Chapter 9

Loop Analysis

Regeneration or feed-back is of considerable importance in many applications of vacuum tubes. The most obvious example is that of vacuum tube oscillators, where the feed-back is carried beyond the singing point. Another application is the 21-circuit test of balance, in which the current due to the unbalance between two impedances is fed back, the gain being increased until singing occurs. Still other applications are cases where portions of the output current of amplifiers are fed back to the input either unintentionally or by design. For the purpose of investigating the stability of such devices they may be looked on as amplifiers whose output is connected to the input through a transducer. This paper deals with the theory of stability of such systems.

Abstract for "Regeneration Theory", Harry Nyquist, 1932 [Nyq32].

In this chapter we study how how stability and robustness of closed loop systems can be determined by investigating how signals propagate around the feedback loop. 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

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 signal grows or decays around the loop. 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,

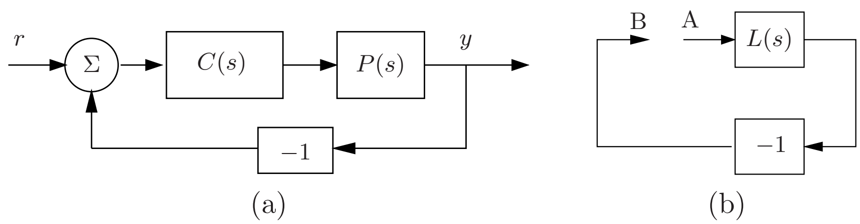

Figure 9.1: Block diagram of a (a) simple feedback system with (b) the loop opened at AB.

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 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

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

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 all 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,*

$$L = PC,$$

which is the transfer function obtained by breaking the feedback loop, as shown in Figure 9.1. The loop transfer function is simply the transfer function from the input at position A to the output at position B.

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 could then connect A to B. Tracing signals around the loop we find that the signals at A and B are identical if

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

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

One of the powerful concepts embedded in Nyquist's approach to stability analysis is that it allows us to determine the stability of the feedback system by looking at properties of the open loop transfer function. This idea will turn out to be very important in how we approach designing transfer functions.

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 showing how it can be used to ascertain stability.

The Nyquist Plot

The frequency response of the loop transfer function can be represented by plotting the complex number $L(j\omega)$ as a function of ω. Such a plot is called a *Nyquist plot* and the curve is called a *Nyquist curve*. An example of a Nyquist plot is given in Figure 9.2. The magnitude $|L(j\omega)|$ is called the *loop gain* because it tells how much the signal is amplified as is passes around the feedback loop.

The condition for oscillation given in equation (9.1) implies that the Nyquist curve of the loop transfer function goes through the point $L = -1$, which is called the *critical point*. Intuitively it seems reasonable that the system is stable if $|L(j\omega_c)| < 1$, which means that the critical point -1 is on the left hand side of the Nyquist curve, as indicated in Figure 9.2.

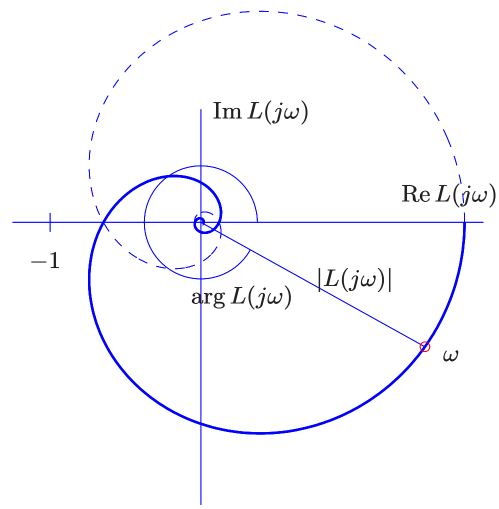

Figure 9.2: Nyquist plot of the transfer function $L(s) = 1.4e^{-s}/(s+1)^2$. The gain and phase at the frequency ω are $g = |L(j\omega)|$ and $\varphi = \arg L(j\omega)$.

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 requires a proper mathematical analysis to clear up, and which we defer until the next section. For now we consider a simplified case, when the loop transfer function is stable.

For loop transfer functions that do not have poles in the right half plane, the precise stability condition is that the complete Nyquist plot does not encircle the critical point -1. The complete Nyquist plot is obtained by adding the plot for negative frequencies shown in the dashed curve in Figure 9.2. This plot is the mirror image of the Nyquist curve about the real axis.

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

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 to the Nyquist plot. Let the end of the string attached to the Nyquist curve traverse the whole curve. There are no encirclements if the cord does not wind up on the pin when the curve is encircled.

Example 9.1 (Cruise control)**.** Consider the speed control system introduced in Section 3.1 and analyzed using state space techniques in Example 6.9. In this example, we study the stability of the system using the Nyquist criterion.

The linearized dynamics around the equilibrium speed v_e and throttle position u_e are given by

$$\dot{\tilde{v}} = a\tilde{v} - g\theta + b\tilde{u}$$
$$y = v = \tilde{v} + v_e,$$

where $\tilde{v} = v - v_e$, $\tilde{u} = u - u_e$, m is the mass of the car and θ is the angle of the road. The constant $a < 0$ depends on the throttle characteristic and is given in Example 5.10.

The transfer function from throttle to speed is given by

$$P(s) = G_{yu}(s) = \frac{b}{s - \alpha}.$$

We consider a controller that is a modified version of the proportional-integral (PI) controller given previously. Assume that the transfer function of the controller is

$$C(s) = G_{ue}(s) = k_p + \frac{k_i}{s + \beta} = \frac{k_p s + k_i + k_p\beta}{s + \beta}$$

giving a loop transfer function of

$$L(s) = b\frac{k_p s + k_i + k_p\beta}{(s + a)(s + \beta)}.$$

The Nyquist plot for the system, using $a = 0.0101$, $b = 1.3203$, $k_p = 0.5$, $k_i = 0.1$ and $\beta = 0.1$, is shown in Figure 9.3. We see from the Nyquist plot that the closed loop system is stable, since there are no net encirclements of the -1 point. ∇

One nice property of the Nyquist stability criterion is that it can be applied to infinite dimensional systems, as is illustrated by the following example.

Example 9.2 (Heat conduction)**.** Consider a temperature control system where the heat conduction process has the transfer function

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

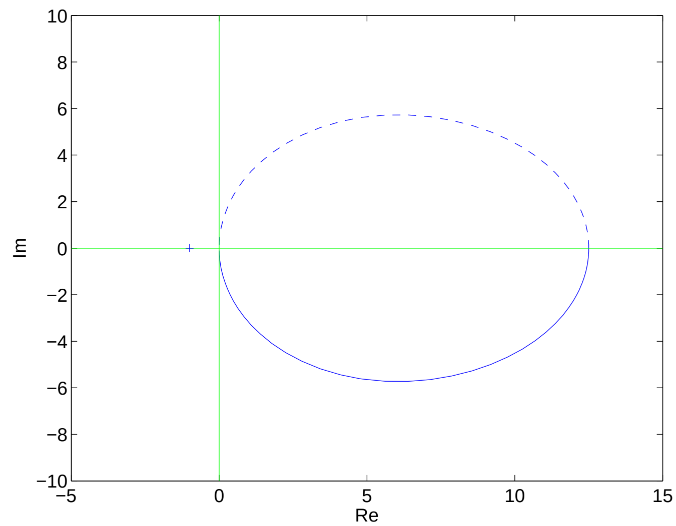

Figure 9.3: Nyquist plot for the speed control system.

and the controller is a proportional controller with gain k. The loop transfer function is $L(s) = ke^{-\sqrt{s}}$ and its Nyquist plot for $k = 1$ is shown in Figure 9.4. To compute the stability condition for the system as a function of the gain k, we analyze the transfer function a bit more carefully. We have

$$P(j\omega) = e^{-\sqrt{j\omega}} = e^{-\sqrt{\omega/2} - i\sqrt{\omega/2}}$$

and hence

$$\log P(j\omega) = -\sqrt{j\omega} = -\frac{\omega\sqrt{2}}{2} - i\frac{\omega\sqrt{2}}{2}$$

and

$$\arg L(j\omega) = -\frac{\omega\sqrt{2}}{2}.$$

The phase is $-\pi$ for $\omega = \omega_c = \pi/\sqrt{2}$ and the gain at that frequency is $ke^{-\pi} \approx 0.0432k$. The Nyquist plot for a system with gain k is obtained simply by multiplying the Nyquist curve in the figure by k. The Nyquist curve reaches the critical point $L = -1$ for $k = e^{\pi} = 23.1$. The complete Nyquist curve in Figure 9.4 shows that the Nyquist curve does not encircle the critical point if $k < e^{\pi}$, giving a stability condition for the system. ∇

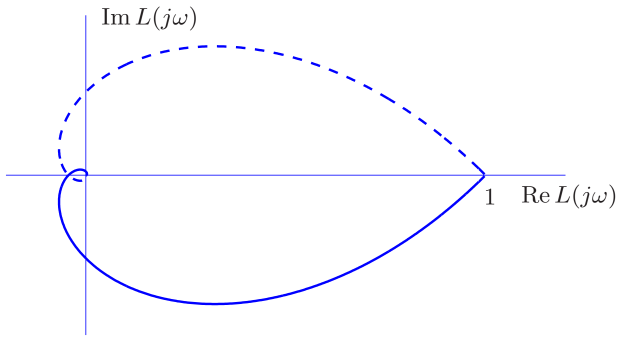

Figure 9.4: Nyquist plot of the transfer function $L(s) = e^{-\sqrt{s}}$

Nyquist's Stability Theorem

We will now state and 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 [?] and the references therein. Since some precision is needed in stating Nyquist's criterion properly, we will also use a more mathematical style of presentation. The key result is the following theorem about functions of complex variables.

Theorem 9.2 (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*

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

where Δ_Γ is the net variation in the angle along the contour Γ, N is the number of zeros and P 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 N. Similarly we find that the sum of the residues of the poles of is $-P$. Furthermore we have

$$\frac{d}{dz}\log f(z) = \frac{f'(z)}{f(z)},$$

which implies that

$$\int_\Gamma \frac{f'(z)}{f(z)} dz = \Delta_\Gamma \log f(z),$$

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

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

Since the variation of $|f(z)|$ around a closed contour is zero we have

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

and the theorem is proven. □

This theorem is useful for 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 9.2 can be used to prove Nyquist's stability theorem by choosing Γ as the Nyquist contour shown in Figure 9.5, which encloses the right half plane. To construct the contour, we start with part of the imaginary axis $-iR \le s \le iR$, 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 $r \to 0$. We call the contour Γ the full Nyquist contour, sometimes call the "D contour".

To see how we used this 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

$$f(s) = 1 + L(s).$$

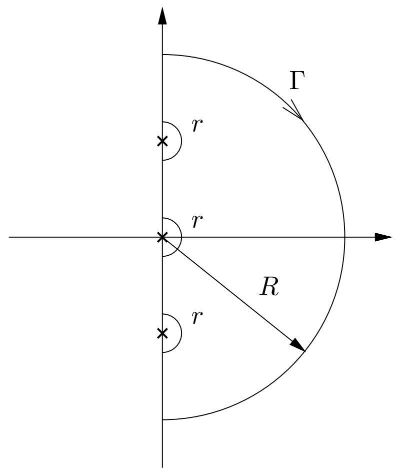

Figure 9.5: The Nyquist contour Γ.

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 clockwise direction. The winding number can conveniently be determined from the Nyquist plot. A direct application of the Theorem 9.2 gives the following result.

Theorem 9.3 (Nyquist's stability theorem). *Consider a closed loop system with the loop transfer function $L(s)$, which which has P poles in the region enclosed by the Nyquist contour. Let w_n be the winding number of the function $f(s) = 1 + L(s)$ when s encircles the Nyquist contour Γ. The closed loop system then has $w_n + P$ poles in the right half plane.*

Since the image of $1+L(s)$ is simply a shifted version of $L(s)$, we usually restate the Nyquist criterion as net encirclements of the -1 point by the image of $L(s)$.

There is a subtlety with the Nyquist plot when the loop transfer function has poles on the imaginary axis because the gain is infinite at the poles. This means that the map of the small semicircles are infinitely large half circles. When plotting Nyquist curves on the computer, one must be careful to see that the 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.

We illustrate Nyquist's theorem by a series of examples.

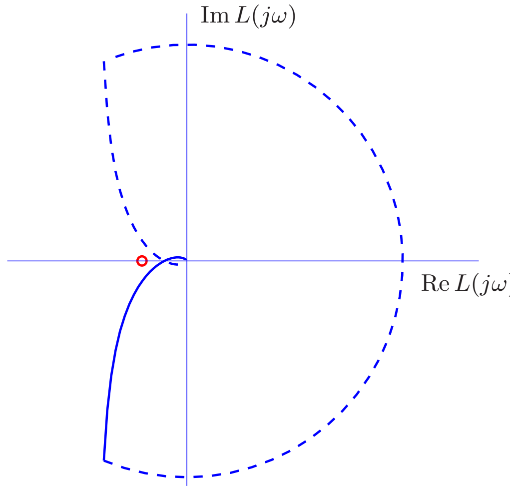

Figure 9.6: The complete Nyquist curve for the loop transfer function $L(s) = \frac{k}{s(s+1)^2}$. The curve is drawn for $k < 2$. The map of the positive imaginary axis is shown in full lines, the map of the negative imaginary axis and the small semi circle at the origin in dashed lines.

Example 9.3. Consider a closed loop system with the loop transfer function

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

Figure 9.6 shows the image of the contour Γ under the map L. The loop transfer function does not have any poles in the region enclosed by the Nyquist contour. By computing the phase of L, one can show that the Nyquist plot intersects the imaginary axis for $\omega = 1$ and the intersection is at $-k/2$. It follows from Figure 9.6 that the winding number is zero if $k < 2$ and 2 if $k > 2$. We can thus conclude that the closed loop system is stable if $k < 2$ and that the closed loop system has two roots in the right half plane if $k > 2$. ∇

Next we will consider a case where the loop transfer function has a pole inside the Nyquist contour.

Example 9.4 (Loop transfer function with RHP pole)**.** Consider a feedback system with the loop transfer function

$$L(s) = \frac{k}{s(s-1)(s+5)}.$$

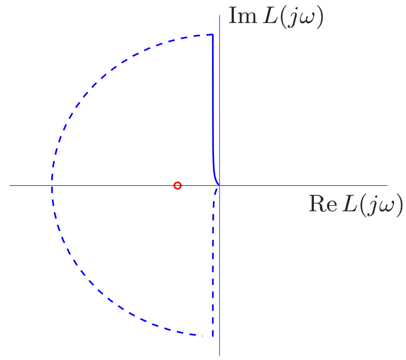

Figure 9.7: Complete Nyquist plot for the loop transfer function $L(s) = \frac{k}{s(s-1)(s+5)}$. The map of the positive imaginary axis is shown in full lines, the map of the negative imaginary axis and the small semi circle at the origin in dashed lines.

This transfer function has a pole at $s = 1$ which is inside the Nyquist contour. The complete Nyquist plot of the loop transfer function is shown in Figure 9.7. Traversing the contour Γ in clockwise we find that the winding number is $w_n = 1$. It follows from the principle of the variation of the argument that the closed loop system has $w_n + P = 2$ poles in the right half plane and hence is unstable. ∇

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 (Conditional stability)**.** Consider a feedback system with the loop transfer function

$$L(s) = \frac{3(s+1)^2}{s(s+6)^2}. \tag{9.2}$$

The Nyquist plot of the loop transfer function is shown in Figure 9.8. 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 9.1 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

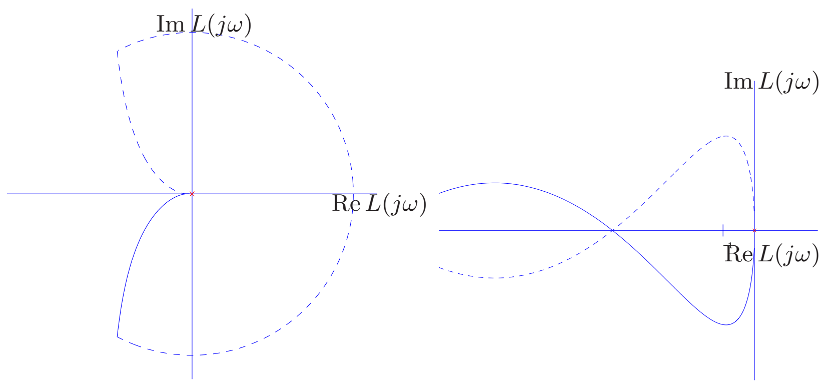

Figure 9.8: Nyquist curve for the loop transfer function $L(s) = \frac{3(s+1)^2}{s(s+6)^2}$. The plot on the right is an enlargement of the area around the origin of the plot on the left.

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 a stable system. It follows from Nyquist's stability criterion that the system is stable because the critical point is to the left of the Nyquist curve when it is traversed for increasing frequencies. ∇

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 is easy to plot of the loop transfer function $L(s)$. An increase of controller gain simply expands the Nyquist plot radially. An increase of the phase of the controller twists the Nyquist plot clockwise. 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 go unstable.

Let ω_{180} be the *phase crossover frequency*, which is the smallest frequency where the phase of the loop transfer function $L(s)$ is $-180°$. The *gain margin* is defined as

$$g_m = \frac{1}{|L(j\omega_{180})|}. \tag{9.3}$$

It tells how much the controller gain can be increased before reaching the stability limit.

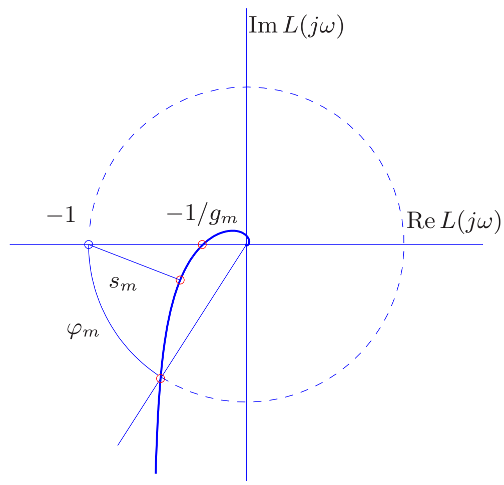

Figure 9.9: Nyquist plot of the loop transfer function L with gain margin g_m, phase margin φ_m and stability margin s_m.

Similarly, let ω_{gc} be the gain crossover frequency, the lowest frequency where the loop transfer function $L(s)$ has unit magnitude. The phase margin is

$$\varphi_m = \pi + \arg L(j\omega_{gc}), \tag{9.4}$$

the amount of phase lag required to reach the stability limit. The margins have simple geometric interpretations in the Nyquist diagram of the loop transfer function as is shown in Figure 9.9.

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 not is 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 also has other nice interpretations as will be discussed in Chapter 12.

When we are 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 $\varphi_m = 30^\circ - 60^\circ$, 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

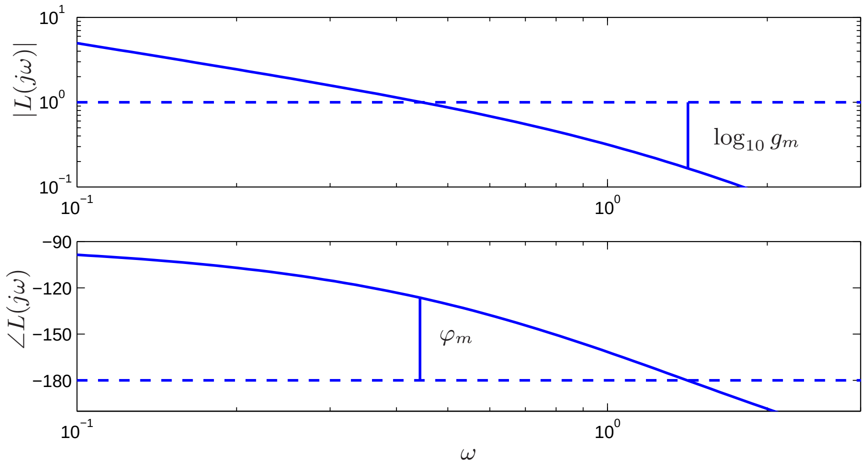

Figure 9.10: Finding gain and phase margins from the Bode plot of the loop transfer function. The loop transfer function is $L(s) = 1/(s(s+1)(s+2))$, the gain margin is $g_m = 6.0$, the gain crossover frequency $\omega_{gc} = 1.42.$, the phase margin is $\varphi_m = 53°$ at the phase crossover frequency $\omega = 0.44$.

the phase margin but for systems where the amplitude ratio of the loop transfer function has several peaks at high frequencies the delay margin is a more relevant measure. A more detailed discussion of robustness measures is given in Chapter 12.

Gain and phase margins can also be determined from the Bode plot of the loop transfer function. A change of controller gain translates the gain curve vertically and it has no effect on the phase curve. To determine the gain margin we first find the phase crossover frequency ω_{180} 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 one. The phase margin is the phase of the loop transfer function at that frequency plus $180°$. Figure 9.10 illustrates how the margins are found in the Bode plot of the loop transfer function. 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.

Example 9.6 (Vehicle steering)**.** Consider the linearized model for vehicle steering with a controller based on state feedback. The transfer function of

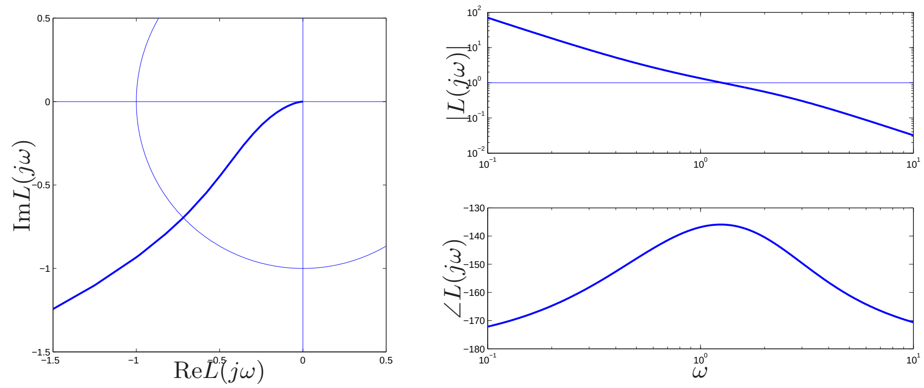

Figure 9.11: Nyquist (left) and Bode (right) plots of the loop transfer function for vehicle steering with a controller based on state feedback and an observer.

the process is

$$P = \frac{\alpha s + 1}{s^2}.$$

and the controller has the transfer function

$$C = \frac{s(k_1 l_1 + k_2 l_2) + k_1 l_2}{s^2 + s(\alpha k_1 + k_2 + l_1) + k_1 + l_2 + k_2 l_1 - \alpha k_2 l_2},$$

as computed in Example 8.4. The Nyquist and Bode plots of the loop transfer function $L = PC$ for the process parameter $a = 0.5$, and a controller characterized by $\omega_c = 1$, $\zeta_c = 0.707$, $\omega_o = 2$, $\zeta_o = 0.707$ are shown in Figure 9.11. The gains of the state feedback are $k_1 = 1$ and $k_2 = 0.914$, and the observer gains are $l_1 = 2.828$ and $l_2 = 4$. The phase margin of the system is 44° and the gain margin is infinite since the phase lag is never greater than 180°, indicating that the closed loop system is robust. ∇

Example 9.7 (Pupillary light reflex dynamics)**.** The pupillary light reflex dynamics was discussed in Example 8.7. Stark found a clever way to artificially increase the loop gain by focusing a narrow beam at the boundary of the pupil. It was possible to increase the gain so much that the pupil started to oscillate. The Bode plot in Figure 9.12b shows that the phase crossover frequency is $\omega_{gc} = 8$ rad/s. This is in good agreement with Stark's experimental investigations which gave an average frequency of 1.35 Hz or 8.5 rad/s. ∇

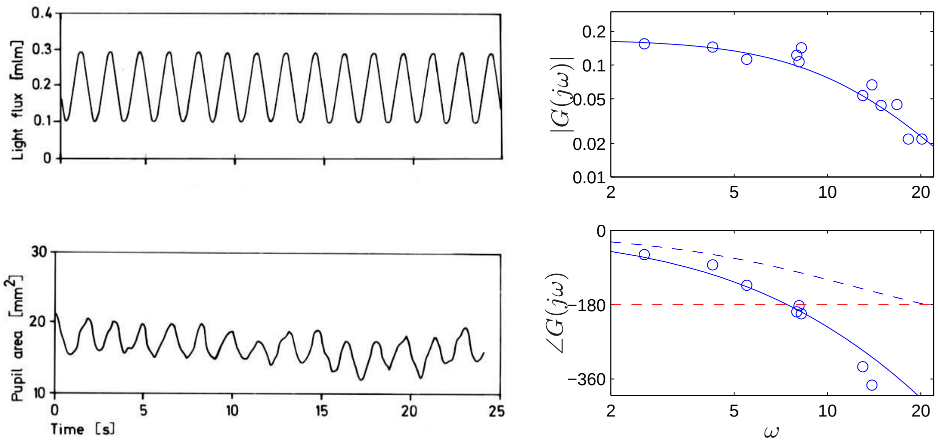

Figure 9.12: Sample curves from open loop frequency response of the eye (left) and Bode plot for the open loop dynamics (right). See Example 8.7 for a detailed description.

9.4 Bode's Relations

An analysis of Bode plots reveals that there appears to be 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 8.10). For the differentiator the slope is $+1$ and the phase is 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 a uniquely given by the gain and vice versa:

$$\begin{aligned}\arg G(j\omega_0) &= \frac{1}{\pi}\int_0^\infty \frac{d\log|G(j\omega)|}{d\log\omega}\log\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right| d\log\omega\\ &\frac{\pi}{2}\int_0^\infty f(\omega)\frac{d\log|G(j\omega)|}{d\log\omega}d\log\omega \approx \frac{\pi}{2}\frac{d\log|G(j\omega)|}{d\log\omega},\end{aligned} \tag{9.5}$$

where f is the weighting kernel

$$f(\omega) = \frac{2}{\pi^2}\log\left|\frac{\omega+\omega_0}{\omega-\omega_0}\right| = \frac{2}{\pi^2}\log\left|\frac{\frac{\omega}{\omega_0}+1}{\frac{\omega}{\omega_0}-1}\right|.$$

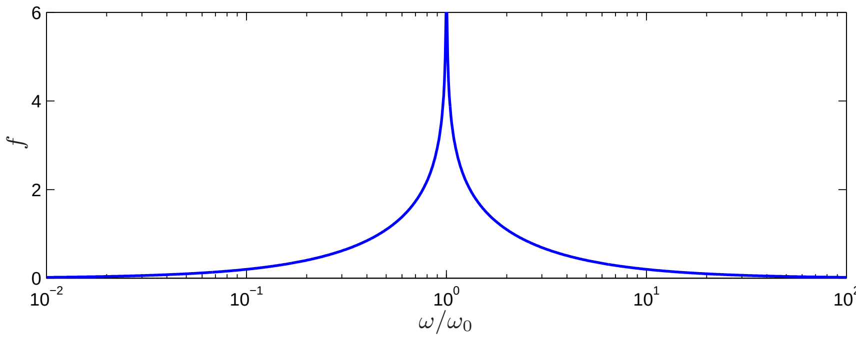

Figure 9.13: The weighting kernel f in Bode's formula for computing the phase curve from the gain curve for minimum phase systems.

The phase curve is thus a weighted average of the derivative of the gain curve. The weight w is shown in Figure 9.13. Notice that the weight falls off rapidly and it is practically zero when the frequency has changed by a factor of ten. It follows from equation (9.5) that a slope of $+1$ corresponds to a phase of $\pi/2$ or $90°$. Compare with Figure 8.10, where the Bode plots have constant slopes -1 and $+1$.

Non-Minimum Phase Systems

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

Example 9.8 (Time delay)**.** The transfer function of a time delay of T units is $G(s) = e^{-sT}$. This transfer function has unit gain, $|G(j\omega)| = 1$, and the phase is

$$\arg G(j\omega) = -\omega T.$$

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 ωT. Notice that the phase lag increases linearly with frequency. Figure 9.14 shows the Bode plot of the transfer function. (Because we use a log scale for frequency, the phase falls off much faster than linearly in the plot.) ∇

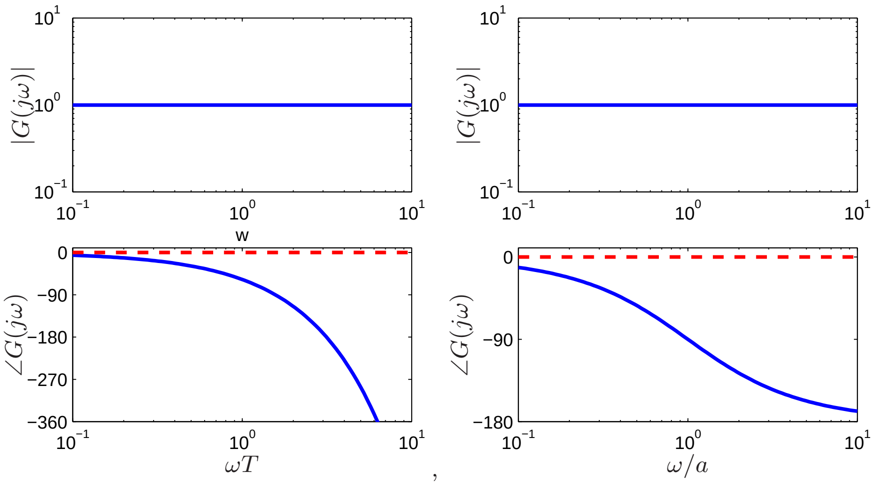

Figure 9.14: Bode plots of a time delay $G(s) = e^{-sT}$ (left) and a system with a right half plane zero $G(s) = (a - s)/(a + s)$ (right). The dashed lines show the phase curves of the corresponding minimum phase systems.

It seems intuitively reasonable that it is impossible to make a system with a time delay respond faster than without the time delay. The presence of a time delay will thus limit the response speed of a system.

Example 9.9 (System with a RHP zero)**.** Consider a system with the transfer function

$$G(s) = \frac{a - s}{a + s}, \quad a > 0,$$

which has a zero $s = a$ in the right half plane. The transfer function has unit gain, $|G(j\omega)| = 1$, and

$$\arg G(j\omega) = -2 \arctan \frac{\omega}{a}.$$

The corresponding minimum phase system with unit gain has the transfer function $G(s) = 1$. Figure 9.14 shows the Bode plot of the transfer function. The Bode plot resembles the Bode plot for a time delay, which is not surprising because the exponential function e^{-sT} can be approximated by

$$e^{-sT} \approx \frac{1 - sT/2}{1 + sT/2}.$$

As far as minimum phase properties are concerned, a right half plane zero at $s = a$ is thus similar to a time delay of $T = 2/a$. Since long time delays

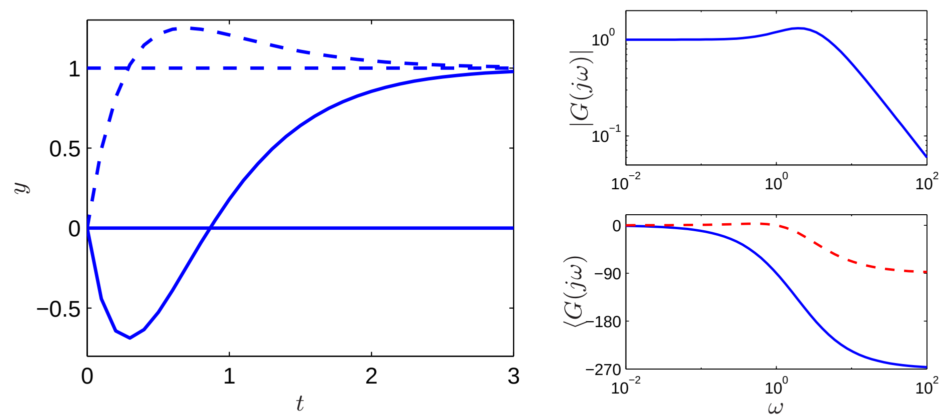

Figure 9.15: Step responses (left) and Bode plots (right) of a system with a zero in the right half plane (full lines) and the corresponding minimum phase system (dashed).

create difficulties in controlling a system we may expect that systems with zeros close to the origin are also difficult to control. ∇

Figure 9.15 shows the step response of a system with the transfer function

$$G(s) = \frac{6(-s+1)}{s^2+5s+6},$$

which has a zero in the right half plane. Notice that the output goes in the wrong direction initially, which is also referred to as an *inverse response*. The figure also shows the step response of the corresponding minimum phase system, which has the transfer function

$$G(s) = \frac{6(s+1)}{s^2+5s+6}.$$

The curves show that the minimum phase system responds much faster. It thus appears that a the non-minimum phase system is more difficult to control. This is indeed the case, as will be shown in Section 11.4.

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 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

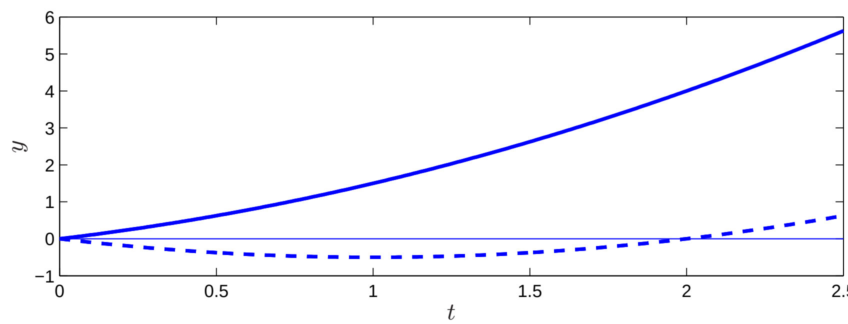

Figure 9.16: Step responses from steer angle to lateral translation for simple kinematics model when driving forward (full) and reverse (dashed).

and actuators or by introducing new sensors and actuators. Non-minimum phase systems are unfortunately not uncommon 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 un-normalized transfer function from steer angle to lateral translation for the simple vehicle model is

$$P(s) = G_{y\delta}(s) = \frac{av_0 s + v_0^2}{bs^2}$$

The transfer function has a zero at $s = v_0/a$. In normal driving this zero is in the left half plane but when reversing the zero moves to the right half plane, which makes the system more difficult to control. Figure 9.16 shows the step response for forward and reverse driving, the parameters are $a = b = 1$, $v_0 = 1$ for forward driving and $v_0 = -1$ for reverse driving. The figure shows that with reverse driving the lateral motion is initially opposite to the desired motion. The action of the zero can also be interpreted as a delay of the control action. ∇

9.5 The Notion of Gain

A key idea in loop analysis it to trace the behavior of signals through a system. The concepts of gain and phase represented by the magnitude and the angle of a transfer function are strongly intuitive because they describe how sinusoidal signals are transmitted. We will now show that the notion of gain can be defined in a much more general way. Something has to be given up to do this and it turns out that it is difficult to define gain for transmission of general signal but that it is easy to define the maximum gain. For this purpose we first define appropriate classes of input and output signals, $u \in \mathcal{U}$ and $u \in \mathcal{Y}$, where $\mathcal{U}$ and $\mathcal{Y}$ are spaces where a notion of magnitude is defined. The gain of a system is defined as

$$\gamma = \sup_{u \in \mathcal{U}} \frac{\|y\|}{\|u\|}, \tag{9.6}$$

where sup is the *supremum,* defined as the smallest number that is larger than its argument. The reason for using supremum is that the maximum may not be defined for $u \in \mathcal{U}$. A correct treatment requires considerable care and space so will will only give a few examples.

Example 9.11 (Linear systems with square integrable inputs)**.** Let the input space $\mathcal{U}$ be square integrable functions, and consider a stable linear system with transfer function $G(s)$. The norm of a signal is given by

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

where the subscript 2 refers to the fact that $\mathcal{U}$ is the set of square integrable functions. Using the same norm for $\mathcal{Y}$, the gain of the system can be shown to be

$$\gamma = \sup_{\omega} |G(j\omega)| := \|G\|_\infty. \tag{9.7}$$

∇

Example 9.12 (Static nonlinear system)**.** Consider a nonlinear static system with scalar inputs and outputs described by $y = f(u)$. The gain obtained γ is a number such that $-\gamma u \le |f(u)| \le \gamma u$. The gain thus defines a sector that encloses the function. ∇

Example 9.13 (Multivariable static system)**.** Consider a static multivariable system $y = Au$, where A is a matrix, whose elements are complex numbers. The matrix does not have to be square. Let the and inputs and outputs be vectors whose elements are complex numbers and use the Euclidean norm

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

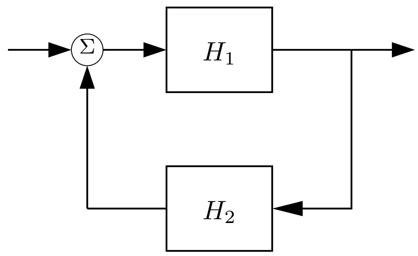

Figure 9.17: A simple feedback loop.

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 eigenvalues $\lambda(A)$ of the matrix A^*A are all real and we have

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

The gain is thus

$$\gamma = \sqrt{\lambda_{min}(A^*A)} \tag{9.8}$$

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

Example 9.14 (Linear multivariable dynamic system)**.** For a linear system multivariable system with a real rational transfer function matrix $G(s)$. Let the input be square integrable functions. The gain of the system is then we have

$$\gamma = \|G(j\omega)\|_\infty = \inf_\omega \bar{\sigma}(G(j\omega)). \tag{9.9}$$

∇

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 one for all frequencies. This result can be extended to much larger class of systems by using the concept of the gain of a system. Consider the closed loop system in Figure 9.17. Let the gains of the systems H_1 and H_2 be γ_1 and γ_2. The *small gain theorem* says that 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.

It also follows from the Nyquist stability theorem that a closed loop system is stable if the phase of the loop transfer function is between $-\pi$ and π. This result can also be extended to nonlinear systems as well. It is called the *passivity theorem* and is closely related to the small gain theorem.

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

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].

9.7 Exercises

1. Use the Nyquist theorem to analyze the stability of the speed control system in Example 9.1, but using the original PI controller from Example 6.9.

2. Discrete time Nyquist

3. Example systems: