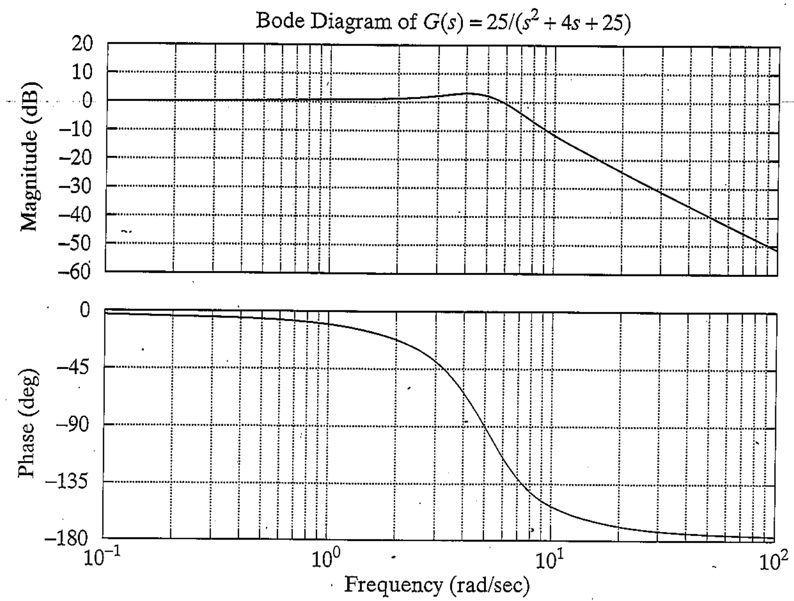


Figure 11–18 Bode diagram of $G(s) = \dfrac{25}{s^2 + 4s + 25}$.

MATLAB Program 11–1

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

11–4 NYQUIST PLOTS AND THE NYQUIST STABILITY CRITERION

In this section, we first discuss Nyquist plots and then the Nyquist stability criterion. We then define the phase margin and gain margin, which are frequently used for determining the relative stability of a control system. Finally, we discuss conditionally stable systems.

Nyquist plots. The *Nyquist plot* of a sinusoidal transfer function $G(j\omega)$ is a plot of the magnitude of $G(j\omega)$ versus the phase angle of $G(j\omega)$ in polar coordinates as ω is varied from zero to infinity. Thus, the polar plot is the locus of vectors $|G(j\omega)|\angle G(j\omega)$ 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 Nyquist plot is often called the *polar plot*. An example of such a

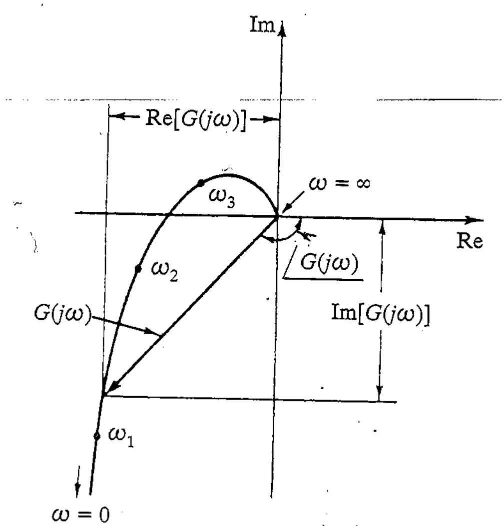


Figure 11–19 Nyquist plot.

plot is shown in Figure 11–19. Each point on the polar plot of $G(j\omega)$ represents the terminal point of a vector at a particular value of ω. The projections of $G(j\omega)$ on the real and imaginary axes are the real and imaginary components of the function.

An advantage in using a Nyquist 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 contribution of each individual factor of the open-loop transfer function.

Table 11–1 shows examples of Nyquist plots of simple transfer functions.

The general shapes of the low-frequency portions of the Nyquist plots of type 0, type 1, and type 2 minimum-phase systems are shown in Figure 11–20(a). 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 $G(j\omega)$ loci converge clockwise to the origin. At $\omega = \infty$, the loci are tangent to one or the other axis, as shown in Figure 11–20(b).

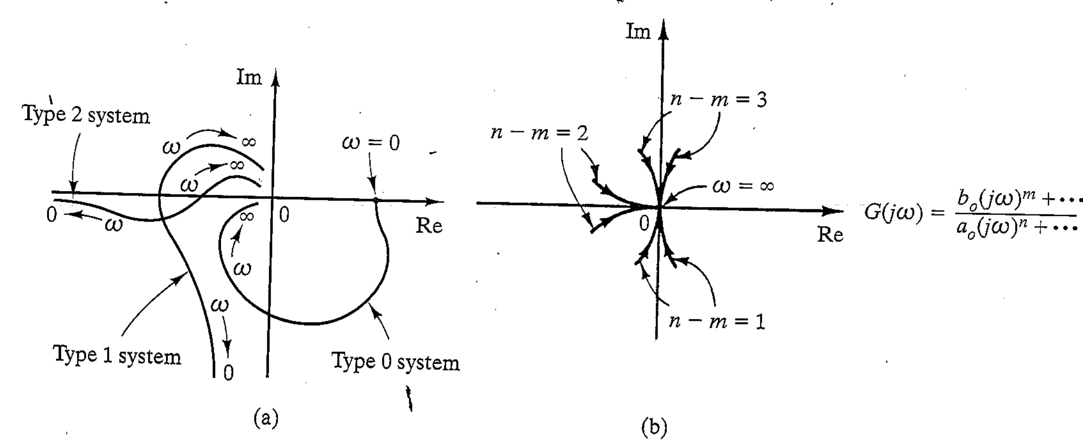
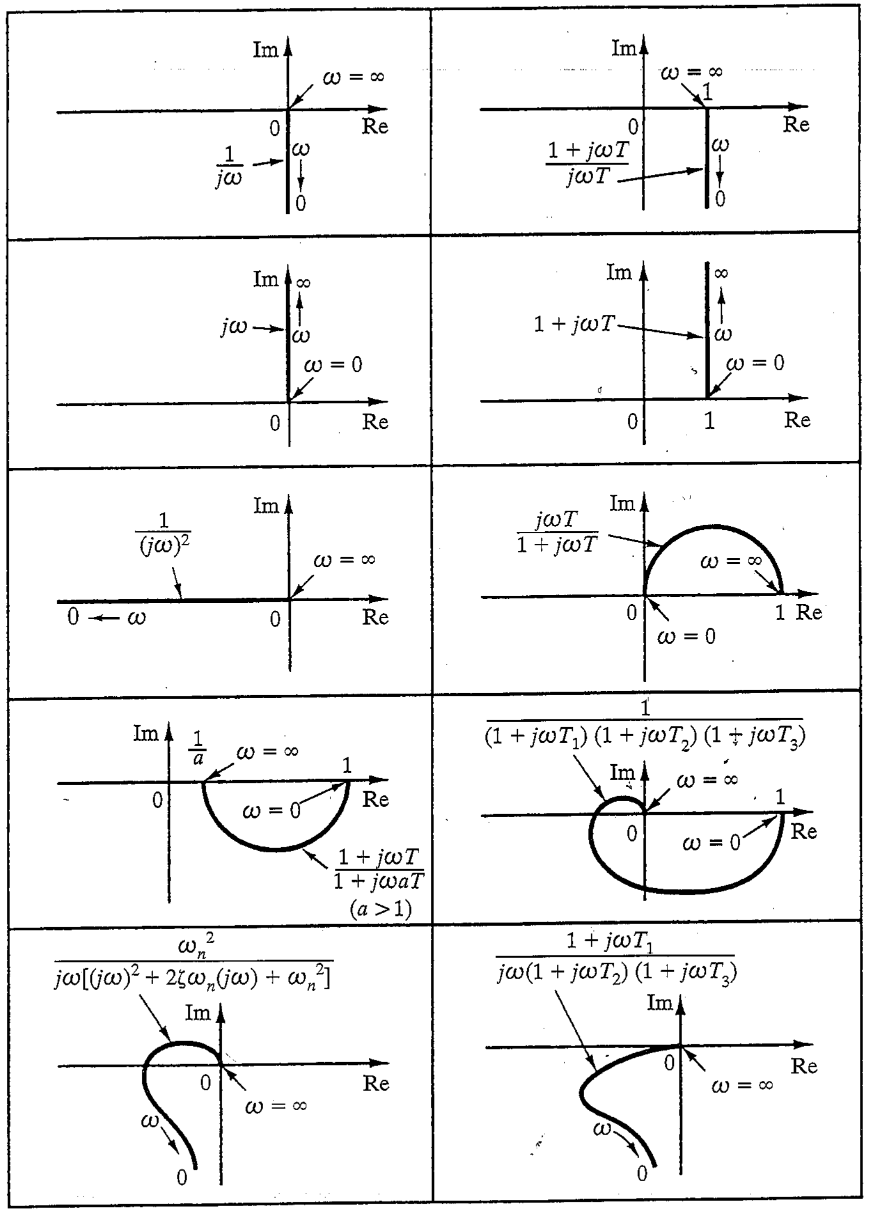


TABLE 11–1 Nyquist Plots of Simple Transfer Functions



For the case where the degrees of the denominator and numerator polynomials of $G(j\omega)$ are the same, the Nyquist plot starts at a finite distance on the real axis and ends at a finite point on the real axis.

Note that any complicated shapes in the Nyquist plot curves are caused by the numerator dynamics—that is, the time constants in the numerator of the transfer function.

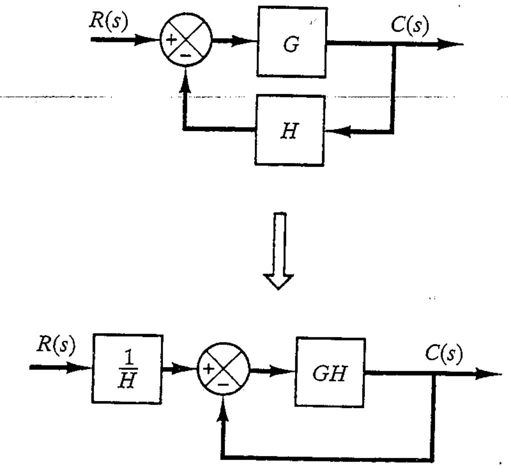


Figure 11–21 Modification of a control system with feedback elements to a unity-feedback control system.

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

In the discussion that follows, we shall assume that the systems considered have unity feedback. Note that it is always possible to reduce a system with feedback elements to a unity-feedback system, as shown in Figure 11–21. Hence, the extension of relative stability analysis for the unity-feedback system to nonunity-feedback systems is possible.

Now consider the system shown in Figure 11–22. The closed-loop transfer function is

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

For stability, all roots of the characteristic equation

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

must lie in the left-half s-plane. The Nyquist stability criterion relates the open-loop frequency response $G(j\omega)$ to the number of zeros and poles of $1 + G(s)$ that lie in the right-half s-plane. This criterion, due to H. Nyquist, is useful in control engineering

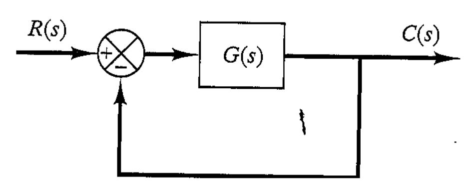


Figure 11–22 Unity-feedback control system

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 experimentally obtained curves, can be used for the stability analysis. This confluence of the two types of curve 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 can be stated as follows:

Nyquist stability criterion: In the system shown in Figure 11–22, if the open-loop transfer function $G(s)$ has P poles in the right-half s-plane, then, for stability, the $G(s)$ locus as a representative point s traces out the Nyquist path in the clockwise direction must encircle the $-1 + j0$ point P times in the counterclockwise direction.

The Nyquist path is a closed contour that 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. Thus, the Nyquist path encloses the entire right-half s-plane. The direction of the path is clockwise.

Remarks on the Nyquist stability criterion

1. The Nyquist stability criterion can be expressed as

$$Z = N + P \tag{11–9}$$

where

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

If P is not zero, then, 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 $G(s)$ does not have any poles in the right-half s-plane, then, from Equation (11–9), we must have $Z = N$ for stability. For example, consider the system with the following open-loop transfer function:

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

Figure 11–23 shows the Nyquist path and $G(s)$ loci for a small and large value of the gain K. Since the number of poles of $G(s)$ in the right-half s-plane is zero, for this system to be stable, it is necessary that $N = Z = 0$, or that the $G(s)$ locus not encircle the $-1 + j0$ point.

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 $G(s)$ 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,

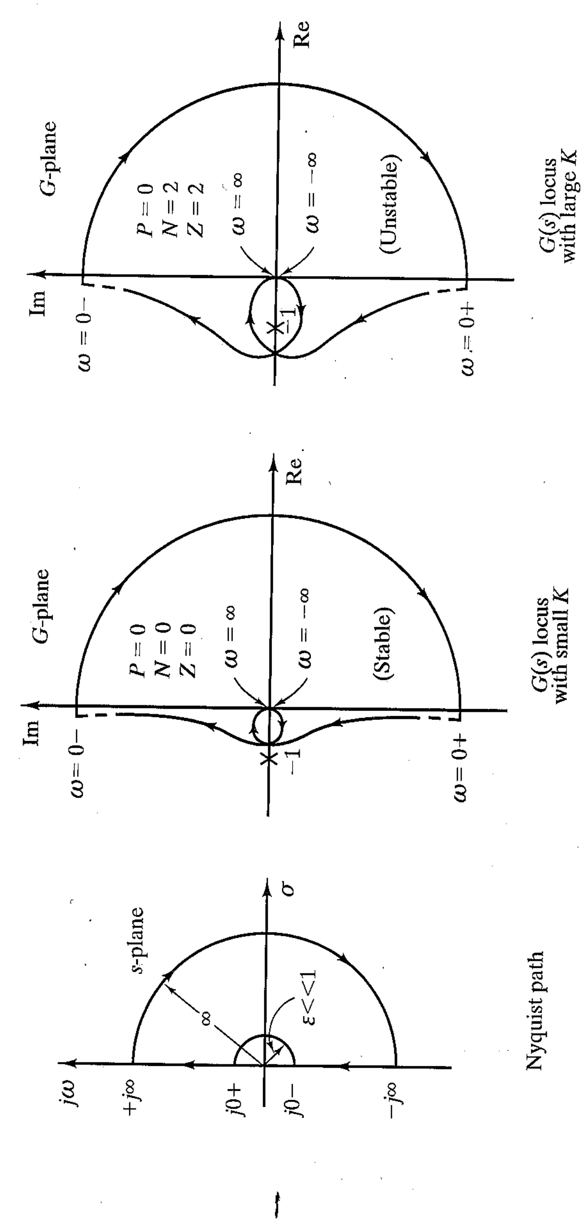


Figure 11–23 Nyquist path and $G(s)$ loci for a small and large value of the gain K.

however, a large value of K causes poor stability or even instability. To compromise between accuracy and stability, it is necessary to insert a compensator into the system.

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

Phase and gain margins. Figure 11–24 shows Nyquist 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, the system is on the verge of instability and will exhibit sustained oscillations. For a small value of the gain K, the system is stable.

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

Phase margin. The *phase margin* is that amount of additional phase lag at the gain crossover frequency required to bring the system to the verge of instability.

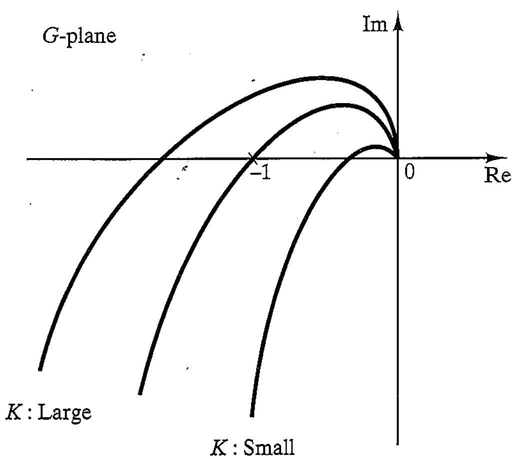


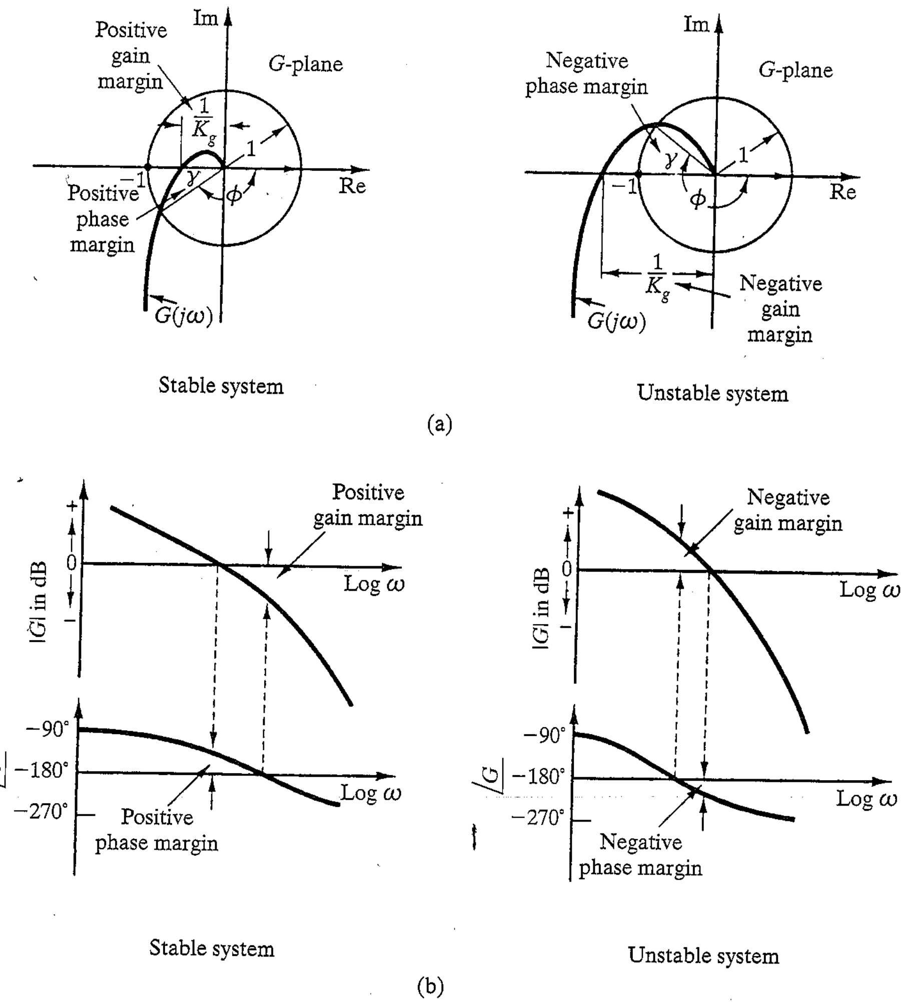
Figure 11–24 Polar plots of $K(1 + j\omega T_a)(1 + j\omega T_b)\cdots$

The gain crossover frequency is the frequency at which $|G(j\omega)|$, the magnitude of the open-loop transfer function, is unity. The phase margin γ is 180° plus the phase angle ϕ of the open-loop transfer function at the gain crossover frequency, or

$$\gamma = 180° + \phi$$

On the Nyquist 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, which is positive for $\gamma > 0$ and negative for $\gamma < 0$. For a minimum-phase system to be stable, the phase margin must be positive.

Figures 11–25(a) and (b) illustrate the phase margins of a stable system and an unstable system in Nyquist plots and Bode diagrams. In the Bode diagrams, the critical point in the complex plane corresponds to the 0-dB line and −180° line.



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,

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

Expressed in decibels, the gain margin is positive if K_g is greater than unity and negative if K_g is smaller than unity. Thus, a positive gain margin (in decibels) means that the system is stable, and a negative gain margin (in decibels) means that the system is unstable. The gain margin is shown in Figures 11–25(a) and (b).

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 margins of first- and second-order minimum-phase systems are infinite, since the Nyquist plots for such systems do not cross the negative real axis. Thus, such first- and second-order systems cannot be unstable.

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 Nyquist plot to the $-1 + j0$ point. Therefore, these margins may be used as design criteria.

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

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

Proper phase and gain margins ensure against variations in a system's components. 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 a 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 is 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 the slope is -40 dB/decade, the system could be either stable or unstable. (Even if the system is stable, however, the phase margin is

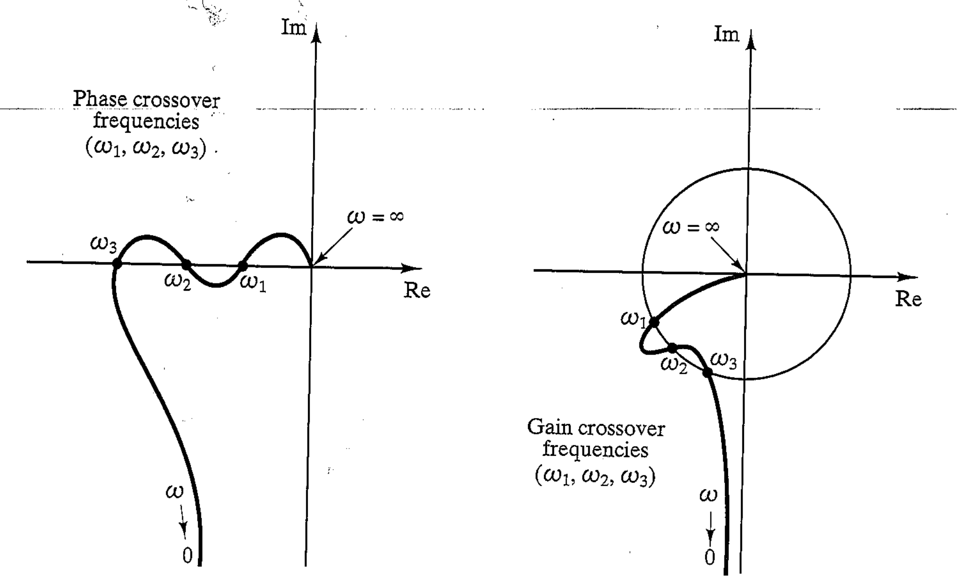


Figure 11–26 Nyquist plots showing more than two phase or gain crossover frequencies.

small.) If the slope at the gain crossover frequency is -60 dB/decade or steeper, the system will be unstable.

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 11–26. For stable systems having two or more gain crossover frequencies, the phase margin is measured at the highest gain crossover frequency.

Conditionally stable systems. Figure 11–27 is an example of a $G(j\omega)$ locus for which the closed-loop system can be made stable or unstable by varying the open-loop gain. If the open-loop gain is increased sufficiently, the $G(j\omega)$ locus encloses the $-1 + j0$ point twice, and the system becomes unstable. If the open-loop gain is decreased sufficiently, again the $G(j\omega)$ locus encloses the $-1 + j0$ point twice. The system is stable only for the limited range of values of the open-loop gain for which the $-1 + j0$ point is completely outside the $G(j\omega)$ locus. Such a system is *conditionally stable.*

Such a conditionally stable system becomes unstable when large input signals are applied, since a large signal may cause saturation, which in turn reduces the open-loop gain of the system.

For stable operation of the conditionally stable system considered here, the critical point $-1 + j0$ must not be located in the regions between OA and BC shown in Figure 11–27.

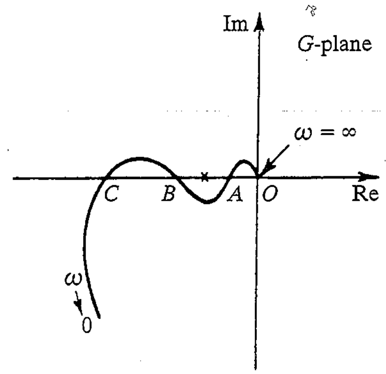


Figure 11–27 Nyquist plot of a conditionally stable system.

11–5 DRAWING NYQUIST PLOTS WITH MATLAB

Nyquist plots, just like Bode diagrams, are commonly used in the frequency-response representation of linear, time-invariant 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. That is, the command

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

draws the Nyquist plot of the transfer function

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

where num and den contain the polynomial coefficients in descending powers of s.

The command

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

employs the user-specified frequency vector w, which gives the frequency points in radians per second at which the frequency response will be calculated.

When invoked with left-hand arguments, as in

```
[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 11–4

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 11–2 produces the Nyquist plot shown in Figure 11–28. 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—we enter the following command into the computer:

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

Alternatively, we may combine these two lines into one as follows:

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

MATLAB Program 11–2

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

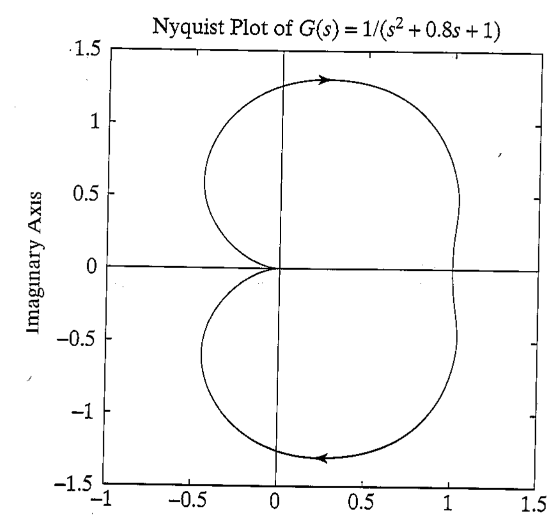


Figure 11–28 Nyquist plot of $G(s) = \frac{1}{\cdots}$

Example 11–5

Draw a Nyquist plot for

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

MATLAB Program 11–3 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 11–29. Notice that the plot 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 commands

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

MATLAB Program 11–3

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

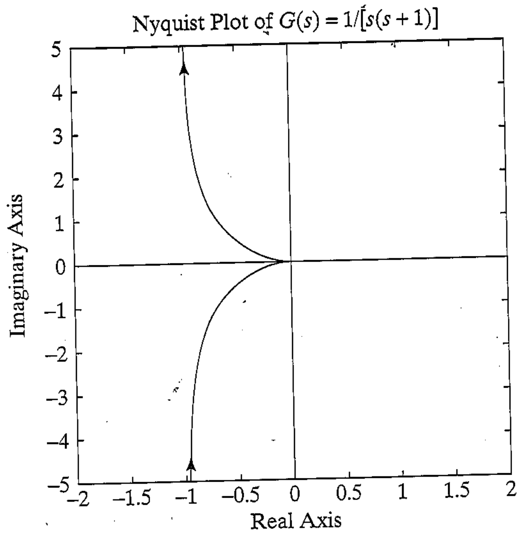


Figure 11–29 Nyquist plot of $G(s) = 1/s(s+1)$. (The plot shows Nyquist loci for both $\omega > 0$ and $\omega < 0$.)

MATLAB Program 11–4 uses these two lines of commands. The resulting Nyquist plot

MATLAB Program 11–4

```
>> num = [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('Imaginary Axis')
```

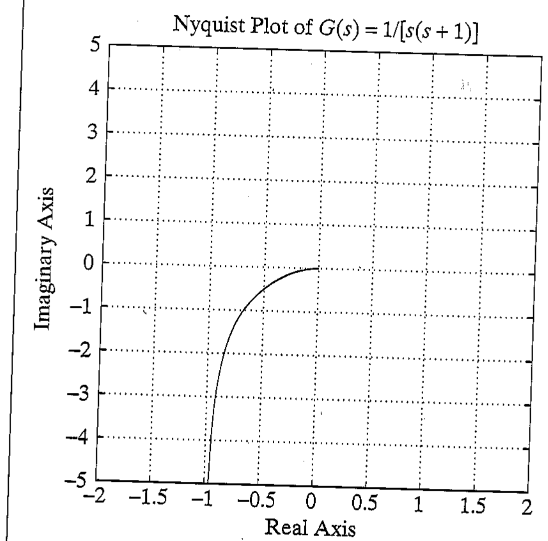


Figure 11–30 Nyquist plot of $G(s) = 1/s(s+1)$. (The plot shows nyquist locus for $\omega > 0$.)

11–6 DESIGN OF CONTROL SYSTEMS IN THE FREQUENCY DOMAIN

This section discusses control systems design based on the Bode diagram approach, an approach that is particularly useful for the following reasons:

1. In the Bode diagram, the low-frequency asymptote of the magnitude curve is indicative of one of the static error constants K_p, K_v, or K_a.
2. Specifications of the transient response can be translated into those of the frequency response in terms of the phase margin, gain margin, bandwidth, and so forth. These specifications can be easily handled in the Bode diagram. In particular, the phase and gain margins can be read directly from the Bode diagram.
3. The design of a compensator or controller to satisfy the given specifications (in terms of the phase margin and gain margin) can be carried out in the Bode