NYQUIST CRITERION, DIAGRAMS, AND STABILITY

H. Nyquist (1889–1976), born in Sweden, is known for his contributions to telephone transmission problems in the 1920s. He is also well known for his contributions in the stability of feedback systems.

In 1927 H. S. Black invented the negative feedback amplifier. Part of the output signal was returned to the amplifier's input to reduce the total system gain. This resulted in a flatter frequency response, a wider bandwidth, and a decrease in the nonlinear distortion, since improperly designed amplifiers were unstable, producing undesired results. In the late 1920s and early 1930s, H. Nyquist and H. W. Bode, a colleague at the Bell Telephone Laboratories, developed a mathematical analysis for feedback amplifiers. Later developments evolved into sinusoidal frequency analysis and design techniques applied to feedback control systems. Although, at that time, the stability criteria for vibrating mechanical systems already existed and had been applied to feedback systems, the idea of algebraic problems on complex roots of polynomials were just arising. The Nyquist criterion offered geometrical solutions and was much easier to apply to amplifiers.

Later, the Nyquist criterion was used to provide vital information on stability essential in the analysis and design of general control systems. It is a graphical method that relates the stability of a closed-loop system to the open-loop frequency response and the locations of poles and zeros. The Nyquist diagram method was found to be very useful in the design of linear feedback systems of all types. An important application in World War II was the feedback control of direction of guns that employed electromechanical feedback-controlled servomechanisms. Before computers became widespread, Nyquist diagrams were largely obtained by calculations and hand-drawn graphics. But, today many companies offer a diverse range of computer software for simulation, analysis and design of control problems. Some of these manufacturers are listed in Table 1. Popular software such as MATLAB, MATHCAD, and SIMULINK include control system tools.

Table 1. List of Manufacturers

AAI Corporation P.O. Box 126 Hunt Valley, MD 21030-0126 Tel: 800-655-2616 Fax: 410-628-8616	Murphy Software Company 1000 Town Center, Suite 1950 Southfield, MI 48075 Tel: 248-351-0900 Fax: 248-351-0906
Cape Software, Inc. 333-T N. Sam Houston Parkway, E. Houston, TX 77060-2403 Tel: 281-448-5177 Fax: 281-448-2607	ProModel Corporation 1875-T S. State Street, Suite 3400 Orem, UT 84097-8075 Tel: 801-223-4600 Fax: 801-226-6046
Design Technology Corporation 5 Suburban Park Drive Billerica (Boston), MA 01821 Tel: 508-663-7000 Fax: 508-663-6841	Scientific and Engineering Software 4301-T Westbank Drive Austin, TX 78746-6546 Tel: 512-328-5544 Fax: 512-327-6646
MicroMath Research Ltd. P.O. Box 71550 Salt Lake City, UT 84171-0550 Tel: 801-483-2949 Fax: 801-483-3025	Wolfram Research, Inc. 100-T Trade Center Drive, Suite 600 Champaign, IL 61820-4853 Tel: 217-398-0700 Fax: 217-398-0747
Mitech Corporation 43-T Nagog Park Acton, MA 01720 Tel: 978-263-7999 Fax: 978-263-8081	

In design and stability analysis, the Nyquist method exhibits distinct features. It provides the same information on the stability of a control system, as does the Routh–Hurwitz criterion. In addition, the Nyquist method indicates the relative stability or instability of the system. Properties of the frequency-domain plots of the loop transfer function $G(s)H(s)$ provide an alternative approach to root locus, and give information on the frequency-domain characteristics of the closed-loop system. One other major feature of the Nyquist method is that it relates a closed-loop system to the transient response and steady-state errors.

ENCIRCLEMENTS AND ENCLOSURES

Before embarking into the Nyquist stability criterion, the concepts of encirclement and enclosures need to be established. These concepts are essential in the interpretation of the Nyquist plots and stability analysis.

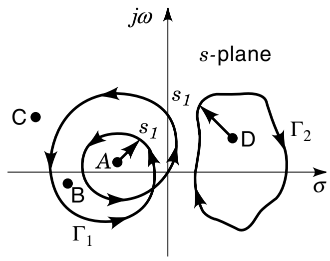

Figure 1. Encirclements and enclosures of points by contours.

In a complex plane, a point is said to be encircled by a closed path if it lies inside the path. If the closed path has a direction assigned for being in a clockwise or counterclockwise direction, then the region inside the path is considered to be encircled in that prescribed direction.

A point or a region is said to be enclosed by a path if it is encircled in the counterclockwise direction. Alternatively, the point or region is enclosed if it lies on the left of the path when the path is traversed in any prescribed direction. The second definition is more useful in situations where only some portion of the closed path is drawn.

As an example, consider the two cases illustrated in Fig. 1. In accordance with the foregoing definitions, the contour Γ_1 encircles the point A twice. The point B is encircled only once, whereas points C and D are not encircled at all by Γ_1. As the contour Γ_1 traverses in the counterclockwise direction, the points A and B lie on the left side; therefore, they are both enclosed, and the points C and D are not enclosed. The contour Γ_2 encircles only the point D in the clockwise direction. Interestingly enough, the contour Γ_2 does not enclose the point D, but it encloses other points A, B, and C. A path Γ can encircle a point N number of times, and the magnitude of N can be determined by drawing a phasor from the point to an arbitrary point s_1 along the path Γ, as illustrated for point A in the contour Γ_1. The point s_1 is traversed along Γ until it returns to its starting position. The net number of revolutions traversed by the phasor is N and the total angle traced is $2\pi N$ radians.

STABILITY CONDITIONS

The stability of linear time-invariant systems depends upon the roots of the characteristic equation on the s-plane. In contrast to root locus plots, the Nyquist criterion does not give the exact locations of the roots of the characteristic equation but indicates the locations of the roots with respect to the left or the right half of the s-plane.

In many control systems, the relations of the bounded-inputs to the bounded-outputs (BIBO) define the stability. The definition states that if bounded inputs yield to bounded outputs, then the system is considered stable. The BIBO relationship can also be related to the roots of the characteristic equation as shown next.

Let's take a linear time-invariant system, as illustrated in Fig. 2, with $r(t)$ as the input, $c(t)$ as the output, and $g(t)$ as the impulse response. The convolution integral relating the function $r(t)$, $c(t)$, and $g(t)$ is given by

$$c(t) = \int_0^\infty r(t-\tau)g(\tau)\,d\tau \tag{1}$$

By taking the absolute value of both sides of Eq. (1) yields

$$|c(t)| = \left|\int_0^\infty r(t-\tau)g(\tau)\,d\tau\right| \tag{2}$$

or

$$|c(t)| \le \int_0^\infty |r(t-\tau)||g(\tau)|\,d\tau \tag{3}$$

Let the input $r(t)$ be bounded by a finite positive number R such that

$$|r(t)| \le R \tag{4}$$

Thus, Eq. (2) becomes

$$|c(t)| \le R\int_0^\infty |g(\tau)|\,d\tau \tag{5}$$

If $c(t)$ is also to be bounded by a positive finite number C, with

$$|c(t)| \le C < \infty \tag{6}$$

then

$$R\int_0^\infty |g(\tau)|\,d\tau \le C < \infty \tag{7}$$

Dividing Eq. (7) through by R and letting C/R equal to Q, a positive finite number, results in

$$\int_0^\infty |g(\tau)|\,d\tau \le Q < \infty \tag{8}$$

For Eq. (8) to hold, the integral of the absolute value of $g(t)$ must be finite.

The Laplace transform may be used to show the relationship between the roots of the characteristic equation and Eq. (8). For $g(t)$

$$L[g(t)] = \int_0^\infty g(t)e^{-st}\,dt = G(s) \tag{9}$$

Where s is a complex number having real and imaginary components on the s-plane as $s = \sigma + j\omega$. Taking the absolute value on both sides of the equation yields

$$|G(s)| = \left|\int_0^\infty g(t)e^{-st}\,dt\right| \le \int_0^\infty |g(t)||e^{-st}|\,dt \tag{10}$$

But since

$$|e^{-st}| = |e^{-\sigma t}| \tag{11}$$

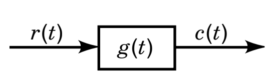

Figure 2. Block diagram representation of a linear time invariant system. The relation between input and output is given by the impulse response transfer function $g(t)$. The $g(t)$ may be a simple gain or a complex function involving derivatives and integrals.

substituting Eq. (11) into Eq. (10) gives

$$\infty \leq \int_0^\infty |g(t)||e^{-\sigma t}|\,dt \tag{12}$$

Note that the imaginary part, $j\omega$, of the complex variable s does not bear any importance in the proof leading to the BIBO stability. All that is needed is the mathematical relation between real parts of the poles in the s-plane, that is whether they lie in the right half or the left half of the complex plane.

If one or more roots of the characteristic equation lies in the right half of the s-plane, or on the $j\omega$-axis if σ is greater than or equal to zero, then

$$|e^{-\sigma t}| \leq R = 1 \tag{13}$$

Substitution of Eq. (13) into Eq. (12) yields

$$\infty \leq \int_0^\infty R|g(t)|\,dt = \int_0^\infty |g(t)|\,dt \tag{14}$$

Note that Eq. (14) does not satisfy the BIBO relation because the equation is not bounded as in Eq. (8). Hence, to satisfy the BIBO stability, the roots of the characteristic equation or the poles of $G(s)$ must lie on the left side of the $j\omega$-axis. A system is classified to be marginally stable if the first-order poles, poles of conjugate pairs, lie on the $j\omega$-axis. However, multiple-order poles or repeating conjugate pairs of poles represent an unstable system. In addition, a system is classified as unstable if more than one pole exists at the origin.

Another definition that is worth mentioning and that helps in the understanding of the Nyquist criterion is the steady-state error. Steady-state error is the difference between the input and the output as $t \to \infty$ for a given test input.

The steady state errors are generally described for three main types of test inputs: the step, the ramp, and the parabolic. Often, control systems are subjected to these inputs to test their ability to give the required outputs. Usually, these test inputs are in the form of electrical signals that have defined waveforms. For example, the parabolic input has a constant second derivative, which represents acceleration with respect to accelerating targets. In general, the output of any system can be represented by the sum of the natural response and the forced response. In relation to the Nyquist stability criterion, a steady state error can be calculated from the closed-loop transfer function of the system $M(s)$ or the loop transfer function $G(s)H(s)$.

THE PRINCIPAL ARGUMENT

The Nyquist criterion is based on a theorem using the theory of complex variables, which leads to the principal argument. The principal argument may be presented in a number of ways. Here, two approaches will be presented.

CASE 1. In this case, a rigorous mathematical approach may be employed by using theories of contours and mappings in complex planes. Assume that $F(s)$ is a function of s and single-valued, that is, for each point in the s-plane there exist a corresponding point, including infinity, in the $F(s)$-plane, and the function consists of a finite number of poles and a finite number of zeros in the s-plane. Now, suppose that there is an arbitrarily chosen continuous closed path Γ_s in the s-plane, then the values of s in Γ_s maps a new closed continuous path Γ_F on the complex $F(s)$-plane. Some examples of mapping are presented in Table 2 and illustrated in Fig. 3. In Fig. 3, it can be seen that for every closed contour on the s-plane there is a corresponding contour on the $F(s)$-plane. If the contour Γ_s traverses in a selected direction say counterclockwise, Fig. 3 shows that the direction of traverse of Γ_F can either be in the same or opposite direction depending on the number of poles and zeros of function $F(s)$ located on contour Γ_s. Lets express $F(s)$ in the following familiar form as seen in the control theory

Table 2. Calculation of Points for the Illustration of Mapping for Contours in Fig. 3

Γ_1	$s = 1 + j$	$s = 1 - j$	$s = -1 - j$	$s = -1 + j$
$F(s)$	$-\frac{7}{5} - j\frac{4}{5}$	$-\frac{7}{5} + j\frac{4}{5}$	$-\frac{7}{13} + j\frac{4}{13}$	$-\frac{7}{13} - j\frac{4}{13}$
Γ_2	$s = 5 + j$	$s = 5 - j$	$s = 3 - j$	$s = 3 + j$
$F(s)$	$5 - j4$	$5 + j4$	$-3 + j4$	$-3 - j4$
Γ_3	$s = -3 + j$	$s = -3 - j$	$s = -5 - j$	$s = -5 + j$
$F(s)$	$-\frac{3}{25} - j\frac{4}{25}$	$-\frac{3}{25} + j\frac{4}{25}$	$\frac{5}{41} + j\frac{4}{41}$	$\frac{5}{41} - j\frac{4}{41}$
Γ_4	$s = 7 + 3j$	$s = 7 - 3j$	$s = -7 - 3j$	$s = -7 + 3j$
$F(s)$	$\frac{7}{3} - j\frac{4}{3}$	$\frac{7}{3} + j\frac{4}{3}$	$\frac{21}{65} + j\frac{12}{65}$	$\frac{21}{65} - j\frac{12}{65}$

$$F(s) = \frac{\prod_{i=1}^{Z}(s+z_i)}{\prod_{k=1}^{P}(s+p_k)} \tag{15}$$

By using the Lucas formula, $F'(s)/F(s)$ can be written as

$$\frac{F'(s)}{F(s)} = \sum_{i=1}^{Z}\frac{1}{(s+z_i)} - \sum_{k=1}^{P}\frac{1}{(s+p_k)} \tag{16}$$

where $F'(s)$ is the first derivative of $F(s)$ with respect to s.

To illustrate this important point, let's take an example

$$F(s) = \frac{s+1}{(s+2)(s+3)} \tag{17}$$

Calculate

$$F'(s) = \frac{-s^2 - 2s + 1}{(s+2)^2(s+3)^2} \tag{18}$$

The ratio of $F'(s)/F(s)$ can be found to be

$$\frac{F'(s)}{F(s)} = \frac{-s^2 - 2s + 1}{(s+1))(s+2)(s+3)} \tag{19}$$

Writing the partial fractions

$$\frac{F'(s)}{F(s)} = \frac{-s^2 - 2s + 1}{(s+1)(s+2)(s+3)} = \frac{A}{(s+1)} + \frac{B}{(s+2)} + \frac{C}{(s+3)} \tag{20}$$

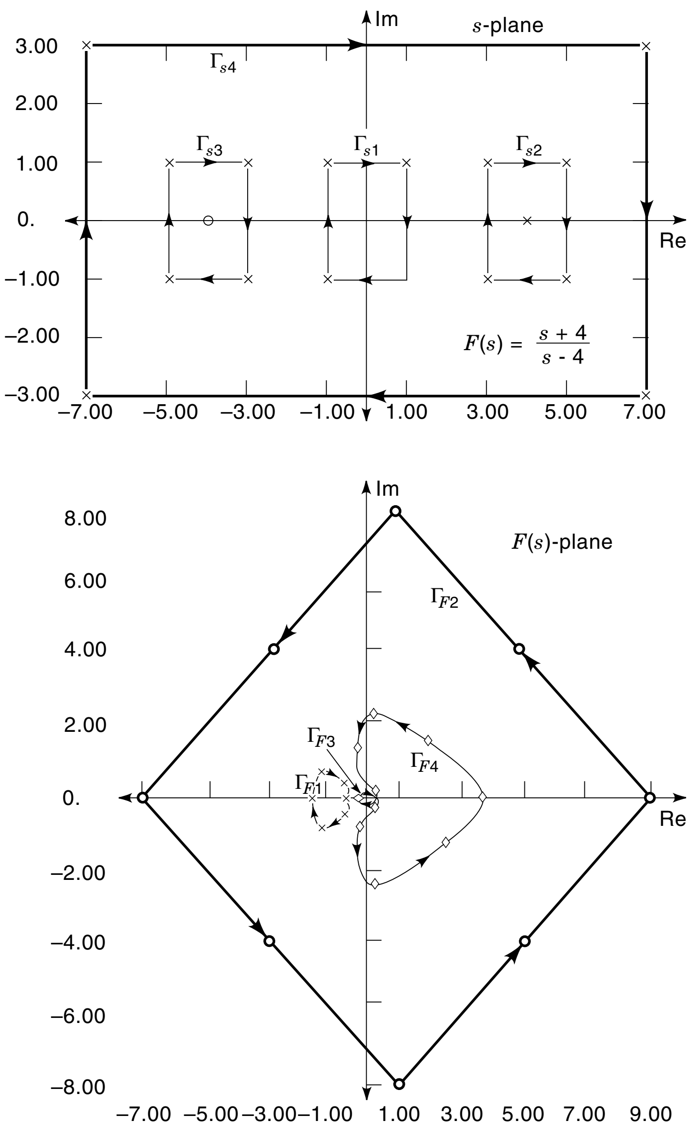

Figure 3. Mapping of s-plane contours to $F(s)$-plane. If the addition of number of poles and zeros of $F(s)$ in the s-plane is other than zero, the contour on the $F(s)$-plane encircles the origin. This is clearly illustrated by the corresponding contours of Γ_2 and Γ_3.

it can be proven that $A = 1, B = -1$, and $C = -1$, thus giving

$$\frac{F'(s)}{F(s)} = \frac{1}{(s+1)} - \frac{1}{(s+2)} - \frac{1}{(s+3)} \tag{21}$$

and verifying the general Eq. (16). Equations (16) and (21) indicate that the zeros of $F(s)$ appear as the denominators of the new equation with positive signs. The poles of $F(s)$ still appear in the denominators, but they have negative signs.

After having stated this important point, we can turn our attention to Cauchy's theorem. Although Cauchy's theorem may be of great mathematical interest, the intention of this article is not to discuss the intricacies involved in the theorem but rather to use it to illustrate relevant points in order to establish a solid understanding of Nyquist stability criteria.

Cauchy's theorem is based on the complex integrals, mapping of points and closed contours between the s-plane and the $f(s)$-plane. The theorem shows that if a complex function $f(s)$ is analytical (differentiable) bounded in a region by a simple closed curve γ, then

$$\int_{\gamma} f(s)\,ds = \oint_{\gamma} f(s)\,ds = 0 \tag{22}$$

However, consider a complex function $f(s) = 1/(s + a)$ with a pole at $-a$ on the s-plane. Let's draw a unit circle centered at $s = -a$ on the s-plane described by $\gamma(t) = -a + e^{j\theta}$ where the angle θ is $0 \leq \theta \leq 2\pi k$, and k is the number of encirclements of the point $-a$ by the unit circle.

Then the integral becomes

$$\int_{\gamma} f(s)\,ds = \int_{\gamma} \frac{1}{(s+a)}\,ds = \ln(s+a)|\gamma \tag{23}$$

The right-hand side of this equation can be evaluated by substituting values of $s(t) = -a + e^{j\theta}$ yielding

$$\ln(s+a)|\gamma = \ln(e^{j\theta} - a + a)|\gamma = j\theta|\gamma \tag{24}$$

Substituting the values of θ as $0 \leq \theta \leq 2\pi k$ gives

$$\int_{\gamma} f(s)\,ds = \int_{\gamma} \frac{1}{(s+a)}\,ds = 2\pi kj \tag{25}$$

On the contrast to Eq. (22) where $f(s)$ is analytic in the s-plane, for $f(s))$ having a singular point in the contour in the s-plane, the resulting closed contour is no longer zero but equal to multiples of $2\pi j$. This indicates that the contour on the $f(s)$-plane containing all the values of s on the s-plane goes through k number of revolutions around the origin. The number of revolutions around the origin depends on the number of times the point $-a$ is encircled in the s-plane. The encirclement k now can be expressed as

$$k = \frac{1}{2\pi j}\int_{\gamma} f(s)\,ds = \frac{1}{2\pi j}\int_{\gamma} \frac{1}{(s+a)}\,ds \tag{26}$$

If only one encirclement has taken place in the s-plane, $k = 1$, the result of this integration equals 1, indicating one revolution around the origin. Previously, a unit circle was considered, but the theory can be generalized for any closed contour around the pole $-a$.

We know that a complex function expressed as a ratio, as in Eq. (15), can be expressed as Eq. (16). Now, let's write Eq. (16) as

$$f(s) = \frac{F'(s)}{F(s)} = \sum_{i=1}^{Z} \frac{1}{(s+z_i)} - \sum_{k=1}^{P} \frac{1}{(s+p_k)} \tag{27}$$

where $f(s)$ can be viewed as the residues of $F(s)$. Now, substitute this Eq. (27) into Eq. (25)

$$\int_{\gamma} f(s)\,ds = \int_{\gamma} \sum_{i=1}^{Z} \frac{1}{(s+z_i)}\,ds - \int_{\gamma} \sum_{k=1}^{P} \frac{1}{(s+p_k)}\,ds \tag{28}$$

Making use of the similarities between Eq. (23) and Eq. (28) will yield the following conclusions. There will be Z number of revolutions around the origin, one for each z_i for $k = 1$.

Also, there will be P number of revolutions around the origin as a result of pole p_k, but this time they will be in the opposite direction, which is indicated by the negative sign of Eq. (28). From Eq. (26), the following may be written

$$\frac{1}{2\pi j}\int_{\gamma} f(s)\,ds = Z - P \tag{29}$$

This equation indicates that the number of times the origin is encircled depends on the difference between the zeros and poles, which are located on the contour of the s-plane.

There are many other approaches to arrive at similar conclusions, and further information may be found in the bibliography at the end of this article. This illustrative approach clarifies for readers many points, which otherwise might have been difficult to understand without intensive mathematical knowledge.

As a result of these explanations, the principle argument can now be stated as follows: Let $F(s)$ be a singled-valued function with a finite number of poles in the s-plane, and let Γ_s be chosen such that it does not pass through the poles or zeros of $F(s)$. Thus the corresponding Γ_F locus mapped in the $F(s)$-plane will encircle the origin given by the formula

$$N = Z - P \tag{30}$$

where

$N \equiv$ number of encirclements of the origin made by the path Γ_F,
$Z \equiv$ number of zeros of $F(s)$ encircled by the path Γ_s,
$P \equiv$ number of poles of $F(s)$ encircled by the path Γ_s.

The values of N can be positive, zero, or negative depending upon the number of zeros and the number of poles of $F(s)$ encircled by Γ_s.

1. $N > 0$ (or $Z > P$). The path Γ_s encircles more zeros than poles of $F(s)$ in either the clockwise or counterclockwise direction, and Γ_F will encircle the origin of the $F(s)$-plane N times in the same direction as that of Γ_s.
2. $N = 0$ (or $Z = P$). The path Γ_s encircles an equal number of poles and zeros, or no poles and zeros of $F(s)$ in Γ_s, and Γ_F will not encircle the origin in the $F(s)$-plane.
3. $N < 0$ (or $Z < P$). This is similar to the case $N > 0$ but Γ_F encircles the origin of the $F(s)$-plane in the opposite direction as that of Γ_s.

In this analysis, for convenience, the origin of the $F(s)$-plane is selected to be the critical point from which the value of N is determined. However, it is possible to designate other points in the complex plane as the critical point depending on the requirement of the application. In the case of the Nyquist criterion, the critical point is the -1 on the real axis of the $F(s)$-plane.

CASE 2. Another way of presenting the principal argument may be to begin with the derivation of the relationship between the open-loop and closed-loop poles or zeros viewed from the characteristic equation. Let's take a closed-loop control system with single input and single output as shown in

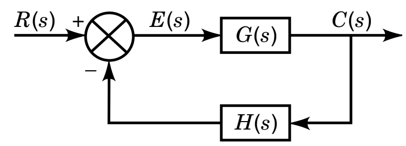

Figure 4. Block diagram of a closed-loop system. A closed-loop system has a forward path transfer function $G(s)$ and a feedback path transfer function $H(s)$. The relation between the input and output can be expressed in terms of these two terms in the form of a system transfer function to be used in analysis and design of systems.

Fig. 4. Writing the transfer functions in Laplace transforms as

$$G(s) = \frac{N_{\mathrm{G}}(s)}{D_{\mathrm{G}}(s)} \tag{31}$$

and

$$H(s) = \frac{N_{\mathrm{H}}(s)}{D_{\mathrm{H}}(s)} \tag{32}$$

then

$$G(s)H(s) = \frac{N_{\mathrm{G}}(s)N_{\mathrm{H}}(s)}{D_{\mathrm{G}}(s)D_{\mathrm{H}}(s)} \tag{33}$$

Hence, the characteristic equation

$$1 + G(s)H(s) = \frac{D_{\mathrm{G}}(s)D_{\mathrm{H}}(s) + N_{\mathrm{G}}(s)N_{\mathrm{H}}(s)}{D_{\mathrm{G}}(s)D_{\mathrm{H}}(s)} \tag{34}$$

and the closed-loop transfer function

$$M(s) = \frac{G(s)}{1 + G(s)H(s)} = \frac{N_{\mathrm{G}}(s)D_{\mathrm{H}}(s)}{D_{\mathrm{G}}(s)D_{\mathrm{H}}(s) + N_{\mathrm{G}}(s)N_{\mathrm{H}}(s)} \tag{35}$$

From Eqs. (31) to (34), it can be observed that the poles of $1 + G(s)H(s)$ are the same as the poles of $G(s)H(s)$. But more importantly, the zeros of the function $1 + G(s)H(s)$ are the same as the poles of $M(s)$ of the closed-loop system. Although simple, this observation bears particular importance in the understanding of the Nyquist criterion.

Assume that $\Delta(s)$ equals $1 + G(s)H(s)$ and has poles and zeros in the s-plane, as shown in Fig. 5. As any point s_1 of contour Γ_s is substituted in $\Delta(s)$, it maps to a point on contour Γ_Δ.

For the purposes of illustration, assume that Γ_s encloses a pole and two zeros. Also, two poles and a zero lie outside or are unbounded by the contour. As the point s_1 moves around the contour in a chosen clockwise direction, each of the pole and zero vectors connected to that point trace angles. Take the general equation

$$\Delta(s) = \frac{(s+z_1)(s+z_2)\cdots(s+z_Z)}{(s+p_1)(s+p_2)\cdots(s+p_P)},\ m, n \in \{1, 2, \ldots\} \tag{36}$$

This is equivalent to

$$\Delta(s) = |\Delta(s)|\angle\Delta(s) \tag{37}$$

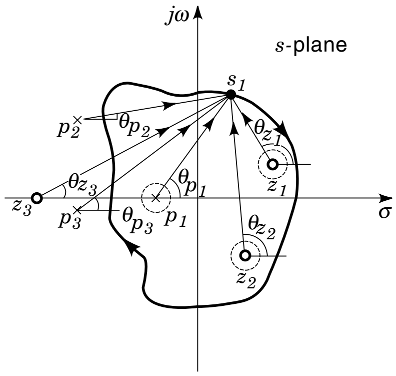

Figure 5. Angles traced by poles and zeros. As the point s_1 traverses around the closed contour, each pole and zero trace 360°. The angles traced by the poles and zeros outside the contour go to a minimum and then to a maximum and then back to the original value tracing the same path in the opposite direction. The net angle traced becomes zero.

where

$$|\Delta(s)| = \frac{|s+z_1||s+z_2|\cdots|s+z_Z|}{|s+p_1||s+p_2|\cdots|s+p_P|} \tag{38}$$

and $|s + z_1| \cdots |s + z_z|$, $|s + p_1| \cdots |s + p_P|$ are length of vectors and the angles are

$$\angle\Delta(s) = \angle(s+z_1) + \cdots + \angle(s+z_Z) - \angle(s+p_1) - \cdots - \angle(s+p_P) \tag{39}$$

From Fig. 5, we can deduce that as the point s_1 traverses around the contour once, the poles and zeros encircled by the contour Γ_s go through a complete rotation, each tracing an angle of 2π radians. On the other hand, the poles or zeros that lie outside of Γ_s undergo a net angular change of 0 radians. Because of the positive and negative signs of the Eq. (39), the net number of rotation is equal to the difference between the number of zeros and poles lying inside contour Γ_s.

As a matter of interest, similar conclusions can be drawn from the Eqs. (27)–(29), taking $\Delta(s)$ analogous to $F(s)$

$$\frac{1}{2\pi j}\int_\gamma f(s)\,ds = \frac{1}{2\pi j}\int_\gamma \frac{F'(s)\,ds}{F(s)} = \frac{1}{2\pi j}\int_\gamma d[\ln F(s)] = Z - P \tag{40}$$

and writing this equation as

$$\frac{1}{2\pi j}\int_\gamma d[\ln F(s)] = \frac{1}{2\pi j}\int_\gamma d[\ln|F(s)|] + j\arg[\ln F(s)] = Z - P \tag{41}$$

For a closed contour, the first part of the integration, $d[\ln F(s)]$ will be zero, but the second term is the 2π times the net encirclement of the origin.

As discussed earlier, the poles of characteristics equation $\Delta(s) = 1 + G(s)H(s)$ are also the poles of $G(s)H(s)$. Because the denominator $D_G(s)D_H(s)$ of Eq. (34) is much simpler than the numerator $D_G(s)D_H(s) + N_G(s)N_H(s)$, the poles of the equation can be determined relatively easily. Also, the zeros enclosed by Γ_s are the zeros of $\Delta(s)$, and they are the unknown poles of the closed-loop system. Therefore, P equals the number of enclosed open-loop poles, and Z equals the number of enclosed closed-loop poles. Thus, $Z = N + P$ indicates that the number of closed-loop poles inside the contour equals the number of open-loop poles of $G(s)H(s)$ inside the contour, hence the number of clockwise rotations of the mapping about the origin.

If the contour in the s-plane includes the entire right half plane (RHP), as illustrated in Fig. 6, the number of closed-loop poles enclosed by the contour determines the stability of the system. Because it is possible to count the number of open-loop poles P (usually by inspection) inside the bounding contour in the RHP, the number of enclosures of the origin N indicates the existence of the closed-loop poles on the RHP.

The poles and zeros of $G(s)H(s)$ are usually known and if the mapping function is taken to be $\Delta(s)$ equals $G(s)H(s)$ instead of $1 + G(s)H(s)$, the resulting contour is the same except that it is translated one unit to the left. Thus the number of rotations about the point -1 in the $G(s)H(s)$-plane may be

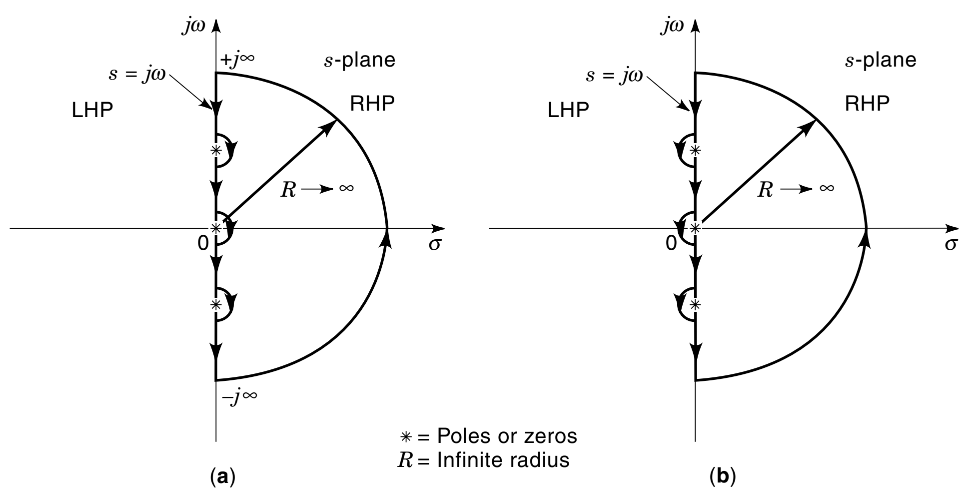

Figure 6. The Nyquist path. The Nyquist path covers the entire right half of the s-plane, avoiding poles and zeros located on the imaginary axis, as shown in (a). The existence of closed-loop poles on the RHP indicates unstable conditions. In (b), poles and zeros on the imaginary axis are included in the Nyquist path.

Table 3. The Possible Outcomes of Number of Poles and Zeros on the s-Plane

		$\Delta(s)$-Plane Locus	
$N = Z - P$	Direction of s-Plane Locus	Encircles of the Origin	Direction of Encirclement
$N > 0$	Clockwise	N	Clockwise
	Counterclockwise	N	Counterclockwise
$N < 0$	Clockwise	N	Counterclockwise
	Counterclockwise	N	Clockwise
$N = 0$	Clockwise	0	None
	Counterclockwise	0	None

counted instead of the origin. Hence the Nyquist criterion may be stated as follows:

If a contour Γ_s encircles the entire right half plane, the number of closed-loop poles Z in the RHP of the s-plane can be determined by the number of open-loop poles P in the RHP and the number of revolutions N of the resulting contour around the point -1 in the $G(s)H(s)$-plane.

This mapping is called the Nyquist diagram or the Nyquist plot of $G(s)H(s)$. A summary of all the possible outcomes of the principle argument is given in Table 3.

The method discussed is also called the frequency response technique. Around contour Γ_s in the RHP the mapping of the points on the $j\omega$-axis through $G(s)H(s)$ is the same as using the substitution s equals $j\omega$, hence forming the frequency response $G(j\omega)H(j\omega)$. Thus the frequency response over the positive $j\omega$-axis from $\omega = 0^+$ to $\omega = \infty$ are used to determine the Nyquist plot. That is, instead of tracing the entire RHP, it is sufficient to use just a part of the contour Γ_s. The Nyquist criteria could have easily been built upon the tracing of the left half plane (LHP); however, the solution is a relative one.

In Fig. 6(a,b) it can be seen that the contour Γ_s encircles the entire right half of the s-plane in the counterclockwise sense. The reason for this is because in mathematics counterclockwise is traditionally defined to be positive.

Observe on Fig. 6(a) that small semicircles are drawn along the $j\omega$-axis because the Nyquist path must not go through any of the poles or zeros of $\Delta(s)$. If any poles or zeros fall on the $j\omega$-axis, then the path Γ_s should detour around these points. Only the poles or zeros that lie in the RHP of the s-plane need to be encircled by the Nyquist path.

From the principal argument the stability of a closed-loop system can be determined, after the Nyquist path is specified, by plotting the function $\Delta(s) = 1 + F(s)$ where $F(s)$ equals to $G(s)H(s)$ and the s variable is chosen along the Nyquist path. The behavior of the $\Delta(s)$ plot, or the new path Γ_Δ, is referred to as the Nyquist plot of $\Delta(s)$, with respect to the critical point, the origin. Because the function $F(s)$ is usually known and is much simpler to construct, the Nyquist plot of $F(s)$ arrives at the same conclusion about the stability of a closed-loop system. This is simply done by shifting the critical point from the origin to the point $(-1, j0)$ on $F(s)$-plane. This is because the origin of the $[1 + F(s)]$-plane corresponds to $(-1, j0)$ of the $F(s)$-plane.

With the new critical point at $(-1, j0)$, it will be necessary to define two sets of N, Z, and P as follows:

$N_0 \equiv$ number of encirclement around the origin made by $F(s)$.

$Z_0 \equiv$ number of zeros $F(s)$ encircled by the Nyquist path in the right half of the s-plane.

$P_0 \equiv$ number of poles $F(s)$ encircled by the Nyquist path in the right half of the s-plane.

$N_{-1} \equiv$ number of encirclement around the point $(-1, j0)$ made by $F(s)$.

$Z_{-1} \equiv$ number of zeros $1 + F(s)$ encircled by the Nyquist path in the right half of the s-plane.

$P_{-1} \equiv$ number of poles $1 + F(s)$ encircled by the Nyquist path in the right half of the s-plane.

When the closed-loop system has only a single feedback system having the loop transfer function of $G(s)H(s)$, then $F(s) = G(s)H(s)$. Now it becomes clear that

$$P_0 = P_{-1} \tag{42}$$

because $F(s)$ and $1 + F(s)$ always have the same poles. The result is similar to the one derived earlier in the discussion. The other stability requirements are that for the closed-loop stability

$$Z_{-1} = 0 \tag{43}$$

and for the open-loop stability

$$P_0 = 0 \tag{44}$$

Z_{-1} must be zero because of the zeros of $1 + G(s)H(s)$ and the poles of the closed-loop transfer function $M(s)$ as discussed earlier; any poles that lie in the left-hand plane causes system instability. For the case of the open-loop stability, the P_0 is the number of poles of $F(s)$ encircled by the Nyquist path in the right half of the s-plane and must be zero for stability conditions.

The discussions presented so far may be summarized as follows:

1. For a given feedback control system, the closed-loop transfer function is given by Eq. (35), and the denominator function represent the closed-loop transfer function as given by Eq. (34), which is equal to $\Delta(s)$. The Nyquist path is defined in accordance with the pole and zero properties of $F(s)$ on the $j\omega$-axis.
2. The Nyquist plot of $F(s)$ is constructed in the $[G(s)H(s)]$-plane.
3. The value of N_0 is determined by observing the behavior of the Nyquist plot of $F(s)$ with respect to the origin. Similarly the value N_1 is determined with respect to the point $(-1, j0)$.
4. After determining N_0 and N_1, the value of P_0, if not already known, can be obtained from

$$N_0 = Z_0 - P_0 \tag{45}$$

if Z_0 is known. With P_0 determined, P_{-1} is also known via Eq. (42), and Z_{-1} can then be calculated with

$$N_{-1} = Z_{-1} - P_{-1} \tag{46}$$

and from Eq. (43) Eq. (46) simplifies to

$$N_{-1} = -P_{-1} \tag{47}$$

Now the Nyquist criterion may also be stated in the following manner.

For a closed-loop system to be stable, the Nyquist plot of $F(s)$ must encircle the critical point $(-1, j0)$ as many times as the number of poles of $F(s)$ that lie in the right half of the s-plane. The encirclements, if any, must be in the clockwise direction when the Nyquist path is defined in the counterclockwise sense.

SIMPLIFIED NYQUIST CRITERIA

The Nyquist criterion discussed previously requires the construction of the Nyquist plot corresponding to the Nyquist path in the s-plane. Complication arises when the $F(s)$-plane poles or zeros that lie on the $j\omega$-axis, as in Fig. 6(a), indicate small indentations around these points. As pointed out in Kuo (1), Yeung and Lai came up with a simplified version of the Nyquist criterion for closed-loop systems that requires only the Nyquist plot corresponding to the positive $j\omega$-axis of the s-plane.

In the development of this simplified criterion, two paths as shown in Fig. 6(a,b) are considered. The first path Γ_{s1} encircles the entire right half of the s-plane excluding all the poles and zeros that lie on the $j\omega$-axis. The second path Γ_{s2} encircles the excluded poles and zeros that may exist. Now new quantities may be defined as follows:

Z_{-1} ≡ number of zeros of $1 + F(s)$ in the right half of the s-plane.

P_{-1} ≡ number of poles of $1 + F(s)$ in the right half of the s-plane, and is equal to P_0, which are poles of $F(s)$ in the right half of the s-plane.

P_ω ≡ number of poles $F(s)$ or $1 + F(s)$ that lie on the $j\omega$-axis including the origin.

$N_{-1,1}$ ≡ number of times the point $(-1, j0)$ of the $F(s)$-plane is encircled by the Nyquist plot of $F(s)$ corresponding to Γ_{s1}.

$N_{-1,2}$ ≡ number of times the point $(-1, j0)$ of the $F(s)$-plane is encircled by the Nyquist plot of $F(s)$ corresponding to Γ_{s2}.

According to the Nyquist criterion,

$$N_{-1,1} = Z_{-1} - P_{-1} \tag{48}$$

and

$$N_{-1,2} = Z_{-1} - P_\omega - P_{-1} \tag{49}$$

where Eq. (49) includes the number of poles of $F(s)$ or $1 + F(s)$ that lie on the $j\omega$-axis.

Also, two quantities θ_1 and θ_2 will be defined to represent the total angle traversed by the Nyquist plots of $F(s)$ with respect to the point $(-1, j0)$, corresponding to the points Γ_{s1} and Γ_{s1}, respectively. Thus the two new quantities may be written as

$$\theta_1 = 2\pi \cdot N_{-1,1} = 2\pi(Z_{-1} - P_{-1}) \tag{50}$$

and

$$\theta_2 = 2\pi \cdot N_{-1,2} = 2\pi(Z_{-1} - P_\omega - P_{-1}) \tag{51}$$

To analyze the Nyquist path, it is best to consider the path having three major sections. The first section is the portion from s equals to $-j\infty$ to $+j\infty$ along the semicircle having an infinite radius, the second portion is the path along the $j\omega$-axis excluding the small indentations, and the final sections include the small indentations. Because the Nyquist plot is symmetrical at $j\omega = 0$, the angles traversed are identical for positive and negative values of ω. Hence

$$\theta_1 = 2\theta_{11} + \theta_{12} + \theta_{13} \tag{52}$$

and

$$\theta_2 = 2\theta_{11} - \theta_{12} + \theta_{13} \tag{53}$$

where

θ_{11} ≡ angle traversed by the Nyquist plot of $F(s)$ with respect to $(-1, j0)$ corresponding to the positive or negative side of the $j\omega$-axis, excluding the small indentations and the factor of two emmerging in Eq. (52) and Eq. (53).

θ_{12} ≡ angle traversed by the Nyquist plot of $F(s)$ with respect to $(-1, j0)$ corresponding to the small indentations along the $j\omega$-axis Γ_{s1}. Also, with the direction of the small indentations of Γ_{s2} different from that of its counterpart, the negative sign emerges in Eq. (53).

θ_{13} ≡ angle traversed by the Nyquist plot of $F(s)$ with respect to $(-1, j0)$ corresponding to the semicircle with infinite radius on the Nyquist path.

Generally, for a physical realizable transfer function, the number of poles cannot exceed the number of zeros of $F(s)$. Therefore, the Nyquist plot of $F(s)$ corresponding to the infinite semicircle must be a point on the real axis or a trajectory around the origin of the $F(s)$-plane. The angle θ_{13} traversed by the phasor from the point at $(-1, j0)$ to the Nyquist plot along the semicircle is always zero.

Combining Eq. (52) and Eq. (53) yields

$$\theta_1 + \theta_2 = 4\theta_{11} \tag{54}$$

since θ_{13} is zero,

$$\theta_1 + \theta_2 = 2\pi(2Z_{-1} - P_\omega - 2P_{-1}) \tag{55}$$

hence

$$\theta_{11} = \pi(Z_{-1} - 0.5P_\omega - P_{-1}) \tag{56}$$

Equation (56) means that the net angle traversed by the phasor from the $(-1, j0)$ point to the $F(s)$ Nyquist plot corresponding to the positive $j\omega$-axis of the s-plane excluding any of the small indentations, that is

The number of zeros of $1 + F(s)$ in the right half of the s-plane minus the sum of half the poles on the $j\omega$-axis and the number of poles of $F(s)$ in the right half of the s-plane multiplied by π radians.

This means that the Nyquist plot can be constructed corresponding to $s = 0$ to $s = j\infty$ portion of the Nyquist path. For an unstable closed-loop system, the number of roots of the characteristic equation that fall in the right half of the s-plane can be determined via Eq. (55).

As mentioned earlier, a closed-loop system is stable only if Z_{-1} is equal to zero.

Hence,

$$\theta_{11} = -\pi(0.5P_\omega + P_{-1}) \tag{57}$$

This indicates that for closed-loop system stability the phase traversed by the Nyquist plot of $F(s)$ where s varies from zero to $j\infty$ with respect to $(-1, j0)$ point cannot be positive because P_ω and P_{-1} cannot be negative.

NYQUIST DIAGRAMS

The Nyquist analysis is based on the assumption that the control systems are linear; hence, the dynamic performances are described by a set of linear differential equations. Because of the nature of feedback control systems, the degrees of numerator of the loop transfer function $F(s) = G(s)H(s)$ is always less than or equal to the degree of the denominator. All the Nyquist diagrams presented here are based on these two assumptions.

As explained previously, when plotting Nyquist diagrams, it is sufficient to assign values for the complex variable s on the $j\omega$-axis avoiding possible poles and zeros on the imaginary axis. The frequency response of $G(s)H(s)$ can be determined by substituting $s = j\omega$ and by finding the imaginary and complex components of $G(j\omega)H(j\omega)$. Alternatively, $G(j\omega)H(j\omega)$ can be written in polar form, and magnitudes and angles are determined for plotting on a polar graph paper. These techniques will be illustrated in the following examples.

Example 1. Plot the Nyquist diagram of a closed-loop control system as in Fig. 4 with a loop transfer function

$$G(s)H(s) = \frac{K(s+1)(s+20)}{250s(s+0.2)(s+0.4)} \tag{58}$$

SOLUTION. To obtain the Nyquist plot, rearrange this equation, substitute $s = j\omega$, and assign the nominal value, $K = 1$,

$$G(j\omega)H(j\omega) = \frac{(1+j\omega)(1+0.05j\omega)}{j\omega(1+5j\omega)(1+2.5j\omega)} \tag{59}$$

Find magnitudes and angles in terms of variable ω as

$$|G(j\omega)H(j\omega)| = \frac{\sqrt{1+\omega^2}\sqrt{1+(0.05\omega)^2}}{\omega\sqrt{1+(5\omega)^2}\sqrt{1+(2.5\omega)^2}} \tag{60}$$

and

$$\langle G(j\omega)H(j\omega) = \tan^{-1}\omega + \tan^{-1}0.05\omega - 90^\circ - \tan^{-1}5\omega - \tan^{-1}2.5\omega \tag{61}$$

Now Nyquist contour may be applied by substituting values of ω from zero to ∞ on the positive part of the imaginary axis on the s-plane. By avoiding the pole located on the origin and substituting small positive values on the positive and negative side of zero, it is possible to observe the following features:

$$\begin{array}{llll}
\omega \to 0^+ & |G(j\omega)H(j\omega)| \to \infty & \text{and} & \langle G(j\omega)H(j\omega) \to -90^\circ \\
\omega \to \infty & |G(j\omega)H(j\omega)| \to 0 & \text{and} & \langle G(j\omega)H(j\omega) \to -90^\circ \\
\omega \to -\infty & |G(j\omega)H(j\omega)| \to 0 & \text{and} & \langle G(j\omega)H(j\omega) \to 90^\circ \\
\omega \to 0^- & |G(j\omega)H(j\omega)| \to \infty & \text{and} & \langle G(j\omega)H(j\omega) \to 90^\circ
\end{array}$$

These features indicate that for a clockwise rotation in the s-plane covering the entire right half plane, the graph starts from the infinity on the imaginary axis (in either the fourth or the third quadrant) and approaches zero again from the -90° for $0^+ \leq \omega \leq +\infty$. Similarly, the graph starts from 0 at an angle $+90^\circ$ and approaches to infinity with the same angle $+90^\circ$ for $-\infty \leq \omega \leq 0^-$.

By substituting intermediate values for ω, the results in Table 4 may be obtained. We can see that the plot goes numerically above -180° between $\omega = 0.4$ rad/s and $\omega = 0.5$ rad/s. It also falls back to be numerically less than -180° after $\omega = 2.0$ rad/s. This means that it crosses the real axis on the negative side of the s-plane twice.

At this point, a polar plot graph paper may be used to sketch the curve outlined in Table 4. Or a second table, which shows the real and imaginary components of $G(j\omega)H(j\omega)$, may be made from Table 4 by using the relation $[Re^{j\theta} = R\cos\theta + Rj\sin\theta]$ as in Table 5. An alternative approach to calculate the real and imaginary components of $G(j\omega)H(j\omega)$ is introduced in the next example.

The Nyquist plot of Table 5 is shown in Fig. 7. It is worth noting that Table 5 could also have been drawn by rearranging Eq. (59) as real and imaginary components. This is a long procedure, but using it allows us to calculate the exact values

Table 4. Calculation of Magnitudes and Angles of $G(j\omega)H(j\omega)$

ω rad/s	0^+	0.1	0.4	0.5	1.0	2.0	4.0	∞
$\langle G(j\omega)H(j\omega)$	-90°	-124°	-175°	-181.6°	-188.9°	-182.8°	-174.2°	-90°
$\lvert G(j\omega)H(j\omega)\rvert$	∞	9.0	0.79	0.55	0.1	0.021	0.005	0

Table 5. Calculation of Real and Imaginary Components of $G(j\omega)H(j\omega)$

ω rad/s	0^+	0.1	0.4	0.5	1.0	2.0	4.0	∞
$\langle G(j\omega)H(j\omega)$	0	−5.03	−0.787	−0.55	−0.099	−0.021	−0.005	0
$\lvert G(j\omega)H(j\omega)\rvert$	$-\infty$	−7.46	−0.069	+0.018	+0.015	0.0001	−0.0005	0

of the gain and phase margins if an approximate estimation from the plot is not permissible.

By using Eqs. (60) and (61), it can be shown that $|G(j\omega)H(j\omega)| = 0.51$ and $\langle G(j\omega)H(j\omega) = -180°$ for $\omega = 0.481$ rad/s. Remember that value is for the nominal K where $K = 1$. For other values of K, Eq. (60) should have been written as

$$|G(j\omega)H(j\omega)| = K\frac{\sqrt{1+\omega^2}\sqrt{1+(0.05\omega)^2}}{\omega\sqrt{1+(5\omega)^2}\sqrt{1+(2.5\omega)^2}} \qquad (62)$$

Therefore, the curve will pass to the left-hand side of the -1 point on the real axis if $|G(j\omega)H(j\omega)| \geq 1$. Hence $K \times 0.51 = 1$ gives the value $K = 1.9$ in which the plot will encircle the $-1 + j0$ point in the clockwise direction, thus leading to instability. This indicates that there is a zero of the characteristic equation on the RHP of the s-plane or a RHP pole of the closed loop transfer function, hence the instability.

From the preceding analysis, it is known that the plot crosses the real axis again at a frequency slightly greater than $\omega = 2.0$ rad/s. By substituting a value $\omega = 2.57$ rad/s, from Eqs. (60) and (61), it can be shown that $|G(j\omega)H(j\omega)| = 0.011$ and $\langle G(j\omega)H(j\omega) = -180°$. As explained previously the corresponding value of $K = 91$ obtained from $K \times 0.011 = 1$. For $K = 91$ and above, the system becomes stable again.

Example 2. Plot the Nyquist diagram of unity feedback control system which has a forward gain transfer function

$$G(s) = \frac{(1+2s)}{s(s-1)} \qquad (63)$$

SOLUTION. In this example, because $G(s)$ has a pole on the RHP of the s-plane, it is open-loop unstable. As before, in order to obtain the Nyquist plot, substitute $s = j\omega$ in the loop

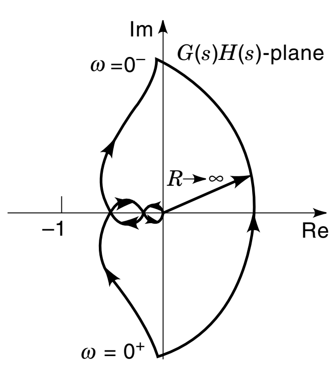

Figure 7. The Nyquist plot of Example 1. As the Nyquist path on the s-plane traverses in the clockwise direction, the corresponding path of $G(s)H(s)$ traverses in the same direction on the $G(s)H(s)$-plane. Because the Nyquist plot does not encircle the critical $-1 + j0$ point, this system is stable.

transfer function $G(j\omega)H(j\omega)$ and assign the nominal value ($K = 1$)

$$G(j\omega) = \frac{(1+j2\omega)}{j\omega(j\omega-1)} \qquad (64)$$

Find magnitudes and angles in terms of variable ω as

$$G(j\omega) = \frac{\sqrt{1+4\omega^2}}{\omega\sqrt{1+\omega^2}} \qquad (65)$$

and

$$\langle G(j\omega) = \tan^{-1}2\omega - 90° - \tan^{-1}\omega/(-1) \qquad (66)$$

Let's consider the nominal value of K. Observing the extreme values for ω in the clockwise direction and starting from 0^+ gives

$$\omega \to 0^+\ |G(j\omega)H(j\omega)| \to \infty \text{ and } \langle G(j\omega)H(j\omega) \to -270° \text{ or } 90°$$
$$\omega \to \infty\ |G(j\omega)H(j\omega)| \to 0 \text{ and } \langle G(j\omega)H(j\omega) \to +270° \text{or } -90°$$
$$\omega \to -\infty\ |G(j\omega)H(j\omega)| \to 0 \text{ and } \langle G(j\omega)H(j\omega) \to +90°$$
$$\omega \to 0^-\ |G(j\omega)H(j\omega)| \to \infty \text{ and } \langle G(j\omega)H(j\omega) \to -90°$$

It is important to highlight the angle equation $-\tan^{-1}\omega/(-1)$ because the negative sign in the denominator indicates what quadrant the angle is for varying ω. These features indicate that for a clockwise rotation of a contour in the s-plane covering the entire right half plane, the graph starts from the infinity on the imaginary axis from the first or second quadrant and approaches zero from the $-90°$ for $0^+ \leq \omega \leq +\infty$. Similarly, the graph starts from 0 at an angle $+90°$ and approaches infinity with the same angle $-90°$ for $-\infty \leq \omega \leq 0^-$.

Table 6 may be obtained by substituting values for ω, but this time for only $0^+ \leq \omega \leq +\infty$. The Nyquist plot is given in Fig. 8. The open-loop transfer function has one pole on the RHP, and therefore $P = 1$. In order for this system to be stable, N must be equal to -1, that is, one counterclockwise encirclement of the $-1 + j0$ point. As shown in Fig. 8, the rotation of the curve is in a counterclockwise direction, and it encircles the origin once; hence, the system is stable.

From Table 6, we can see that the graph crosses the real axis between $\omega = 0.6$ rad/s and $\omega = 1.0$ rad/s. By guessing and by using the trial-and-error method, this crossover frequency may be determined as $\omega = 0.7$ rad/s. At this frequency, the magnitude is 2.0. As in the case of Example 1, the critical value of the gain can be found from $K \times 2.0 = 1$ to be $K = 0.5$.

An alternative mathematical approach can be employed to find the real and imaginary components of the loop transfer function as

$$G(j\omega) = \frac{(1+j2\omega)}{j\omega(j\omega-1)} = \frac{(1+j2\omega)}{\omega^2 - j\omega} = \frac{-3\omega^2 + j(\omega - 2\omega^3)}{\omega^2+\omega^4} \qquad (67)$$

Table 6. Calculation of Magnitudes and Angles of $G(j\omega)H(j\omega)$

ω rad/s	0^+	0.1	0.4	0.6	1.0	4.0	10.0	∞
$\langle G(j\omega)H(j\omega)$	90°	107.0°	150.5°	171.16°	198.4°	248.8°	261.4°	270°
$\|G(j\omega)H(j\omega)\|$	∞	10.15	2.97	2.23	1.58	0.49	0.02	0

The graph crosses the real axis when the imaginary part is equal to zero, that is $(\omega - 2\omega^3) = 0$ or $\omega = 1/\sqrt{2}$ rad/s. The intersection can be calculated by substituting the value of ω in the real component of $G(j\omega)H(j\omega)$ as

$$\frac{-3\omega^2}{\omega^2 + \omega^4} = \frac{-3 \times (1/2)}{(1/2) + (1/4)} = -2 \tag{68}$$

There are many examples of Nyquist plots in control engineering books (1–5). Figure 9 illustrates typical Nyquist plots of some of the selected control systems.

Stability Margins

In the preceding examples, we have demonstrated that the Nyquist plots of the loop transfer function, $G(j\omega)H(j\omega)$, depends on the values of K. This is illustrated in Fig. 10. As the K is increased or decreased, as the case may be, at a certain value the locus passes through $-1 + j0$ point. At this point, the system exhibits sustained oscillations. As K increases further, the system becomes unstable.

Generally, oscillations increase as the locus of $G(j\omega)H(j\omega)$ gets closer to the $-1 + j0$ point. The closeness of the locus to the critical point is measured by the stability margins usually expressed in the form of phase and gain margins, as illustrated in Fig. 10. These margins indicate relative stability and hence help the design of control systems to achieve desired responses. The gain and phase margins may be defined as follows.

Gain margin is the amount of gain that can be allowed to increase before the closed loop system becomes unstable.

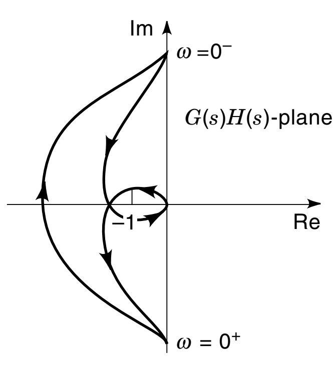

Figure 8. The Nyquist plot of Example 2. The open-loop transfer function of this control system has a pole on the RHP; hence, the system is open-loop unstable. However, the Nyquist plot encircles the critical $-1 + j0$ point once in the counterclockwise direction, indicating that there are no closed-loop poles on the RHP. Therefore, the system is stable.

Phase margin is the angle, in degrees, by which the locus must be rotated in order that gain crossover point passes through $-1 + j0$.

The gain margin is measured in decibels and expressed in phase-crossover frequency as

$$\text{GM} = 20\log_{10} = \frac{1}{|G(j\omega_c)H(j\omega_c)|}\ \text{dB} \tag{69}$$

When the loop transfer function $G(j\omega)H(j\omega)$ passes through the $-1 + j0$ point, the gain margin is 0 dB. The negative or positive value of gain margin depends on the number of poles and zeros of $G(j\omega)H(j\omega)$ on the RHP. If the stability is evaluated when the locus crosses the real axis on the right of the $-1 + j0$ point (Example 1), the gain margin is positive. If the stability is evaluated on the left of the $-1 + j0$ point (Example 2), then the gain margin is negative.

The phase margin can be determined by calculating the gain-crossover frequency when the gain of the $G(j\omega)H(j\omega) = 1$ and by evaluating the phase angle of the system at that frequency. That is

$$\text{PM} = \langle G(j\omega)H(j\omega) - 180^\circ \tag{70}$$

As in the case of gain margin, the sign of phase margin is relative to stability condition and the shape of the locus. In Example 1, a negative value for the PM indicates unstable condition whereas in Example 2, the negative value implies stability.

Effects of Adding Poles and Zeros

Control systems are often designed by introducing additional poles and zeros to the system. Extra poles and zeros in the system change the shape of the Nyquist diagrams and alter phase and gain margins. The influences of additional poles and zeros on the Nyquist locus can be evaluated by comparing the loci of different systems given in Fig. 9. Some observations may be made as follows.

The mathematical difference between parts (a) and (b) in Fig. 9 is the additional pole. In this case, the Nyquist locus is shifted by $-90°$ as $\omega \to \infty$, occupying quadrants 3 and 4 instead of quadrant 4 only. Adding an extra pole introduces further $-90°$, and the locus occupies three quadrants. In this case, the risk of instability exists because the possibility of encirclement of $-1 + j0$ is introduced.

In Fig. 9, the effect of adding a pole at $s = 0$ can be seen by observing parts (a) and (d) or (b) and (e). In both cases, Nyquist loci are rotated by $-90°$ for all frequencies. Adding a finite pole increases the risk of instability.

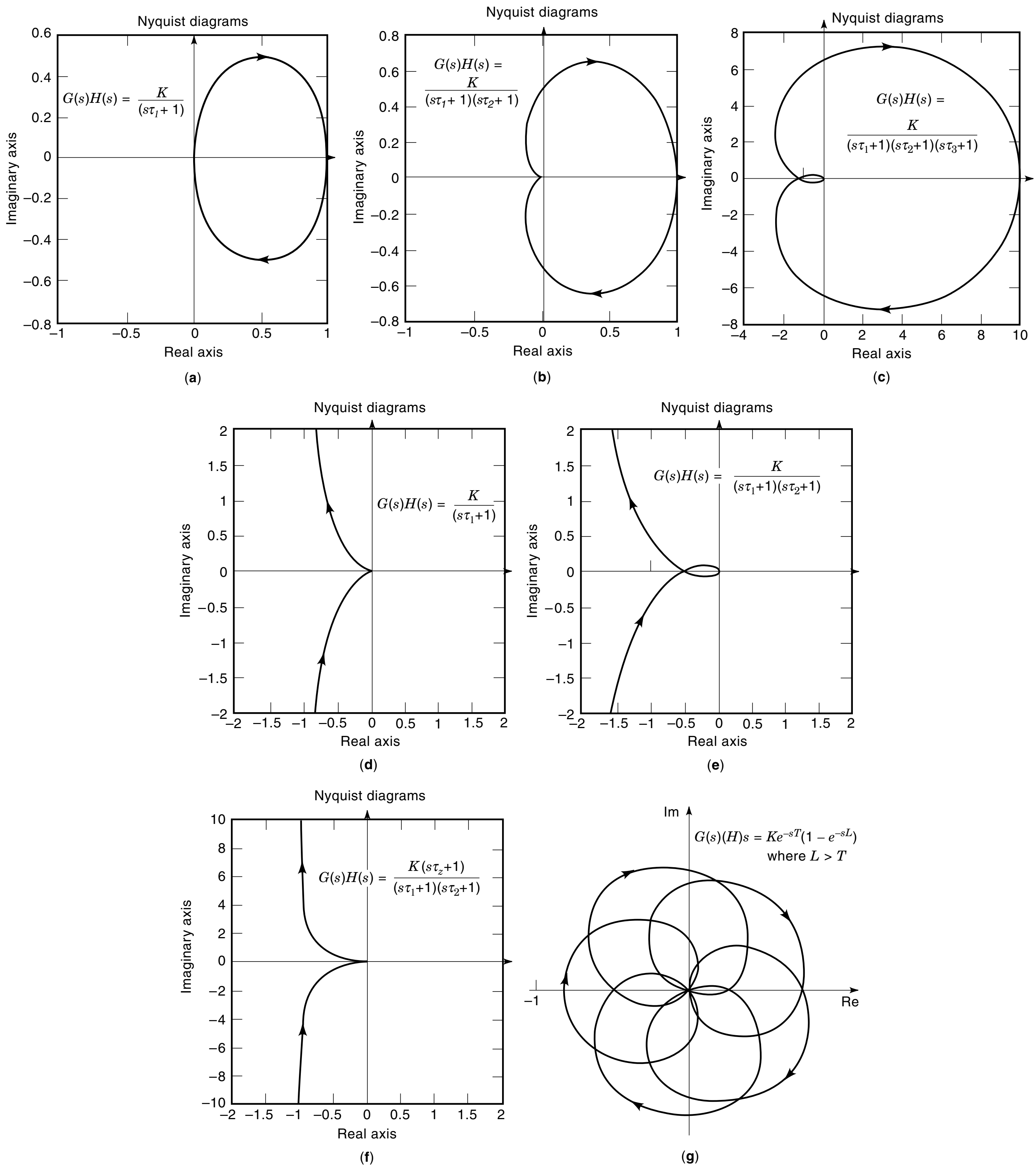

Figure 9. The Nyquist plots of selected control systems.

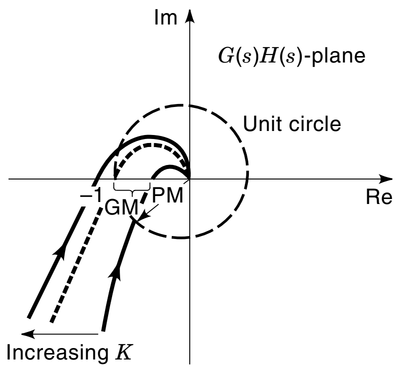

Figure 10. Phase and gain margins. The closeness of the locus to the critical $-1 + j0$ point is measured by the margins. Good gain and phase margins are obtained for K_1. As K increases, both gain and phase margins become zero (for K_2) indicating critical stability. Further increase in K leads to unstable conditions. Note the changes in the gain and phase margins for varying K.

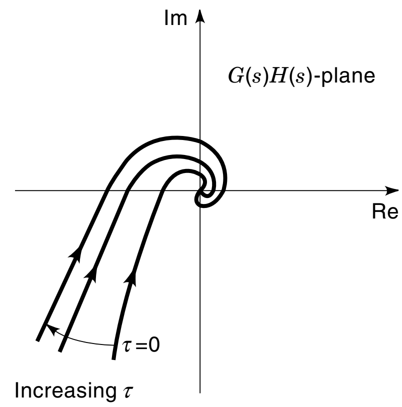

Figure 11. The effect of time delays. Pure time delays do not introduce any extra poles and zeros into the system. However, the magnitude is equal to unity for all frequencies, the phase ($= -\omega T$) affects the stability. For large values of time delay T, the system may be unstable.

The effect of adding a zero into the system can be seen in parts (c) and (f) in Fig. 9. In this case, the loop transfer function increases the phase of $G(s)H(s)$ by $+90°$ as $\omega \to \infty$. This result confirms the general knowledge that addition of a derivative control or a zero makes the system more stable.

EFFECTS OF TIME DELAYS

The Nyquist criterion can be utilized to evaluate the effects of time delays on the relative stability of feedback control systems. With the pure time delays, the loop transfer function may be written as

$$G(s)H(s) = e^{-sT} G_1(s)H_1(s) \tag{71}$$

where T is the time delay. The term e^{-sT} does not introduce any additional poles or zeros within the contour. However, it adds a phase shift to the frequency response without altering the magnitude of the curve. This is because

$$\begin{aligned} |G(j\omega)H(j\omega)| &= |e^{-j\omega T}||G_1(j\omega)H_1(j\omega)| \\ &= |\cos(\omega) - j\sin(\omega)||G_1(j\omega)H_1(j\omega)| \end{aligned} \tag{72}$$

The term containing the time delay is $|\cos(\omega) - j\ \sin(\omega)| = 1$, but the phase is $\tan^{-1}(-\sin\ \omega T/\cos\ \omega T) = -\omega T$. This shows that the phase grows increasingly negative in proportion to the frequency. A plot of the effect of time delay is given in Fig. 11. Because of the addition of the phase shift, the stability of the system is affected for large values of T.

NYQUIST STABILITY CRITERION FOR DIGITAL SYSTEMS

The Nyquist stability criterion can equally be applied to linear continuous-data discrete time systems to graphically determine the stability. Generally, the closed-loop transfer function of a single loop, single input and single output of a system may be written as

$$M(z) = \frac{C(z)}{R(z)} = \frac{G(z)}{1 + GH(z)} \tag{73}$$

where z is the z-transform defined as $z = e^{sT}$.

The stability of the system can be studied by investigating the zeros of the characteristic equation $1 + GH(z) = 0$. For the system to be stable, all the roots of the characteristic equation must be inside the unit circle in the z-plane. As in the continuous-time systems, the investigation of the Nyquist plot of $GH(z)$ with respect to critical point, $-1 + j0$, indicates the system stability. The general theory presented for continuous time control systems is applicable to discrete time systems with minor modifications. Here, an example will be given to illustrate the use of Nyquist in discrete time control systems.

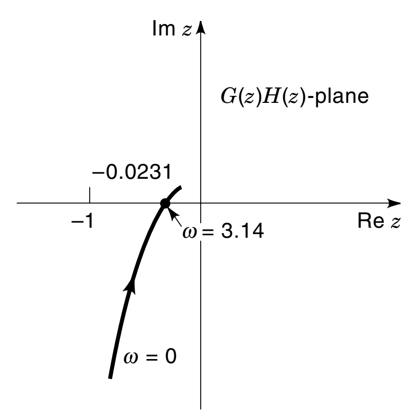

Figure 12. Nyquist plot of Example 3. The Nyquist path on the z-plane must have small indention at $z = 1$ on the unit circle. The Nyquist plot of path of $GH(z)$ in the $GH(z)$-plane intersects the negative real axis at -0.231 when $\omega = 3.14$ rad/s. For stability, the value of K must be less than 4.33.

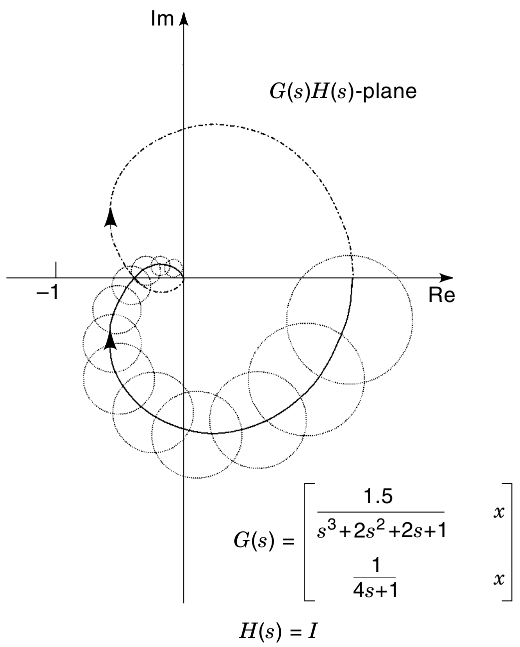

Figure 13. Examples of Nyquist plots of multivariable systems. The Nyquist plot for multivariable systems carries similar information as in the single-input–single-output systems. The number and the direction of encirclements of the critical $-1 + j0$ point conveys the message about the stability. But rigorous mathematical analysis is necessary because matrices are involved.

Example 3. Show the Nyquist plot of a discrete time system with transfer function of

$$GH(z) = \frac{0.632z}{(z-1)(z-0.368z)} \tag{74}$$

for a sampling period of $T = 1$ s.

SOLUTION. The loop transfer function $GH(z)$ does not have any poles outside the unit circle, but it has one pole on the unit circle. As in the case of s-plane zeros on the imaginary axis, the Nyquist path on the z-plane must have small indention at $z = 1$ on the unit circle. The Nyquist path, shown in Fig. 12, intersects the negative real axis of the $GH(z)$-plane at -0.231 when the value of $\omega = 3.14$ rad/s. The critical $-1 + j0$ point may be encircled if $0.231K = 1$, that is $K = 4.33$.

THE INVERSE NYQUIST AND NYQUIST PLOT FOR MULTIVARIABLE SYSTEMS

Inverse Nyquist is simply the reciprocal of the complex quantity in the Nyquist plot. They find applications particularly in multiple loop and multivariable systems where graphical analysis may be preferred.

The Nyquist stability criterion applied to inverse plots can be stated as a closed loop system stable, if the encirclement of the critical $-1 + j0$ point by the $1/G(s)H(s)$ is in the counterclockwise direction for a clockwise Nyquist path in the s-plane. As in the case of a normal Nyquist, the number of encirclements must equal the number of poles of $1/G(s)H(s)$ that lie in the right half of the s-plane.

Inverse Nyquist plots is particularly useful in the analysis of multi-input–multi-output control systems. In the multivariable feedback control systems, the relations between inputs and outputs may be expressed in matrix form as

$$\mathbf{C}(s) = [\mathbf{I} + \mathbf{KG}(s)\mathbf{H}(s)]^{-1}\mathbf{G}(s)\mathbf{KR}(s) \tag{75}$$

where $\mathbf{G}(s)$, $\mathbf{H}(s)$ and $\mathbf{K}$ are $n \times n$ matrices.

Similar to single-input–single-output systems, the output is exponentially stable iff det$[\mathbf{I} + \mathbf{KG}(s)\mathbf{H}(s)]^{-1}$ has no poles in the right half of the s-plane. The Nyquist diagrams can be obtained by appropriately considering the $\mathbf{K}$ values as $K =$ diag$\{k_1, k_2, \ldots, k_n\}$ and the $g_{ij}(s)$ elements of $\mathbf{G}(s)$. A typical example of a Nyquist diagram of a multivariable control systems is shown in Fig. 13. This example is also given by Westphal (5); interested readers could refer to that book for further details.

BIBLIOGRAPHY

1. B. J. Kuo, *Automatic Control Systems,* 6th ed., Englewood Cliffs, NJ: Prentice-Hall, 1991.
2. E. Kreyszing, *Advanced Engineering Mathematics,* 7th ed., New York: John Wiley, 1993.
3. W. S. Levine, *The Control Handbook,* Boca Raton, FL: CRC Press, 1996.
4. K. Ogata, *Modern Control Engineering,* 3rd ed., Upper Saddle River, NJ: Prentice-Hall, 1997.
5. L. C. Westphal, *Sourcebook of Control Systems Engineering,* Cambridge, UK: Chapman & Hall, 1995.

HALIT EREN
BERT WEI JUET WONG
Curtin University of Technology

NYQUIST STABILITY. See NYQUIST CRITERION, DIAGRAMS, AND STABILITY.