Chapter Nine
Frequency Domain Analysis

> *Mr. Black proposed a negative feedback repeater and proved by tests that it possessed the advantages which he had predicted for it. In particular, its gain was constant to a high degree, and it was linear enough so that spurious signals caused by the interaction of the various channels could be kept within permissible limits. For best results the feedback factor $\mu\beta$ had to be numerically much larger than unity. The possibility of stability with a feedback factor larger than unity was puzzling.*
>
> Harry Nyquist, "The Regeneration Theory," 1956 [Nyq56].

In this chapter we study how the stability and robustness of closed loop systems can be determined by investigating how sinusoidal signals of different frequencies propagate around the feedback loop. This technique allows us to reason about the closed loop behavior of a system through the frequency domain properties of the open loop transfer function. The Nyquist stability theorem is a key result that provides a way to analyze stability and introduce measures of degrees of stability.

9.1 The Loop Transfer Function

Determining the stability of systems interconnected by feedback can be tricky because each system influences the other, leading to potentially circular reasoning. Indeed, as the quote from Nyquist above illustrates, the behavior of feedback systems can often be puzzling. However, using the mathematical framework of transfer functions provides an elegant way to reason about such systems, which we call *loop analysis*.

The basic idea of loop analysis is to trace how a sinusoidal signal propagates in the feedback loop and explore the resulting stability by investigating if the propagated signal grows or decays. This is easy to do because the transmission of sinusoidal signals through a linear dynamical system is characterized by the frequency response of the system. The key result is the Nyquist stability theorem, which provides a great deal of insight regarding the stability of a system. Unlike proving stability with Lyapunov functions, studied in Chapter 4, the Nyquist criterion allows us to determine more than just whether a system is stable or unstable. It provides a measure of the degree of stability through the definition of stability margins. The Nyquist theorem also indicates how an unstable system should be changed to make it stable, which we shall study in detail in Chapters 10–12.

Consider the system in Figure 9.1a. The traditional way to determine if the closed loop system is stable is to investigate if the closed loop characteristic polynomial has all its roots in the left half-plane. If the process and the controller have rational

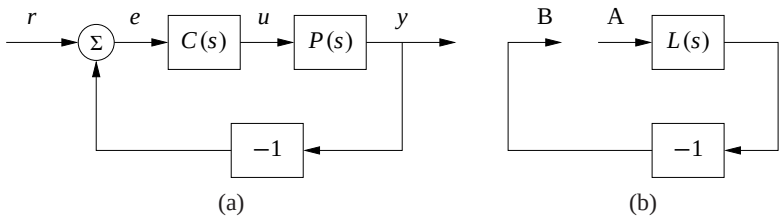


Figure 9.1: The loop transfer function. The stability of the feedback system (a) can be determined by tracing signals around the loop. Letting $L = PC$ represent the loop transfer function, we break the loop in (b) and ask whether a signal injected at the point A has the same magnitude and phase when it reaches point B.

transfer functions $P(s) = n_p(s)/d_p(s)$ and $C(s) = n_c(s)/d_c(s)$, then the closed loop system has the transfer function

$$G_{yr}(s) = \frac{PC}{1 + PC} = \frac{n_p(s)n_c(s)}{d_p(s)d_c(s) + n_p(s)n_c(s)},$$

and the characteristic polynomial is

$$\lambda(s) = d_p(s)d_c(s) + n_p(s)n_c(s).$$

To check stability, we simply compute the roots of the characteristic polynomial and verify that they each have negative real part. This approach is straightforward but it gives little guidance for design: it is not easy to tell how the controller should be modified to make an unstable system stable.

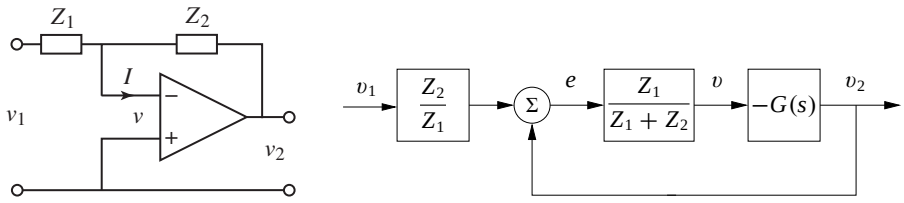
Nyquist's idea was to investigate conditions under which oscillations can occur in a feedback loop. To study this, we introduce the *loop transfer function* $L(s) = P(s)C(s)$, which is the transfer function obtained by breaking the feedback loop, as shown in Figure 9.1b. The loop transfer function is simply the transfer function from the input at position A to the output at position B multiplied by -1 (to account for the usual convention of negative feedback).

We will first determine conditions for having a periodic oscillation in the loop. Assume that a sinusoid of frequency ω_0 is injected at point A. In steady state the signal at point B will also be a sinusoid with the frequency ω_0. It seems reasonable that an oscillation can be maintained if the signal at B has the same amplitude and phase as the injected signal because we can then disconnect the injected signal and connect A to B. Tracing signals around the loop, we find that the signals at A and B are identical if

$$L(i\omega_0) = -1, \tag{9.1}$$

which then provides a condition for maintaining an oscillation. The key idea of the Nyquist stability criterion is to understand when this can happen in a general setting. As we shall see, this basic argument becomes more subtle when the loop transfer function has poles in the right half-plane.

Example 9.1 Operational amplifier circuit
Consider the op amp circuit in Figure 9.2a, where Z_1 and Z_2 are the transfer func-



(a) Amplifier circuit (b) Block diagram

Figure 9.2: Loop transfer function for an op amp. The op amp circuit (a) has a nominal transfer function $v_2/v_1 = Z_2(s)/Z_1(s)$, where Z_1 and Z_2 are the impedances of the circuit elements. The system can be represented by its block diagram (b), where we now include the op amp dynamics $G(s)$. The loop transfer function is $L = Z_1 G/(Z_1 + Z_2)$.

tions of the feedback elements from voltage to current. There is feedback because voltage v_2 is related to voltage v through the transfer function $-G$ describing the op amp dynamics and voltage v is related to voltage v_2 through the transfer function $Z_1/(Z_1 + Z_2)$. The loop transfer function is thus

$$L = \frac{GZ_1}{Z_1 + Z_2}. \tag{9.2}$$

Assuming that the current I is zero, the current through the elements Z_1 and Z_2 is the same, which implies

$$\frac{v_1 - v}{Z_1} = \frac{v - v_2}{Z_2}.$$

Solving for v gives

$$v = \frac{Z_2 v_1 + Z_1 v_2}{Z_1 + Z_2} = \frac{Z_2 v_1 - Z_1 G v}{Z_1 + Z_2} = \frac{Z_2}{Z_1}\frac{L}{G} v_1 - Lv.$$

Since $v_2 = -Gv$ the input/output relation for the circuit becomes

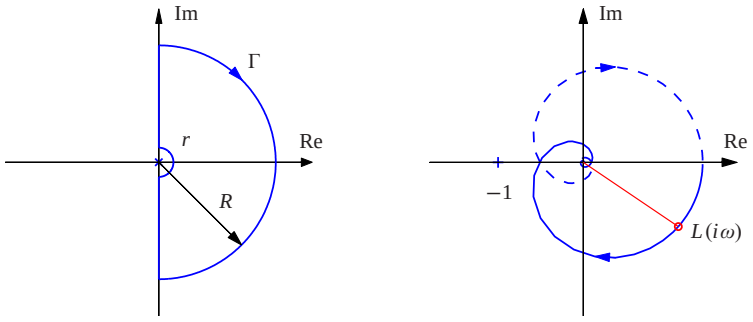
$$G_{v_2 v_1} = -\frac{Z_2}{Z_1}\frac{L}{1 + L}.$$

A block diagram is shown in Figure 9.2b. It follows from (9.1) that the condition for oscillation of the op amp circuit is

$$L(i\omega) = \frac{Z_1(i\omega)G(i\omega)}{Z_1(i\omega) + Z_2(i\omega)} = -1 \tag{9.3}$$

∇

One of the powerful concepts embedded in Nyquist's approach to stability analysis is that it allows us to study the stability of the feedback system by looking at properties of the loop transfer function. The advantage of doing this is that it is easy to see how the controller should be chosen to obtain a desired loop transfer function. For example, if we change the gain of the controller, the loop transfer function will be scaled accordingly. A simple way to stabilize an unstable system is then to reduce the gain so that the -1 point is avoided. Another way is to introduce



(a) Nyquist D contour (b) Nyquist plot

Figure 9.3: The Nyquist contour Γ and the Nyquist plot. The Nyquist contour (a) encloses the right half-plane, with a small semicircle around any poles of $L(s)$ on the imaginary axis (illustrated here at the origin) and an arc at infinity, represented by $R \to \infty$. The Nyquist plot (b) is the image of the loop transfer function $L(s)$ when s traverses Γ in the clockwise direction. The solid line corresponds to $\omega > 0$, and the dashed line to $\omega < 0$. The gain and phase at the frequency ω are $g = |L(i\omega)|$ and $\varphi = \angle L(i\omega)$. The curve is generated for $L(s) = 1.4e^{-s}/(s+1)^2$.

a controller with the property that it bends the loop transfer function away from the critical point, as we shall see in the next section. Different ways to do this, called loop shaping, will be developed and will be discussed in Chapter 11.

9.2 The Nyquist Criterion

In this section we present Nyquist's criterion for determining the stability of a feedback system through analysis of the loop transfer function. We begin by introducing a convenient graphical tool, the Nyquist plot, and show how it can be used to ascertain stability.

The Nyquist Plot

We saw in the last chapter that the dynamics of a linear system can be represented by its frequency response and graphically illustrated by a Bode plot. To study the stability of a system, we will make use of a different representation of the frequency response called a *Nyquist plot*. The Nyquist plot of the loop transfer function $L(s)$ is formed by tracing $s \in \mathbb{C}$ around the Nyquist "D contour," consisting of the imaginary axis combined with an arc at infinity connecting the endpoints of the imaginary axis. The contour, denoted as $\Gamma \in \mathbb{C}$, is illustrated in Figure 9.3a. The image of $L(s)$ when s traverses Γ gives a closed curve in the complex plane and is referred to as the Nyquist plot for $L(s)$, as shown in Figure 9.3b. Note that if the transfer function $L(s)$ goes to zero as s gets large (the usual case), then the portion of the contour "at infinity" maps to the origin. Furthermore, the portion of the plot corresponding to $\omega < 0$ is the mirror image of the portion with $\omega > 0$.

There is a subtlety in the Nyquist plot when the loop transfer function has poles on the imaginary axis because the gain is infinite at the poles. To solve this problem, we modify the contour Γ to include small deviations that avoid any poles on the imaginary axis, as illustrated in Figure 9.3a (assuming a pole of $L(s)$ at the origin). The deviation consists of a small semicircle to the right of the imaginary axis pole location.

The condition for oscillation given in equation (9.1) implies that the Nyquist plot of the loop transfer function go through the point $L = -1$, which is called the *critical point*. Let ω_c represent a frequency at which $\angle L(i\omega_c) = 180^\circ$, corresponding to the Nyquist curve crossing the negative real axis. Intuitively it seems reasonable that the system is stable if $|L(i\omega_c)| < 1$, which means that the critical point -1 is on the left-hand side of the Nyquist curve, as indicated in Figure 9.3b. This means that the signal at point B will have smaller amplitude than the injected signal. This is essentially true, but there are several subtleties that require a proper mathematical analysis to clear up. We defer the details for now and state the Nyquist condition for the special case where $L(s)$ is a stable transfer function.

Theorem 9.1 (Simplified Nyquist criterion)**.** *Let $L(s)$ be the loop transfer function for a negative feedback system (as shown in Figure 9.1a) and assume that L has no poles in the closed right half-plane (*$\operatorname{Re} s \geq 0$*) except for single poles on the imaginary axis. Then the closed loop system is stable if and only if the closed contour given by $\Omega = \{L(i\omega) : -\infty < \omega < \infty\} \subset \mathbb{C}$ has no net encirclements of the critical point $s = -1$.*

The following conceptual procedure can be used to determine that there are no encirclements. Fix a pin at the critical point $s = -1$, orthogonal to the plane. Attach a string with one end at the critical point and the other on the Nyquist plot. Let the end of the string attached to the Nyquist curve traverse the whole curve. There are no encirclements if the string does not wind up on the pin when the curve is encircled.

Example 9.2 Third-order system
Consider a third-order transfer function

$$L(s) = \frac{1}{(s+a)^3}.$$

To compute the Nyquist plot we start by evaluating points on the imaginary axis $s = i\omega$, which yields

$$L(i\omega) = \frac{1}{(i\omega+a)^3} = \frac{(a-i\omega)^3}{(a^2+\omega^2)^3} = \frac{a^3-3a\omega^2}{(a^2+\omega^2)^3} + i\,\frac{\omega^3-3a^2\omega}{(a^2+\omega^2)^3}.$$

This is plotted in the complex plane in Figure 9.4, with the points corresponding to $\omega > 0$ drawn as a solid line and $\omega < 0$ as a dashed line. Notice that these curves are mirror images of each other.

To complete the Nyquist plot, we compute $L(s)$ for s on the outer arc of the

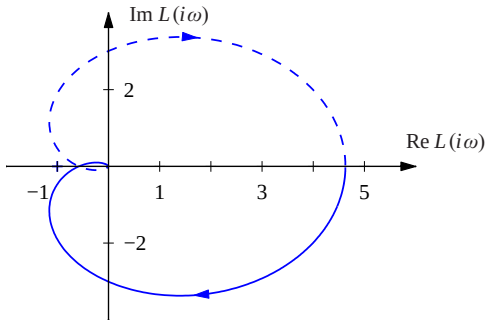


Figure 9.4: Nyquist plot for a third-order transfer function. The Nyquist plot consists of a trace of the loop transfer function $L(s) = 1/(s+a)^3$. The solid line represents the portion of the transfer function along the positive imaginary axis, and the dashed line the negative imaginary axis. The outer arc of the D contour maps to the origin.

Nyquist D contour. This arc has the form $s = Re^{i\theta}$ for $R \to \infty$. This gives

$$L(Re^{i\theta}) = \frac{1}{(Re^{i\theta}+a)^3} \to 0 \quad \text{as} \quad R \to \infty.$$

Thus the outer arc of the D contour maps to the origin on the Nyquist plot. ∇

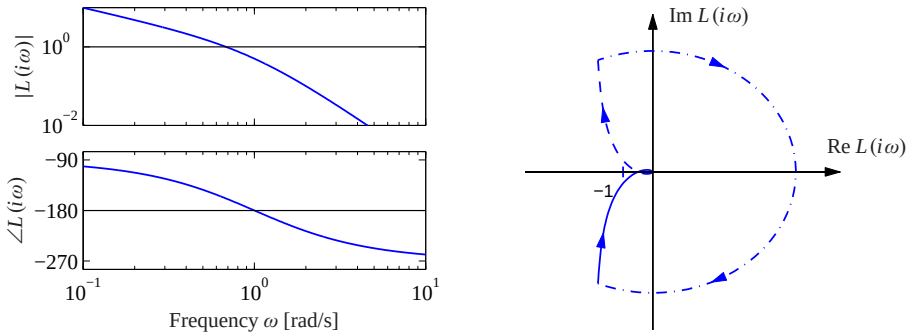
An alternative to computing the Nyquist plot explicitly is to determine the plot from the frequency response (Bode plot), which gives the Nyquist curve for $s = i\omega$, $\omega > 0$. We start by plotting $G(i\omega)$ from $\omega = 0$ to $\omega = \infty$, which can be read off from the magnitude and phase of the transfer function. We then plot $G(Re^{i\theta})$ with $\theta \in [-\pi/2, \pi/2]$ and $R \to \infty$, which almost always maps to zero. The remaining parts of the plot can be determined by taking the mirror image of the curve thus far (normally plotted using a dashed line). The plot can then be labeled with arrows corresponding to a clockwise traversal around the D contour (the same direction in which the first portion of the curve was plotted).

Example 9.3 Third-order system with a pole at the origin
Consider the transfer function

$$L(s) = \frac{k}{s(s+1)^2},$$

where the gain has the nominal value $k = 1$. The Bode plot is shown in Figure 9.5a. The system has a single pole at $s = 0$ and a double pole at $s = -1$. The gain curve of the Bode plot thus has the slope -1 for low frequencies, and at the double pole $s = 1$ the slope changes to -3. For small s we have $L \approx k/s$, which means that the low-frequency asymptote intersects the unit gain line at $\omega = k$. The phase curve starts at -90° for low frequencies, it is -180° at the breakpoint $\omega = 1$ and it is -270° at high frequencies.

Having obtained the Bode plot, we can now sketch the Nyquist plot, shown in Figure 9.5b. It starts with a phase of -90° for low frequencies, intersects the negative real axis at the breakpoint $\omega = 1$ where $L(i) = 0.5$ and goes to zero along



(a) Bode plot (b) Nyquist plot

Figure 9.5: Sketching Nyquist and Bode plots. The loop transfer function is $L(s) = 1/(s(s+1)^2)$. The large semicircle is the map of the small semicircle of the Γ contour around the pole at the origin. The closed loop is stable because the Nyquist curve does not encircle the critical point. The point where the phase is -180° is marked with a circle in the Bode plot.

the imaginary axis for high frequencies. The small half-circle of the Γ contour at the origin is mapped on a large circle enclosing the right half-plane. The Nyquist curve does not encircle the critical point, and it follows from the simplified Nyquist theorem that the closed loop is stable. Since $L(i) = -k/2$, we find the system becomes unstable if the gain is increased to $k = 2$ or beyond. ∇

The Nyquist criterion does not require that $|L(i\omega_c)| < 1$ for all ω_c corresponding to a crossing of the negative real axis. Rather, it says that the number of encirclements must be zero, allowing for the possibility that the Nyquist curve could cross the negative real axis and cross back at magnitudes greater than 1. The fact that it was possible to have high feedback gains surprised the early designers of feedback amplifiers, as mentioned in the quote in the beginning of this chapter.

One advantage of the Nyquist criterion is that it tells us how a system is influenced by changes of the controller parameters. For example, it is very easy to visualize what happens when the gain is changed since this just scales the Nyquist curve.

Example 9.4 Congestion control

Consider the Internet congestion control system described in Section 3.4. Suppose we have N identical sources and a disturbance d representing an external data source, as shown in Figure 9.6a. We let w represent the individual window size for a source, q represent the end-to-end probability of a dropped packet, b represent the number of packets in the router's buffer and p represent the probability that that a packet is dropped by the router. We write $\bar{w}$ for the total number of packets being received from all N sources. We also include a time delay between the router and the senders, representing the time delays between the sender and receiver.

To analyze the stability of the system, we use the transfer functions computed

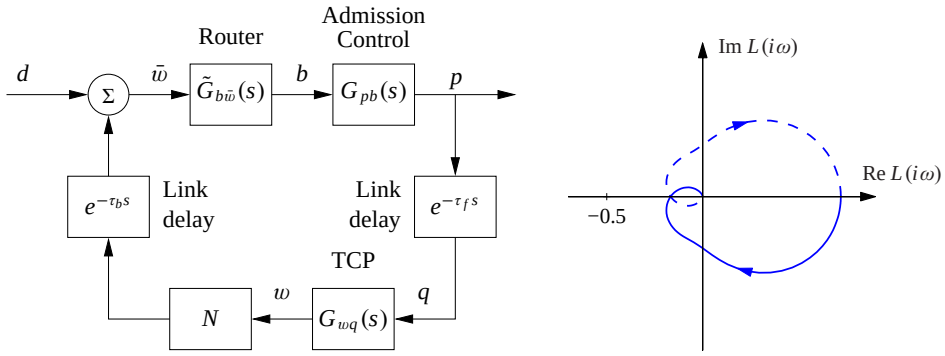


Figure 9.6: Internet congestion control. A set of N sources using TCP/Reno send messages through a single router with admission control (left). Link delays are included for the forward and backward directions. The Nyquist plot for the loop transfer function is shown on the right.

in Exercise 94:

$$\tilde{G}_{b\bar{w}}(s) = \frac{1}{\tau_e s + e^{-\tau_f s}}, \qquad G_{wq}(s) = -\frac{1}{q_e(\tau_e s + q_e w_e)}, \qquad G_{pb}(s) = \rho,$$

where (w_e, b_e) is the equilibrium point for the system, N is the number of sources, τ_e is the steady-state round-trip time and τ_f is the forward propagation time. We use $\tilde{G}_{b\bar{w}}$ to represent the transfer function with the forward time delay removed since this is accounted for as a separate block in Figure 9.6a. Similarly, $G_{wq} = G_{\bar{w}q}/N$ since we have pulled out the multiplier N as a separate block as well.

The loop transfer function is given by

$$L(s) = \rho \cdot \frac{N}{\tau_e s + e^{-\tau_f s}} \cdot \frac{1}{q_e(\tau_e s + q_e w_e)} e^{-\tau_e s}.$$

Using the fact that $q_e \approx 2N/w_e^2 = 2N^3/(\tau_e c)^2$ and $w_e = b_e/N = \tau_e c/N$ from equation (3.22), we can show that

$$L(s) = \rho \cdot \frac{N}{\tau_e s + e^{-\tau_f s}} \cdot \frac{c^3 \tau_e^3}{2N^3(c\tau_e^2 s + 2N^2)} e^{-\tau_e s}.$$

Note that we have chosen the sign of $L(s)$ to use the same sign convention as in Figure 9.1b. The exponential term representing the time delay gives significant phase above $\omega = 1/\tau_e$, and the gain at the crossover frequency will determine stability.

To check stability, we require that the gain be sufficiently small at crossover. If we assume that the pole due to the queue dynamics is sufficiently fast that the TCP dynamics are dominant, the gain at the crossover frequency ω_c is given by

$$|L(i\omega_c)| = \rho \cdot N \cdot \frac{c^3 \tau_e^3}{2N^3 c\tau_e^2 \omega_c} = \frac{\rho c^2 \tau_e}{2N\omega_c}.$$

Using the Nyquist criterion, the closed loop system will be unstable if this quantity

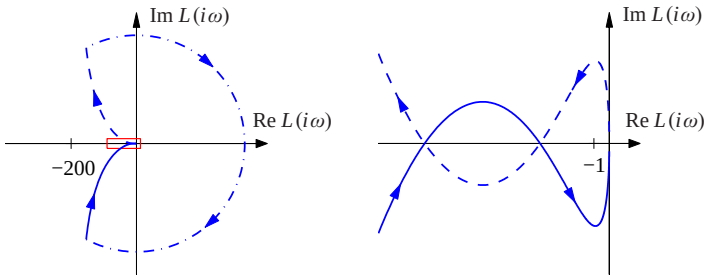


Figure 9.7: Nyquist curve for the loop transfer function $L(s) = \frac{3(s+1)^2}{s(s+6)^2}$. The plot on the right is an enlargement of the box around the origin of the plot on the left. The Nyquist curve intersections the negative real axis twice but has no net encirclements of -1.

is greater than 1. In particular, for a fixed time delay, the system will become unstable as the link capacity c is increased. This indicates that the TCP protocol may not be scalable to high-capacity networks, as pointed out by Low et al. [LPD02]. Exercise 104 provides some ideas of how this might be overcome. ∇

Conditional Stability

Normally, we find that unstable systems can be stabilized simply by reducing the loop gain. There are, however, situations where a system can be stabilized by increasing the gain. This was first encountered by electrical engineers in the design of feedback amplifiers, who coined the term *conditional stability*. The problem was actually a strong motivation for Nyquist to develop his theory. We will illustrate by an example.

Example 9.5 Third-order system
Consider a feedback system with the loop transfer function

$$L(s) = \frac{3(s+6)^2}{s(s+1)^2}. \tag{9.4}$$

The Nyquist plot of the loop transfer function is shown in Figure 9.7. Notice that the Nyquist curve intersects the negative real axis twice. The first intersection occurs at $L = -12$ for $\omega = 2$, and the second at $L = -4.5$ for $\omega = 3$. The intuitive argument based on signal tracing around the loop in Figure 9.1b is strongly misleading in this case. Injection of a sinusoid with frequency 2 rad/s and amplitude 1 at A gives, in steady state, an oscillation at B that is in phase with the input and has amplitude 12. Intuitively it is seems unlikely that closing of the loop will result in a stable system. However, it follows from Nyquist's stability criterion that the system is stable because there are no net encirclements of the critical point. Note, however, that if we *decrease* the gain, then we can get an encirclement, implying that the gain must be sufficiently large for stability. ∇

General Nyquist Criterion

Theorem 9.1 requires that $L(s)$ have no poles in the closed right half-plane. In some situations this is not the case and a more general result is required. Nyquist originally considered this general case, which we summarize as a theorem.

Theorem 9.2 (Nyquist's stability theorem)**.** *Consider a closed loop system with the loop transfer function $L(s)$ that has P poles in the region enclosed by the Nyquist contour. Let N be the net number of clockwise encirclements of -1 by $L(s)$ when s encircles the Nyquist contour Γ in the clockwise direction. The closed loop system then has $Z = N + P$ poles in the right half-plane.*

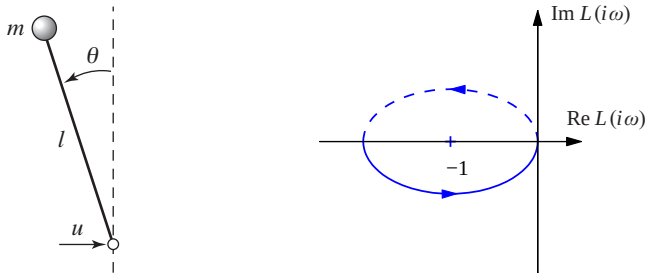
The full Nyquist criterion states that if $L(s)$ has P poles in the right half-plane, then the Nyquist curve for $L(s)$ should have P counterclockwise encirclements of -1 (so that $N = -P$). In particular, this *requires* that $|L(i\omega_c)| > 1$ for some ω_c corresponding to a crossing of the negative real axis. Care has to be taken to get the right sign of the encirclements. The Nyquist contour has to be traversed clockwise, which means that ω moves from $-\infty$ to ∞ and N is positive if the Nyquist curve winds clockwise. If the Nyquist curve winds counterclockwise, then N will be negative (the desired case if $P \neq 0$).

As in the case of the simplified Nyquist criterion, we use small semicircles of radius r to avoid any poles on the imaginary axis. By letting $r \to 0$, we can use Theorem 9.2 to reason about stability. Note that the image of the small semicircles generates a section of the Nyquist curve whose magnitude approaches infinity, requiring care in computing the winding number. When plotting Nyquist curves on the computer, one must be careful to see that such poles are properly handled, and often one must sketch those portions of the Nyquist plot by hand, being careful to loop the right way around the poles.

Example 9.6 Stabilized inverted pendulum
The linearized dynamics of a normalized inverted pendulum can be represented by the transfer function $P(s) = 1/(s^2 - 1)$, where the input is acceleration of the pivot and the output is the pendulum angle θ, as shown in Figure 9.8 (Exercise 85). We attempt to stabilize the pendulum with a proportional-derivative (PD) controller having the transfer function $C(s) = k(s + 2)$. The loop transfer function is

$$L(s) = \frac{k(s+2)}{s^2 - 1}.$$

The Nyquist plot of the loop transfer function is shown in Figure 9.8b. We have $L(0) = -k$ and $L(\infty) = 0$. If $k > 1$, the Nyquist curve encircles the critical point $s = -1$ in the counterclockwise direction when the Nyquist contour γ is encircled in the clockwise direction. The number of encirclements is thus $N = -1$. Since the loop transfer function has one pole in the right half-plane ($P = 1$), we find that $Z = N + P = 0$ and the system is thus stable for $k > 1$. If $k < 1$, there is no encirclement and the closed loop will have one pole in the right half-plane. ∇



(a) Inverted pendulum (b) Nyquist plot

Figure 9.8: PD control of an inverted pendulum. (a) The system consists of a mass that is balanced by applying a force at the pivot point. A proportional-derivative controller with transfer function $C(s) = k(s + 2)$ is used to command u based on θ. (b) A Nyquist plot of the loop transfer function for gain $k = 2$. There is one counterclockwise encirclement of the critical point, giving $N = -1$ clockwise encirclements.

Derivation of Nyquist's Stability Theorem

We will now prove the Nyquist stability theorem for a general loop transfer function $L(s)$. This requires some results from the theory of complex variables, for which the reader can consult Ahlfors [Ahl66]. Since some precision is needed in stating Nyquist's criterion properly, we will use a more mathematical style of presentation. We also follow the mathematical convention of counting encirclements in the counterclockwise direction for the remainder of this section. The key result is the following theorem about functions of complex variables.

Theorem 9.3 (Principle of variation of the argument)**.** *Let D be a closed region in the complex plane and let Γ be the boundary of the region. Assume the function $f : \mathbb{C} \to \mathbb{C}$ is analytic in D and on Γ, except at a finite number of poles and zeros. Then the* winding number w_n *is given by*

$$w_n = \frac{1}{2\pi}\Delta_\Gamma \arg f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f'(z)}{f(z)}dz = Z - P,$$

where Δ_Γ is the net variation in the angle when z traverses the contour Γ in the counterclockwise direction, Z is the number of zeros in D and P is the number of poles in D. Poles and zeros of multiplicity m are counted m times.

Proof. Assume that $z = a$ is a zero of multiplicity m. In the neighborhood of $z = a$ we have

$$f(z) = (z - a)^m g(z),$$

where the function g is analytic and different from zero. The ratio of the derivative of f to itself is then given by

$$\frac{f'(z)}{f(z)} = \frac{m}{z - a} + \frac{g'(z)}{g(z)},$$

and the second term is analytic at $z = a$. The function f'/f thus has a single pole

at $z = a$ with the residue m. The sum of the residues at the zeros of the function is Z. Similarly, we find that the sum of the residues of the poles of is $-P$, and hence

$$Z - P = \frac{1}{2\pi i} \int_\Gamma \frac{f'(z)}{f(z)}\, dz = \frac{1}{2\pi i} \int_\Gamma \frac{d}{dz} \log f(z)\, dz = \frac{1}{2\pi i} \Delta_\Gamma \log f(z),$$

where Δ_Γ again denotes the variation along the contour Γ. We have

$$\log f(z) = \log |f(z)| + i \arg f(z),$$

and since the variation of $|f(z)|$ around a closed contour is zero it follows that

$$\Delta_\Gamma \log f(z) = i\, \Delta_\Gamma \arg f(z),$$

and the theorem is proved. □

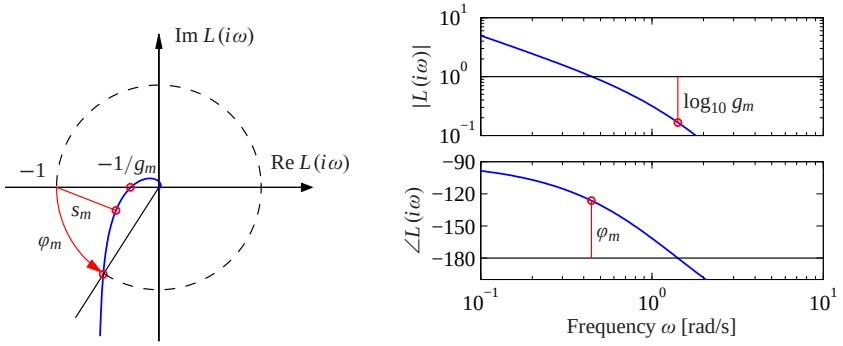
This theorem is useful in determining the number of poles and zeros of a function of complex variables in a given region. By choosing an appropriate closed region D with boundary Γ, we can determine the difference between the number of poles and zeros through computation of the winding number.

Theorem 9.3 can be used to prove Nyquist's stability theorem by choosing Γ as the Nyquist contour shown in Figure 9.3a, which encloses the right half-plane. To construct the contour, we start with part of the imaginary axis $-jR \leq s \leq jR$ and a semicircle to the right with radius R. If the function f has poles on the imaginary axis, we introduce small semicircles with radii r to the right of the poles as shown in the figure. The Nyquist contour is obtained by letting $R \to \infty$ and $r \to 0$. Note that Γ has orientation *opposite* that shown in Figure 9.3a. (The convention in engineering is to traverse the Nyquist contour in the clockwise direction since this corresponds to moving upwards along the imaginary axis, which makes it easy to sketch the Nyquist contour from a Bode plot.)

To see how we use the principle of variation of the argument to compute stability, consider a closed loop system with the loop transfer function $L(s)$. The closed loop poles of the system are the zeros of the function $f(s) = 1 + L(s)$. To find the number of zeros in the right half-plane, we investigate the winding number of the function $f(s) = 1 + L(s)$ as s moves along the Nyquist contour Γ in the *counterclockwise* direction. The winding number can conveniently be determined from the Nyquist plot. A direct application of Theorem 9.3 gives the Nyquist criterion, taking care to flip the orientation. Since the image of $1 + L(s)$ is a shifted version of $L(s)$, we usually state the Nyquist criterion as net encirclements of the -1 point by the image of $L(s)$.

9.3 Stability Margins

In practice it is not enough that a system is stable. There must also be some margins of stability that describe how stable the system is and its robustness to perturbations. There are many ways to express this, but one of the most common is the use of gain and phase margins, inspired by Nyquist's stability criterion. The key idea is that it



(a) Nyquist plot (b) Bode plot

Figure 9.9: Stability margins. The gain margin g_m and phase margin φ_m are shown on the the Nyquist plot (a) and the Bode plot (b). The gain margin corresponds to the smallest increase in gain that creates an encirclement, and the phase margin is the smallest change in phase that creates an encirclement. The Nyquist plot also shows the stability margin s_m, which is the shortest distance to the critical point -1.

is easy to plot the loop transfer function $L(s)$. An increase in controller gain simply expands the Nyquist plot radially. An increase in the phase of the controller twists the Nyquist plot. Hence from the Nyquist plot we can easily pick off the amount of gain or phase that can be added without causing the system to become unstable.

Formally, the *gain margin* g_m of a system is defined as the smallest amount that the open loop gain can be increased before the closed loop system goes unstable. For a system whose phase decreases monotonically as a function of frequency starting at 0°, the gain margin can be computed based on the smallest frequency where the phase of the loop transfer function $L(s)$ is -180°. Let ω_{pc} represent this frequency, called the *phase crossover frequency*. Then the gain margin for the system is given by

$$g_m = \frac{1}{|L(i\omega_{\text{pc}})|}. \tag{9.5}$$

Similarly, the *phase margin* is the amount of phase lag required to reach the stability limit. Let ω_{gc} be the *gain crossover frequency*, the smallest frequency where the loop transfer function $L(s)$ has unit magnitude. Then for a system with monotonically decreasing gain, the phase margin is given by

$$\varphi_m = \pi + \arg L(i\omega_{\text{gc}}). \tag{9.6}$$

These margins have simple geometric interpretations on the Nyquist diagram of the loop transfer function, as shown in Figure 9.9a, where we have plotted the portion of the curve corresponding to $\omega > 0$. The gain margin is given by the inverse of the distance to the nearest point between -1 and 0 where the loop transfer function crosses the negative real axis. The phase margin is given by the smallest angle on the unit circle between -1 and the loop transfer function. When the gain or phase is monotonic, this geometric interpretation agrees with the formulas above.

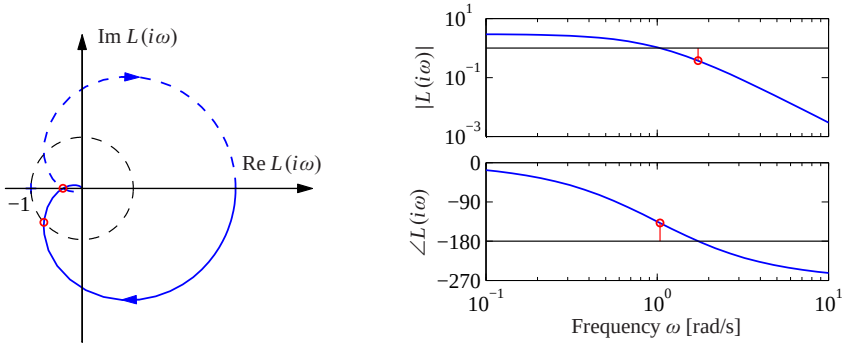


Figure 9.10: Stability margins for a third-order transfer function. The Nyquist plot on the left allows the gain, phase and stability margins to be determined by measuring the distances of relevant features. The gain and phase margins can also be read off of the Bode plot on the right.

A drawback with gain and phase margins is that it is necessary to give both of them in order to guarantee that the Nyquist curve is not close to the critical point. An alternative way to express margins is by a single number, the *stability margin* s_m, which is the shortest distance from the Nyquist curve to the critical point. This number is related to disturbance attenuation, as will be discussed in Section 11.3.

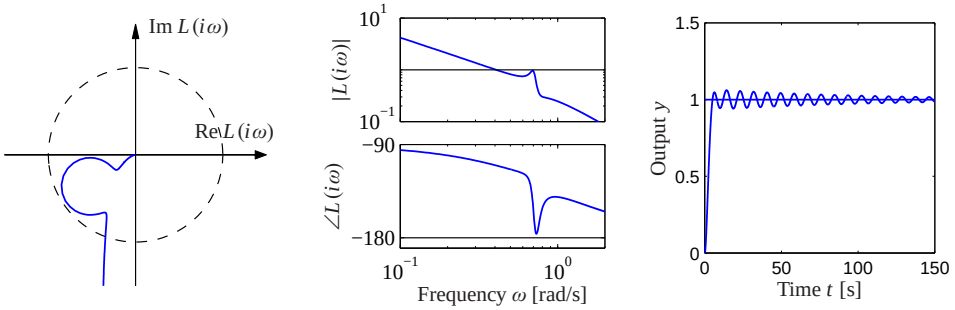
For many systems, the gain and phase margins can be determined from the Bode plot of the loop transfer function. To find the gain margin we first find the phase crossover frequency ω_{pc} where the phase is -180°. The gain margin is the inverse of the gain at that frequency. To determine the phase margin we first determine the gain crossover frequency ω_{gc}, i.e., the frequency where the gain of the loop transfer function is 1. The phase margin is the phase of the loop transfer function at that frequency plus 180°. Figure 9.9b illustrates how the margins are found in the Bode plot of the loop transfer function. Note that the Bode plot interpretation of the gain and phase margins can be incorrect if there are multiple frequencies at which the gain is equal to 1 or the phase is equal to -180°.

Example 9.7 Third-order system
Consider a loop transfer function $L(s) = 3/(s+1)^3$. The Nyquist and Bode plots are shown in Figure 9.10. To compute the gain, phase and stability margins, we can use the Nyquist plot shown in Figure 9.10. This yields the following values:

$$g_m = 2.67, \qquad \varphi_m = 41.7^\circ, \qquad s_m = 0.464.$$

The gain and phase margins can also be determined from the Bode plot. ∇

The gain and phase margins are classical robustness measures that have been used for a long time in control system design. The gain margin is well defined if the Nyquist curve intersects the negative real axis once. Analogously, the phase margin is well defined if the Nyquist curve intersects the unit circle at only one point. Other more general robustness measures will be introduced in Chapter 12.



(a) (b) (c)

Figure 9.11: System with good gain and phase margins but a poor stability margin. Nyquist (a) and Bode (b) plots of the loop transfer function and step response (c) for a system with good gain and phase margins but with a poor stability margin. The Nyquist plot shows on the portion of the curve corresponding to $\omega > 0$.

Even if both the gain and phase margins are reasonable, the system may still not be robust, as is illustrated by the following example.

Example 9.8 Good gain and phase margins but poor stability margins
Consider a system with the loop transfer function

$$L(s) = \frac{0.38(s^2 + 0.1s + 0.55)}{s(s+1)(s^2 + 0.06s + 0.5)}.$$

A numerical calculation gives the gain margin as $g_m = 266$, and the phase margin is 70°. These values indicate that the system is robust, but the Nyquist curve is still close to the critical point, as shown in Figure 9.11. The stability margin is $s_m = 0.27$, which is very low. The closed loop system has two resonant modes, one with damping ratio $\zeta = 0.81$ and the other with $\zeta = 0.014$. The step response of the system is highly oscillatory, as shown in Figure 9.11c. ∇

The stability margin cannot easily be found from the Bode plot of the loop transfer function. There are, however, other Bode plots that will give s_m; these will be discussed in Chapter 12. In general, it is best to use the Nyquist plot to check stability since this provides more complete information than the Bode plot.

When designing feedback systems, it will often be useful to define the robustness of the system using gain, phase and stability margins. These numbers tell us how much the system can vary from our nominal model and still be stable. Reasonable values of the margins are phase margin $\varphi_m = 30^\circ$–60°, gain margin $g_m = 2$–5 and stability margin $s_m = 0.5$–0.8.

There are also other stability measures, such as the *delay margin*, which is the smallest time delay required to make the system unstable. For loop transfer functions that decay quickly, the delay margin is closely related to the phase margin, but for systems where the gain curve of the loop transfer function has several peaks at high frequencies, the delay margin is a more relevant measure.

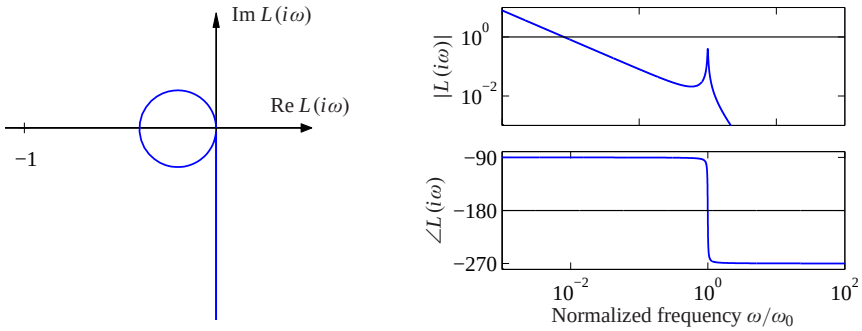


Figure 9.12: Nyquist and Bode plots of the loop transfer function for the AFM system (9.7) with an integral controller. The frequency in the Bode plot is normalized by a. The parameters are $\zeta = 0.01$ and $k_i = 0.008$.

Example 9.9 Nanopositioning system for an atomic force microscope
Consider the system for horizontal positioning of the sample in an atomic force microscope. The system has oscillatory dynamics, and a simple model is a spring–mass system with low damping. The normalized transfer function is given by

$$P(s) = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}, \tag{9.7}$$

where the damping ratio typically is a very small number, e.g., $\zeta = 0.1$.

We will start with a controller that has only integral action. The resulting loop transfer function is

$$L(s) = \frac{k_i\omega_0^2}{s(s^2 + 2\zeta\omega_0 s + \omega_0^2)},$$

where k_i is the gain of the controller. Nyquist and Bode plots of the loop transfer function are shown in Figure 9.12. Notice that the part of the Nyquist curve that is close to the critical point -1 is approximately circular.

From the Bode plot in Figure 9.12b, we see that the phase crossover frequency is $\omega_{pc} = a$, which will be independent of the gain k_i. Evaluating the loop transfer function at this frequency, we have $L(i\omega_0) = -k_i/(2\zeta\omega_0)$, which means that the gain margin is $g_m = 1 - k_i/(2\zeta\omega_0)$. To have a desired gain margin of g_m the integral gain should be chosen as

$$k_i = 2\omega_0\zeta(1 - g_m).$$

Figure 9.12 shows Nyquist and Bode plots for the system with gain margin $g_m = 1.67$ and stability margin $s_m = 0.597$. The gain curve in the Bode plot is almost a straight line for low frequencies and has a resonant peak at $\omega = \omega_0$. The gain crossover frequency is approximately equal to k_i. The phase decreases monotonically from -90° to -270°: it is equal to -180° at $\omega = \omega_0$. The curve can be shifted vertically by changing k_i: increasing k_i shifts the gain curve upward and increases the gain crossover frequency. Since the phase is -180° at the resonant peak, it is necessary that the peak not touch the line $|L(i\omega)| = 1$. ∇

9.4 Bode's Relations and Minimum Phase Systems

An analysis of Bode plots reveals that there appears to be a relation between the gain curve and the phase curve. Consider, for example, the Bode plots for the differentiator and the integrator (shown in Figure 8.12). For the differentiator the slope is $+1$ and the phase is a constant $\pi/2$ radians. For the integrator the slope is -1 and the phase is $-\pi/2$. For the first-order system $G(s) = s + a$, the amplitude curve has the slope 0 for small frequencies and the slope $+1$ for high frequencies, and the phase is 0 for low frequencies and $\pi/2$ for high frequencies.

Bode investigated the relations between the curves for systems with no poles and zeros in the right half-plane. He found that the phase was uniquely given by the shape of the gain curve, and vice versa:

$$\arg G(i\omega_0) = \frac{\pi}{2}\int_0^\infty f(\omega)\frac{d\log|G(i\omega)|}{d\log\omega}\,d\log\omega \approx \frac{\pi}{2}\frac{d\log|G(i\omega)|}{d\log\omega}, \qquad (9.8)$$

where f is the weighting kernel

$$f(\omega) = \frac{2}{\pi^2}\log\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right|.$$

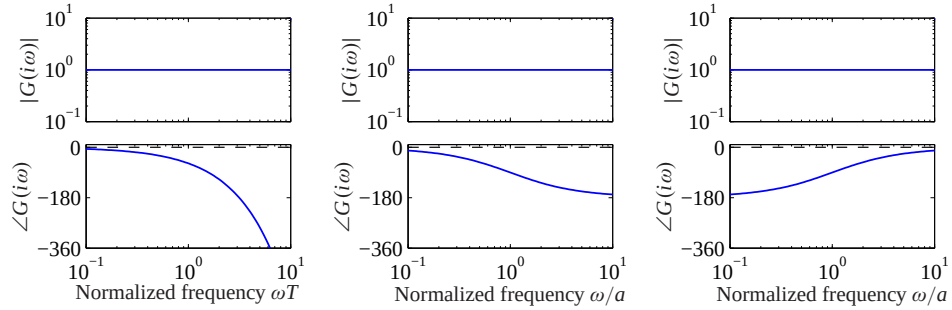
The phase curve is thus a weighted average of the derivative of the gain curve. If the gain curve has constant slope n, the phase curve has constant value $n\pi/2$.

Bode's relations (9.8) hold for systems that do not have poles and zeros in the right half-plane. Such systems are called *minimum phase systems* because systems with poles and zeros in the right half-plane have a larger phase lag. The distinction is important in practice because minimum phase systems are easier to control than systems with a larger phase lag. We will now give a few examples of nonminimum phase transfer functions.

The transfer function of a time delay of τ units is $G(s) = e^{-s\tau}$. This transfer function has unit gain $|G(i\omega)| = 1$, and the phase is $\arg G(i\omega) = -\omega\tau$. The corresponding minimum phase system with unit gain has the transfer function $G(s) = 1$. The time delay thus has an additional phase lag of $\omega\tau$. Notice that the phase lag increases linearly with frequency. Figure 9.13a shows the Bode plot of the transfer function. (Because we use a log scale for frequency, the phase falls off exponentially in the plot.)

Consider a system with the transfer function $G(s) = (a - s)/(a + s)$ with $a > 0$, which has a zero $s = a$ in the right half-plane. The transfer function has unit gain $|G(i\omega)| = 1$, and the phase is $\arg G(i\omega) = -2\arctan(\omega/a)$. The corresponding minimum phase system with unit gain has the transfer function $G(s) = 1$. Figure 9.13b shows the Bode plot of the transfer function. A similar analysis of the transfer function $G(s) = (s + a)/s - a)$ with $a > 0$, which has a pole in the right half-plane, shows that its phase is $\arg G(i\omega) = -2\arctan(a/\omega)$. The Bode plot is shown in Figure 9.13c.

The presence of poles and zeros in the right half-plane imposes severe limitations on the achievable performance. Dynamics of this type should be avoided by redesign of the system whenever possible. While the poles are intrinsic properties of the



(a) Time delay (b) RHP zero (c) RHP pole

Figure 9.13: Bode plots of systems that are not minimum phase. (a) Time delay $G(s) = e^{-sT}$, (b) system with a right half-plane (RHP) zero $G(s) = (a - s)/(a + s)$ and (c) system with right half-plane pole. The corresponding minimum phase system has the transfer function $G(s) = 1$ in all cases, the phase curves for that system are shown as dashed lines.

system and they do not depend on sensors and actuators, the zeros depend on how inputs and outputs of a system are coupled to the states. Zeros can thus be changed by moving sensors and actuators or by introducing new sensors and actuators. Nonminimum phase systems are unfortunately quite common in practice.

The following example gives a system theoretic interpretation of the common experience that it is more difficult to drive in reverse gear and illustrates some of the properties of transfer functions in terms of their poles and zeros.

Example 9.10 Vehicle steering

The nonnormalized transfer function from steering angle to lateral velocity for the simple vehicle model is

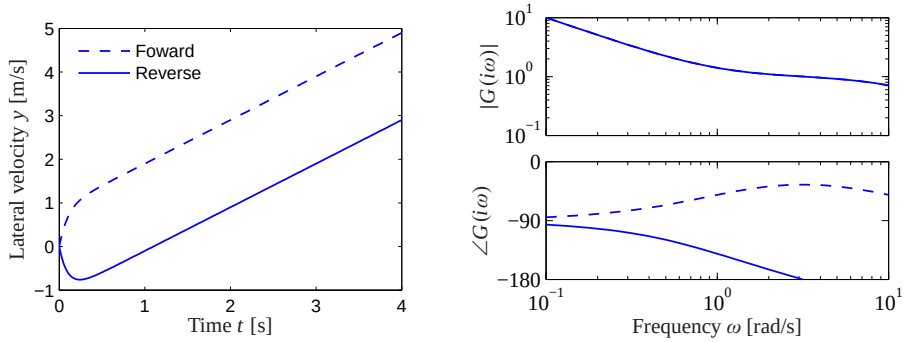
$$G(s) = \frac{av_0 s + v_0^2}{bs},$$

where v_0 is the velocity of the vehicle and $a, b > 0$ (see Example 5.12). The transfer function has a zero at $s = v_0/a$. In normal driving this zero is in the left half-plane, but it is in the right half-plane when driving in reverse, $v_0 < 0$. The unit step response is

$$y(t) = \frac{av_0}{b} + \frac{av_0^2 t}{b}.$$

The lateral velocity thus responds immediately to a steering command. For reverse steering v_0 is negative and the initial response is in the wrong direction, a behavior that is representative for nonminimum phase systems (called an *inverse response*).

Figure 9.14 shows the step response for forward and reverse driving. In this simulation we have added an extra pole with the time constant T to approximately account for the dynamics in the steering system. The parameters are $a = b = 1$, $T = 0.1$, $v_0 = 1$ for forward driving and $v_0 = -1$ for reverse driving. Notice that for $t > t_0 = a/v_0$, where t_0 is the time required to drive the distance a, the step response for reverse driving is that of forward driving with the time delay t_0. The



(a) Step response (b) Frequency response

Figure 9.14: Vehicle steering for driving in reverse. (a) Step responses from steering angle to lateral translation for a simple kinematics model when driving forward (dashed) and reverse (solid). With rear-wheel steering the center of mass first moves in the wrong direction and that the overall response with rear-wheel steering is significantly delayed compared with that for front-wheel steering. (b) Frequency response for driving forward (dashed) and reverse (solid). Notice that the gain curves are identical, but the phase curve for driving in reverse has nonminimum phase.

position of the zero v_0/a depends on the location of the sensor. In our calculation we have assumed that the sensor is at the center of mass. The zero in the transfer function disappears if the sensor is located at the rear wheel. The difficulty with zeros in the right half-plane can thus be visualized by a thought experiment where we drive a car in forward and reverse and observe the lateral position through a hole in the floor of the car. ∇

9.5 Generalized Notions of Gain and Phase

A key idea in frequency domain analysis is to trace the behavior of sinusoidal signals through a system. The concepts of gain and phase represented by the transfer function are strongly intuitive because they describe amplitude and phase relations between input and output. In this section we will see how to extend the concepts of gain and phase to more general systems, including some nonlinear systems. We will also show that there are analogs of Nyquist's stability criterion if signals are approximately sinusoidal.

System Gain

We begin by considering the case of a static linear system $y = Au$, where A is a matrix whose elements are complex numbers. The matrix does not have to be square. Let the inputs and outputs be vectors whose elements are complex numbers and use the Euclidean norm

$$\|u\| = \sqrt{\Sigma |u_i|^2}. \tag{9.9}$$

The norm of the output is

$$\|y\|^2 = u^* A^* A u,$$

where $*$ denotes the complex conjugate transpose. The matrix A^*A is symmetric and positive semidefinite, and the right-hand side is a quadratic form. The square root of eigenvalues of the matrix A^*A are all real, and we have

$$\|y\|^2 \leq \lambda_{\max}(A^*A)\|u\|^2.$$

The gain of the system can then be defined as the maximum ratio of the output to the input over all possible inputs:

$$\gamma = \max_u \frac{\|y\|}{\|u\|} = \sqrt{\lambda_{\max}(A^*A)}. \tag{9.10}$$

The square root of the eigenvalues of the matrix A^*A are called the *singular values* of the matrix A, and the largest singular value is denoted $\bar{\sigma}(A)$.

To generalize this to the case of an input/output dynamical system, we need to think of the inputs and outputs not as vectors of real numbers but as vectors of *signals*. For simplicity, consider first the case of scalar signals and let the signal space L_2 be square-integrable functions with the norm

$$\|u\|_2 = \sqrt{\int_0^\infty |u|^2(\tau)\, d\tau}.$$

This definition can be generalized to vector signals by replacing the absolute value with the vector norm (9.9). We can now formally define the gain of a system taking inputs $u \in L_2$ and producing outputs $y \in L_2$ as

$$\gamma = \sup_{u \in L_2} \frac{\|y\|}{\|u\|}, \tag{9.11}$$

where sup is the *supremum*, defined as the smallest number that is larger than its argument. The reason for using the supremum is that the maximum may not be defined for $u \in L_2$. This definition of the system gain is quite general and can even be used for some classes of nonlinear systems, though one needs to be careful about how initial conditions and global nonlinearities are handled.

The norm (9.11) has some nice properties in the case of linear systems. In particular, given a single-input, single-output stable linear system with transfer function $G(s)$, it can be shown that the norm of the system is given by

$$\gamma = \sup_\omega |G(i\omega)| =: \|G\|_\infty. \tag{9.12}$$

In other words, the gain of the system corresponds to the peak value of the frequency response. This corresponds to our intuition that an input produces the largest output when we are at the resonant frequencies of the system. $\|G\|_\infty$ is called the *infinity norm* of the transfer function $G(s)$.

This notion of gain can be generalized to the multi-input, multi-output case as well. For a linear multivariable system with a real rational transfer function matrix

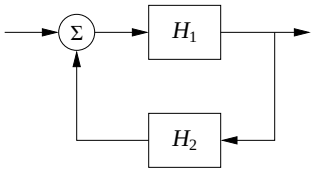


Figure 9.15: A feedback connection of two general nonlinear systems H_1 and H_2. The stability of the system can be explored using the small gain theorem.

$G(s)$ we can define the gain as

$$\gamma = \|G\|_\infty = \sup_\omega \bar{\sigma}(G(i\omega)). \tag{9.13}$$

Thus we can combine the idea of the gain of a matrix with the idea of the gain of a linear system by looking at the maximum singular value over all frequencies.

Small Gain and Passivity

For linear systems it follows from Nyquist's theorem that the closed loop is stable if the gain of the loop transfer function is less than 1 for all frequencies. This result can be extended to a larger class of systems by using the concept of the system gain defined in equation (9.11).

Theorem 9.4 (Small gain theorem)**.** *Consider the closed loop system shown in Figure 9.15, where H_1 and H_2 are stable systems and the signal spaces are properly defined. Let the gains of the systems H_1 and H_2 be γ_1 and γ_2. Then the closed loop system is input/output stable if $\gamma_1\gamma_2 < 1$, and the gain of the closed loop system is*

$$\gamma = \frac{\gamma_1}{1 - \gamma_1\gamma_2}.$$

Notice that if systems H_1 and H_2 are linear, it follows from the Nyquist stability theorem that the closed loop is stable because if $\gamma_1\gamma_2 < 1$, the Nyquist curve is always inside the unit circle. The small gain theorem is thus an extension of the Nyquist stability theorem.

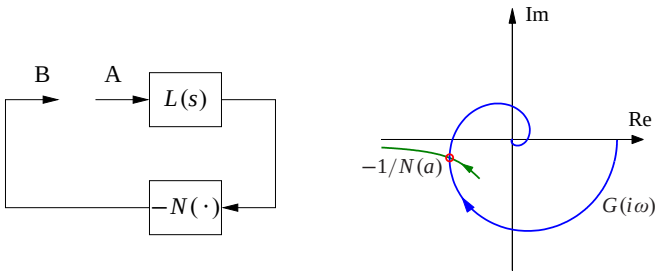
Although we have focused on linear systems, the small gain theorem also holds for nonlinear input/output systems. The definition of gain in equation (9.11) holds for nonlinear systems as well, with some care needed in handling the initial condition.

The main limitation of the small gain theorem is that it does not consider the phasing of signals around the loop, so it can be very conservative. To define the notion of phase we require that there be a scalar product. For square-integrable functions this can be defined as

$$\langle u, y \rangle = \int_0^\infty u(\tau)y(\tau)\,d\tau.$$

The phase φ between two signals can now be defined as

$$\langle u, y \rangle = \|u\|\|y\|\cos(\varphi).$$



(a) Block diagram (b) Nyquist plot

Figure 9.16: Describing function analysis. A feedback connection between a static nonlinearity and a linear system is shown in (a). The linear system is characterized by its transfer function $L(s)$, which depends on frequency, and the nonlinearity by its describing function $N(a)$, which depends on the amplitude a of its input. The Nyquist plot of $L(i\omega)$ and the plot of the $-1/N(a)$ are shown in (b). The intersection of the curves represents a possible limit cycle.

Systems where the phase between inputs and outputs is 90° or less for all inputs are called *passive systems*. It follows from the Nyquist stability theorem that a closed loop linear system is stable if the phase of the loop transfer function is between $-\pi$ and π. This result can be extended to nonlinear systems as well. It is called the *passivity theorem* and is closely related to the small gain theorem. See Khalil [Kha01] for a more detailed description.

Additional applications of the small gain theorem and its application to robust stability are given in Chapter 12.

Describing Functions

For special nonlinear systems like the one shown in Figure 9.16a, which consists of a feedback connection between a linear system and a static nonlinearity, it is possible to obtain a generalization of Nyquist's stability criterion based on the idea of *describing functions*. Following the approach of the Nyquist stability condition, we will investigate the conditions for maintaining an oscillation in the system. If the linear subsystem has low-pass character, its output is approximately sinusoidal even if its input is highly irregular. The condition for oscillation can then be found by exploring the propagation of a sinusoid that corresponds to the first harmonic.

To carry out this analysis, we have to analyze how a sinusoidal signal propagates through a static nonlinear system. In particular we investigate how the first harmonic of the output of the nonlinearity is related to its (sinusoidal) input. Letting F represent the nonlinear function, we expand $F(e^{i\omega t})$ in terms of its harmonics:

$$F(ae^{i\omega t}) = \sum_{n=0}^{\infty} M_n(a)e^{i(n\omega t+\varphi_n(a))},$$

where $M_n(a)$ and $\varphi_n(a)$ represent the gain and phase of the nth harmonic, which depend on the input amplitude since the function F is nonlinear. We define the

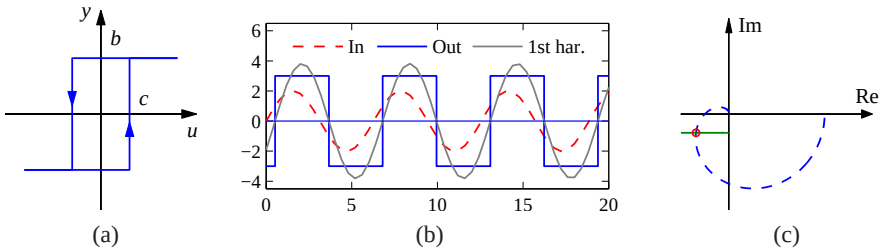


Figure 9.17: Describing function analysis for a relay with hysteresis. The input/output relation of the hysteresis is shown in (a) and the input with amplitude $a = 2$, the output and its first harmonic are shown in (b). The Nyquist plots of the transfer function $L(s) = (s+1)^{-4}$ and the negative of the inverse describing function for the relay with $b = 3$ and $c = 1$ are shown in (c).

describing function to be the complex gain of the first harmonic:

$$N(a) = M_1(a)e^{i\varphi_n(a)}. \tag{9.14}$$

The function can also be computed by assuming that the input is a sinusoid and using the first term in the Fourier series of the resulting output.

Arguing as we did when deriving Nyquist's stability criterion, we find that an oscillation can be maintained if

$$L(i\omega)N(a) = -1. \tag{9.15}$$

This equation means that if we inject a sinusoid at A in Figure 9.16, the same signal will appear at B and an oscillation can be maintained by connecting the points. Equation (9.15) gives two conditions for finding the frequency ω of the oscillation and its amplitude a: the phase must be 180°, and the magnitude must be unity. A convenient way to solve the equation is to plot $L(i\omega)$ and $-1/N(a)$ on the same diagram as shown in Figure 9.16b. The diagram is similar to the Nyquist plot where the critical point -1 is replaced by the curve $-1/N(a)$ and a ranges from 0 to ∞.

It is possible to define describing functions for types of inputs other than sinusoids. Describing function analysis is a simple method, but it is approximate because it assumes that higher harmonics can be neglected. Excellent treatments of describing function techniques can be found in the texts by Atherton [Ath75] and Graham and McRuer [GM61].

Example 9.11 Relay with hysteresis

Consider a linear system with a nonlinearity consisting of a relay with hysteresis. The output has amplitude b and the relay switches when the input is $\pm c$, as shown in Figure 9.17a. Assuming that the input is $u = a\sin(\omega t)$, we find that the output is zero if $a \leq c$, and if $a > c$, the output is a square wave with amplitude b that switches at times $\omega t = \arcsin(c/a) + n\pi$. The first harmonic is then $y(t) = (4b/\pi)\sin(\omega t - \alpha)$, where $\sin\alpha = c/a$. For $a > c$ the describing function and its inverse are

$$N(a) = \frac{4b}{a\pi}\left(\sqrt{1 - \frac{c^2}{a^2}} - i\frac{c}{a}\right), \qquad \frac{1}{N(a)} = \frac{\pi\sqrt{a^2 - c^2}}{4b} + i\frac{\pi c}{4b},$$

where the inverse is obtained after simple calculations. Figure 9.17b shows the response of the relay to a sinusoidal input with the first harmonic of the output shown as a dashed line. Describing function analysis is illustrated in Figure 9.17c, which shows the Nyquist plot of the transfer function $L(s) = 2/(s+1)^4$ (dashed line) and the negative inverse describing function of a relay with $b = 1$ and $c = 0.5$. The curves intersect for $a = 1$ and $\omega = 0.77$ rad/s, indicating the amplitude and frequency for a possible oscillation if the process and the relay are connected in a a feedback loop. ∇

9.6 Further Reading

Nyquist's original paper giving his now famous stability criterion was published in the *Bell Systems Technical Journal* in 1932 [Nyq32]. More accessible versions are found in the book [BK64], which also includes other interesting early papers on control. Nyquist's paper is also reprinted in an IEEE collection of seminal papers on control [Bas01]. Nyquist used $+1$ as the critical point, but Bode changed it to -1, which is now the standard notation. Interesting perspectives on early developments are given by Black [Bla77], Bode [Bod60] and Bennett [Ben93]. Nyquist did a direct calculation based on his insight into the propagation of sinusoidal signals through systems; he did not use results from the theory of complex functions. The idea that a short proof can be given by using the principle of variation of the argument is presented in the delightful book by MacColl [Mac45]. Bode made extensive use of complex function theory in his book [Bod45], which laid the foundation for frequency response analysis where the notion of minimum phase was treated in detail. A good source for complex function theory is the classic by Ahlfors [Ahl66]. Frequency response analysis was a key element in the emergence of control theory as described in the early texts by James et al. [JNP47], Brown and Campbell [BC48] and Oldenburger [Old56], and it became one of the cornerstones of early control theory. Frequency response underwent a resurgence when robust control emerged in the 1980s, as will be discussed in Chapter 12.

Exercises

98 (Operational amplifier) Consider an op amp circuit with $Z_1 = Z_2$ that gives a closed loop system with nominally unit gain. Let the transfer function of the operational amplifier be

$$G(s) = \frac{ka_1a_2}{(s+a)(s+a_1)(s+a_2)},$$

where $a_1, a_2 \gg a$. Show that the condition for oscillation is $k < a_1 + a_2$ and compute the gain margin of the system. Hint: Assume $a = 0$.

99 (Atomic force microscope) The dynamics of the tapping mode of an atomic force microscope are dominated by the damping of the cantilever vibrations and the system that averages the vibrations. Modeling the cantilever as a spring–mass

system with low damping, we find that the amplitude of the vibrations decays as $\exp(-\zeta\omega t)$, where ζ is the damping ratio and ω is the undamped natural frequency of the cantilever. The cantilever dynamics can thus be modeled by the transfer function

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

where $a = \zeta\omega_0$. The averaging process can be modeled by the input/output relation

$$y(t) = \frac{1}{\tau}\int_{t-\tau}^{t} u(v)dv,$$

where the averaging time is a multiple n of the period of the oscillation $2\pi/\omega$. The dynamics of the piezo scanner can be neglected in the first approximation because they are typically much faster than a. A simple model for the complete system is thus given by the transfer function

$$P(s) = \frac{a(1-e^{-s\tau})}{s\tau(s+a)}.$$

Plot the Nyquist curve of the system and determine the gain of a proportional controller that brings the system to the boundary of stability.

100 (Heat conduction) A simple model for heat conduction in a solid is given by the transfer function

$$P(s) = ke^{-\sqrt{s}}.$$

Sketch the Nyquist plot of the system. Determine the frequency where the phase of the process is -180° and the gain at that frequency. Show that the gain required to bring the system to the stability boundary is $k = e^{\pi}$.

101 (Vectored thrust aircraft) Consider the state space controller designed for the vectored thrust aircraft in Examples 6.8 and 7.5. The controller consists of two components: an optimal estimator to compute the state of the system from the output and a state feedback compensator that computes the input given the (estimated) state. Compute the loop transfer function for the system and determine the gain, phase and stability margins for the closed loop dynamics.

102 (Vehicle steering) Consider the linearized model for vehicle steering with a controller based on state feedback discussed in Example 7.4. The transfer functions for the process and controller are given by

$$P(s) = \frac{\gamma s + 1}{s^2}, \quad C(s) = \frac{s(k_1 l_1 + k_2 l_2) + k_1 l_2}{s^2 + s(\gamma k_1 + k_2 + l_1) + k_1 + l_2 + k_2 l_1 - \gamma k_2 l_2},$$

as computed in Example 8.6. Let the process parameter be $\gamma = 0.5$ and assume that the state feedback gains are $k_1 = 1$ and $k_2 = 0.914$ and that the observer gains are $l_1 = 2.828$ and $l_2 = 4$. Compute the stability margins numerically.

103 (Stability margins for second-order systems) A process whose dynamics is described by a double integrator is controlled by an ideal PD controller with the

transfer function $C(s) = k_d s + k_p$, where the gains are $k_d = 2\zeta\omega_0$ and $k_p = \omega_0^2$. Calculate and plot the gain, phase and stability margins as a function ζ.

104 (Congestion control in overload conditions) A strongly simplified flow model of a TCP loop under overload conditions is given by the loop transfer function

$$L(s) = \frac{k}{s} e^{-s\tau},$$

where the queuing dynamics are modeled by an integrator, the TCP window control is a time delay τ and the controller is simply a proportional controller. A major difficulty is that the time delay may change significantly during the operation of the system. Show that if we can measure the time delay, it is possible to choose a gain that gives a stability margin of $s_n \geq 0.6$ for all time delays τ.

105 (Bode's formula) Consider Bode's formula (9.8) for the relation between gain and phase for a transfer function that has all its singularities in the left half-plane. Plot the weighting function and make an assessment of the frequencies where the approximation $\arg G \approx (\pi/2) d \log |G| / d \log \omega$ is valid.

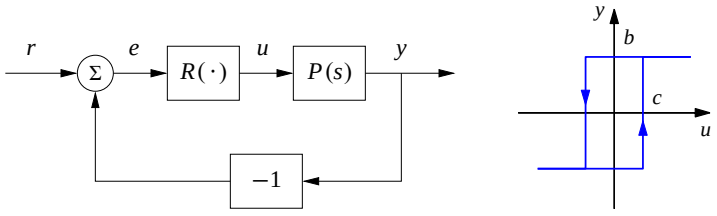
106 (Padé approximation to a time delay) Consider the transfer functions

$$G_1(s) = e^{-s\tau}, \qquad G_2(s) = e^{-s\tau} \approx \frac{1 - s\tau/2}{1 + s\tau/2}. \tag{9.16}$$

Show that the minimum phase properties of the transfer functions are similar for frequencies $\omega < 1/\tau$. A long time delay τ is thus equivalent to a small right half-plane zero. The approximation (9.16) is called a first-order *Padé approximation*.

107 (Inverse response) Consider a system whose input/output response is modeled by $G(s) = 6(-s+1)/(s^2+5s+6)$, which has a zero in the right half-plane. Compute the step response for the system, and show that the output goes in the wrong direction initially, which is also referred to as an *inverse response*. Compare the response to a minimum phase system by replacing the zero at $s = 1$ with a zero at $s = -1$.

108 (Describing function analysis) . Consider the system with the block diagram shown on the left below.



The block R is a relay with hysteresis whose input/output response is shown on the right and the process transfer function is $P(s) = e^{-s\tau}/s$. Use describing function analysis to determine frequency and amplitude of possible limit cycles. Simulate the system and compare with the results of the describing function analysis.