ANNEX G

Nyquist Criterion

Nyquist (1932) established the result which bears his name during his work on problems relating to the development of the US telephone system. It is a wonderful yet simple general result, a relation among the poles and zeros of a complex function, established by applying Cauchy's residue theorem. In Section G.5 we repeat a standard proof to clarify what is implied by use of the criterion in applications to control theory. We treat only linear single-input-single-output (SISO) systems in detail; at the end of this annex we make a few remarks on multi-input-multi-output (MIMO) systems.

There are many books devoted to feedback control, but surprisingly few contain proofs of Nyquist's Criterion. That's a pity because with a little effort one gains a great deal of understanding. Besides Bode's book (1945) volumes that I have found helpful include Franklin *et al.* (2002), McFarlane and Glover (1992), Ogata (1990), and DiStefano *et al.* (1990). For five years I shared with several faculty, a first course in feedback control of dynamical systems. That experience was enormously helpful to me.

In practice, the system or plant dynamics, written in terms of the Laplace transfer variable s and represented by its transfer function $G(s)$, are known. To meet performance specifications (such as steady-state error for a step input), a controller is required, having dynamics $H(s)$. Hence the open-loop transfer function $H(s)G(s)$ is known, practically always in factored form, so the open-loop performance and stability are readily determined.[1]

Performance specifications are often such that they cannot be met with open-loop operation. Feedback is then added, the performance specifications are placed on the closed-loop system, and we must be concerned with stability of the system shown in Figure G.1, for which the transfer function is

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

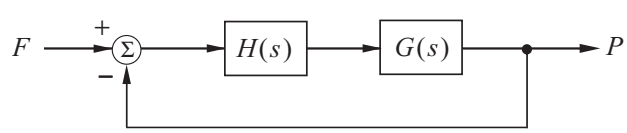


FIGURE G.1. The system $G(s)$ with controller $H(s)$ and negative unity feedback.

Stability now is determined by the zeros of $1 + HG$ and we are usually faced with two problems in the analysis and design of a feedback system:

(1) given $H(s)G(s)$, find the roots of $1 + H(s)G(s)$;
(2) for the given transfer function $G(s)$ of the plant, find the transfer function $H(s)$ of a controller required to meet the desired closed-loop performance.

[1] Unlike the convention followed in the remainder of this book, we assume in this annex that the time dependence for steady waves is $e^{+i\omega}$, instead of $e^{-i\omega}$, to be consistent with standard usage in control theory. Also, the complex variable, $s = \sigma + i\omega$, which originates as the Laplace transform variable in control theory, takes the part of $z = x + iy$.

The actual performance depends on the zeros of $1 + HG$, which cannot be determined until $H(s)$ is known. Hence there is clearly a design problem which can be solved only if methods of analysis are available, the matter we now address. The design process involves some sort of 'cut-and-try' or interactive process, at each stage of which the methods of analysis are applied to check the actual performance and stability.

G.1. The Criterion Stated Without Proof

The closed-loop transfer function will often have the form of 1 plus a rational polynomial,

$$1 + HG = 1 + K\frac{\prod^{i}(s - z_i)}{s^{\ell}\prod^{j}(s - p_j)},$$

where we allow for a pole of order ℓ at the origin in HG. Stability of the *open*-loop system depends on the *poles* of HG, but the stability of the *closed*-loop system depends on the *zeros* of $1+HG$. Hence it is possible that the closed-loop system may be stable even if the open-loop system is unstable, one of the advantages of feedback.

For practical purposes, it is usually difficult to factor $1+HG$, so closed-loop stability is not easily found. The Nyquist Criterion allows one to determine *without* factoring $1 + HG$, whether or not the closed-loop system is stable, and, if it is unstable, the number of unstable poles (i.e., zeros of $1+HG$ in the right-half of the s-plane). This remarkable result is based on a polar plot of the open-loop transfer function, proceeding in the following steps:

(i) locate the *poles* of $H(s)G(s)$ on the imaginary axis in the s-plane, including the origin;

(ii) construct the Nyquist *path* in the s-plane (Figure G.2), traversing the imaginary axis *excluding* the poles of HG, and closed by a semi-circle in the right-half s-plane. Poles and zeros of HG within the path are not excluded. Hence the interior of the closed figure lies to the *right* of the path;[2]

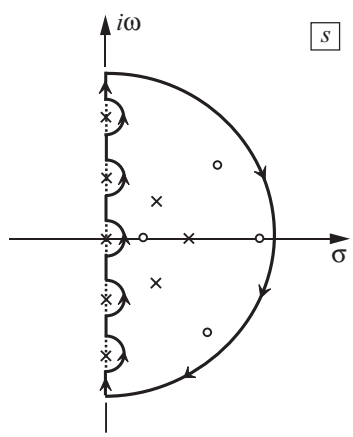


FIGURE G.2. Definition of the Nyquist path excluding poles of HG on the imaginary axis. Examples of poles and zeros within the path are not excluded.

[2]This is probably the most common definition of the Nyquist path, excluding poles on $Re(s) = 0$. Alternatively, poles on the imaginary axis can be explicitly included by closing the semi-circles to the left of those poles. The details of the following argument would change, and the statement of Nyquist's Criterion would be different, but the substance of the result would remain.

(iii) construct the Nyquist contour or polar plot, i.e. the mapping, of the Nyquist path by the open-loop transfer function $H(s)G(s)$;

(iv) count the number N of encirclements of the point -1 by the polar plot ($N > 0$ if -1 lies inside the contour and $N < 0$ if -1 lies outside, i.e. the contour makes counter-clockwise encirclements of -1);

(v) let $\mathcal{P}$ be the number of poles of HG in the right-half s-plane (right-half s-plane [RHP] means $Re(s) > 0$); $\mathcal{P} > 0$ if the open-loop system is unstable; $\mathcal{P} = 0$ if the open-loop system is stable, so all poles are in the left-half s-plane, or if HG has poles on the imaginary axis. Thus $\mathcal{P}$ is either zero or a positive integer.

(vi) then the *Nyquist Criterion* is:

(1) if $N > 0$, the number $\mathcal{Z}$ of zeros of $1 + HG$ in the *right*-half s-plane is

$$\mathcal{Z} = N + \mathcal{P} \tag{G.1}$$

(2) the closed-loop system is *stable* if and only if the number of *counter-clockwise* encirclements equals the number of poles in the right-half s-plane:

$$N = -\mathcal{P} \leq 0 \tag{G.2}$$

Note that the Nyquist Criterion gives information about the *absolute* stability of the closed-loop system but provides no measure of *relative* stability—i.e., how stable or unstable it is. However, we will see in Section G.7 that the Nyquist polar plot provides a clear basis for defining measures of relative stability.

G.2. Some General Properties of Polar Plots

In the present context we view a polar plot as a mapping—in the cases which we deal with a conformal transformation—of a contour in the s-plane to the image figure in the G-plane by $G(s)$. If the contour avoids poles of G, the mapping is conformal and angles are preserved. In particular, a right angle in the s-plane of the contour in the s-plane is mapped to the corresponding right angle in the G-plane:

The polar plot of $G(s) + g$, where $g = g_r + ig_i$, is a complex number, is the polar plot of $G(s)$ with origin shifted to $-g$. This property follows by direct calculation,

$$G + g = (G_r + g_r) + i(G_i + g_i)$$

which can be interpreted with a sketch of a simple example. See Figure G.8 for a case when $g = g_r = 1$.

For a time-invariant linear system the polar plot of $G(s)$ antisymmetric about the real axis for $s = i\omega$,

$$ImG(-i\omega) = -ImG(i\omega) \tag{G.3}$$

To show this, suppose that $G(s)$ is a rational polynomial with a pole of order ℓ at the origin,

$$G(s) = \frac{K}{s^\ell}\frac{\Pi(s - z_i)}{\Pi(s - p_j)} = \frac{K}{s^\ell}\frac{s^{n_N} + a_{N-1}s^{n_N-1} + \cdots}{s^{n_D} + a_{n_D-1}s^{n_D-1} + \cdots}$$

in which the coefficient in the polynomials, as well as the constant K, are real. On the imaginary axis of s,

$$G(i\omega) = \frac{K}{(i\omega)^\ell}\frac{\cdots(i\omega - z_n)(i\omega - z_n^*)\cdots}{\cdots(i\omega - p_n)(i\omega - p_n^*)\cdots}$$

and the complex conjugate is

$$G^*(i\omega) = G(-i\omega) = G_r - iG_i = \frac{K}{(-i\omega)^\ell}\frac{\cdots(-i\omega - z_n^*)(-i\omega - z_n)\cdots}{\cdots(-i\omega - p_n^*)(-i\omega - p_n)\cdots}$$

Consider the pair of factors shown in the numerators of G and G^*,

$$\begin{aligned} G: \quad (i\omega - z_n)(i\omega - z_n^*) &= (|z_n|^2 - \omega^2) - i(2\omega z_n^r) \\ G^*: \quad (-i\omega - z_n^*)(-i\omega - z_n) &= (|z_n|^2 - \omega^2) + i(2\omega z_n^r) \end{aligned}$$

These are typical factors so we conclude that $G(-i\omega)$ has imaginary part having the same numerical value, but opposite sign compared with that of $G(i\omega)$ if we ignore the pole of order ℓ at the origin, i.e. take $\ell = 0$. In this case, $G(i\omega)$ is antisymmetric about the real axis, as asserted by (G.1).

When $\ell = 1, 2, 3, \ldots$, G has the additional multiplying factor $\frac{1}{s}$, $\frac{1}{s^2}$, $\frac{1}{s^3}$, $\cdots$ or $\frac{1}{i\omega}$, $\frac{1}{(i\omega)^2}$, $\frac{1}{(i\omega)^3}$, $\cdots$, that is $\frac{1}{i\omega}$, $-\frac{1}{\omega^2}$, $-\frac{1}{i\omega^3}$, $\cdots$. Thus G^* contains the corresponding factors $-\frac{1}{i\omega}$, $-\frac{1}{\omega^2}$, $\frac{1}{i\omega^3}$, $\cdots$. Hence, when ℓ is even, the pole has no effect on the above argument and (G.1) still holds. But when ℓ is odd, the polar plot of G is symmetric about the real axis and the values of $G(-i\omega)$ overlay those of $G(i\omega)$.

G.3. Construction of Nyquist Contours (Polar Plots)

The Nyquist contour, or polar plot, is the mapping of the Nyquist path, Figure G.2, to the HG plane. Poles on the path are avoided by indenting the path, as shown in Figure G.3, with a small semi-circle of radius ρ ($\rho \to 0$ eventually); the path is closed by the large semi-circle of radius R ($R \to \infty$ eventually). The indentations on the imaginary s-axis exclude poles representing undamped motions, for example, oscillations of constant amplitude.

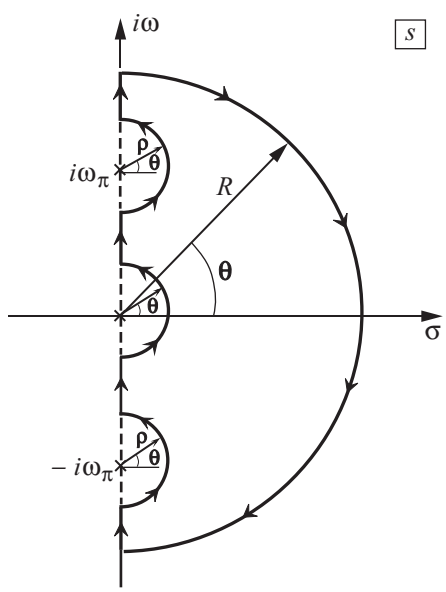


FIGURE G.3. Indentations around poles on the imaginary axis. Poles and zeros within the contour are supressed.

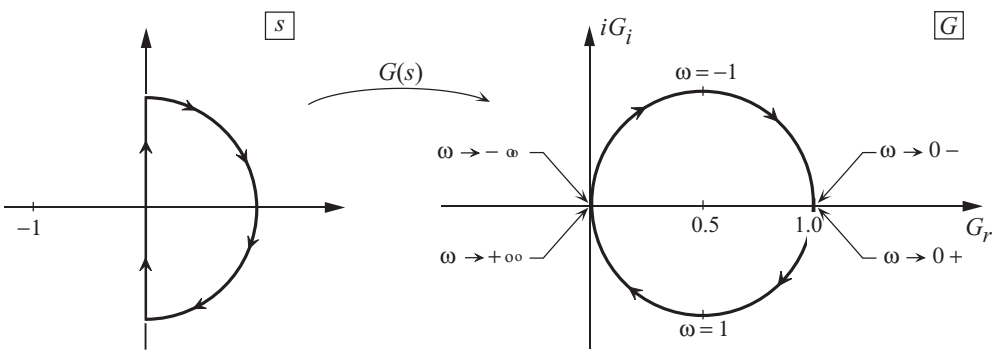
As an aid to constructing the mapping, note that on the various portions of the Nyquist path, s assumes the following values:

(i)	on the positive imaginary axis:	$s = i\omega$	
(ii)	on an indentation around a pole on the positive imaginary axis:	$s = i\omega_p = \rho e^{i\theta}$	$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$
(iii)	on the large semi-circle:	$s = Re^{i\theta}$	$\frac{\pi}{2} \le \theta \le -\frac{\pi}{2}$
(iv)	on an indentation around a pole on the negative imaginary axis:	$s = i\omega_p + \rho e^{i\theta}$	$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$
(v)	on the indentation around the origin:	$s = \rho e^{i\theta}$	$-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$

For simplicity in the following examples we set $H = 1$.[3]

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

There are no poles on the Nyquist path and the polar plot is sketched below.



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

This is example (a) plus a simple pole at the origin. The path in the s-plane must be indented to avoid the pole. On this indentation we set $s = \rho e^{i\theta}$ and G is

$$G = \frac{1}{\rho e^{i\theta}(1 + \rho e^{i\theta})} \underset{\rho \to 0}{\longrightarrow} \frac{1}{\rho} e^{-i\theta} \quad \left(-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}\right) \Longrightarrow \left(\frac{\pi}{2} \ge \phi \ge -\frac{\pi}{2}\right)$$

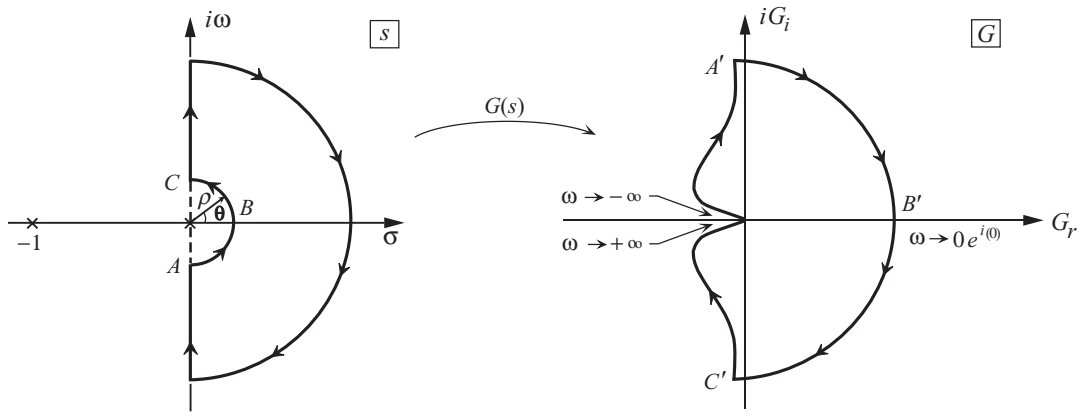
Hence $|G| \to \infty$ as the radius of the semi-circle shrinks, and the polar angle changes from $-\left(-\frac{\pi}{2}\right) = \frac{\pi}{2}$ to $-\frac{\pi}{2}$, traversing a large semi-circle in the counter-clockwise direction, as sketched below. (Points $A \to A'$, $B \to B'$, and $C \to C'$.)

On the large semi-circle of the path, $s = Re^{i\theta}$ and

$$G = \frac{1}{Re^{i\theta}\,(1 + Re^{i\theta})} \underset{R \to \infty}{\longrightarrow} \frac{1}{R^2} e^{-i2\theta} \quad \left(\frac{\pi}{2} \ge \theta \ge -\frac{\pi}{2}\right) \Longrightarrow (-\pi \le \phi \le \pi)$$

[3]For the examples here and in Section G.6, I have relied heavily on DiStefano, Stubberud and Williams (1990), Franklin *et al.* (2002) and Ogata (1990).

For $\theta \to \frac{\pi}{2}$ ($\omega \to \infty$ on the positive imaginary axis), $G \to \frac{1}{R^2}e^{-i\pi}$; and for $\theta \to -\frac{\pi}{2}(\omega \to -\infty$ on the negative imaginary axis), $G \to \frac{1}{R^2}e^{i\pi}$. Hence we find the mapping of the Nyquist path to the Nyquist polar plot shown below.



Compare this result with the figure for example (a). The pole at the origin causes the Nyquist contour to approach the origin along the negative real axis instead of tangent to the imaginary axis, and a portion of the contour therefore lies in the left half-plane.

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

Now we have a third-order pole at the origin, which will have a significant effect on the Nyquist contour at infinity. The path in the s-plane is the same as in example (b). For s on the indentation at the origin,

$$G = \frac{1}{\rho^3 e^{i3\theta}} \frac{1}{(1+\rho e^{i\theta})} \sim \frac{1}{\rho^3} e^{-i3\theta} \quad \left(-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}\right)$$

Thus as θ increases from $-\frac{\pi}{2}$ through zero to $+\frac{\pi}{2}$, $\arg G$ decreases from $+\frac{3\pi}{2}$ to $-\frac{3\pi}{2}$. However, care is required to get the proper behavior. A clue is found by computing the total change of $\arg G$ for the change in θ, $\Delta\theta = \pi$:

$$\Delta(\arg G) = -3\Delta\theta = -3\pi$$

Hence the contour makes an encirclement of the origin, a result that is clarified further by constructing the following table for $\arg G$ as a function of θ, plotted in the adjacent sketch.

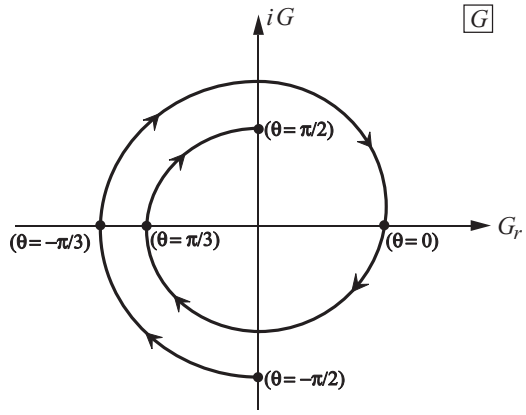
The question now is—how does the contour close? For s on the large semi-circle, we have

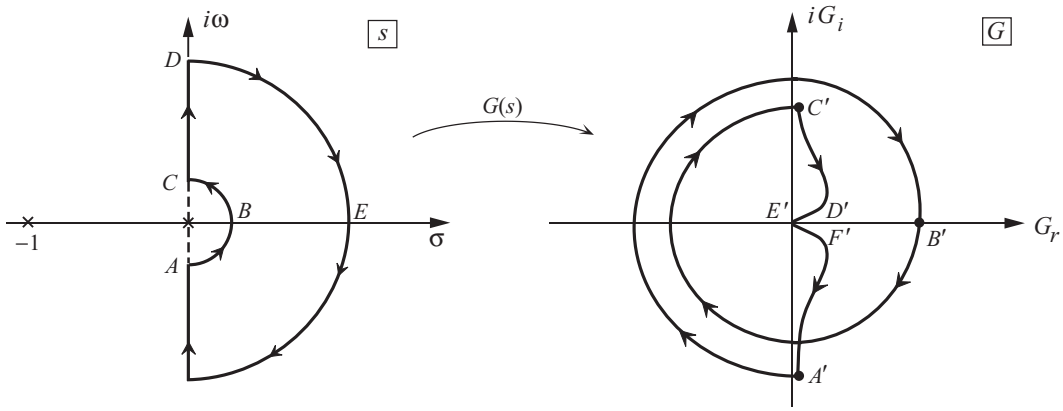
$$G \sim \frac{1}{R^4} e^{-i4\theta} \quad \left(\frac{\pi}{2} \ge \theta \ge -\frac{\pi}{2}\right)$$

so $\arg G$ changes from -2π for $\omega \to \infty$ on the positive imaginary axis, to $+2\pi$ for $\omega \to -\infty$ on the negative imaginary axis. Hence the contour comes into the origin tangent to the positive real axis, in the first quadrant ($\arg G = -2\pi$) for $\omega \to -\infty$ and in the fourth quadrant for $\omega \to -\infty$.

The Nyquist plot has the form shown in the figure, again with a few corresponding points indicated in the s- and G-planes. The dashed line is completely the image of the portion of the Nyquist path near the origin! Note especially that while the path in the s-plane is traversed once, the image polar plot encircles the origin twice in the G-plane. This is an example of

θ	$\arg G = -3\theta$
$-\pi/2$	$3\pi/2$
$-\pi/3$	π
0	0
$\pi/3$	$-\pi$
$\pi/2$	$-3\pi/2$





the importance of the behavior of $G(s)$ for $s \to 0$. In view of the final value theorem, this reflects the behavior that we can expect in the time domain for large times. So it may not be surprising that encirclements by the contours should be closely related to stability, one aspect of the long-time behavior of a system. In fact, stability is the only characteristic of long-time behavior that is independent of the forcing function or input—and nothing we are doing here is related to the input. These remarks suggest that we should be more precise with our definition of encirclement.

G.4. Definition of Encirclement

A closed contour in the HG plane is said to make N_0 *positive* encirclements of the origin if a line drawn from the origin to a point on the contour rotates *clockwise* through $2\pi N_0$ radians as the point traverses the contour. It is essential that the definition of positive be maintained, as we shall see with the proof of Nyquist's criterion, and emphasized in Figure G.4. The reason for this definition of positive follows from the chosen direction for positive traversal of the Nyquist contour in the s-plane.

As a practical matter, two points should be noted. First, polar plots for the systems we are concerned with are symmetrical about the real axis. That is particularly to be kept in mind when drawing or viewing contours making several encirclements of the origin—it is a matter of convenience in drawing that the portions far from the origin seem not to satisfy that symmetry, as is the case for example (c).

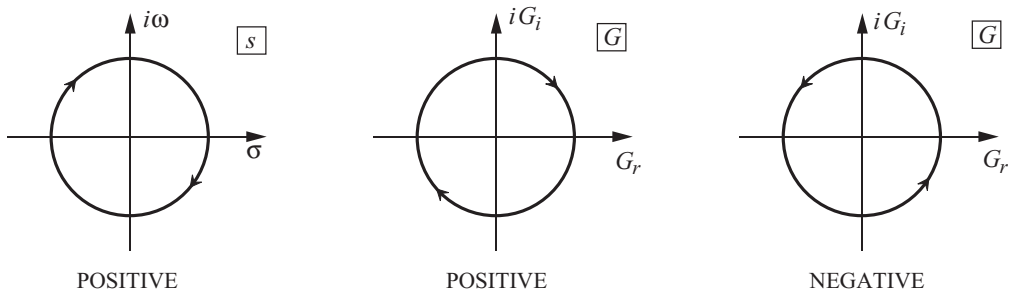


FIGURE G.4. Definitions of positive and negative encirclements.

Second, an operational hint, another way of determining the number of encirclements of the origin by a contour is to draw a line from the origin to infinity and count the *net* number of crossings of that line by the contour, taking into account the definition of 'positive' (hence of 'negative') given above.

G.5. Proof of Nyquist's Criterion

G.5.1. Principle of the Argument. We first establish an important relation among the numbers of poles and zeros of a function by applying Cauchy's residue theorem.[4] Consider a simple closed contour $\mathcal{C}$ in the s-plane and a function $f(s)$ analytic within $\mathcal{C}$ except for isolated poles, none of which are on $\mathcal{C}$. Moreover, we assume that $f(s)$ may have zeros in the region enclosed by $\mathcal{C}$ but does not vanish on $\mathcal{C}$. Now apply Cauchy's residue theorem to the function f'/f:

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f'}{f} ds = \sum \text{ Residues of } \frac{f'}{f}$$

The residues of f'/f are the coefficients of $1/(s - s_k)$ in the Laurent-series expansion. In this case, residues are associated with both poles and zeros of f. First, near a zero of order m_i of f, we can approximate $f(s)$ with the form

$$f(s) = (s - z_i)^{m_i} g(s)$$

where $g(s)$ is non-zero and analytic at $s = z_i$. Then

$$\frac{f'}{f} = \frac{m_i}{s - z_i} + \frac{g'}{g}$$

and the residue is m_i at $s = z_i$.

Similarly, near a pole of order n_j,

$$f(s) = \frac{h}{(s - p_j)^{n_j}}$$

where h is the analytic near p_j; the logarithmic derivative is

$$\frac{f'}{f} = \frac{-n_j}{s - p_j} + \frac{h'}{h}$$

so the residue is $-n_j$. Hence by the residue theorem we find

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f'}{f} ds = \sum m_i - \sum n_j \tag{G.4}$$

[4]We assume for this discussion sufficient knowledge of functions of a complex variable, at the level of a well-educated undergraduate. See, e.g., the Caltech undergraduate course ACM 95.

This result is true independently of the direction in which the contour is traversed: there is no ambiguity of sign.

If we regard each multiple pole (zero) as equivalent to $n_j(m_i)$ simple poles (zeros) we can write

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f'}{f} ds = \mathcal{Z} - \mathcal{P} \tag{G.5}$$

where $\mathcal{Z}$ and $\mathcal{P}$ are the numbers of zeros and poles respectively within $\mathcal{C}$.

Now carry out the integral explicitly:

$$\begin{aligned}\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f'}{f} ds &= \frac{1}{2\pi i}\left[\log f\right]_{\mathcal{C}} \\ &= \frac{1}{2\pi i}\left[\log |f|\right]_{\mathcal{C}} + \frac{1}{2\pi}\left[\arg |f|\right]_{\mathcal{C}}\end{aligned}$$

The change of a quantity in one passage around the path $\mathcal{C}$ in the s-plane is denoted by the brackets. Hence the first term vanishes, because $|f|$ returns to its initial value, and the second term is the increment in the argument of f, $\Delta_c(\arg f)$. When the value of s traverses the closed path once in the s-plane, the values of $f(s)$ will also trace a closed contour in the f-plane, which may be simple, or may intersect itself. If the contour in the f-plane encircles the origin once, then the argument of f changes by 2π; for N_0 encirclements of the origin, the change of argument is $2\pi N_0$. Figure G.5 illustrates two possibilities ($N_0 = 0$ and $N_0 = -2$).

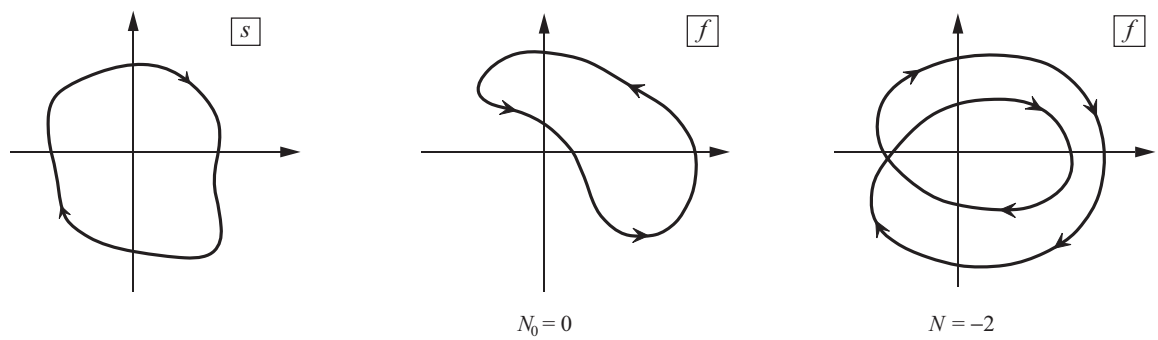


FIGURE G.5. Illustrating no encirclement ($N_0 = 0$) and two negative encirclements ($N_0 = -2$) by $f(s)$ when s executes a closed contour encircling the origin.

Therefore, the value of the integral may be written generally

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f'}{f} ds = \frac{1}{2\pi}\Delta_c(\arg f) = N_0 \tag{G.6}$$

Equating the right-hand sides of (G.5) and (G.6), we have the *principle of the argument:*

$$N_0 = \mathcal{Z} - \mathcal{P} \tag{G.7}$$

We emphasize that the result is *independent of the direction in which the path* $\mathcal{C}$ *is traversed* in the s-plane, but the direction (i.e., the definition of "inside" the path) must be consistently maintained in the f-plane. Figure G.6 shows examples of the point, two cases of clockwise and counter-clockwise traversal of the path in the s-plane each giving rise to two of the many possible cases in the f-plane. Which case arises depends of course on the function f. The shading denotes the interior of a path: the interior of a path in the s-plane maps to the interior of a contour in the f-plane because the mapping is conformal, preserving both the magnitude and sense of angles locally.

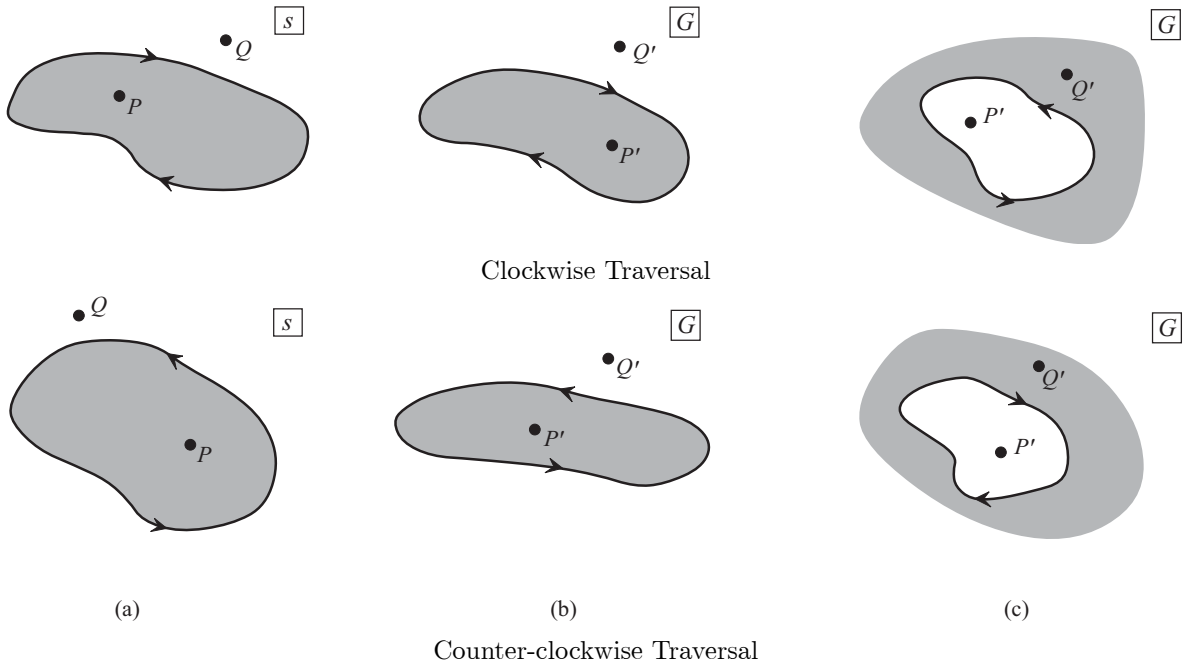


FIGURE G.6. Examples showing consistent maintenance of the definition of 'interior' of a path under mapping from the s-plane to the G-plane.

The definition of positive sense of a path is defined in the s-plane. Either case in Figure G.6 is valid: if the path is traversed in the clockwise sense, the interior is to the right, and if in the counter-clockwise sense the interior is to the left. The chosen definition is maintained in the G-plane, so the 'interior' regions are mapped as shown. Then an encirclement of a point in the G-plane is positive if that point belongs to the image of the interior of the path in the s-plane and negative if the point belongs to the image of the exterior. If the point lies outside the path in the G-plane, then the net phase change is zero for a vector drawn from the point to a point executing the path. By definition, the encirclement is then zero.

G.5.2. Proof of Nyquist's Criterion. Returning to the demonstration of Nyquist's criterion, we set $f = HG$, the forward-path transfer function. Equation (G.7) has then established that the number of encirclements of the origin by the Nyquist contour in the HG-plane is equal to the number of zeros of HG minus the number of poles of HG in the right half-plane. Because we chose to execute the Nyquist path in the clockwise direction in the s-plane, a positive encirclement in the HG- plane is also defined as a 2π increase of $\arg HG$ in the clockwise sense. Figure G.7 illustrates the result.

What we really want to determine is the number of zeros of $1 + HG$ in the right half-plane, a small extension of the reasoning leading to equation (G.7) and achieved in the following steps.

(i) Let $f(s) = 1 + HG$. Then the origin for the mapping $1 + HG$ in the HG-plane is at $HG = -1$. (Figure G.8 shows the process.)

(ii) Let N be the number of encirclements of -1 by the *mapping* of the Nyquist path by $HG(s)$.

(iii) Then according to equation (G.7), $N = \mathcal{Z} - \mathcal{P}$, where $\mathcal{Z}$, $\mathcal{P}$ *are the numbers of zeros and poles of* $1 + HG$ enclosed by the Nyquist path in the s-plane (i.e., in the right-half s-plane).

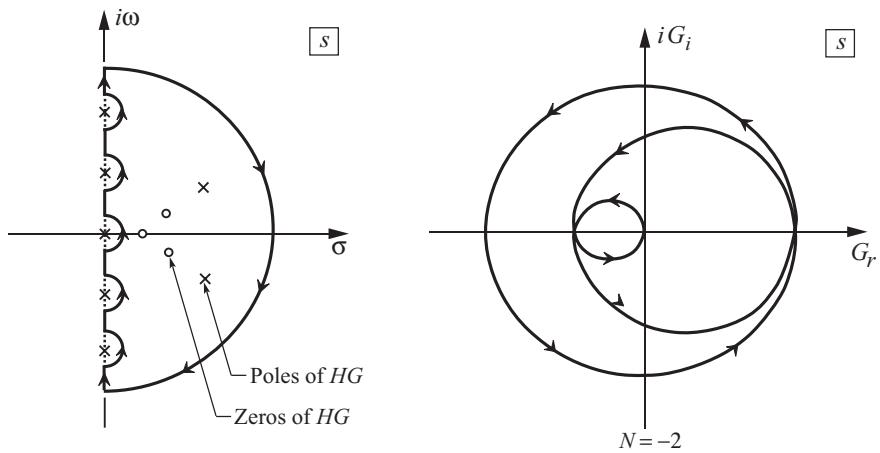


FIGURE G.7. An example showing two negative encirclements of the origin by the path obtained by mapping the Nyquist contour.

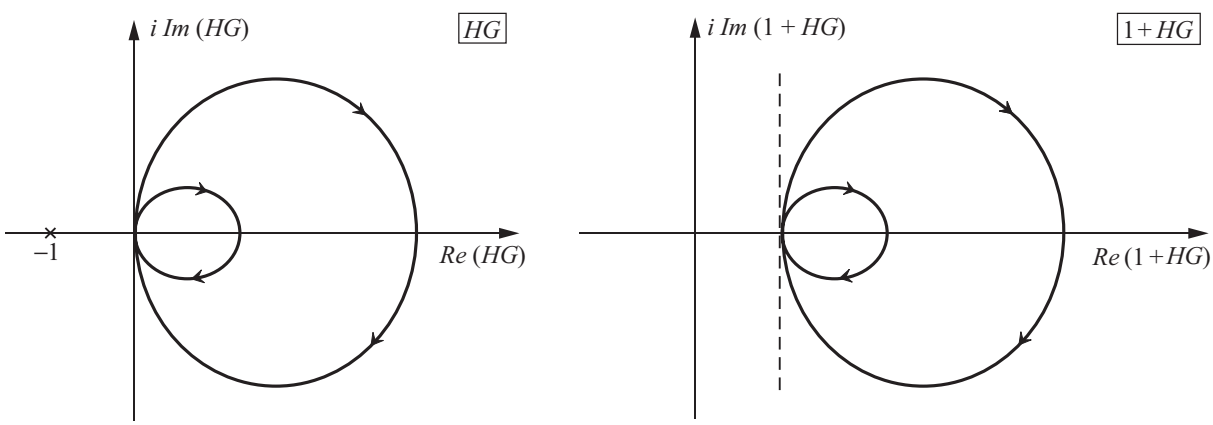


FIGURE G.8. Mappings of the Nyquist contour in the HG and $1 + HG$ planes.

(iv) The poles of $1 + HG$ are the same as the poles of HG. Hence the number of zeros $\mathcal{Z}$ of HG in the right-half s-plane is

$$\mathcal{Z} = N + \mathcal{P} \tag{G.8}$$

$\mathcal{Z}$: number of zeros of $1 + HG$ in the right-half s-plane, i.e. the number of unstable roots of the closed-loop system

N: number of encirclements of -1 by the mapping of HG of the Nyquist path

$\mathcal{P}$: number of poles of HG (hence of $1 + HG$) in the right-half s-plane

Thus we have shown how to determine the number of zeros of the *closed-loop* transfer function in the right half-plane from a polar plot of the *open-loop* transfer function applied to the Nyquist contour.

Note that if the closed loop is stable, there are no zeros of $1 + HG$ in the right half-plane, $\mathcal{Z} = 0$, and

$$N = -\mathcal{P} \geq 0 \tag{G.9}$$

which is zero or negative because by definition $\mathcal{P}$ is 0 or a positive number. Conversely, if $N = -\mathcal{P}$, then $\mathcal{Z} = 0$ and the closed-loop system is stable. Hence (G.9) *is the necessary and sufficient condition* that the closed-loop system be stable, where 'stable' means that the transient motions are bounded for large times (i.e., neutral stability is included). For N to be negative the point -1 is encircled negatively by the Nyquist path and hence by definition lies outside the path (see Figure G.6 and accompanying remarks). An equivalent statement is: if the mapping of the right-half s-plane under $1 + HG(s)$ does not include the point -1, then the system is stable.

If $\mathcal{P} = 0$ (no poles of HG in the right half-plane), then the system is stable if and only if $N = 0$, for then $\mathcal{Z} = 0$.

If $N > 0$, then the point -1 lies inside the Nyquist contour and there must be at least one zero of $1+HG$ in the right-half s-plane. Figure G.9 illustrates positive and negative encirclements of -1.

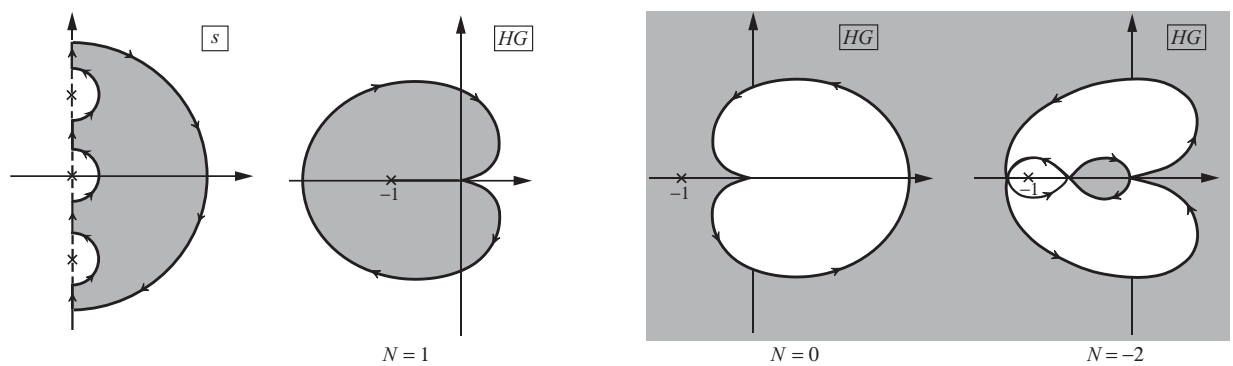


FIGURE G.9. Examples of positive and negative encirclements of -1 by mappings of the Nyquist contour.

As a final remark, we note that the portion of the Nyquist path for $s \to \infty$ *must* map to the origin in the HG-plane. Otherwise, the initial value theorem gives for the response near $t = 0$

$$F(t = 0+) = \lim_{s\to\infty} sHG\,P(s)$$

and for an impulse of amplitude A, $P(s \to \infty) = A$, so

$$F(t = 0+) = \lim_{s\to\infty} sHG\,A$$

Thus the initial response becomes infinitely large even for infinitesimally small impulses, unless $HG \sim 1/s^{1+\delta}$, where $\delta \geq 0$. Hence for any practically realizable system, the open-loop transfer function must have the behavior

$$HG \sim \frac{1}{s^n} \qquad (s \to \infty, n \geq 1) \tag{G.10}$$

Thus $HG \to 0$ for $s \to \infty$ and the Nyquist polar plot approaches the origin for $\omega \to +\infty$ and leaves the origin for ω decreasing from $-\infty$.

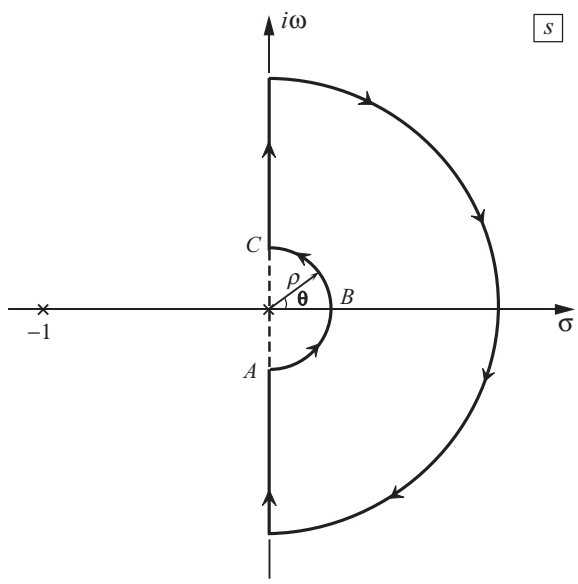
It follows that the character of the Nyquist polar plot (the mapping of the Nyquist path by HG) depends heavily on the portion of the Nyquist path on the imaginary axis. But note that it is not only the array of poles of HG on the axis that matters, for we are really using the mapping of the Nyquist path by the function $HG(s)$ so all poles and zeros matter. The behavior required by equation (G.10) is an example.

G.6. Examples of Nyquist's Criterion

The procedure for applying Nyquist's criterion is straightforward, although the details may cause minor difficulties. According to our discussion in the preceding sections, there are two preliminary tasks: (1) construct the Nyquist plot as the mapping of the Nyquist contour; and (2) count the number N of encirclements of the -1 point. If N is equal to the number $\mathcal{P}$ of poles of the open-loop transfer function in the right half-plane and the encirclements are in the sense opposite to the traversal of the Nyquist contour, then the closed-loop system is stable. Otherwise the closed-loop system is unstable, having $N + \mathcal{P}$ poles in the right half-plane.

(a) $HG = \frac{K}{s(s+1)} \qquad (K > 0)$

The Nyquist polar plot sketched in example (b), Section G.3, is repeated here. The plot does



not encircle the point -1, so $N = 0$; there are no open-loop poles in the right half-plane, $\mathcal{P} = 0$, and we have $N = \mathcal{P} = 0 = \mathcal{Z}$: the closed-loop system is stable for all K. This result is easily confirmed from the closed-loop transfer function:

$$\frac{F}{P} = \frac{HG}{1 + HG} = \frac{\frac{K}{s(s+1)}}{1 + \frac{K}{s(s+1)}} = \frac{K}{s^2 + s + K}$$

The roots of the denominator always have negative real parts if $K > 0$.

$$\begin{aligned} s_{1,2} &= -\frac{1}{2} \pm \frac{1}{2}\sqrt{1 - 4K} \qquad (K \le \frac{1}{2}) \\ &= -\frac{1}{2} \pm i\frac{1}{2}\sqrt{4K - 1} \qquad (K \ge \frac{1}{2}) \end{aligned}$$

If K is negative, the plot actually closes in the left half-plane, always enclosing the point -1; the direction of traversal is clockwise, and $N = 1$. Because $\mathcal{P} = 0$, we then find correctly that $\mathcal{Z} = 1$.

(b) $HG = \frac{K}{s^3(s+1)} \qquad (K > 0)$

The Nyquist polar plot is sketched in example (c), Section G.3. The plot encircles -1 twice in the positive direction, $N = 2$. There are no poles of HG in the right half-plane and the closed-loop system is unstable with $\mathcal{Z} = N + \mathcal{P} = 2$ zeros in the right half-plane. The denominator of the closed-loop transfer function is $s^4 + s^3 + K$, not easily factored, so we cannot confirm the result directly.

(c) $HG = \frac{K}{s(s+p_1)(s+p_2)}$ $\qquad (K > 0; p_1, p_2 > 0)$

This transfer function may represent use of an integrator with a second-order system having two real stable poles. Now we need to construct the Nyquist plot, sketched in Figure G.10.

We have assumed p_1 and p_2 positive. This assumption sets the position of the asymptote in the left half-plane. To see this, write HG for $\omega \to 0$ as

$$\begin{aligned}\frac{1}{K}HG &= \frac{1}{i\omega(i\omega + p_1)(i\omega + p_2)} \\ &= -\frac{i}{\omega}\frac{(p_1 - i\omega)(p_2 - i\omega)}{\omega(p_1^2 + \omega^2)(p_2^2 + \omega^2)} \\ &= -\frac{i}{\omega}\frac{(p_1 p_2 - \omega^2) - i\omega(p_1 + p_2)}{(p_1^2 + \omega^2)(p_2^2 + \omega^2)}\end{aligned}$$

Now let $\omega \to 0$ to find

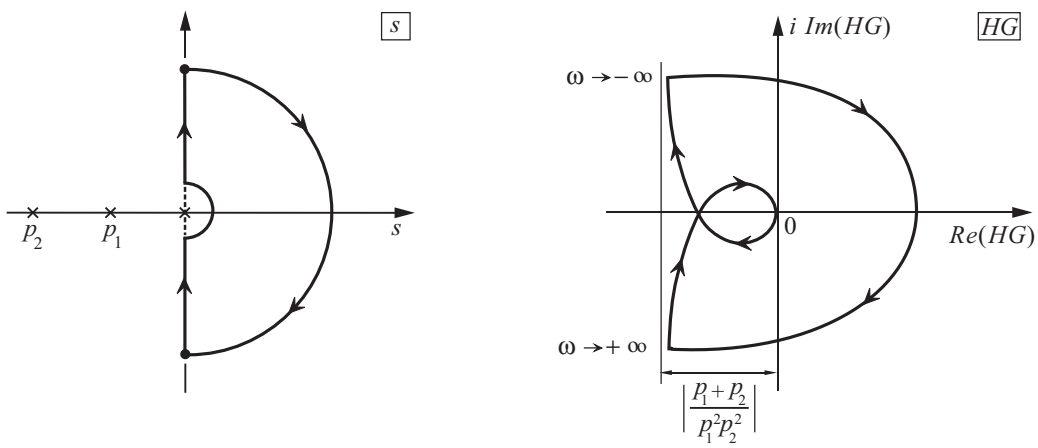


FIGURE G.10. Forming the Nyquist plot for $HG = \frac{K}{s(s+p_1)(s+p_2)}$.

$$\frac{1}{K}HG \to -\frac{(p_1 + p_2)}{p_1^2 p_2^2} - i\frac{1}{\omega}\frac{1}{p_1 p_2}$$

which shows the asymptote.

Whether or not the plot encircles -1 depends on where the point A is. We can find A by using the condition that at A, the phase of HG, as ω is increasing from 0^+, is $-\pi$. Then from the definition of HG, we have

$$\frac{1}{K}\arg HG \equiv -\frac{\pi}{K} = -\frac{\pi}{2} - \tan^{-1}\frac{\omega_A}{p_1} - \tan^{-1}\frac{\omega_A}{p_2} \tag{G.11}$$

where ω_A is the frequency at A. Given p_1, p_2 and K, we solve this equation for ω_A. Then calculate $|HG(\omega_A)|$ to find the position of A; if $|HG(\omega_A)| > 1$, then the plot makes one positive

encirclement of -1. Hence $\mathcal{Z} = 1 + \mathcal{P} = 1$, because we assume that both poles lie in the left half-plane (p_1, $p_2 > 0$ so the values of s_1, s_2 are negative at the poles). Hence the imaginary part is

$$\mathcal{I}m\{HG\} = -\frac{K(p_1p_2 - \omega^2)}{\omega(p_1^2+\omega^2)(p_2^2+\omega^2)} \tag{G.12}$$

which vanishes when $\omega = \omega_A = \sqrt{p_1p_2}$. The magnitude of the real part is then

$$\begin{aligned}
|HG| &= K\left|\frac{-(p_1+p_2)}{(p_1^2+\omega^2)(p_2^2+\omega^2)}\right|_{\omega_A} \\
&= K\frac{(p_1+p_2)}{p_1^2+p_2^2+(p_1^2+p_1^2)(p_1+p_2)+p_1^2+p_2^2} \\
&= K\frac{p_1p_2}{(p_1^2+2p_1p_2+p_2^2)} \\
&= K\frac{(p_1+p_2)}{(p_1p_2)(p_1+p_2)^2} \\
&= \frac{K}{p_1p_2(p_1+p_2)}
\end{aligned}$$

Hence the system is unstable if

$$\frac{K}{p_1p_2(p_1+p_2)} > 1$$

or

$$K > p_1p_2(p_1+p_2) \tag{G.13}$$

G.7. Relative Stability; Gain and Phase Margins

A great advantage of the Nyquist method is that it suggests a quantitative assessment of *relative stability* —i.e., how far is the closed-loop system from being unstable. This result appears in practice as a restriction on the magnitude of the gain arising from the compensator or controller represented by H. We have already seen in example (c) of the preceding section how increasing the gain can cause the Nyquist plot to expand and enclose the -1 point, indicating that the system has become unstable.

To reiterate the point, consider the control system in which a plant having two stable poles at -1 is subject to integral control, $H = K/s$, Figure G.11. The Nyquist plot is the same as Figure G.10, but now with $p_1 = p_2 = 1$, and is re-drawn in Figure G.12. We have already shown, with equation (G.13), that the plot encircles the point -1 if $K > 2$. To make quantitative statements about how close the system is to being unstable (i.e., in this case when $K < 2$), consider the portion of the plot for positive ω and $K < 2$.

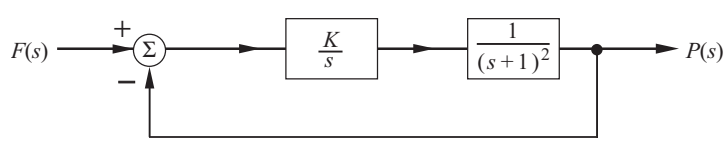


FIGURE G.11. A stable second order system with an integrator.

The *gain margin* is defined as a measure of the change of gain necessary to cause point A to reach -1. A common 'general' definition is

$$GM = \frac{1}{|HG|_\pi} \tag{G.14}$$

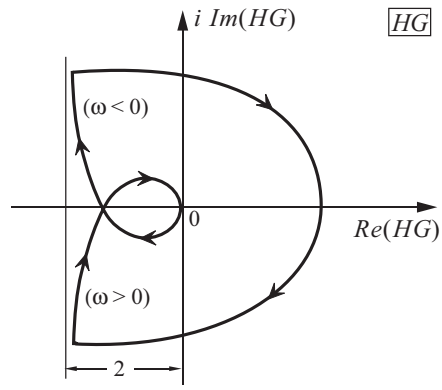


FIGURE G.12. The Nyquist plot for the system shown in Figure G.11.

where $|HG|_\pi$, is the magnitude of HG when $\arg HG = \pi$. The meaning is clear if we set $H = Kh$, where h is now independent of gain, and rewrite (G.14) as

$$(GM)K|hG|_\pi = 1$$

Hence GM is the multiplier of the gain required to make $|HG|_\pi = 1$, so the contour passes through the point -1 (i.e., when $\arg(HG) = \pi$).

Another measure of relative stability is the *phase margin* defined as the difference between π and the phase of HG where $|HG| = 1$; hence for the case drawn here (note that $\arg HG$ is *negative*)

$$PM = \arg HG(i\omega_1) + \pi \tag{G.15}$$

where ω_1 is the frequency at which $|HG| = 1$, the condition denoted $|HG|_\pi = 1$.

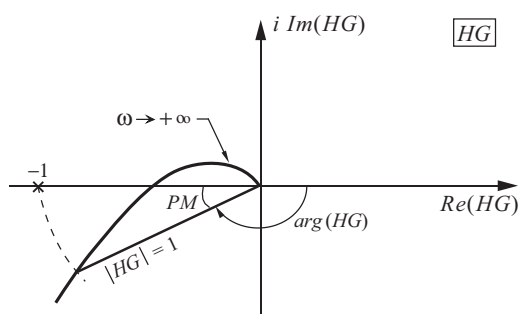


FIGURE G.13. The Nyquist plot for 'any' HG near $|HG|_\pi$, i.e., the magnitude of HG when the phase or arg of HG is π.

For the example $HG = K/s(s+1)^2$ giving Figure G.11, when $s = i\omega$,

$$HG = \frac{K}{\omega(1+\omega^2)} e^{-i\left(\frac{\pi}{2} + 2\tan^{-1}\omega\right)} \tag{G.16}$$

The contour crosses the axis when $\arg(HG) = -\pi$;

$$-\left(\frac{\pi}{2} + 2\tan^{-1}\omega\right) = -\pi$$

or

$$2\tan^{-1}\omega = \frac{\pi}{2}$$

Hence $\tan^{-1}\omega = \pi/4$, or $\omega = 1$. If we set $\arg(HG) = -\pi$, we find $\omega = -1$, a result also immediately evident from the symmetry of the contour about the real axis.

When $\omega = 1$,

$$|HG|_\pi = \left.\frac{K}{\omega(1+\omega^2}\right|_{\omega=1} = \frac{K}{2}$$

Hence from the definition (G.14) we have

$$GM = \frac{2}{K}$$

which is less than unity, and the system is unstable if $K > 2$, the value given by equation (G.13) for $p_1 = p_2 = 1$. That is, if $K > 2$, then $|HG|_\pi = K/2$ is greater than 1 and the contour encloses 1.

It is more difficult to determine the phase margin from the Nyquist plot, for we require the value of ω satisfying the transcendental equation

$$|HG| = 1 = \frac{K}{\omega(1+\omega^2)} \tag{G.17}$$

Here, ω is the solution to

$$\omega^3 + \omega - K = 0, \tag{G.18}$$

and this is a *simple* example!

Consequently, while the ideas and definitions of the gain and phase margins are suggested by the Nyquist plot, their values are more easily found from Bode plots. To see how the procedure works we continue with the example given in Figure G.11,

$$HG(i\omega) = \frac{K}{i\omega}\frac{1}{(1+i\omega)^2}$$

The Bode plot is the graphical representation of the magnitude and phase of HG for $\omega > 0$:

$$20\log_{10}|HG| = 20\log_{10}K - 20\log_{10}\omega - 20\log_{10}(1+\omega^2)$$

$$\arg HG = -\left(\frac{\pi}{2} + 2\tan^{-1}\omega\right)$$

These are sketched in Figure G.14; the magnitude is drawn for $K = 1$.

To compute the gain margin, find the value of $|HG|$ for $\arg HG = -\pi$, giving $|HG|_\pi = 0.5$. Hence from equation (G.14), $GM = 1/|HG|_\pi = 2$, which is here the gain margin for $K = 1$. Similarly, to compute the phase margin, find the value of $\arg HG$ at which $|HG| = 1$, giving the value here $PM = 22°$. It should be evident that Bode plots prepared with readily available computer programs can be used to find the gain and phase margins.

The definitions of phase margin and gain margin seemed quite natural for the simple example, Figure G.11, treated here. One should wonder whether more complicated systems will have Nyquist polar plots having such a character that the same ideas can be applied. The answer—perhaps surprisingly—is yes. Although the entire Nyquist plot may be very complicated indeed, nevertheless, in the vicinity of the point -1, both the Nyquist plot and the Bode plots commonly behave much like this example.

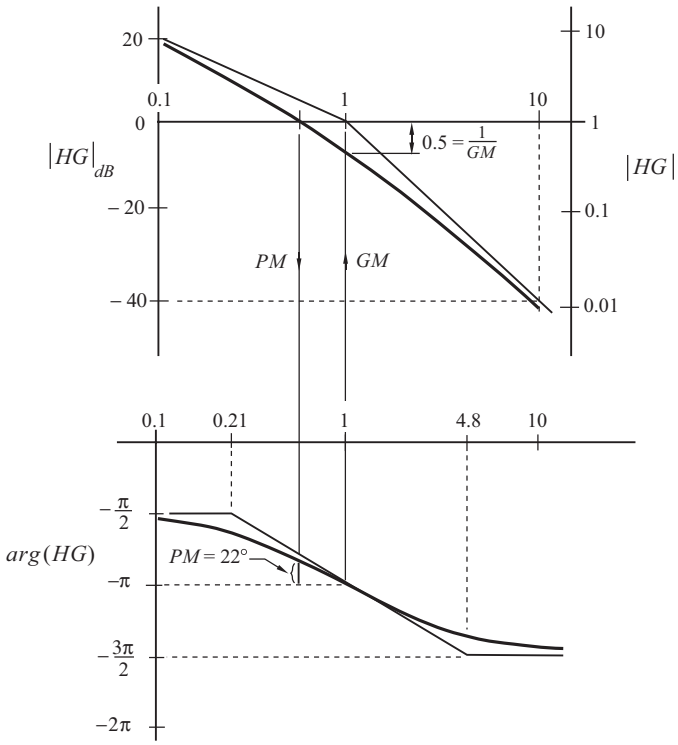


FIGURE G.14. The Bode plot for the system shown in Figure G.11, $K = 1$.