac{\beta}{T_1} = 53.35$$

See Figure 7–34. The value of β is thus determined as

$$\beta = 14.419$$

For the lag portion of the compensator, we may choose

$$\frac{1}{\beta T_2} = 0.01$$

Then

$$\frac{1}{T_2} = 0.1442$$

Noting that

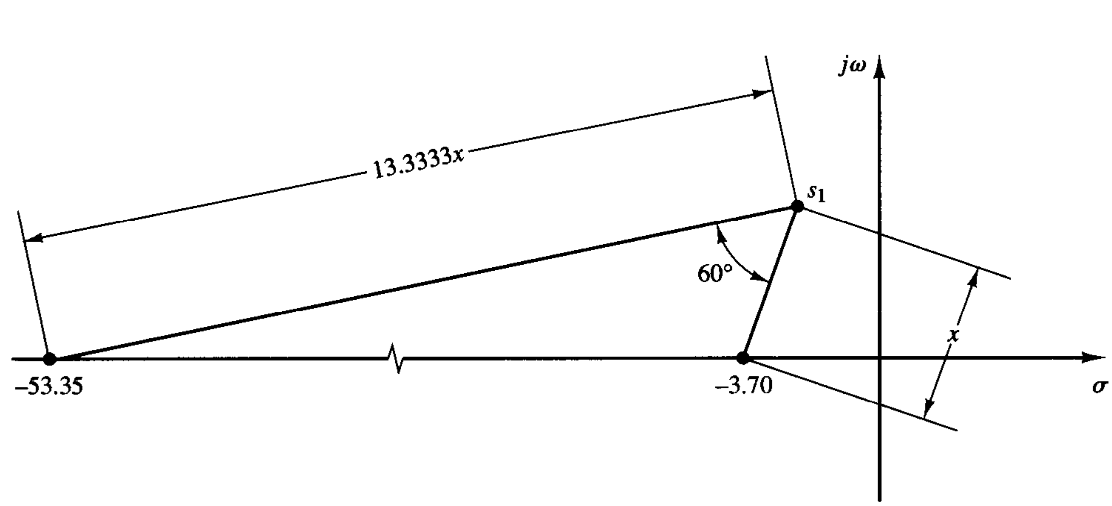


Figure 7–34 Graphical determination of the zero and pole of the lead portion of the compensator.

$$\left|\frac{s_1 + 0.1442}{s_1 + 0.01}\right|_{s_1 = -2 + j2\sqrt{3}} = 0.9837$$

$$\left/\frac{s_1 + 0.1442}{s_1 + 0.01}\right|_{s_1 = -2 + j2\sqrt{3}} = -1.697°$$

the angle contribution of the lag portion is $-1.697°$ and the magnitude contribution is 0.9837. This means that the dominant closed-loop poles lie close to the desired location $s = -2 \pm j2\sqrt{3}$. Thus the compensator designed,

$$G_c(s) = 128\left(\frac{s + 3.70}{s + 53.35}\right)\left(\frac{s + 0.1442}{s + 0.01}\right)$$

is acceptable. The feedforward transfer function of the compensated system becomes

$$G_c(s)G(s) = \frac{1280(s + 3.7)(s + 0.1442)}{s(s + 53.35)(s + 0.01)(s + 2)(s + 8)}$$

A root-locus plot of the compensated system is shown in Figure 7–35(a). An enlarged root-locus plot near the origin is shown in Figure 7–35(b).

To verify the improved system performance of the compensated system, see the unit-step responses and unit-ramp responses of the compensated and uncompensated systems shown in Figures 7–36 (a) and (b), respectively.

A-7-9. Consider the system shown in Figure 7–37. Design a lag–lead compensator such that the static velocity error constant K_v is 50 sec^{-1} and the damping ratio ζ of the dominant closed-loop poles is 0.5. (Choose the zero of the lead portion of the lag–lead compensator to cancel the pole at $s = -1$ of the plant.) Determine all closed-loop poles of the compensated system.

Solution. Let us employ the lag–lead compensator given by

$$G_c(s) = K_c\left(\frac{s + \dfrac{1}{T_1}}{s + \dfrac{\beta}{T_1}}\right)\left(\frac{s + \dfrac{1}{T_2}}{s + \dfrac{1}{\beta T_2}}\right) = K_c\frac{(T_1 s + 1)(T_2 s + 1)}{\left(\dfrac{T_1}{\beta}s + 1\right)(\beta T_2 s + 1)}$$

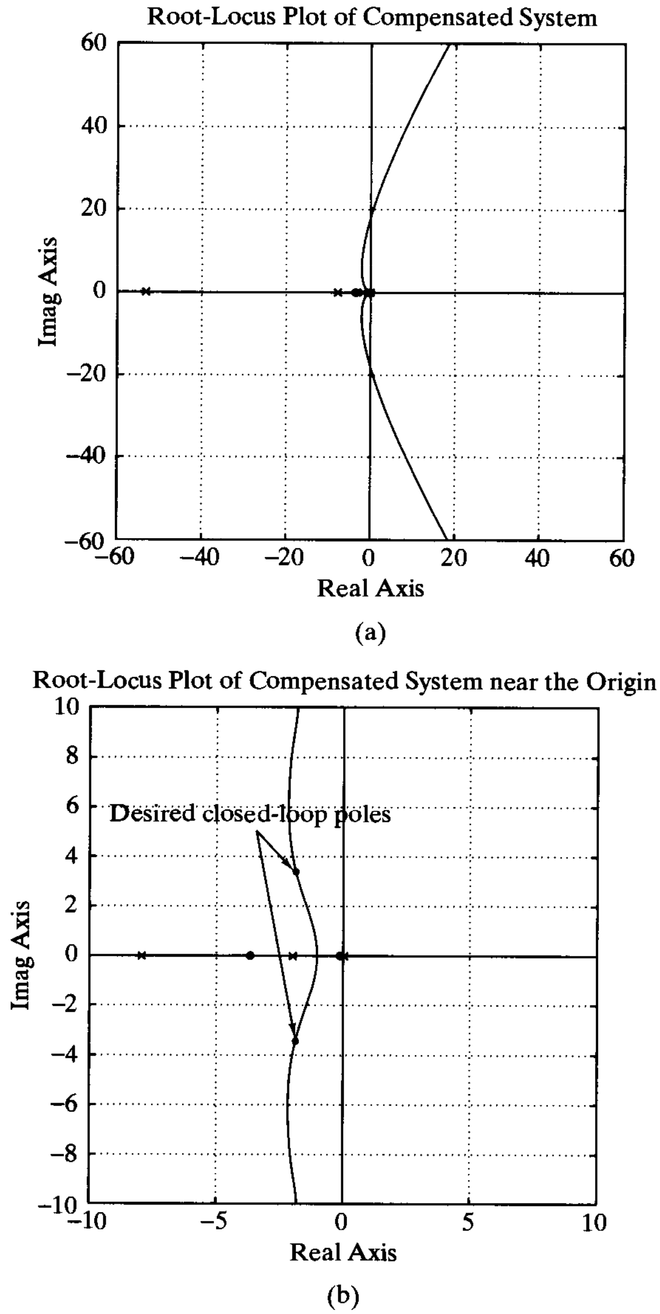


Figure 7–35
(a) Root-locus plot of compensated system; (b) root-locus plot near the origin.

where $\beta > 1$. Then

$$K_v = \lim_{s \to 0} sG_c(s)G(s)$$

$$= \lim_{s \to 0} s \frac{K_c(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta}s + 1\right)\left(\beta T_2 s + 1\right)} \frac{1}{s(s+1)(s+5)}$$

$$= \frac{K_c}{5}$$

The specification that $K_v = 50 \text{ sec}^{-1}$ determines the value of K_c, or

$$K_c = 250$$

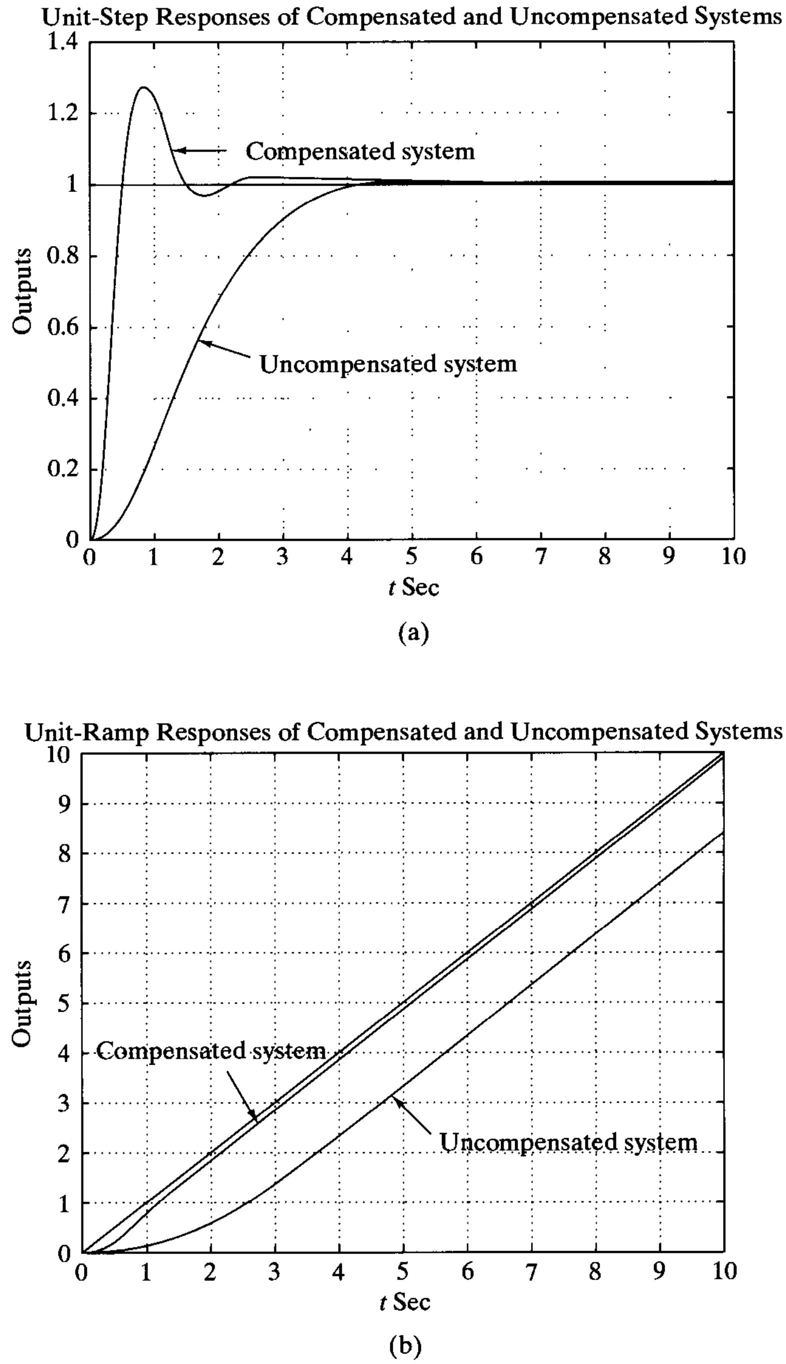


Figure 7–36
(a) Unit-step responses of compensated and uncompensated systems;
(b) unit-ramp responses of both systems.

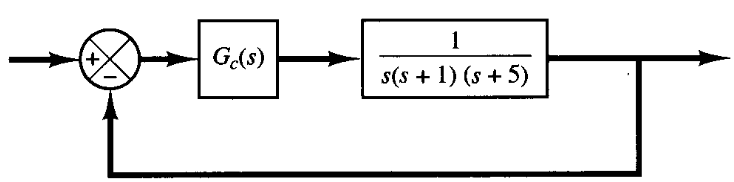


Figure 7–37
Control system.

We now choose $T_1 = 1$ so that $s + (1/T_1)$ will cancel the $(s + 1)$ term of the plant. The lead portion then becomes

$$\frac{s + 1}{s + \beta}$$

For the lag portion of the lag–lead compensator we require

$$\left|\frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{1}{\beta T_2}}\right| \doteq 1, \qquad -5^\circ < \angle\frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{1}{\beta T_2}} < 0^\circ$$

where $s = s_1$ is one of the dominant closed-loop poles. For $s = s_1$, the open-loop transfer function becomes

$$G_c(s_1)G(s_1) \doteq K_c\left(\frac{s_1 + 1}{s_1 + \beta}\right)\frac{1}{s_1(s_1 + 1)(s_1 + 5)} = K_c\frac{1}{(s_1 + \beta)s_1(s_1 + 5)}$$

Noting that at $s = s_1$ the magnitude and angle conditions are satisfied, we have

$$\left|K_c\frac{1}{s_1(s_1 + \beta)(s_1 + 5)}\right| = 1 \tag{7–8}$$

$$\angle K_c\frac{1}{s_1(s + \beta)(s_1 + 5)} = \pm 180^\circ(2k + 1) \tag{7–9}$$

where $k = 0,1,2,\ldots$. In Equations (7–8) and (7–9), β and s_1 are unknowns. Since the damping ratio ζ of the dominant closed-loop poles is specified as 0.5, the closed-loop pole $s = s_1$ can be written as

$$s_1 = -x + j\sqrt{3}\,x$$

where x is as yet undetermined.

Notice that the magnitude condition, Equation (7–8), can be rewritten as

$$\left|\frac{K_c}{(-x + j\sqrt{3}\,x)(-x + \beta + j\sqrt{3}\,x)(-x + 5 + j\sqrt{3}\,x)}\right| = 1$$

Noting that $K_c = 250$, we have

$$x\sqrt{(\beta - x)^2 + 3x^2}\,\sqrt{(5 - x)^2 + 3x^2} = 125 \tag{7–10}$$

The angle condition, Equation (7–9), can be rewritten as

$$\angle K_c\frac{1}{(-x + j\sqrt{3}\,x)(-x + \beta + j\sqrt{3}\,x)(-x + 5 + j\sqrt{3}\,x)}$$

$$= -120^\circ - \tan^{-1}\left(\frac{\sqrt{3}\,x}{-x + \beta}\right) - \tan^{-1}\left(\frac{\sqrt{3}\,x}{-x + 5}\right) = -180^\circ$$

or

$$\tan^{-1}\left(\frac{\sqrt{3}\,x}{-x + \beta}\right) + \tan^{-1}\left(\frac{\sqrt{3}\,x}{-x + 5}\right) = 60^\circ \tag{7–11}$$

We need to solve Equations (7–10) and (7–11) for β and x. By several trial-and-error calculations, it can be found that

$$\beta = 16.025, \qquad x = 1.9054$$

Thus

$$s_1 = -1.9054 + j\sqrt{3}\,(1.9054) = -1.9054 + j3.3002$$

The lag portion of the lag–lead compensator can be determined as follows: Noting that the pole and zero of the lag portion of the compensator must be located near the origin, we may choose

$$\frac{1}{\beta T_2} = 0.01$$

That is,

$$\frac{1}{T_2} = 0.16025 \qquad \text{or} \qquad T_2 = 6.25$$

With the choice of $T_2 = 6.25$, we find

$$\left| \frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{1}{\beta T_2}} \right| = \left| \frac{-1.9054 + j3.3002 + 0.16025}{-1.9054 + j3.3002 + 0.01} \right|$$

$$= \left| \frac{-1.74515 + j3.3002}{-1.89054 + j3.3002} \right| = 0.98 \doteqdot 1 \tag{7–12}$$

and

$$\angle \frac{s_1 + \dfrac{1}{T_2}}{s_1 + \dfrac{\gamma}{\beta T_2}} = \angle \frac{-1.9054 + j3.3002 + 0.16025}{-1.9054 + j3.3002 + 0.01}$$

$$= \tan^{-1}\left(\frac{3.3002}{-1.74515}\right) - \tan^{-1}\left(\frac{3.3002}{-1.89054}\right) = -1.937° \tag{7–13}$$

Since

$$-5° < -1.937° < 0°$$

our choice of $T_2 = 6.25$ is acceptable. Then the lag–lead compensator just designed can be written as

$$G_c(s) = 250\left(\frac{s + 1}{s + 16.025}\right)\left(\frac{s + 0.16025}{s + 0.01}\right)$$

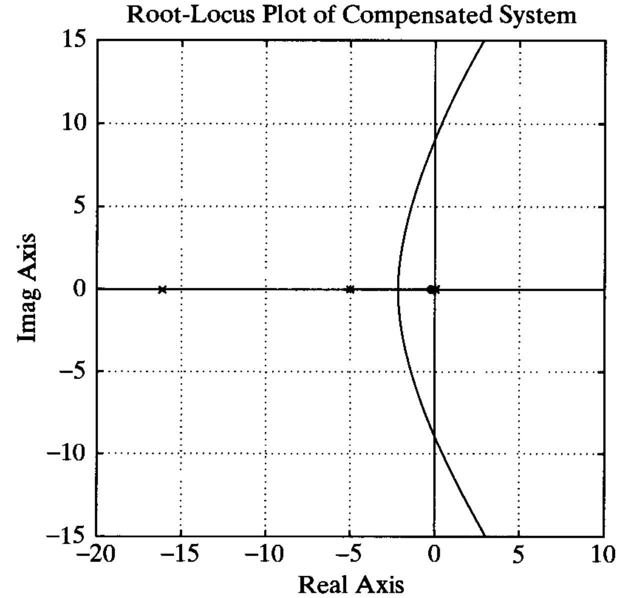
Therefore, the compensated system has the following open-loop transfer function:

$$G_c(s)G(s) = \frac{250(s + 0.16025)}{s(s + 0.01)(s + 5)(s + 16.025)}$$

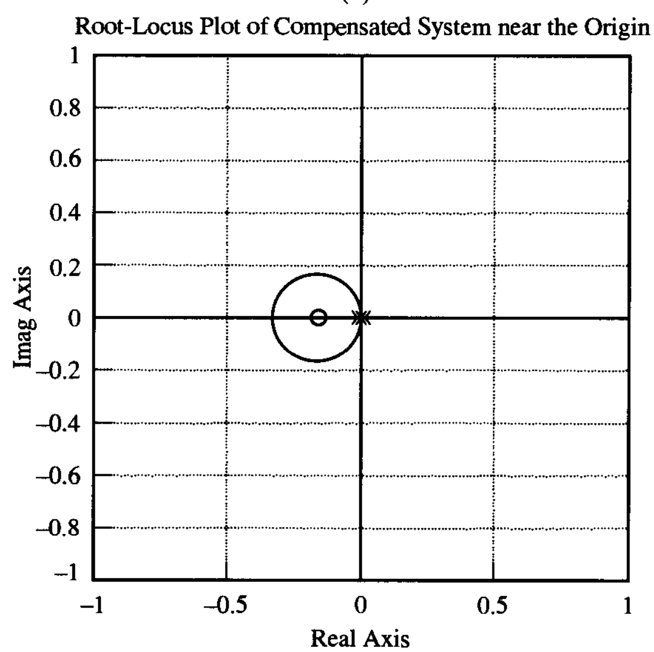
A root-locus plot of the compensated system is shown in Figure 7–38(a). An enlarged root-locus plot near the origin is shown in Figure 7–38(b).

The closed loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{250(s + 0.16025)}{s(s + 0.01)(s + 5)(s + 16.025) + 250(s + 0.16025)}$$



(a)



(b)

Figure 7–38
(a) Root-locus plot of compensated system; (b) root-locus plot near the origin.

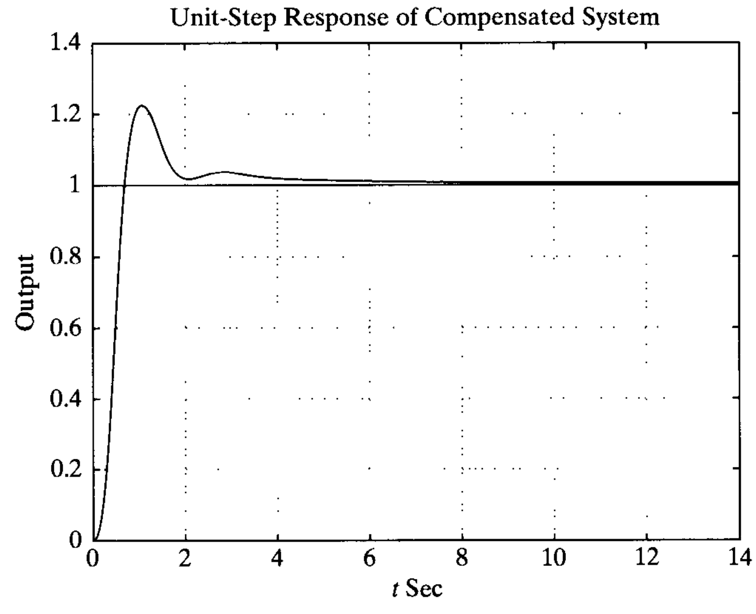
The closed-loop poles are located at

$$s = -1.8308 \pm j3.2359$$

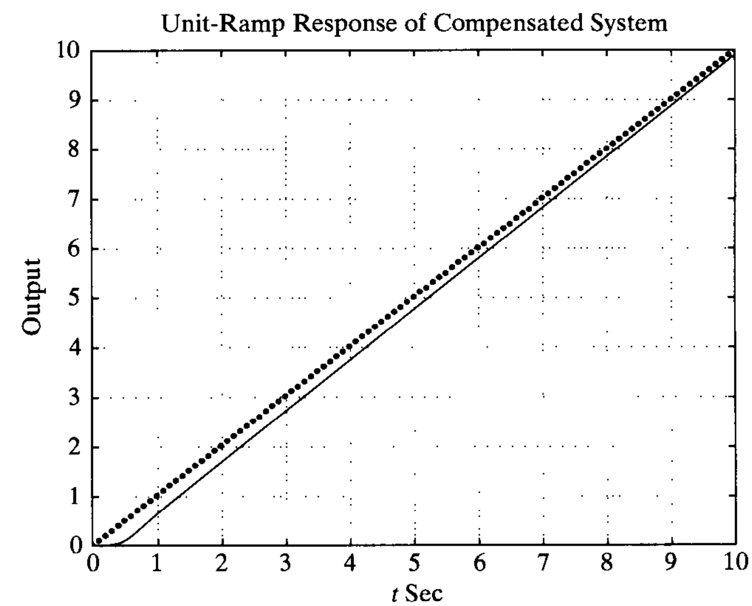
$$s = -0.1684$$

$$s = -17.205$$

Notice that the dominant closed-loop poles $s = -1.8308 \pm j3.2359$ differ from the dominant closed-loop poles $s = \pm s_1$ assumed in the computation of β and T_2. Small deviations of the dominant closed-loop poles $s = -1.8308 \pm j3.2359$ from $s = \pm s_1 = -1.9054 \pm j3.3002$ are due to the approximations involved in determining the lag portion of the compensator [See Equations (7–12) and (7–13)].



(a)



(b)

Figure 7–39
(a) Unit-step response of the compensated system; (b) unit-ramp response of the compensated system.

Figures 7–39(a) and (b) show the unit-step response and unit-ramp response of the designed system, respectively. Note that the closed-loop pole at $s = -0.1684$ almost cancels the zero at $s = -0.16025$. However, this pair of closed-loop pole and zero located near the origin produces a long tail of small amplitude. Since the closed-loop pole at $s = -17.205$ is located very much farther to the left compared to the closed-loop poles at $s = -1.8308 \pm j3.2359$, the effect of this real pole on the system response is also very small. Therefore, the closed-loop poles at $s = -1.8308 \pm j3.2359$ are indeed dominant closed-loop poles that determine the response characteristics of the closed-loop system. In the unit-ramp response, the steady-state error in following the unit-ramp input eventually becomes $1/K_v = \frac{1}{50} = 0.02$.

A-7-10. Consider the system shown in Figure 7–40. 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,

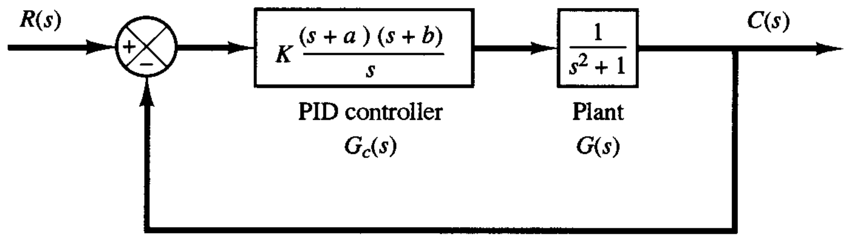


Figure 7–40 PID-controlled system.

choose $a = 1$ and then determine the values of K and b. Sketch the root-locus diagram for the designed system.

Solution. Since

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

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 7–41 is a root-locus plot.

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 7–42. The closed-loop pole at $s = -0.3333$ and a zero at $s = -0.5714$ produces a long tail of small amplitude.

A-7-11. Figure 7–43(a) is a block diagram of a model for an attitude-rate control system. The closed-loop transfer function for this system is

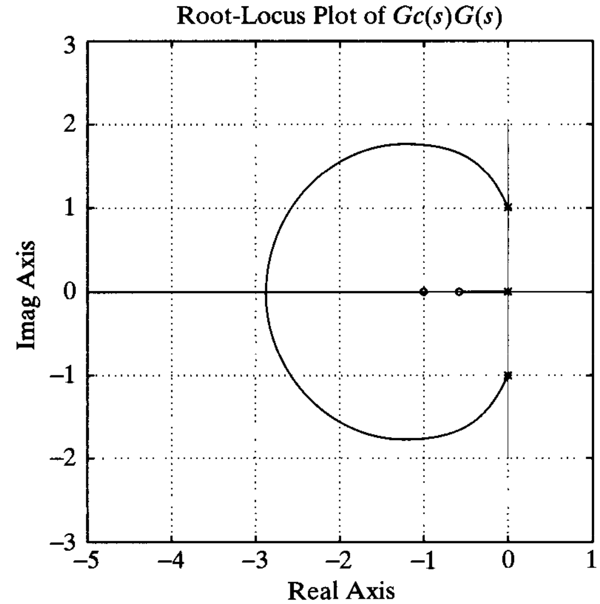


Figure 7–41 Root-locus plot of the compensated system (Problem A-7-10).

$$\frac{C(s)}{R(s)} = \frac{2s + 0.1}{s^3 + 0.1s^2 + 6s + 0.1}$$

$$= \frac{2(s + 0.05)}{(s + 0.0417 + j2.4489)(s + 0.0417 - j2.4489)(s + 0.0167)}$$

The unit-step response of this system is shown in Figure 7–43(b). The response shows high-frequency oscillations at the beginning of the response due to the poles at $s = -0.0417 \pm j2.4489$. The response is dominated by the pole at $s = -0.0167$. The settling time is approximately 240 sec.

It is desired to speed up the response and also eliminate the oscillatory mode at the beginning of the response. Design a suitable compensator such that the dominant closed-loop poles are at $s = -2 \pm j2\sqrt{3}$.

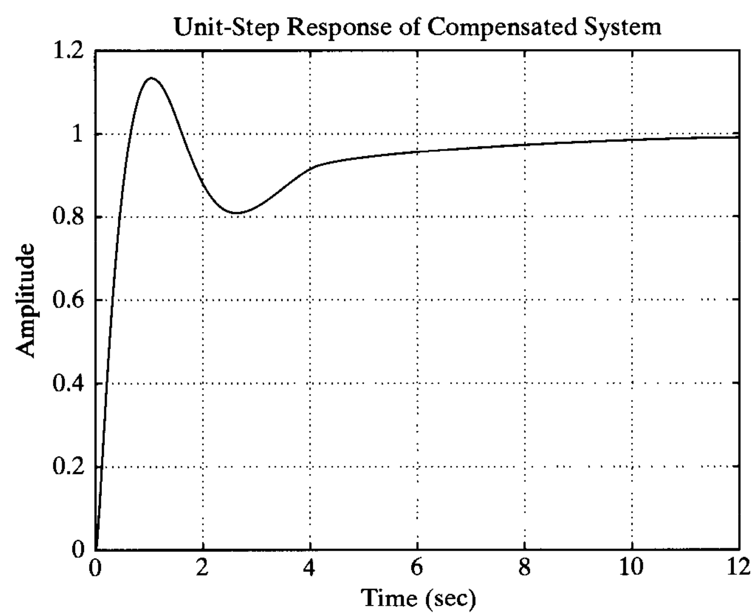
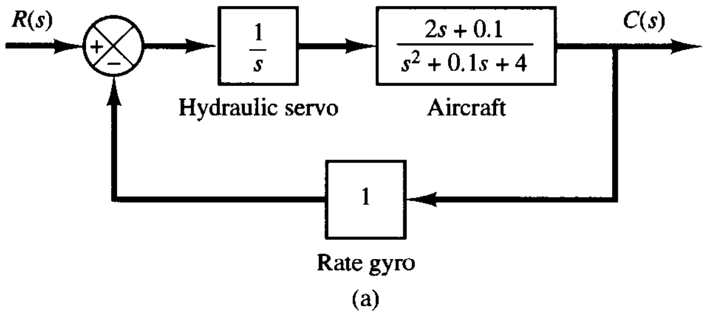


Figure 7–42 Unit-step response of the compensated system (Problem A-7-10).



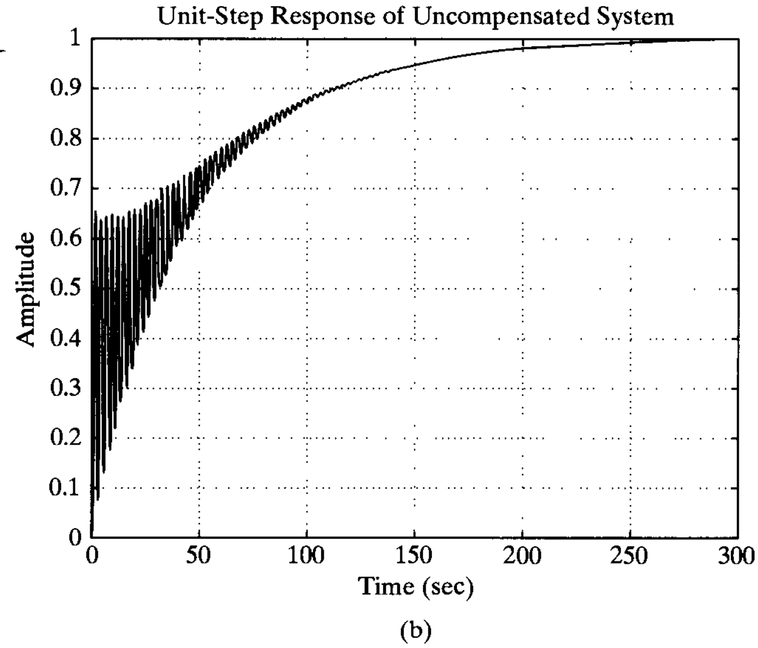


Figure 7–43
(a) Attitude-rate control system;
(b) unit-step response.

Solution. Figure 7–44 shows a block diagram for the compensated system. Note that the open-loop zero at $s = -0.05$ and the open-loop pole at $s = 0$ generate a closed-loop pole between $s = 0$ and $s = -0.05$. Such a closed-loop pole becomes a dominant closed-loop pole and make the response quite slow. Hence, it is necessary to replace this zero by a zero that is located far away from the $j\omega$ axis, for example, a zero at $s = -4$.

We now choose the compensator in the following form:

$$G_c(s) = \hat{G}_c(s)\frac{s + 4}{2s + 0.1}$$

Then the open-loop transfer function of the compensated system becomes

$$G_c(s)G(s) = \hat{G}_c(s)\frac{s + 4}{2s + 0.1}\frac{1}{s}\frac{2s + 0.1}{s^2 + 0.1s + 4}$$

$$= \hat{G}_c(s)\frac{s + 4}{s(s^2 + 0.1s + 4)}$$

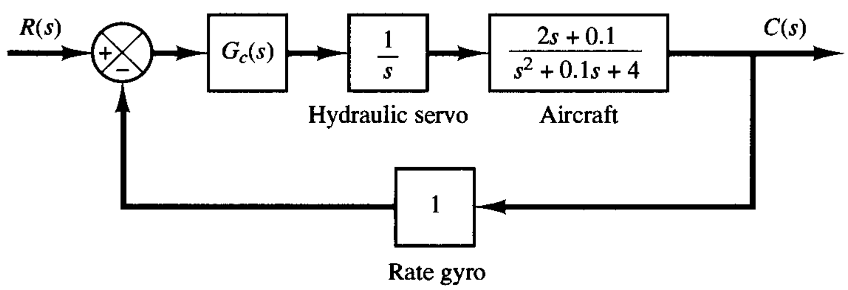


Figure 7–44
Compensated attitude-rate control system.

To determine $\hat{G}_c(s)$ by the root-locus method, we need to find the angle deficiency at the desired closed-loop pole $s = -2 + j2\sqrt{3}$. The angle deficiency can be found as follows:

$$\begin{aligned}\text{Angle deficiency} &= -143.088^\circ - 120^\circ - 109.642^\circ + 60^\circ + 180^\circ \\ &= -132.73^\circ\end{aligned}$$

Hence, the lead compensator $\hat{G}_c(s)$ must provide 132.73°. Since the angle deficiency is 132.73°, we need two lead compensators, each providing 66.365°. Thus $G_c(s)$ will have the following form:

$$G_c(s) = K_c\left(\frac{s + s_z}{s + s_p}\right)^2$$

Suppose that we choose two zeros at $s = -2$. Then the two poles of the lead compensators can be obtained from

$$\frac{3.4641}{s_p - 2} = \tan(90^\circ - 66.365^\circ) = 0.4376169$$

or

$$\begin{aligned}s_p &= 2 + \frac{3.4641}{0.4376169} \\ &= 9.9158\end{aligned}$$

Hence,

$$\hat{G}_c(s) = K_c\left(\frac{s + 2}{s + 9.9158}\right)^2$$

The entire compensator $G_c(s)$ for the system becomes

$$G_c(s) = \hat{G}_c(s)\frac{s + 4}{2s + 0.1} = K_c\frac{(s + 2)^2}{(s + 9.9158)^2}\frac{s + 4}{2s + 0.1}$$

The value of K_c can be determined from the magnitude condition. Since the open-loop transfer function is

$$G_c(s)G(s) = K_c\frac{(s + 2)^2(s + 4)}{(s + 9.9158)^2 s(s^2 + 0.1s + 4)}$$

the magnitude condition becomes

$$\left|K_c\frac{(s + 2)^2(s + 4)}{(s + 9.9158)^2 s(s^2 + 0.1s + 4)}\right|_{s=-2+j2\sqrt{3}} = 1$$

Hence,

$$\begin{aligned}K_c &= \left|\frac{(s + 9.9158)^2 s(s^2 + 0.1s + 4)}{(s + 2)^2(s + 4)}\right|_{s=-2+j2\sqrt{3}} \\ &= 88.0227\end{aligned}$$

Thus the compensator $G_c(s)$ becomes

$$G_c(s) = 88.0227 \frac{(s+2)^2(s+4)}{(s+9.9158)^2(2s+0.1)}$$

The open-loop transfer function is given by

$$G_c(s)G(s) = \frac{88.0227(s+2)^2(s+4)}{(s+9.9158)^2 s(s^2+0.1s+4)}$$

A root-locus plot for the compensated system is shown in Figure 7–45. The closed-loop poles for the compensated system are indicated in the plot. The closed-loop poles, the roots of the characteristic equation

$$(s+9.9158)^2 s(s^2+0.1s+4) + 88.0227(s+2)^2(s+4) = 0$$

are as follows:

$$s = -2.0000 \pm j3.4641$$

$$s = -7.5224 \pm j6.5326$$

$$s = -0.8868$$

Now that the compensator has been designed, we shall examine the transient response characteristics with MATLAB. The closed-loop transfer function is given by

$$\frac{C(s)}{R(s)} = \frac{88.0227(s+2)^2(s+4)}{(s+9.9158)^2 s(s^2+0.1s+4) + 88.0227(s+2)^2(s+4)}$$

Figures 7–46(a) and (b) show the plots of the unit-step response and unit-ramp response of the compensated system. These response curves show that the designed system is acceptable.

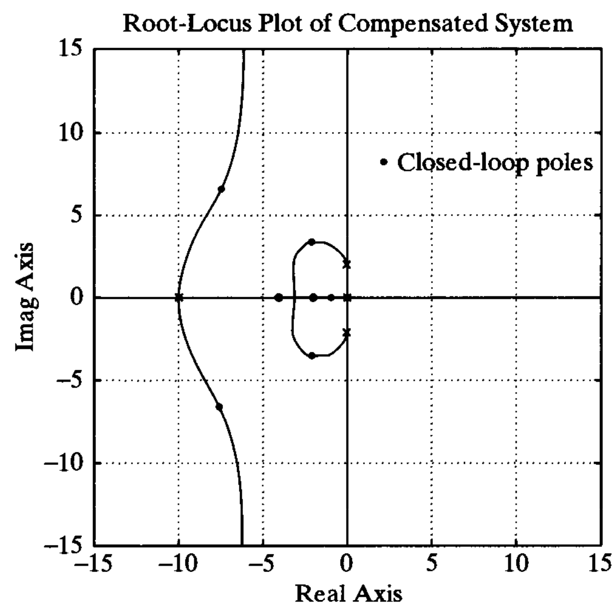


Figure 7–45 Root-locus plot of the compensated system.

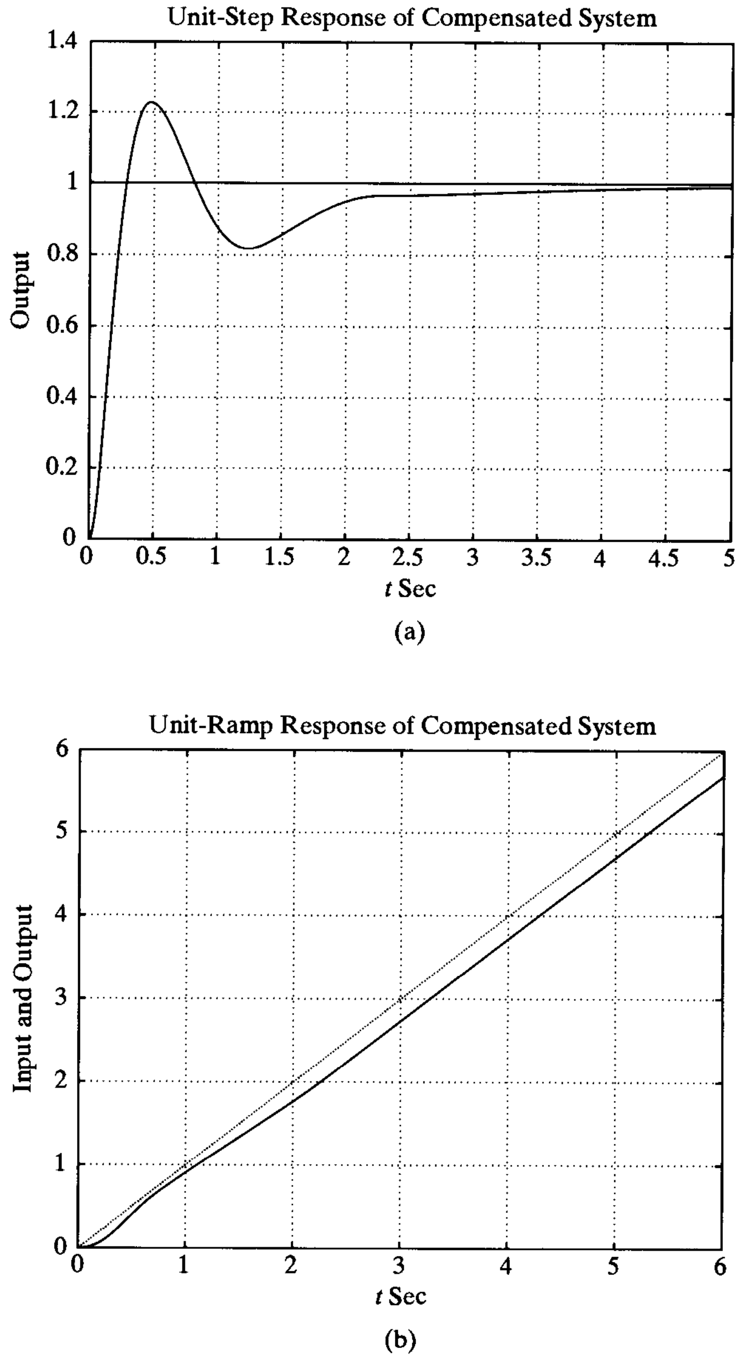


Figure 7–46
(a) Unit-step response of the compensated system; (b) unit-ramp response of the compensated system (Problem A-7-11).

A-7-12. Consider the model for a space vehicle control system shown in Figure 7–47. Design a lead compensator $G_c(s)$ such that the damping ratio ζ and the undamped natural frequency ω_n of the dominant closed-loop poles are 0.5 and 2 rad/sec, respectively.

Solution

First attempt: Assume the lead compensator $G_c(s)$ to be

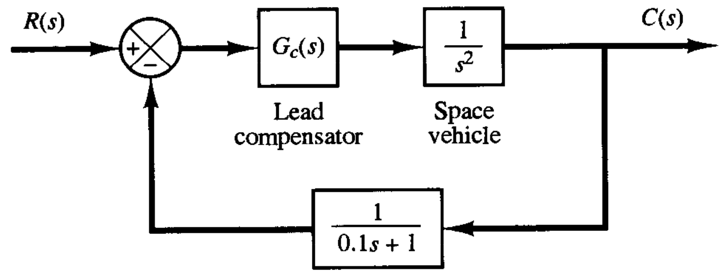


Figure 7–47 Space vehicle control system.

$$G_c(s) = K_c\left(\frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}}\right) \qquad (0 < \alpha < 1)$$

From the given specifications, $\zeta = 0.5$ and $\omega_n = 2$ rad/sec, the dominant closed-loop poles must be located at

$$s = -1 \pm j\sqrt{3}$$

We first calculate the angle deficiency at this closed-loop pole.

$$\begin{aligned}\text{Angle deficiency} &= -120^\circ - 120^\circ - 10.8934^\circ + 180^\circ \\ &= -70.8934^\circ\end{aligned}$$

This angle deficiency must be compensated by the lead compensator. There are many ways to determine the locations of the pole and zero of the lead network. Let us choose the zero of the compensator at $s = -1$. Then, referring to Figure 7–48, we have the following equation:

$$\frac{1.73205}{x - 1} = \tan(90^\circ - 70.8934^\circ) = 0.34641$$

or

$$x = 1 + \frac{1.73205}{0.34641} = 6$$

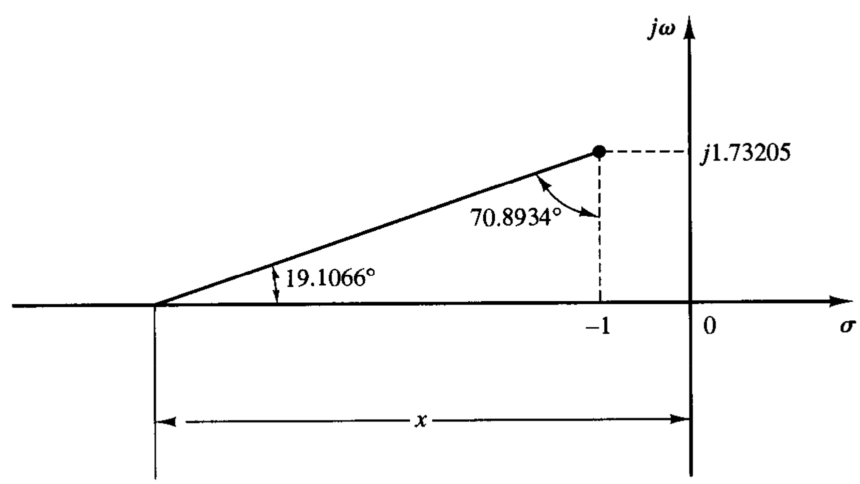


Figure 7–48 Determination of the pole of the lead network.

Hence,

$$G_c(s) = K_c \frac{s+1}{s+6}$$

The value of K_c can be determined from the magnitude condition

$$K_c \left| \frac{s+1}{s+6} \frac{1}{s^2} \frac{1}{0.1s+1} \right|_{s=-1+j\sqrt{3}} = 1$$

as follows:

$$K_c = \left| \frac{(s+6)s^2(0.1s+1)}{s+1} \right|_{s=-1+j\sqrt{3}} = 11.2000$$

Thus

$$G_c(s) = 11.2 \frac{s+1}{s+6}$$

Since the open-loop transfer function becomes

$$G_c(s)G(s)H(s) = 11.2 \frac{s+1}{(s+6)s^2(0.1s+1)}$$

$$= \frac{11.2(s+1)}{0.1s^4 + 1.6s^3 + 6s^2}$$

a root-locus plot of the compensated system can be obtained easily with MATLAB by entering num and den and using rlocus command. The result is shown in Figure 7–49.

The closed-loop transfer function for the compensated system becomes

$$\frac{C(s)}{R(s)} = \frac{11.2(s+1)(0.1s+1)}{(s+6)s^2(0.1s+1) + 11.2(s+1)}$$

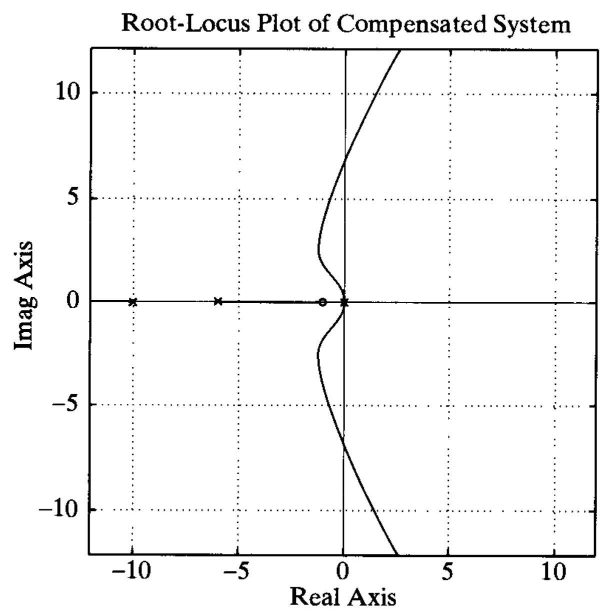


Figure 7–49 Root-locus plot of the compensated system.

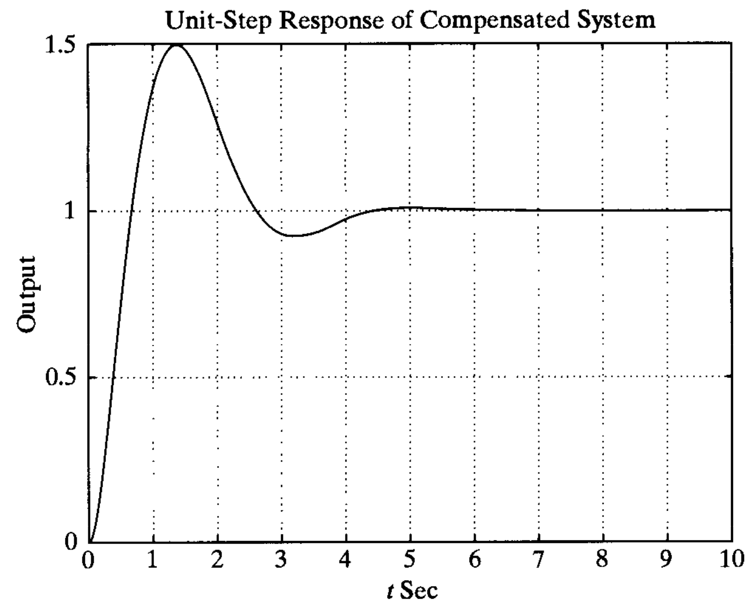


Figure 7–50 Unit-step response of the compensated system.

Figure 7–50 shows the unit-step response curve. It shows a fairly large overshoot (50% overshoot). It is desirable to modify the compensator and make the maximum overshoot smaller. A close look at the root-locus plot reveals that the presence of the zero at $s = -1$ is adding the amount of the maximum overshoot. One way to avoid this situation is to modify the lead compensator as presented in the following second attempt.

Second attempt: To modify the shape of the root loci, we may use two lead networks, each contributing half the necessary lead angle, which is $70.8934°/2 = 35.4467°$. Let us choose the location of the zeros at $s = -3$. (This is an arbitrary choice. Other choices such as $s = -2.5$ and $s = -4$ may be made.)

Once we choose two zeros at $s = -3$, the necessary location of the poles can be determined as shown in Figure 7–51, or

$$\frac{1.73205}{y - 1} = \tan(40.89334° - 35.4467°)$$
$$= \tan 5.4466° = 0.09535$$

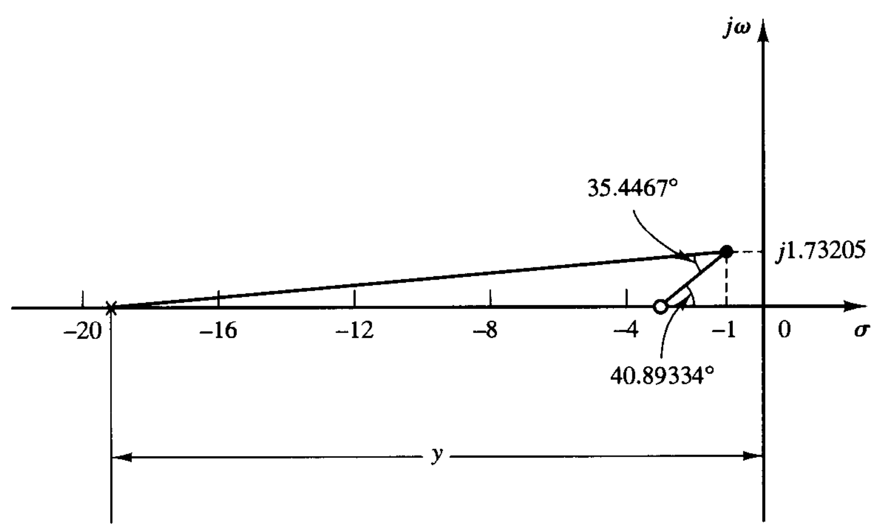


Figure 7–51 Determination of the pole of the lead network.

which yields

$$y = 1 + \frac{1.73205}{0.09535} = 19.1652$$

Hence, the lead compensator will have the following transfer function:

$$G_c(s) = K_c\left(\frac{s+3}{s+19.1652}\right)^2$$

The value of K_c can be determined from the magnitude condition as follows:

$$\left|K_c\left(\frac{s+3}{s+19.1652}\right)^2 \frac{1}{s^2}\frac{1}{0.1s+1}\right|_{s=-1+j\sqrt{3}} = 1$$

or

$$K_c = 174.3864$$

Then the lead compensator just designed is

$$G_c(s) = 174.3864\left(\frac{s+3}{s+19.1652}\right)^2$$

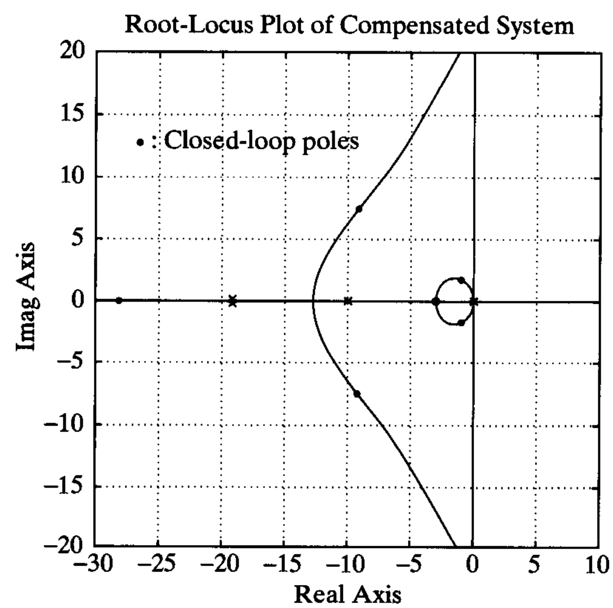
Then the open-loop transfer function becomes

$$G_c(s)G(s)H(s) = 174.3864\left(\frac{s+3}{s+19.1652}\right)^2 \frac{1}{s^2}\frac{1}{0.1s+1}$$

A root-locus plot for the compensated system is shown in Figure 7–52(a). Notice that there is no closed-loop zero near the origin. An expanded view of the root-locus plot near the origin is shown in Figure 7–52(b).

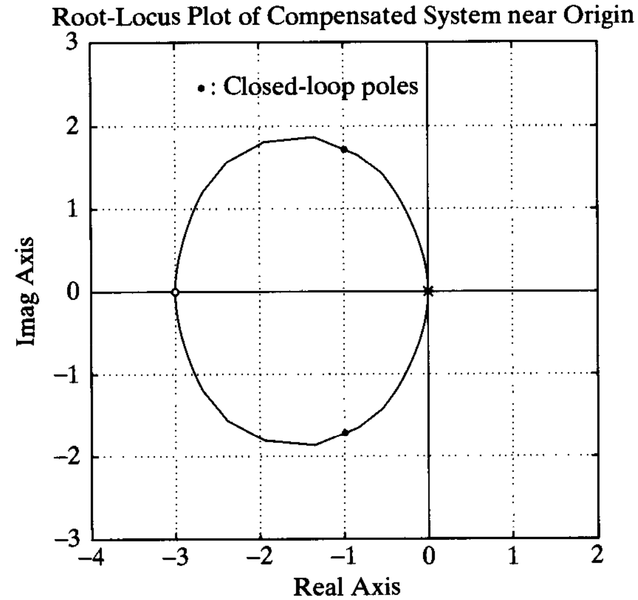
The closed-loop transfer function becomes

$$\frac{C(s)}{R(s)} = \frac{174.3864(s+3)^2(0.1s+1)}{(s+19.1652)^2 s^2(0.1s+1) + 174.3864(s+3)^2}$$



(a)

Figure 7–52
(a) Root-locus plot of compensated system; (b) root-locus plot near the origin.



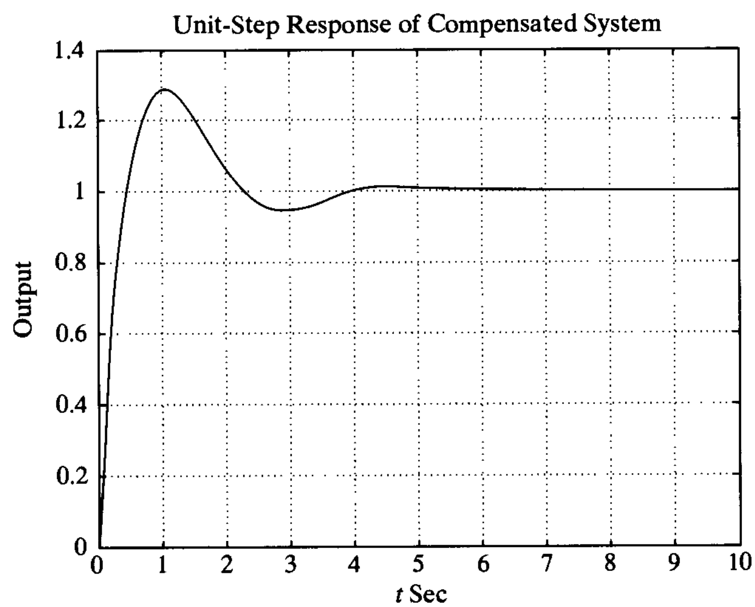
(b)

Figure 7–52
(Continued)

The closed-loop poles are found as follows:

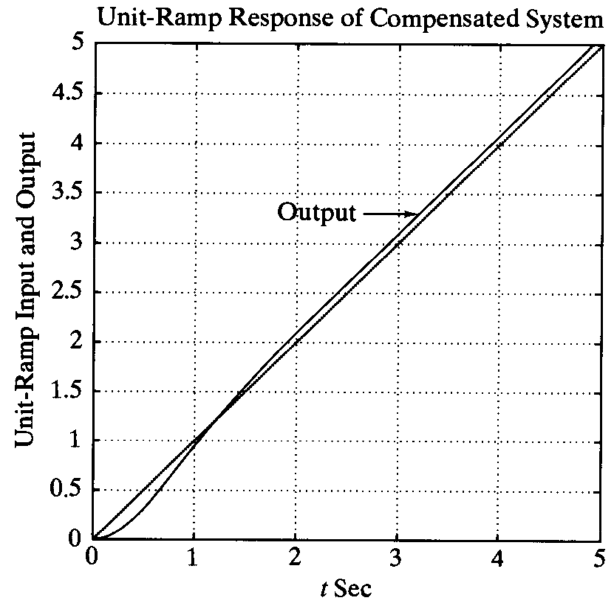
$$s = -1 \pm j1.73205$$
$$s = -9.1847 \pm j7.4814$$
$$s = -27.9606$$

Figures 7–53(a) and (b) show the unit-step response and unit-ramp response of the compensated system. The unit-step response curve is reasonable and the unit-ramp response looks acceptable. Notice that in the unit-ramp response the output leads the input by a small amount. This is because the system has a feedback transfer function $1/(0.1s + 1)$. If the feedback signal versus t is plotted, together with the unit-ramp input, the former will not lead the input ramp at steady state. See Figure 7–53(c).

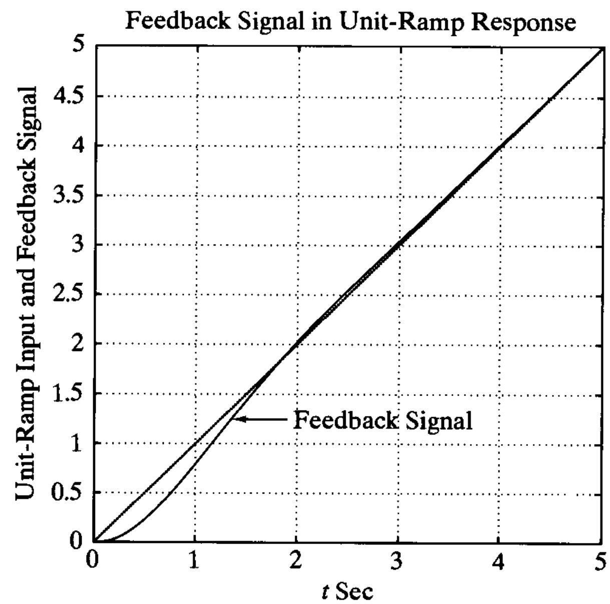


(a)

Figure 7–53
(a) Unit-step response of the compensated system;
(b) unit-ramp response of the compensated system;
(c) a plot of feedback signal versus t in the unit-ramp response.



(b)



(c)

Figure 7–53
(Continued)

PROBLEMS

B-7-1. Consider the mechanical system shown in Figure 7–54. It consists of a spring and two dashpots. Obtain the transfer function of the system. The displacement x_i is the input and displacement x_o is the output. Is this system a mechanical lead network or lag network?

B-7-2. Obtain the transfer function of the electrical network shown in Figure 7–55. Show that it is a lag network.

B-7-3. Consider the system shown in Figure 7–56. Plot the root loci for the system. Determine the value of K such that the damping ratio ζ of the dominant closed-loop poles is 0.5. Then determine all closed-loop poles. Plot the unit-step response curve with MATLAB.

B-7-4. Determine the values of K, T_1, and T_2 of the system shown in Figure 7–57 so that the dominant closed-loop

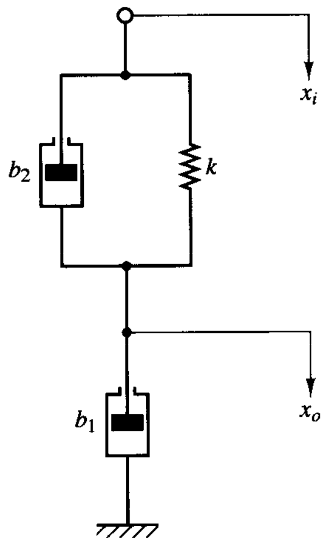


Figure 7–54
Mechanical system.

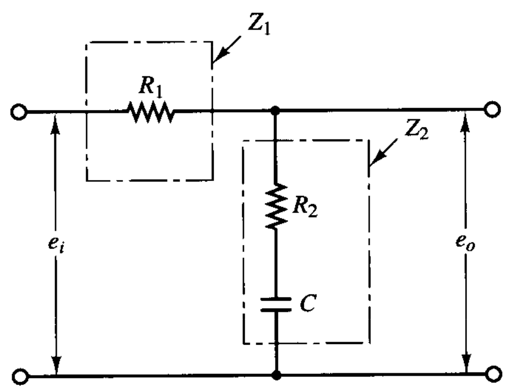


Figure 7–55
Electrical network.

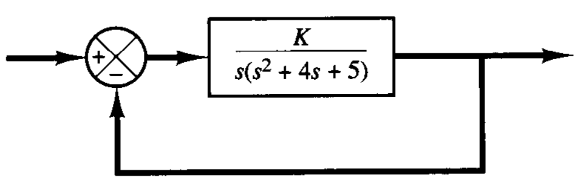


Figure 7–56
Control system.

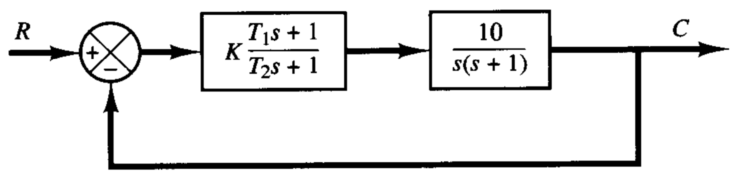


Figure 7–57
Control system.

poles have the damping ratio $\zeta = 0.5$ and the undamped natural frequency $\omega_n = 3$ rad/sec.

B-7-5. Consider the system shown in Figure 7–58, which involves velocity feedback. Determine the values of the amplifier gain K and the velocity feedback gain K_h so that the following specifications are satisfied:

1. Damping ratio of the closed-loop poles is 0.5
2. Settling time ≤ 2 sec
3. Static velocity error constant $K_v \geq 50\ \text{sec}^{-1}$
4. $0 < K_h < 1$

B-7-6. Consider the system shown in Figure 7–59. Design a lead compensator such that the dominant closed-loop poles

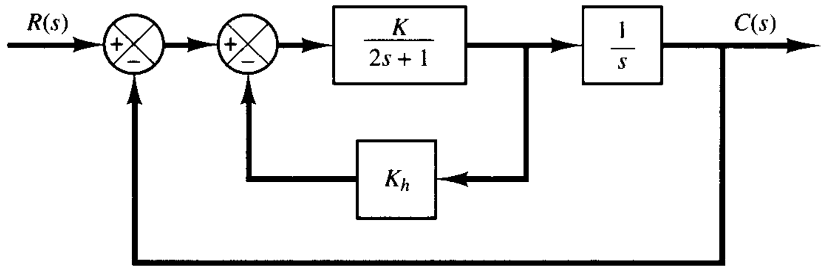


Figure 7–58
Control system.

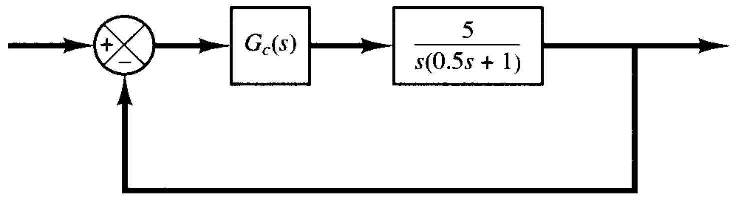


Figure 7–59
Control system.

are located at $s = -2 \pm j2\sqrt{3}$. Plot the unit-step response curve of the designed system with MATLAB.

B-7-7. Consider the system shown in Figure 7–60. Design a compensator such that the dominant closed-loop poles are located at $s = -1 \pm j1$.

B-7-8. Referring to the system shown in Figure 7–61, design a compensator such that the static velocity error constant K_v is 20 sec^{-1} without appreciably changing the original location ($s = -2 \pm j2\sqrt{3}$) of a pair of the complex-conjugate closed-loop poles.

B-7-9. Consider the angular-positional system shown in Figure 7–62. The dominant closed-loop poles are located at $s = -3.60 \pm j4.80$. The damping ratio ζ of the dominant closed-loop poles is 0.6. The static velocity error constant K_v is 4.1 sec^{-1}, which means that for a ramp input of 360°/sec the steady-state error in following the ramp input is

$$e_v = \frac{\theta_i}{K_v} = \frac{360°/\text{sec}}{4.1\ \text{sec}^{-1}} = 87.8°$$

It is desired to decrease e_v to one-tenth of the present value, or to increase the value of the static velocity error constant K_v to 41 sec^{-1}. It is also desired to keep the damping ratio ζ of the dominant closed-loop poles at 0.6. A small change in the undamped natural frequency ω_n of the dominant closed-loop poles is permissible. Design a suitable lag compensator to increase the static velocity error constant as desired.

B-7-10. Consider the control system shown in Figure 7–63. Design a compensator such that the dominant closed-loop poles are located at $s = -2 \pm j2\sqrt{3}$ and the static velocity error constant K_v is 50 sec^{-1}.

B-7-11. Consider the same system as considered in Problem A-7-10. It is desired to design a PID controller $G_c(s)$

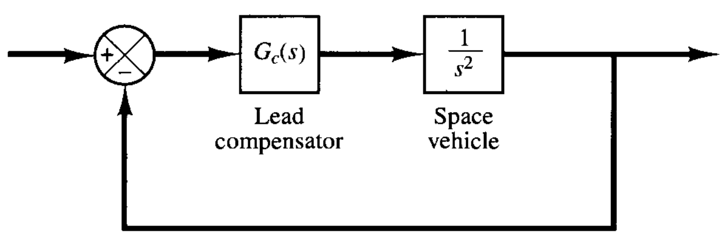


Figure 7–60
Control system.

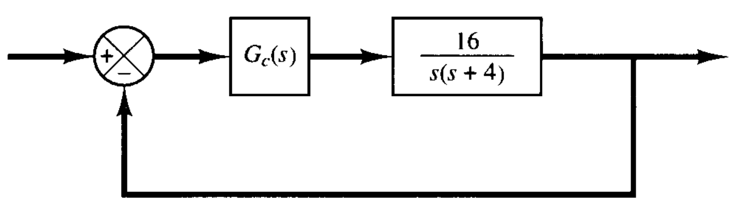


Figure 7–61
Control system.

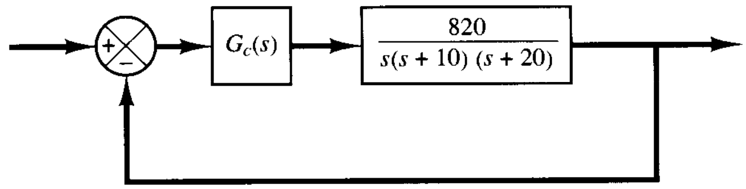


Figure 7–62
Angular-positional system.

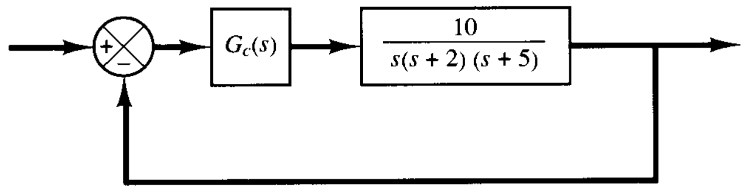


Figure 7–63
Control system.

such that the dominant closed-loop poles are located at $s = -1 \pm j\sqrt{3}$. For the PID controller, choose $a = 0.5$ (instead of $a = 1$ as discussed in Problem A-7-10) and then determine the values of K and b. Sketch the root-locus plot for the designed system. Also, obtain the unit-step response curve with MATLAB.

B-7-12. Consider the control system shown in Figure 7–64. The plant is critically stable in the sense that oscillations will continue indefinitely. Design a suitable compensator such that the unit-step response will exhibit maximum overshoot of less than 40% and settling time of 5 sec or less.

B-7-13. Consider the control system shown in Figure 7–65. Design a compensator such that the unit-step response curve will exhibit maximum overshoot of 30% or less and settling time of 3 sec or less.

B-7-14. Consider the control system shown in Figure 7–66. Design a compensator such that the unit-step response curve will exhibit maximum overshoot of 25% or less and settling time of 5 sec or less.

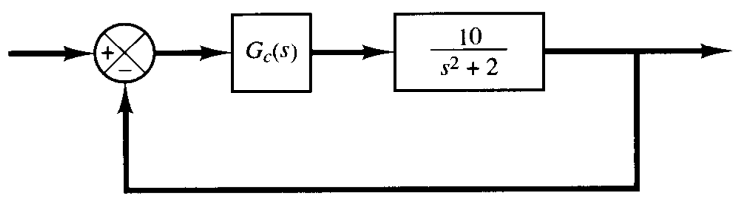


Figure 7–64
Control system.

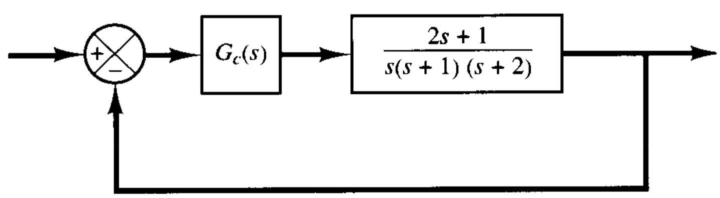


Figure 7–65
Control system.

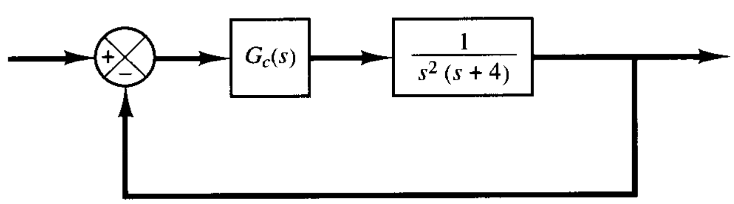


Figure 7–66
Control system.

8

Frequency-Response Analysis

8–1 INTRODUCTION

By the term *frequency response,* we mean the steady-state response of a system to a sinusoidal input. In frequency-response methods, we vary the frequency of the input signal over a certain range and study the resulting response.

The Nyquist stability criterion enables us to investigate both the absolute and relative stabilities of linear closed-loop systems from a knowledge of their open-loop frequency-response characteristics. An advantage of the frequency-response approach is that frequency-response tests are, in general, simple and can be made accurately by use of readily available sinusoidal signal generators and precise measurement equipment. Often the transfer functions of complicated components can be determined experimentally by frequency-response tests. In addition, the frequency-response approach has the advantages that a system may be designed so that the effects of undesirable noise are negligible and that such analysis and design can be extended to certain nonlinear control systems.

Although the frequency response of a control system presents a qualitative picture of the transient response, the correlation between frequency and transient responses is indirect, except for the case of second-order systems. In designing a closed-loop system, we adjust the frequency-response characteristic of the open-loop transfer function by using several design criteria in order to obtain acceptable transient-response characteristics for the system.

Steady-state output to sinusoidal input. Consider the linear time-invariant system shown in Figure 8–1. For this system

$$\frac{Y(s)}{X(s)} = G(s)$$

The input $x(t)$ is sinusoidal and is given by

$$x(t) = X \sin \omega t$$

As presented in Chapter 5, if the system is stable, then the output $y(t)$ can be given by

$$y(t) = Y \sin(\omega t + \phi)$$

where

$$Y = X|G(j\omega)|$$

and

$$\phi = \underline{/G(j\omega)} = \tan^{-1}\left[\frac{\text{imaginary part of } G(j\omega)}{\text{real part of } G(j\omega)}\right]$$

A stable linear time-invariant system subjected to a sinusoidal input will, at steady state, have a sinusoidal output of the same frequency as the input. But the amplitude and phase of the output will, in general, be different from those of the input. In fact, the amplitude of the output is given by the product of that of the input and $|G(j\omega)|$, while the phase angle differs from that of the input by the amount $\phi = \underline{/G(j\omega)}$. An example of input and output sinusoidal signals is shown in Figure 8–2.

Note that for sinusoidal inputs

$$|G(j\omega)| = \left|\frac{Y(j\omega)}{X(j\omega)}\right| = \text{amplitude ratio of the output sinusoid to the input sinusoid}$$

$$\underline{/G(j\omega)} = \underline{/\frac{Y(j\omega)}{X(j\omega)}} = \text{phase shift of the output sinusoid with respect to the input sinusoid}$$

Hence, the response characteristics of a system to a sinusoidal input can be obtained directly from

$$\frac{Y(j\omega)}{X(j\omega)} = G(j\omega)$$

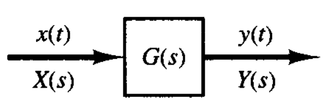


Figure 8–1
Linear time-invariant system.

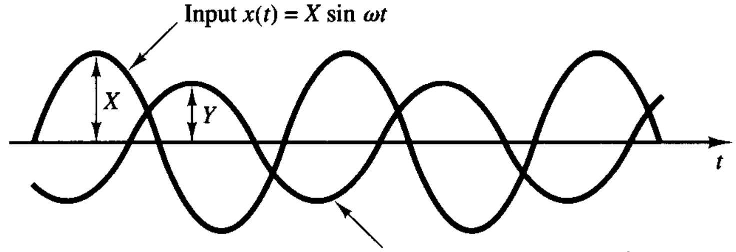


Figure 8–2
Input and output sinusoidal signals.

The sinusoidal transfer function $G(j\omega)$, the ratio of $Y(j\omega)$ to $X(j\omega)$, is a complex quantity and can be represented by the magnitude and phase angle with frequency as a parameter. (A negative phase angle is called *phase lag,* and a positive phase angle is called *phase lead.*) The sinusoidal transfer function of any linear system is obtained by substituting $j\omega$ for s in the transfer function of the system.

Presenting frequency-response characteristics in graphical forms. The sinusoidal transfer function, a complex function of the frequency ω, is characterized by its magnitude and phase angle, with frequency as the parameter. There are three commonly used representations of sinusoidal transfer functions:

1. Bode diagram or logarithmic plot
2. Nyquist plot or polar plot
3. Log-magnitude versus phase plot

We shall discuss these representations in detail in this chapter. We shall also discuss the MATLAB approach to obtain Bode diagrams and Nyquist plots.

Outline of the chapter. Section 8–1 has presented introductory material on the frequency response. Section 8–2 presents Bode diagrams of various transfer-function systems. Section 8–3 discusses a computational approach to obtain Bode diagrams with MATLAB, Section 8–4 treats polar plots of sinusoidal transfer functions, and Section 8–5 discusses drawing Nyquist plots with MATLAB. Section 8–6 briefly presents log-magnitude versus phase plots. Section 8–7 gives a detailed account of Nyquist stability criterion, Section 8–8 discusses the stability analysis of closed-loop systems using the Nyquist stability criterion, and Section 8–9 treats the relative stability analysis of closed-loop systems. Measures of relative stability such as phase margin and gain margin are introduced here. The correlation between the transient response and frequency response is also discussed. Section 8–10 presents a method for obtaining the closed-loop frequency response from the open-loop frequency response by use of the M and N circles. Use of the Nichols chart is also discussed for obtaining the closed-loop frequency response. Finally, Section 8–11 deals with the determination of the transfer function based on an experimental Bode diagram.

8–2 BODE DIAGRAMS

Bode diagrams or logarithmic plots. A sinusoidal transfer function may be represented by two separate plots, one giving the magnitude versus frequency and the other the phase angle (in degrees) versus frequency. A Bode diagram consists of two graphs: One is a plot of the logarithm of the magnitude of a sinusoidal transfer function; the other is a plot of the phase angle; both are plotted against the frequency in logarithmic scale.

The standard representation of the logarithmic magnitude of $G(j\omega)$ is $20 \log |G(j\omega)|$, where the base of the logarithm is 10. The unit used in this representation of the magnitude is the decibel, usually abbreviated dB. In the logarithmic representation, the curves are drawn on semilog paper, using the log scale for frequency and the linear scale for either magnitude (but in decibels) or phase angle (in degrees). (The frequency range of interest determines the number of logarithmic cycles required on the abscissa.)

The main advantage of using the Bode diagram is that multiplication of magnitudes can be converted into addition. Furthermore, a simple method for sketching an approximate log-magnitude curve is available. It is based on asymptotic approximations. Such approximation by straight-line asymptotes is sufficient if only rough information on the frequency-response characteristics is needed. Should exact curve be desired, corrections can be made easily to these basic asymptotic ones. The phase-angle curves can be drawn easily if a template for the phase-angle curve of $1 + j\omega$ is available. Expanding the low-frequency range by use of a logarithmic scale for the frequency is very advantageous since characteristics at low frequencies are most important in practical systems. Although it is not possible to plot the curves right down to zero frequency because of the logarithmic frequency ($\log 0 = -\infty$), this does not create a serious problem.

Note that the experimental determination of a transfer function can be made simple if frequency-response data are presented in the form of a Bode diagram.

Basic factors of $G(j\omega)H(j\omega)$. As stated earlier, the main advantage in using the logarithmic plot is the relative ease of plotting frequency-response curves. The basic factors that very frequently occur in an arbitrary transfer function $G(j\omega)H(j\omega)$ are

1. Gain K
2. Integral and derivative factors $(j\omega)^{\mp 1}$
3. First-order factors $(1 + j\omega T)^{\mp 1}$
4. Quadratic factors $[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]^{\mp 1}$

Once we become familiar with the logarithmic plots of these basic factors, it is possible to utilize them in constructing a composite logarithmic plot for any general form of $G(j\omega)H(j\omega)$ by sketching the curves for each factor and adding individual curves graphically, because adding the logarithms of the gains corresponds to multiplying them together.

The process of obtaining the logarithmic plot can be further simplified by using asymptotic approximations to the curves for each factor. (If necessary, corrections can be made easily to an approximate plot to obtain an accurate one.)

The gain K. A number greater than unity has a positive value in decibels, while a number smaller than unity has a negative value. The log-magnitude curve for a constant gain K is a horizontal straight line at the magnitude of $20 \log K$ decibels. The phase angle of the gain K is zero. The effect of varying the gain K in the transfer function is that it raises or lowers the log-magnitude curve of the transfer function by the corresponding constant amount, but it has no effect on the phase curve.

A number–decibel conversion line is given in Figure 8–3. The decibel value of any number can be obtained from this line. As a number increases by a factor of 10, the corresponding decibel value increases by a factor of 20. This may be seen from the following:

$$20 \log(K \times 10) = 20 \log K + 20$$

Similarly,

$$20 \log (K \times 10^n) = 20 \log K + 20n$$

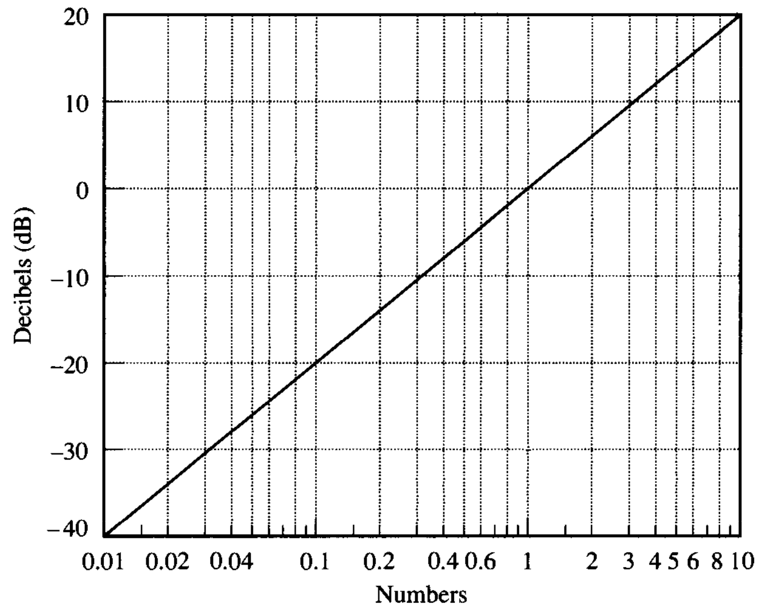


Figure 8–3 Number–decibel conversion line.

Note that, when expressed in decibels, the reciprocal of a number differs from its value only in sign; that is, for the number K,

$$20 \log K = -20 \log \frac{1}{K}$$

Integral and derivative factors $(j\omega)^{\mp 1}$. The logarithmic magnitude of $1/j\omega$ in decibels is

$$20 \log \left| \frac{1}{j\omega} \right| = -20 \log \omega \text{ dB}$$

The phase angle of $1/j\omega$ is constant and equal to $-90°$.

In Bode diagrams, frequency ratios are expressed in terms of octaves or decades. An octave is a frequency band from ω_1 to $2\omega_1$, where ω_1 is any frequency value. A decade is a frequency band from ω_1 to $10\omega_1$, where again ω_1 is any frequency. (On the logarithmic scale of semilog paper, any given frequency ratio can be represented by the same horizontal distance. For example, the horizontal distance from $\omega = 1$ to $\omega = 10$ is equal to that from $\omega = 3$ to $\omega = 30$.)

If the log magnitude $-20 \log \omega$ dB is plotted against ω on a logarithmic scale, it is a straight line. To draw this straight line, we need to locate one point (0 dB, $\omega = 1$) on it. Since

$$(-20 \log 10\omega) \text{ dB} = (-20 \log \omega - 20) \text{ dB}$$

the slope of the line is -20 dB/decade (or -6 dB/octave).

Similarly, the log magnitude of $j\omega$ in decibels is

$$20 \log |j\omega| = 20 \log \omega \text{ dB}$$

The phase angle of $j\omega$ is constant and equal to $90°$. The log-magnitude curve is a straight line with a slope of 20 dB/decade. Figures 8–4 (a) and (b) show frequency-response

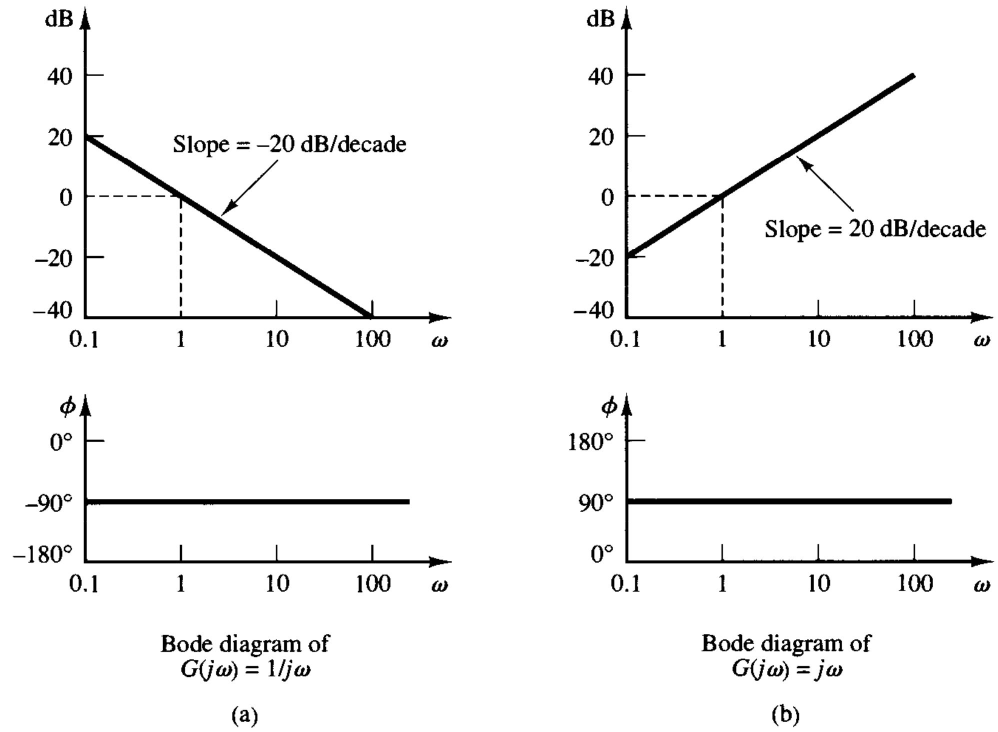


Figure 8–4
(a) Bode diagram of $G(j\omega) = 1/j\omega$;
(b) Bode diagram of $G(j\omega) = j\omega$.

curves for $1/j\omega$ and $j\omega$, respectively. We can clearly see that the differences in the frequency responses of the factors $1/j\omega$ and $j\omega$ lie in the signs of the slopes of the log-magnitude curves and in the signs of the phase angles. Both log magnitudes become equal to 0 dB at $\omega = 1$.

If the transfer function contains the factor $(1/j\omega)^n$ or $(j\omega)^n$, the log magnitude becomes, respectively,

$$20 \log \left| \frac{1}{(j\omega)^n} \right| = -n \times 20 \log |j\omega| = -20n \log \omega \text{ dB}$$

or

$$20 \log |(j\omega)^n| = n \times 20 \log |j\omega| = 20n \log \omega \text{ dB}$$

The slopes of the log-magnitude curves for the factors $(1/j\omega)^n$ and $(j\omega)^n$ are thus $-20n$ dB/decade and $20n$ dB/decade, respectively. The phase angle of $(1/j\omega)^n$ is equal to $-90° \times n$ over the entire frequency range, while that of $(j\omega)^n$ is equal to $90° \times n$ over the entire frequency range. The magnitude curves will pass through the point (0 dB, $\omega = 1$).

First-order factors $(1 + j\omega T)^{\mp 1}$. The log magnitude of the first-order factor $1/(1 + j\omega T)$ is

$$20 \log \left| \frac{1}{1 + j\omega T} \right| = -20 \log \sqrt{1 + \omega^2 T^2} \text{ dB}$$

For low frequencies, such that $\omega \ll 1/T$, the log magnitude may be approximated by

$$-20 \log \sqrt{1 + \omega^2 T^2} \doteqdot -20 \log 1 = 0 \text{ dB}$$

Thus, the log-magnitude curve at low frequencies is the constant 0-dB line. For high frequencies, such that $\omega \gg 1/T$,

$$-20 \log \sqrt{1 + \omega^2 T^2} \doteqdot -20 \log \omega T \text{ dB}$$

This is an approximate expression for the high-frequency range. At $\omega = 1/T$, the log magnitude equals 0 dB; at $\omega = 10/T$, the log magnitude is -20 dB. Thus, the value of $-20 \log \omega T$ dB decreases by 20 dB for every decade of ω. For $\omega \gg 1/T$, the log-magnitude curve is thus a straight line with a slope of -20 dB/decade (or -6 dB/octave).

Our analysis show that the logarithmic representation of the frequency-response curve of the factor $1/(1 + j\omega T)$ can be approximated by two straight-line asymptotes, one a straight line at 0 dB for the frequency range $0 < \omega < 1/T$ and the other a straight line with slope -20 dB/decade (or -6 dB/octave) for the frequency range $1/T < \omega < \infty$. The exact log-magnitude curve, the asymptotes, and the exact phase-angle curve are shown in Figure 8–5.

The frequency at which the two asymptotes meet is called the *corner* frequency or *break* frequency. For the factor $1/(1 + j\omega T)$, the frequency $\omega = 1/T$ is the corner frequency since at $\omega = 1/T$ the two asymptotes have the same value. (The low-frequency asymptotic expression at $\omega = 1/T$ is 20 log 1 dB = 0 dB, and the high-frequency asymptotic expression at $\omega = 1/T$ is also 20 log 1 dB = 0 dB.) The corner frequency divides the frequency-response curve into two regions, a curve for the low-frequency region and a curve for the high-frequency region. The corner frequency is very important in sketching logarithmic frequency-response curves.

The exact phase angle ϕ of the factor $1/(1 + j\omega T)$ is

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

At zero frequency, the phase angle is 0°. At the corner frequency, the phase angle is

$$\phi = -\tan^{-1} \frac{T}{T} = -\tan^{-1} 1 = -45^\circ$$

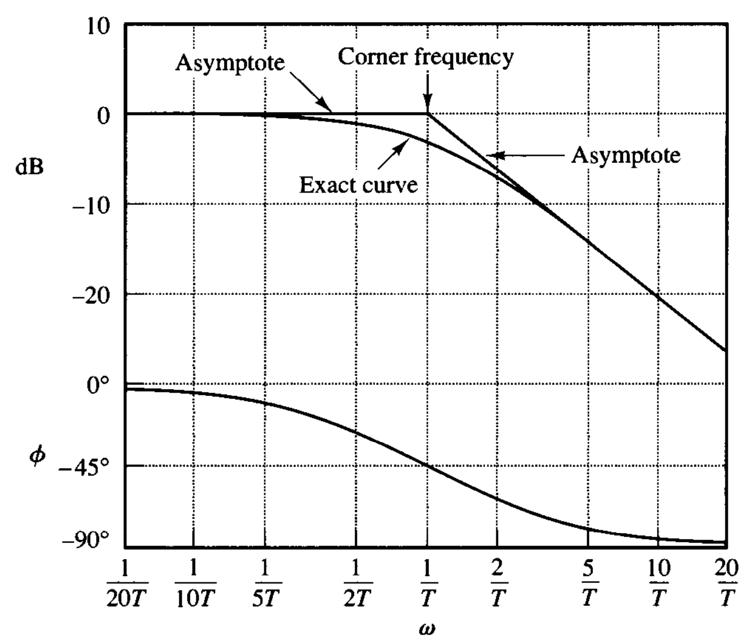


Figure 8–5
Log-magnitude curve together with the asymptotes and phase angle curve of $1/(1 + j\omega T)$.

At infinity, the phase angle becomes $-90°$. Since the phase angle is given by an inverse-tangent function, the phase angle is skew symmetric about the inflection point at $\phi = -45°$.

The error in the magnitude curve caused by the use of asymptotes can be calculated. The maximum error occurs at the corner frequency and is approximately equal to -3 dB since

$$-20 \log \sqrt{1 + 1} + 20 \log 1 = -10 \log 2 = -3.03 \text{ dB}$$

The error at the frequency one octave below the corner frequency, that is, at $\omega = 1/(2T)$, is

$$-20 \log \sqrt{\frac{1}{4} + 1} + 20 \log 1 = -20 \log \frac{\sqrt{5}}{2} = -0.97 \text{ dB}$$

The error at the frequency one octave above the corner frequency, that is, at $\omega = 2/T$, is

$$-20 \log \sqrt{2^2 + 1} + 20 \log 2 = -20 \log \frac{\sqrt{5}}{2} = -0.97 \text{ dB}$$

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 8–6. 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 8–6. 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.

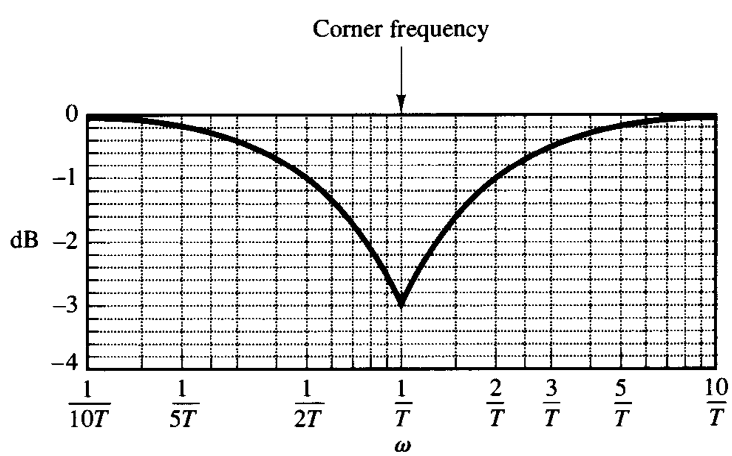


Figure 8–6 Log-magnitude error in the asymptotic expression of the frequency-response curve of $1/(1 + j\omega T)$.

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

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

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

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 8–7.

The shapes of phase-angle curves are the same for any factor of the form $(1 + j\omega T)^{\mp 1}$. Hence, it is convenient to have a template for the phase-angle curve on cardboard. Then such a template may be used repeatedly for constructing phase-angle

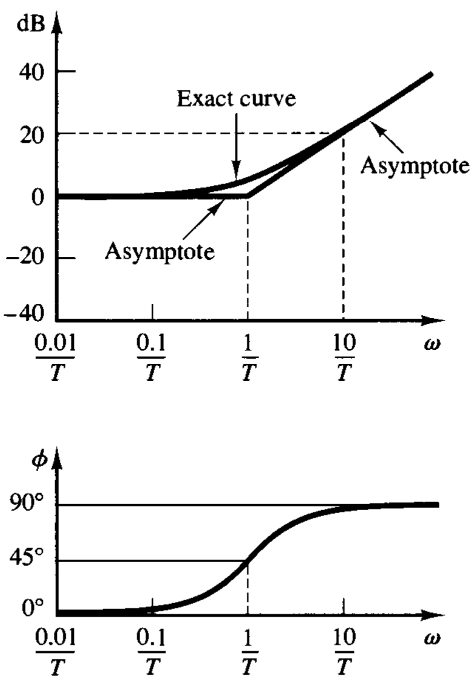


Figure 8–7
Log-magnitude curve together with the asymptotes and phase-angle curve for $1 + j\omega T$.

curves for any function of the form $(1 + j\omega T)^{\mp 1}$. If such a template is not available, we have to locate several points on the curve. The phase angles of $(1 + j\omega T)^{\mp 1}$ are

$$\begin{aligned}
&\mp 45^\circ \quad &\text{at} \quad &\omega = \frac{1}{T} \\
&\mp 26.6^\circ \quad &\text{at} \quad &\omega = \frac{1}{2T} \\
&\mp 5.7^\circ \quad &\text{at} \quad &\omega = \frac{1}{10T} \\
&\mp 63.4^\circ \quad &\text{at} \quad &\omega = \frac{2}{T} \\
&\mp 84.3^\circ \quad &\text{at} \quad &\omega = \frac{10}{T}
\end{aligned}$$

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 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

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

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

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

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 (8–1). 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 8–8 shows the exact log-magnitude curves together with the straight-line asymptotes and the exact phase-angle curves for the quadratic factor given by (8–1) 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 8–8.

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

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

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 8–8.

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

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

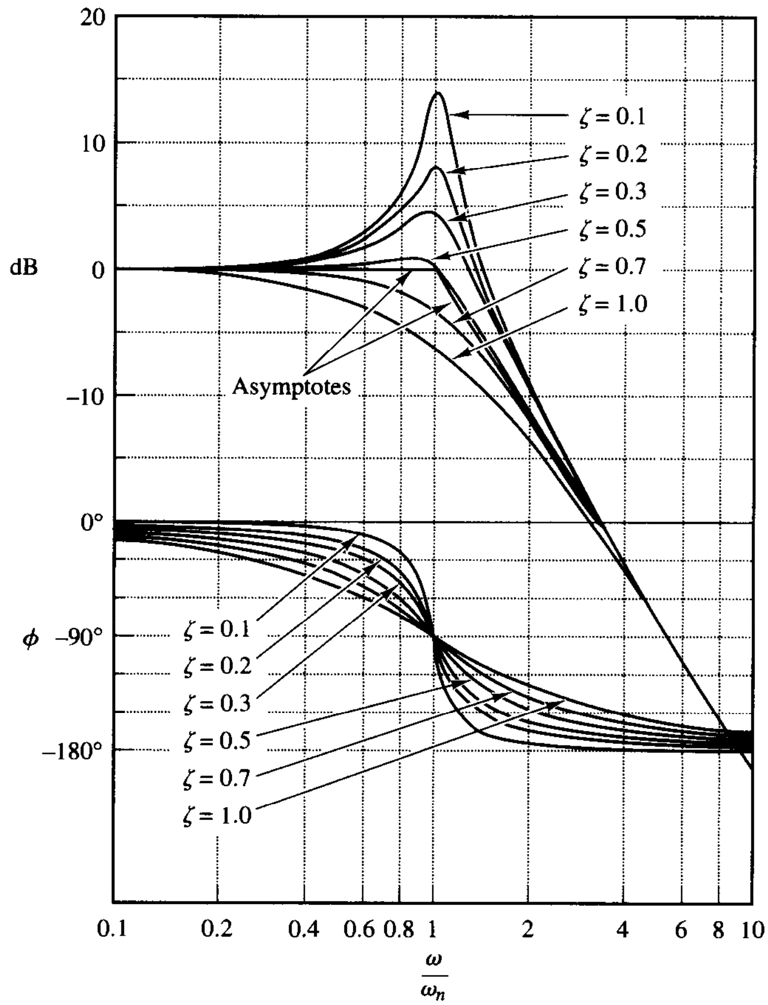


Figure 8–8 Log-magnitude curves together with the asymptotes and phase-angle curves of the quadratic transfer function given by (8–1).

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 8–8, 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

$$|G(j\omega)| = \frac{1}{\sqrt{\left(1 - \frac{\omega^2}{\omega_n^2}\right)^2 + \left(2\zeta\frac{\omega}{\omega_n}\right)^2}} \tag{8-3}$$

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

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

is a minimum. Since Equation (8–4) can be written

$$g(\omega) = \left[\frac{\omega^2 - \omega_n^2(1 - 2\zeta^2)}{\omega_n^2}\right]^2 + 4\zeta^2(1 - \zeta^2) \tag{8-5}$$

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

$$\omega_r = \omega_n\sqrt{1 - 2\zeta^2}, \qquad \text{for } 0 \le \zeta \le 0.707 \tag{8-6}$$

As the damping ratio ζ approaches zero, the resonant frequency approaches ω_n. For $0 < \zeta \leq 0.707$, 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 (8–6) 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.)

The magnitude of the resonant peak M_r can be found by substituting Equation (8–6) into Equation (8–3). For $0 \le \zeta \le 0.707$,

$$M_r = |G(j\omega)|_{\max} = |G(j\omega_r)| = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \tag{8-7}$$

For $\zeta > 0.707$,

$$M_r = 1 \tag{8-8}$$

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 8–9.

The phase angle of $G(j\omega)$ at the frequency where the resonant peak occurs can be obtained by substituting Equation (8–6) into Equation (8–2). Thus, at the resonant frequency ω_r,

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

General procedure for plotting Bode diagrams. First rewrite the sinusoidal-transfer function $G(j\omega)H(j\omega)$ as a product of basic factors discussed above. Then

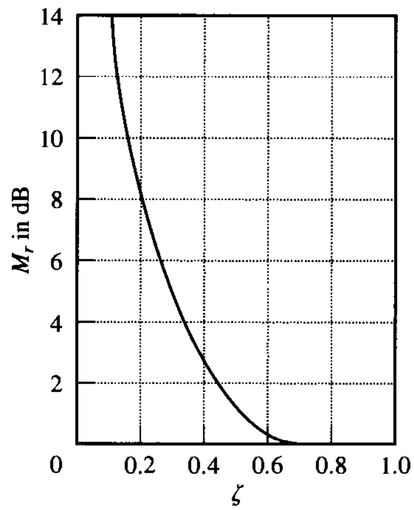


Figure 8-9
M_r versus ζ curve for the second-order system $1/[1 + 2\zeta(j\omega/\omega_n) + (j\omega/\omega_n)^2]$.

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.

EXAMPLE 8–1 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.

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

This function is composed of the following factors:

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

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 8–10. The composite curve is then obtained by adding algebraically the individual curves, also shown in Figure 8–10. 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 8–8. The exact log-magnitude curve for $G(j\omega)$ is shown by a dashed curve in Figure 8–10.

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

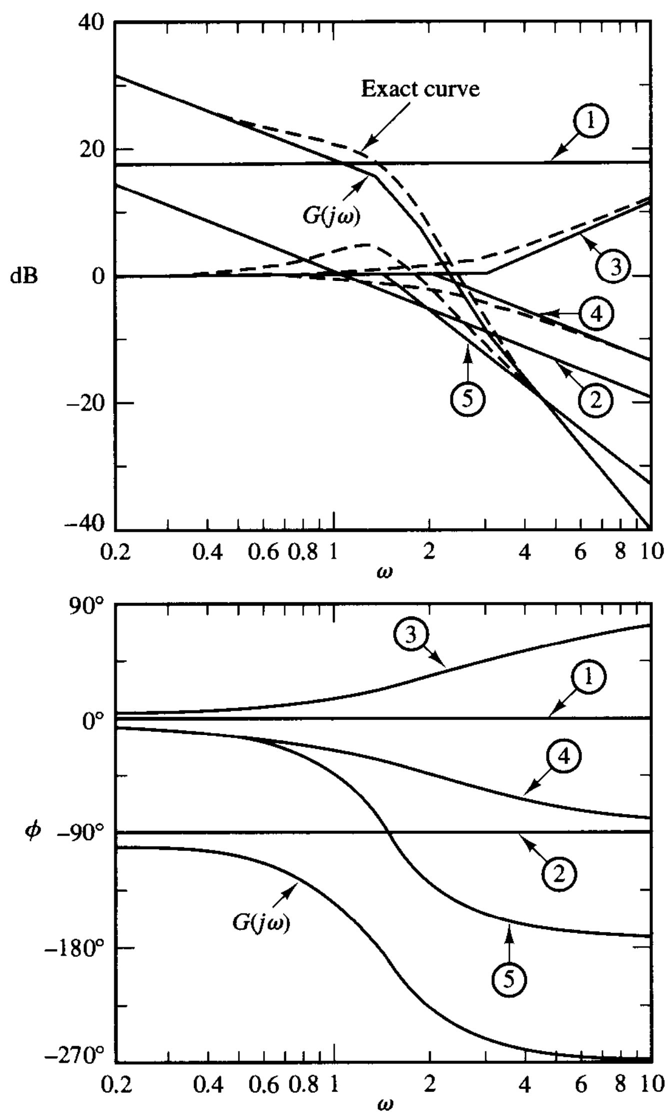


Figure 8–10
Bode diagram of the system considered in Example 8–1.

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 8–10.

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,

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

The pole–zero configurations of these systems are shown in Figure 8–11. The two sinusoidal transfer functions have the same magnitude characteristics, but they have different phase-angle characteristics, as shown in Figure 8–12. These two systems differ from each other by the factor

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

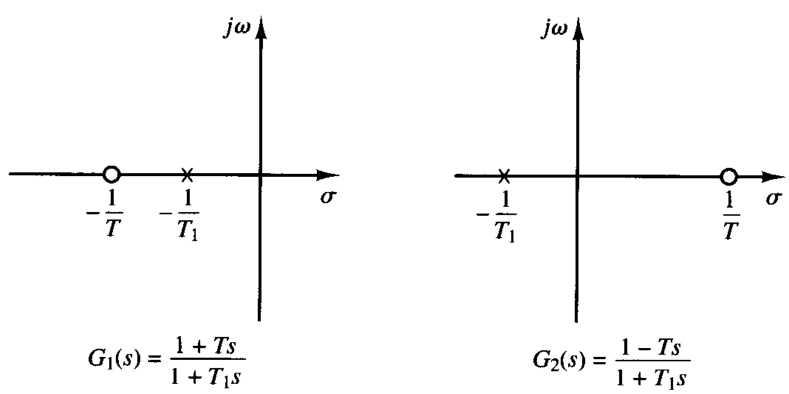


Figure 8–11 Pole–zero configurations of a minimum-phase system $G_1(s)$ and nonminimum-phase system $G_2(s)$.

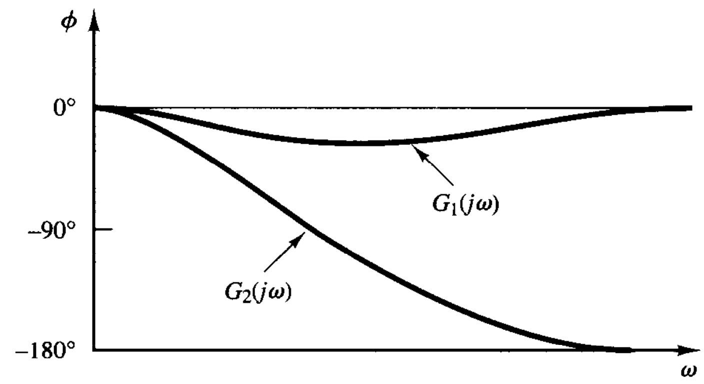


Figure 8–12 Phase-angle characteristics of the systems $G_1(s)$ and $G_2(s)$ shown in Figure 8–11.

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.

Nonminimum-phase systems are slow in response because of their faulty behavior at the start of 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 system is transport lag.)

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 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 8–13.

EXAMPLE 8–2 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

$$\begin{aligned} 20 \log|G(j\omega)| &= 20 \log|e^{-j\omega L}| + 20 \log\left|\frac{1}{1 + j\omega T}\right| \\ &= 0 + 20 \log\left|\frac{1}{1 + j\omega T}\right| \end{aligned}$$

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

$$\begin{aligned} \underline{/G(j\omega)} &= \underline{/e^{-j\omega L}} + \underline{\Big/\frac{1}{1 + j\omega T}} \\ &= -\omega L - \tan^{-1} \omega T \end{aligned}$$

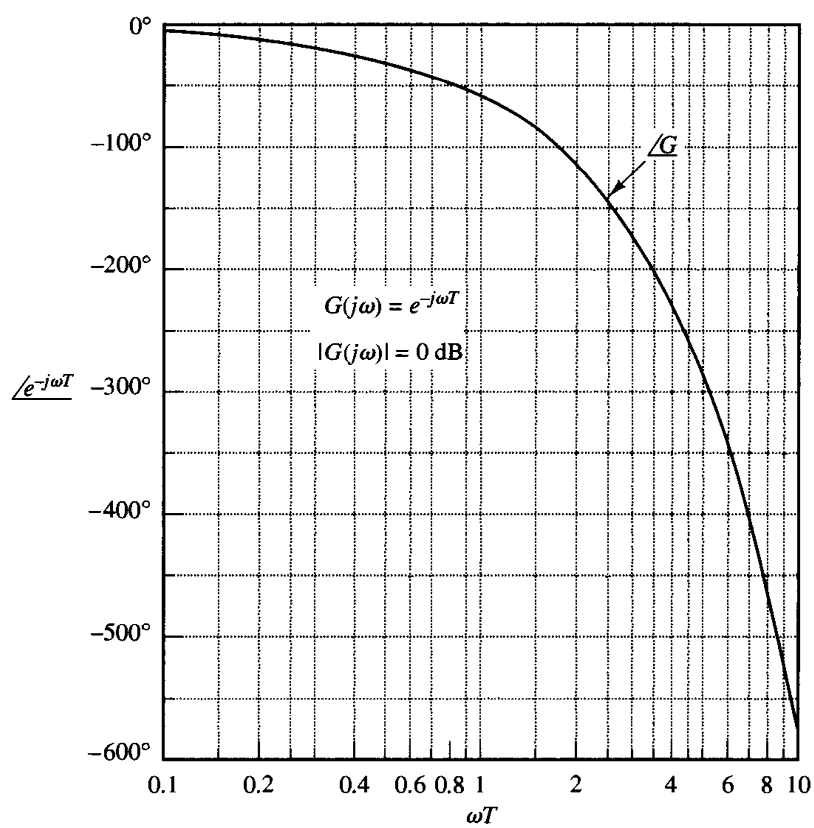


Figure 8–13 Phase-angle characteristic of transport lag.

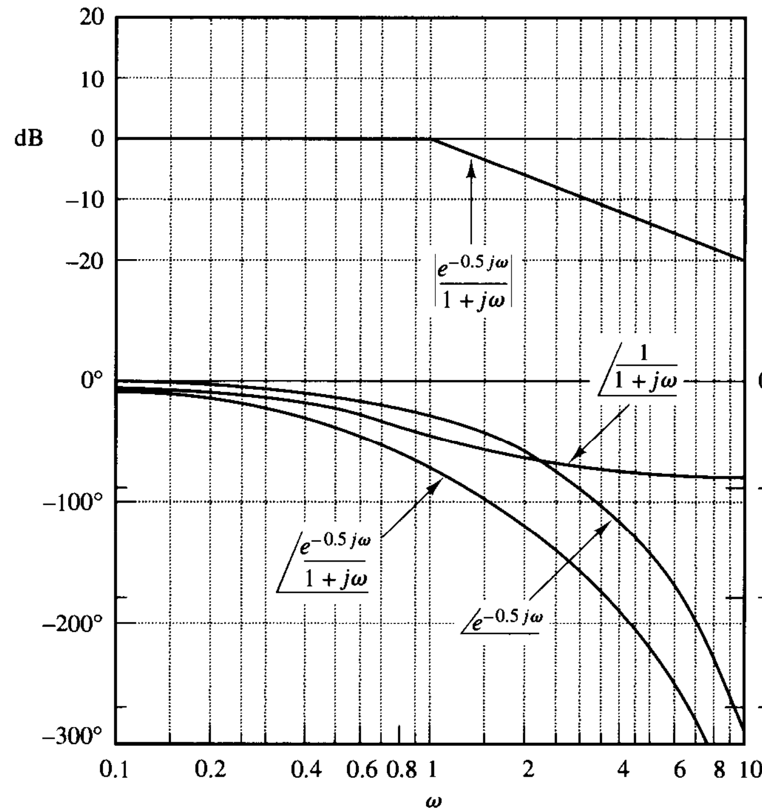


Figure 8–14
Bode diagram for the system $e^{-j\omega L}/(1 + j\omega T)$ with $L = 0.5$ and $T = 1$.

The log-magnitude and phase-angle curves for this transfer function with $L = 0.5$ and $T = 1$ are shown in Figure 8–14.

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 8–15. Assume that the open-loop transfer function is given by

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

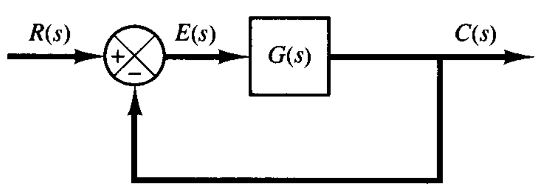


Figure 8–15
Unity-feedback control system.

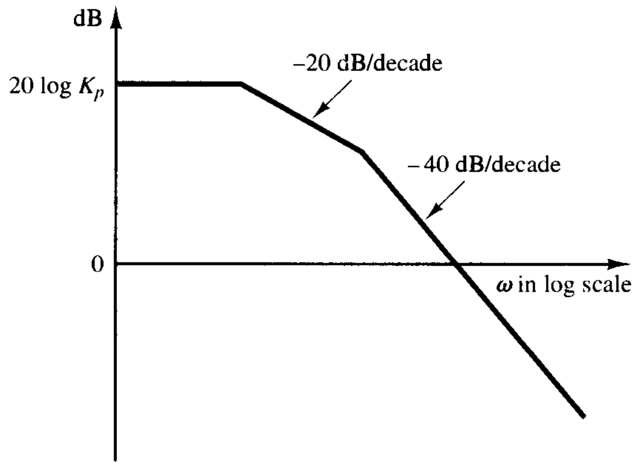


Figure 8–16
Log-magnitude curve of a type 0 system.

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 8–16 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

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

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

Determination of static velocity error constants. Consider the unity-feedback control system shown in Figure 8–15. Figure 8–17 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

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

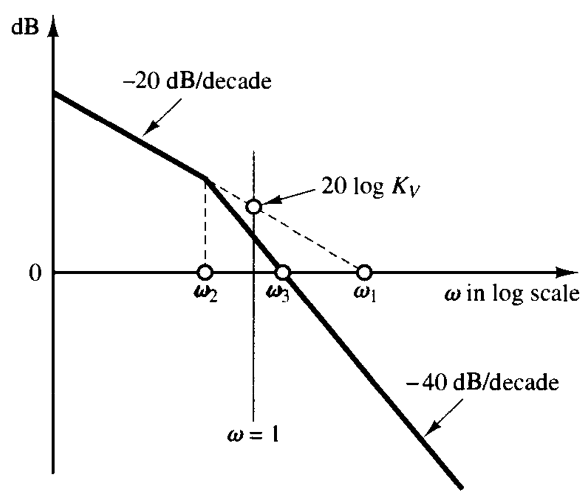


Figure 8–17
Log-magnitude curve of a type 1 system.

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

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

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 8–15. Figure 8–18 shows an example of the log-magnitude plot of a type 2 system. The intersection of the initial -40-dB/decade

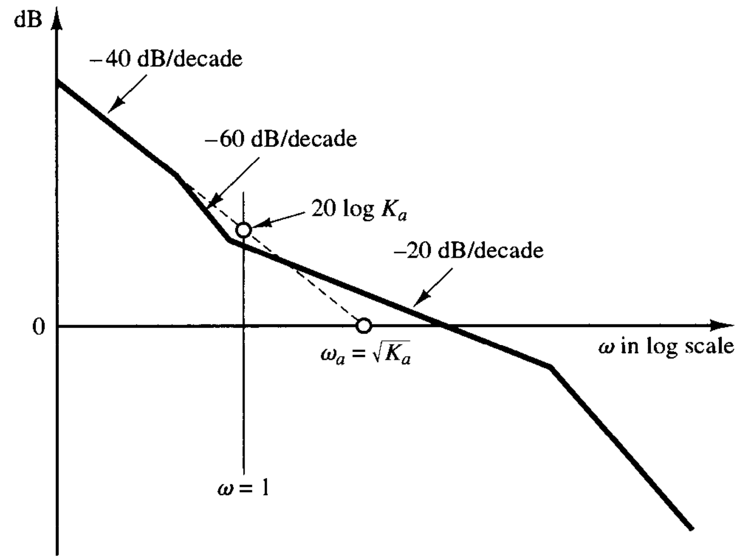


Figure 8–18 Log-magnitude curve of a type 2 system.

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

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

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.

When invoked with left-hand arguments,

```
[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)
```

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}. That is, to generate 50 points between 0.1 rad/sec and 100 rad/sec, enter the command

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

logspace(d1,d2,n) generates n points logarithmically equally spaced between decades 10^{d1} and 10^{d2}. 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 these frequency points when plotting Bode diagrams, use the command bode(num,den,w) or bode(A,B,C,D,iu,w). These commands use the user-specified frequency vector w.

EXAMPLE 8–3 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 8–19.

MATLAB Program 8–1

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

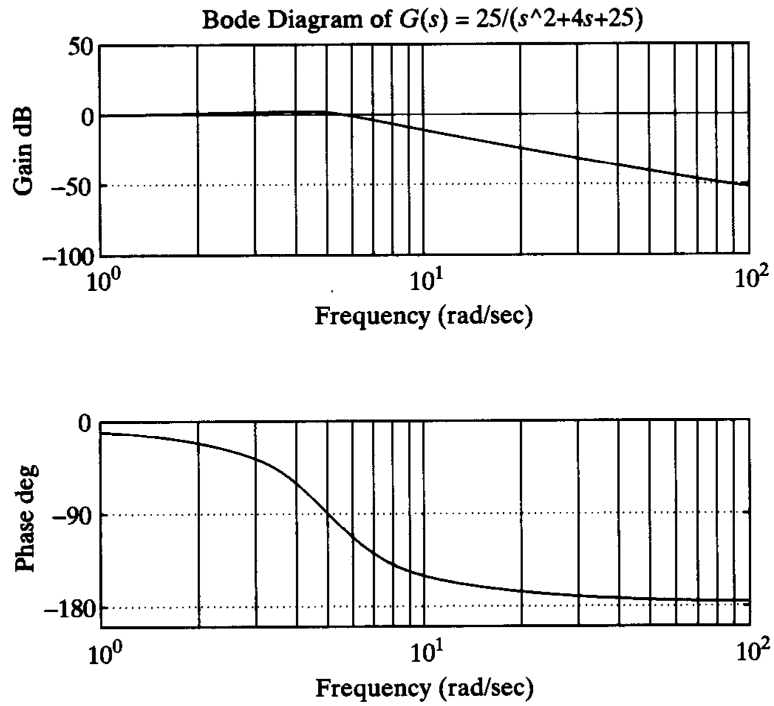


Figure 8–19
Bode diagram of
$$G(s) = \frac{25}{s^2 + 4s + 25}.$$

EXAMPLE 8–4 Consider the system shown in Figure 8–20. 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 8–21. The frequency range in this case is automatically determined to be from 0.1 to 10 rad/sec.

MATLAB Program 8–2

```
num = [0   9   1.8   9];
den = [1   1.2   9   0];
bode(num,den)
subplot(2,1,1);
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)
```

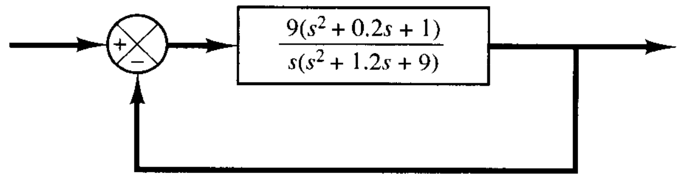


Figure 8–20
Control system.

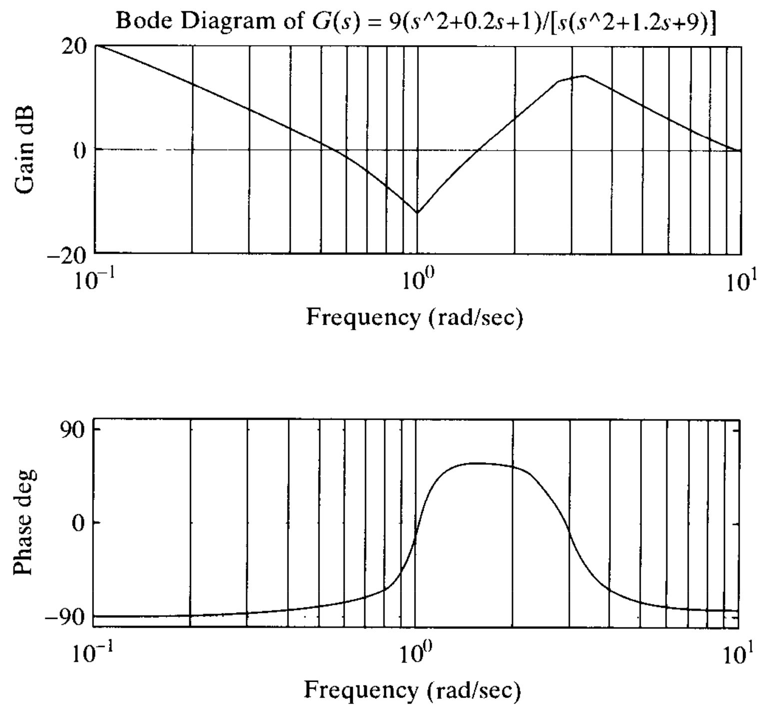


Figure 8–21
Bode diagram of $G(s) = \dfrac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}$.

This command generates 100 points logarithmically equally spaced between 0.01 and 1000 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)
```

then the frequency range is as 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 8–22.

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

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)
subplot(2,1,1);
title('Bode Diagram of G(s) = 9(s^2 + 0.2s + 1)/[s(s^2 + 1.2s + 9)]')
```

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

If we wish to specify the magnitude range to be, for example, at least between −45 dB and +45 dB, then enter invisible lines at −45 dB and +45 dB in the plot by specifying dBmax (maximum magnitude) and dBmin (minimum magnitude) as follows:

```
dBmax = 45*ones(1,100);
dBmin = -45*ones(1,100);
```

Then enter the following semilog plot command:

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

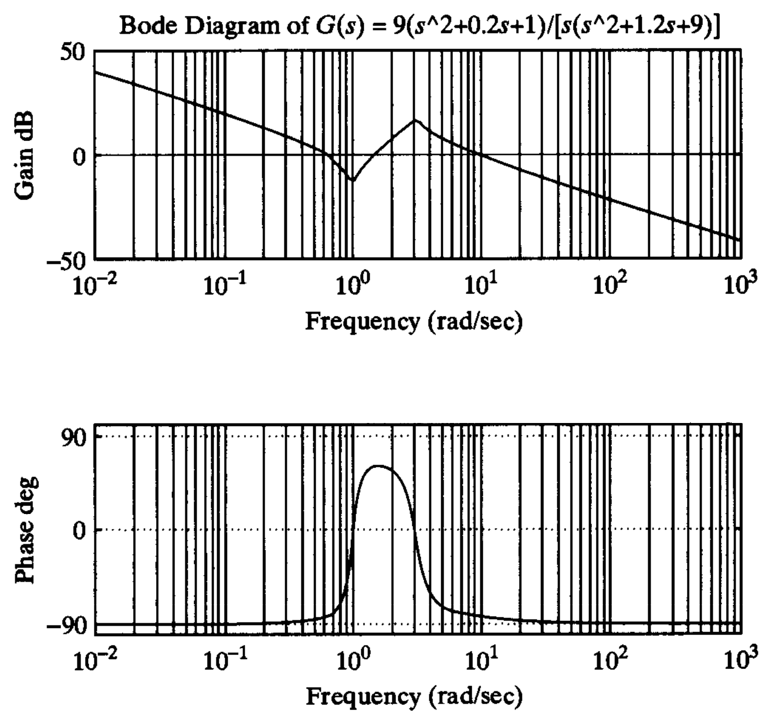


Figure 8–22 Bode diagram of $G(s) = \dfrac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}$.

(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. (Straight lines at +45 dB and −45 dB are invisible.)

Note that 'i' is an invisible color. For example, 'og' will show small circles in green color and 'oi' will show small circles in 'invisible' color: that is, you will not see small circles in the screen. By changing a portion of the preceding semilogx command from

```
w,dBmax,'--i',w,dBmin,':i'
```

to

```
w,dBmax,'--',w,dBmin,':'
```

the +45-dB line and −45-dB line will become visible on the screen.

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.

For the phase angle, if we wish to specify the range to be, for example, at least between −145° and +115°, we enter invisible lines at −145° and +115° in the program by specifying pmax (maximum phase angle) and pmin (minimum phase angle) as follows:

```
pmax = 115*ones(1,100)
pmin = −145*ones(1,100)
```

Then enter the semilog plot command:

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

(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. Straight lines at +115° and −145° are invisible.

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 8–23 shows the Bode diagram obtained by use of MATLAB Program 8–4.

What happens to the Bode diagram if the gain becomes infinite at a certain frequency point? If there is a system pole on the $j\omega$ axis and the w vector happens to contain this frequency point, the gain becomes infinite at this frequency. In such a case, MATLAB produces warning messages. Consider the following example.

MATLAB Program 8–4

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

% ***** In this program we shall obtain Bode diagram of
% transfer-function system using user-specified frequency
% range *****

% ***** 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 -45 dB
% and +45 dB. Enter dBmax and dBmin in the program and
% draw dBmax line and dBmin line in invisible color. To
% plot the magdB curve and invisible lines, enter the
% following dBmax, dBmin, and semilogx command *****

dBmax = 45*ones(1,100);
dBmin = -45*ones(1,100);
semilogx(w,magdB,'o',w,magdB,'-',w,dBmax,'--i',w,dBmin,':i')

% ***** 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 -145 degrees
% and +115 degrees. Enter pmax and pmin in the program and
% draw pmax line and pmin line in invisible color. To plot
% the phase curve and invisible lines, enter the following
% pmax, pmin, and semilogx command *****
```

```
pmax = 115*ones(1,100);
pmin = -145*ones(1,100);
semilogx(w,phase,'o',w,phase,'-',w,pmax,'--i',w,pmin,':i')

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

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

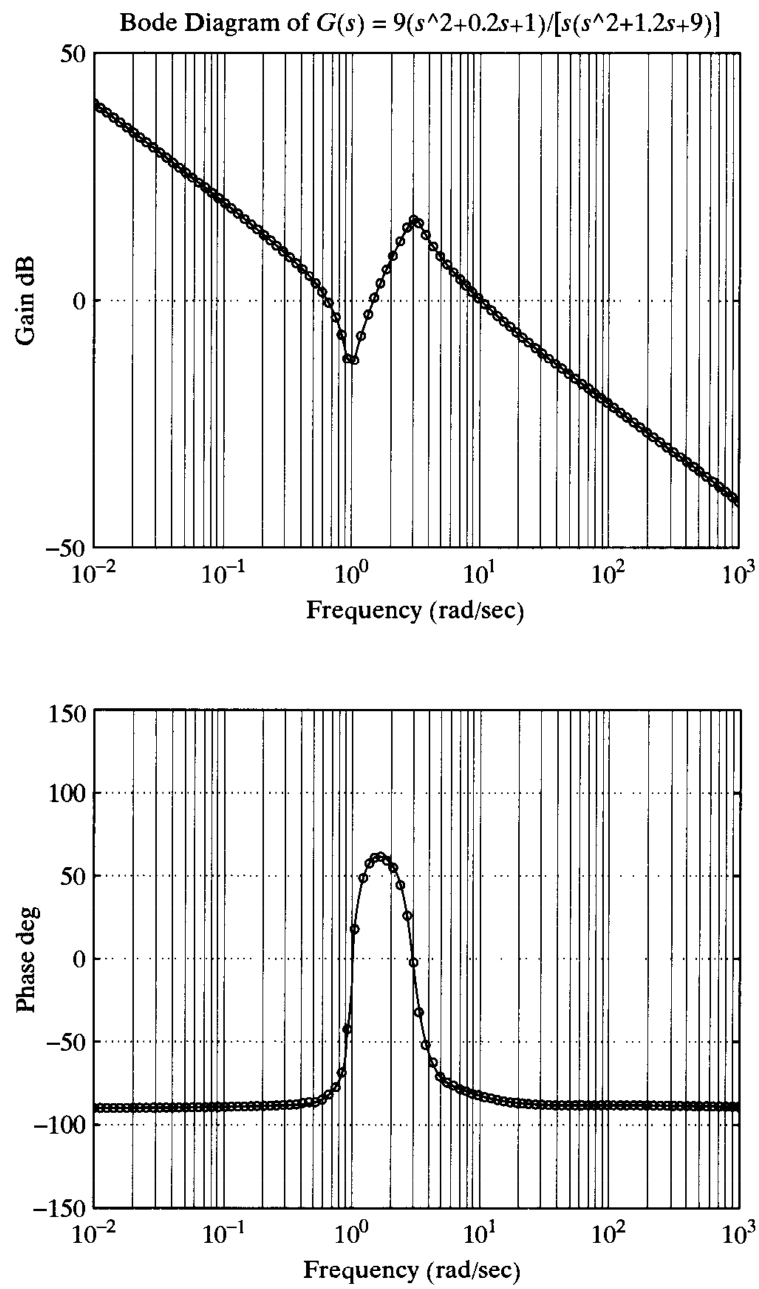


Figure 8–23 Bode diagram of $G(s) = \dfrac{9(s^2 + 0.2s + 1)}{s(s^2 + 1.2s + 9)}$.

EXAMPLE 8–5 Consider a system with the following open-loop transfer function:

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

This open-loop transfer function has poles on the $j\omega$ axis at $\pm j$.

MATLAB Program 8–5 may be used to plot the Bode diagram for this system. The resulting plot is shown in Figure 8–24. Theoretically, the magnitude becomes infinite at a frequency point where $\omega = 1$ rad/sec. However, this frequency point is not among the computing frequency points. In the plot the peak magnitude is shown to be approximately 50 dB. This value is computed near, but not exactly at, $\omega = 1$ rad/sec.

MATLAB Program 8–5

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

If, however, one of the computing frequency points coincides with the pole at $\omega = 1$, then the magnitude becomes infinite at this point. MATLAB sends out warning messages. See MATLAB

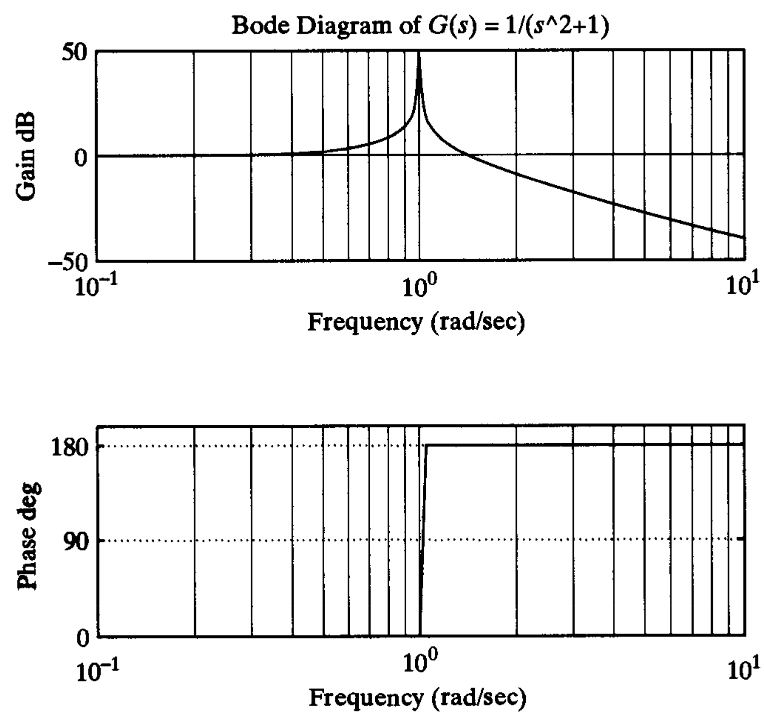


Figure 8–24 Bode diagram of $G(s) = \frac{1}{s^2 + 1}$.

Program 8–6 where computing points include the point at $\omega = 1$ rad/sec. (There are 101 computing points in this case. The computing points range from $\omega = 0.1$ to $\omega = 10$. The fifty-first point is at $\omega = 1$.) When MATLAB Program 8–6 is entered into the computer, warning messages appear, as shown. The resulting Bode diagram, shown in Figure 8–25, does not include the computed magnitude at $\omega = 1$. (Theoretically, this magnitude is infinite.) The magnitude curve shows the peak value at about 20 dB. The phase curve shows a gradual change in the phase angle from 0° to +180 ° near point $\omega = 1$. (Theoretically, the change in the phase angle from 0° to +180° should be abrupt at $\omega = 1$.) Obviously, the Bode diagram shown in Figure 8–25 is incorrect.

If the w vector contains such a frequency point, where the gain becomes infinite, change the number of frequency points, for example, from 101 to 100. Normally, a small change in the number of frequency points will avoid this kind of problem.

MATLAB Program 8–6

```
num = [0  0  1];
den = [1  0  1];
w = logspace(-1,1,101);
bode(num,den,w)

Warning: Divide by zero
subplot(2,1,1);
title('Incorrect Bode Diagram')
```

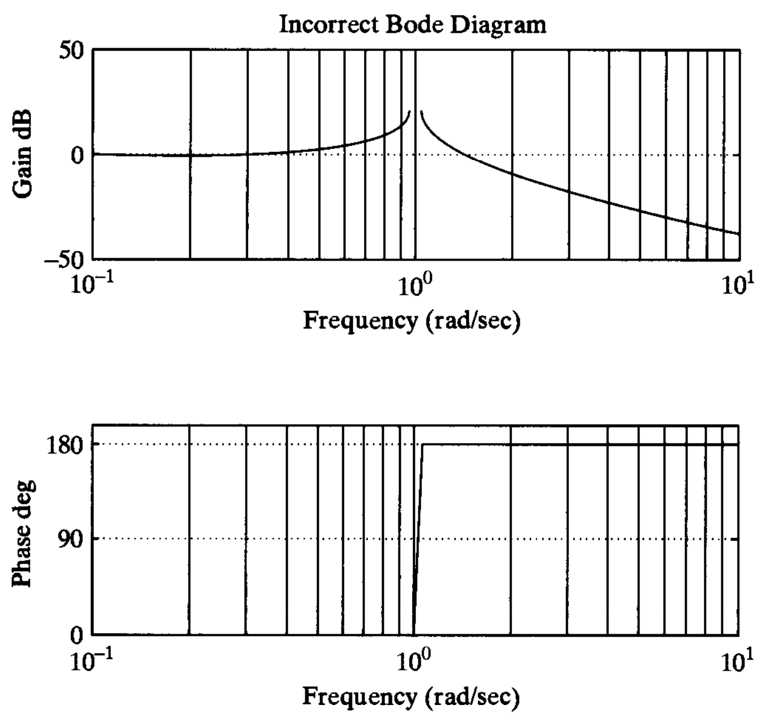


Figure 8–25 Incorrect Bode diagram of $G(s) = \dfrac{1}{s^2 + 1}$.

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

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

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 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 8–6 Consider the following system:

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

$$y = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

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

bode(A,B,C,D)

and entering MATLAB Program 8–7 into the computer, we obtain the Bode diagram shown in Figure 8–26.

MATLAB Program 8–7

```
A = [0   1;-25   -4];
B = [0;25];
C = [1   0];
D = [0];
bode(A,B,C,D)
subplot(2,1,1);
title('Bode Diagram')
```

If we replace the command bode(A,B,C,D) in MATLAB Program 8–7 with

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

then MATLAB will produce the Bode diagram identical to that shown in Figure 8–26.

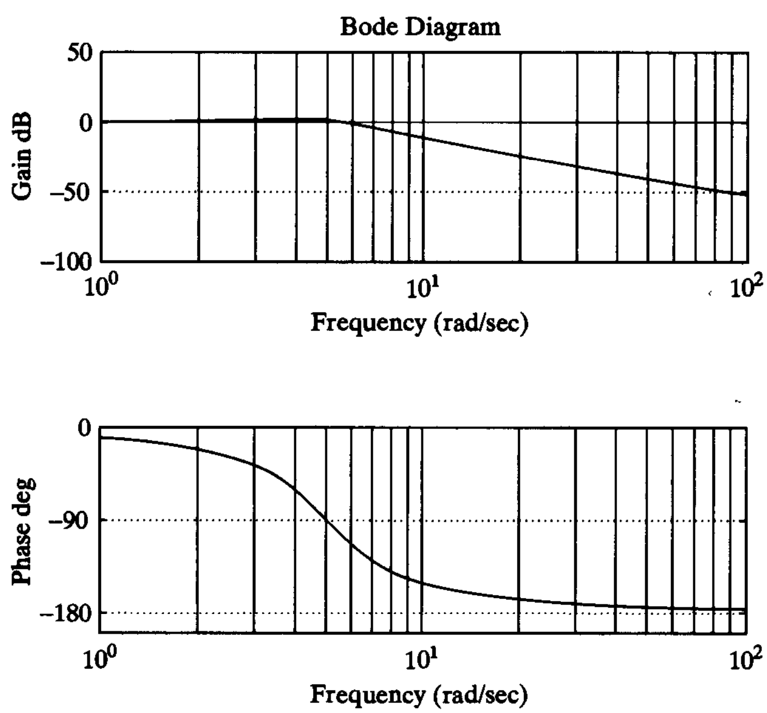


Figure 8–26
Bode diagram of the system considered in Example 8–6.

Note that if we use, by mistake, the command

```
bode(A,B,C,D,2)
```

MATLAB produces an error message, because the present system has only one input and iu should be set to '1', not '2' or any other number.

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 8–27. 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. Both the magnitude $|G(j\omega)|$ and phase angle $\underline{/G(j\omega)}$ must be calculated directly for each frequency ω in order to construct polar plots. Since the logarithmic plot is easy to construct, however, the data necessary for plotting the polar plot may be obtained directly from the logarithmic plot if the latter is drawn first and decibels are converted into ordinary magnitude. Or, of course, MATLAB may be used to obtain a polar plot $G(j\omega)$ or to obtain $|G(j\omega)|$ and $\underline{/G(j\omega)}$ 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.

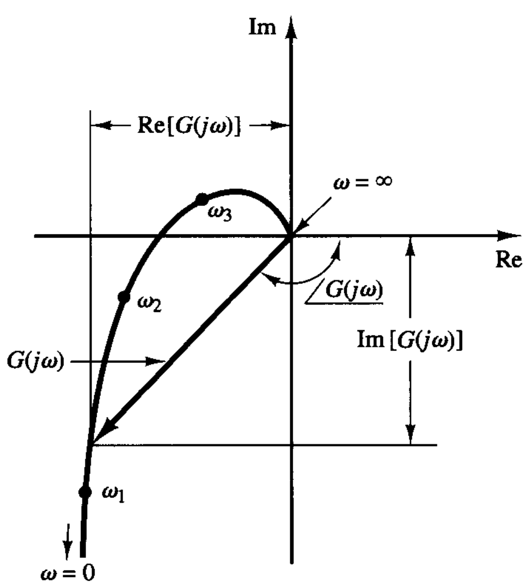


Figure 8–27
Polar plot.

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

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

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

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

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

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

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 8–28(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 is a semicircle, define

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

where

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

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

Then we obtain

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

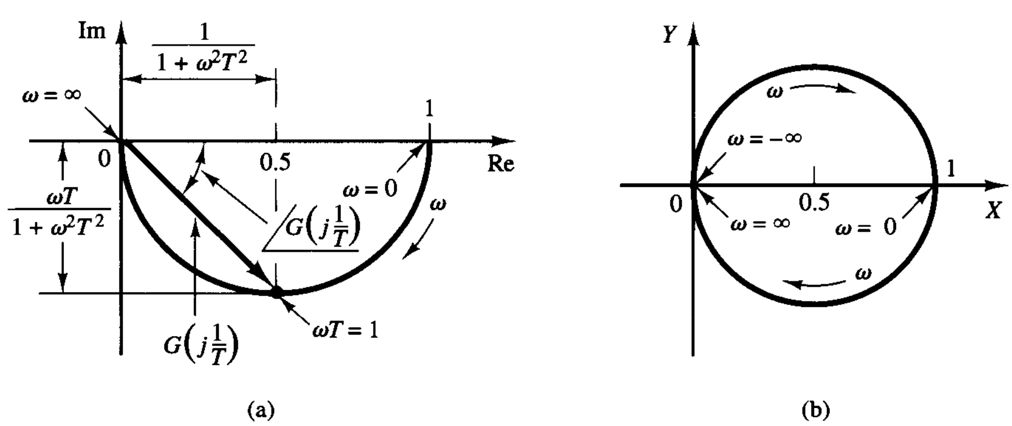


Figure 8–28
(a) Polar plot of $1/(1 + j\omega T)$; (b) plot of $G(j\omega)$ in X–Y plane.

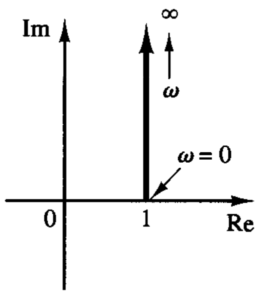


Figure 8–29
Polar plot of $1 + j\omega T$.

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 8–28(b). The lower semicircle corresponds to $0 \leq \omega \leq \infty$, and the upper semicircle corresponds to $-\infty \leq \omega \leq 0$.

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 8–29. 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 lot 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

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

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. The values of $G(j\omega)$ in the frequency range of interest can be calculated directly or by use of the Bode diagram or by use of MATLAB.

Examples of polar plots of the transfer function just considered are shown in Figure 8–30. 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

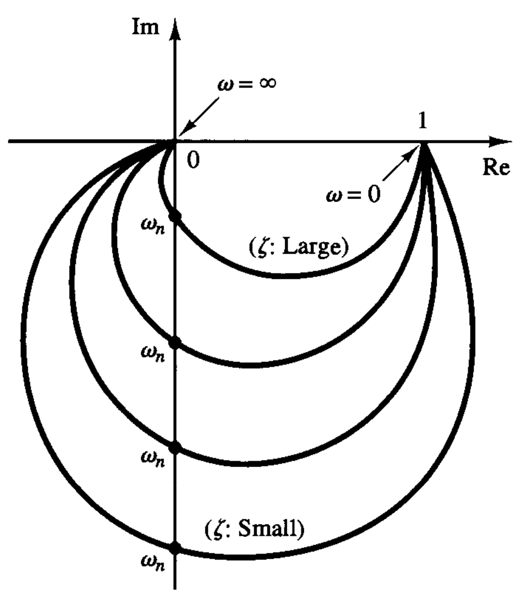


Figure 8–30
Polar plots of
$$\frac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2},$$
for $\zeta > 0$.

$G(j\omega)$ 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 8–31.

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.

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 8–32. The phase angle is between 0° and 180°.

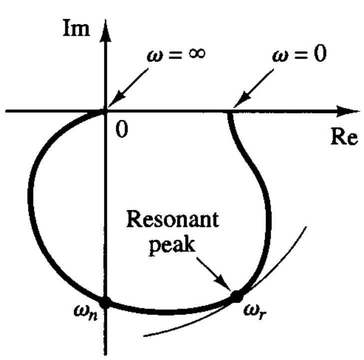


Figure 8–31
Polar plot showing the resonant peak and resonant frequency ω_r.

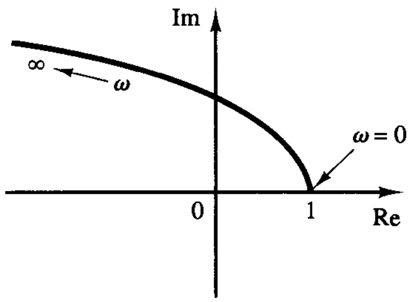


Figure 8–32
Polar plot of $1 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + \left(j\frac{\omega}{\omega_n}\right)^2$, for $\zeta > 0$.

EXAMPLE 8–7 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

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

the low-frequency portion of the polar plot becomes

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

and the high-frequency portion becomes

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

The general shape of the polar plot of $G(j\omega)$ is shown in Figure 8–33. 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.

Transport lag. The transport lag

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

can be written

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

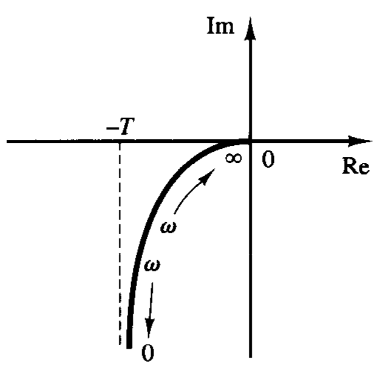


Figure 8–33
Polar plot of $1/[j\omega(1 + j\omega T)]$.

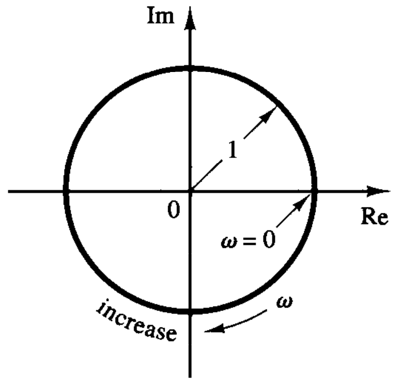


Figure 8–34
Polar plot of transport lag.

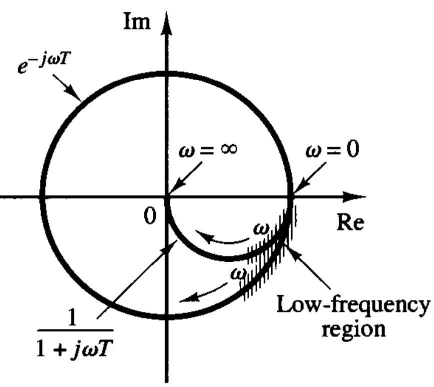


Figure 8–35
Polar plots of $e^{-j\omega T}$ and $1/(1 + j\omega T)$.

Since the magnitude of $G(j\omega)$ is always unity and the phase angle varies linearly with ω, the polar plot of the transport lag is a unit circle, as shown in Figure 8–34.

At low frequencies, the transport lag $e^{-j\omega T}$ and the first-order lag $1/(1 + j\omega T)$ behave similarly, as shown in Figure 8–35. The polar plots of $e^{-j\omega T}$ and $1/(1 + j\omega T)$ are tangent to each other at $\omega = 0$. This may be seen from the fact that, for $\omega \ll 1/T$,

$$e^{-j\omega T} \doteqdot 1 - j\omega T \qquad \text{and} \qquad \frac{1}{1 + j\omega T} \doteqdot 1 - j\omega T$$

For $\omega \gg 1/T$, however, an essential difference exists between $e^{-j\omega T}$ and $1/(1 + j\omega T)$, as may also be seen from Figure 8–35.

EXAMPLE 8–8 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

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

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

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

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 8–36.

General shapes of polar plots. The polar plots of a transfer function of the form

$$\begin{aligned} G(j\omega) &= \frac{K(1 + j\omega T_a)(1 + j\omega T_b)\cdots}{(j\omega)^\lambda(1 + j\omega T_1)(1 + j\omega T_2)\cdots} \\ &= \frac{b_0(j\omega)^m + b_1(j\omega)^{m-1} + \cdots}{a_0(j\omega)^n + a_1(j\omega)^{n-1} + \cdots} \end{aligned}$$

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 $0 \le \omega \le \infty$. 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 $0 \le \omega \le \infty$. 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 8–37. 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 $G(j\omega)$

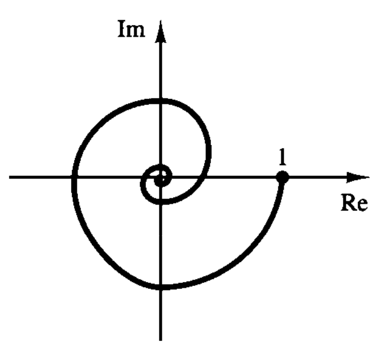


Figure 8–36
Polar plot of $e^{-j\omega L}/(1 + j\omega T)$.

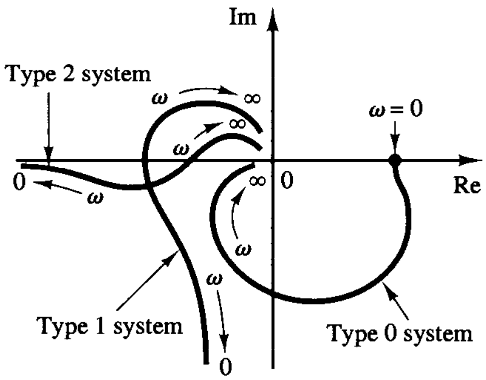
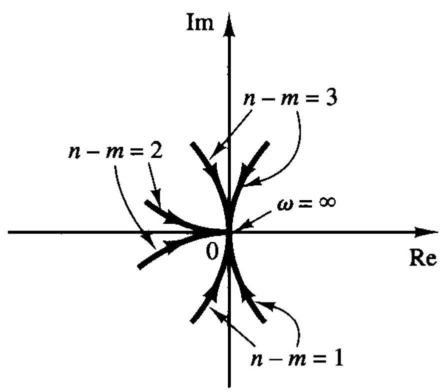


Figure 8–37
Polar plots of type 0, type 1, and type 2 systems.

loci converge to the origin clockwise. At $\omega = \infty$, the loci are tangent to one or the other axes, as shown in Figure 8–38.

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 8–39 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.

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



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

Figure 8–38
Polar plots in the high-frequency range.

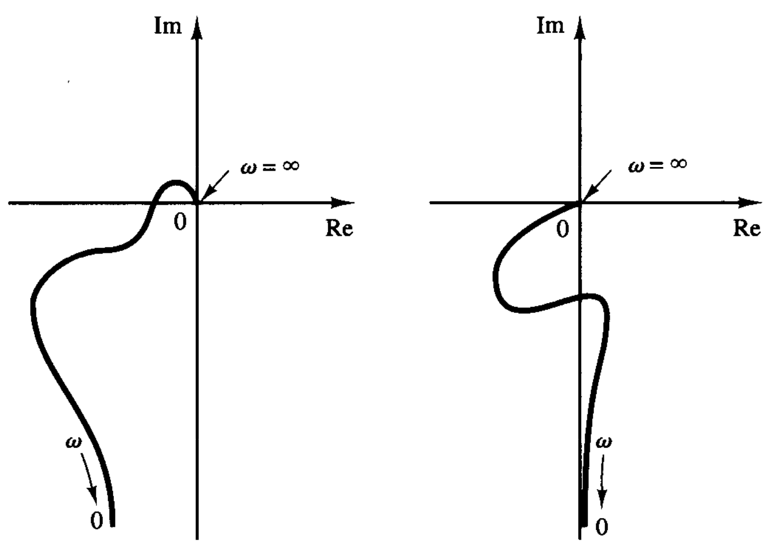
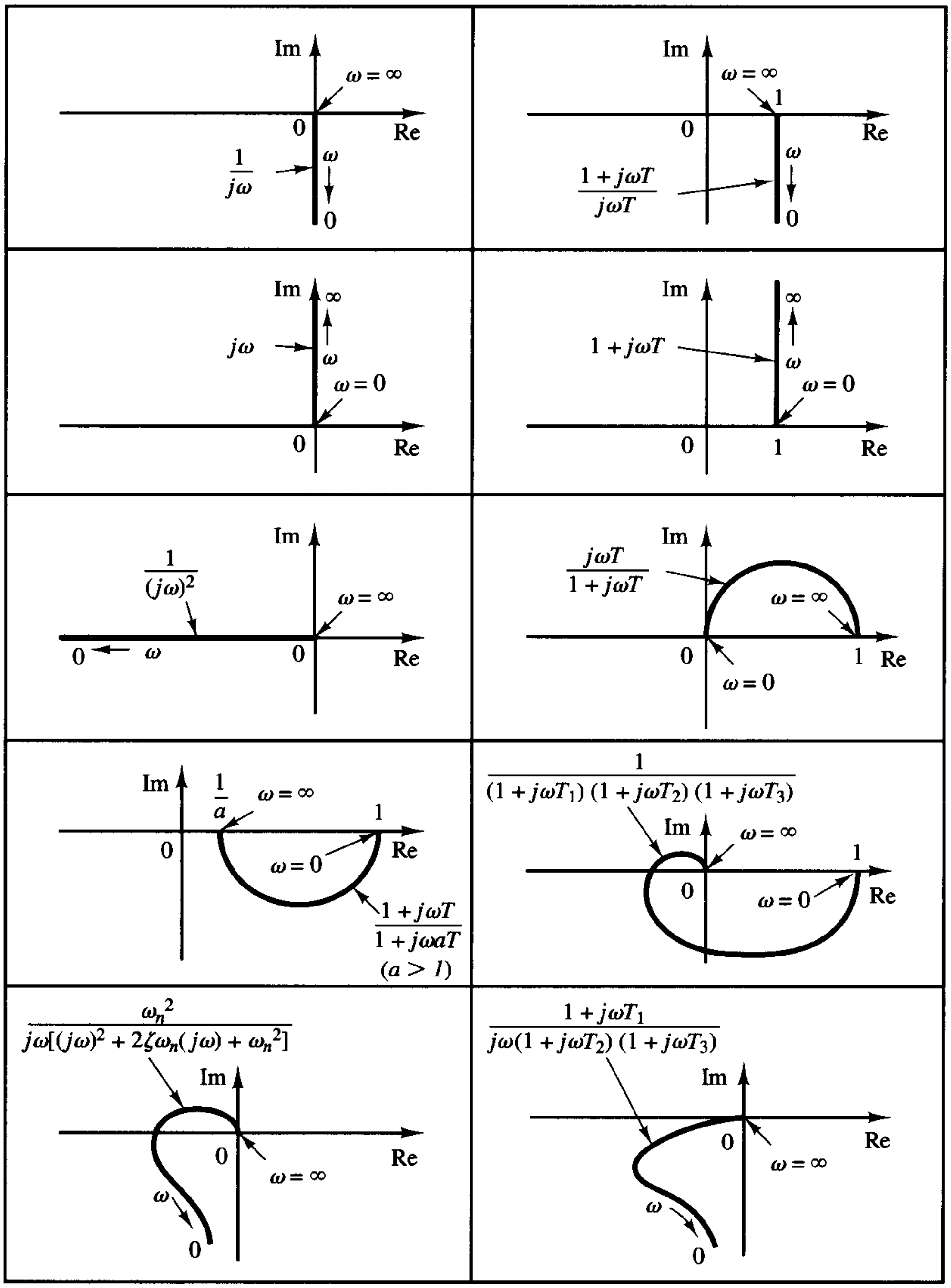


Figure 8–39
Polar plots of transfer functions with numerator dynamics.

Table 8–1 Polar Plots of Simple 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 convenient for a particular operation, but a given operation can always be carried out in either plot.

The 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.

The command

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

uses the user-specified frequency vector w. The vector w specifies the frequency points in radians per second at which the frequency response will be calculated.

When invoked with the left-hand arguments

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

or

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

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 8–9 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)
```

may be used to draw a Nyquist plot. MATLAB Program 8–8 produces the Nyquist plot shown in Figure 8–40. In this plot the ranges for the real axis and imaginary axis are automatically determined.

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:

```
MATLAB Program 8-8

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)')
```

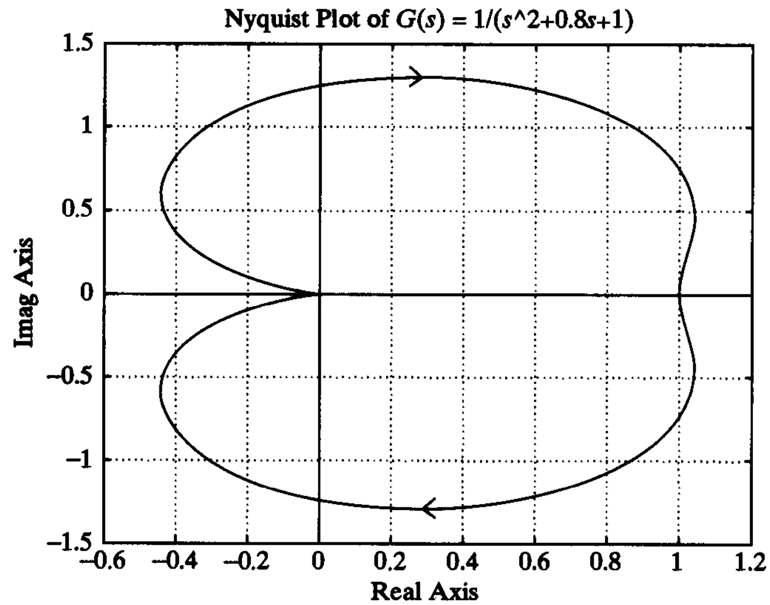


Figure 8-40 Nyquist plot of $G(s) = \dfrac{1}{s^2 + 0.8s + 1}$.

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

or, combining these two lines into one,

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

See MATLAB Program 8-9 and the resulting Nyquist plot shown in Figure 8-41.

```
MATLAB Program 8-9

% ---------- 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)')
```

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

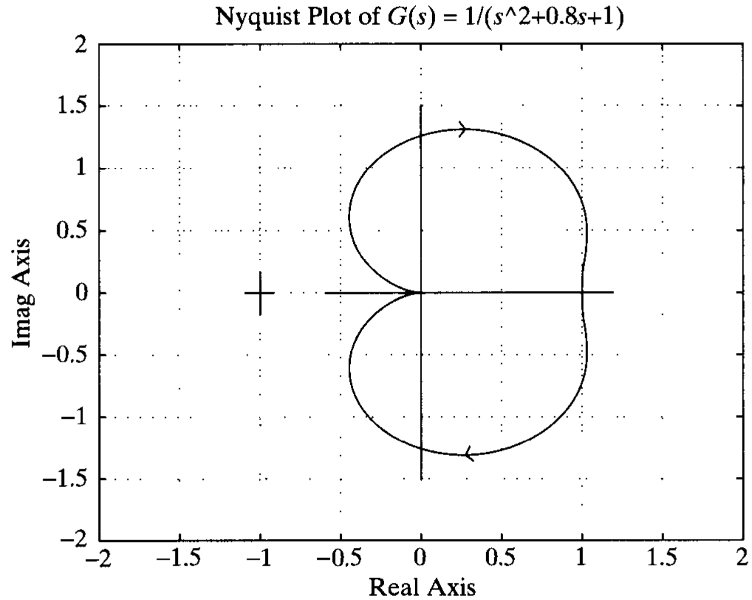


Figure 8–41
Nyquist plot of $G(s) = \dfrac{1}{s^2 + 0.8s + 1}$.

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

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 8–42. 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

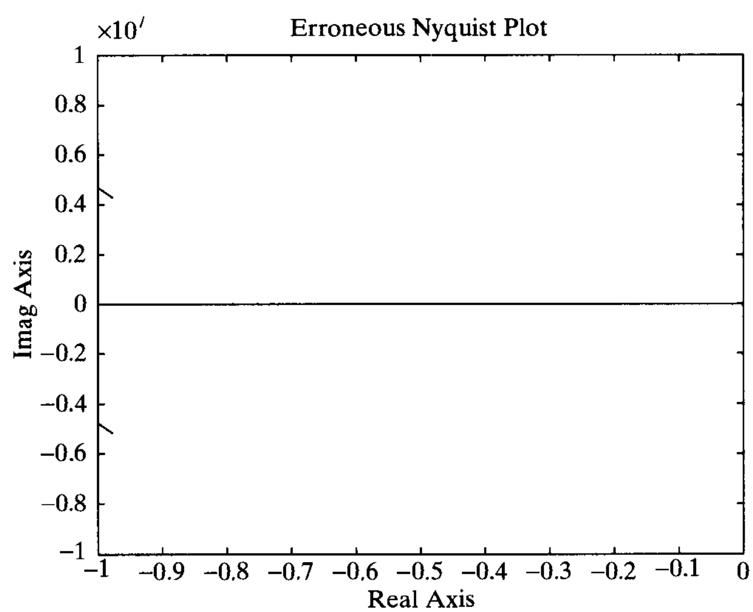


Figure 8–42
Erroneous Nyquist plot.

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

in the computer, then a correct Nyquist plot can be obtained. See Example 8–10.

EXAMPLE 8–10 Draw a Nyquist plot for the following $G(s)$:

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

MATLAB Program 8–10 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 8–43.

MATLAB Program 8–10

```
% --------- 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 8–43 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 8–11. The resulting Nyquist plot is presented in Figure 8–44.

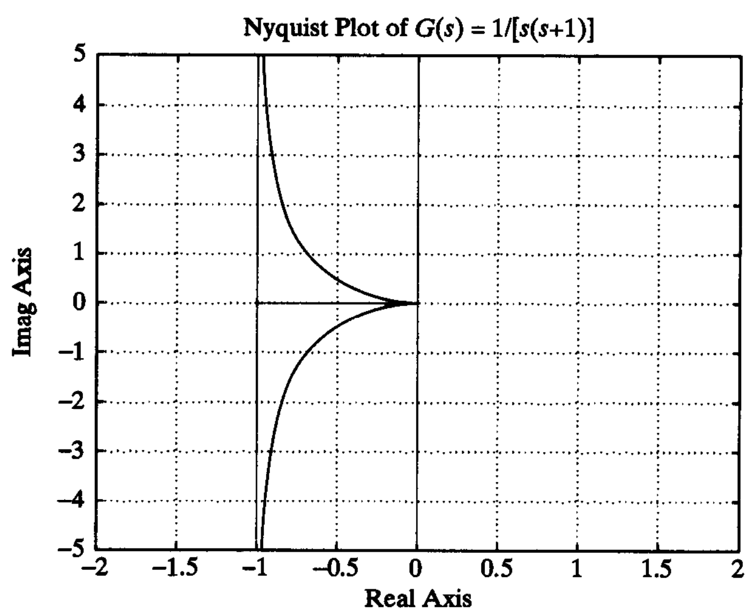


Figure 8–43 Nyquist plot of $G(s) = \frac{1}{s(s+1)}$.

MATLAB Program 8-11

```
% ---------- 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')
```

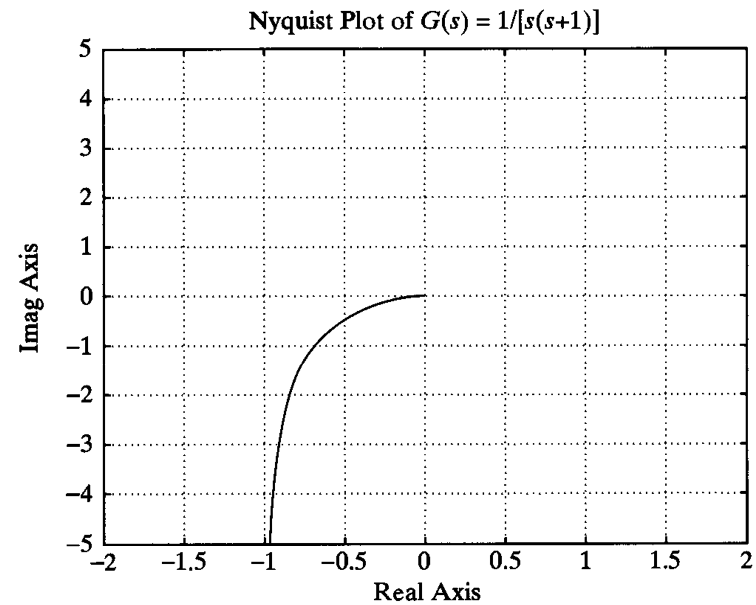


Figure 8–44 Nyquist plot of $G(s) = \dfrac{1}{s(s+1)}$.

Drawing Nyquist plots of a system defined in state space. Consider the system defined by

$$\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$$

$$\mathbf{y} = \mathbf{Cx} + \mathbf{Du}$$

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 use of the command

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

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 8–11 Consider the system defined by

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

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

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 8–12 will provide the Nyquist plot. (Note that we obtain the identical result by using either of these two commands.) Figure 8–45 shows the Nyquist plot produced by MATLAB Program 8–12.

MATLAB Program 8–12

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

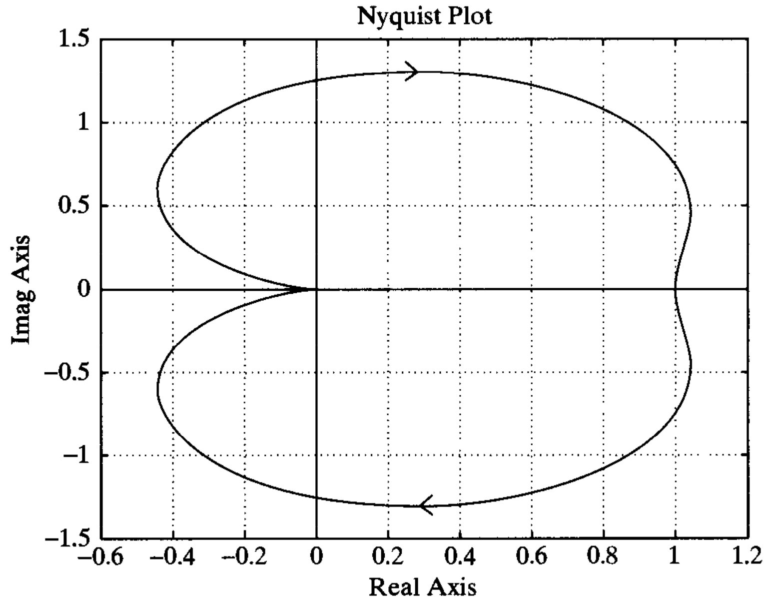


Figure 8–45 Nyquist plot of system considered in Example 8–11.

EXAMPLE 8–12 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 use of the command

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

MATLAB Program 8–13 produces the four Nyquist plots. They are shown in Figure 8–46.

MATLAB Program 8–13

```
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)
```

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

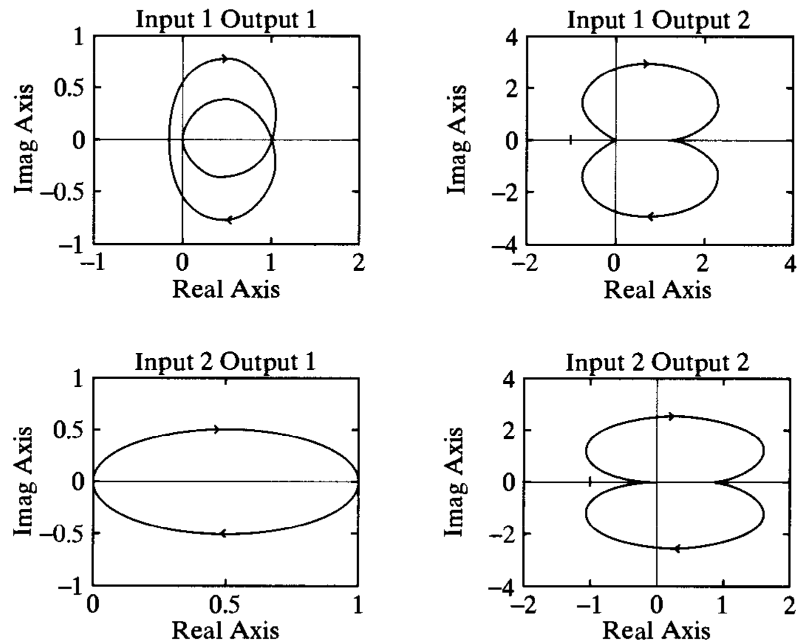


Figure 8–46 Nyquist plots of system considered in Example 8–12.

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. 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 plots for the sinusoidal transfer function $G(j\omega)$ and $1/G(j\omega)$ are skew symmetrical about the origin since

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

and

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

Figure 8–47 compares frequency-response curves of

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

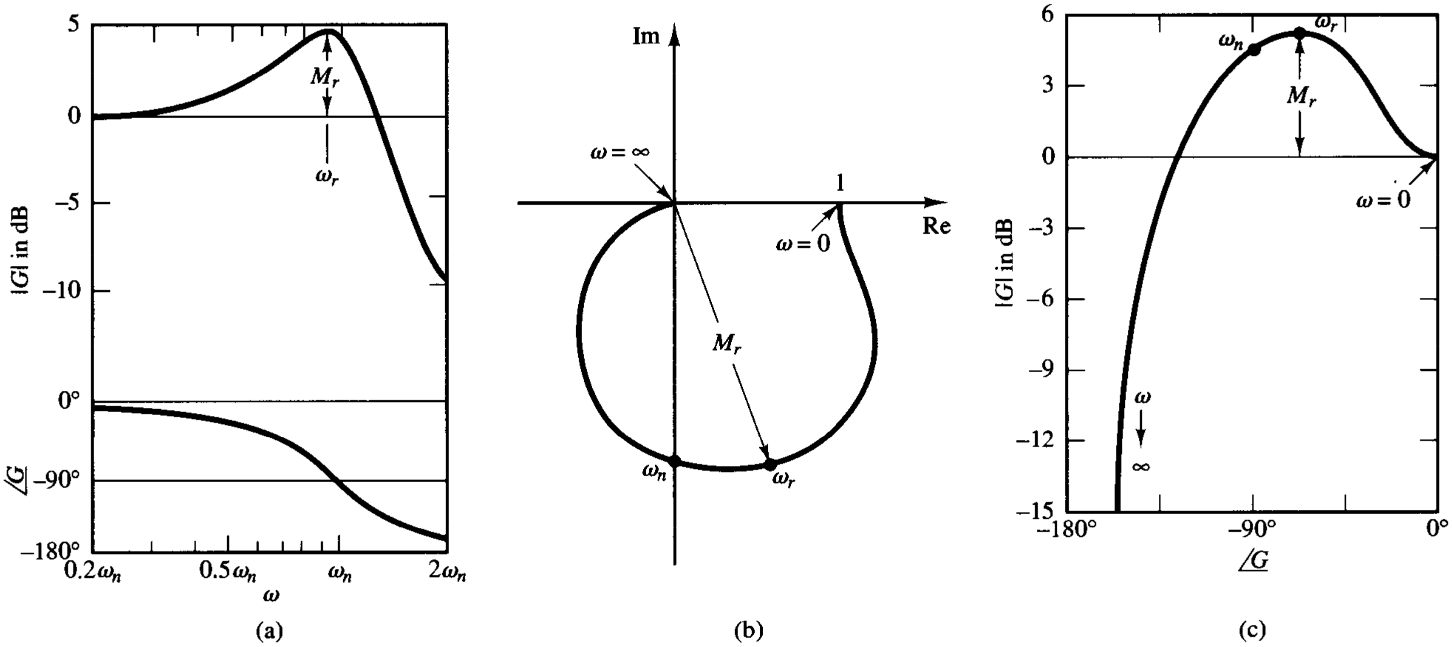


Figure 8–47
Three representations of the frequency response of $\dfrac{1}{1 + 2\zeta\left(j\dfrac{\omega}{\omega_n}\right) + \left(j\dfrac{\omega}{\omega_n}\right)^2}$, for $\zeta > 0$.
(a) Bode diagram; (b) polar plot; (c) log-magnitude versus phase plot.

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.

8–7 NYQUIST STABILITY CRITERION

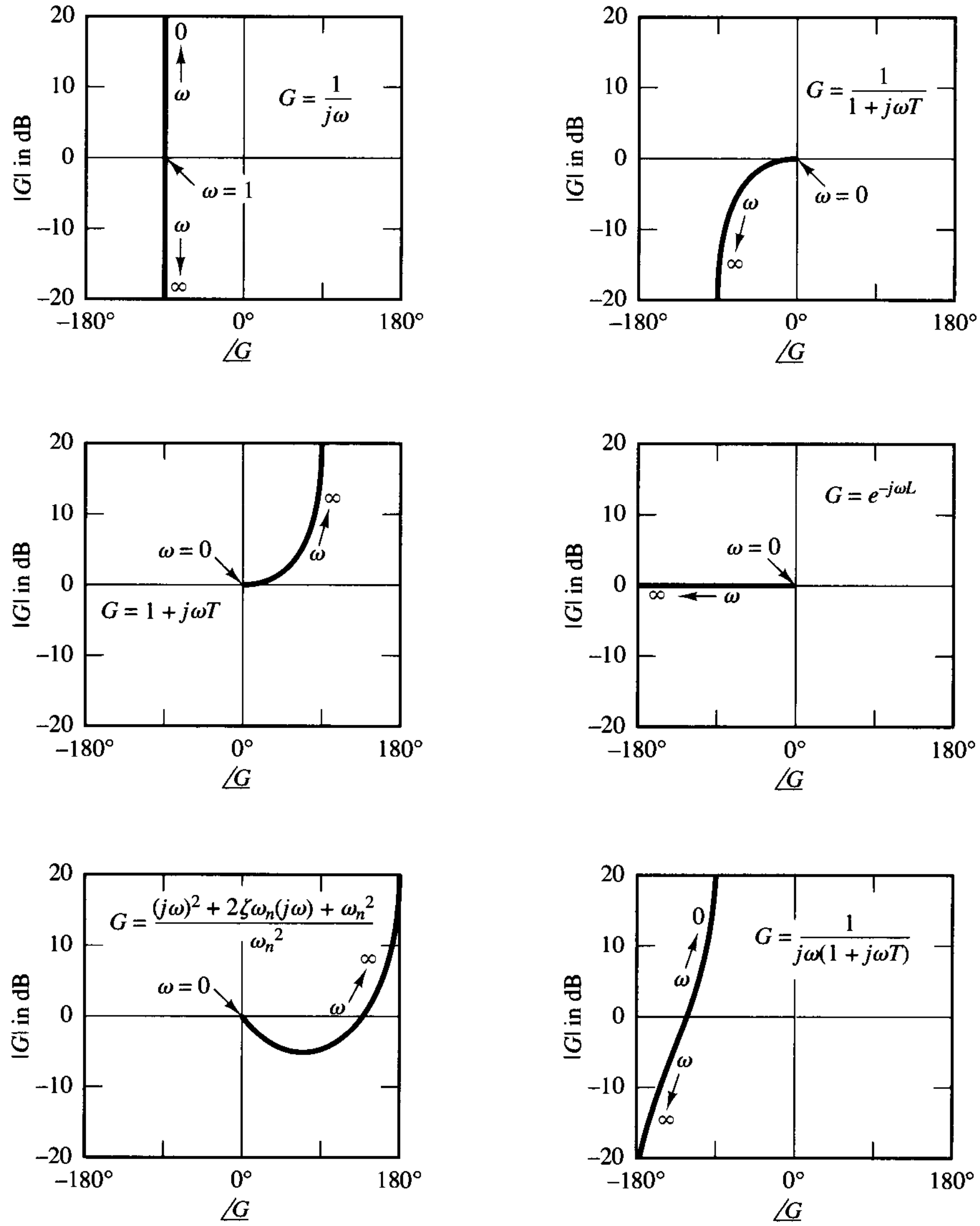
This section presents the Nyquist stability criterion and associated mathematical backgrounds. Consider the closed-loop system shown in Figure 8–48. The closed-loop transfer function is

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

For stability, all roots of the characteristic equation

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

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



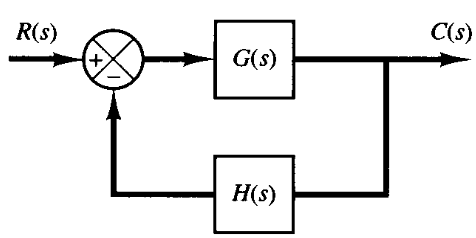


Figure 8–48
Closed-loop system.

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 8–48 is

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

We shall show that 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. The number and direction of encirclements of the origin of the $F(s)$ plane by the closed curve plays 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:

$$G(s)H(s) = \frac{6}{(s + 1)(s + 2)}$$

The characteristic equation is

$$F(s) = 1 + G(s)H(s) = 1 + \frac{6}{(s + 1)(s + 2)}$$

$$= \frac{(s + 1.5 + j2.4)(s + 1.5 - j2.4)}{(s + 1)(s + 2)} = 0$$

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 $s = 1 + j2$, then $F(s)$ becomes

$$F(1 + j2) = 1 + \frac{6}{(2 + j2)(3 + j2)} = 1.115 - j0.577$$

Thus, the point $s = 1 + j2$ in the s plane maps into the point $1.115 - j0.577$ 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. Figure 8–49 (a) shows conformal mappings of the lines $\omega = 0,1,2,3$ and the lines $\sigma = 1, 0, -1, -2, -3, -4$ in the upper-half s plane into the $F(s)$ plane. For example, the line $s = j\omega$ in the upper-half s plane ($\omega \geq 0$) maps into the curve denoted by $\sigma = 0$ in the $F(s)$ plane. Figure 8–49(b) shows conformal mappings of the lines $\omega = 0, -1, -2, -3$ and the lines $\sigma = 1, 0, -1, -2, -3, -4$ in the lower-half s plane into the $F(s)$ plane. Notice that for a given σ the curve for negative frequencies is symmetrical about the real axis with the curve for positive frequencies. Referring to Figures 8–49(a) and (b), we see that for the path $ABCD$ in the s plane traversed in the clockwise direction the corresponding curve in the $F(s)$ plane is $A'B'C'D'$. The arrows on the curves indicate directions of traversal. Similarly, the path $DEFA$ in the s plane maps into the curve $D'E'F'A'$ in the $F(s)$ plane. Because of the property of conformal mapping, the corresponding angles in the s plane and $F(s)$ plane are equal and have the same sense. [For example, since lines AB and BC intersect at right angles to each other in the s plane, curves $A'B'$ and $B'C'$ also intersect at right angles at point B' in the $F(s)$ plane.] Referring to Figure 8–49(c), we see that on the closed contour $ABCDEFA$ in the s plane the variable s starts at point A and assumes values on this path in a clockwise direction until it returns to the starting point A. The corresponding curve in the $F(s)$ plane is denoted $A'B'C'D'E'F'A'$. If we define the area to the right of the contour when a representative point s moves in the clockwise direction to be the inside of the contour and the area to the left to be the outside, then the shaded area in Figure 8–49(c) is enclosed by the contour $ABCDEFA$ and is inside it. From Figure 8–49(c), it can be seen that when the contour in the s plane encloses two poles of $F(s)$ the locus of $F(s)$ encircles the origin of the $F(s)$ plane twice in the counterclockwise direction.

The number of encirclements of the origin of the $F(s)$ plane depends on the closed contour in the s plane. If this contour encloses two zeros and two poles of $F(s)$, then the corresponding $F(s)$ locus does not encircle the origin, as shown in Figure 8–49(d). If this contour encloses only one zero, the corresponding locus of $F(s)$ encircles the origin once in the clockwise direction. This is shown in Figure 8–49(e). Finally, if the closed contour in the s plane encloses neither zeros nor poles, then the locus of $F(s)$ does not encircle the origin of the $F(s)$ plane at all. This is also shown in Figure 8–49(e).

Note that for each point in the s plane, except for the singular points, there is only one corresponding point in the $F(s)$ plane; that is, the mapping from the s plane into the $F(s)$ plane is one to one. The mapping from the $F(s)$ plane into the s plane may not be one to one, however, so that a given point in the $F(s)$ plane may correspond to more than one point in the s plane. For example, point B' in the $F(s)$ plane in Figure 8–49(d) corresponds to both point $(-3, 3)$ and point $(0, -3)$ in the s plane.

From the foregoing analysis, we can see that the direction of encirclement of the origin of the $F(s)$ plane 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 k zeros and k poles ($k = 0, 1, 2, \ldots$), that is, an equal number of each, 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

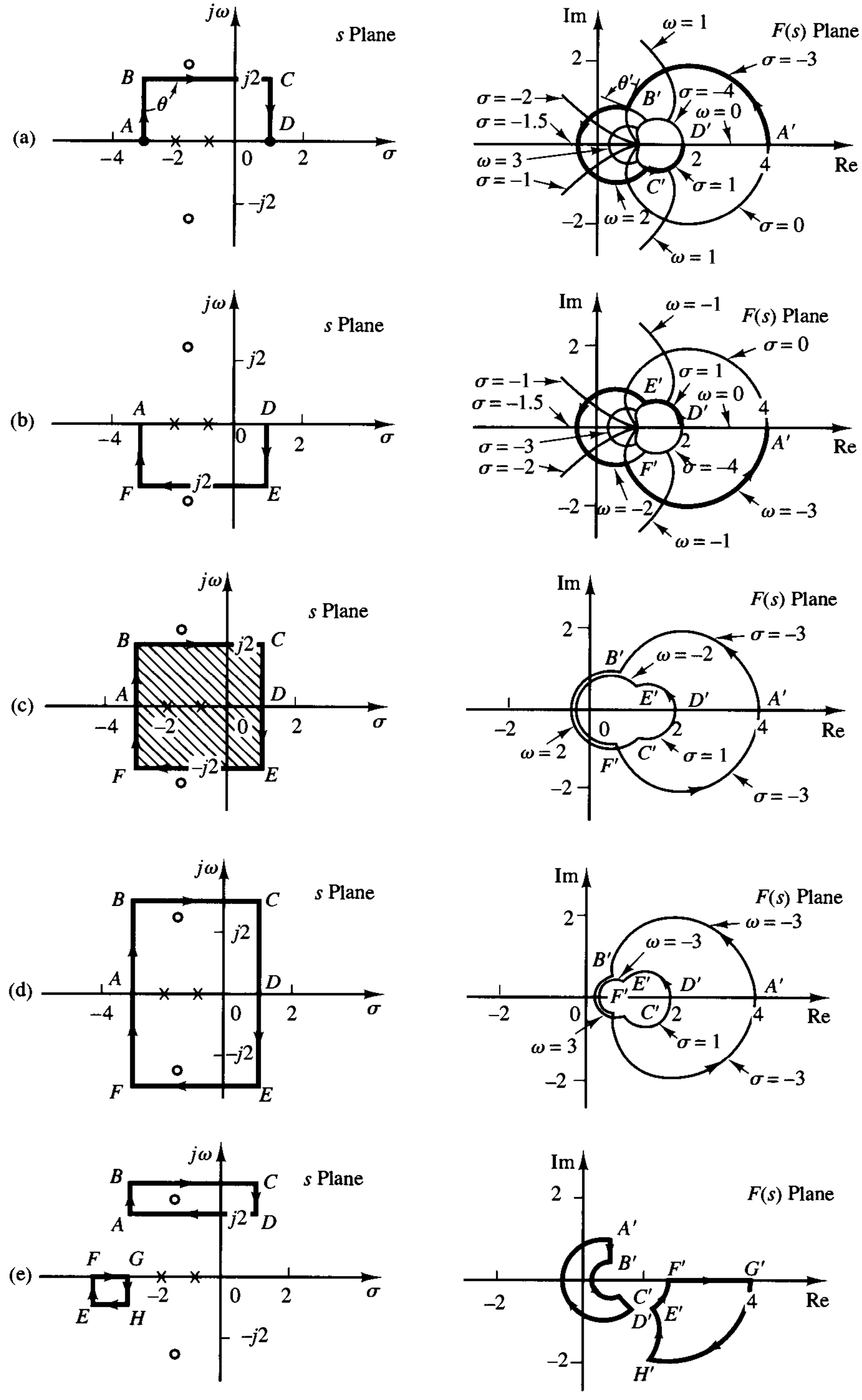


Figure 8–49
Conformal mapping of s plane grids into the $F(s)$ plane.

graphical explanation of the mapping theorem, which is the basis for the Nyquist stability criterion.

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 this 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.)

We shall not present a formal proof of this theorem here but leave the proof to Problem A-8-10. 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 8–50, 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

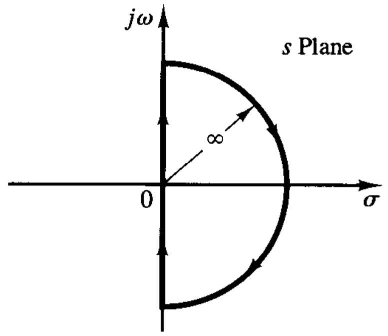


Figure 8–50
Closed contour in the s plane.

$1 + G(s)H(s)$ 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 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 8–51. 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, 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

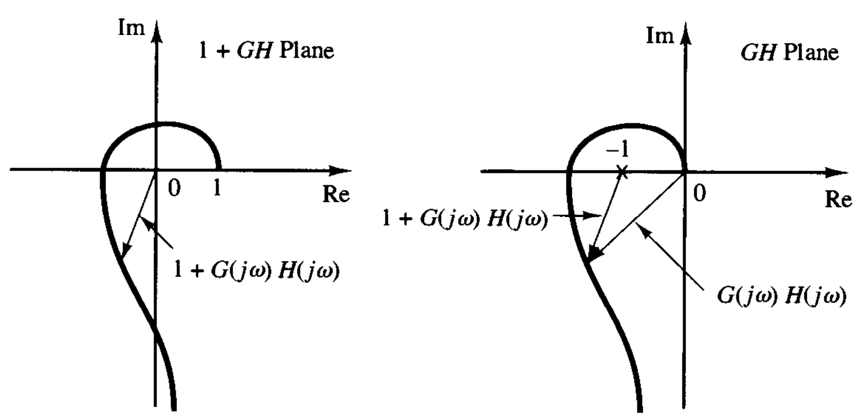


Figure 8–51
Plots of $1 + G(j\omega)H(j\omega)$ in the $1 + GH$ plane and GH plane.

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 8–48, if the open-loop transfer function $G(s)H(s)$ has k poles in the right-half s plane and $\lim_{s\to\infty} G(s)H(s) =$ 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.

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 8–52. 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 in-

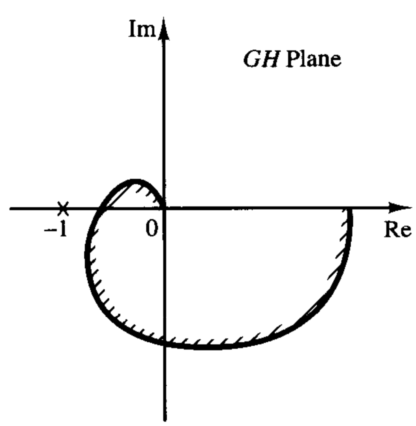


Figure 8–52
Region enclosed by a Nyquist plot.

spection 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. One form of a series expansion of e^{-Ts} was given in Chapter 5 and repeated here:

$$e^{-Ts} = \frac{1 - \dfrac{Ts}{2} + \dfrac{(Ts)^2}{8} - \dfrac{(Ts)^3}{48} + \cdots}{1 + \dfrac{Ts}{2} + \dfrac{(Ts)^2}{8} + \dfrac{(Ts)^3}{48} + \cdots}$$

As a first approximation, we may take only the first two terms in the numerator and denominator, respectively, or

$$e^{-Ts} \doteqdot \frac{1 - \dfrac{Ts}{2}}{1 + \dfrac{Ts}{2}} = \frac{2 - Ts}{2 + Ts}$$

This gives a good approximation to transport lag for the frequency range $0 \leq \omega \leq (0.5/T)$. [Note that the magnitude of $(2 - j\omega T)(2 + j\omega T)$ is always unity, and the phase lag of $(2 - j\omega T)/(2 + j\omega T)$ closely approximates that of transport lag within the stated frequency range.]

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 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 8–53. 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 $\epsilon \ll 1$) 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. 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 $\epsilon \ll 1$), the complex variable s can be written

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

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

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

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. 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 maps into the GH plane as a semicircle of infinite radius. Figure 8–54 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 8–54, 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 $\epsilon \ll 1$). For example, consider the following open-loop transfer function:

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

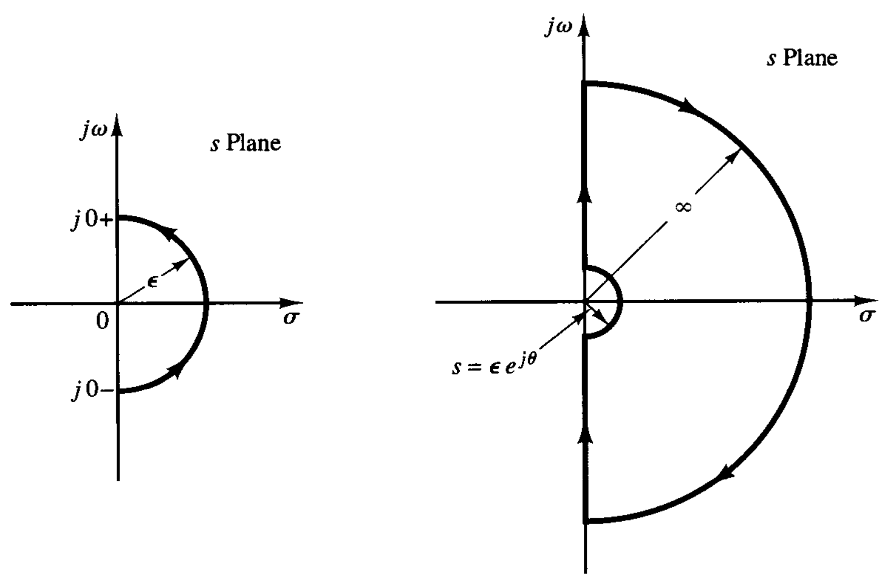


Figure 8–53
Closed contours in the s plane avoiding poles and zeros at the origin.

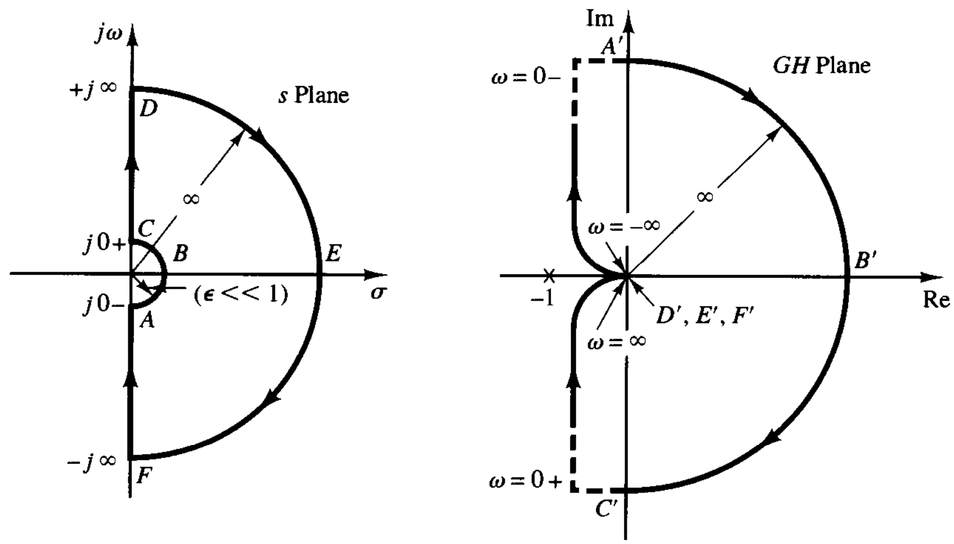


Figure 8–54
s-Plane contour and the $G(s)H(s)$ locus in the GH plane, where $G(s)H(s) = K/[s(Ts + 1)]$.

Then

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

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 8–55. 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.

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 8–48, 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

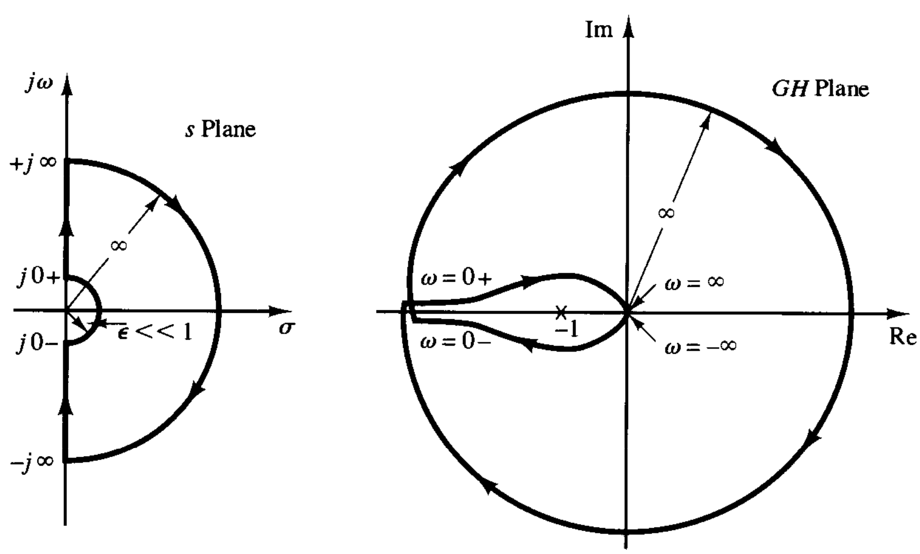


Figure 8–55
s-Plane contour and the $G(s)H(s)$ locus in the GH plane, where $G(s)H(s) = K/[s^2(Ts + 1)]$.

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 is a counterclockwise encirclement or 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 $G(s)H(s)$ in the right-half s plane; otherwise, the system is unstable.
3. There is a clockwise encirclement or 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.

EXAMPLE 8–13 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 8–56. 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.

EXAMPLE 8–14 Consider the system with the following open-loop transfer function:

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

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 8–57. 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.

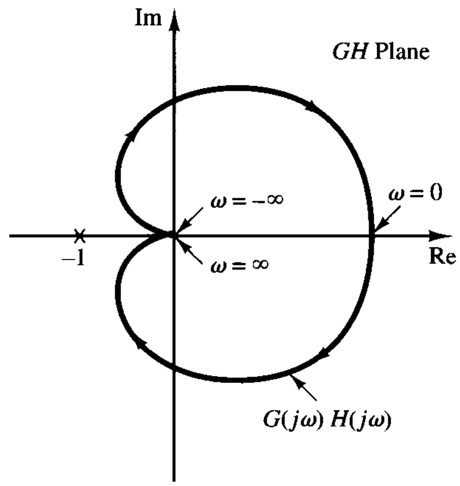


Figure 8–56
Polar plot of $G(j\omega)H(j\omega)$ considered in Example 8–13.

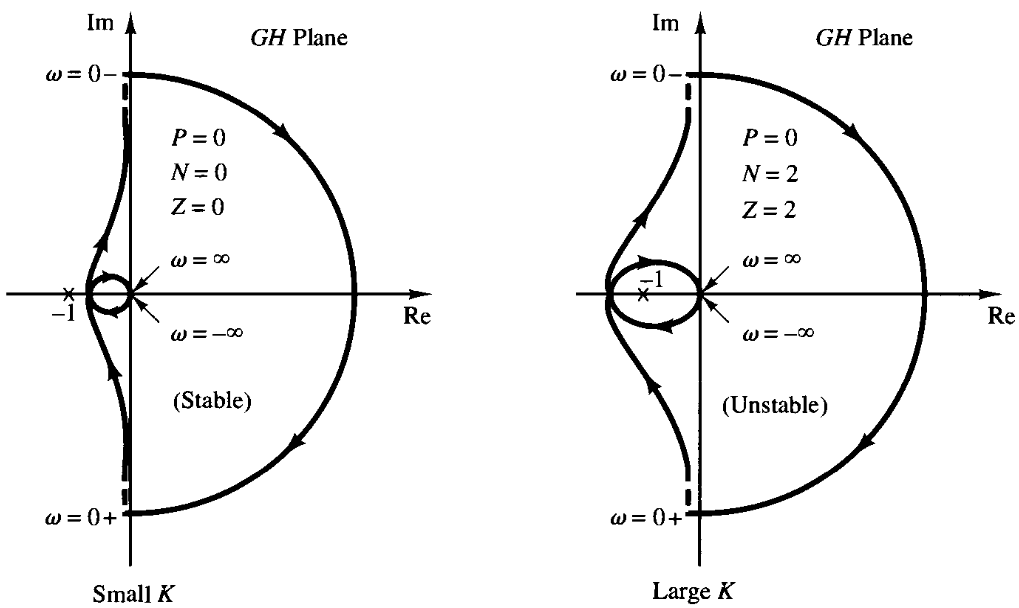


Figure 8–57
Polar plots of the system considered in Example 8–14.

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 8–15 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 8–58. 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

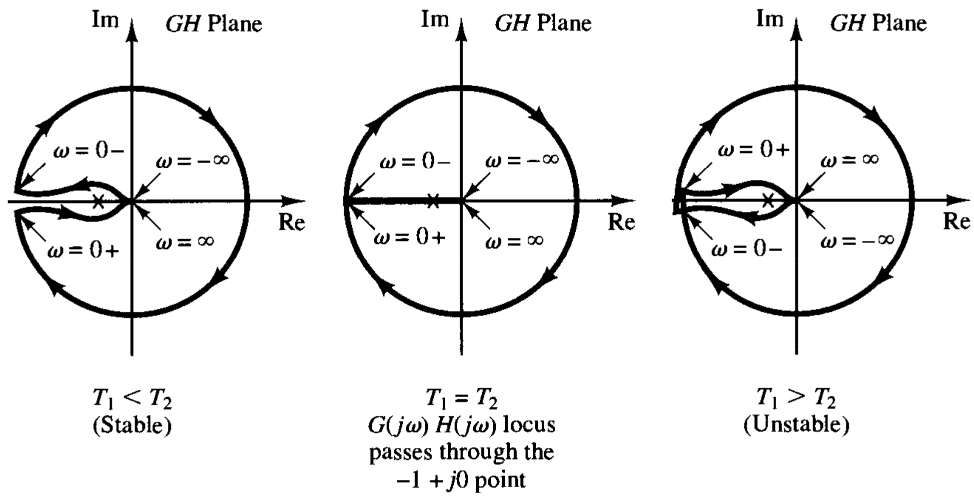


Figure 8–58
Polar plots of the system considered in Example 8–15.

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 8–16 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 8–59 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 8–17 Investigate the stability of a closed-loop system with the following open-loop transfer function:

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

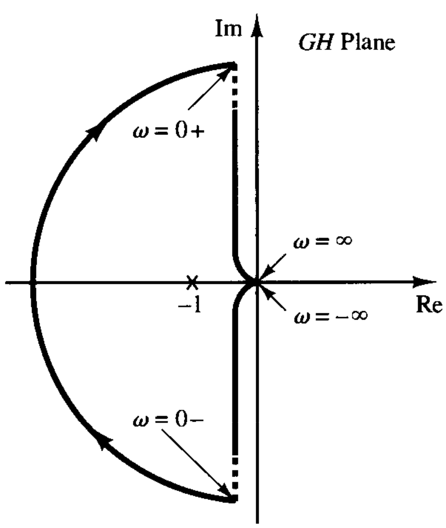


Figure 8–59
Polar plot of the system considered in Example 8–16.

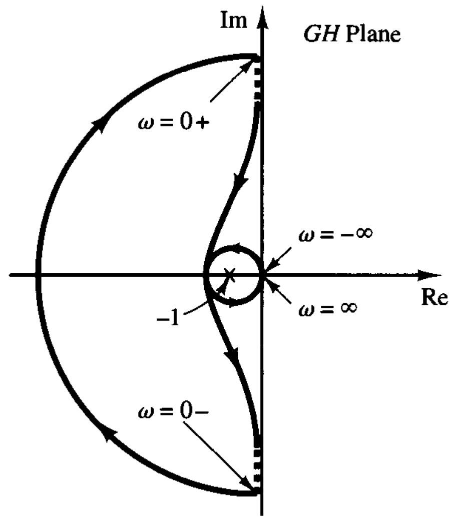


Figure 8–60
Polar plot of the system considered in Example 8–17.

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 8–60 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 8–61 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 8–61. 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.

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

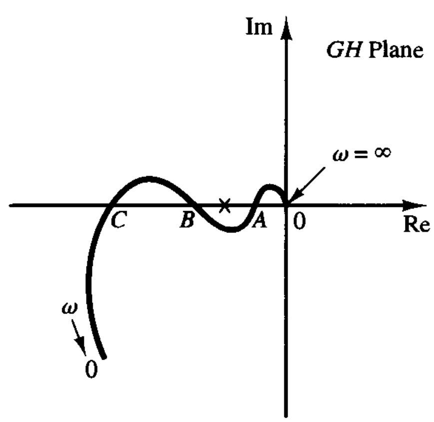


Figure 8–61
Polar plot of a conditionally stable system.

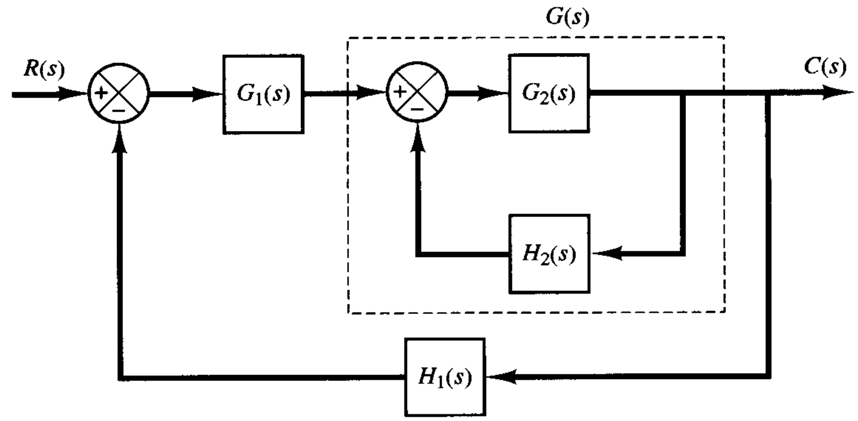


Figure 8–62
Multiple-loop system.

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 8–62. 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 this portion of the 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 $G_1(s)G(s)H_1(s)$.

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 8–18 Consider the control system shown in Figure 8–63. The system involves two loops. Determine the range of gain K for stability of the system by use of the Nyquist stability criterion. (The gain K is positive).

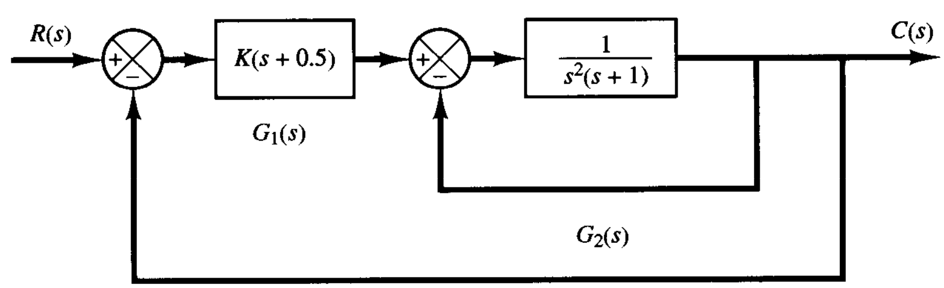


Figure 8–63
Control system.

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:

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

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

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

Our problem is to determine the range of 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 8–64 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 $P_1 = 2$. Noting that

$$Z_1 = N_1 + P_1$$

for stability, we require $Z_1 = 0$ or $N_1 = -2$. That is, the Nyquist locus of $G(j\omega)$ must encircle the $-1 + j0$ point twice counterclockwise. From Figure 8–64, 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 gain K for stability is

$$2 < K$$

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 use of the Routh stability criterion.] If the open-loop transfer function $G(s)H(s)$ has no zeros in the right-half s

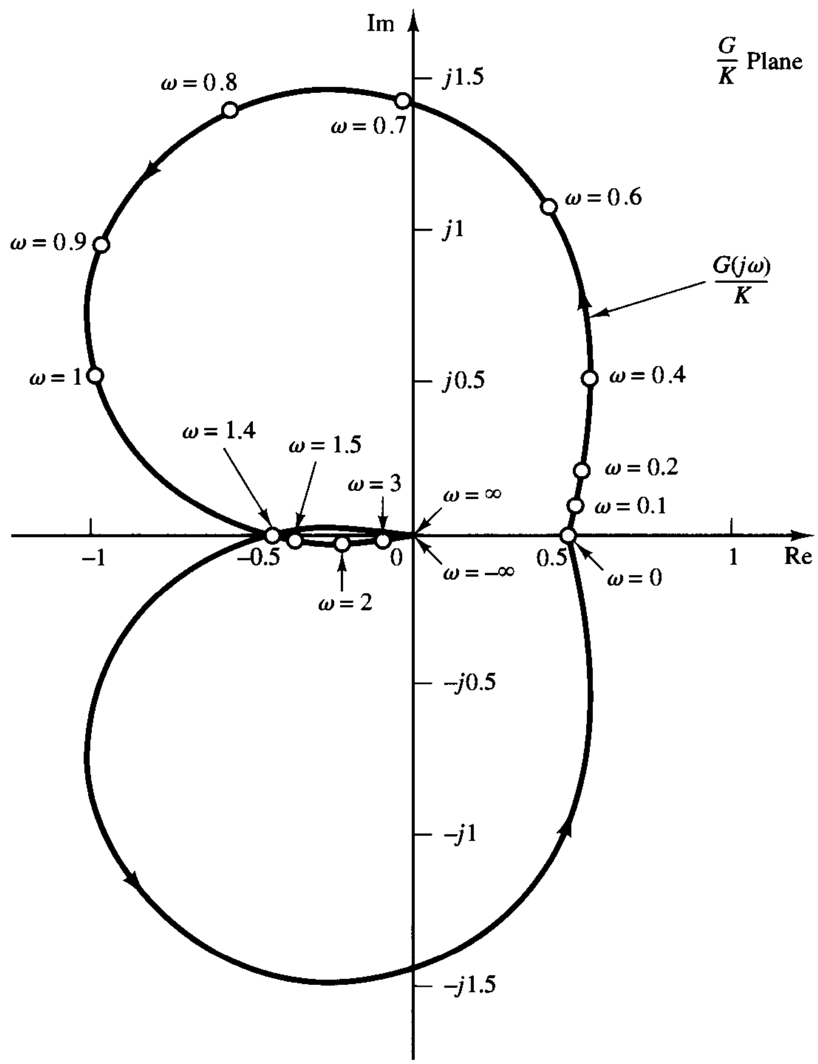


Figure 8–64
Polar plot of $G(j\omega)/K$.

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).

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 8–19 Consider the control system shown in Figure 8–63. (Refer to Example 8–18.) 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 8–65 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 8–65 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 8–18.

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. For the following second-order characteristic equation,

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0 \qquad (0 < \zeta < 1)$$

the roots are complex conjugates and are

$$s_1 = -\zeta\omega_n + j\omega_n\sqrt{1 - \zeta^2}, \qquad s_2 = -\zeta\omega_n - j\omega_n\sqrt{1 - \zeta^2}$$

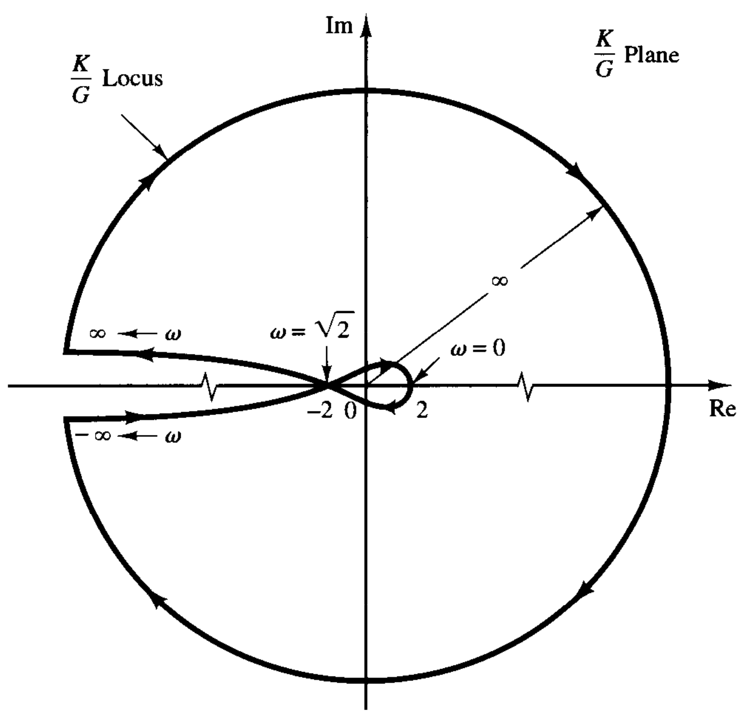


Figure 8–65 Polar plot of $K/G(j\omega)$.

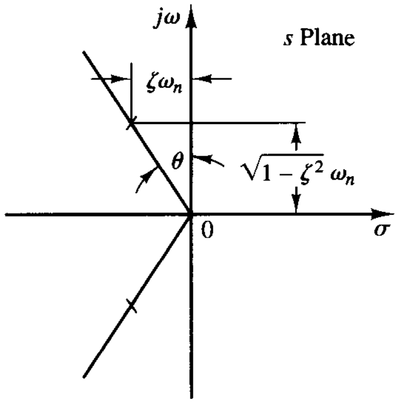


Figure 8–66
Plot of complex-conjugate roots in the s plane.

If these roots are plotted in the s plane, as shown in Figure 8–66, then we see that sin $\theta = \zeta$, or the angle θ is indicative of the damping ratio ζ. As θ becomes smaller, so does the value of ζ.

If we modify the Nyquist path and use radial lines with angle θ_x, instead of the $j\omega$ axis, as shown in Figure 8–67, then it can be said, following the same reasoning as in the case of the Nyquist stability criterion, that if the $G(s)H(s)$ locus corresponding to the modified s-plane contour does not encircle the $-1 + j0$ point and none of the poles of $G(s)H(s)$ lie within the closed s-plane contour then this contour does not enclose any zeros of $1 + G(s)H(s)$. The characteristic equation, $1 + G(s)H(s) = 0$, then does not have any roots within the modified s-plane contour. If no closed-loop poles of a higher-order system are enclosed by this contour, we can say that the damping ratio of each pair of complex-conjugate closed-loop poles of the system is greater than sin θ_x.

Suppose that the s-plane contour consists of a line to the left of and parallel to the $j\omega$ axis at a distance $-\sigma_o$ (or the line $s = -\sigma_o + j\omega$) 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 $s = -\sigma_o + j\omega$ and $s = j\omega$, as shown in Figure 8–68(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 $s = -\sigma_o + j\omega$. An example of a $G(-\sigma_o + j\omega)H(-\sigma_o + j\omega)$ locus, together with a $G(j\omega)H(j\omega)$ locus, is shown in Figure 8–68(b). The magnitude $1/\sigma_o$ is indicative of the time constant of the dominant closed-loop poles. If all

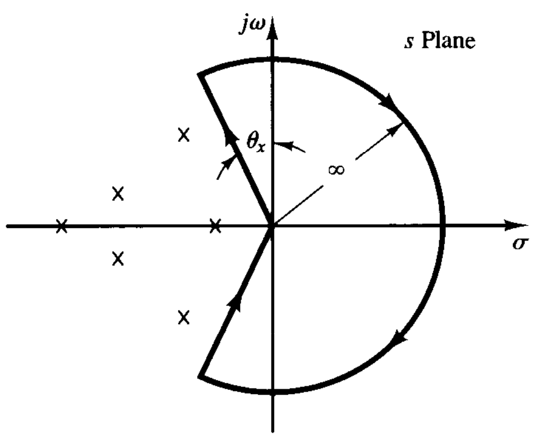


Figure 8–67
Modified Nyquist path.

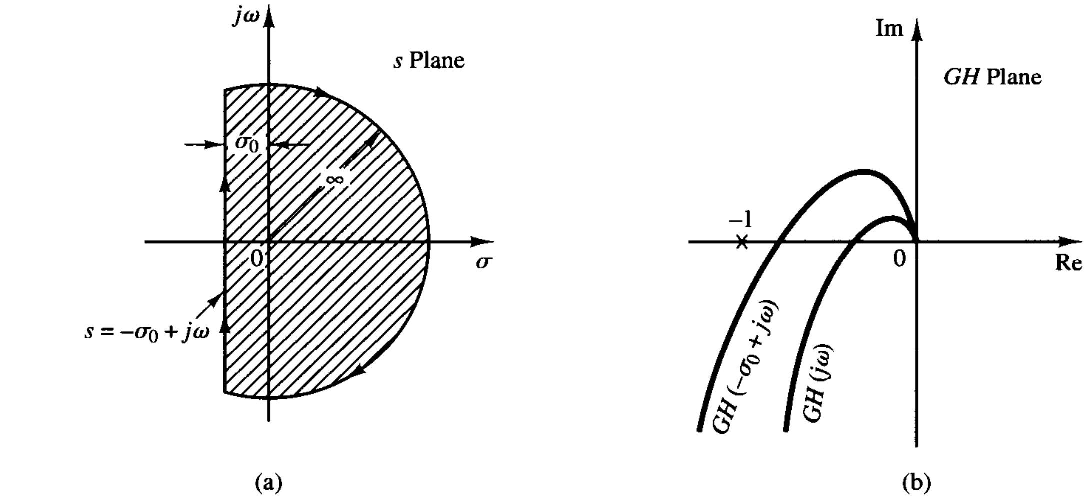


Figure 8–68
(a) Modified Nyquist path; (b) polar plots of $G(-\sigma_o + j\omega)$ $H(-\sigma_o + j\omega)$ locus and $G(j\omega)H(j\omega)$ locus in the GH plane.

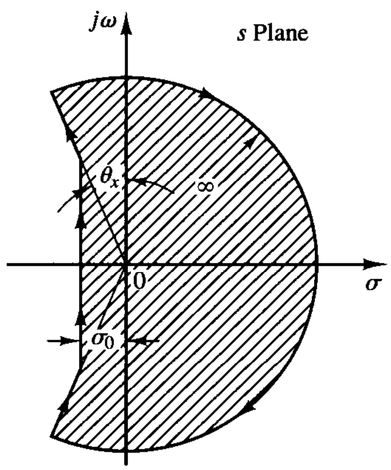


Figure 8–69
Modified Nyquist path.

roots lie outside the s-plane contour, all time constants of the closed-loop transfer function are less than $1/\sigma_o$. If the s-plane contour is chosen as shown in Figure 8–69, then the test of encirclements of the $-1 + j0$ point reveals the existence or nonexistence of the roots of the characteristic equation of the closed-loop system within this s-plane contour. If the test reveals that no roots lie in the s-plane contour, then it is clear that all the closed-loop poles have damping ratios greater than ζ_x and time constants less than $1/\sigma_o$. Thus, by taking an appropriate s-plane contour, we can investigate time constants and damping ratios of closed-loop poles from the open-loop transfer function.

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

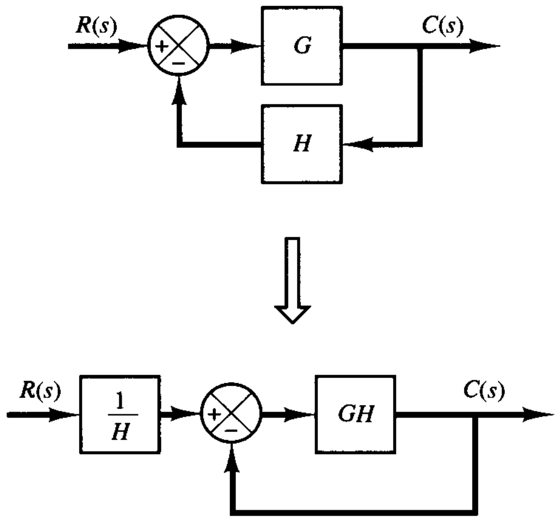


Figure 8–70
Modification of a system with feedback elements to a unity-feedback system.

a unity-feedback system, as shown in Figure 8–70. 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 $G(s)$ 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.

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 8–71. The constant-ω lines in the s plane map into curves, also shown in Figure 8–71.

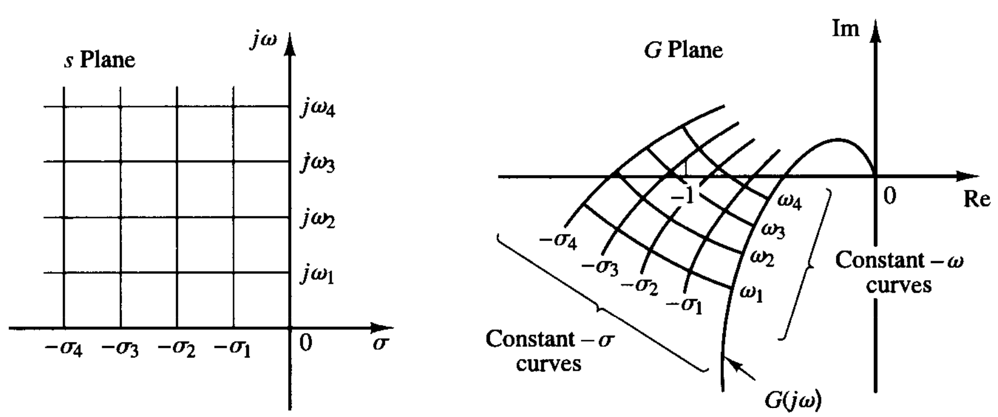


Figure 8–71
Conformal mapping of s-plane grids into the $G(s)$ plane.

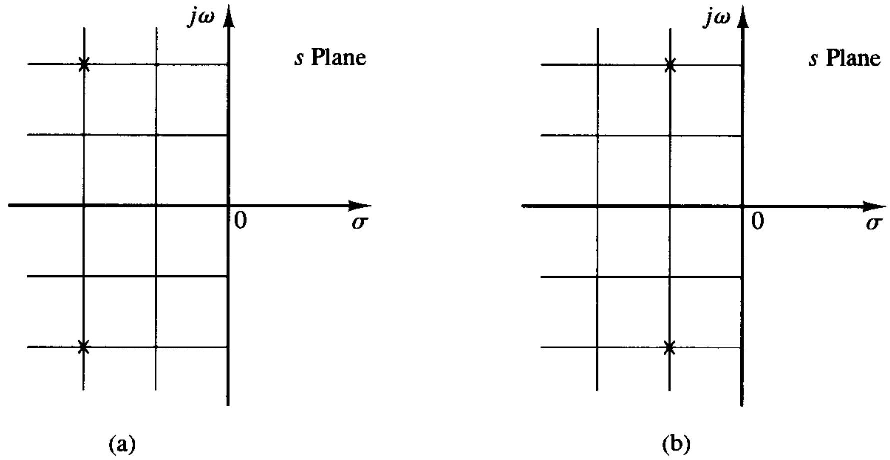


Figure 8–72 Two systems with two closed-loop poles.

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 Figure 8–72(a) and (b). (In Figure 8–72, the $\times$'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 8–73(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.

Phase and gain margins. Figure 8–74 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,

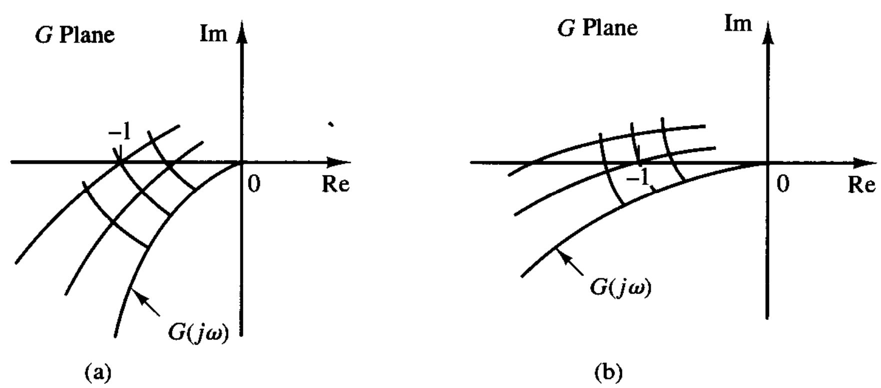


Figure 8–73 Conformal mappings of s-plane grids for the systems shown in Figure 8–72 into the $G(s)$ plane.

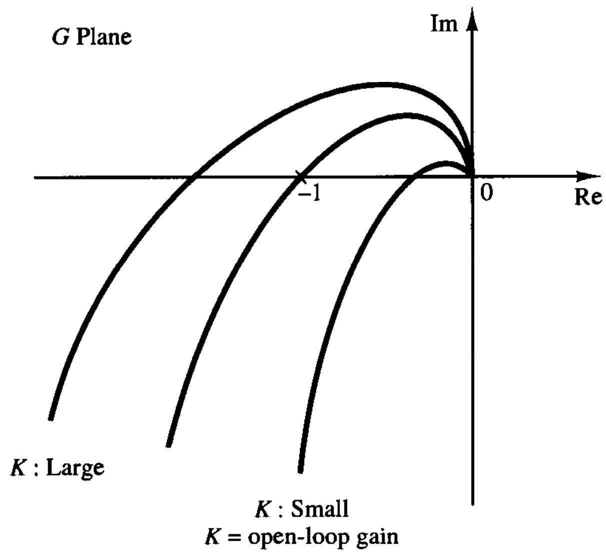


Figure 8–74
Polar plots of
$$\frac{K(1 + j\omega T_a)(1 + j\omega T_b)\cdots}{(j\omega)(1 + j\omega T_1)(1 + j\omega T_2)\cdots}.$$

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

$$\gamma = 180^\circ + \phi$$

Figures 8–75(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 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

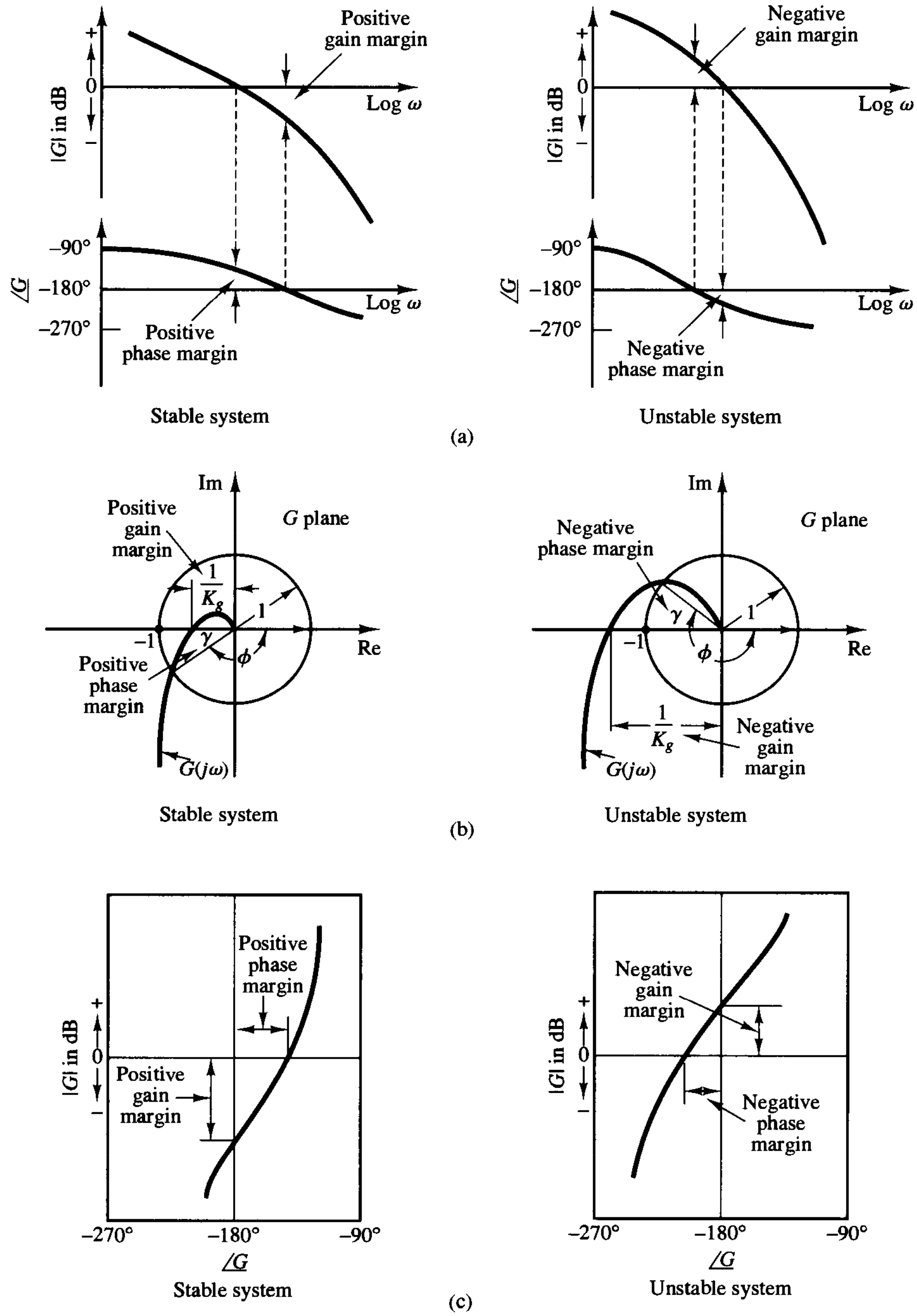


Figure 8–75
Phase and gain margins of stable and unstable systems.
(a) Bode diagrams;
(b) polar plots;
(c) log-magnitude versus phase plots.

system is stable, and a negative gain margin (in decibels) means that the system is unstable. The gain margin is shown in Figures 8–75(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 8–76. 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.

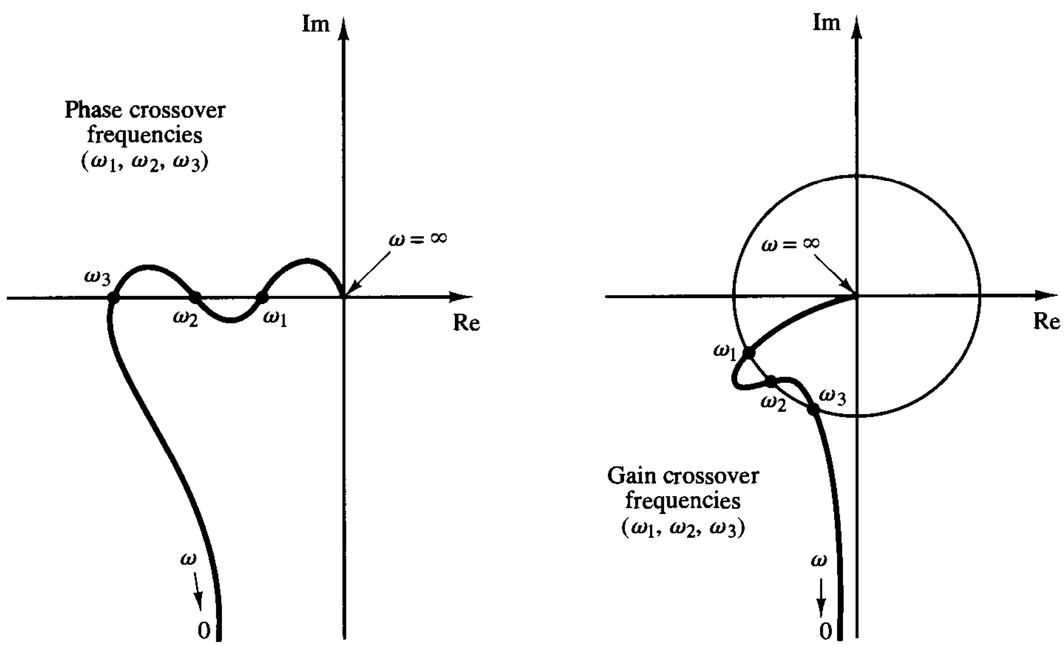


Figure 8–76
Polar plots showing more than two phase or gain crossover frequencies.

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 values of frequency. The two values bound the behavior of the 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 system 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.

EXAMPLE 8–20 Obtain the phase and gain margins of the system shown in Figure 8–77 for the two cases where $K = 10$ and $K = 100$.

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 8–78(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 8–78(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° ~ 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.

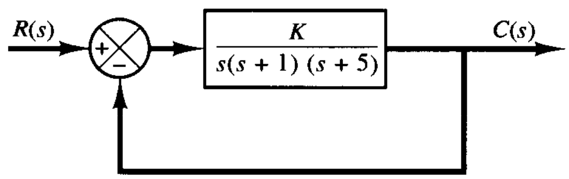


Figure 8–77
Control system.

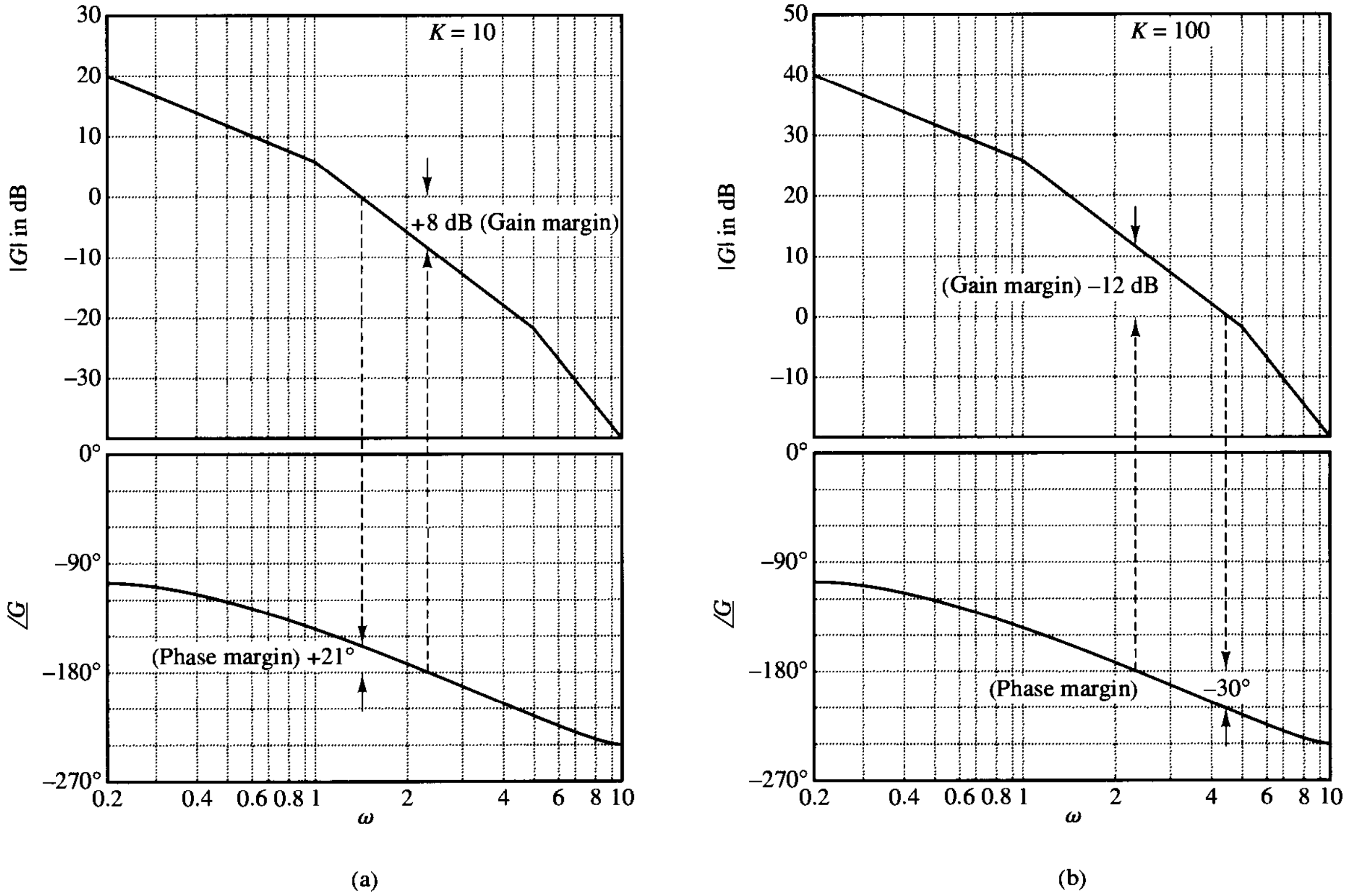


Figure 8–78
Bode diagrams of the system shown in Figure 8–77 (a) with $K = 10$ and (b) with $K = 100$.

Resonant peak magnitude M_r and resonant peak frequency ω_r. Consider the system shown in Figure 8–79. The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{8–9}$$

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

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

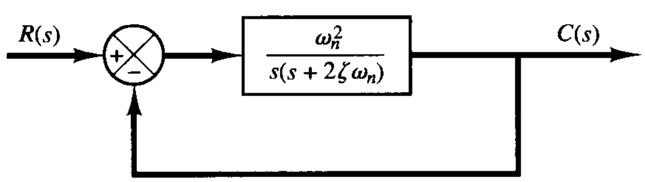


Figure 8–79
Control system.

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 (8–6), for $0 \leq \zeta \leq 0.707$ the maximum value of M occurs at the frequency ω_r, where

$$\omega_r = \omega_n \sqrt{1 - 2\zeta^2} = \omega_n \sqrt{\cos 2\theta} \tag{8–10}$$

The angle θ is defined in Figure 8–80. The frequency ω_r is the resonant frequency. At the resonant frequency, the value of M is maximum and is given by Equation (8–7), rewritten

$$M_r = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} = \frac{1}{\sin 2\theta} \tag{8–11}$$

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 (8–8).] 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.

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 8–79, can be exactly correlated

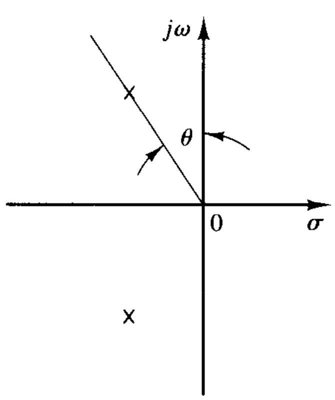


Figure 8–80
Definition of the angle θ.

to the resonant peak magnitude in the frequency response. Hence, essentially the same information of 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 8–79 is given by Equation (4–21), 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

$$\omega_d = \omega_n\sqrt{1-\zeta^2} = \omega_n\cos\theta \tag{8–12}$$

On the other hand, the maximum overshoot M_p for the unit-step response is given by Equation (4–30), or

$$M_p = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \tag{8–13}$$

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 8–79 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

$$\begin{aligned}\gamma &= 180^\circ + \underline{/G(j\omega)} \\ &= 90^\circ - \tan^{-1}\frac{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}{2\zeta} \\ &= \tan^{-1}\frac{2\zeta}{\sqrt{\sqrt{1+4\zeta^4} - 2\zeta^2}}\end{aligned} \tag{8–14}$$

Equation (8–14) 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.)

In the following, we shall summarize the correlation between the step transient response and frequency response of the second-order system given by Equation (8–9):

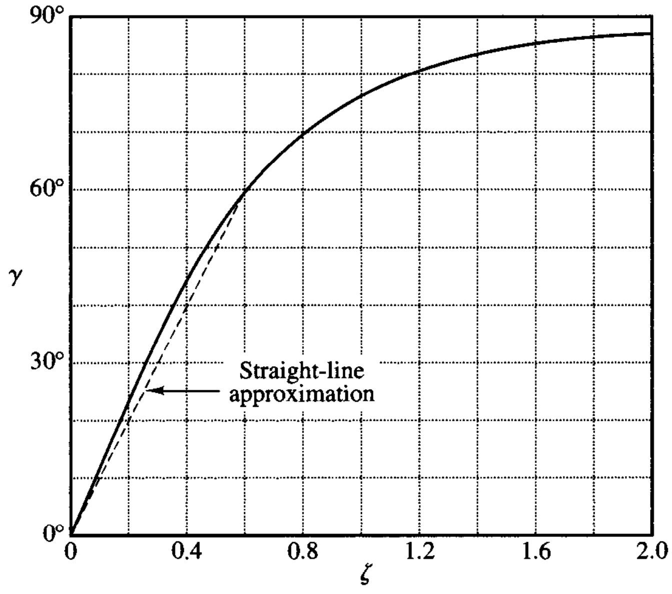


Figure 8–81
Curve γ (phase margin) versus ζ for the system shown in Figure 8–79.

1. The phase margin and the damping ratio are directly related. Figure 8–81 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 8–79 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 transient response (that is, the damping ratio) from the frequency response.

2. Referring to Equations (8–10) and (8–12), 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 (8–11) and (8–13), 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 8–82. 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 frequency response. The main reason for this is the relative simplicity of this approach as compared to others. Since in many applications it is the transient response of the

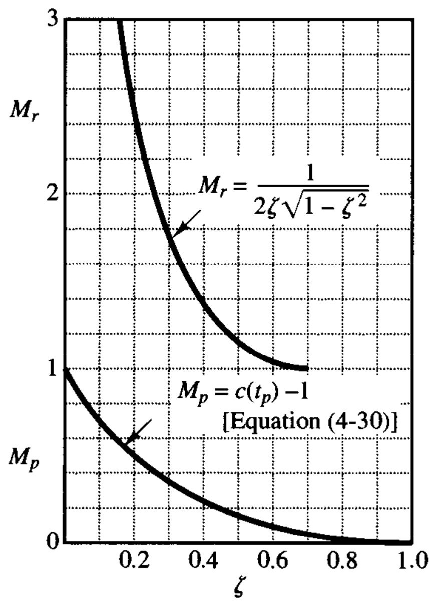


Figure 8–82
Curves M_r versus ζ and M_p versus ζ for the system shown in Figure 8–79.

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 second-order system shown in Figure 8–79, mathematical relationships correlating the step transient response and frequency response can be obtained easily. The time response of a second-order system can be exactly predicted from a knowledge of the M_r and ω_r of its closed-loop frequency response.

For higher-order systems, the correlation is more complex, and transient response may not be predicted easily from frequency response because additional poles may change the correlation between the step transient response and frequency response existing for a 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 second-order system shown in Figure 8–79 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 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 dB <

$M_r < 3$ 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.

3. The resonant peak frequency ω_r and the damped natural frequency ω_d for 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 they can be approximated by a 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 peak 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 8–83, 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

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

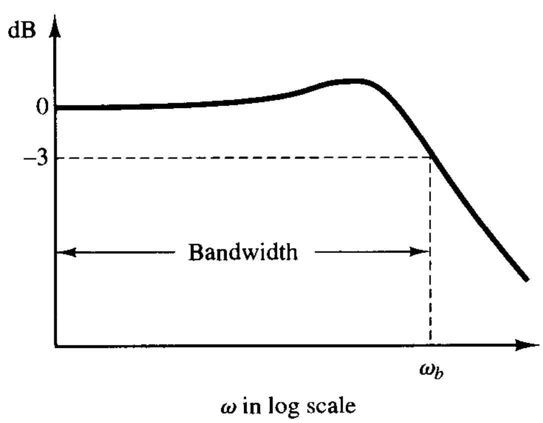


Figure 8–83
Plot of a closed-loop frequency response curve showing cutoff frequency ω_b and bandwidth.

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 $0 \leq \omega \leq \omega_b$ 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 of 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:

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.
2. The necessary filtering characteristics for high-frequency noise.

For the system to follow arbitrary inputs accurately, it is necessary that the system 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 relatively small stability margin.

EXAMPLE 8–21 Consider the following two systems:

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

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 a smaller bandwidth.

Figure 8–84(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 $0 \leq \omega \leq 1$ rad/sec and that of system II is $0 \leq \omega \leq 0.33$ rad/sec. Figures 8–84(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.

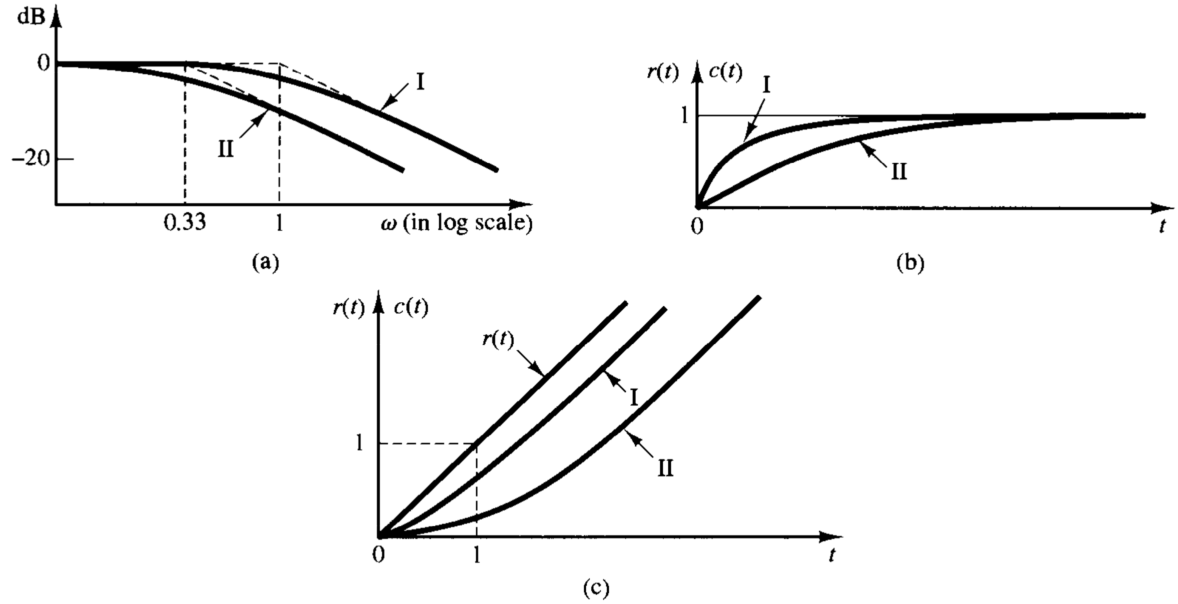


Figure 8–84 Comparison of dynamic characteristics of the two systems considered in Example 8–21. (a) Closed-loop frequency-response curves; (b) unit-step response curves; (c) unit-ramp response curves.

8–10 CLOSED-LOOP FREQUENCY RESPONSE

Closed-loop frequency response of unity-feedback systems. For a stable closed-loop system, the frequency response can be obtained easily from that of the open loop. Consider the unity-feedback system shown in Figure 8–85(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 8–85(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

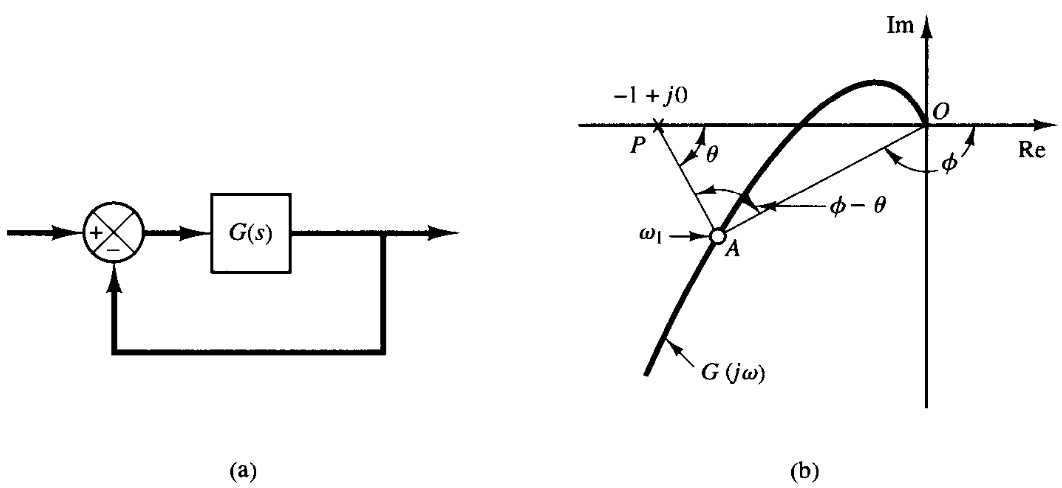


Figure 8–85 (a) Unity-feedback system; (b) determination of closed-loop frequency response from open-loop frequency response.

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

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 8–85(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

$$X^2(1 - M^2) - 2M^2X - M^2 + (1 - M^2)Y^2 = 0 \tag{8–15}$$

If $M = 1$, then from Equation (8–15) we obtain $X = -\frac{1}{2}$. This is the equation of a straight line parallel to the Y axis and passing through the point $(-\frac{1}{2}, 0)$.

If $M \neq 1$, Equation (8–15) 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

$$\left(X + \frac{M^2}{M^2 - 1}\right)^2 + Y^2 = \frac{M^2}{(M^2 - 1)^2} \tag{8–16}$$

Equation (8–16) 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 8–86. 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.

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

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

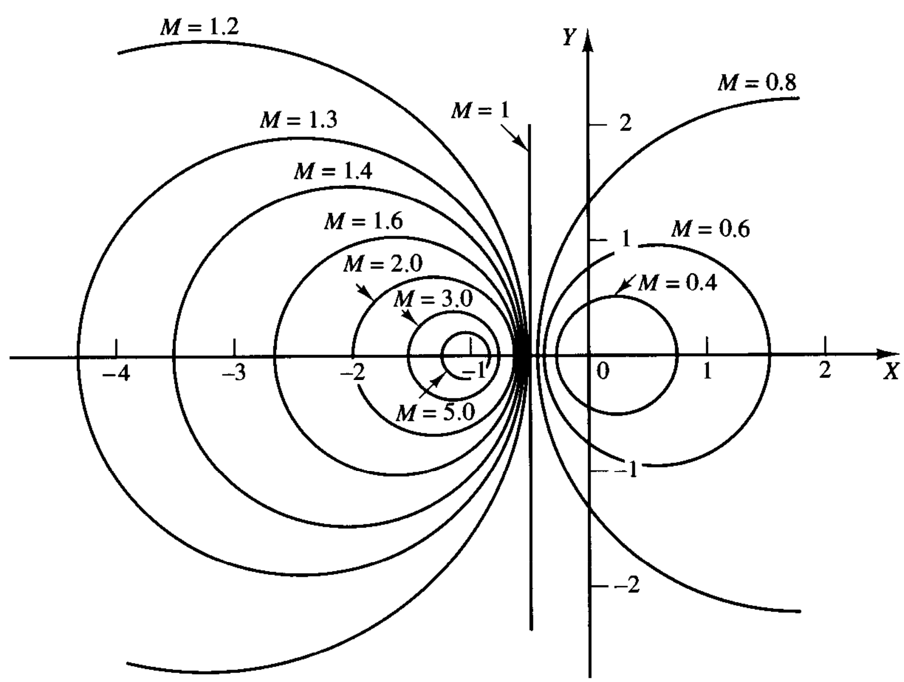


Figure 8–86
A family of constant M circles.

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 (¼) + $1/(2N)^2$ to both sides of this last equation yields

$$\left(X + \frac{1}{2}\right)^2 + \left(Y - \frac{1}{2N}\right)^2 = \frac{1}{4} + \left(\frac{1}{2N}\right)^2 \tag{8–17}$$

This is an equation of a circle with center at $X = -\frac{1}{2}$, $Y = 1/(2N)$ and with radius $\sqrt{(1/4) + 1/(2N)^2}$. 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 (8–17) 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 8–87 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 multiplies thereof) is added to the angle.

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 gives 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.

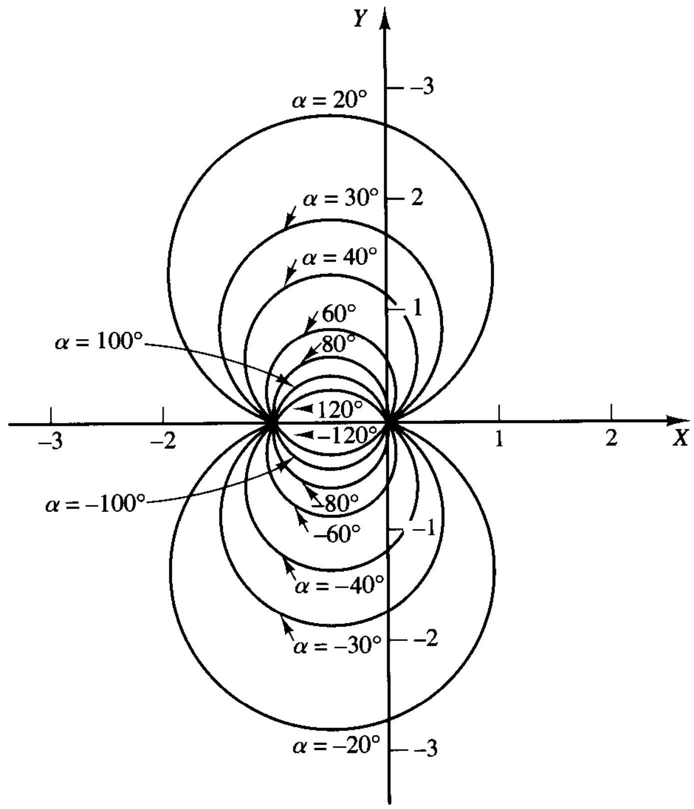


Figure 8–87 A family of constant N circles.

Figure 8–88(a) shows the $G(j\omega)$ locus superimposed on a family of M circles. Figure 8–88(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 8–88(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 8–88(c) shows the 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 $\omega_r = \omega_4$.)

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. This chart is shown in Figure 8–89, 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 peak frequency, and bandwidth of the closed-loop system from the plot of the open-loop locus, $G(j\omega)$.

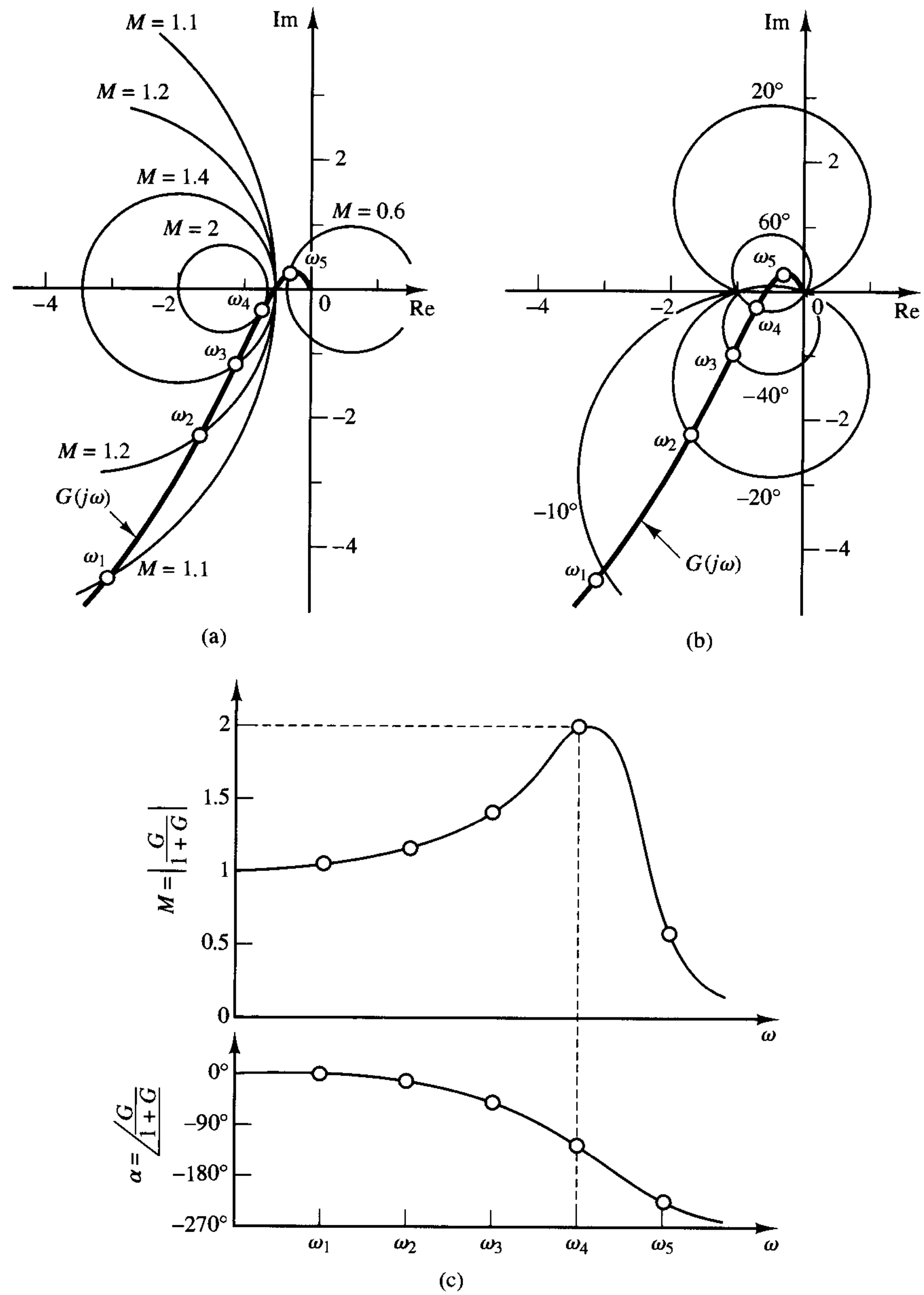


Figure 8–88
(a) $G(j\omega)$ locus superimposed on a family of M circles; (b) $G(j\omega)$ locus superimposed on a family of N circles; (c) closed-loop frequency-response curves.

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

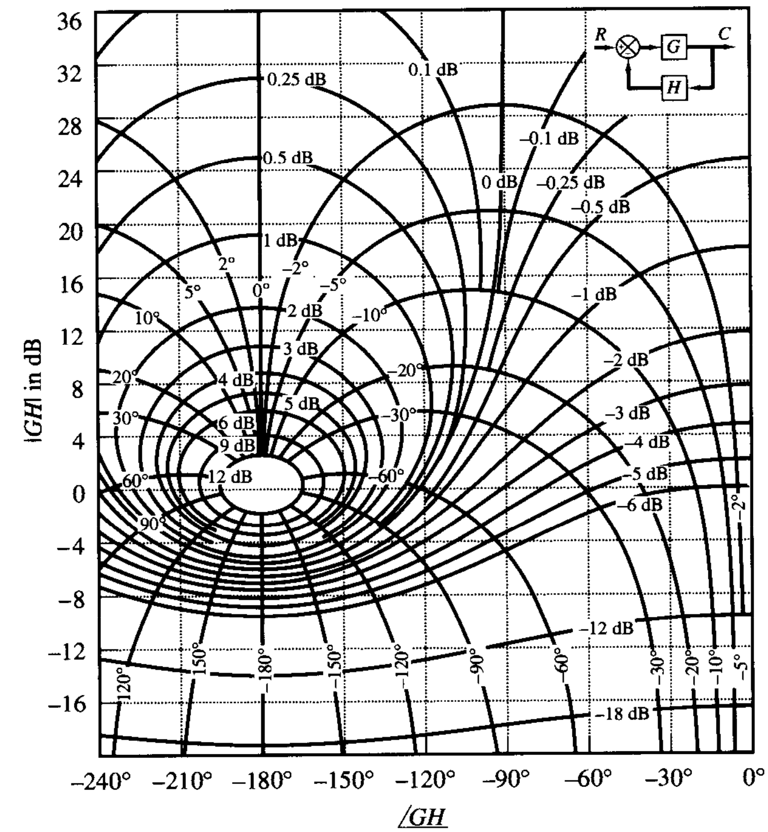


Figure 8–89
Nichols chart.

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 peak frequency is given by the frequency at the point of tangency.

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 from the Bode diagram. The use of the Bode diagram eliminates the lengthy numerical calculation of $G(j\omega)$. Figure 8–90(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 8–90(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, $\omega = 0.76$

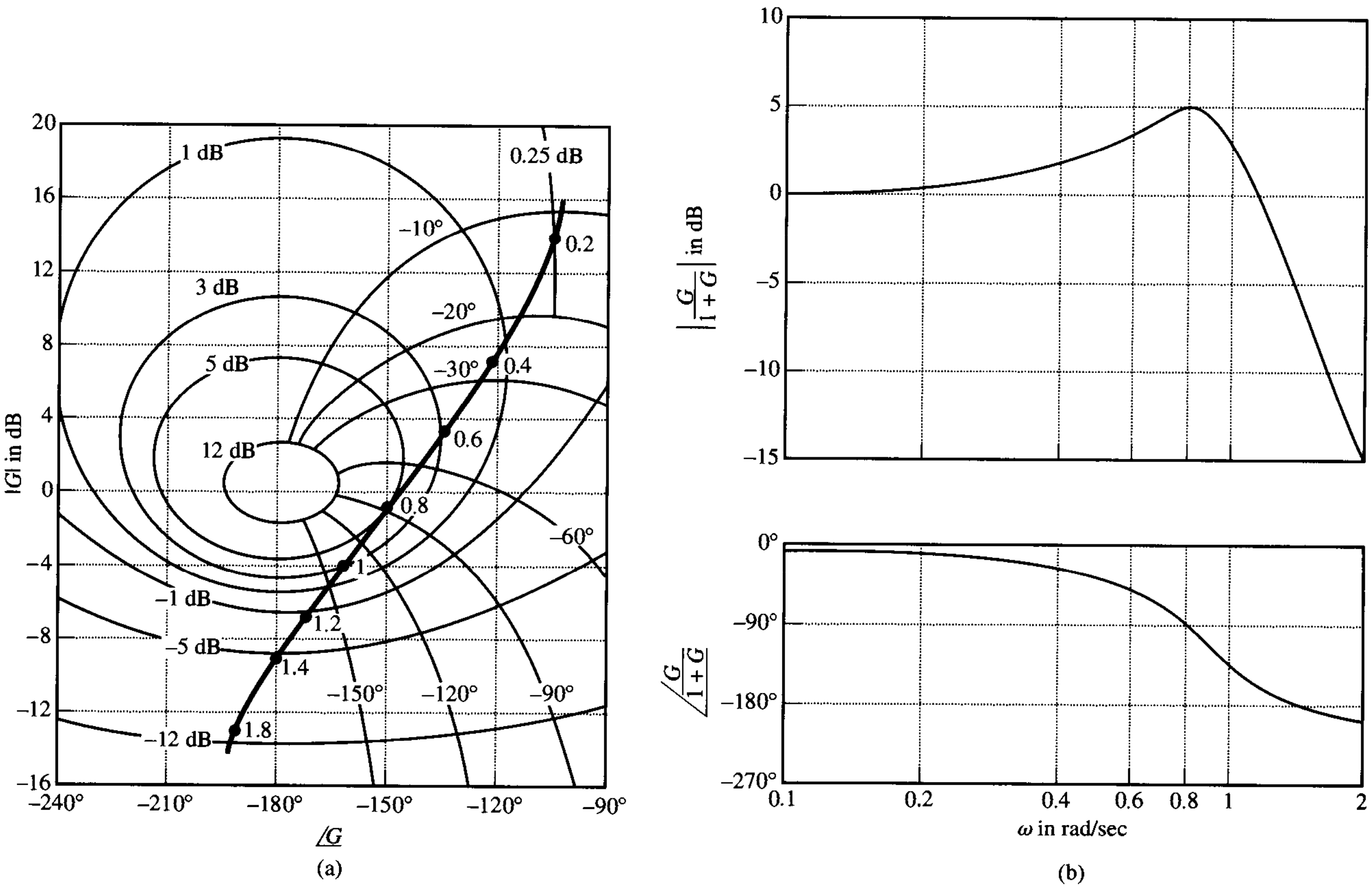


Figure 8–90
(a) Plot of $G(j\omega)$ superimposed on Nichols chart; (b) closed-loop frequency-response curves.

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.

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.

Closed-loop frequency response for nonunity-feedback systems. In the preceding sections, our discussions were limited to closed-loop systems with unity feedback. The constant M and N loci and the Nichols chart cannot be directly applied to control systems with nonunity feedback, but rather require a slight modification.

If the closed-loop system involves a nonunity-feedback transfer function, then the closed-loop transfer function may be written

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

where $G(s)$ is the feedforward transfer function and $H(s)$ is the feedback transfer function. Then $C(j\omega)/R(j\omega)$ can be written

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

The magnitude and phase angle of

$$\frac{G_1(j\omega)}{1 + G_1(j\omega)}$$

where $G_1(j\omega) = G(j\omega)H(j\omega)$, may be obtained easily by plotting the $G_1(j\omega)$ locus on the Nichols chart and reading the values of M and N at various frequency points. The closed-loop frequency response $C(j\omega)/R(j\omega)$ may then be obtained by multiplying $G_1(j\omega)/[1 + G_1(j\omega)]$ by $1/H(j\omega)$. This multiplication can be made without difficulty if we draw Bode diagrams for $G_1(j\omega)/[1 + G_1(j\omega)]$ and $H(j\omega)$ and then graphically subtract the magnitude of $H(j\omega)$ from that of $G_1(j\omega)/[1 + G_1(j\omega)]$ and also graphically subtract the phase angle of $H(j\omega)$ from that of $G_1(j\omega)/[1 + G_1(j\omega)]$. Then the resulting log-magnitude curve and phase-angle curve give the closed-loop frequency response $C(j\omega)/R(j\omega)$.

To obtain acceptable values of M_r, ω_r, and ω_b for $|C(j\omega)/R(j\omega)|$, a trial-and-error process may be necessary. In each trial, the $G_1(j\omega)$ locus is varied in shape. Then Bode diagrams for $G_1(j\omega)/[1 + G_1(j\omega)]$ and $H(j\omega)$ are drawn, and the closed-loop frequency response $C(j\omega)R(j\omega)$ is obtained. The values of M_r, ω_r, and ω_b are checked until they are acceptable.

Gain adjustments. The concept of M circles will now be applied to the design of control systems. In obtaining suitable performance, the adjustment of gain is usually the first consideration. The adjustment may be based on a desirable value for the resonant peak.

In the following, we shall demonstrate a method for determining the gain K so that the system will have some maximum value M_r, not exceeded over the entire frequency range.

Referring to Figure 8–91, we see that the tangent line drawn from the origin to the desired M_r circle has an angle of ψ, as shown, if M_r is greater than unity. The value of $\sin\psi$ is

$$\sin\psi = \left| \frac{\dfrac{M_r}{M_r^2 - 1}}{\dfrac{M_r^2}{M_r^2 - 1}} \right| = \frac{1}{M_r} \tag{8–18}$$

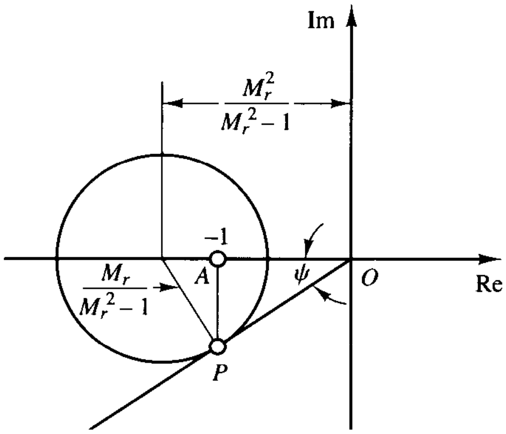


Figure 8–91
M circle.

Let us call the tangency point of the tangent line and the M_r circle as point P. It can easily be proved that the line drawn from point P, perpendicular to the negative real axis, intersects this axis at the $-1 + j0$ point.

Consider the system shown in Figure 8–92. The procedure for determining the gain K so that $G(j\omega) = KG_1(j\omega)$ will have a desired value of M_r (where $M_r > 1$) can be summarized as follows:

1. Draw the polar plot of the normalized open-loop transfer function $G_1(j\omega) = G(j\omega)/K$.
2. Draw from the origin the line that makes an angle of $\psi = \sin^{-1}(1/M_r)$ with the negative real axis.
3. Fit a circle with center on the negative real axis tangent to both the $G_1(j\omega)$ locus and the line PO.
4. Draw a perpendicular line to the negative real axis from point P, the point of tangency of this circle with the line PO. The perpendicular line PA intersects the negative real axis at point A.
5. For the circle just drawn to correspond to the desired M_r circle, point A should be the $-1 + j0$ point.
6. The desired value of the gain K is that value which changes the scale so that point A becomes the $-1 + j0$ point. Thus, $K = 1/\overline{OA}$.

Note that the resonant frequency ω_r is the frequency of the point at which the circle is tangent to the $G_1(j\omega)$ locus. The present procedure may not yield a satisfactory value for ω_r. If this is the case, the system must be compensated in order to increase the value

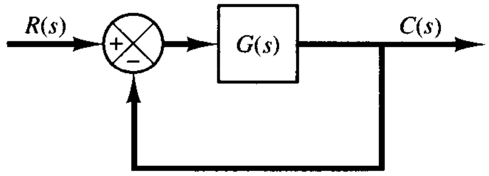


Figure 8–92
Control system.

of ω_r without changing the value of M_r. (For the compensation of control systems by frequency-response methods, see Chapter 9.)

Note also that if the system has nonunity feedback then the method requires some cut-and-try steps.

EXAMPLE 8–22 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

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

as shown in Figure 8–93. The value of ψ corresponding to $M_r = 1.4$ is obtained from

$$\psi = \sin^{-1}\frac{1}{M_r} = \sin^{-1}\frac{1}{1.4} = 45.6°$$

The next step is to draw the line OP that makes an angle $\psi = 45.6°$ with the negative real axis. Then draw the circle that is tangent to both the $G(j\omega)/K$ locus and the line OP. Define the point where the circle is tangent to the 45.6° line as point P. The perpendicular line drawn from point P intersects the negative real axis at $(-0.63, 0)$. Then the gain K of the system is determined as follows:

$$K = \frac{1}{0.63} = 1.59$$

It should be noted that such determination of the gain can also be done easily on the log-magnitude versus phase plot. In what follows, we shall demonstrate how the log-magnitude versus phase diagram can be used to determine the gain K so that the system will have a desired value of M_r.

Figure 8–94 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 $K < 1$.

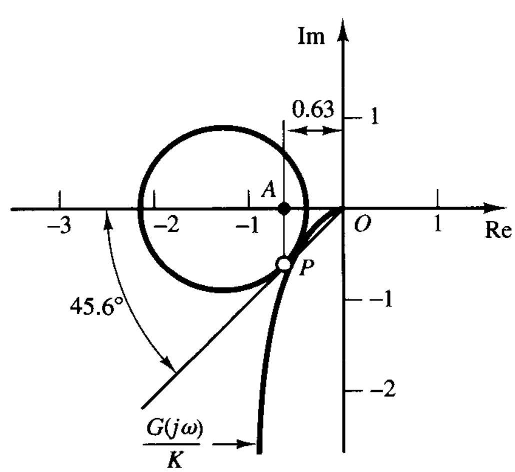


Figure 8–93
Determination of the gain K using an M circle.

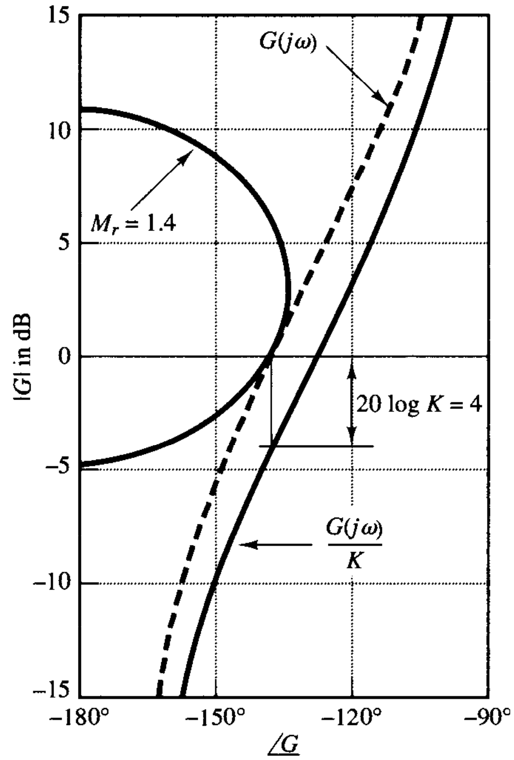


Figure 8–94
Determination of the gain K using the Nichols chart.

In Figure 8–94, 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 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

$$K = 1.59$$

Thus, we have the same result as obtained earlier.

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

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

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 8–8.

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 8–95, 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 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

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

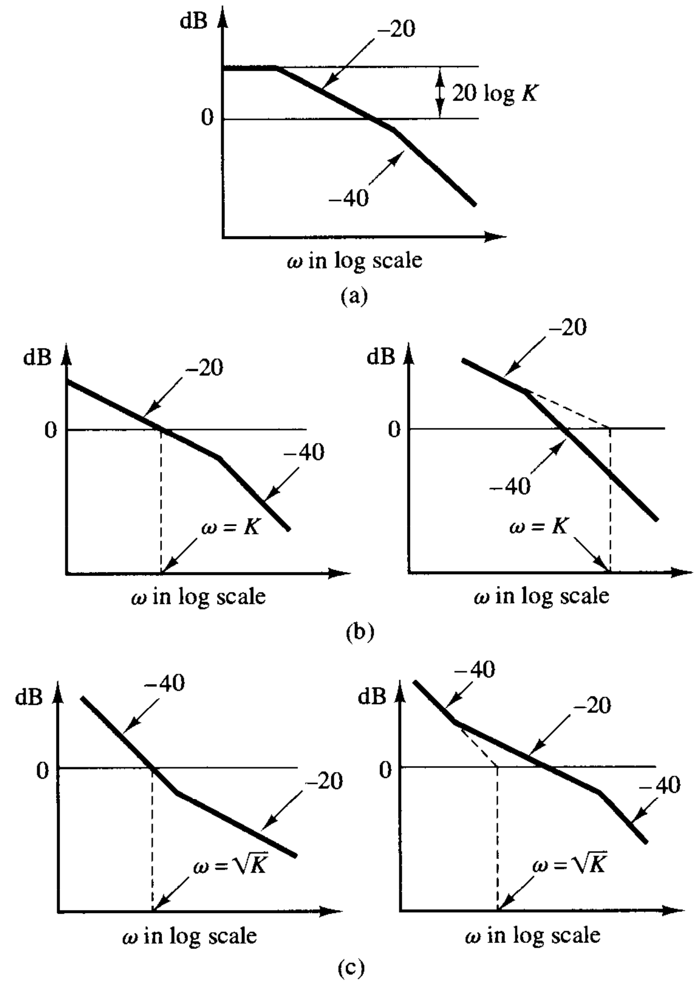


Figure 8–95
(a) Log-magnitude curve of a type 0 system; (b) log-magnitude curves of type 1 systems; (c) log-magnitude curves of type 2 systems. (The slopes shown are in dB/decade.)

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

$$\begin{aligned}\lim_{\omega\to\infty} \frac{d}{d\omega} \underline{/G(j\omega)e^{-j\omega T}} &= \lim_{\omega\to\infty} \frac{d}{d\omega}\left[\underline{/G(j\omega)} + \underline{/e^{-j\omega T}}\right] \\ &= \lim_{\omega\to\infty} \frac{d}{d\omega}\left[\underline{/G(j\omega)} - \omega T\right] \\ &= 0 - T = -T\end{aligned}$$

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 curves. 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 8–23 Determine the transfer function of the system whose experimental frequency-response curves are as shown in Figure 8–96.

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 8–96. We then estimate the corner frequencies. For the system shown in Figure 8–96, the following form of the transfer function is estimated.

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

The value of the damping ratio ζ is estimated by examining the peak resonance near $\omega = 6$ rad/sec. Referring to Figure 8–8, 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

$$G(j\omega) = \frac{10(1 + 0.5j\omega)}{j\omega(1 + j\omega)\left[1 + \left(j\dfrac{\omega}{8}\right) + \left(j\dfrac{\omega}{8}\right)^2\right]}$$

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.

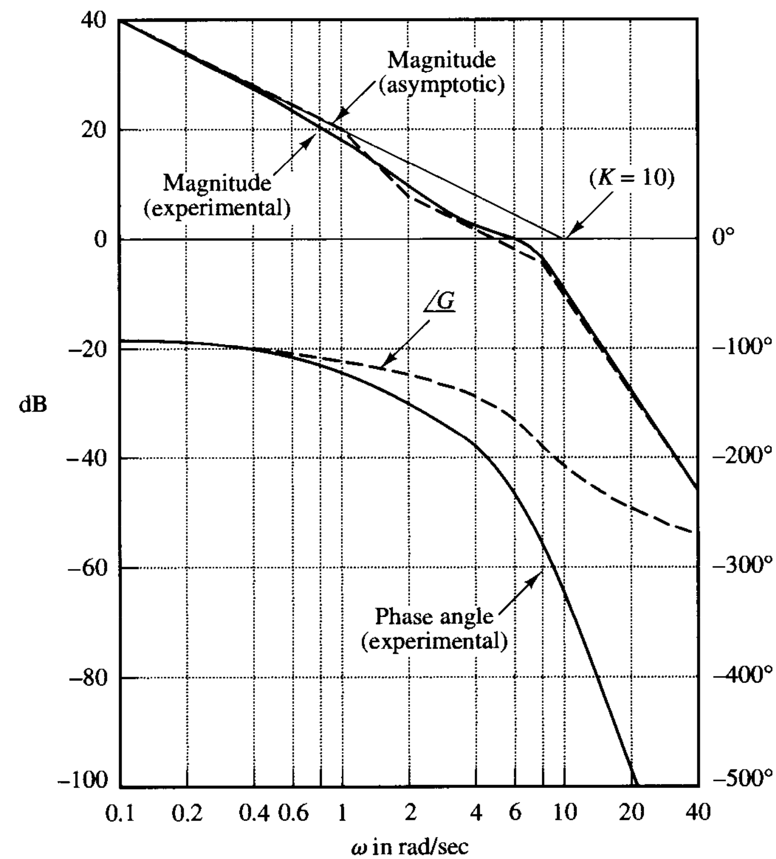


Figure 8–96
Bode diagram of a system. (Solid curves are experimentally obtained curves.)

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 8–96. From Figure 8–96 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.

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

$$G(s)e^{-Ts} = \frac{320(s+2)e^{-0.2s}}{s(s+1)(s^2+8s+64)}$$

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-6-9.) 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 6–51.)

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 then unity.

Solution. Figure 8–97 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.)

A-8-2. Plot a Bode diagram for the following open-loop transfer function $G(s)$:

$$G(s) = \frac{20(s^2 + s + 0.5)}{s(s+1)(s+10)}$$

Solution. By substituting $s = j\omega$ into $G(s)$, we have

$$G(j\omega) = \frac{20[(j\omega)^2 + (j\omega) + 0.5]}{j\omega(j\omega+1)(j\omega+10)}$$

Notice that ω_n and ζ of the quadratic term in the numerator are

$$\omega_n = \sqrt{0.5} \qquad \text{and} \qquad \zeta = 0.707$$

This quadratic term can be written as

$$\omega_n^2\left[\left(j\frac{\omega}{\omega_n}\right)^2 + 2\zeta\left(j\frac{\omega}{\omega_n}\right) + 1\right] = (\sqrt{0.5})^2\left[\left(j\frac{\omega}{\sqrt{0.5}}\right)^2 + 2\times 0.707\left(j\frac{\omega}{\sqrt{0.5}}\right) + 1\right]$$

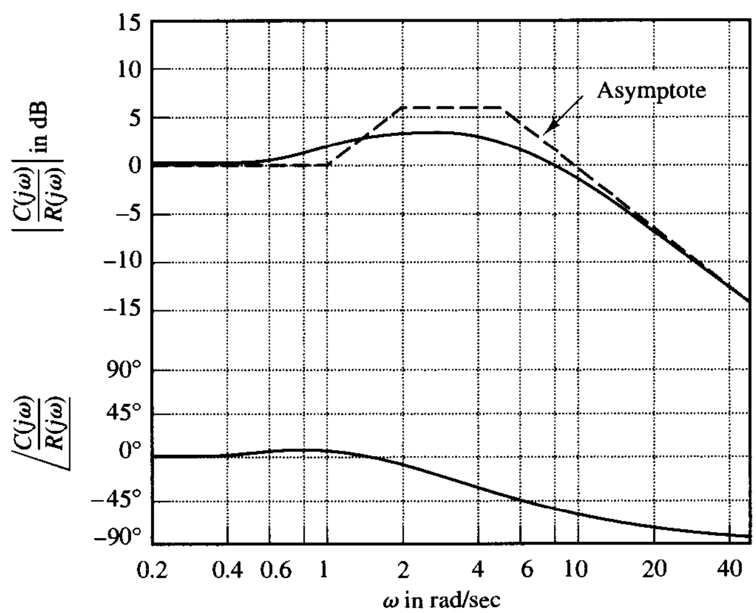


Figure 8–97 Bode diagram for $10(1+j\omega)/[(2+j\omega)(5+j\omega)]$

Note that the corner frequency is at $\omega = \sqrt{0.5} = 0.707$ rad/sec. Now $G(j\omega)$ can be written as

$$G(j\omega) = \frac{\left(j\dfrac{\omega}{\sqrt{0.5}}\right)^2 + 1.414\left(j\dfrac{\omega}{\sqrt{0.5}}\right) + 1}{j\omega(j\omega + 1)(0.1j\omega + 1)}$$

The Bode diagram for $G(j\omega)$ is shown in Figure 8–98.

A-8-3. Draw the Bode diagram of the following nonminimum-phase system:

$$\frac{C(s)}{R(s)} = 1 - Ts$$

Obtain the unit-ramp response of the system and plot $c(t)$ versus t.

Solution. The Bode diagram of the system is shown in Figure 8–99. For a unit-ramp input, $R(s) = 1/s^2$, we have

$$C(s) = \frac{1 - Ts}{s^2} = \frac{1}{s^2} - \frac{T}{s}$$

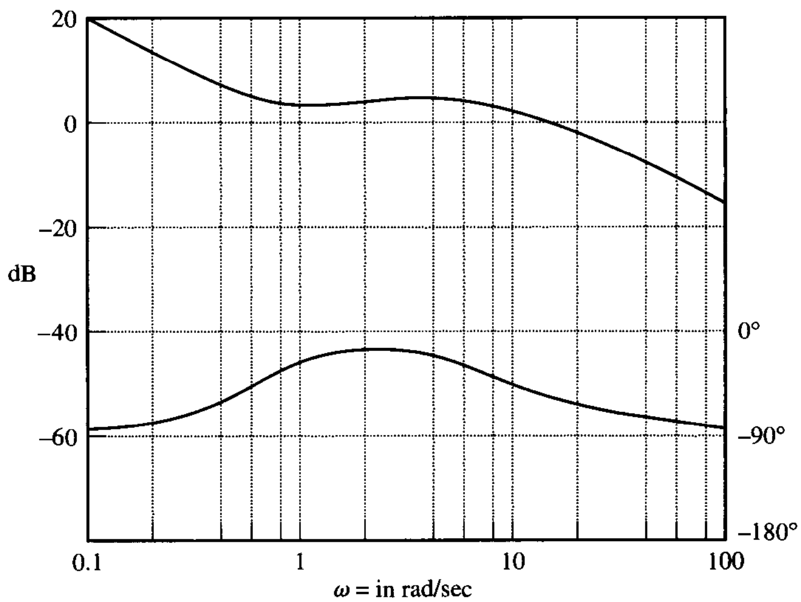


Figure 8–98
Bode diagram for $G(j\omega)$ of Problem A-8-2.

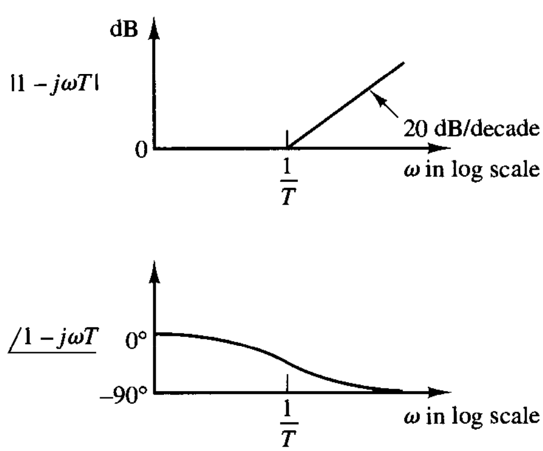


Figure 8–99
Bode diagram of $1 - j\omega T$.

The inverse Laplace transform of $C(s)$ gives

$$c(t) = t - T, \qquad \text{for } t \geq 0$$

Figure 8–100 shows the response curve $c(t)$ versus t. (Note the faulty behavior at the start of the response.) A characteristic property of such a nonminimum-phase system is that the transient response starts out in the opposite direction to the input but eventually comes back in the same direction.

A-8-4. 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 $U_1(j\omega) = 0$. Obtain also the Bode diagrams of these four transfer functions with MATLAB.

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

$$\begin{aligned}
\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} + \mathbf{D} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} s & -1 \\ 25 & s+4 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\
&= \frac{1}{s^2 + 4s + 25} \begin{bmatrix} s+4 & 1 \\ -25 & s \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{s+4}{s^2 + 4s + 25} & \dfrac{s+5}{s^2 + 4s + 25} \\ \dfrac{-25}{s^2 + 4s + 25} & \dfrac{s-25}{s^2 + 4s + 25} \end{bmatrix}
\end{aligned}$$

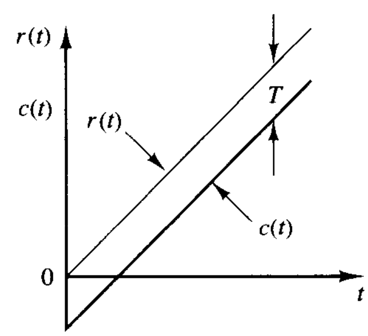


Figure 8–100
Unit-ramp response of the system considered in Problem A-8-3.

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.

To plot Bode diagrams for $Y_1(j\omega)/U_1(j\omega)$, $Y_2(j\omega)/U_1(j\omega)$, $Y_1(j\omega)/U_2(j\omega)$, and $Y_2(j\omega)/U_2(j\omega)$ with MATLAB, we may use the command

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

Then MATLAB produces Bode diagrams when u_1 is the input and u_2 is zero and when u_2 is the input and u_1 is zero. See MATLAB Program 8–14 and the resulting Bode diagrams shown in Figure 8–101. [Note that MATLAB produces two sets of figures (called Figure 1 and Figure 2) on the screen. Figure 8–101 consists of these two sets of Bode diagrams.]

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)
```

A-8-5. Referring to Problem A-8-4, consider an alternative way to plot Bode diagrams of the system. One way to plot the Bode diagrams is to use the command

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

to obtain Bode diagrams for $Y_1(j\omega)/U_1(j\omega)$ and $Y_2(j\omega)/U_1(j\omega)$. To obtain Bode diagrams for $Y_1(j\omega)/U_2(j\omega)$ and $Y_2(j\omega)/U_2(j\omega)$, use the command

```
bode(A,B,C,D,2)
```

Write a MATLAB program to obtain Bode diagrams using these bode commands.

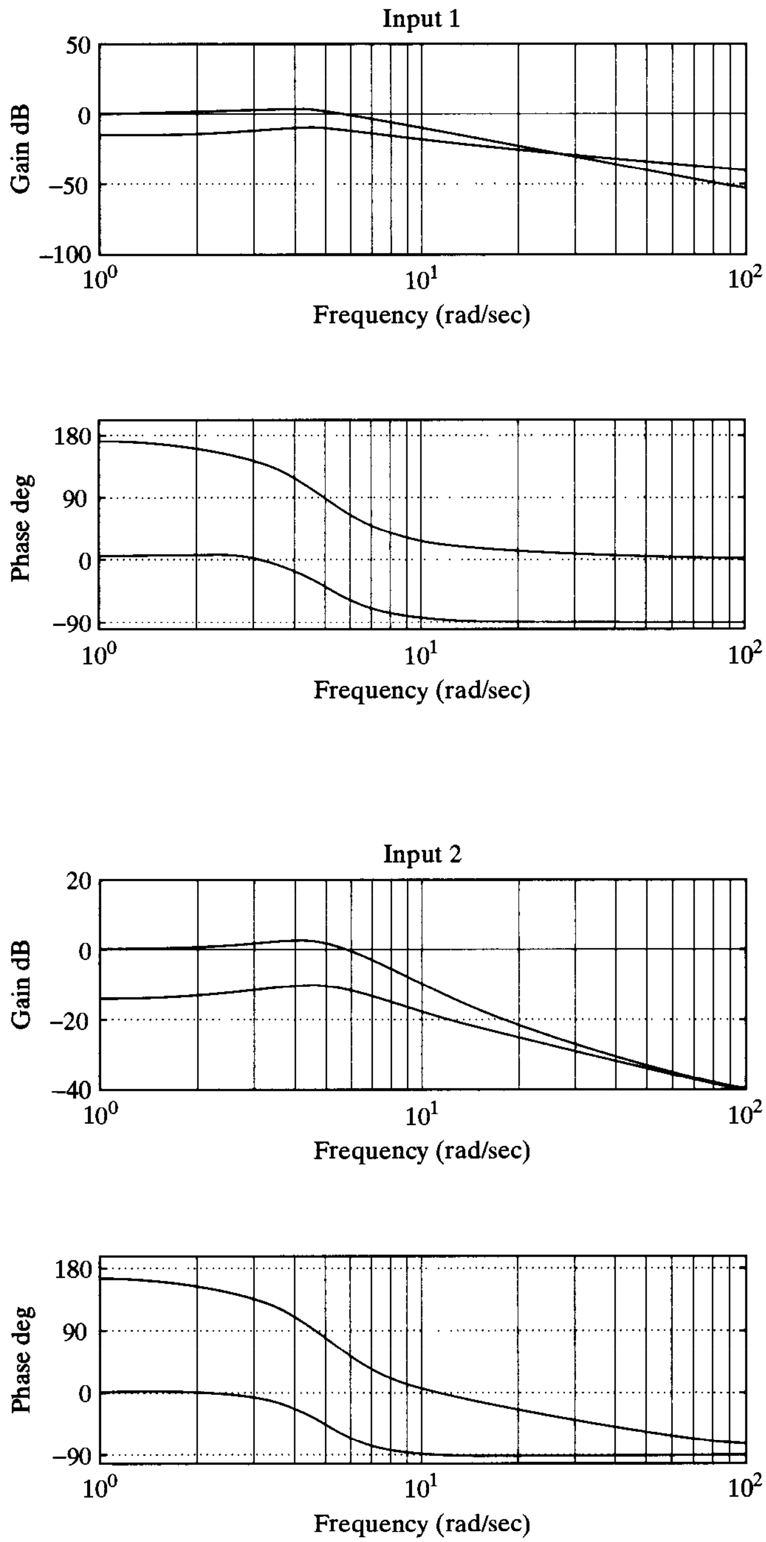


Figure 8–101
Bode diagrams of the system considered in Problem A-8-4.

```
MATLAB Program 8-15

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,1)
subplot(2,1,1), title('Bode Diagrams; Input = u1 (u2 = 0)')
bode(A,B,C,D,2)
subplot(2,1,1), title ('Bode Diagrams; Input = u2 (u1 = 0)')
```

Solution. The program for this problem is given as MATLAB Program 8–15. The Bode diagrams produced by this program are shown in Figure 8–102. In these diagrams it may not be easy to identify which curves are for $Y_1(j\omega)$ or $Y_2(j\omega)$. Ordinarily the text command may be used to identify curves. However, the text command does not apply to the present bode commands. To use the text command, we may use the following command:

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

See Problem A-8-6 for the details.

A-8-6. Referring to Problems A-8-4 and A-8-5, consider plotting Bode diagrams of the same system as discussed in these problems. Use the text command to distinguish curves in the diagrams. Write a possible MATLAB program for plotting Bode diagrams using the following command:

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

Solution. In using the command specified, note that the matrices mag and phase contain the magnitudes of $Y_1(j\omega)$ and $Y_2(j\omega)$ and phase angles for $Y_1(j\omega)$ and $Y_2(j\omega)$ evaluated at each frequency point considered. To get the magnitude of $Y_1(j\omega)$, use the following command:

```
Y1 = mag*[1;0]
```

To convert the magnitude to decibels, use the statement

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

Hence, to convert Y1 to decibels, enter the statement

```
Y1dB = 20*log10(Y1)
```

Similarly, to plot the magnitude of Y2 in decibels, use the following command:

```
Y2 = mag*[0;1]
Y2dB = 20*log10(Y2)
```

Then enter the command

```
semilogx(w,Y1dB,'o',w,Y1dB,'-',w,Y2dB,'x',w,Y2dB,'-')
```

The text command can now be used to write text in the figure. See MATLAB Program 8–16.

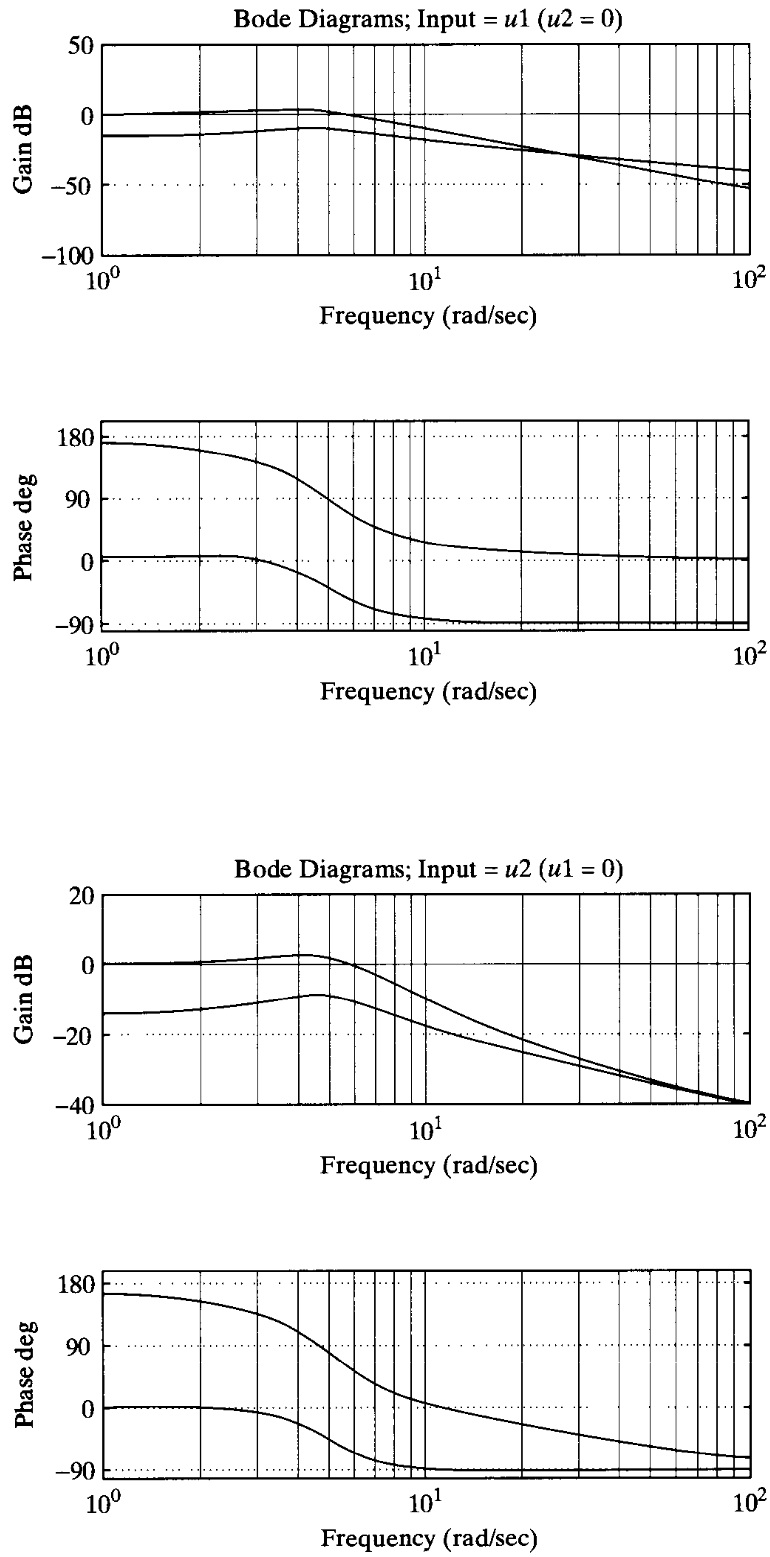


Figure 8–102
Bode diagrams.

MATLAB Program 8–16

```
% ***** We shall first obtain the frequency response when
% u2 = 0; that is, we obtain Y1(jw)/U1(jw) and Y2(jw)/U1(jw) *****

A = [0   1;-25   -4];
B = [1   1;0   1];
C = [1   0;0   1];
D = [0   0;0   0];
w = logspace(-1,3,100);
[mag1, phase1, w] = bode(A,B,C,D,1,w);
Y1 = mag1*[1;0]; Y1dB = 20*log10(Y1);
Y2 = mag1*[0;1]; Y2dB = 20*log10(Y2);
semilogx(w,Y1dB,'o',w,Y1dB,'-',w,Y2dB,'x',w,Y2dB,'-')
grid
subplot(2,1,1);
title('Bode Diagrams: Input = u1 (u2 = 0)')
xlabel('Frequency (rad/sec)')
ylabel('Gain dB')
text(1.2,-10,'Y1'); text(1.2,6,'Y2')

Y1p = phase1*[1;0];
Y2p = phase1*[0;1];
semilogx(w,Y1p,'o',w,Y1p,'-',w,Y2p,'x',w,Y2p,'-')
grid
xlabel('Frequency (rad/sec)')
ylabel('Phase deg')
text(1.2,25,'Y1');text(1.2,188,'Y2')

% ***** In the following, we shall obtain the frequency response
% when u1 = 0; that is, we obtain Y1(jw)/U2(jw) and Y2(jw)/U2(jw) *****

[mag2,phase2,w] = bode(A,B,C,D,2,w);
YY1 = mag2*[1;0]; YY1dB = 20*log10(YY1);
YY2 = mag2*[0;1]; YY2dB = 20*log10(YY2);
semilogx(w,YY1dB,'o',w,YY1dB,'-',w,YY2dB,'x',w,YY2dB,'-')
grid
subplot(2,1,1);
title('Bode Diagrams: Input = u2 (u1 = 0)')
xlabel('Frequency (rad/sec)')
ylabel('Gain dB')
text(1.2, -17,'Y1'); text(1.2,5,'Y2')

YY1p = phase2*[1;0];
YY2p = phase2*[0;1];
semilogx(w,YY1p,'o',w,YY1p,'-',w,YY2p,'x',w,YY2p,'-')
grid
xlabel('Frequency (rad/sec)')
ylabel('Phase deg')
text(1.2,20,'Y1'); text(1.2,182,'Y2')
```

Similarly, to plot the phase angles for $Y_1(j\omega)$ and $Y_2(j\omega)$, use the following commands:

```
Y1p = phase*[1;0];
Y2p = phase*[0;1];
semilogx(w,Y1p,'o',w,Y1p,'-',w,Y2p,'x',w,Y2p,'-')
```

Bode diagrams obtained by use of MATLAB Program 8–16 are shown in Figures 8–103 and 8–104.

A-8-7. 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.

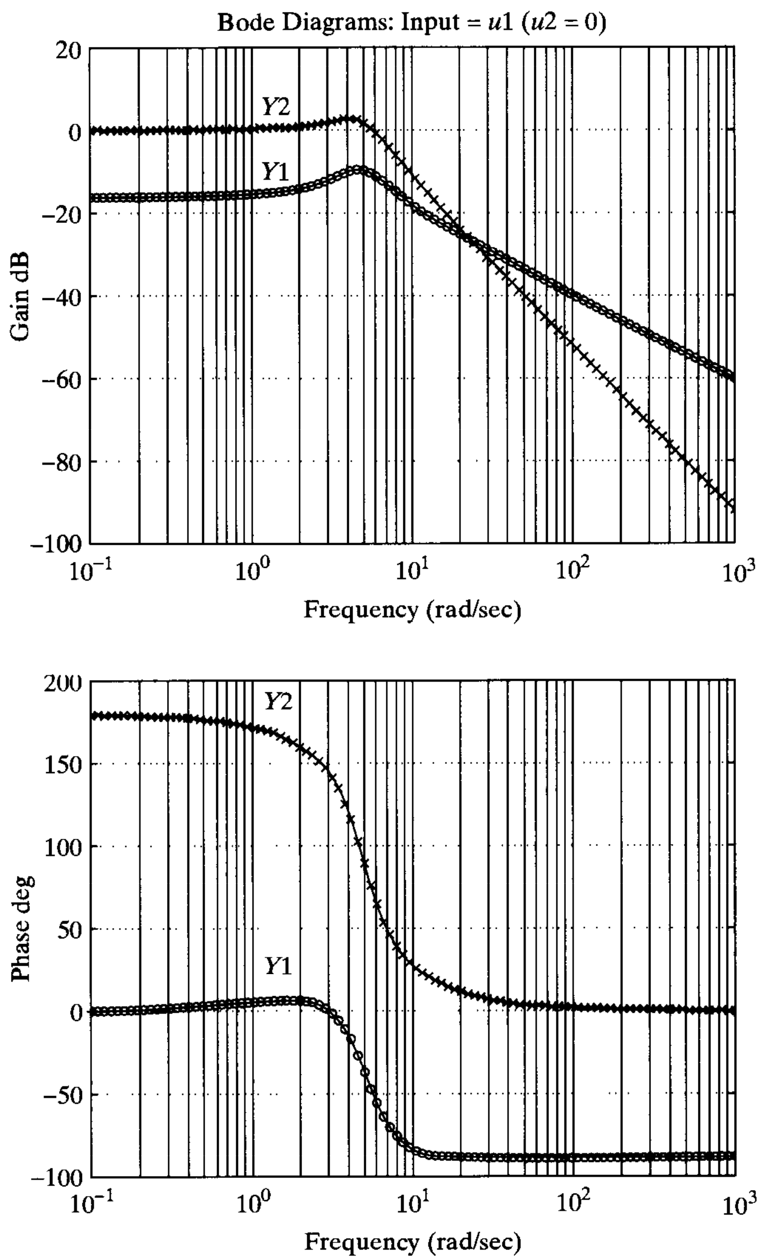


Figure 8–103
Bode diagrams.

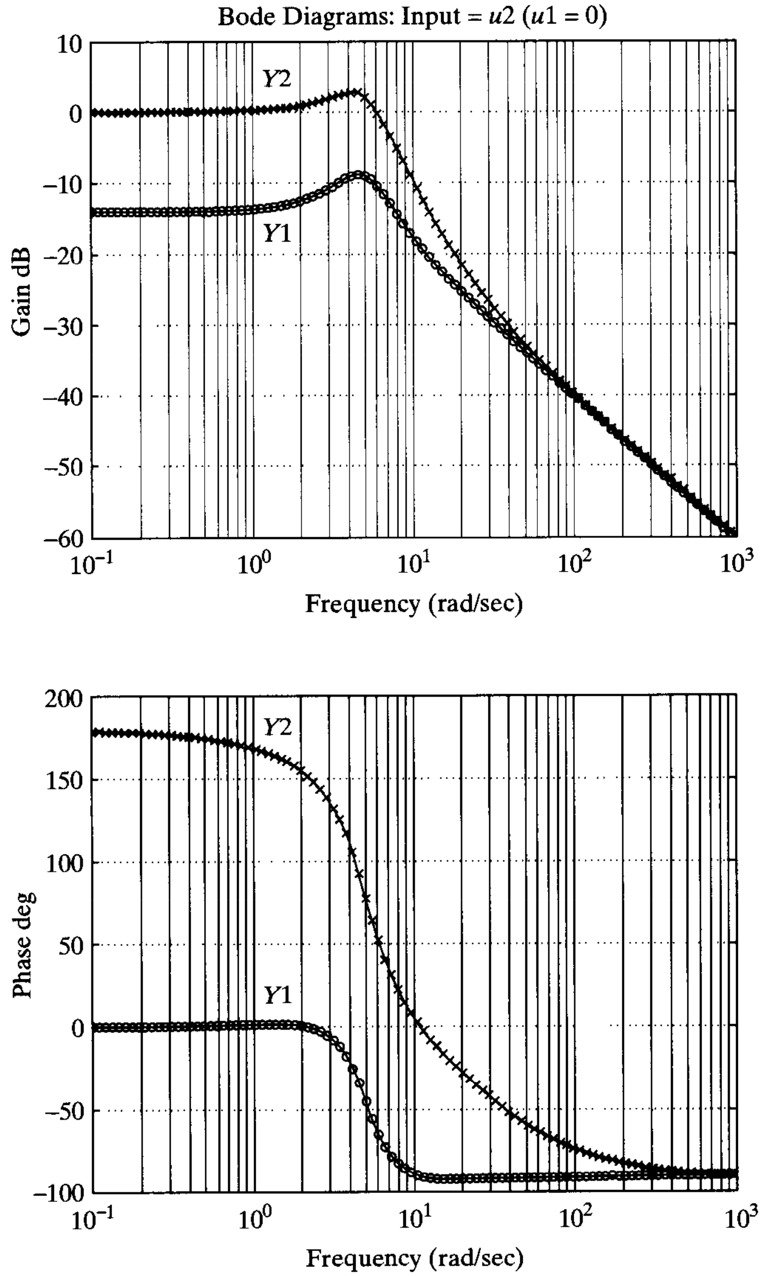


Figure 8–104
Bode diagrams.

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-8-8. Referring to Problem A-8-2, 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 $\omega = 0.1, 0.2, 0.4, 0.6, 1.0, 2.0, 4.0, 6.0, 10.0, 20.0$, and 40.0 rad/sec.

Solution. Noting that

$$G(j\omega) = \frac{2(-\omega^2 + j\omega + 0.5)}{j\omega(j\omega + 1)(0.1\,j\omega + 1)}$$

we have

$$|G(j\omega)| = \frac{2\sqrt{(0.5 - \omega^2)^2 + \omega^2}}{\omega\sqrt{1 + \omega^2}\,\sqrt{1 + 0.01\,\omega^2}}$$

$$\underline{/G(j\omega)} = \tan^{-1}\left(\frac{\omega}{0.5 - \omega^2}\right) -90^\circ - \tan^{-1}\omega - \tan^{-1}(0.1\omega)$$

The magnitude and phase angle may be obtained as shown in Table 8–3. (Note that the magnitude in decibels and phase angle in degrees can be easily read from Figure 8–98.) The magnitude in decibels can also be easily converted to a number. Figure 8–105 shows the polar plot. Notice the existence of a loop in the polar locus.

Table 8–3 Magnitude and Phase of $G(j\omega)$ Considered in Problem A-8-8

ω	$\lvert G(j\omega)\rvert$	$\underline{/G(j\omega)}$
0.1	9.952	−84.75°
0.2	4.918	−78.96°
0.4	2.435	−64.46°
0.6	1.758	−47.53°
1.0	1.573	−24.15°
2.0	1.768	−14.49°
4.0	1.801	−22.24°
6.0	1.692	−31.10°
10.0	1.407	−45.03°
20.0	0.893	−63.44°
40.0	0.485	−75.96°

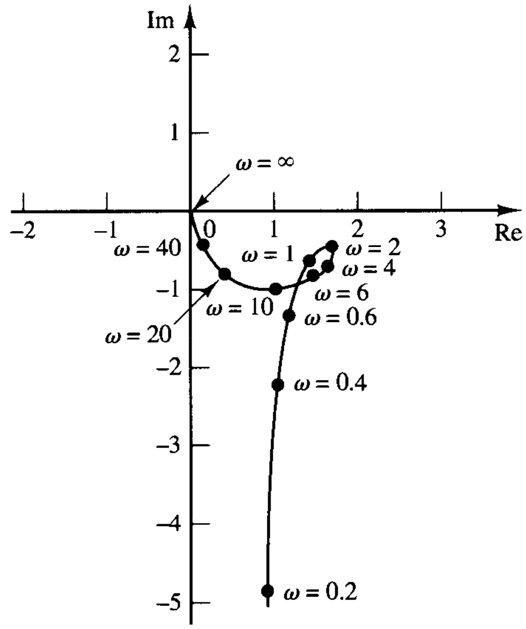


Figure 8–105
Polar plot of
$G(j\omega)$ given in Problem A-8-8.

A-8-9. Consider the function

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

The conformal mapping of the lines $\omega = 0, \pm 1, \pm 2$ and the lines $\sigma = 0, \pm 1, \pm 2$ yield circles in the $F(s)$ plane, as shown in Figure 8–106. Show that 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 in the counterclockwise direction. 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 in the clockwise direction. If the contour in the s plane encloses both the zero and pole or if the contour encloses neither the zero nor pole, then there is no encirclement of the origin of the $F(s)$ plane by the locus of $F(s)$. (Note that in the s plane a representative point s traces out a contour in the clockwise direction.)

Solution. A graphical solution is given in Figure 8–107; this shows closed contours in the s plane and their corresponding closed curves in the $F(s)$ plane.

A-8-10. 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, 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

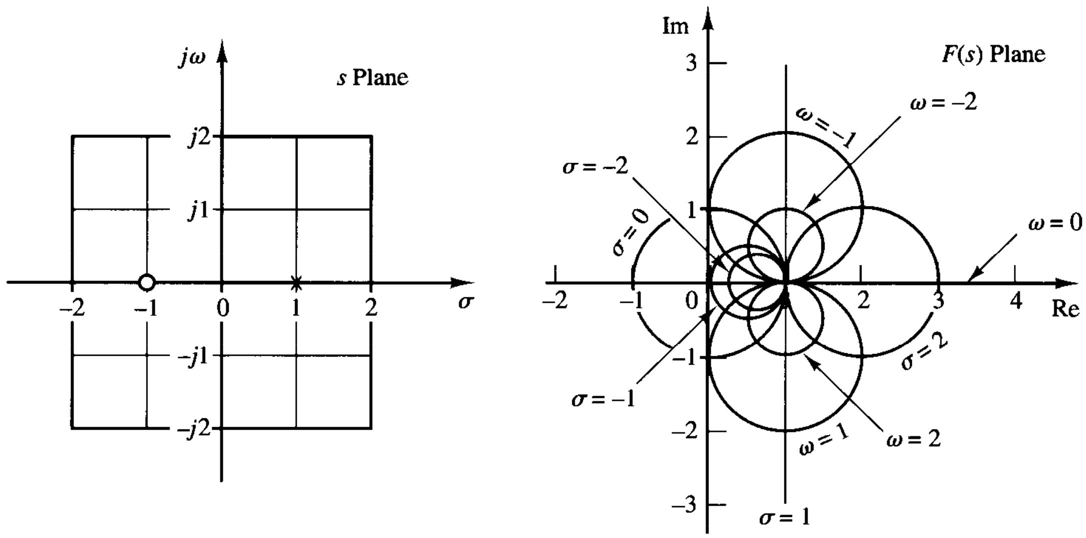


Figure 8–106 Conformal mapping of the s-plane grids into the $F(s)$ plane, where $F(s) = (s + 1)/(s - 1)$.

$$\oint F(s)\, ds = 0$$

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

$$\frac{F'(s)}{F(s)} = \left(\frac{k_1}{s + z_1} + \frac{k_2}{s + z_2} + \cdots\right) - \left(\frac{m_1}{s + p_1} + \frac{m_2}{s + p_2} + \cdots\right) + \frac{X'(s)}{X(s)} \tag{8–19}$$

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,

$$\frac{F'(s)}{F(s)} = \frac{k}{s + z_1} + \frac{X'(s)}{X(s)} \tag{8–20}$$

We see that by taking the ratio $F'(s)/F(s)$ the kth-order zero of $F(s)$ becomes a simple pole of $F'(s)/F(s)$.

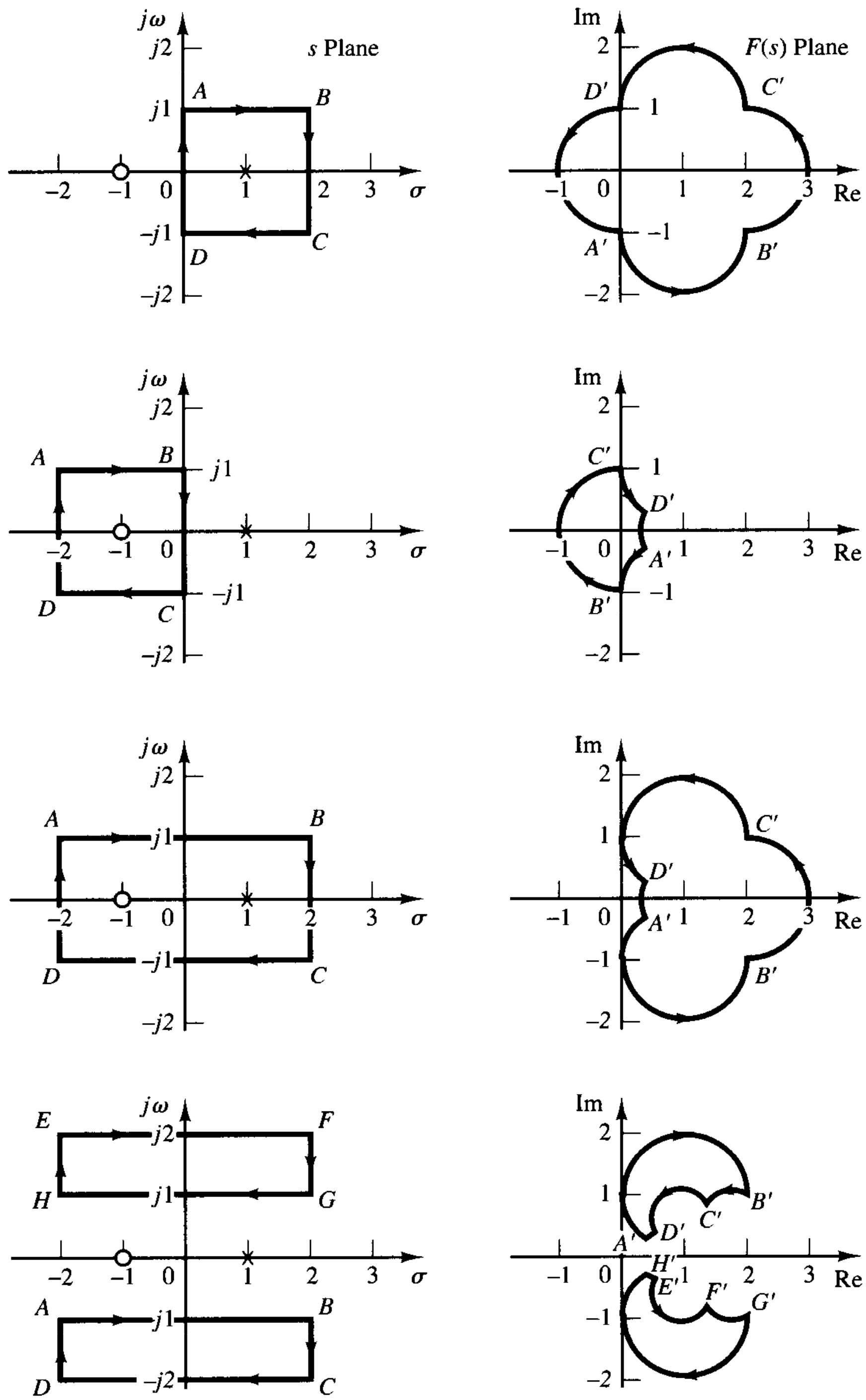


Figure 8–107 Conformal mapping of the s-plane contours into the $F(s)$ plane where $F(s) = (s + 1)/(s - 1)$.

If the last term on the right-hand side of Equation (8–20) 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 (8–19) 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

$$\oint \frac{F'(s)}{F(s)}\,ds = -2\pi j(\Sigma \text{ residues})$$

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

$$\frac{\theta_2 - \theta_1}{2\pi} = -N$$

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 8–108 (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-8-11. The Nyquist plot (polar plot) of the open-loop frequency response of a unity-feedback control system is shown in Figure 8–109. 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?

Solution. Figure 8–110 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-8-12. 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.

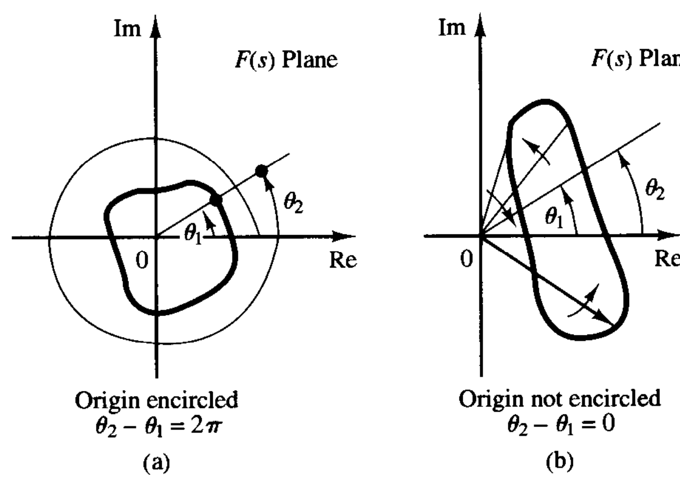


Figure 8–108 Determination of encirclement of the origin of $F(s)$ plane.

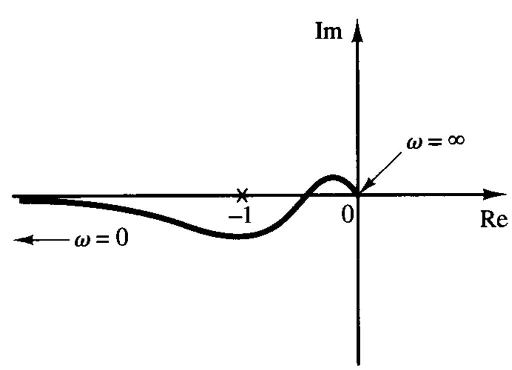


Figure 8–109 Nyquist plot.

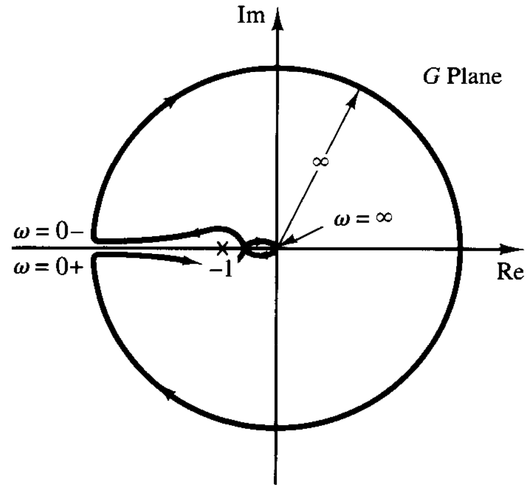


Figure 8–110
Complete Nyquist plot in the G plane.

Solution. The open-loop transfer function is

$$G(j\omega)H(j\omega) = \frac{K}{j\omega(j\omega + 1)(2j\omega + 1)}$$
$$= \frac{K}{-3\omega^2 + j\omega(1 - 2\omega^2)}$$

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.

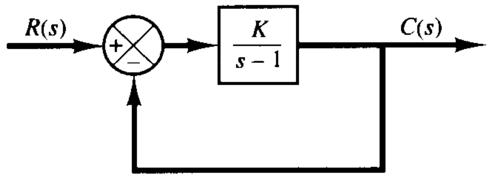


Figure 8–111
Closed-loop system.

A-8-13. Consider the closed-loop system shown in Figure 8–111. Determine the critical value of K for stability by 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 8–112 (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 8–112(b) shows both stable and unstable cases of $G(j\omega)$ plots.

A-8-14. Consider a unity-feedback system whose open-loop transfer function is

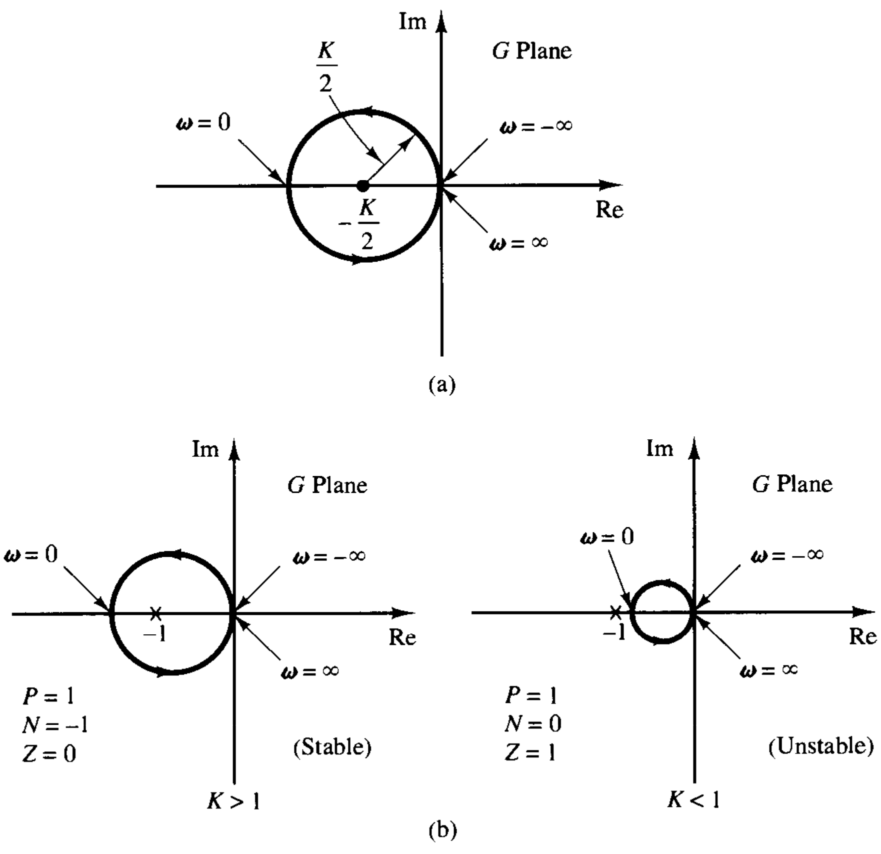


Figure 8–112
(a) Polar plot of $K/(j\omega - 1)$; (b) polar plots of $K/(j\omega - 1)$ for stable and unstable cases.

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

Using the Nyquist plot, determine the critical value of K for stability.

Solution. For this system,

$$\begin{aligned}
G(j\omega) &= \frac{Ke^{-0.8j\omega}}{j\omega + 1} \\
&= \frac{K(\cos 0.8\omega - j\sin 0.8\omega)(1 - j\omega)}{1+\omega^2} \\
&= \frac{K}{1+\omega^2}[(\cos 0.8\omega - \omega \sin 0.8\omega) - j(\sin 0.8\omega + \omega\cos 0.8\omega)]
\end{aligned}$$

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 8–113 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 transport lag of 0.8 sec becomes unstable for $K > 2.65$.

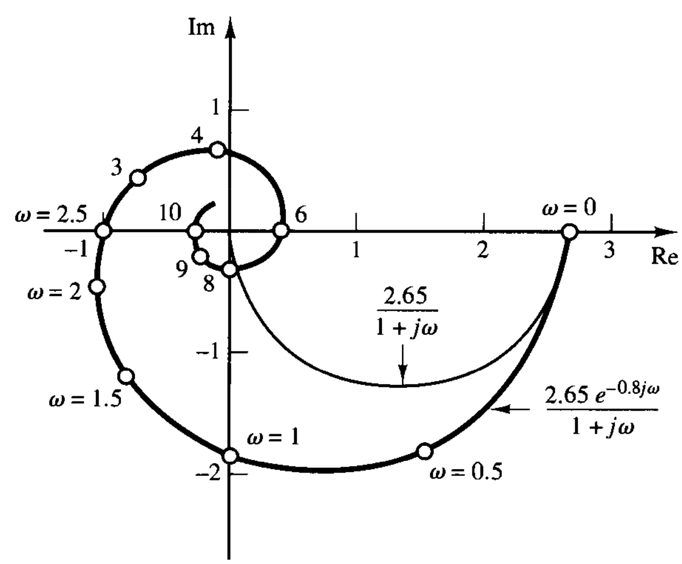


Figure 8–113
Polar plots of $2.65e^{-0.8j\omega}/(1 + j\omega)$ and $2.65/(1 + j\omega)$.

A-8-15. 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. We first enter MATLAB Program 8–17 into the computer. Because in this system MATLAB involves "Divide by zero" in the computation, the resulting Nyquist plot is erroneous, as shown in Figure 8–114.

MATLAB Program 8–17

```
num = [0   20   20   10];
den = [1   11   10   0];
nyquist(num,den)
```

This erroneous Nyquist plot can be corrected by entering the axis command, as shown in MATLAB Program 8–18. The resulting Nyquist plot is shown in Figure 8–115.

MATLAB Program 8–18

```
num = [0   20   20   10];
den = [1   11   10   0];
nyquist(num,d.en)
v = [-2   3   -3   3]; axis(v)
grid
title('Nyquist Plot of G(s) = 20(s^2 + s + 0.5)/[s(s + 1)(s + 10)]')
```

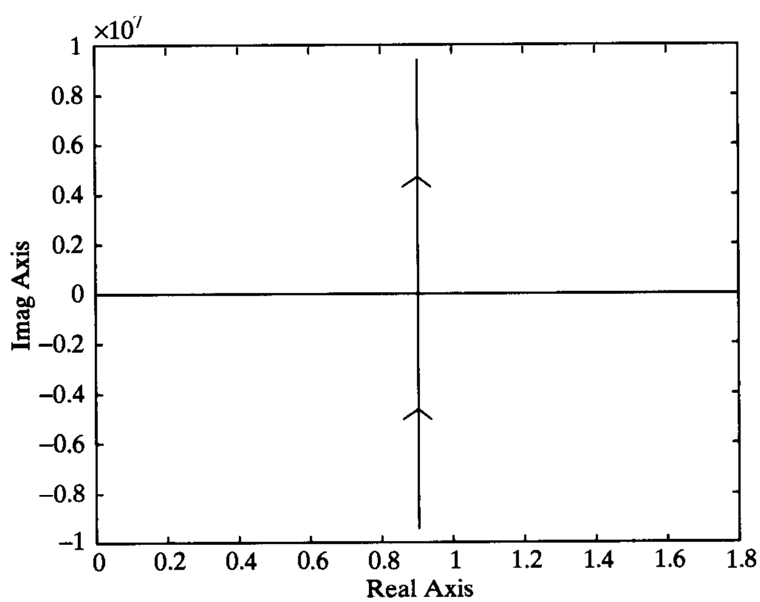


Figure 8–114 Erroneous Nyquist plot.

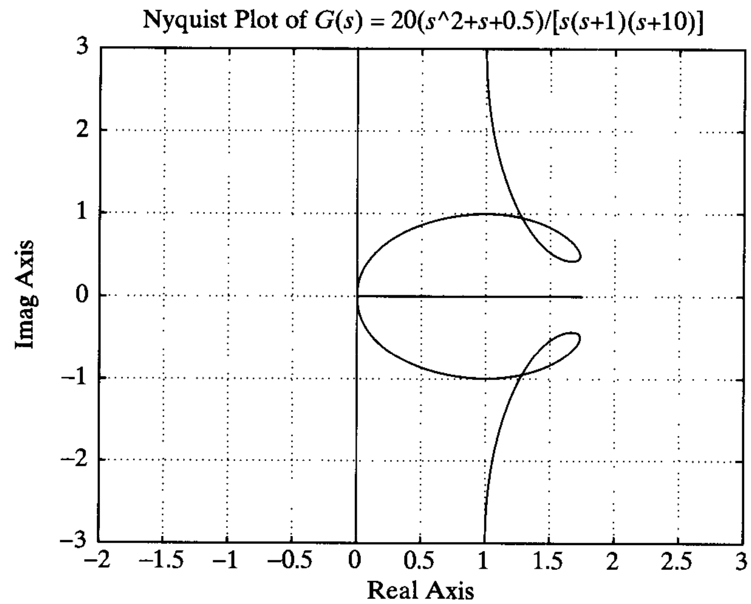


Figure 8–115 Nyquist plot of $G(s) = \dfrac{20(s^2 + s + 0.5)}{s(s + 1)(s + 10)}$.

Since no open-loop poles lie in the right-half s plane, $P = 0$ in the Nyquist stability criterion. From Figure 8–115 we see that the Nyquist plot does not encircle the $-1 + j0$ point. Hence, the closed-loop system is stable.

A-8-16. Consider the same system as discussed in Problem A-8-15. 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 use of the following command:

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

The frequency region may be divided into several subregions 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]
```

MATLAB Program 8–19 uses this frequency region. Using this program, we obtain the Nyquist plot shown in Figure 8–116.

A-8-17. 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]
```

MATLAB Program 8–19

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

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')
```

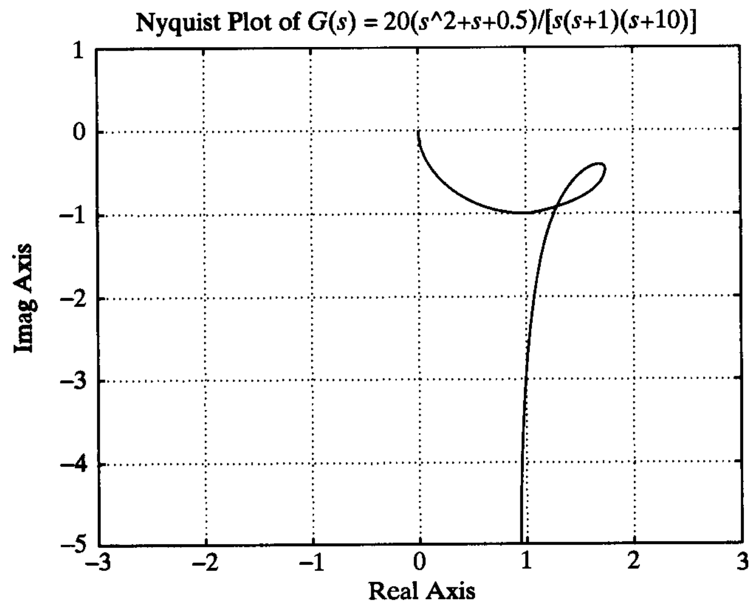


Figure 8–116 Nyquist plot for the positive frequency region.

and using the command nyquist(num,den). MATLAB Program 8–20 produces the Nyquist plot, as shown in Figure 8–117.

MATLAB Program 8–20

```
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)')
```

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

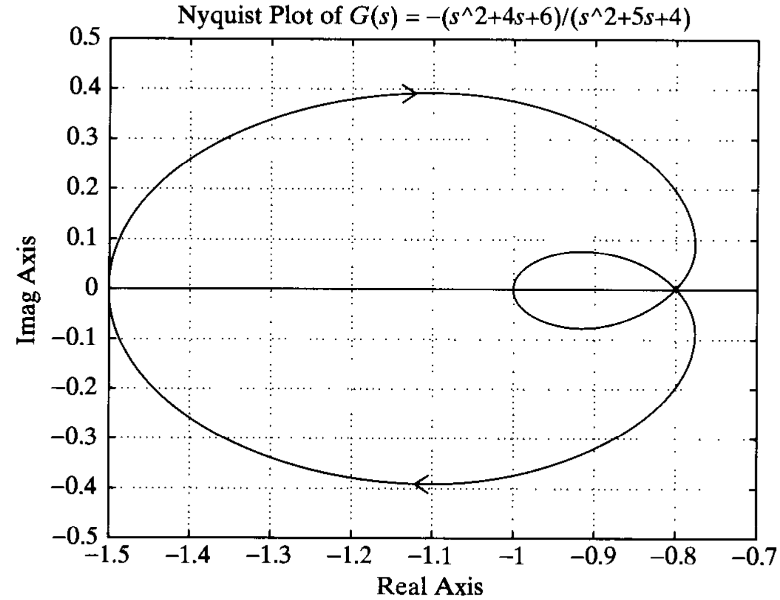


Figure 8–117 Nyquist plot for positive-feedback system.

is an unstable first-order system. See the following closed-loop transfer function of the positive-feedback system:

$$\frac{C(s)}{R(s)} = \frac{s^2 + 4s + 6}{s^2 + 5s + 4 - (s^2 + 4s + 6)}$$
$$= \frac{s^2 + 4s + 6}{s - 2}$$

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 8–118, which was obtained by use of MATLAB Program 8–21.

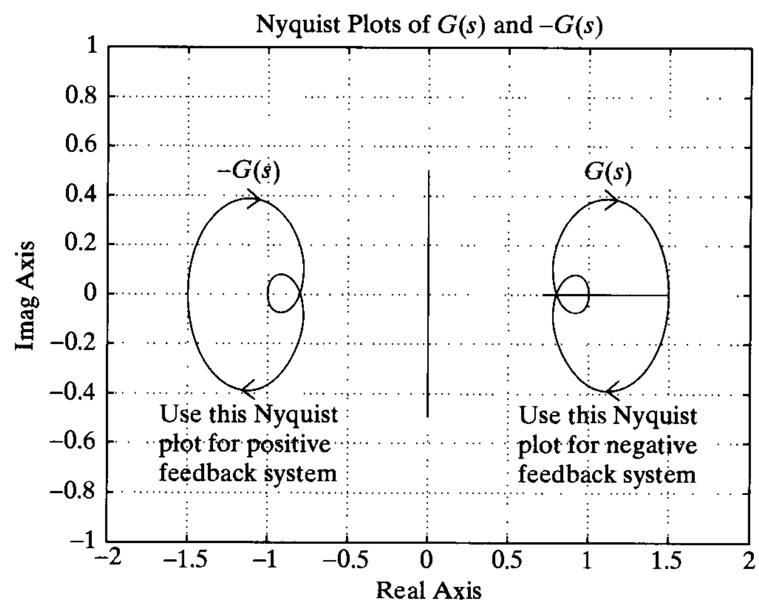


Figure 8–118 Nyquist plots for positive-feedback system and negative-feedback system.

MATLAB Program 8–21

```
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(0.95,0.5,'G(s)')
text(0.57,-0.48,'Use this Nyquist')
text(0.57,-0.62,'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,-0.62,'plot for positive')
text(-1.7,-0.73,'feedback system')
```

A-8-18. Suppose that a system possesses at least one pair of complex-conjugate closed-loop poles. If the $-1 + j0$ point is found at the intersection of a constant-σ curve and a constant-ω curve in the $G(s)$ plane, then those particular values of σ and ω, which we define as $-\sigma_c$ and ω_c, respectively, characterize the closed-loop pole closest to the $j\omega$ axis in the upper-half s plane. (Note that $-\sigma_c$ represents the exponential decay and ω_c represents the damped natural frequency of the step transient-response term due to the pair of the closed-loop poles closest to the $j\omega$ axis.) Probable values of $-\sigma_c$ and ω_c may be estimated from the plot, as shown in Figure 8–119. Thus, the pair of complex-conjugate closed-loop poles that lies closest to the $j\omega$ axis can be determined graphically. It should be noted that all closed-loop poles are mapped into the $-1 + j0$ point in the $G(s)$ plane. Although the complex-conjugate closed-loop poles closest to the $j\omega$ axis can be found easily by this technique, finding other closed-loop poles, if any, by this technique is practically impossible.

If the data on $G(j\omega)$ are experimental, then a curvilinear square near the $-1 + j0$ point can be constructed by extrapolation. Referring to Figure 8–120, we can find the location of the dominant closed-loop poles in the s plane, or the damping ratio ζ and the damped natural frequency ω_d, by drawing the line AB that connects the $-1 + j0$ point (point A) and point B, the nearest approach to the $-1 + j0$ point, and then constructing a curvilinear square $CDEF$, as shown in

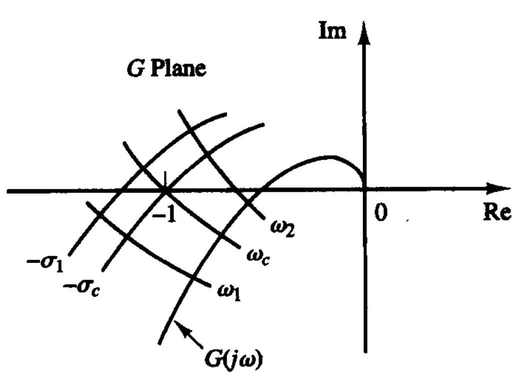


Figure 8–119
Estimation of $-\sigma_c$ and ω_c.

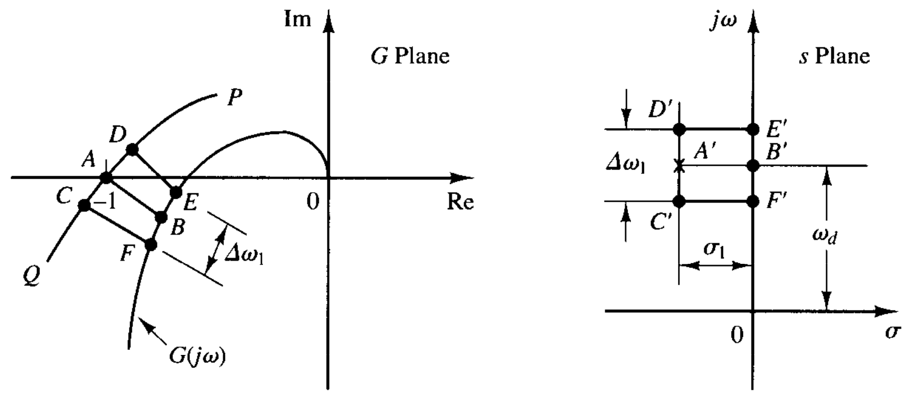


Figure 8–120 Conformal mapping of a curvilinear square near the $-1 + j0$ point in the $G(s)$ plane into the s plane.

Figure 8–120. This curvilinear square $CDEF$ may be constructed by drawing the most likely curve PQ (where PQ is the conformal mapping of a line parallel to the $j\omega$ axis in the s plane) passing through the $-1 + j0$ point and "parallel" to the $G(j\omega)$ locus, and adjusting the points $C, D, E,$ and F such that $\widehat{FB} = \widehat{BE}$, $\widehat{CA} = \widehat{AD}$, and $\widehat{FE} + \widehat{CD} = \widehat{FC} + \widehat{ED}$. The corresponding s-plane contour $C'D'E'F'$, together with point A', the closed-loop pole closest to the $j\omega$ axis, is shown in Figure 8–120. The value of the frequency interval $\Delta\omega_1$ between points E and F is approximately equal to the value of σ_1 shown in Figure 8–120. The frequency at point B is approximately equal to the damped natural frequency ω_d. The closed-loop poles closest to the $j\omega$ axis are then estimated as

$$s = -\sigma_1 \pm j\omega_d$$

Then the damping ratio ζ of these closed-loop poles can be obtained from

$$\frac{\zeta}{\sqrt{1 - \zeta^2}} = \frac{\sigma_1}{\omega_d} = \frac{\Delta\omega_1}{\omega_d}$$

It should be noted that the damped natural frequency ω_d of the step transient response actually is on the frequency contour that passes through the $-1 + j0$ point and is not necessarily the point of nearest approach to the $G(j\omega)$ locus. Therefore, the value ω_d obtained by the technique above is somewhat in error.

From our analysis, we may conclude that it is possible to estimate the closed-loop poles closest to the $j\omega$ axis from the closeness of approach of the $G(j\omega)$ locus to the $-1 + j0$ point, the frequency at the point of nearest approach, and the frequency graduation near this point.

Referring to the frequency-response plot of $G(j\omega)$ of a unity-feedback system, as shown in Figure 8–121, find the closed-loop poles closest to the $j\omega$ axis.

Solution. The line connecting the $-1 + j0$ point and the point of nearest approach of the $G(j\omega)$ locus to the $-1 + j0$ point is drawn first. Then the curvilinear square $ABCD$ is constructed. Since the frequency at the point of nearest approach is $\omega = 2.9$, the damped natural frequency is approximately 2.9, or $\omega_d = 2.9$. From the curvilinear square $ABCD$, it is found that

$$\Delta\omega = \omega_D - \omega_A = 3.4 - 2.4 = 1.0$$

The closed-loop poles closest to the $j\omega$ axis are then estimated as

$$s = -1 \pm j2.9$$

The $G(j\omega)$ locus shown in Figure 8–121 is actually a plot of the following open-loop transfer function:

$$G(s) = \frac{5(s + 20)}{s(s + 4.59)(s^2 + 3.41s + 16.35)}$$

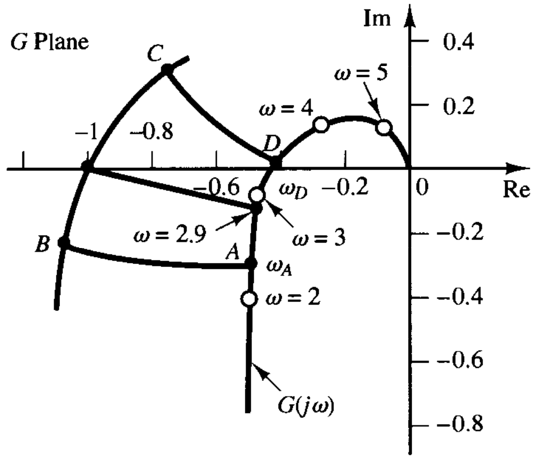


Figure 8–121
Polar plot and a curvilinear square.

The exact closed-loop poles of this system are $s = -1 \pm j3$ and $s = -3 \pm j1$. The closed loop poles closest to the $j\omega$ axis are $s = -1 \pm j3$. In this particular example, we see that the error involved is fairly small. In general, this error depends on a particular $G(j\omega)$ curve. The nearer the $G(j\omega)$ locus is to the $-1 + j0$ point, the smaller the error.

A-8-19. Figure 8–122 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

$$\angle G(j\omega) = \angle j\omega + 2 - 2\angle j\omega = \tan^{-1}\frac{\omega}{2} - 180^\circ$$

The requirement that the phase margin be 50° means that $\angle G(j\omega_c)$ must be equal to -130°, where ω_c is the gain crossover frequency, or

$$\angle G(j\omega_c) = -130^\circ$$

Hence, we set

$$\tan^{-1}\frac{\omega_c}{2} = 50^\circ$$

from which we obtain

$$\omega_c = 2.3835 \text{ rad/sec}$$

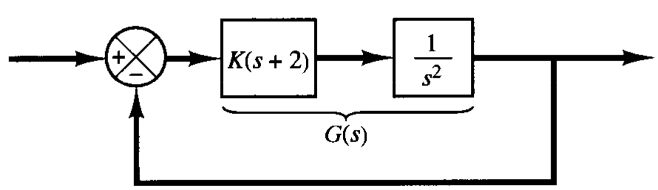


Figure 8–122
Space vehicle control system.

Since the phase curve never crosses the $-180°$ line, the gain margin is $+\infty$ dB. Noting that the magnitude of $G(j\omega)$ must be equal to 0 dB at $\omega = 2.3835$, we have

$$\left|\frac{K(j\omega + 2)}{(j\omega)^2}\right|_{\omega=2.3835} = 1$$

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-8-20. Draw a Bode diagram of the open-loop transfer function $G(s)$ of the closed-loop system shown in Figure 8–123. Determine the phase margin and gain margin.

Solution. Note that

$$G(j\omega) = \frac{20(j\omega + 1)}{j\omega(j\omega + 5)[(j\omega)^2 + 2j\omega + 10]}$$
$$= \frac{0.4(j\omega + 1)}{j\omega(0.2j\omega + 1)\left[\left(\frac{j\omega}{\sqrt{10}}\right)^2 + \frac{2}{10}j\omega + 1\right]}$$

The quadratic term in the denominator has the corner frequency of $\sqrt{10}$ rad/sec and the damping ratio ζ of 0.3162, or

$$\omega_n = \sqrt{10}, \qquad \zeta = 0.3162$$

The Bode diagram of $G(j\omega)$ is shown in Figure 8–124. From this diagram we find the phase margin to be 100° and the gain margin to be +13.3 dB.

A-8-21. 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

$$\omega_b = \omega_n(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2})^{1/2}$$

Note that ω_b/ω_n is a function only of ζ. Plot a curve 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,

$$\left|\frac{C(j\omega_b)}{R(j\omega_b)}\right| = \left|\frac{\omega_n^2}{(j\omega_b)^2 + 2\zeta\omega_n(j\omega_b) + \omega_n^2}\right| = 0.707$$

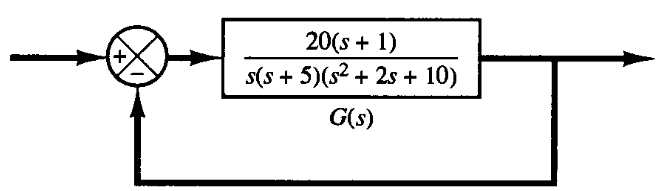


Figure 8–123
Closed-loop system.

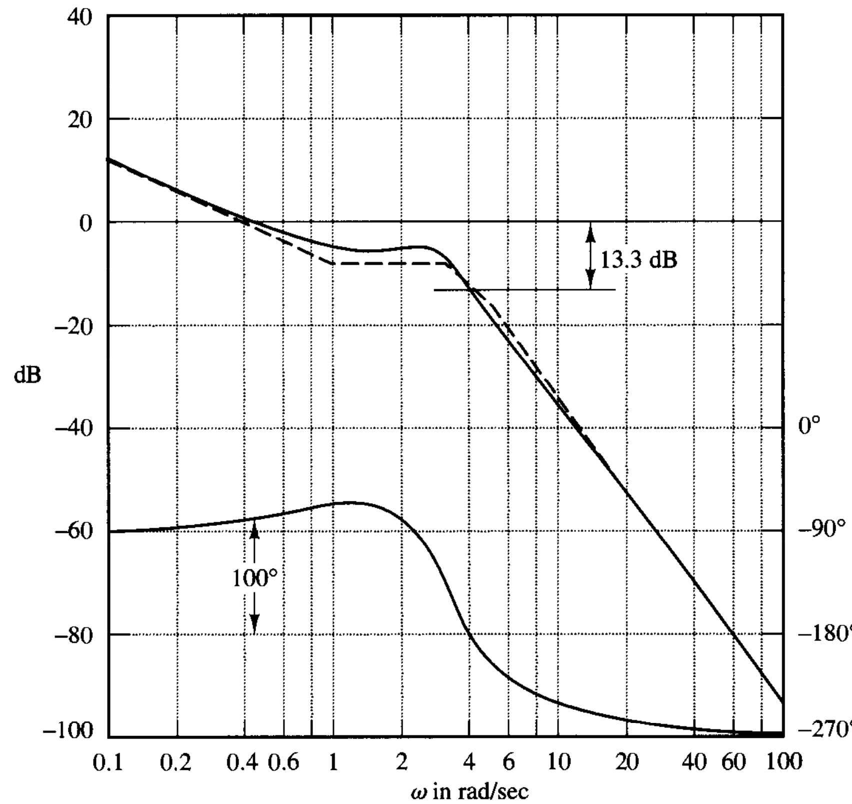


Figure 8–124 Bode diagram of $G(j\omega)$ of system shown in Figure 8–123.

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

$$\omega_n^4 = 0.5[(\omega_n^2 - \omega_b^2)^2 + 4\zeta^2\omega_n^2\omega_b^2]$$

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

$$\omega_b^2 = \omega_n^2(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2})$$

or

$$\omega_b = \omega_n(1 - 2\zeta^2 + \sqrt{4\zeta^4 - 4\zeta^2 + 2})^{1/2}$$

Figure 8–125 shows a curve relating ω_b/ω_n versus ζ.

A-8-22. A unity-feedback control system has the following open-loop transfer function:

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

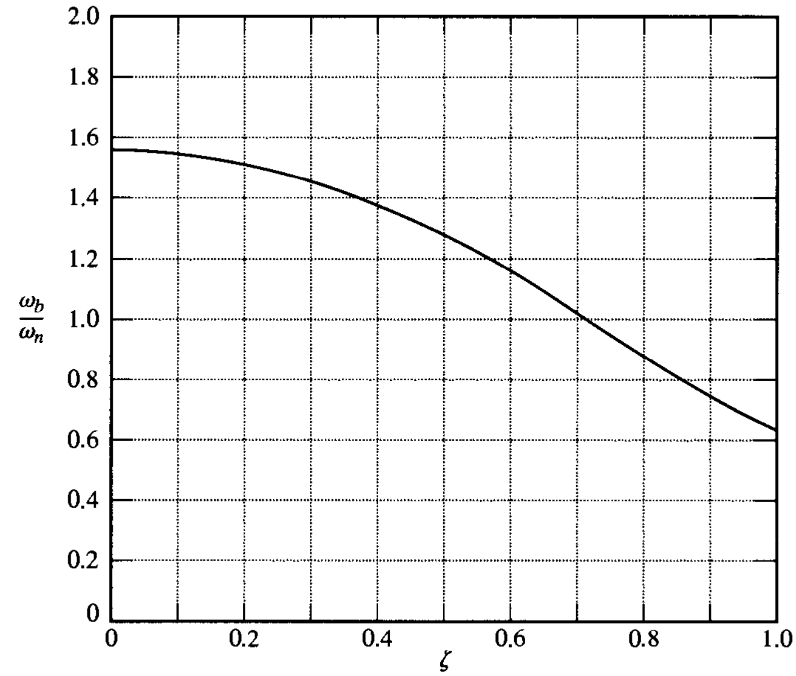


Figure 8–125
Curve ω_b/ω_n versus ζ, where ω_b is the bandwidth.

Consider the frequency response of this system. Plot a polar locus of $G(j\omega)/K$. Then determine the value of gain K such that the resonant peak magnitude M_r, of the closed-loop frequency response is 2.

Solution. A plot of $G(j\omega)/K$ is shown in Figure 8–126. The value of angle ψ corresponding to $M_r = 2$ obtained from Equation (8–18) is

$$\psi = \sin^{-1}\frac{1}{2} = 30°$$

Hence, we draw the line $\overline{OP}$ that passes through the origin and makes an angle of 30° with the negative real axis as shown in Figure 8–126. We then draw the circle that is tangent to both the $G(j\omega)/K$ locus and line $\overline{OP}$. Define the point where the circle and line $\overline{OP}$ are tangent as point P. The perpendicular line drawn from point P intersects the negative real axis at $(-0.445,0)$. Then the gain K is determined as

$$K = \frac{1}{0.445} = 2.247$$

From Figure 8–126 we notice that the resonant frequency is approximately $\omega = 0.83$ rad/sec.

A-8-23. Figure 8–127 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

$$G(s) = \frac{80e^{-0.1s}}{s(s+4)(s+10)} = \frac{2e^{-0.1s}}{s(0.25s+1)(0.1s+1)}$$

we have

$$G(j\omega) = \frac{2e^{-0.1j\omega}}{j\omega(0.25j\omega+1)(0.1j\omega+1)}$$

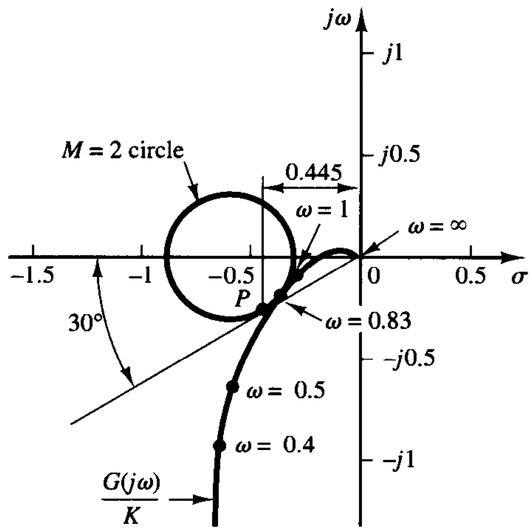


Figure 8–126
Plot of $G(j\omega)/K$ of the system considered in Problem A-8-22.

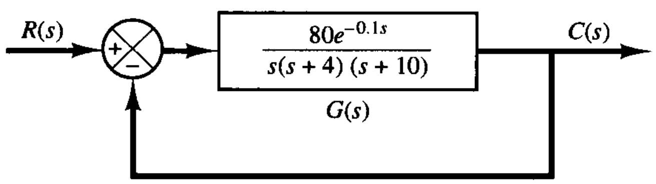


Figure 8–127
Block diagram of a chemical reactor system.

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 8–128.

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 8–129 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 8–130 depicts the Bode diagram of the closed-loop frequency response $G(j\omega)/[1 + G(j\omega)]$.

A-8-24. A Bode diagram of the open-loop transfer function $G(s)$ of a unity-feedback control system is shown in Figure 8–131. 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 8–8 and examining the Bode diagram of Figure 8–131, we find the damping ratio ζ and undamped natural frequency ω_n of the quadratic term to be

$$\zeta = 0.1, \qquad \omega_n = 2 \text{ rad/sec}$$

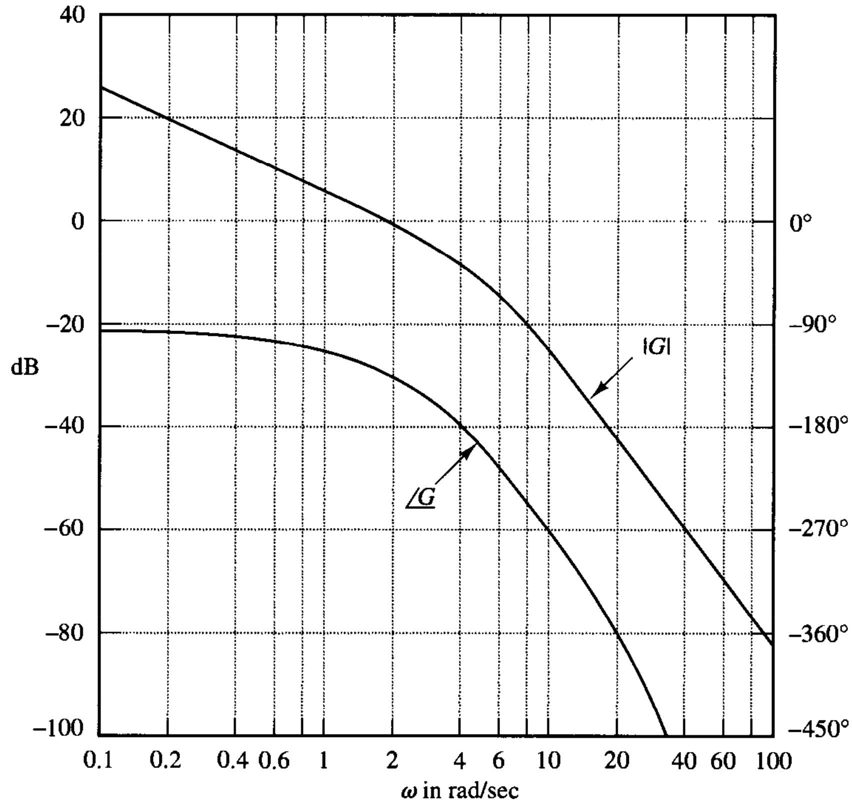


Figure 8–128 Bode diagram of $G(j\omega)$ of the system shown in Figure 8–127.

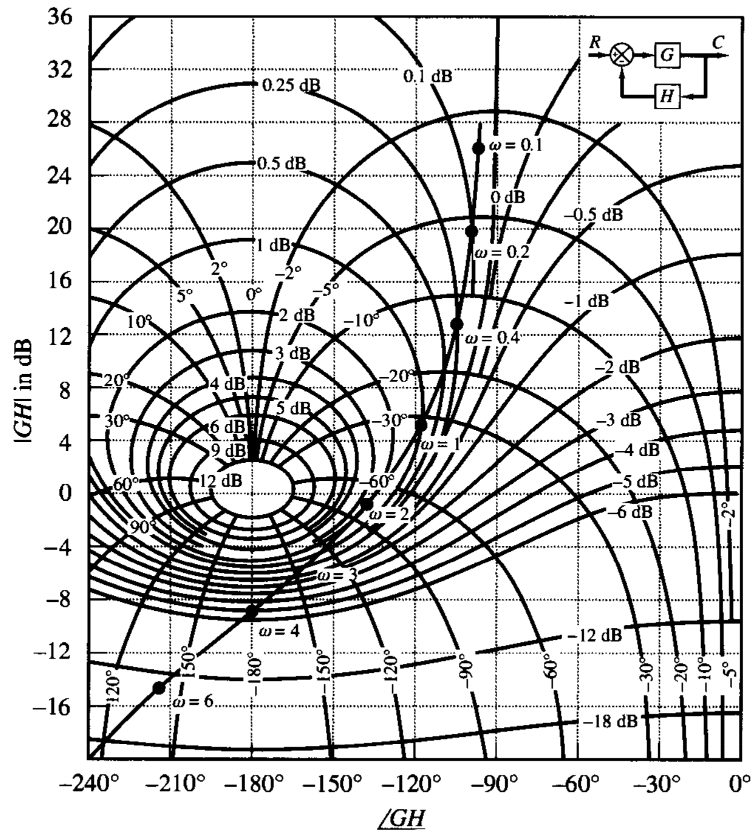


Figure 8–129 $G(j\omega)$ locus superimposed on Nichols chart (Problem A-8-23).

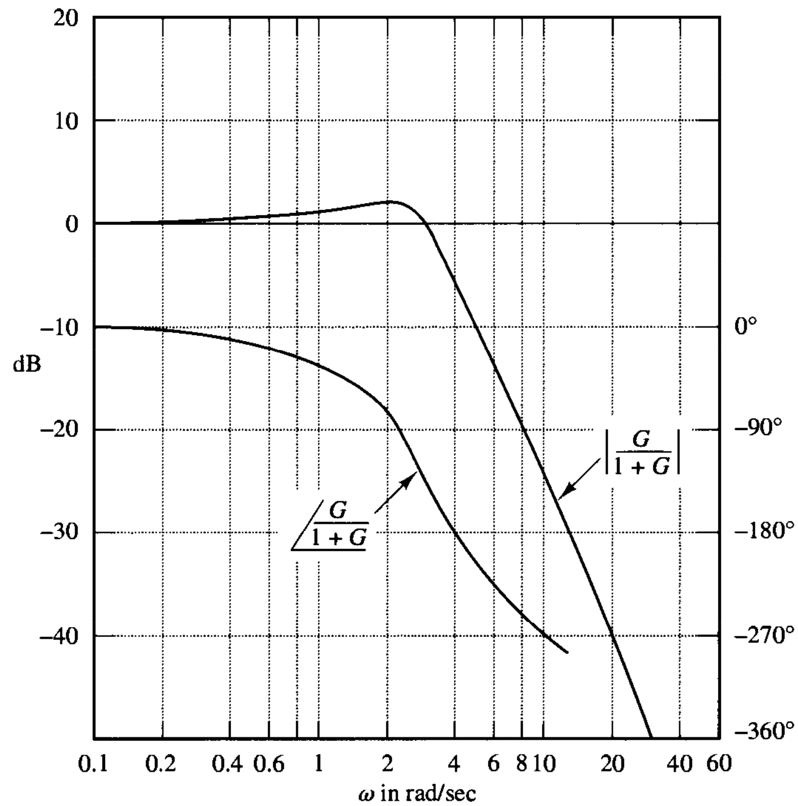


Figure 8–130
Bode diagram of the closed-loop frequency response (Problem A-8-23).

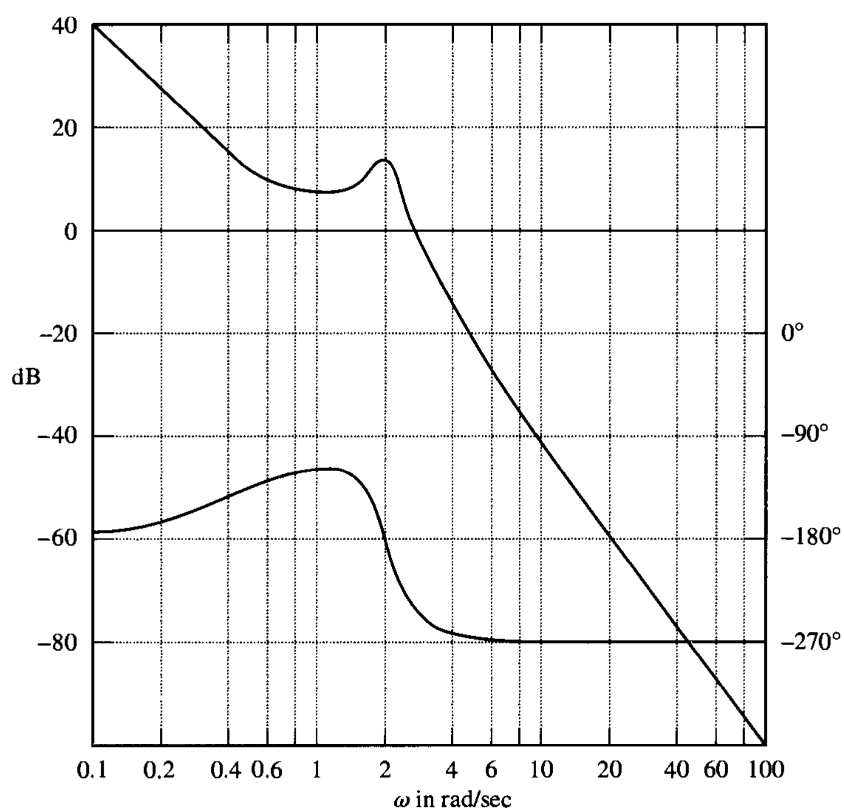


Figure 8–131
Bode diagram of the open-loop transfer function of a unity-feedback control system.

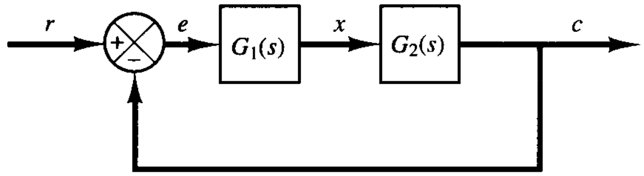


Figure 8–132
Control system.

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:

$$G(j\omega) = \frac{K\left(\frac{j\omega}{0.5} + 1\right)}{(j\omega)^2\left[\left(\frac{j\omega}{2}\right)^2 + 0.1(j\omega) + 1\right]}$$

Since from Figure 8–131 we find $|G(j0.1)| = 40$ dB, the gain value K can be determined as unity. Also, the calculated phase curve, $\angle G(j\omega)$ versus ω, agrees with the given phase curve. Hence, the transfer function $G(s)$ can be determined as

$$G(s) = \frac{4(2s + 1)}{s^2(s^2 + 0.4s + 4)}$$

A-8-25. 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 8–132. 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 functions.

$$G(s) = \frac{10}{s + 1}$$

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. Sketch the Bode diagrams of the following three transfer functions:

(a) $G(s) = \dfrac{T_1 s + 1}{T_2 s + 1} \quad (T_1 > T_2 > 0)$

(b) $G(s) = \dfrac{T_1 s - 1}{T_2 s + 1} \quad (T_1 > T_2 > 0)$

(c) $G(s) = \dfrac{-T_1 s + 1}{T_2 s + 1} \quad (T_1 > T_2 > 0)$

B-8-4. Plot the Bode diagram of

$$G(s) = \frac{10(s^2 + 0.4s + 1)}{s(s^2 + 0.8s + 9)}$$

B-8-5. Given

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

show that

$$|G(j\omega_n)| = \frac{1}{2\zeta}$$

B-8-6. 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-8-7. 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-8-8. The pole–zero configurations of complex functions $F_1(s)$ and $F_2(s)$ are shown in Figures 8–133 (a) and (b), respectively. Assume that the closed contours in the s plane are those shown in Figures 8–133 (a) and (b). Sketch qualitatively the corresponding closed contours in the $F_1(s)$ plane and $F_2(s)$ plane.

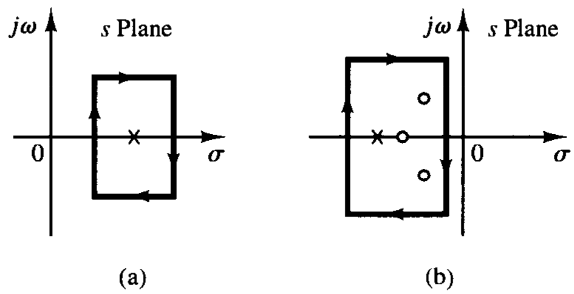


Figure 8–133 (a) s-Plane representation of complex function $F_1(s)$ and a closed contour; (b) s-Plane representation of complex function $F_2(s)$ and a closed contour.

B-8-9. 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-8-10. 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-8-11. 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-8-12. 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-8-13. Draw a Nyquist plot of the following $G(s)$:

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

B-8-14. 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-8-15. 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.

B-8-16. 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-8-17. Referring to Problem B-8-16, 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-8-18. 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-8-19. Consider the system shown in Figure 8–134. Draw a Bode diagram of the open-loop transfer function $G(s)$. Determine the phase margin and gain margin.

B-8-20. Consider the system shown in Figure 8–135. Draw a Bode diagram of the open-loop transfer function $G(s)$. Determine the phase margin and gain margin.

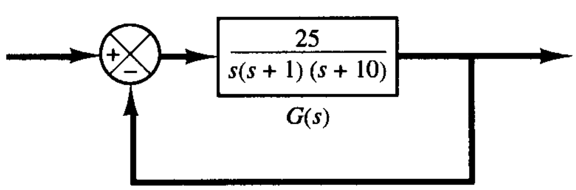


Figure 8–134 Control system.

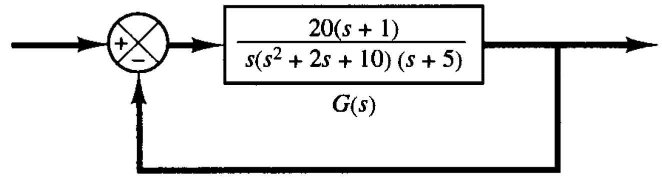


Figure 8–135 Control system.

B-8-21. 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-8-22. Consider the system shown in Figure 8–136. 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?

B-8-23. 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-8-24. Figure 8–137 shows a block diagram of a process control system. Determine the range of gain K for stability.

B-8-25. 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 dead time T.

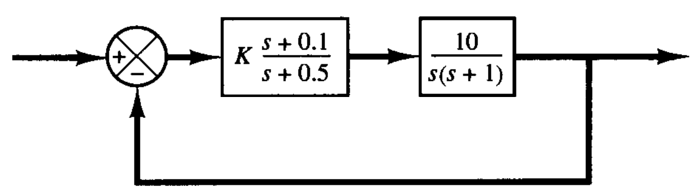


Figure 8–136 Control system.

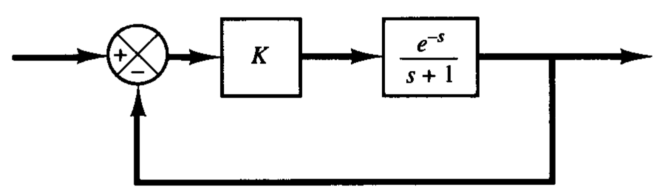


Figure 8–137 Process control system.

B-8-26. 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-8-27. Figure 8–138 shows a Bode diagram of a transfer function $G(s)$. Determine this transfer function.

B-8-28. The experimentally determined Bode diagram of a system $G(j\omega)$ is shown in Figure 8–139. Determine the transfer function $G(s)$.

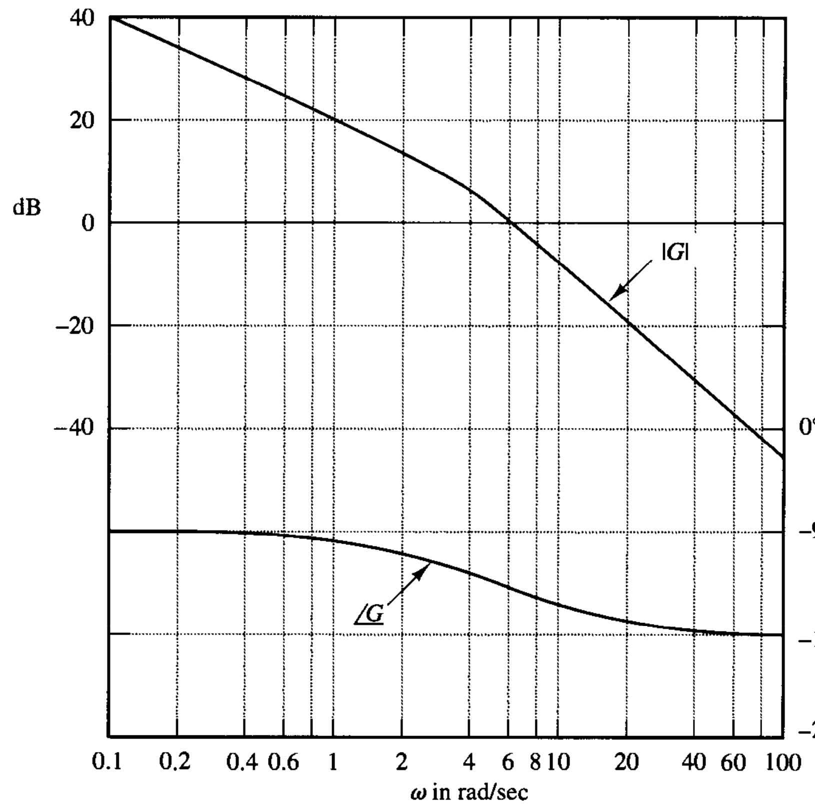


Figure 8–138 Bode diagram of a transfer function $G(s)$.

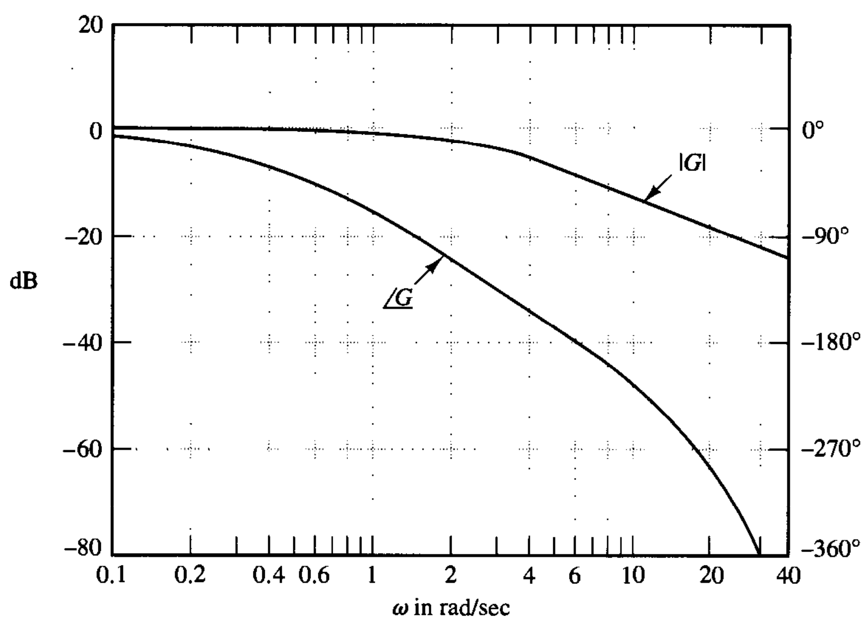


Figure 8–139 Experimentally determined Bode diagram of a system.

9

Control Systems Design by Frequency Response

9–1 INTRODUCTION

The primary objective of this chapter is to present procedures for the design and compensation of single-input–single-output, linear, time-invariant control systems by the frequency-response approach.

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.

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 contradictory to each other.

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.

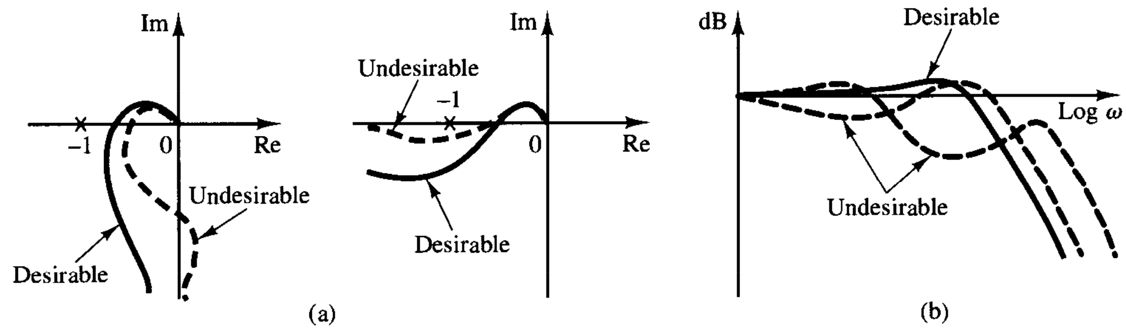
Requirements on open-loop frequency response. We might say that, in many practical cases, compensation is essentially a compromise between steady-state accuracy and relative stability.

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 also, near the gain crossover frequency, the slope of the log-magnitude curve in the Bode diagram should 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.

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.



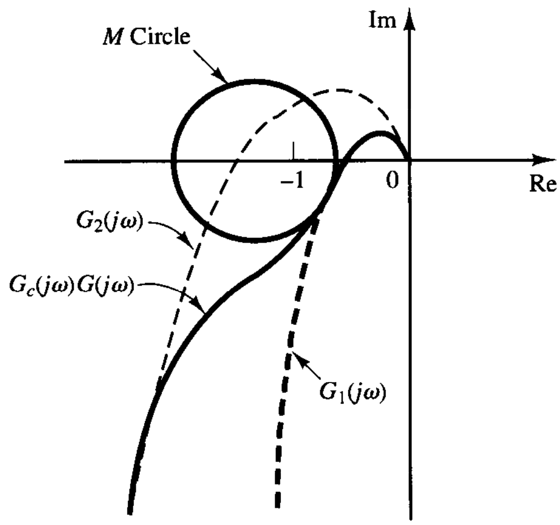


Figure 9–2
Reshaping of the open-loop frequency-response curve.

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 transfer 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:

$$K_c\alpha\frac{Ts+1}{\alpha Ts+1} = K_c\frac{s+\dfrac{1}{T}}{s+\dfrac{1}{\alpha T}} \qquad (0<\alpha<1)$$

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°.)

Figure 9–3 shows the polar plot of

$$K_c\alpha\frac{j\omega T+1}{j\omega\alpha T+1} \qquad (0<\alpha<1)$$

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,

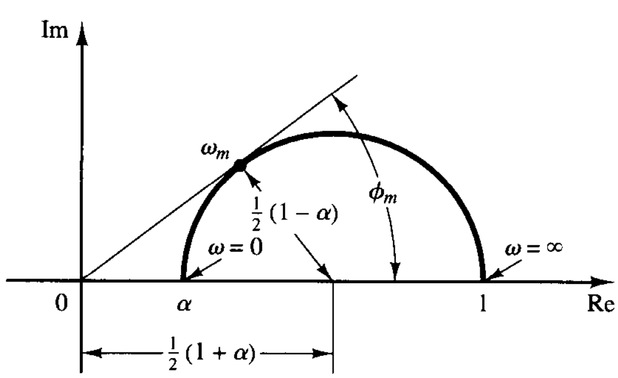


Figure 9–3
Polar plot of a lead compensator $\alpha(j\omega T + 1)/(j\omega\alpha T + 1)$, where $0 < \alpha < 1$.

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:

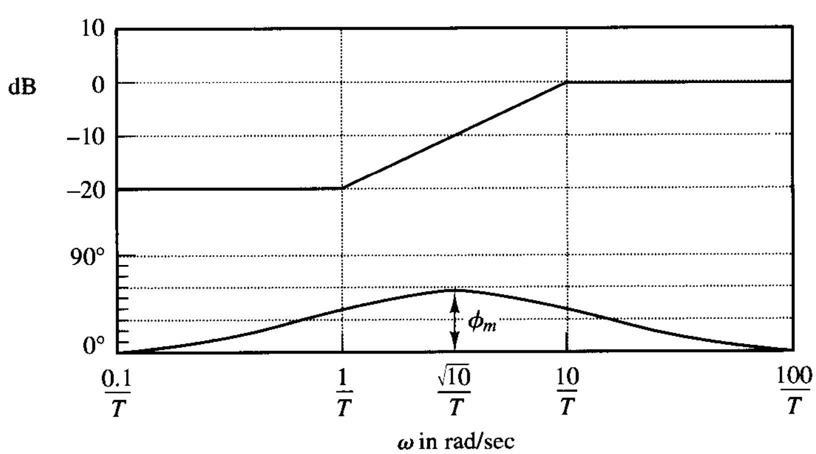


Figure 9–4 Bode diagram of a lead compensator $\alpha(j\omega T + 1)/(j\omega\alpha T + 1)$, where $\alpha = 0.1$.

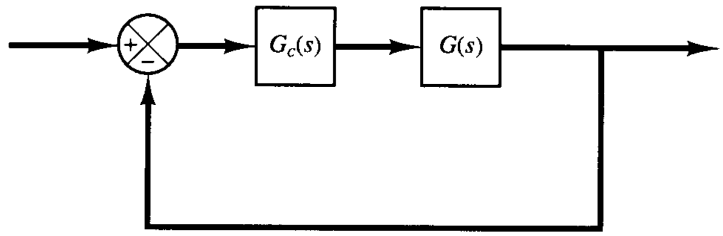


Figure 9–5
Control system.

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 ϕ to be added to the system.

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:

$$\text{Zero of lead compensator:} \quad \omega = \frac{1}{T}$$

$$\text{Pole of lead compensator:} \quad \omega = \frac{1}{\alpha T}$$

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

$$G_c(s) = K_c\alpha\frac{Ts + 1}{\alpha Ts + 1} = K_c\frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}}$$

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 9–7 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

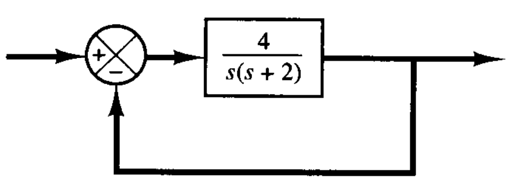


Figure 9–6 Control system.

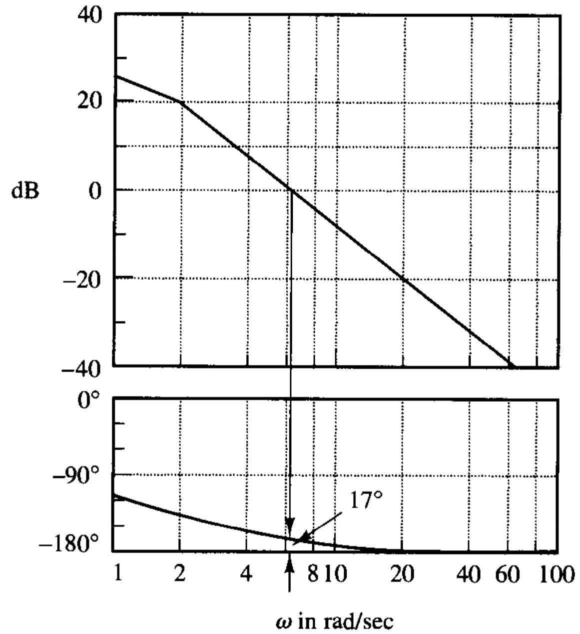


Figure 9–7
Bode diagram for
$G_1(j\omega) = 10G(j\omega) = 40/[j\omega(j\omega + 2)]$

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

$$\left|\frac{1 + j\omega T}{1 + j\omega\alpha T}\right|_{\omega=1/(\sqrt{\alpha}T)} = \left|\frac{1 + j\dfrac{1}{\sqrt{\alpha}}}{1 + j\alpha\dfrac{1}{\sqrt{\alpha}}}\right| = \frac{1}{\sqrt{\alpha}}$$

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. 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.

Figure 9–10 shows the polar plots of the uncompensated system $G_1(j\omega) = 10G(j\omega)$ and the compensated system $G_c(j\omega)G(j\omega)$. 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

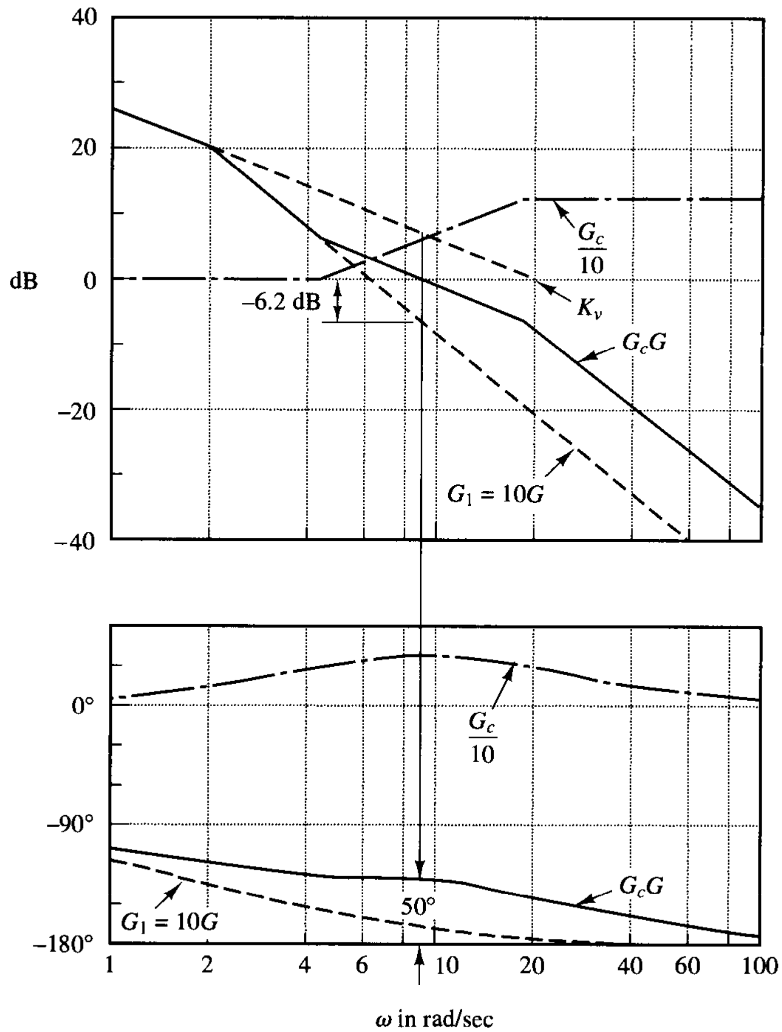


Figure 9–8
Bode diagram for the compensated system.

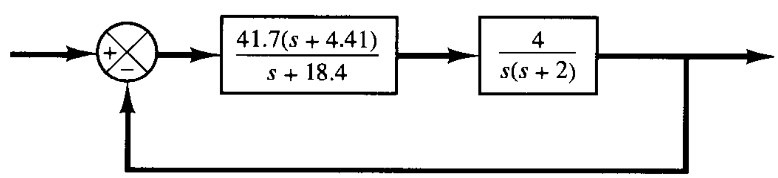


Figure 9–9
Compensated system.

the value of M_r may be obtained easily by transferring the data from the Bode diagram to the Nichol's chart.)

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 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 un-

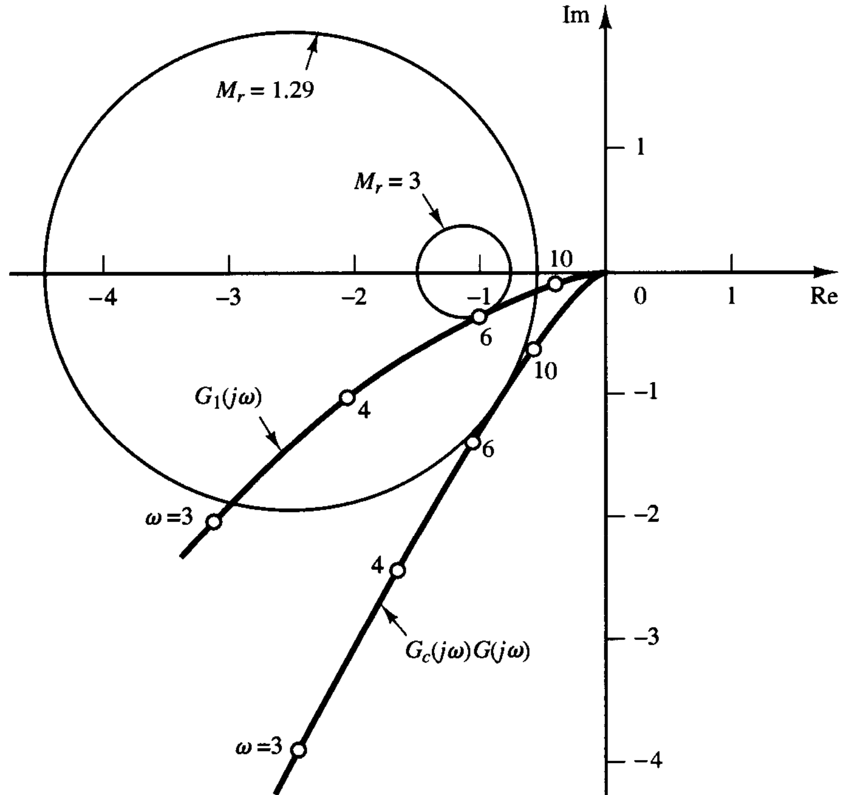


Figure 9–10
Polar plots of the uncompensated and compensated open-loop transfer function. (G_1: uncompensated system; G_cG: compensated system.)

compensated 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 and Figure 9–12 depicts the unit-ramp response curves. These response curves indicate that the designed system is satisfactory.

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(0.35,1.3,'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')
```

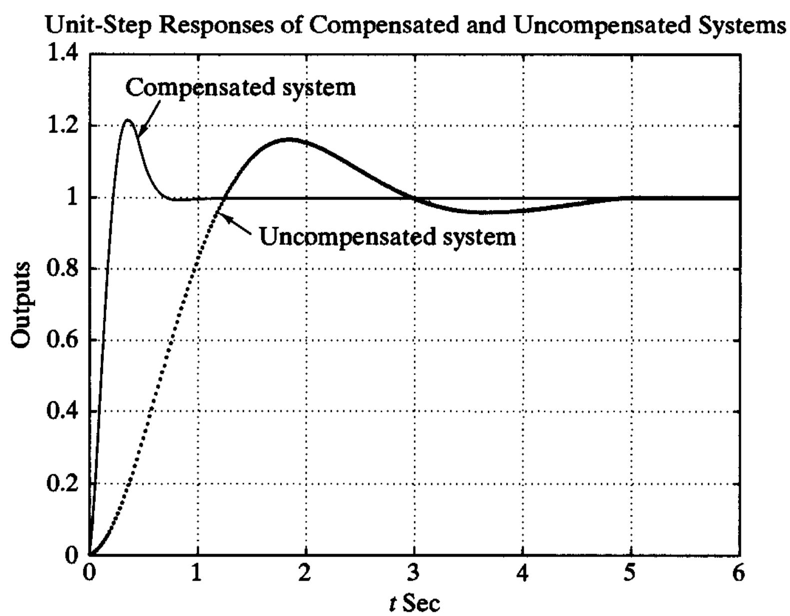


Figure 9–11
Unit-step response curves of the compensated and uncompensated systems.

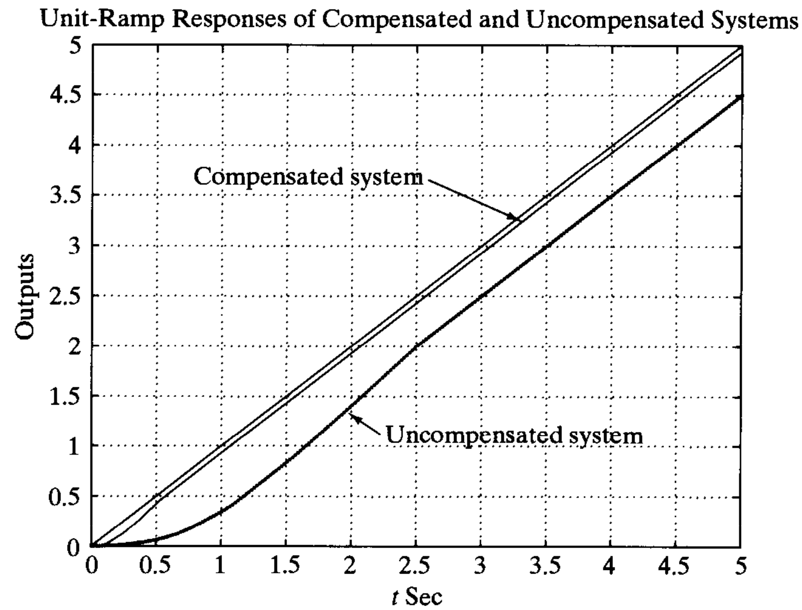


Figure 9–12 Unit-ramp response curves of the compensated and uncompensated systems.

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.

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:

$$G_c(s) = K_c\beta\frac{Ts + 1}{\beta Ts + 1} = K_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}} \qquad (\beta > 1)$$

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.

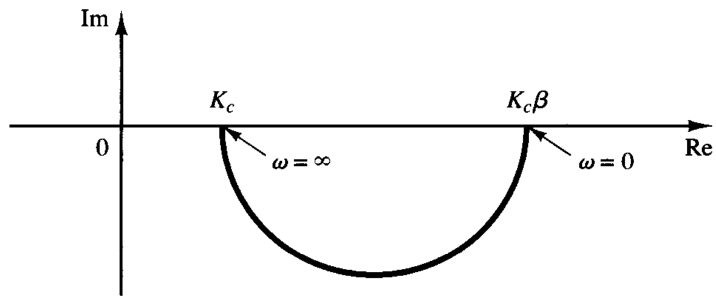


Figure 9–13 Polar plot of a lag compensator $K_c\beta(j\omega T + 1)/(j\omega\beta T + 1)$.

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:

$$G_c(s) = K_c\beta\frac{Ts + 1}{\beta Ts + 1} = K_c\frac{s + \dfrac{1}{T}}{s + \dfrac{1}{\beta T}} \qquad (\beta > 1)$$

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

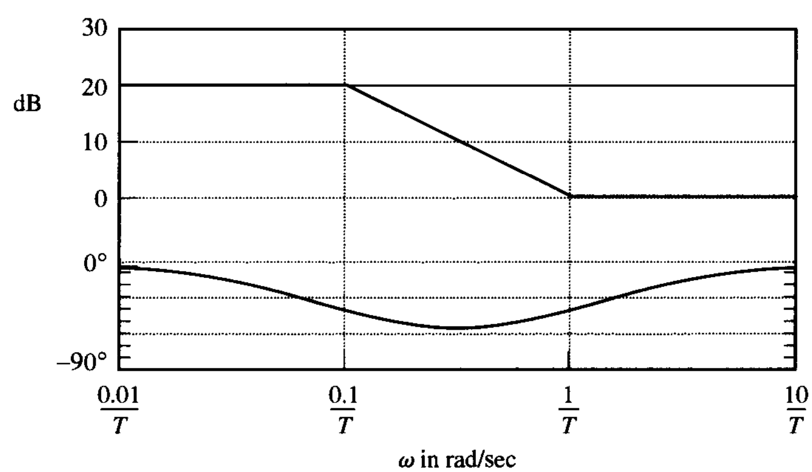


Figure 9–14 Bode diagram of a lag compensator $\beta(j\omega T + 1)/(j\omega\beta T + 1)$, with $\beta = 10$.

$$G_c(s)G(s) = K\frac{Ts+1}{\beta Ts+1}G(s) = \frac{Ts+1}{\beta Ts+1}KG(s) = \frac{Ts+1}{\beta Ts+1}G_1(s)$$

where

$$G_1(s) = KG(s)$$

Determine gain K to satisfy the requirement on the given static error constant.

2. If the 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.)

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 5, calculate constant K_c from

$$K_c = \frac{K}{\beta}$$

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

$$G_c(s) = K_c\beta\frac{Ts+1}{\beta Ts+1} = K_c\frac{s+\frac{1}{T}}{s+\frac{1}{\beta T}} \qquad (\beta > 1)$$

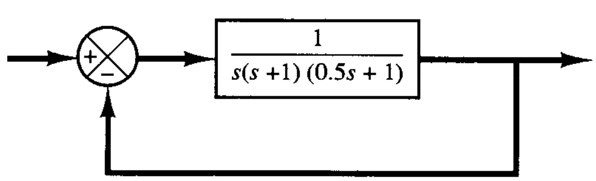


Figure 9–15
Control system.

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

$$G_1(j\omega) = \frac{5}{j\omega(j\omega + 1)(0.5j\omega + 1)}$$

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 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

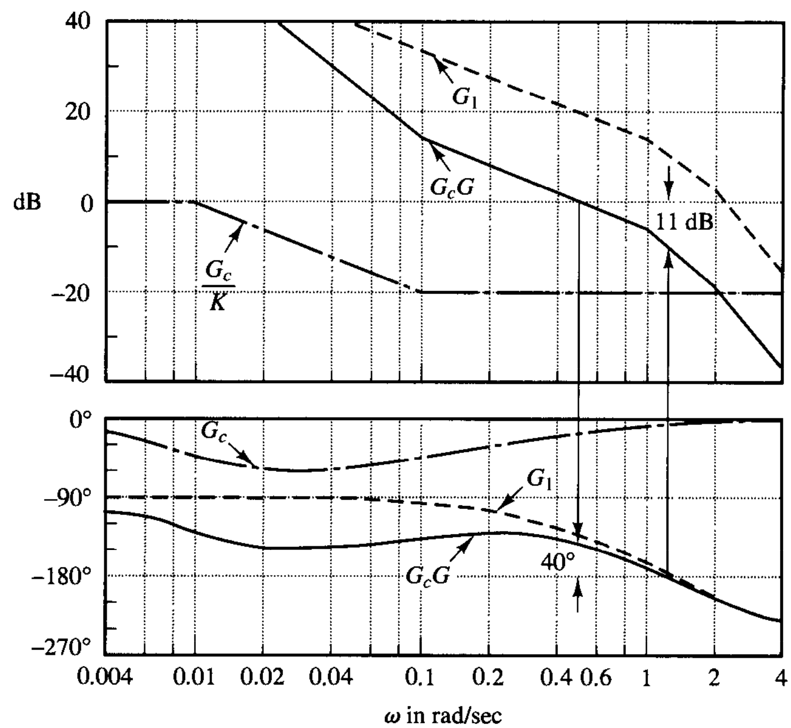


Figure 9–16
Bode diagrams for the uncompensated system, the compensator, and the compensated system. (G_1: uncompensated system, G_c: compensator, G_cG: compensated system.)

$$\frac{1}{\beta T} = 0.01 \text{ rad/sec}$$

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.

Finally, we shall examine the unit-step response and unit-ramp response of the compensated system and the original uncompensated system. 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.

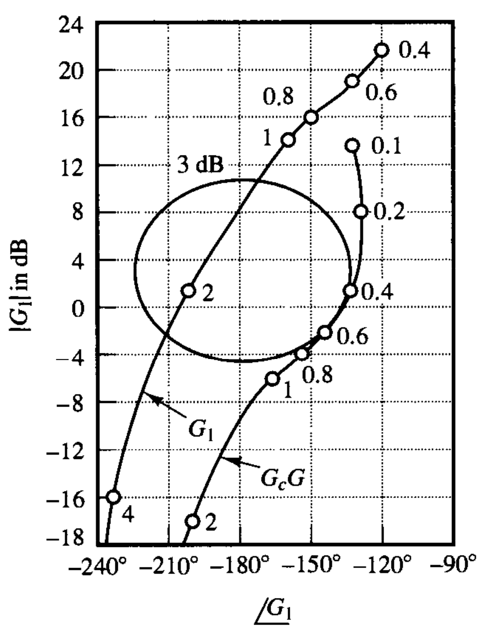


Figure 9–17
Log-magnitude versus phase plots of the uncompensated system and the compensated system. (G_1: uncompensated system, G_cG: compensated system.)

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(12.2,1.27,'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(8.4,3,'Compensated system')
text(8.4,5,'Uncompensated system')
```

Note that the zero and poles of the designed closed-loop systems are as follows:

$$\text{Zero at } s = -0.1$$

$$\text{Poles at } s = -0.2859 \pm j0.5196, \qquad s = -0.1228, \qquad s = -2.3155$$

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.

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)

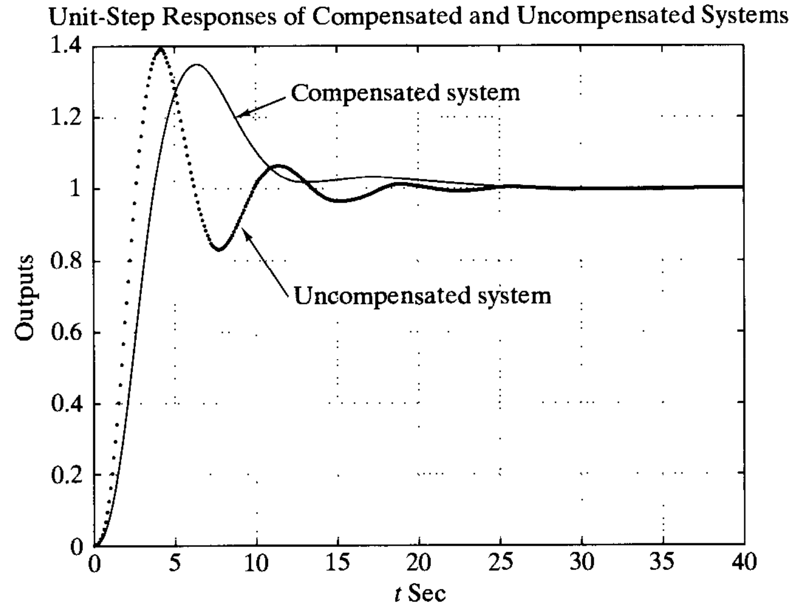


Figure 9–18
Unit-step response curves for the compensated and uncompensated systems (Example 9–2).

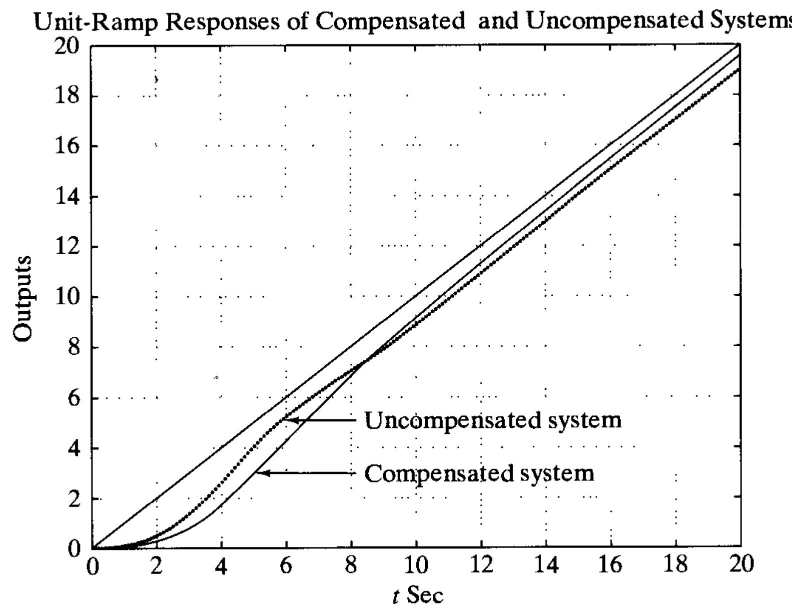


Figure 9–19
Unit-ramp response curves for the compensated and uncompensated systems (Example 9–2).

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

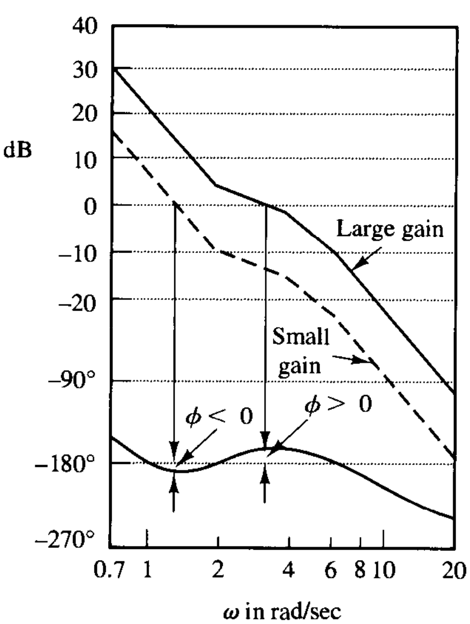


Figure 9–20
Bode diagram of a conditionally stable system.

the amplitude of the input to the saturating element is small. (This can be done by means of minor feedback-loop compensation.)

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.

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

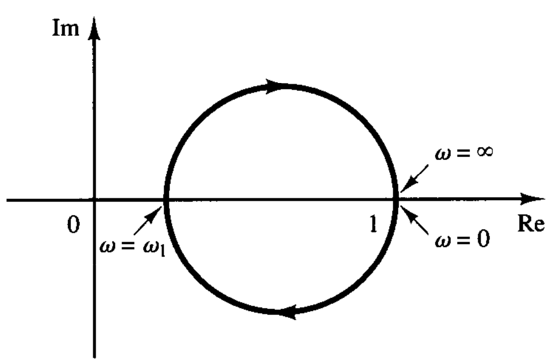


Figure 9–21
Polar plot of a lag–lead compensator given by Equation (9–3), with $K_c = 1$ and $\gamma = \beta$.

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.

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:

$$G_c(s) = K_c \frac{(T_1 s + 1)(T_2 s + 1)}{\left(\frac{T_1}{\beta} s + 1\right)\left(\beta T_2 s + 1\right)} = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)} \qquad (9\text{–}4)$$

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.

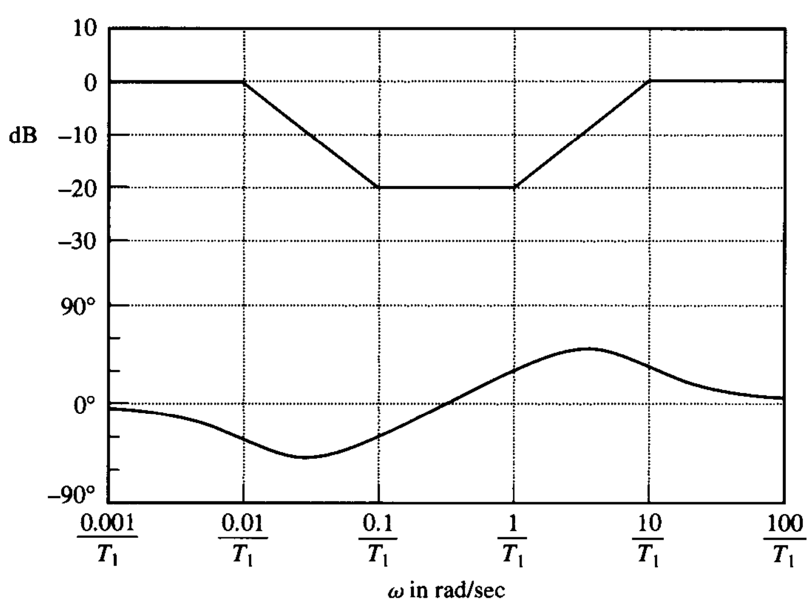


Figure 9–22 Bode diagram of a lag–lead compensator given by Equation (9–3) with $K_c = 1, \gamma = \beta = 10$, and $T_2 = 10T_1$.

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). 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 uncompensated system is found to be $-32°$, which indicates that the 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 $\underline{/G(j\omega)} = -180°$ 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.

Recall that for the lead compensator the maximum phase lead angle ϕ_m is given by Equation (9–1), where α in Equation (9–1) 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 compen-

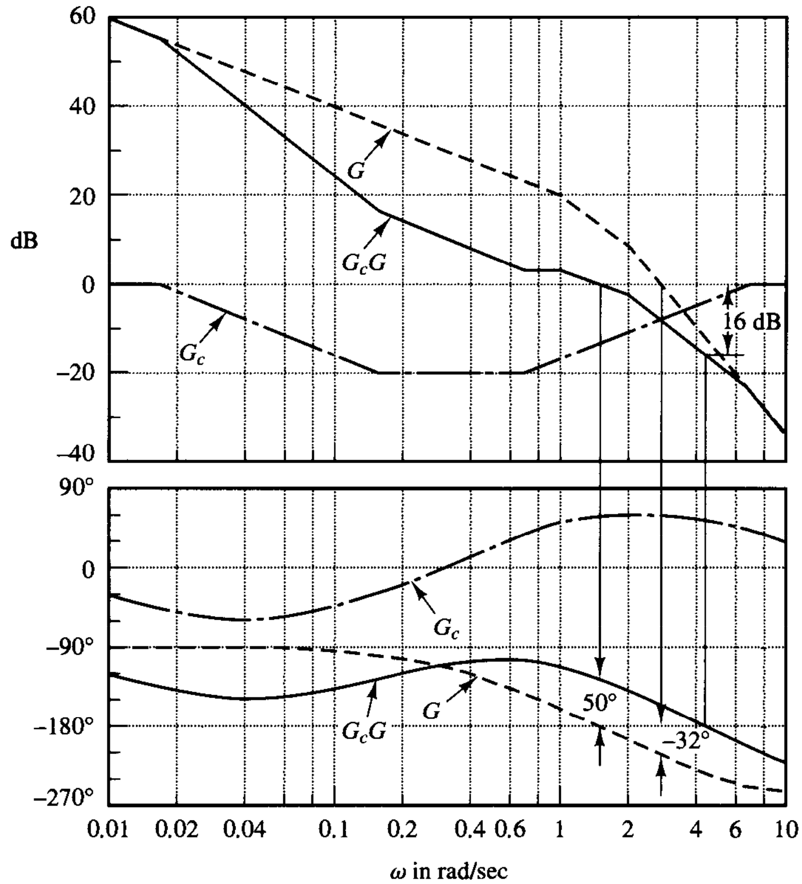


Figure 9–23
Bode diagrams for the uncompensated system, the compensator, and the compensated system. (G: uncompensated system, G_c: compensator, G_cG: compensated system.)

sator 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 (-13 dB, 1.5 rad/sec). 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 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

$$G_c(s)G(s) = \frac{(s + 0.7)(s + 0.15)20}{(s + 7)(s + 0.015)s(s + 1)(s + 2)}$$

$$= \frac{10(1.43s + 1)(6.67s + 1)}{s(0.143s + 1)(66.7s + 1)(s + 1)(0.5s + 1)} \tag{9–5}$$

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 the uncompensated system and compensated system. 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.

In the following we shall examine the transient-response characteristics of the compensated system. (The 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:

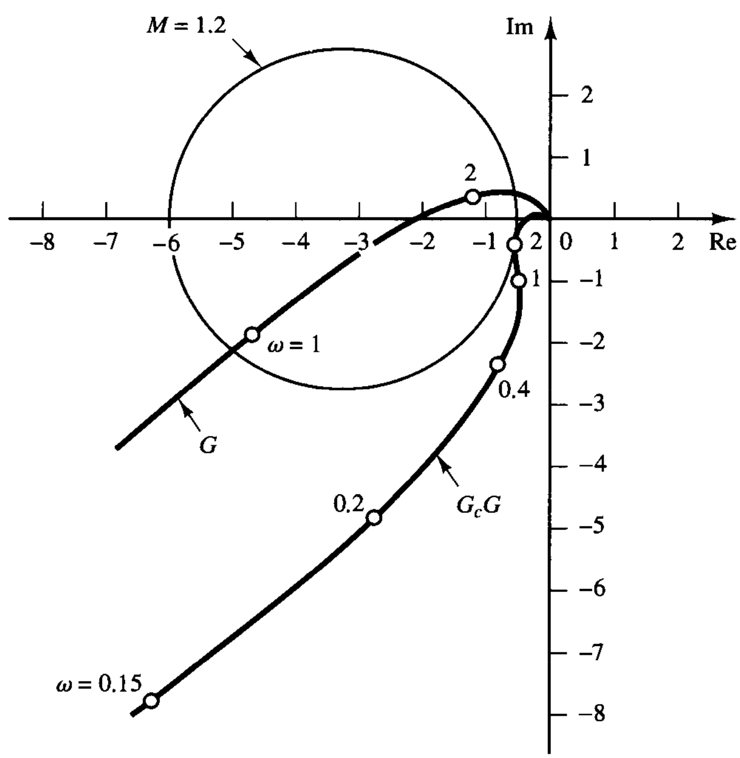


Figure 9–24
Polar plots of the uncompensated system and the compensated system. (G: uncompensated system, G_cG: compensated system.)

$$\text{Zeros at } s = -0.1499, \quad s = -0.6993$$

$$\text{Poles at } s = -0.8973 \pm j1.4439$$

$$s = -0.1785, \quad s = -0.5425, \quad s = -7.4923$$

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 an amplitude to the long tail.

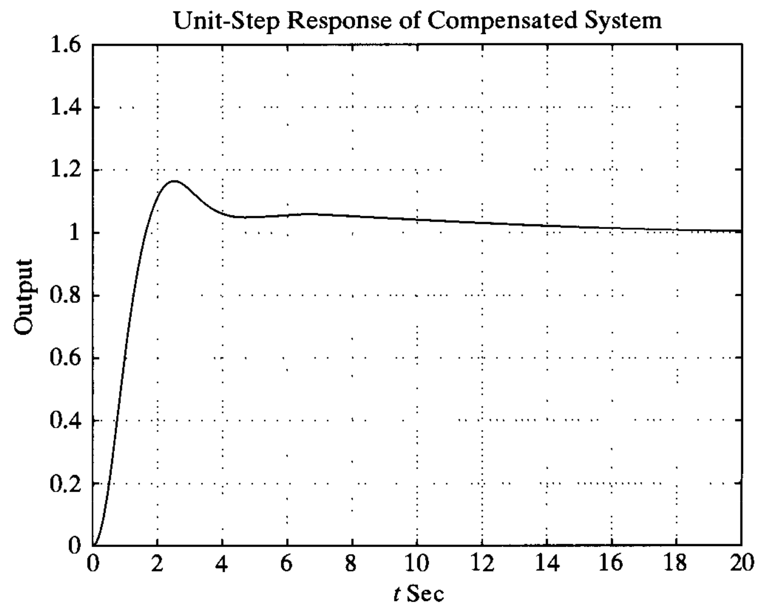


Figure 9–25
Unit-step response of the compensated system (Example 9–3).

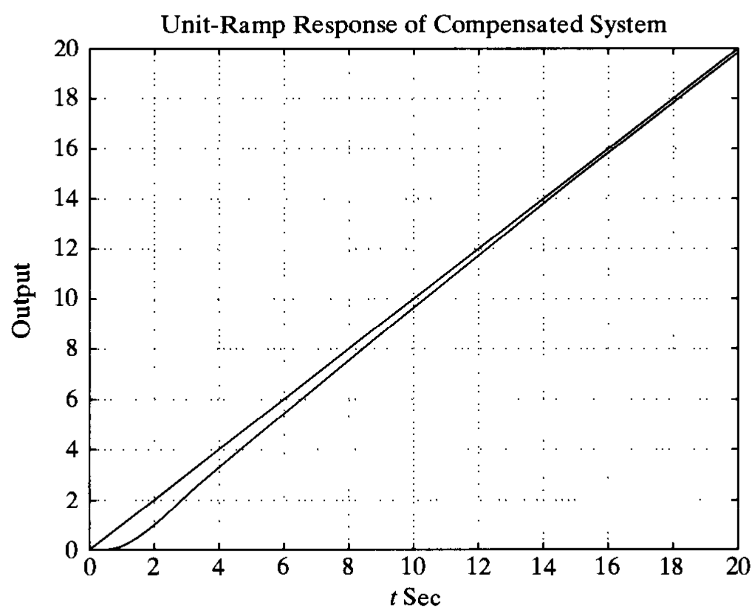


Figure 9–26
Unit-ramp response of the compensated system (Example 9–3).

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. Note that a satisfactory design of a compensator for a complex system may require a creative application of these basic design principles.

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 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 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. 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.

5. 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.

6. 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 network, respectively, are shown in Figures 9–27(b), (c), and (d). The system with a lead

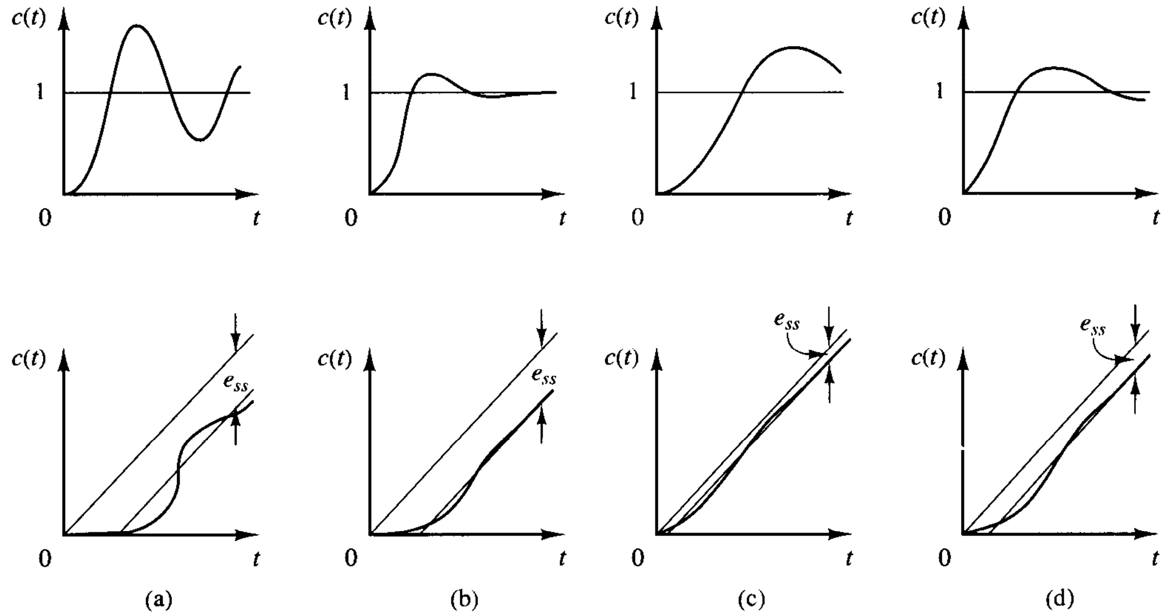


Figure 9–27
Unit-step response curves and unit-ramp response curves. (a) Uncompensated system; (b) lead compensated system; (c) lag compensated system; (d) lag–lead compensated system.

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.

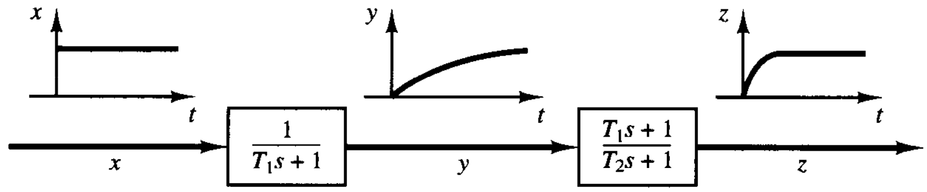
Velocity feedback using a tachometer is very commonly used in positional servo systems. It is noted that, if the system is subjected to noise signals, velocity feedback may generate some difficulty if a particular velocity feedback scheme performs differentiation of the output signal. (The result is the accentuation of the noise effects.)

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.

Figure 9–28
Step-response curves showing the effect of canceling a large time constant.



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 quite 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.

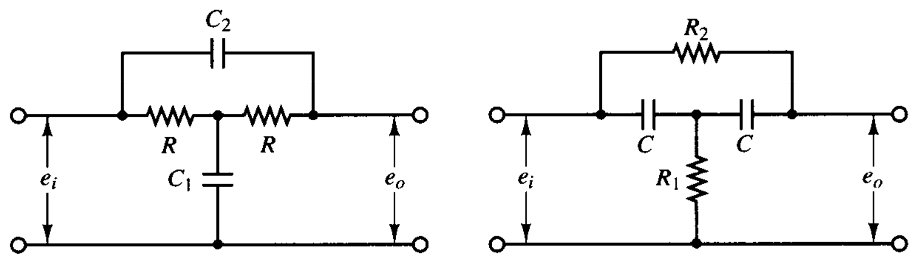
Cancellation of undesirable complex-conjugate poles. If the transfer function of a plant contains one or more pairs of complex-conjugate poles, then a lead, lag, or lag–lead compensator may not give satisfactory results. In such a case, a network that has two zeros and two poles may prove to be useful. 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

$$\frac{1}{s^2 + 2\zeta_1\omega_1 s + \omega_1^2}$$

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.)



$$\frac{E_o(s)}{E_i(s)} = \frac{RC_1RC_2s^2 + 2RC_2s + 1}{RC_1RC_2s^2 + (RC_1 + 2RC_2)s + 1}$$

(a)

$$\frac{E_o(s)}{E_i(s)} = \frac{R_1CR_2Cs^2 + 2R_1Cs + 1}{R_1CR_2Cs^2 + (R_2C + 2R_1C)s + 1}$$

(b)

Figure 9–29
Bridged-*T* networks.

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.

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 acts as proportional-plus-derivative control (in the region of small ω) and proportional-plus-integral control (in the region of large ω), respectively.

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).

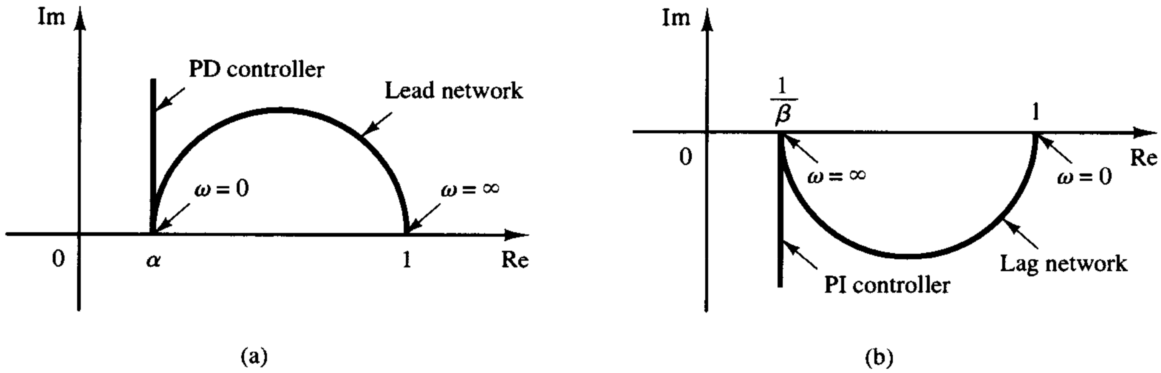


Figure 9–30
(a) Polar plots of a lead network and a proportional-plus-derivative controller; (b) polar plots of a lag network and a proportional-plus-integral controller.

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

$$\angle G_c(j\omega) = \angle j\omega + \frac{1}{T_1} + \angle j\omega + \frac{1}{T_2} - \angle j\omega + \frac{\beta}{T_1} - \angle j\omega + \frac{1}{\beta T_2}$$

$$= \tan^{-1}\omega T_1 + \tan^{-1}\omega T_2 - \tan^{-1}\omega T_1/\beta - \tan^{-1}\omega T_2\beta$$

At $\omega = \omega_1 = 1/\sqrt{T_1 T_2}$, we have

$$\angle G_c(j\omega_1) = \tan^{-1}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\sqrt{\frac{T_2}{T_1}} - \tan^{-1}\frac{1}{\beta}\sqrt{\frac{T_1}{T_2}} - \tan^{-1}\beta\sqrt{\frac{T_2}{T_1}}$$

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

$$\tan^{-1}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\sqrt{\frac{T_2}{T_1}} = 90°$$

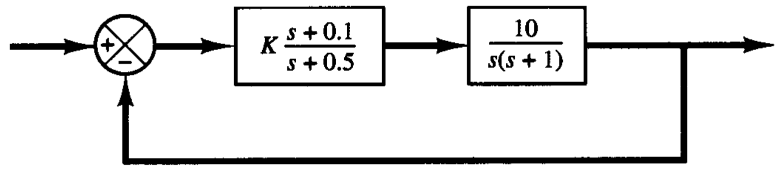


Figure 9–31
Control system.

and also

$$\tan^{-1}\frac{1}{\beta}\sqrt{\frac{T_1}{T_2}} + \tan^{-1}\beta\sqrt{\frac{T_2}{T_1}} = 90°$$

we have

$$\underline{/G_c(j\omega_1)} = 0°$$

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°.

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:

$$20 \log K = -14.5 \text{ dB}$$

MATLAB Program 9–3

```
num = [0   0   10   1];
den = [1   1.5   0.5   0];
bode(num,den)
subplot(2,1,1);
title('Bode Diagram of G(s) = (10s + 1)/[s(s + 0.5)(s + 1)]')
```

or

$$K = 0.188$$

Thus, we have determined the value of gain K.

To verify the results, let us draw a Nyquist plot of G for the frequency range

$$w = 0.1{:}0.01{:}1.15$$

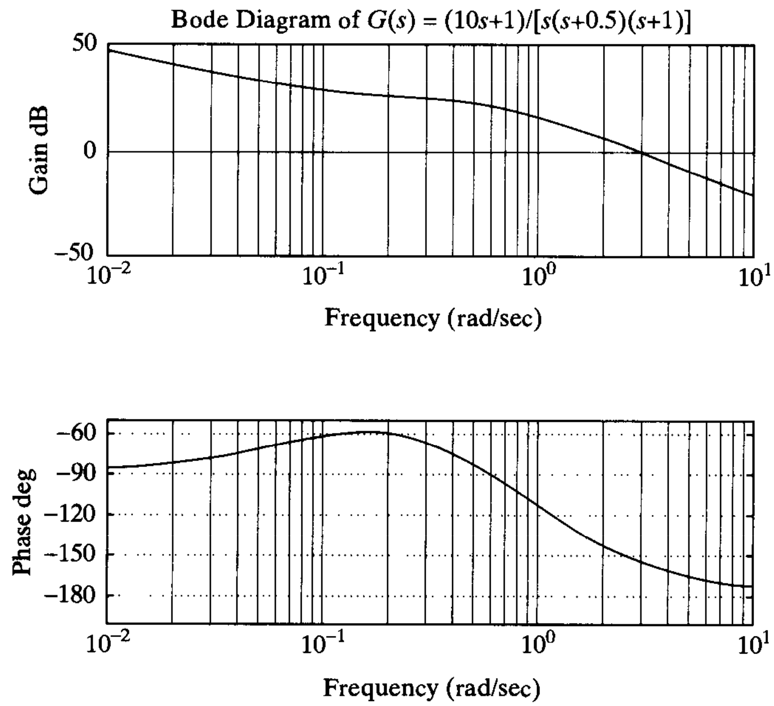


Figure 9–32
Bode diagram of $G(s) = \dfrac{10s + 1}{s(s + 0.5)(s + 1)}$.

The end point of the locus ($w = 1.15$ 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

$$z = \mathrm{re} + \mathrm{i*im} = \mathrm{re}^{i\theta}$$

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.

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.

MATLAB Program 9–4

```
%*****Nyquist plot in polar 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)
title('Check of Phase Margin')
text(0.1,-1.5,'Nyquist plot')
text(-2.25,-0.3,'Phase margin')
text(-2.25,-0.7,'is 60 degrees.')
```

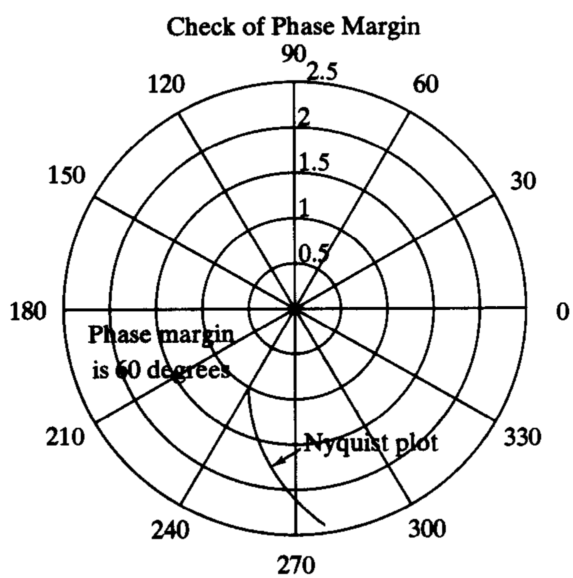


Figure 9–33
Nyquist plot of $G(j\omega)$ showing that the phase margin is 60°.

(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.

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.

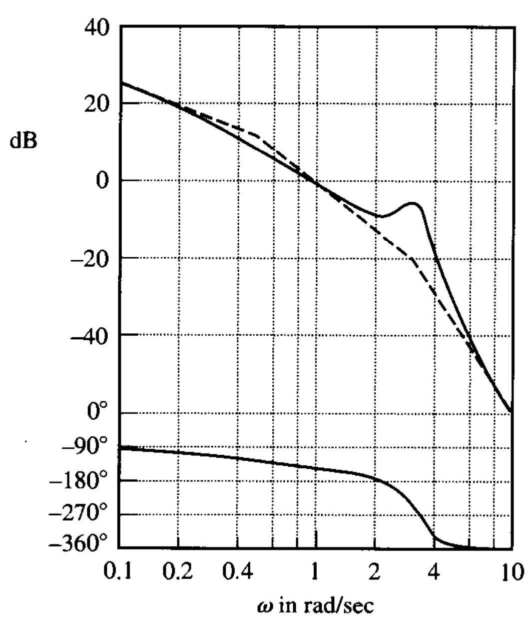


Figure 9–34 Bode diagram of $G(j\omega)$ given by Equation (9–6).

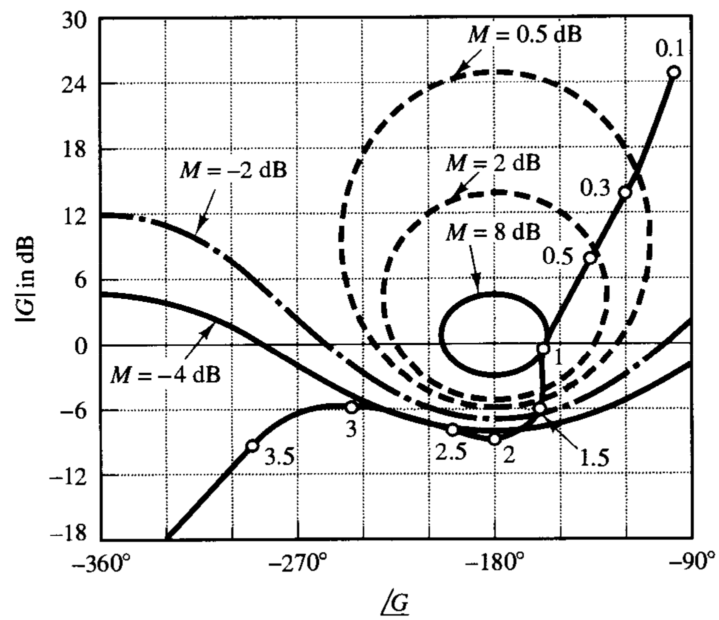


Figure 9–35
Log-magnitude versus phase plot of $G(j\omega)$ given by Equation (9–6).

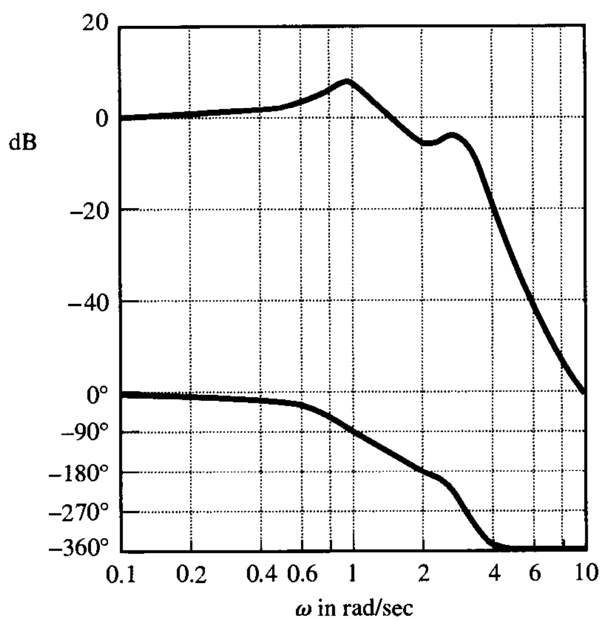


Figure 9–36
Bode diagram of $G(j\omega)/[1 + G(j\omega)]$, where $G(j\omega)$ is given by Equation (9–6).

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

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

$$= \frac{9}{(s^2 + 0.487s + 1)(s^2 + 0.613s + 9)}$$

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:

$$G_c(s) = K_c \frac{Ts + 1}{\alpha Ts + 1} \quad (\alpha < 1)$$

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

$$\sin 45° = \frac{1 - \alpha}{1 + \alpha}$$

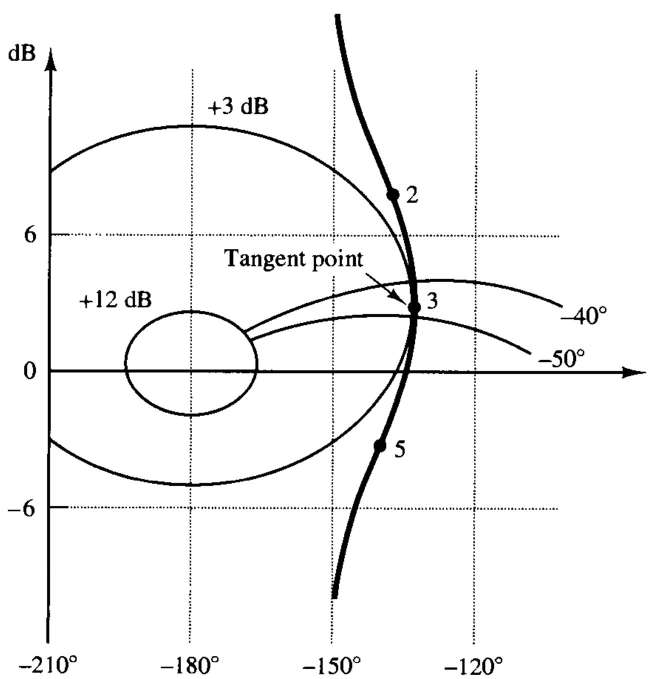


Figure 9–37
Nichols chart showing that the $G_c(j\omega)G(j\omega)$ locus is tangent to the $M = 3$-dB locus at $\omega = 3$ rad/sec.

or $\alpha = 0.172 \doteqdot \frac{1}{6}$. 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 ω = 3 rad/sec, we obtain

$$20 \log |G_c(j\omega)G(j\omega)|_{\omega=3} = 20 \log |G_c(j3)| + 20 \log |G(j3)|$$
$$= 20 \log |G_c(j3)| + 20 \log \left|\frac{1}{9}\right| = 3 \text{ dB}$$

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?

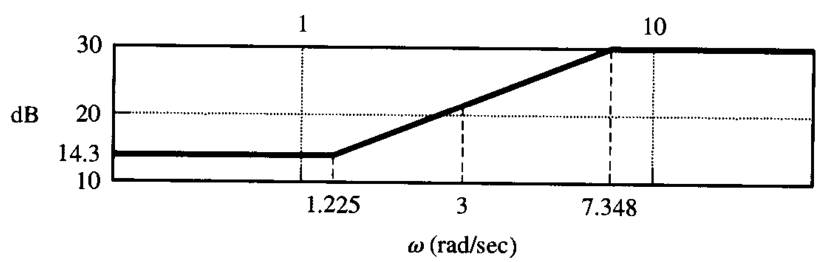


Figure 9–38 Bode diagram of the lead compensator designed in Problem A-9-5.

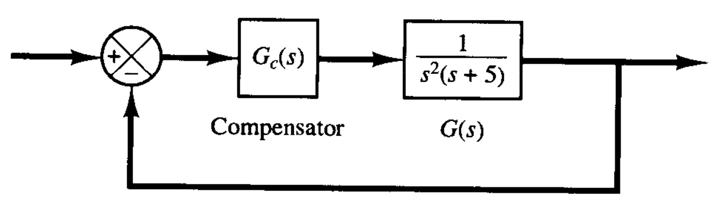


Figure 9–39 Closed-loop system.

Solution. Notice that

$$G_c(j\omega)G(j\omega) = K_c\alpha\left(\frac{Tj\omega + 1}{\alpha Tj\omega + 1}\right)\frac{0.2}{(j\omega)^2(0.2j\omega + 1)}$$

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.

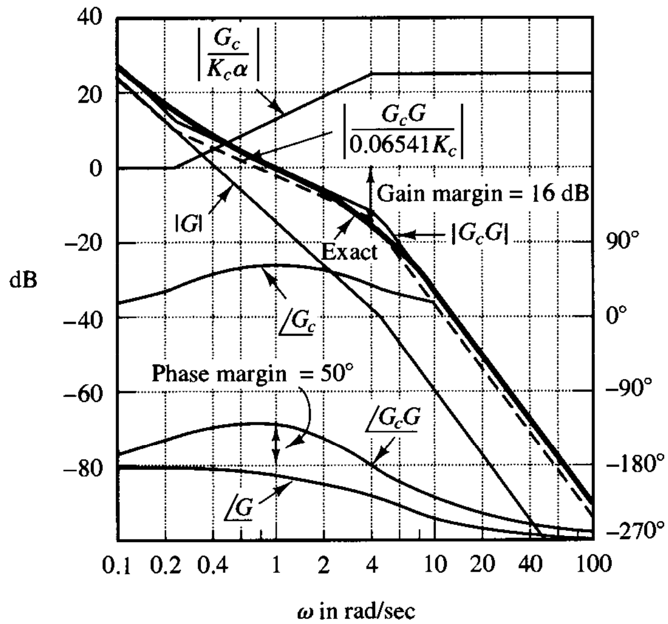


Figure 9–40
Bode diagram of the system shown in Figure 9–39.

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:

$$M_r = 2.13 \text{ dB}, \qquad \omega_r = 0.58 \text{ rad/sec}$$

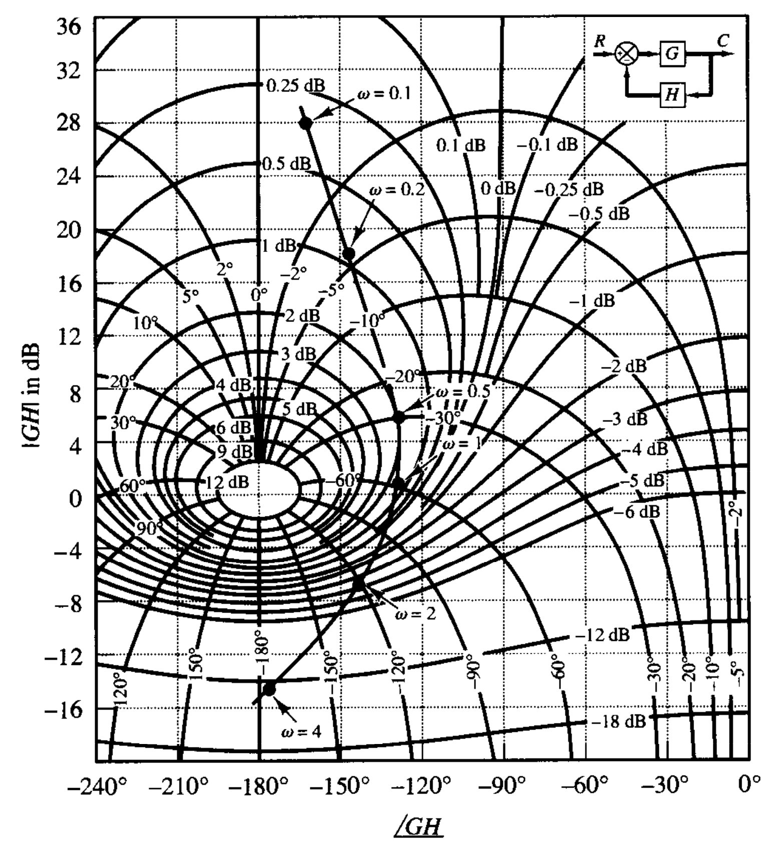


Figure 9–41
$G_c(j\omega)G(j\omega)$ locus superimposed on Nichols chart (Problem A-9-6).

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 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 on
[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')
title('Nyquist Plots of G(jw), G1(jw), and Gc(jw)G(jw)')
text(0.7,-8.8,'G(jw)')
text(-11.7,-8.7,'G1(jw)')
text(-6.6,-12.3,'Gc(jw)G(jw)')
```

Nyquest Plots of $G(jw)$, $G1(jw)$, and $Gc(jw)G(jw)$

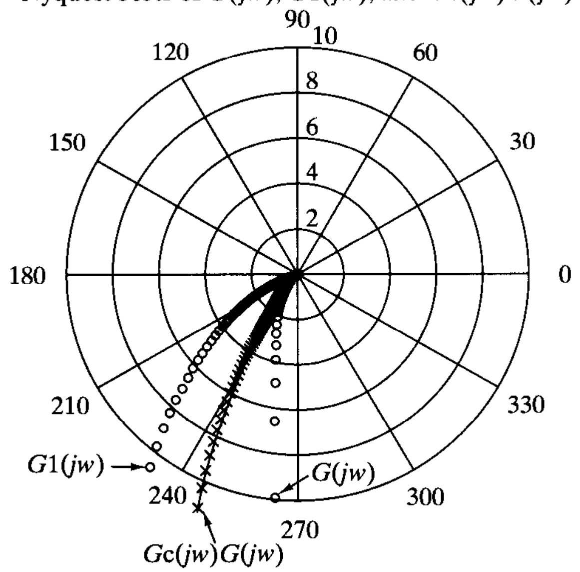


Figure 9–42
Nyquist plots of $G(j\omega)$, $G_1(j\omega)$, and $G_c(j\omega)G(j\omega)$.

A-9-8. Consider the system shown in Figure 9–43(a). Design a compensator such that the closed-loop system will satisfy the following requirements:

$$\text{Static velocity error constant} = 20 \text{ sec}^{-1}$$

$$\text{Phase margin} = 50^\circ$$

$$\text{Gain margin} \geqq 10 \text{ dB}$$

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).

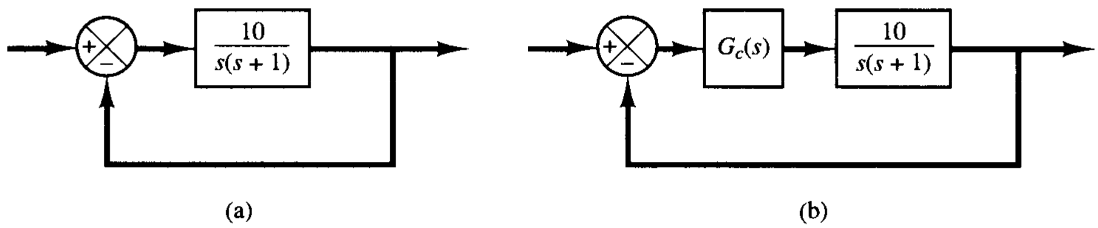


Figure 9–43
(a) Control system;
(b) compensated system.

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

$$\begin{aligned}
K_v &= \lim_{s\to 0} sG_c(s)G(s) \\
&= \lim_{s\to 0} s\frac{Ts+1}{\alpha Ts+1}G_1(s) \\
&= \lim_{s\to 0} \frac{s10K}{s(s+1)} \\
&= 10K = 20
\end{aligned}$$

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)
subplot(2,1,1);
title('Bode Diagram of G1(s) = 20/[s(s + 1)]')
```

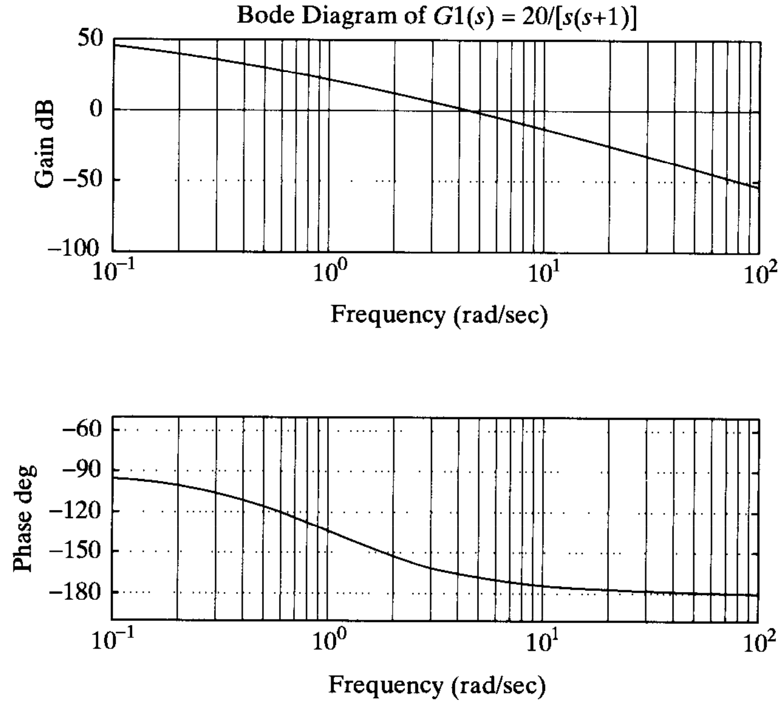


Figure 9–44 Bode diagram of $G_1(s)$.

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

$$\frac{1}{\sqrt{a}} = \frac{1}{\sqrt{0.21}} = 6.7778 \text{ dB}$$

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

MATLAB Program 9–7

```
num = [0  0  20];
den = [1  1  0];
w = logspace(0,1,100);
bode(num,den,w)
subplot(2,1,1);
title('Bode Diagram of G1(s) = 20/[s(s + 1)]')
```

Figure 9–45. From this diagram, we find the frequency point where $|G_1(j\omega)| = -6.7778$ dB 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 $1/(\sqrt{a}T)$, or

$$\omega_c = \frac{1}{\sqrt{a}T}$$

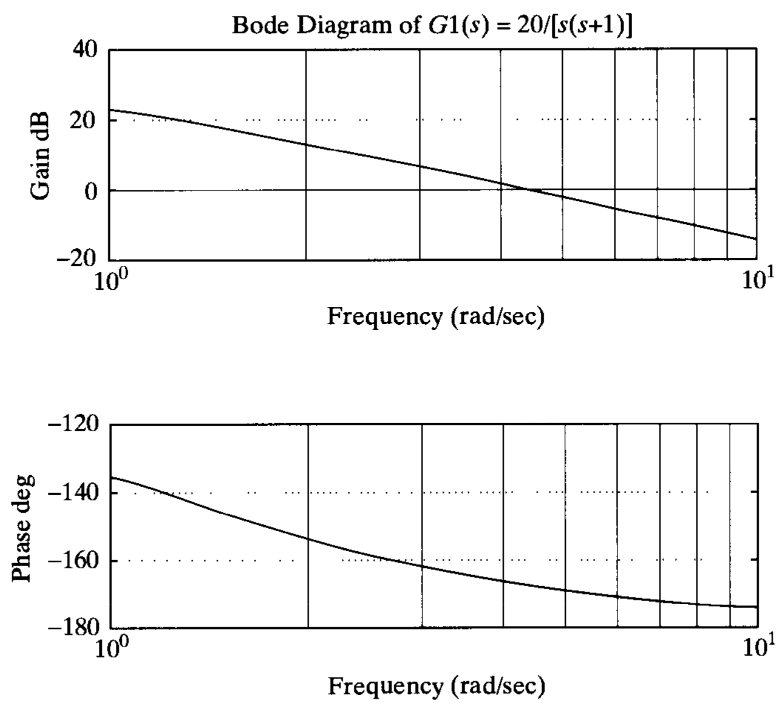


Figure 9–45 Bode diagram of $G_1(s)$.

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.

MATLAB Program 9–8

```
numc = [9.5238   28.3685];
denc = [1   14.1842];
w = logspace(-1,3,100);
bode(numc,denc,w)
subplot(2,1,1);
title('Bode Diagram of Gc(s) = 9.5238(s + 2.9787)/(s + 14.1842)')
```

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.

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.

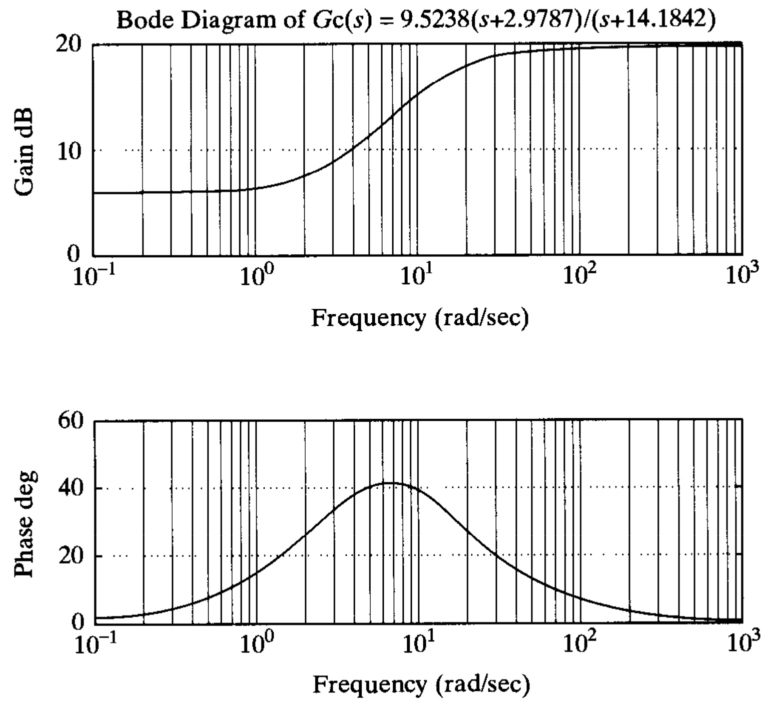


Figure 9–46
Bode diagram of $G_c(s)$.

Unit-step response: We shall compare the unit-step response of the compensated system and that of the original uncompensated system.

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)
subplot(2,1,1);
title('Bode Diagram of Gc(s)G(s)')
```

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:

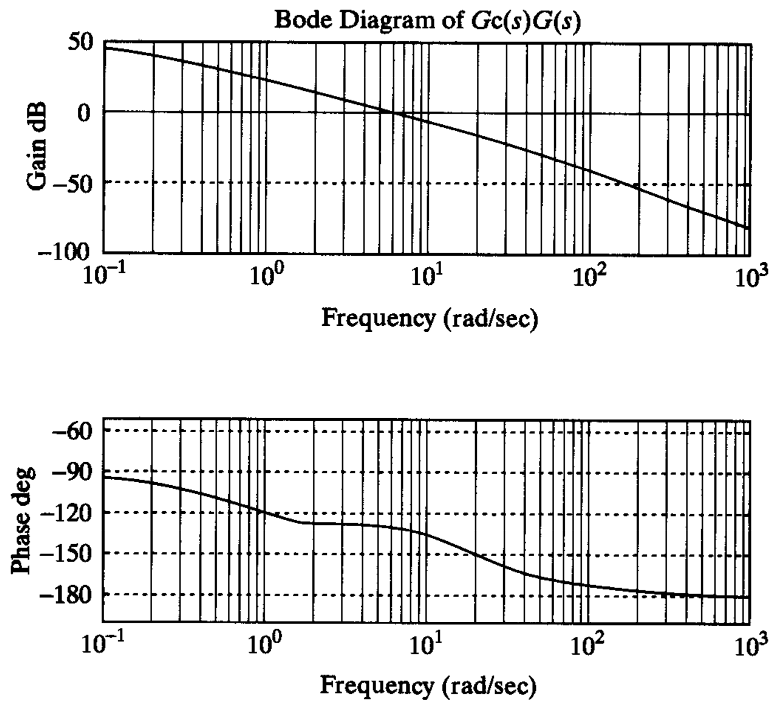


Figure 9–47 Bode diagram of $G_c(s)G(s)$.

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;
[cl,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')
```

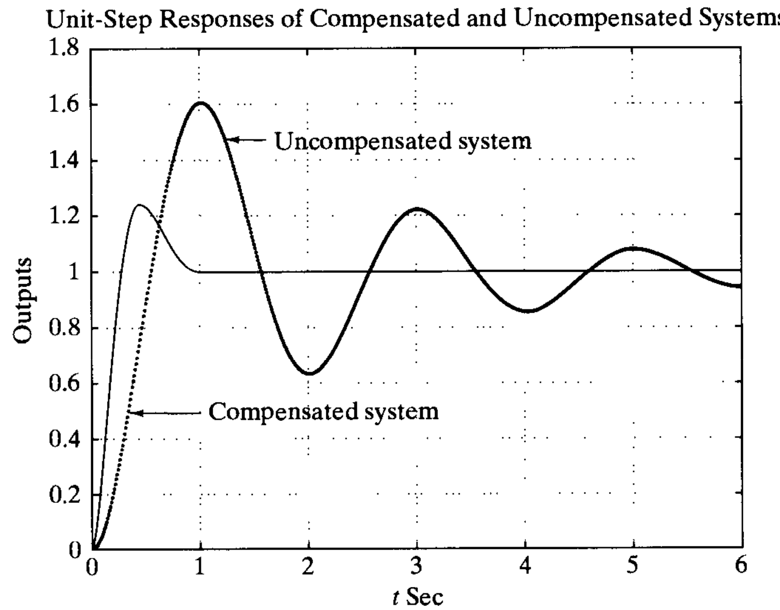


Figure 9–48
Unit-step responses of the compensated and uncompensated systems.

Zero at $s = -2.9787$

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 $K_v = 20\ \text{sec}^{-1}$, 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 $10\ \text{sec}^{-1}$. Hence, the original uncompensated system will have twice as large a steady-state error in following the unit-ramp input.

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.

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 compensator $G_c(s)$ such that the static velocity error constant is $10\ \text{sec}^{-1}$, the phase margin is 50°, and the gain margin is 10 dB or more.

Solution. We shall design a lag–lead compensator of the form

$$G_c(s) = K_c \frac{\left(s + \frac{1}{T_1}\right)\left(s + \frac{1}{T_2}\right)}{\left(s + \frac{\beta}{T_1}\right)\left(s + \frac{1}{\beta T_2}\right)}$$

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(0.07,1.3,'Compensated system')
text(1.2,0.65,'Uncompensated system')
```

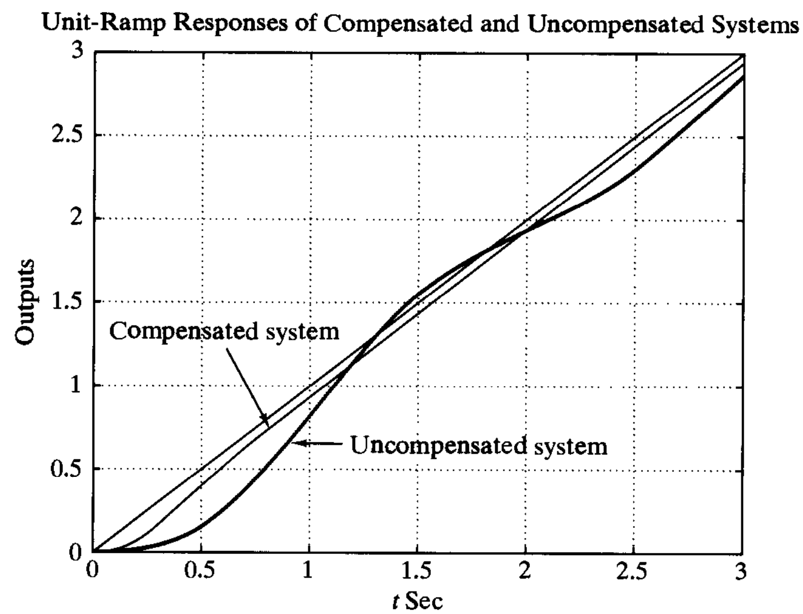


Figure 9–49 Unit-ramp responses of the compensated and uncompensated systems.

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)
subpole(2,1,1);
title('Bode Diagram of G(s) = 40/[s(s + 1)(s + 4)]')
```

From Figure 9–50, the phase margin of the uncompensated system is found to be $-16°$, which indicates that the 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 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.

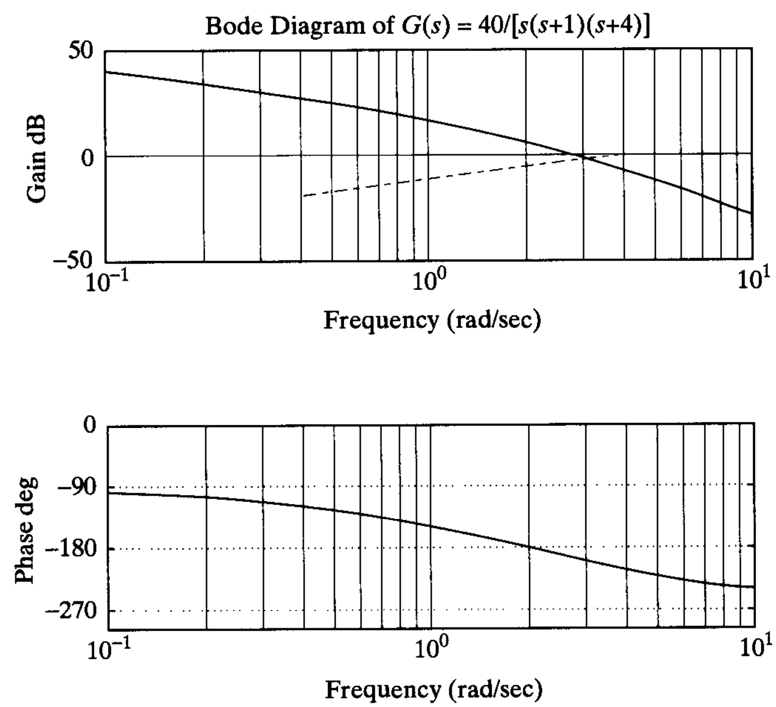


Figure 9–50 Bode diagram of $G(s) = 40/[s(s + 1)(s + 4)]$.

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 $\omega = 1$ 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 (-6 dB, 2 rad/sec). (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.

MATLAB Program 9–13

```
numc = [1   0.6   0.08];
denc = [1   4.02   0.08];
bode(numc,denc)
subplot(2,1,1);
title('Bode Diagram of Lag–Lead Compensator')
```

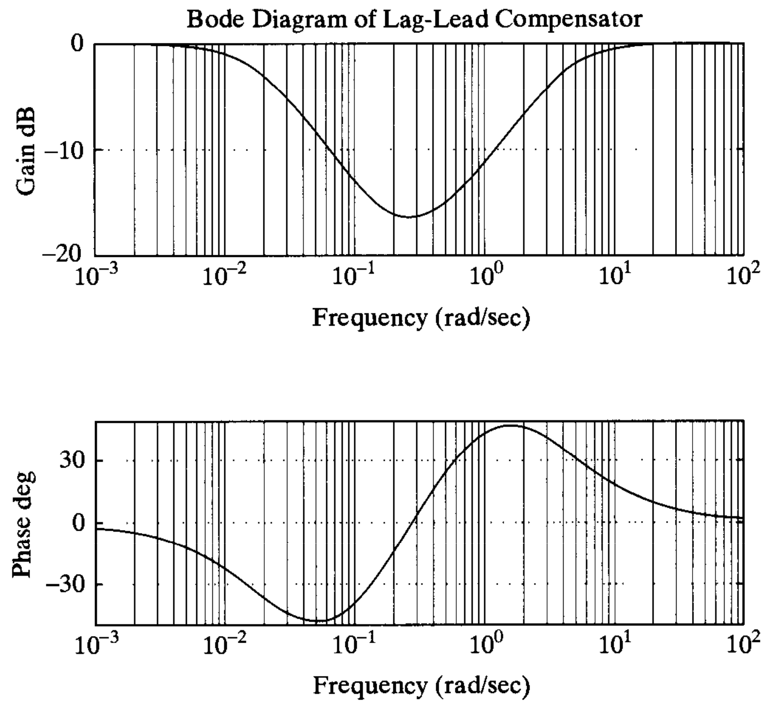


Figure 9–51
Bode diagram of the designed lag–lead compensator.

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 den 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
```

From Figure 9–52, we see that the actual gain crossover frequency is slightly shifted from 2 rad/sec to a lower value. The actual gain crossover frequency can be found by plotting the Bode diagram in the region $1 \leq \omega \leq 10$. It is found to be $\omega = 1.86$ rad/sec. [Such a small shift of the gain crossover frequency from the assumed gain crossover frequency (2 rad/sec in this case) always occurs in the present method of design.]

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)
subplot(2,1,1);
title('Bode Diagram of Gc(s)G(s)')
```

Since the phase margin of the compensated system is 50°, the gain margin is 12.5 dB, and the static velocity error constant is 10 $\sec^{-1}$, all the requirements are met.

Figure 9–53 shows the Nyquist plots of $G(j\omega)$ (uncompensated case) and $G_c(j\omega)G(j\omega)$ (compensated case). MATLAB Program 9–15 was used to obtain Figure 9–53.

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

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

$$= \frac{40(s+0.4)(s+0.2)}{(s+4)(s+0.02)s(s+1)(s+4)+40(s+0.4)(s+0.2)}$$

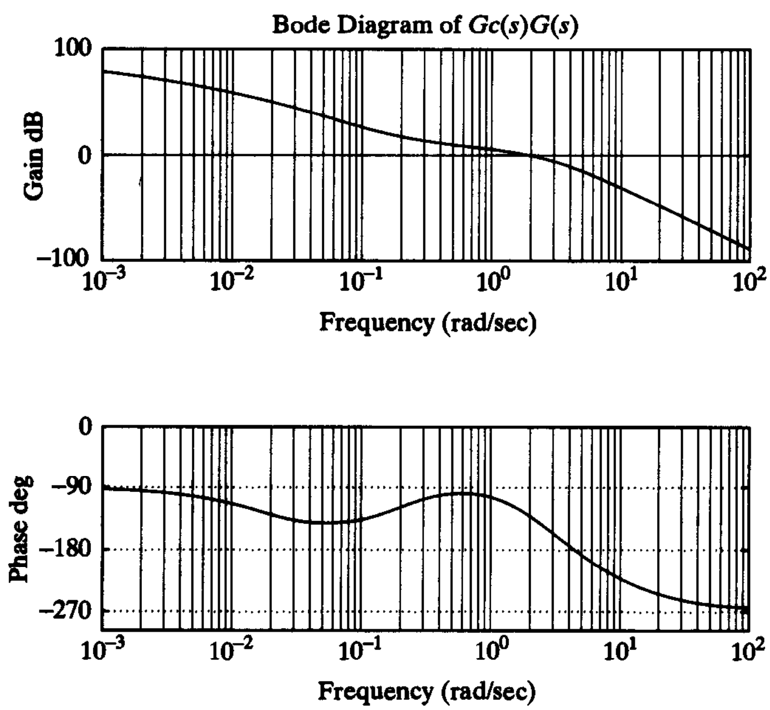


Figure 9–52
Bode diagram of the open-loop transfer function $G_c(s)G(s)$ of the compensated system.

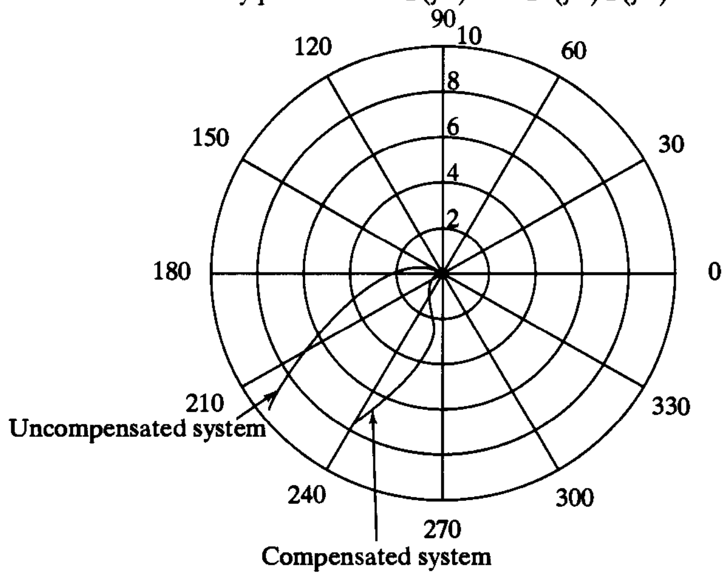


Figure 9–53
Nyquist plots of $G(j\omega)$ and $G_c(j\omega)G(j\omega)$.

MATLAB Program 9–15

```
%*****Nyquist plots*****

num0 = [0   0   0   40];
den0 = [1   5   4   0];
num1 = [0   0   0   40   24   3.2];
den1 = [1   9.02   24.18   16.48   0.32   0];
w = 0.8:0.02:10;
ww = 0.2:0.02:10;
[re0,im0,w] = nyquist(num0,den0,w);
z0 = re0 + i*im0;
r0 = abs(z0);
theta0 = angle(z0);
polar(theta0,r0)
hold on
[re1,im1,ww] = nyquist(num1,den1,ww);
z1 = re1 + i*im1;
r1 = abs(z1);
theta1 = angle(z1);
polar(theta1,r1)
text(-8, -12.7,'Compensated system')
text(-18.8, -7,'Uncompensated system')
title('Nyquist Plots of G(jw) and Gc(jw)G(jw)')
```

To determine the denominator polynomial with MATLAB, we may proceed as follows: Define

$$a(s) = (s + 4)(s + 0.02) = s^2 + 4.02s + 0.08$$

$$b(s) = s(s + 1)(s + 4) = s^3 + 5s^2 + 4s$$

$$c(s) = 40(s + 0.4)(s + 0.2) = 40s^2 + 24s + 3.2$$

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 9–16 is used to obtain the unit-step response of the compensated system. The resulting unit-step response curve is shown in Figure 9–54. (Note that the uncompensated system is unstable.)

```
MATLAB Program 9–16

%*****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')
```

Unit-ramp response: The unit-ramp response of this system may be obtained by entering MATLAB Program 9–17 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 9–55.

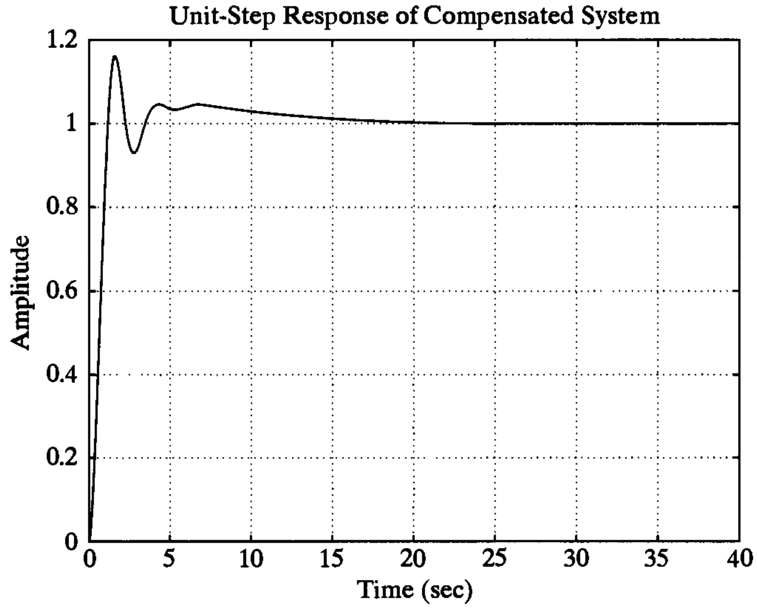


Figure 9–54
Unit-step response curve of the compensated system.

MATLAB Program 9–17

```
%*****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('Amplitude')
```

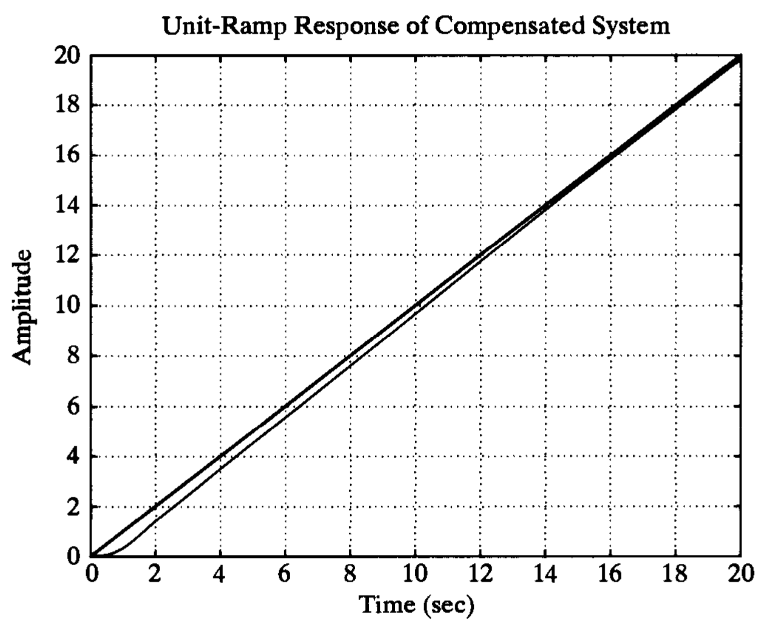


Figure 9–55
Unit-ramp response of the compensated system.

B-9-1. Draw Bode diagrams of the lead network and lag network shown in Figures 9–56 (a) and (b), respectively.

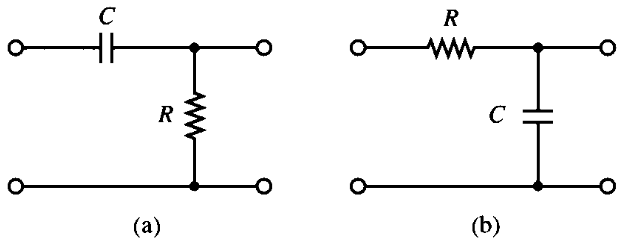


Figure 9–56 (a) Lead network; (b) lag network.

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 9–57 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.

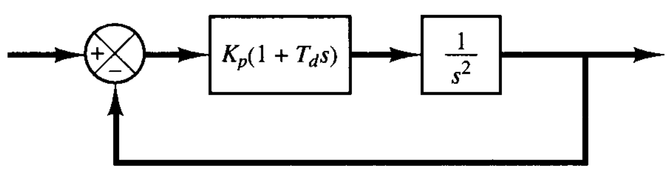


Figure 9–57 Block diagram of space vehicle attitude control system.

B-9-5. Referring to the closed-loop system shown in Figure 9–58, 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.

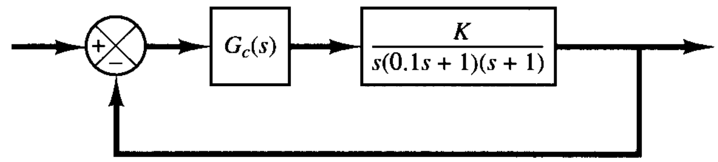


Figure 9–58 Closed-loop system

B-9-6. Consider the system shown in Figure 9–59. 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.

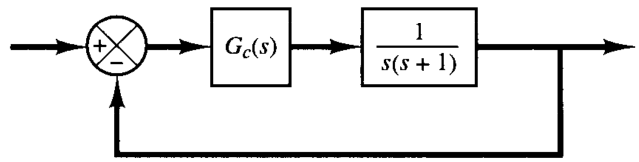


Figure 9–59 Control system.

B-9-7. Consider the system shown in Figure 9–60. 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.

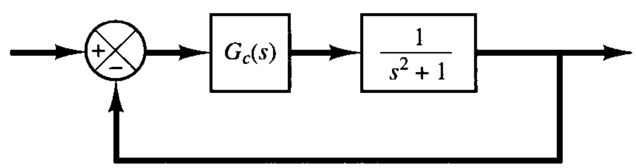


Figure 9–60 Control system.

B-9-8. Consider the system shown in Figure 9–61. It is desired to design a compensator such that the static velocity

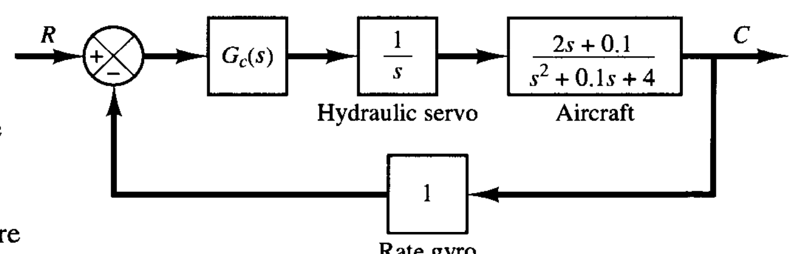


Figure 9–61 Control system.

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.

B-9-9. Consider the system shown in Figure 9–62. 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.

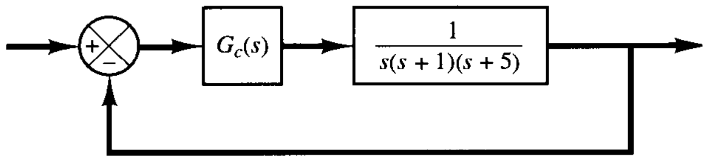


Figure 9–62 Control system.

10

PID Controls and Introduction to Robust Control

10–1 INTRODUCTION

In previous chapters we occasionally discussed the basic PID control schemes. For example, in Chapter 5 we presented hydraulic, pneumatic, and electronic PID controllers. In Chapters 7 and 9 we designed control systems where PID controllers were involved.

In this chapter we first present tuning rules for the basic PID controllers and then discuss modified forms of PID control schemes, including PI-D control, I-PD control, and two-degrees-of-freedom PID control. Finally, we introduce the concept of robust design.

It is interesting to note that more than half of the industrial controllers in use today utilize PID or modified PID control schemes. Analog PID controllers are mostly hydraulic, pneumatic, electric, and electronic types or their combinations. Currently, many of these are transformed into digital forms through the use of microprocessors.

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 the field of process control systems, it is a well-known fact that the basic and modified PID control schemes have proved their usefulness in providing satisfactory control, although they may not provide optimal control in many given situations.

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 modified PID control schemes, such as PI-D control and I-PD control. Section 10–4 introduces two-degrees-of-freedom PID control schemes. Section 10–5 introduces the concept of robust control using a two-degrees-of-freedom control system as an example.

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 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 the proportional control action is used. Ziegler–Nichols rules, which are presented in the following, are very convenient when mathematical models of plants are not known. (These rules can, of course, be applied to the design of systems with known mathematical models.)

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. Here, however, we introduce only the Ziegler–Nichols tuning rules.)

There are two methods called Ziegler–Nichols tuning rules. In both methods, they aimed at obtaining 25% maximum overshoot in step response (see Figure 10–2).

First method. In the first method, we obtain experimentally the response of the plant to a unit-step input, as shown in Figure 10–3. If the plant involves neither inte-

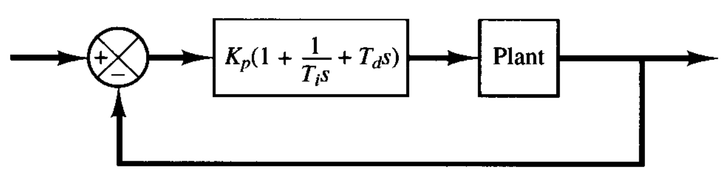


Figure 10–1 PID control of a plant.

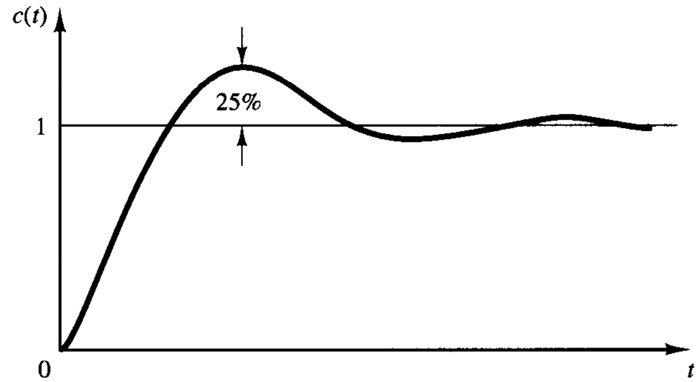


Figure 10–2
Unit-step response curve showing 25% maximum overshoot.

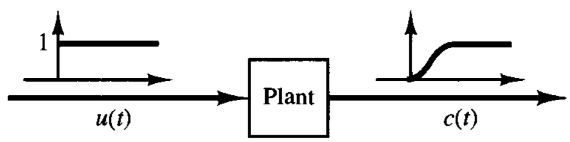


Figure 10–3
Unit-step response of a plant.

grator(s) nor dominant complex-conjugate poles, then such a unit-step response curve may look like an S-shaped curve, as shown in Figure 10–4. (If the response does not exhibit an S-shaped curve, this method does not apply.) 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 10–4. The transfer 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

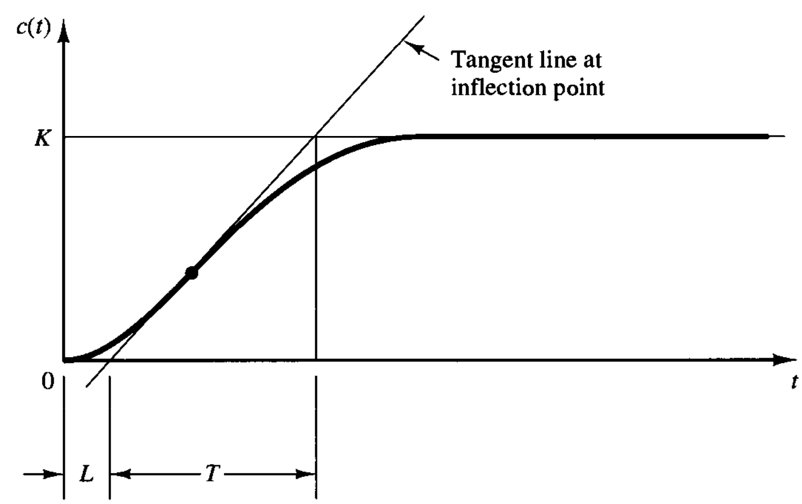


Figure 10–4
S-shaped response curve.

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	$1.2\frac{T}{L}$	$2L$	$0.5L$

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 1.2\frac{T}{L}\left(1 + \frac{1}{2Ls} + 0.5Ls\right) \\ &= 0.6T\frac{\left(s + \frac{1}{L}\right)^2}{s} \end{aligned}$$

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 10–5), increase K_p from 0 to a critical value K_{cr} where 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 determined (see Figure 10–6). 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.

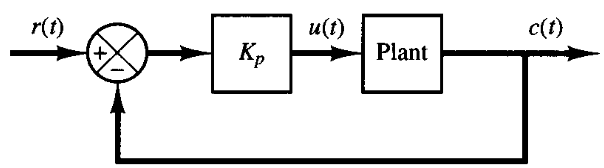


Figure 10–5 Closed-loop system with a proportional controller.

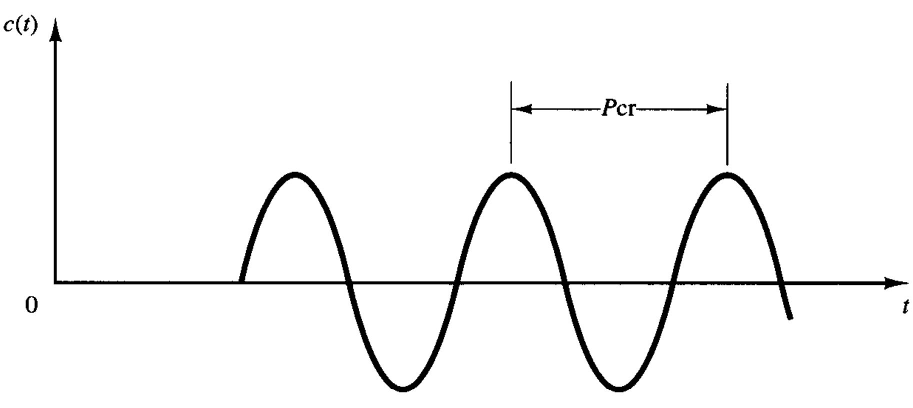


Figure 10–6 Sustained oscillation with period P_{cr}.

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

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 0.6K_{cr}\left(1 + \frac{1}{0.5P_{cr}s} + 0.125\, P_{cr}s\right) \\ &= 0.075K_{cr}P_{cr}\frac{\left(s + \dfrac{4}{P_{cr}}\right)^2}{s} \end{aligned}$$

Thus, the PID controller has a pole at the origin and double zeros at $s = -4/P_{cr}$.

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 plant dynamics are known, many analytical and graphical approaches to the design of PID controllers are available, in addition to Ziegler–Nichols tuning rules.)

If the transfer function of the plant is known, a unit-step response may be calculated or the critical gain K_{cr} and critical period P_{cr} may be calculated. Then, using those calculated values, it is possible to determine the parameters K_p, T_i, and T_d from Table 10–1 or 10–2. However, the real usefulness of Ziegler–Nichols tuning rules (and other tuning rules) becomes apparent when the plant dynamics are not known so that no analytical or graphical approaches to the design of controllers are available.

Generally, for plants with complicated dynamics but no integrators, Ziegler–Nichols tuning rules can be applied. However, if the plant has an integrator, these rules may not be applied in some cases. To illustrate such a case where Ziegler–Nichols rules do not apply, consider the following case: Suppose that a unity-feedback control system has a plant whose transfer function is

$$G(s) = \frac{(s + 2)(s + 3)}{s(s + 1)(s + 5)}$$

Because of the presence of an integrator, the first method does not apply. Referring to

Figure 10–3, the step response of this plant will not have an S-shaped response curve; rather, the response increases with time. Also, if the second method is attempted (see Figure 10–5), the closed-loop system with a proportional controller will not exhibit sustained oscillations whatever value the gain K_p may take. This can be seen from the following analysis. Since the characteristic equation is

$$s(s + 1)(s + 5) + K_p(s + 2)(s + 3) = 0$$

or

$$s^3 + (6 + K_p)s^2 + (5 + 5K_p)s + 6K_p = 0$$

the Routh array becomes

$$\begin{array}{lcc} s^3 & 1 & 5 + 5K_p \\ s^2 & 6 + K_p & 6K_p \\ s^1 & \dfrac{30 + 29K_p + 5K_p^2}{6 + K_p} & 0 \\ s^0 & 6K_p & \end{array}$$

The coefficients in the first column are positive for all values of positive K_p. Thus, in the present case the closed-loop system will not exhibit sustained oscillations and, therefore, the critical gain value K_{cr} does not exist. Hence, the second method does not apply.

If the plant is such that Ziegler–Nichols rules can be applied, then the plant with a PID controller tuned by Ziegler–Nichols rules will exhibit approximately 10% ~ 60% maximum overshoot in step response. On the average (experimented on many different plants), the maximum overshoot is approximately 25%. (This is quite understandable because the values suggested in Tables 10–1 and 10–2 are based on the average.) In a given case, if the maximum overshoot is excessive, it is always possible (experimentally or otherwise) to make fine tuning so that the closed-loop system will exhibit satisfactory transient responses. In fact, Ziegler–Nichols tuning rules give an educated guess for the parameter values and provide a starting point for fine tuning.

EXAMPLE 10–1 Consider the control system shown in Figure 10–7 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%.

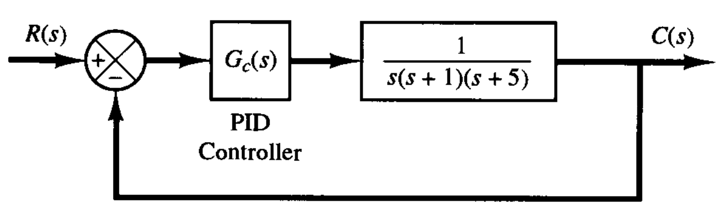


Figure 10–7 PID-controlled system.

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:

$$\begin{array}{lcc} s^3 & 1 & 5 \\ s^2 & 6 & K_p \\ s^1 & \dfrac{30 - K_p}{6} & \\ s^0 & K_p & \end{array}$$

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

$$s^3 + 6s^2 + 5s + 30 = 0$$

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

$$P_{cr} = \frac{2\pi}{\omega} = \frac{2\pi}{\sqrt{5}} = 2.8099$$

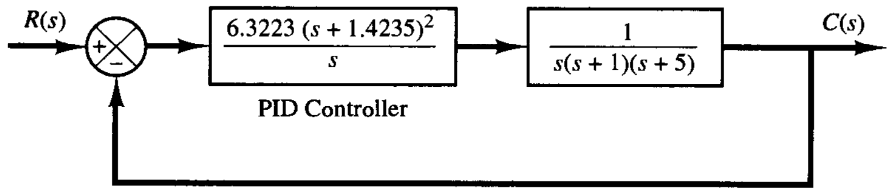
Referring to Table 10–2, we determine K_p, T_i, and T_d as follows:

$$K_p = 0.6K_{cr} = 18$$
$$T_i = 0.5P_{cr} = 1.405$$
$$T_d = 0.125P_{cr} = 0.35124$$

The transfer function of the PID controller is thus

$$\begin{aligned} G_c(s) &= K_p\left(1 + \frac{1}{T_i s} + T_d s\right) \\ &= 18\left(1 + \frac{1}{1.405s} + 0.35124s\right) \\ &= \frac{6.3223(s + 1.4235)^2}{s} \end{aligned}$$

Figure 10–8
Block diagram of the system with PID controller designed by use of Ziegler–Nichols tuning rule (second method).



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 10–8.

Next, let us examine the unit-step response of the system. The closed-loop transfer function $C(s)/R(s)$ is given by

$$\frac{C(s)}{R(s)} = \frac{6.3223s^2 + 18s + 12.811}{s^4 + 6s^3 + 11.3223s^2 + 18s + 12.811}$$

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 10–9. 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

$$G_c(s) = 18\left(1 + \frac{1}{3.077s} + 0.7692s\right) = 13.846\frac{(s + 0.65)^2}{s} \qquad (10\text{–}1)$$

the maximum overshoot in the unit-step response can be reduced to approximately 18% (see Figure 10–10). 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

$$G_c(s) = 39.42\left(1 + \frac{1}{3.077s} + 0.7692s\right) = 30.322\frac{(s + 0.65)^2}{s} \qquad (10\text{–}2)$$

then the speed of response is increased, but the maximum overshoot is also increased to approximately 28%, as shown in Figure 10–11. 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 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

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')
```

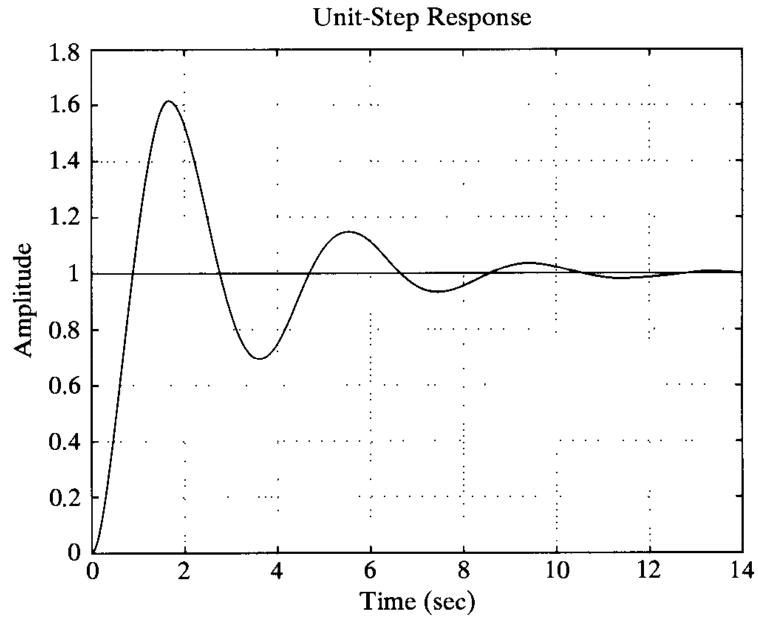


Figure 10–9 Unit-step response curve of PID-controlled system designed by use of Ziegler–Nichols tuning rule (second method).

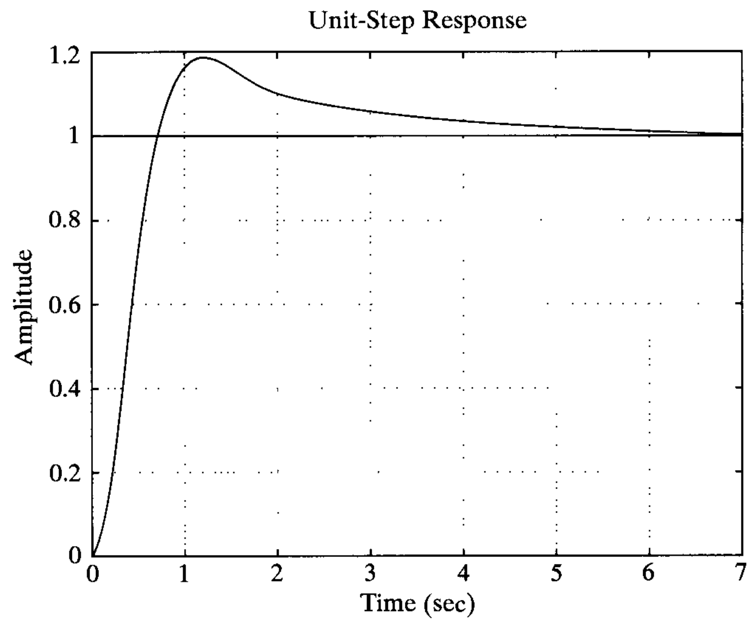


Figure 10–10 Unit-step response of the system shown in Figure 10–7 with PID controller having parameters $K_p = 18$, $T_i = 3.077$, and $T_d = 0.7692$.

overshoot is concerned, varying gain K_p has very little effect. The reason for this may be seen from the root-locus analysis. Figure 10–12 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 10–13 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.