Chapter Eleven
Frequency Domain Design

Sensitivity improvements in one frequency range must be paid for with sensitivity deteriorations in another frequency range, and the price is higher if the plant is open-loop unstable. This applies to every controller, no matter how it was designed.

Gunter Stein in the inaugural IEEE Bode Lecture in 1989 [181].

In this chapter we continue to explore the use of frequency domain techniques with a focus on design of feedback systems. We begin with a more thorough description of the performance specifications for control systems, and then introduce the concept of “loop shaping” as a mechanism for designing controllers in the frequency domain. We also introduce some fundamental limitations to performance for systems with right half plane poles and zeros.

11.1 SENSITIVITY FUNCTIONS

In the previous chapter, we considered the use of PID feedback as a mechanism for designing a feedback controller for a given process. In this chapter we will expand our approach to include a richer repertoire of tools for shaping the frequency response of the closed loop system.

One of the key ideas in this chapter is that we can design the behavior of the closed loop system by focusing on the open loop transfer function. This same approach was used in studying stability using the Nyquist criterion: we plotted the Nyquist plot for the *open* loop transfer function to determine the stability of the *closed* loop system. From a design perspective, the use of loop analysis tools is very powerful: since the loop transfer function is $L = PC$, if we can specify the desired performance in terms of properties of L, we can directly see the impact of changes in the controller C. This is much easier, for example, than trying to reason directly about the tracking response of the closed loop system, whose transfer function is given by $G_{yr} = PC/(1+PC)$.

We will start by investigating some key properties of the feedback loop. A block diagram of a basic feedback loop is shown in Figure 11.1. The system loop is composed of two components, the process and the controller. The controller has two blocks: the feedback block C and the feedforward block F. There are two disturbances acting on the process, the load disturbance d, and the measurement noise n. The load disturbance represents disturbances that drive the process away from its desired behavior, while the measurement noise represents the disturbances that corrupt the information about the process given by the sensors. In the figure,

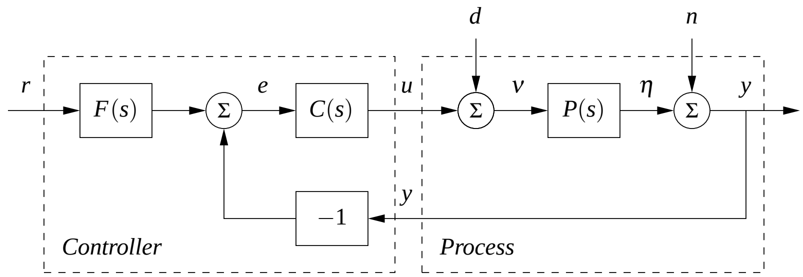


Figure 11.1: Block diagram of a basic feedback loop with two degrees of freedom. The controller has a feedback block C and a feedforward block F. The external signals are the command signal r, the load disturbance d and the measurement noise n. The process output is y and the control signal is u.

the load disturbance is assumed to act on the process input. This is a simplification, since disturbances often enter the process in many different ways, but it allows us to streamline the presentation without significant loss of generality.

The process output η is the real variable that we want to control. Control is based on the measured signal y, where the measurements are corrupted by measurement noise n. The process is influenced by the controller via the control variable u. The process is thus a system with three inputs—the control variable u, the load disturbance d and the measurement noise n—and one output—the measured signal y. The controller is a system with two inputs and one output. The inputs are the measured signal y and the reference signal r and the output is the control signal u. Note that the control signal u is an input to the process and the output of the controller, and that the measured signal y is the output of the process and an input to the controller.

The feedback loop in Figure 11.1 is influenced by three external signals, the reference r, the load disturbance d and the measurement noise n. Any of the remaining signals can be of interest in controller design, depending on the particular application. Since the system is linear, the relations between the inputs and the interesting signals can be expressed in terms of the transfer functions. The following relations are obtained from the block diagram in Figure 11.1:

$$\begin{pmatrix} y \\ \eta \\ v \\ u \\ e \end{pmatrix} = \begin{pmatrix} \dfrac{PCF}{1+PC} & \dfrac{P}{1+PC} & \dfrac{1}{1+PC} \\ \dfrac{PCF}{1+PC} & \dfrac{P}{1+PC} & \dfrac{-PC}{1+PC} \\ \dfrac{CF}{1+PC} & \dfrac{1}{1+PC} & \dfrac{-C}{1+PC} \\ \dfrac{CF}{1+PC} & \dfrac{-PC}{1+PC} & \dfrac{-C}{1+PC} \\ \dfrac{F}{1+PC} & \dfrac{-P}{1+PC} & \dfrac{-1}{1+PC} \end{pmatrix} \begin{pmatrix} r \\ d \\ n \end{pmatrix}. \tag{11.1}$$

In addition, we can write the transfer function for the error between the reference

r and the output η (not an explicit signal in the diagram), which satisfies

$$\varepsilon = r - \eta = \Big(1 - \frac{PCF}{1+PC}\Big) r + \frac{-P}{1+PC} d + \frac{PC}{1+PC} n.$$

There are several interesting conclusions we can draw from these equations. First we can observe that several transfer functions are the same and that all of the important relations are given by the following set of six transfer functions, which we call the *Gang of Six*:

$$\begin{array}{ccc} \dfrac{PCF}{1+PC} & \dfrac{PC}{1+PC} & \dfrac{P}{1+PC} \\ \dfrac{CF}{1+PC} & \dfrac{C}{1+PC} & \dfrac{1}{1+PC}. \end{array} \tag{11.2}$$

The transfer functions in the first column give the response of the process output and control signal to the setpoint. The second column contains the response of the control variable to load disturbance and noise and the final column gives the response of the process output to those two inputs. Notice that only four transfer functions are required to describe how the system reacts to load disturbances and the measurement noise, and that two additional transfer functions are required to describe how the system responds to setpoint changes.

The linear behavior of the system is determined by the six transfer functions in equation (11.2) and specifications can be expressed in terms of these transfer functions. The special case when $F = 1$ is called a system with (pure) error feedback. In this case all control actions are based on feedback from the error only and the system is completely characterized by four transfer functions, namely the four rightmost transfer functions in equation (11.2), which have specific names:

$$\begin{aligned} S &= \frac{1}{1+PC} && \text{sensitivity function} \\ T &= \frac{PC}{1+PC} && \text{complementary sensitivity function} \\ PS &= \frac{P}{1+PC} && \text{load sensitivity function} \\ CS &= \frac{C}{1+PC} && \text{noise sensitivity function} \end{aligned} \tag{11.3}$$

These transfer functions and their equivalent systems are called the *Gang of Four*. The load disturbance sensitivity function is sometimes called the input sensitivity function and the noise sensitivity function is sometimes called the output sensitivity function. These transfer functions have many interesting properties that will be discussed in detail in the rest of the chapter. Good insight into these properties is essential for understanding the performance of feedback systems both for the purpose of use and design.

Analyzing the Gang of Six we find that the feedback controller C influences the

the effects of load disturbances and measurement noise. Notice that measurement noise enters the process via the feedback. In Section 12.2 it will be shown that the controller influences the sensitivity of the closed loop to process variations. The feedforward part F of the controller only influences the response to command signals.

In Chapter 9 we focused on the loop transfer function and we found that its properties gave useful insight into the properties of a system. To make a proper assessment of a feedback system it is necessary to consider the properties of all transfer functions (11.2) in the Gang of Six or Gang of Four for error feedback, as illustrated in the following example.

Example 11.1 The loop transfer function only gives limited insight
Consider a process with the transfer function $P(s) = 1/(s-a)$ controlled by a PI controller with error feedback having the transfer function $C(s) = k(s-a)/s$. The loop transfer function is $L = k/s$, and the sensitivity functions are

$$T = \frac{PC}{1+PC} = \frac{k}{s+k} \qquad PS = \frac{P}{1+PC} = \frac{s}{(s-a)(s+k)}$$
$$CS = \frac{C}{1+PC} = \frac{k(s-a)}{s+k} \qquad S = \frac{1}{1+PC} = \frac{s}{s+k}.$$

Notice that the factor $s-a$ is canceled when computing the loop transfer function and that this factor also does not appear in the sensitivity function or complementary sensitivity function. However, cancellation of the factor is very serious if $a > 0$ since the transfer function PS relating load disturbances to process output is then unstable. In particular, a small disturbance d can lead to an unbounded output, which is clearly not desirable. ∇

The system in Figure 11.1 represents a special case because it is assumed that the load disturbance enters at the process input and that the measured output is the sum of the process variable and measurement noise. Disturbances can enter in many different ways and the sensors may have dynamics. A more abstract way to capture the general case is shown in Figure 11.2, which only has two blocks representing the process ($\mathscr{P}$) and the controller ($\mathscr{C}$). The process has two inputs, the control signal u and a vector of disturbances w, and two outputs, the measured signal y and a vector of signals z that is used to specify performance. The system in Figure 11.1 can be captured by choosing $w = (d,n)$ and $z = (\eta, \nu, e, \varepsilon)$. The process transfer function $\mathscr{P}$ is a 2×2 block matrix and the controller transfer function $\mathscr{C}$ is a 1×2 block matrix; see Exercise 11.3. Processes with multiple inputs and outputs can also be considered by regarding u and y as vectors. Representations at these higher levels of abstraction are useful for the development of theory because they make it possible to focus on fundamentals and to solve general problems with a wide range of applications. However, care must be exercised to maintain the coupling to the real world control problems we intend to solve.

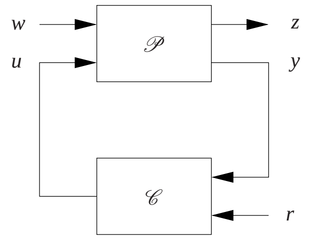


Figure 11.2: A more general representation of a feedback system. The process input u represents the control signal, which can be manipulated, and the process input w represents other signals that influence the process. The process output y is the measured variables and z are other interesting signals of interest.

11.2 FEEDFORWARD DESIGN

Most of our analysis and design tools up to this point have focused on the role of feedback and its effect on the dynamics of the system. Feedforward is a simple and powerful technique that complements feedback. It can be used both to improve the response to reference signals and to reduce the effect of measurable disturbances. Feedforward compensation admits perfect elimination of disturbances but it is much more sensitive than feedback. A general scheme for feedforward was discussed in Section 7.5 on page 225 using Figure 7.10. A simple form of feedforward for PID controllers was discussed in Section 10.5. The controller in Figure 11.1 also has a feedforward block to improve response to reference signals. An alternative version of feedforward is shown in Figure 11.3, which we will use in this section to understand some of the tradeoffs between feedforward and feedback.

Systems with two degrees of freedom (feedforward and feedback) have the advantage that the response to reference signals can be designed independently of the design for disturbance attenuation and robustness. We will first consider the response to reference signals and we will therefore initially assume that the load

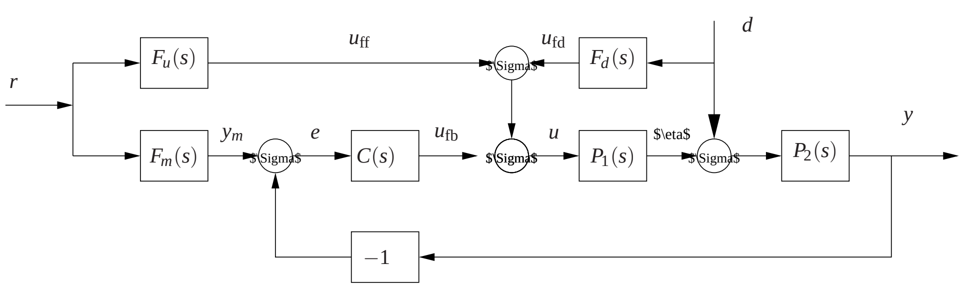


Figure 11.3: Block diagram of a system with feedforward compensation for improved response to reference signals and measured disturbances. Three feedforward elements are present: $F_m(s)$ sets the desired output value, $F_u(s)$ generates the feedforward command u_{ff} and $F_d(s)$ attempts to cancel disturbances.

disturbance d is zero. Let F_m represent the ideal response of the system to reference signals. The feedforward compensator is characterized by the transfer functions F_u and F_m. When the set point is changed the transfer function F_u generates the signal u_{ff}, which is chosen to give the desired output when applied as input to the process. Under ideal conditions the output y is then equal to y_m, the error signal is zero and there will be no feedback action. If there are disturbances or modeling errors, the signal y_m and y will differ. The feedback then attempts to bring the error to zero.

To make a formal analysis we compute the transfer function from reference to process output

$$G_{yr}(s) = \frac{P(CF_m + F_u)}{1+PC} = F_m + \frac{PF_u - F_m}{1+PC}, \tag{11.4}$$

where $P = P_2P_1$. The first term represents the desired transfer function. The second term can be made small in two ways. Feedforward compensation can be used to make $PF_u - F_m$ small or feedback compensation can be used to make $1+PC$ large. Perfect feedforward compensation is obtained by choosing

$$F_m = PF_u. \tag{11.5}$$

Notice the different character of feedback and feedforward. With feedforward we attempt to match two transfer functions, and with feedback we attempt attempt to make the error small by dividing it by a large number. For a controller having integral action, the loop gain is large for small frequencies and it is thus sufficient to make sure that the condition for ideal feedforward holds at higher frequencies. This is easier than trying to satisfy the condition (11.5) for all frequencies.

We will now consider reduction of effects of the load disturbance d in Figure 11.3. We consider the case where we are able to measure the disturbance signal and assume that the disturbance enters the process dynamics in a known way (captured by P_1 and P_2). The effect of the disturbance can be reduced by feeding the measured signal through a dynamical system with the transfer function F_d. Assuming that the reference r is zero, we can use block diagram algebra to find that the transfer function from disturbance to process output is

$$G_{yd} = \frac{P_2(1+F_dP_1)}{1+PC}, \tag{11.6}$$

where $P = P_1P_2$. The effect of the disturbance can be reduced by making $1 - F_dP_1$ small (feedforward) or by making $1+PC$ large (feedback). Perfect compensation is obtained by choosing

$$F_d = -P_1^{-1}. \tag{11.7}$$

Notice that the feedforward disturbance compensator is the inverse of the transfer function P_1, requiring precise knowledge of the process dynamics.

As in the case of reference tracking, disturbance rejection can be accomplished by combining Feedback and feedforward controllers. Since low frequency disturbances can be effectively eliminated by feedback we only require the use of feedforward for high frequency disturbances, and the transfer function F_d in equation (11.7) can then be computed using an approximation of P_1 for high frequen-

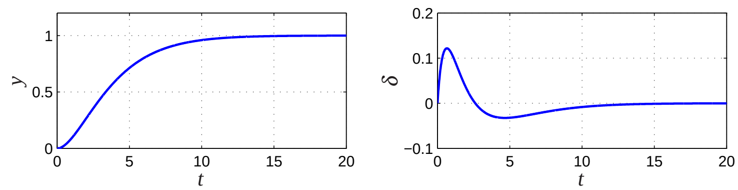


Figure 11.4: Feedforward control for vehicle steering. Lateral deviation y and steering angle δ for smooth lane change control using feedforward.

cies.

Equations (11.5) and (11.7) give analytic expressions for the feedforward compensator. To obtain a transfer function that can be implemented without difficulties we require that the feedforward compensator is stable and that is does not require differentiation. Therefore there may be constraints on possible choices of the desired response F_m and approximations are needed if the process has zeros in the right half plane.

Example 11.2 Vehicle steering

A linearized model for vehicle steering was given in Example 6.4. The normalized transfer function from steering angle to lateral deviation is

$$P(s) = \frac{\gamma s + 1}{s^2}.$$

For a lane transfer system we would like to have a nice response without overshoot and we therefore choose the desired response as

$$F_m = \frac{a^2}{(s+a)^2},$$

where the response speed or aggressiveness of the steering is governed by the parameter a. Equation (11.5) gives

$$F_u = \frac{F_m}{P} = \frac{a^2 s^2}{(\gamma s+1)(s+a)^2},$$

which is a stable transfer function as long as $\gamma > 0$. Figure 11.4 shows the responses of the system for $a = 1.5$. The figure shows that a lane change is accomplished in about 10 vehicle lengths with smooth steering angles. The largest steering angle is a little bit more than 0.1 rad (6 deg). Using the scaled variables the curve showing lateral deviations can also be interpreted as the vehicle path with vehicle length as the length unit. ∇

A major advantage of controllers with two degrees of freedom that combine feedback and feedforward is that the control design problem can be split in two parts. The feedback controller C can be designed to give good robustness and effective disturbance attenuation and the feedforward part can be designed independently to give the desired response to command signals.

11.3 PERFORMANCE SPECIFICATIONS

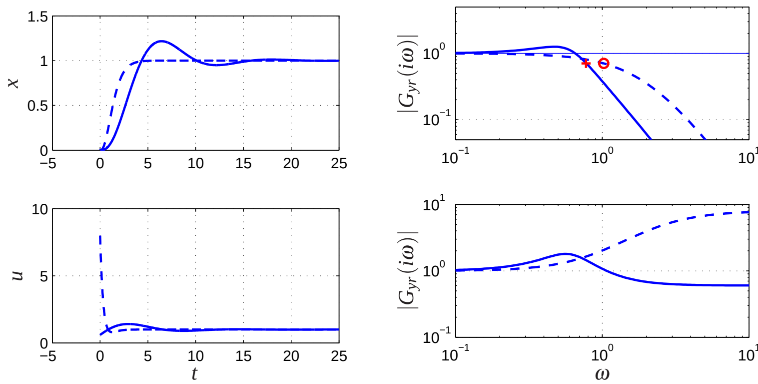
A key element of the control design process is how we specify the desired performance of the system. It is also important for users to understand performance specifications so that they know what to ask for and how to test a system. Specifications are often given in terms of robustness to process variations and responses to reference signals and disturbances. They can be given both in terms of time and frequency responses. Specifications on the step response to reference signals was given in Figure 5.9 in Section 5.3 and in Section 6.3. Robustness specifications based on the loop transfer function and the sensitivity functions were discussed in Section 9.3 and will be discussed more in Chapter 12. The specifications discussed previously were based on the loop transfer function. Since we found in Section 11.1 that a single transfer functions did not always characterize the properties of the closed loop completely we will give a more complete discussion of specifications in this section, based on the full Gang of Six.

The transfer function gives a good characterization of the linear behavior of a system. To give specifications is is desirable to capture the characteristic properties of a system with a few parameters. Common features for time responses are overshoot, rise time and settling time, as shown in Figure 5.9 on page 157. Common features of frequency responses are resonance peak, peak frequency, crossover frequency and bandwidth. The crossover frequency is defined as the frequency where the gain is equal to the low frequency gain for low-pass systems or the high frequency gain for high-pass systems. The bandwidth is defined as the frequencies where the gain is $1/\sqrt{2}$ of the low frequency (low-pass), mid frequency (band-pass) or high frequency gain (high-pass). There are interesting relations between specifications in the time and frequency domain. Roughly speaking, the behavior of time responses for short times is related to behavior of frequency responses at high frequencies and vice versa. The precise relations are not trivial to derive.

Response to Reference Signals

Consider the basic feedback loop in Figure 11.1. The response to reference signals is described by the transfer functions $G_{yr} = PCF/(1+PC)$ and $G_{ur} = CF/(1+PC)$ ($F = 1$ for systems with error feedback). Notice that it is useful to consider both the response of the output and that of the control signal. In particular, the control signal response allows us to judge the magnitude and rate of the control signal required to obtain the output response.

The time response of process output can be characterized by rise time T_r, overshoot M_p and settling time T_s. The response of the control signal can be characterized by the largest value of the control signal or the overshoot. The frequency response G_{yr} can be characterized by the resonance peak M_r, the largest value of the frequency response; the peak frequency ω_{mr}, the frequency where the maximum occurs; and the bandwidth ω_b, the frequency where the gain has decreased to $1/\sqrt{2}$. The transfer function G_{ur} can be characterized by the largest value of $|G_{ur}(i\omega)|$.



(a) **(b)**

Figure 11.5: Reference signal responses. The responses in process output y and control signal u to a unit step in the reference signal r is shown in (a) and the gain curves of G_{yr} and G_{ur} are shown in (b). Results with PI control with error feedback are shown in full lines, the dashed lines show results for a controller with a feedforward compensator.

Example 11.3 Response to reference signals
Consider a process with the transfer function $P(s) = (s+1)^{-3}$ and a PI controller with error feedback having the gains $k_p = 0.6$ and $k_i = 0.5$. The responses are illustrated in Figure 11.5. The full lines show results for a PI controller with error feedback. The dashed lines show results for a controller with feedforward designed to give the transfer function $G_{yr} = (0.5s+1)^{-3}$. Looking at the time responses we find that the controller with feedforward gives a faster response with no overshoot. However, much larger control signals are required to obtain the fast response. The largest value of the control signal is 8 compared to 1.2 for the regular PI controller. The controller with feedforward has a larger bandwidth (marked with $\circ$) and no resonance peak. The transfer function G_{ur} also has higher gain at high frequencies. ∇

Response to Load Disturbances and Measurement Noise

A simple criterion for disturbance attenuation is to compare the output of the closed loop system in Figure 11.1 with the output of the corresponding open loop system obtained by setting $C = 0$. If we let the disturbances for the open and closed loop systems be identical, the output of the closed loop system is then obtained simply by passing the open loop output through a system with the transfer function S. The sensitivity function tells how the variations in the output are influenced by feedback (Exercise 11.10). Disturbances with frequencies such that $|S(i\omega)| < 1$ are attenuated but disturbances with frequencies such that $|S(i\omega)| > 1$ are amplified by feedback. The maximum sensitivity M_s, which occurs at the sensitivity crossover frequency ω_{sc}, is thus a measure of the largest amplification of the disturbances. The maximum magnitude of $1/(1+L)$ is also the minimum of $|1+L|$, which is precisely the stability margin s_m defined in Section 9.3, so that

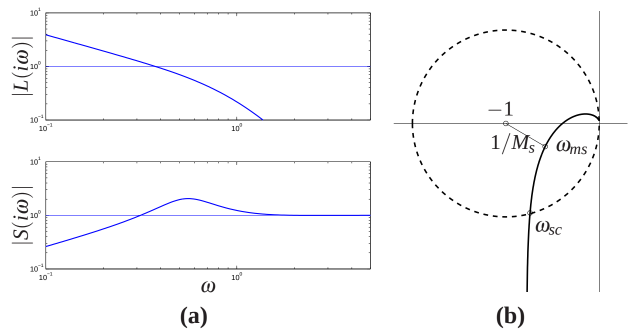


Figure 11.6: Graphical interpretation of the sensitivity function. Gain curves of the loop transfer function and the sensitivity function on the right (a) can be used to calculated the properties of the sensitivity function through the relation $S = 1/(1+L)$. The sensitivity crossover frequency ω_{sc} and the frequency ω_{ms} where the sensitivity has its largest value are indicated in the figure. All points inside the dashed circle have sensitivities greater than 1.

$M_s = 1/s_m$. The maximum sensitivity is therefore also a robustness measure.

If the sensitivity function is known, the potential improvements by feedback can be evaluated simply by recording a typical output and filtering it through the sensitivity function. A plot of the gain curve of the sensitivity function is a good way to make an assessment of disturbance attenuation. Since the sensitivity function only depends on the loop transfer function its properties can also be visualized graphically using the Nyquist plot of the loop transfer function. This is illustrated in Figure 11.6. The complex number $1 + L(i\omega)$ can be represented as the vector from the point -1 to the point $L(i\omega)$ on the Nyquist curve. The sensitivity is thus less than one for all points outside a circle with radius 1 and center at -1. Disturbances with frequencies in this range are attenuated by the feedback.

The transfer function G_{yd} from load disturbance d to process output y for the system in Figure 11.1 is

$$G_{yd} = \frac{P}{1+PC} = PS = \frac{T}{C}. \tag{11.8}$$

Since load disturbances typically have low frequencies, it is natural to focus on the behavior of the transfer function at low frequencies. For a system with $P(0) \neq 0$ and a controller with integral action, the controller gain goes to infinity for small frequencies and we have the following approximation for small s:

$$G_{yd} = \frac{T}{C} \approx \frac{1}{C} \approx \frac{s}{k_i}, \tag{11.9}$$

where k_i is the integral gain. Since the sensitivity function S goes to 1 for large s we have the approximation $G_{yd} \approx P$ for high frequencies.

Measurement noise, which typically has high frequencies, generates rapid variations in the control variable that are detrimental because they cause wear in many actuators and can even saturate an actuator. It is thus important to keep the varia-

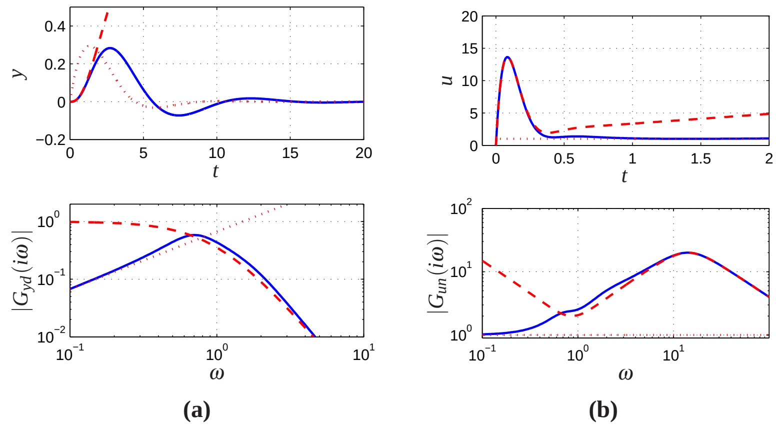


Figure 11.7: Disturbance responses. Time and frequency responses of process output y to load disturbance d are shown in (a) and responses of the control signal u to measurement noise d are shown in (b). The low frequency approximation is shown with dotted lines and the high frequency approximations by dashed lines.

tions in the control signal due to measurement noise at reasonable levels—a typical requirement is that the variations are only a fraction of the span of the control signal. The variations can be influenced by filtering and by proper design of the high frequency properties of the controller.

The effects of measurement noise are captured by the transfer function from measurement noise to the control signal,

$$-G_{un} = \frac{C}{1+PC} = CS = \frac{T}{P}. \tag{11.10}$$

For low frequencies the transfer function the sensitivity function equals 1 and G_{un} can be approximated by $1/P$. For high frequencies PC is small and G_{un} can be approximated as $G_{un} \approx C$.

Example 11.4 Response to disturbances

Consider a process with the transfer function $P(s) = (s+1)^{-3}$ and a PID controller with gains $k = 0.6$, $k_i = 0.5$ and $k_d = 2.0$. We augment the controller with a second order noise filter with $T_f = 0.1$ so that its transfer function is

$$C(s) = \frac{k_d s^2 + ks + k_i}{s(s^2 T_f^2/2 + sT_f + 1)}.$$

The responses are illustrated in Figure 11.7. The system response to a step in the load disturbance in the top part of Figure 11.7a has a peak of 0.28 at time $t = 2.73$, and the initial part of the response is well captured by the high frequency approximation $G_{yd} \approx P$ (dashed). The magnitude of the peak is also well approximated by the low frequency approximation $G_{yd} \approx 1/C$ (dotted), but the peak time is not. The frequency response in Figure 11.7a shows that the gain has a maximum 0.58 at $\omega = 0.7$. The figure shows that the gain curve is well captured by the approximations.

The response of the control signal to a step in measurement noise is shown in Figure 11.7b. The high frequency roll-off of the transfer function $G_{un}(i\omega)$ is due to filtering; without that the gain curve in Figure 11.7b would continue to rise after 20 rad/s. The step response has a peak of 13 at $t = 0.08$ which is well captured by the high frequency approximation (dashed). The frequency response has its peak 20 at $\omega = 14$, which is also well captured by the high frequency approximation (dashed). Notice that the peak occurs far above the peak of the response to load disturbances and far above the gain crossover frequency $\omega_{gc} = 0.78$. An approximation derived in Exercise 11.11 gives $\max|CS(i\omega)| \approx k_d/T_f = 20$ which occurs at $\omega = \sqrt{2}/T_d = 14.1$. ∇

11.4 FEEDBACK DESIGN VIA LOOP SHAPING

One advantage of the Nyquist stability theorem is that it is based on the loop transfer function, which is related to the controller transfer function through $L = PC$. It is thus easy to see how the controller influences the loop transfer function. To make an unstable system stable we simply have to bend the Nyquist curve away from the critical point.

This simple idea is the basis of several different design methods, collectively called *loop shaping*. The methods are based on choosing a compensator that gives a loop transfer function with a desired shape. One possibility is to determine a loop transfer function that gives a closed loop system with the desired properties and to compute the controller as $C = L/P$. Another is to start with the process transfer function change its gain and then add poles and zeros until the desired shape is obtained. In this section we will explore different loop shaping methods for control law design.

Design Considerations

We will first discuss a suitable shape of the loop transfer function that gives good performance and good stability margins. Figure 11.8 shows a typical loop transfer function. Good robustness requires good stability margins (or good gain and phase margins) which imposes requirements on the loop transfer function around the crossover frequencies ω_{pc} and ω_{gc}. The gain of L at low frequencies must be large in order to have good tracking of command signals and good rejection of low frequency disturbances. Since $S = 1/(1+L)$ it follows that for frequencies where $|L| > 100$ disturbances will be attenuated by a factor of 100 and the tracking error is less than 1%. It is therefore desirable to have a large crossover frequency and a steep (negative) slope of the gain curve. The gain at low frequencies can be increased by a controller with integral action which is also called *lag compensation*. To avoid injecting too much measurement noise into the system it is desirable that the loop transfer function have a low gain at frequencies high frequencies *high fre-*

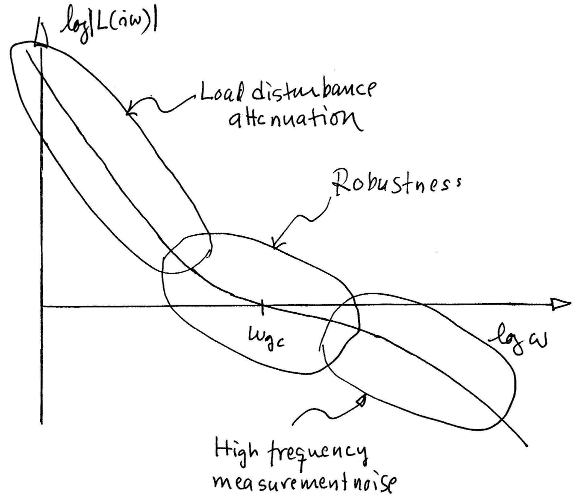


Figure 11.8: Gain curve of the Bode plot for a typical loop transfer function. The gain crossover frequency ω_{gc} and the slope n_{gc} of the gain curve at crossover are important parameters that determine the robustness of the closed lop systems. At low frequency, a large magnitude for L provides good load disturbance rejection and reference tracking, while at high frequency a small loop gain is used to avoid amplifying measurement noise.

quency roll-off. The choice of gain crossover frequency is a compromise between attenuation of load disturbances, injection of measurement noise and robustness.

Bode's relations (see Section 9.4) impose restrictions on the shape of the loop transfer function. Equation (9.8) implies that the slope of the gain curve at gain crossover cannot be too steep. If the gain curve is constant slope, we have the following relation between slope n_{gc} and phase margin φ_m:

$$n_{gc} = -2 + \frac{2\varphi_m}{\pi}\,[\text{rad}] = -180^\circ + \varphi_m\,[\text{deg}]. \tag{11.11}$$

This formula is a reasonable approximation when the gain curve does not deviate too much from a straight line. It follows from equation (11.11) that the phase margins 30°, 45° and 60° correspond to the slopes -5/3, -3/2 and -4/3.

Loop shaping is a trial and error procedure. We typically start with a Bode plot of the process transfer function. We then attempt to shape the loop transfer function by changing controller gain and adding poles and zeros of the controller. Different performance specifications are evaluated for each controller as we attempt to balance many different requirements by adjusting controller parameters and complexity. Loop shaping is straightforward to apply to single-input, single output systems. It can also be applied to systems with one input and many outputs by closing the loops one at a time starting with the innermost loop. The only limitation for minimum phase systems is that large phase leads and high controller gains may be required to obtain closed loop systems with fast response. Many specific procedures are available: they all require experience but they also give a

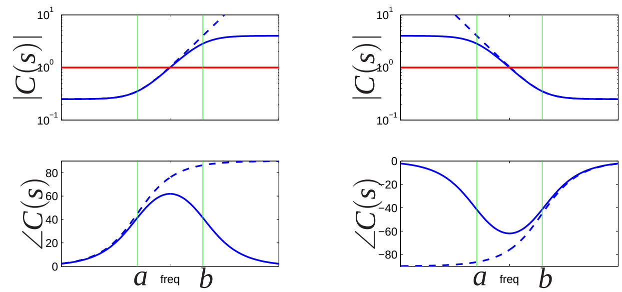


Figure 11.9: Frequency response for a lead and lag compensators, $C(s) = k(s+a)/(s+b)$. Lead compensation occurs when $a < b$ (left) and provides phase lead between $\omega = a$ and $\omega = b$. Lag compensation corresponds to $a > b$ and provides low frequency gain. PI control is a special case of lag compensation and PD control is a special case of lead compensations. Frequency responses are shown in dashed curves.

good insight into the conflicting requirements. There are fundamental limitations to what can be achieved for systems that are not minimum phase; they will be discussed in the next section.

Lead and Lag Compensation

A simple way to do loop shaping is to start with the transfer function of the process and to add simple compensators with the transfer function

$$C(s) = k\frac{s+a}{s+b}. \tag{11.12}$$

The compensator is called a lead compensator if $a < b$ and a lag compensator if $a > b$. The PI controller is a special case of lag compensator with $b = 0$ and the ideal PD controller is a special case of a lead compensator with $a = 0$. Bode plots of lead and lag compensators are shown in Figure 11.9. Lag compensation increases the gain at low frequencies. It is typically used to improve tracking performance and disturbance attenuation at low frequencies. The following example gives an illustration.

Example 11.5 Atomic force microscope in tapping mode

A simple model of the dynamics of the vertical motion of an atomic force microscope in tapping mode was given in Exercise 9.5. The transfer function for the system dynamics is

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

where $a = \zeta\omega_0$, and $\tau = 2\pi n/\omega_0$ and the gain has been normalized to 1. A Bode plot of this transfer function for the parameters $a = 1$ and is shown in dashed curves in Figure 11.10a. To improve attenuation of load disturbances we increase the low frequency gain by introducing an integrating controller. The loop transfer function then becomes $L = k_i P(s)/s$ and we adjust the gain so that the phase margin is zero,

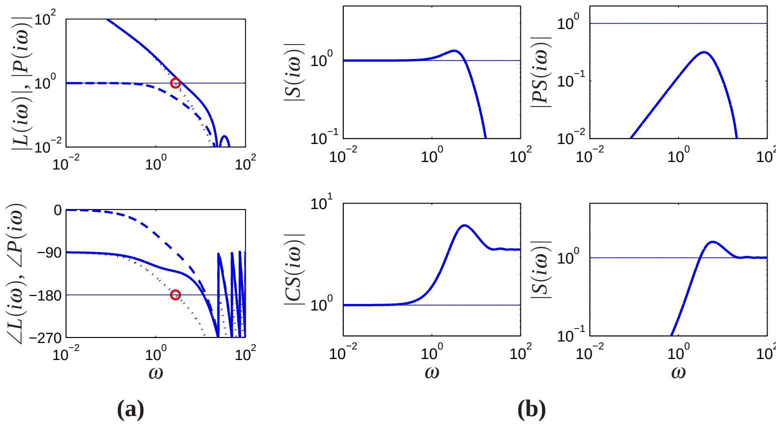


Figure 11.10: Loop shaping design of a controller for an atomic force microscope in tapping mode. Figure 11.10a shows Bode plots of the process (dashed), the loop transfer function with an integral controller with critical gain (dotted) and a PI controller adjusted to give reasonable robustness. Figure 11.10b shows the gain curves for the Gang of Four for the system.

giving $k_i = 8.3$. Notice the increase of the gain at low frequencies. The Bode plot is shown by the dotted line in Figure 11.10a where the critical point is indicated by ○. To improve the phase margin we introduce proportional action and we increase the proportional gain k_p gradually until reasonable values of the sensitivities are obtained. The value $k_p = 3.5$ gives $M_s = 1.6$ and $M_t = 1.3$. The loop transfer function is shown in full lines in Figure 11.10a. Notice the significant increase of the phase margin compared with the purely integrating controller (dotted line).

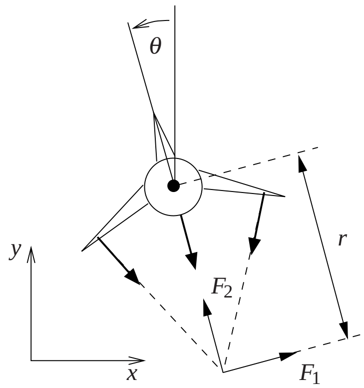
To evaluate the design we also compute the gain curves of the transfer functions in the Gang of Four. They are shown in Figure 11.10b. The peak of the sensitivity curves are reasonable and the plot of *PS* shows that the largest value of *PS* is 0.3 which implies that load disturbances are well attenuated. The plot of *CS* shows that the largest controller gain is 6. The controller has a gain of 3.5 at high frequencies and hence we may consider adding high frequency roll off. ∇

A common problem in design of feedback systems is that the phase lag of the system at the desired crossover frequency is not high enough to allow either proportional or integral feedback to be used effectively. Instead, one may have a situation where you need to add phase *lead* to the system, so that the crossover frequency can be increased.

A standard way to accomplish this is to use a *lead compensator*, which has the form

$$C(s) = k\frac{s+a}{s+b} \qquad a < b. \tag{11.13}$$

A key feature of the lead compensator is that it adds phase *lead* in the frequency range between the pole/zero pair (and extending approximately 10X in frequency in each direction). By appropriately choosing the location of this phase lead, we



(a)

Symbol	Description	Value
m	vehicle mass	4.0 kg
J	vehicle inertia, φ_3 axis	0.0475 kg m^2
r	force moment arm	26.0 cm
d	angular damping	0.001 kg m/s
g	gravitational constant	9.8 m/s^2

(b)

Figure 11.11: Roll control of a vectored thrust aircraft. The roll angle θ is controlled by applying maneuvering thrusters, resulting in a moment generated by F_z. The table to the right lists the parameter values for a laboratory version of the system.

can provide additional phase margin at the gain crossover frequency.

Because the phase of a transfer function is related to the slope of the magnitude, increasing the phase requires increasing the gain of the loop transfer function over the frequency range in which the lead compensation is applied. Hence we can also think of the lead compensator as changing the slope of the transfer function and thus shaping the loop transfer function in the crossover region (although it can be applied elsewhere as well).

Example 11.6 Roll control for a vectored thrust aircraft

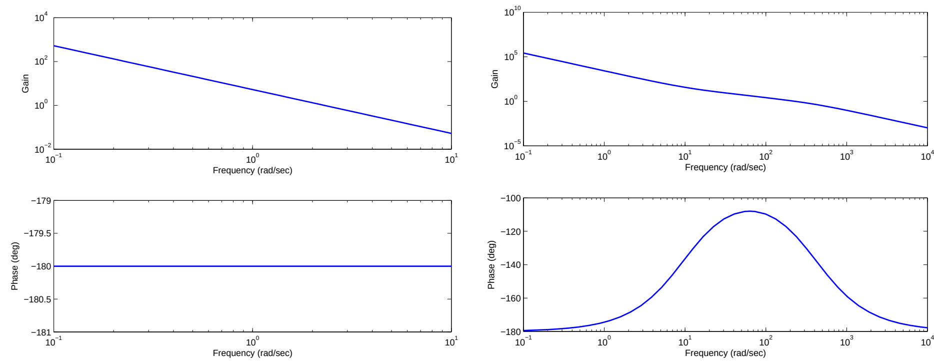
Consider the control of the roll of a vectored thrust aircraft, such as the one illustrated in Figure 11.11. Following exercise 8.11, we model the system with a second order transfer function of the form

$$P(s) = \frac{r}{Js^2 + cs},$$

with the parameters given in Figure 11.11b. We take as our performance specification that we would like less than 1% error in steady state and less than 10% tracking error up to 10 rad/sec.

The open loop transfer function is shown in Figure 11.12a. To achieve our performance specification, we would like to have a gain of at least 10 at a frequency of 10 rad/sec, requiring the gain crossover frequency to be at a higher frequency. We see from the loop shape that in order to achieve the desired performance we cannot simply increase the gain, since this would give a very low phase margin. Instead, we must increase the phase at the desired crossover frequency.

To accomplish this, we use a lead compensator (11.13) with $a = 2$ and $b = 50$. We then set the gain of the system to provide a large loop gain up to the desired bandwidth, as shown in Figure 11.12b. We see that this system has a gain of greater than 10 at all frequencies up to 10 rad/sec and that it has over 40° degrees of phase margin. ∇



(a) (b)

Figure 11.12: Control design for a vectored thrust aircraft using lead compensation. The Bode plot for the open loop process P is shown on the left and the loop transfer function $L = PC$ using a lead compensator on the right. Note the phase lead in the crossover region near $\omega = 100$ rad/s.

The action of a lead compensator is essentially the same as that of the derivative portion of a PID controller. As described in Section 10.5, we often use a filter for the derivative action of a PID controller to limit the high frequency gain. This same effect is present in a lead compensator through the pole at $s = b$.

Equation (11.13) is a first order lead compensator and can provide up to 90° of phase lead. Higher levels of phase lead can be provided by using a second order lead compensator:

$$C(s) = k\frac{(s+a)^2}{(s+b)^2} \qquad a < b.$$

11.5 FUNDAMENTAL LIMITATIONS

Although loop shaping gives us a great deal of flexibility in designing the closed loop response of a system, there are certain fundamental limits on what can be achieved. We consider here some of the primary performance limitations that can occur due to difficult dynamics; additional limitations having to do with robustness are considered in the next chapter.

Right Half Plane Poles and Zeros and Time Delays

There are linear systems that are inherently difficult to control. The limitations are related to poles and zeros in the right half plane and time delays. To explore the limitations caused by poles and zeros in the right half plane we factor the process transfer function as

$$P(s) = P_{mp}(s)P_{ap}(s), \tag{11.14}$$

where P_{mp} is the minimum phase part and P_{ap} is the non-minimum phase part. The factorization is normalized so that $|P_{ap}(i\omega)| = 1$ and the sign is chosen so that P_{ap}

has negative phase. The transfer function P_{ap} is called an *all-pass system* because it has unit gain for all frequencies. Requiring that the phase margin is φ_m we get

$$\arg L(i\omega_{gc}) = \arg P_{ap}(i\omega_{gc}) + \arg P_{mp}(i\omega_{gc}) + \arg C(i\omega_{gc}) \geq -\pi + \varphi_m, \quad (11.15)$$

where C is the controller transfer function. Let n_{gc} be the slope of the gain curve at the crossover frequency. Since $|P_{ap}(i\omega)| = 1$ it follows that

$$n_{gc} = \frac{d\log|L(i\omega)|}{d\log\omega}\bigg|_{\omega=\omega_{gc}} = \frac{d\log|P_{mp}(i\omega)C(i\omega)|}{d\log\omega}\bigg|_{\omega=\omega_{gc}}.$$

Assuming that the slope n_{gc} is negative it has to be larger than -2 for the system to be stable. It follows from Bode's relations, equation (9.8), that

$$\arg P_{mp}(i\omega) + \arg C(i\omega) \approx n_{gc}\frac{\pi}{2}.$$

Combining this with equation (11.15) gives the following inequality for the allowable phase lag

$$-\arg P_{ap}(i\omega_{gc}) \leq \pi - \varphi_m + n_{gc}\frac{\pi}{2} =: \varphi_l. \quad (11.16)$$

This condition, which we call the *crossover frequency inequality*, shows that the gain crossover frequency must be chosen so that the phase lag of the non-minimum phase component is not too large. For systems with high robustness requirements we may choose a phase margin of 60° ($\varphi_m = \pi/3$) and a slope $n_{gc} = -1$, which gives an admissible phase lag $\varphi_l = \pi/6 = 0.52$ rad (30°). For systems where we can accept a lower robustness we may choose a phase margin of 45° ($\varphi_m = \pi/4$) and the slope $n_{gc} = -1/2$, which gives an admissible phase lag $\varphi_l = \pi/2 = 1.57$ rad (90°).

The crossover frequency inequality shows that non-minimum phase components impose severe restrictions on possible crossover frequencies. It also means that there are systems that cannot be controlled with sufficient stability margins. The conditions are more stringent if the process has an uncertainty $\Delta P(i\omega_{gc})$, as we shall see in the next chapter. We illustrate the limitations in a number of commonly encountered situations.

Example 11.7 Zero in the right half plane

The non-minimum phase part of the process transfer function for a system with a right half plane zero is

$$P_{ap}(s) = \frac{z-s}{z+s},$$

where $z > 0$. The phase lag of the non-minimum phase part is

$$-\arg P_{ap}(i\omega) = 2\arctan\frac{\omega}{z}.$$

Since the phase of P_{ap} decreases with frequency, the inequality (11.16) gives the following bound on the crossover frequency:

$$\omega_{gc} < z\tan(\varphi_l/2). \quad (11.17)$$

With $\varphi_l = \pi/3$ we get $\omega_{gc} < 0.6a$. Slow zeros (z small) therefore give stricter restrictions on possible gain crossover frequencies than fast zeros. ∇

Time delays also impose limitations similar to those given by zeros in the right half plane. We can understand this intuitively from the approximation

$$e^{-s\tau} \approx \frac{1-s\tau}{1+s\tau}.$$

Example 11.8 Pole in the right half plane

The non-minimum phase part of the transfer function for a system with a pole in the right half plane is

$$P_{ap}(s) = \frac{s+p}{s-p},$$

where $p > 0$. The phase lag of the non-minimum phase part is

$$\varphi_l = -\arg P_{ap}(i\omega) = 2\arctan\frac{p}{\omega}$$

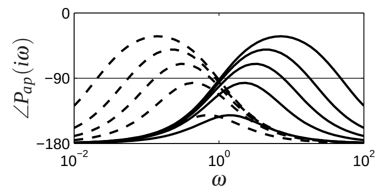
and the crossover frequency inequality becomes

$$\omega_{gc} > \frac{p}{\tan(\varphi_l/2)}. \tag{11.18}$$

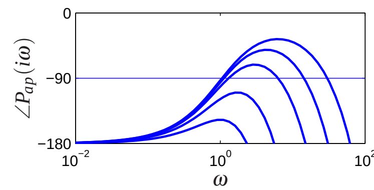
Right half plane poles thus require that the closed loop system have sufficiently high bandwidth. With $\varphi_l = \pi/3$ we get $\omega_{gc} > 1.7p$. Fast right half plane poles (p large) therefore gives stricter restrictions on possible gain crossover frequencies than slow poles. Control of unstable systems imposes requirements for process actuators and sensors. ∇

Since a zero in the right half plane gives an upper limit to the achievable gain crossover frequency it follows that zeros far to the right give small limitations but that zeros close to the origin imposes severe limitations. The situation with right half plane poles is different because a pole imposes a lower limit to the gain crossover frequency and poles far to the right require systems with a high gain crossover frequency. It can thus be expected that systems with poles and zeros cannot be controlled robustly if the poles and zeros are too close.

A straightforward way to use the crossover frequency inequality is to plot the phase of the non-minimum phase factor P_{ap} of the process transfer function. Such a plot will immediately show the permissible gain crossover frequencies. An illustration is given in Figure 11.13 which shows the phase of P_{ap} for systems with a right half plane pole-zero pair and systems with a right half plane pole and a time delay. If we require that the phase lag φ_l of the nonminimum phase factor should be less than 90 deg we must require that the ratio z/p is larger than 6 or smaller than 1/6 for system with right half plane poles and zeros and that the product $p\tau$ is less than 0.15 for systems with a time delay and a right half plane pole. Notice the symmetry in the problem for $z > p$ and $z < p$: in either case the zeros and the poles must be sufficiently far apart (Exercise 11.13). Also notice that possible values of the gain crossover frequency ω_{gc} are quite limited.



(a) RHP pole/zero pair



(b) RHP pole + time delay

Figure 11.13: Example limitations due to the crossover frequency inequality. The figure illustrates limitations by showing the phase of the minimum phase factor P_{ap} of transfer functions. All systems have a right half plane pole at $s = 1$. The system in (a) has zeros at $s = 2$, 5, 10, 20 and 50 (full lines) and at $s = 0.5$, 0.02 0.1 0.05 and 0.02 (dashed lines). The system in (b) has time delays $\tau = 0.05$ 0.1, 0.2 0.5 and 1.

As the examples above show, right half plane poles and zeros significantly limit the achievable performance of a system, hence one would like to avoid these whenever possible. The poles of a system depend on the intrinsic dynamics of the system and are given by the eigenvalues of the dynamics matrix A of a linear system. Sensors and actuators have no effect on the poles; the only way to change poles is to redesign the system. Notice that this does not imply that unstable systems should be avoided. Unstable system may actually have advantages; one example is high performance supersonic aircraft.

The zeros of a system depend on the how sensors and actuators are coupled to the states. The zeros depend on all the matrices A, B, C and D in a linear system. The zeros can thus be influenced by moving sensors and actuators or by adding sensors and actuators. Notice that a fully actuated system $B = I$ does not have any zeros.

Example 11.9 Balance system

As an example of a system with both right half plane poles and zeros, consider the balance system with zero damping, whose dynamics are given by

$$H_{\theta F} = \frac{ml}{-(M_tJ_t - m^2l^2)s^2 + mglM_t}$$
$$H_{pF} = \frac{-J_ts^2 + mgl}{s^2\big(-(M_tJ_t - m^2l^2)s^2 + mglM_t\big)}.$$

Assume that we want to stabilize the pendulum by using the cart position as the measured signal. The transfer function from the input force F to the cart position p has poles $\{0, 0, \pm\sqrt{mglM_t/(M_tJ_t - m^2l^2)}\}$ and zeros $\{\pm\sqrt{mgl/J_t}\}$. Using the parameters in Example 6.7, the right half plane pole is at $p = 2.68$ and the zero is at $z = 2.09$. The pole is so close to the zero that the system cannot be controlled robustly. Using Figure 11.13, we see that the amount of achievable phase margin for the system is very small if we desire a bandwidth in the range of 2–4 rad/s.

The right half plane zero of the system can be eliminated by changing the output of the system. For example, if we choose the output to correspond to a position at a distance r along the pendulum, we have $y = p - r\sin\theta$ and the transfer function for the linearized output becomes

$$H_{y,F} = H_{pF} - rH_{\theta F} = \frac{(mlr - J_t)s^2 + mgl}{s^2\big(-(M_tJ_t - m^2l^2)s^2 + mglM_t\big)}.$$

If we choose r sufficiently large then $mlr - J_t > 0$ and we eliminate the right half plane zero, obtaining instead two pure imaginary zeros. Note that $J_t = J + ml^2$ and so if the inertia of the pendulum J is nonzero then $mlr - J_t > 0$ requires $r > l$, indicating that our output must correspond to a point above the center of mass of the pendulum.

If we choose r such that $mlr - J_t > 0$ then the crossover inequality is based just on the right half plane pole (Example 11.8). If our desired phase lag is $\varphi_l = 45^\circ$ then our gain crossover must satisfy

$$\omega_{\text{gc}} > \frac{p}{\tan\varphi_l/2} = 2.68.$$

Assuming that our actuators have sufficiently high bandwidth, say a factor of 10 above ω_{gc} or roughly 4 Hz, then we can provide robust tracking up to this frequency.

∇

Bode's Integral Formula

In addition to providing adequate phase margin for robust stability, a typical control design will have to satisfy performance conditions on the sensitivity functions (Gang of Four). In particular the sensitivity function $S = 1/(1+PC)$ represents disturbance attenuation and also relates the tracking error e to the reference signal: we usually want the sensitivity to be small over the range of frequencies where we want small tracking error and good disturbance attenuation. A basic problem is to investigate if S can be made small over a large frequency range. We will start by investigating an example.

Example 11.10 System that admits small sensitivities
Consider a closed loop system consisting of a first order process and a proportional controller. Let the loop transfer function be

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

where parameter k is the controller gain. The sensitivity function is

$$S(s) = \frac{s+1}{s+1+k}$$

and we have

$$|S(i\omega)| = \sqrt{\frac{1+\omega^2}{1+2k+k^2+\omega^2}}.$$

This implies that $|S(i\omega)| < 1$ for all finite frequencies and that the sensitivity can be made arbitrary small for any finite frequency by making k sufficiently large. ∇

The system in Example 11.10 is unfortunately an exception. The key feature of the system is that the Nyquist curve of the process is completely contained in the right half plane. Such systems are called *positive real*. For these systems the Nyquist curve never enters the unit disk centered at -1 (the region is shown in Figure 11.6) where the sensitivity is greater than one.

For typical control systems there are unfortunately severe constraints on the sensitivity function. The following theorem, due to Bode, provides insights into the limits of performance under feedback.

Theorem 11.1 (Bode's integral formula)**.** *Let $S(s)$ be the sensitivity function for a feedback system and assume that it goes to zero faster than $1/s$ for large s. If the loop transfer function has poles p_k in the right half plane then the sensitivity function satisfies the following integral:*

$$\int_0^\infty \log|S(i\omega)|\,d\omega = \int_0^\infty \log\frac{1}{|1+L(i\omega)|}\,d\omega = \pi\sum p_k. \tag{11.19}$$

Equation (11.19) implies that there are fundamental limitations to what can be achieved by control and that control design can be viewed as a redistribution of disturbance attenuation over different frequencies. In particular, this equation shows that if the sensitivity function is made smaller for some frequencies it must increase at other frequencies so that the integral of $\log|S(i\omega)|$ remains constant. This means that if disturbance attenuation is improved in one frequency range it will be worse in another, a property sometime referred to as the *waterbed effect*. It also follows that systems with open loop poles in the right half plane have larger overall sensitivity than stable systems.

Equation (11.19) can be regarded as a *conservation law*: if the loop transfer function has no poles in the right half plane the equation simplifies to

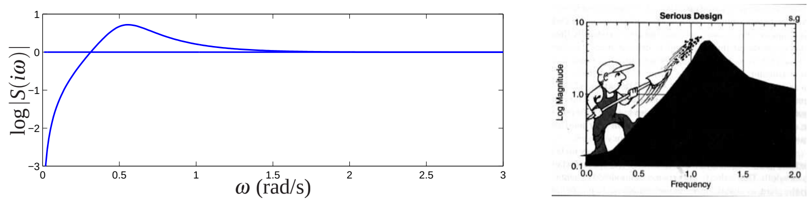
$$\int_0^\infty \log|S(i\omega)|d\omega = 0.$$

This formula can be given a nice geometric interpretation as illustrated in Figure 11.14, which shows $\log|S(i\omega)|$ as a function of ω. The area over the horizontal axis must be equal to the area under the axis when frequency is plotted on a *linear* scale. Thus if we wish to make the sensitivity smaller up to some frequency ω_{sc} we must balance this by increased sensitivity above ω_{sc}. Control system design can be viewed as trading the disturbance attenuation at some frequencies for disturbance amplification at other frequencies.

There is an analogous result for the complementary sensitivity function which tells that

$$\int_0^\infty \frac{log|T(i\omega)|}{\omega^2}\,d\omega = \pi\sum\frac{1}{z_i}, \tag{11.20}$$

where the summation is over all right half plane zeros. Notice that slow right half plane zeros are worse than fast ones and that fast right half plane poles are worse



(a) **(b)**

Figure 11.14: Interpretation of the *waterbed effect*. The function $\log|S(i\omega)|$ is plotted versus ω in linear scales in (a). According to Bode's integral formula (11.19) the area of $\log|S(i\omega)|$ above zero must be equal to the area below zero. Gunter Stein's interpretation of design as a trade-off of sensitivities at different frequencies is shown in (b) (from [181]).

than slow ones.

Example 11.11 X29 aircraft

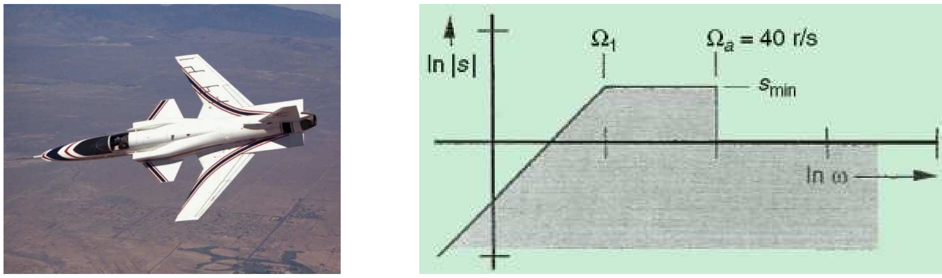
As an example of the application of Bode's integral formula, we present an analysis of the control system for the X-29 aircraft (see Figure 11.15), which has an unusual configuration of aerodynamic surfaces that are designed to enhance its maneuverability. This analysis was originally carried out by Gunter Stein in his article "Respect the Unstable" [181], which is also the source of the quote at the beginning of this chapter.

To analyze this system, we make use of a small set of parameters that describe the key properties of the system. The X-29 has longitudinal dynamics that are very similar to the inverted pendulum dynamics (Example **??**) and, in particular, have a pair of poles at approximately $p = \pm 6$ and a zero at $z = 26$. The actuators that stabilize the pitch have a bandwidth of $\omega_a = 40$ rad/s and the desired bandwidth of the pitch control loop is $\omega_1 = 3$ rad/s. Since the ratio of the zero to the pole is only 4.3 we may expect that is may be difficult to achieve the specifications.

To evaluate the achievable performance, we seek to choose the control law such that the sensitivity function is small up to the desired bandwidth and has a value of no greater than M_s beyond that value. Because of the Bode integral formula, we know that M_s must be greater than 1 to balance the small sensitivity at low frequency. We thus ask whether or not we can find a controller that has the shape shown in Figure 11.15b and seek to find the smallest value of M_s that achieves this. Note that the sensitivity above the frequency ω_a is not specified since we have no actuator authority at that frequency. However, assuming that the process dynamics fall off at high frequency, the sensitivity at high frequency will approach 1. Thus, we desire to design a closed loop system that has low sensitivity at frequencies below ω_1 and sensitivity that is not too large between ω_1 and ω_a.

From Bode's integral formula, we know that whatever controller we choose, equation (11.19) must hold. We will assume that the sensitivity function is given by

$$|S(i\omega)| = \begin{cases} \frac{\omega M_s}{\omega_1} & \omega \le \omega_1 \\ M_s & \omega_1 \le \omega \le \omega_a, \end{cases}$$



(a) **(b)**

Figure 11.15: X-29 flight control system. The aircraft makes use of forward swept wings and a set of canards on the fuselage to achieve high maneuverability. The figure on the right shows the desired sensitivity for the closed loop system. We seek to use our control authority to shape the sensitivity curve so that we have low sensitivity (good performance) up to frequency ω_1 by creating higher sensitivity up to our actuator bandwidth ω_a.

corresponding Figure 11.15b. If we further assume that $|L(s)| < \delta/\omega^2$ for frequencies larger than the actuator bandwidth, Bode's integral becomes

$$\begin{aligned}\int_0^{\infty} \log|S(i\omega)|\,d\omega &= \int_0^{\omega_a} \log|S(i\omega)|\,d\omega + \delta \\ &= \int_0^{\omega_1} \log\frac{\omega M_s}{\omega_1}\,d\omega + (\omega_a - \omega_1)\log M_s + \delta = \pi p.\end{aligned}$$

If we ignore the small contribution from δ, we can solve for M_s in terms of the remaining parameters of the system,

$$M_s = e^{(\pi p + \omega_1)/\omega_a}.$$

This formula tells us what the achievable value of M_s will be for the given control specifications. In particular, using $p = 6$, $\omega_1 = 3$ and $\omega_a = 40$ rad/s we get that $M_s = 1.75$, which means that in the range of frequencies between ω_1 and ω_a, disturbances at the input to the process dynamics (such as wind) will be amplified by a factor of 1.75 in terms of their effect on the aircraft.

Another way to view these results is to compute the phase margin that corresponds to the given level of sensitivity. Since the peak sensitivity normally occurs at or near the crossover frequency, we can compute the phase margin corresponding to $M_s = 1.75$. As shown in Exercise 11.16 the maximum achievable phase margin for this system is approximately 35°, which is below the usual design limit in aerospace systems of 45°. Hence for this system it is not possible to obtain high performance and robustness at the same time, unless more actuator authority is available.

∇

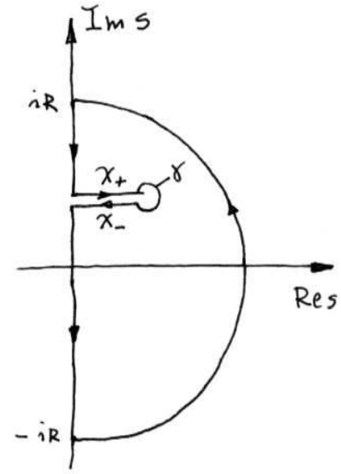


Figure 11.16: Contour used to prove Bode's theorem. For each right half plane pole we create a path from the imaginary axis that encircles the pole as shown in the figure To avoid clutter we have shown only one of the paths that enclose one right half plane.

Derivation of Bode's Formula

This is a technical section which requires some knowledge of the theory of complex variables, in particular contour integration. Assume that the loop transfer function has distinct poles at $s = p_k$ in the right half plane and that $L(s)$ goes to zero faster than $1/s$ for large values of s.

Consider the integral of the logarithm of the sensitivity function $S(s) = 1/(1+L(s))$ over the contour shown in Figure 11.16. The contour encloses the right half plane except the points $s = p_k$ where the loop transfer function $L(s) = P(s)C(s)$ has poles and the sensitivity function $S(s)$ has zeros. The direction of the contour is counter-clockwise.

The integral of the log of the sensitivity function around this contour is given by

$$\begin{aligned}\int_\Gamma \log(S(s))\,ds &= \int_{iR}^{-iR} \log(S(s))\,ds + \int_R \log(S(s))\,ds + \sum_k \int_\gamma \log(S(s))\,ds\\ &= I_1 + I_2 + I_3 = 0,\end{aligned}$$

where R is a large semicircle on the right and γ_k is the contour starting on the imaginary axis at $s = \operatorname{Im} p_k$ and a small circle enclosing the pole p_k. The integral is zero because the function $\log S(s)$ is regular inside the contour. We have

$$I_1 = -i\int_{-iR}^{iR} \log(S(i\omega))d\omega = -2i\int_0^{iR} \log(|S(i\omega)|)d\omega$$

because the real part of $\log S(i\omega)$ is an even function and the imaginary part is an odd function. Furthermore we have

$$I_2 = \int_R \log(S(s))\,ds = \int_R \log(1+L(s))\,ds \approx \int_R L(s)\,ds.$$

Since $L(s)$ goes to zero faster than $1/s$ for large s the integral goes to zero when the radius of the circle goes to infinity.

Next we consider the integral I_3. For this purpose we split the contour into three parts X_+, γ and X_- as indicated in Figure 11.16. We can then write the integral as

$$I_3 = \int_{X_+} \log S(s)\,ds + \int_\gamma \log S(s)\,ds + \int_{X-} \log S(s)\,ds.$$

The contour γ is a small circle with radius r around the pole p_k. The magnitude of the integrand is of the order $\log r$ and the length of the path is $2\pi r$. The integral thus goes to zero as the radius r goes to zero. Furthermore, making use of the fact that X_- is oriented oppositely from X_+, we have

$$\int_{X_+} \log S(s)\,ds + \int_{X_-} \log S(s)\,ds = \int_{X_+} \big(\log S(s) - \log S(s-2\pi i)\big)\,ds = 2\pi p_k.$$

Since $|S(s)| = |S(s-2\pi i)|$ we have

$$\log S(s) - \log S(s-2\pi i) = \arg S(s) - \arg S(s-2\pi i) = 2\pi$$

and we find that

$$I_3 = 2\pi\Sigma p_k$$

Letting the small circles go to zero and the large circle go to infinity and adding the contributions from all right half plane poles p_k gives

$$I_1 + I_2 + I_3 = -2i\int_0^R \log|S(i\omega)|d\omega + \sum_k 2\pi p_k = 0.$$

which is Bode's formula (11.19).

11.6 DESIGN EXAMPLE

In this section we carry out a detailed design example that illustrates the main techniques in this chapter.

Example 11.12 Lateral control of a vectored thrust aircraft
The problem of controlling the motion of a vertical take off and landing (VTOL) aircraft was introduced in Example 2.9 and in Example 11.6, where we designed a controller for the roll dynamics. We now wish to control the position of the aircraft, a problem that requires stabilization of both the attitude and position. To control the lateral dynamics of the vectored thrust aircraft, we make use of a "inner/outer" loop design methodology, as illustrated in Figure 11.17. This diagram shows the process dynamics and controller divided into two components: an "inner loop" consisting of the roll dynamics and control and an "outer loop" consisting of the

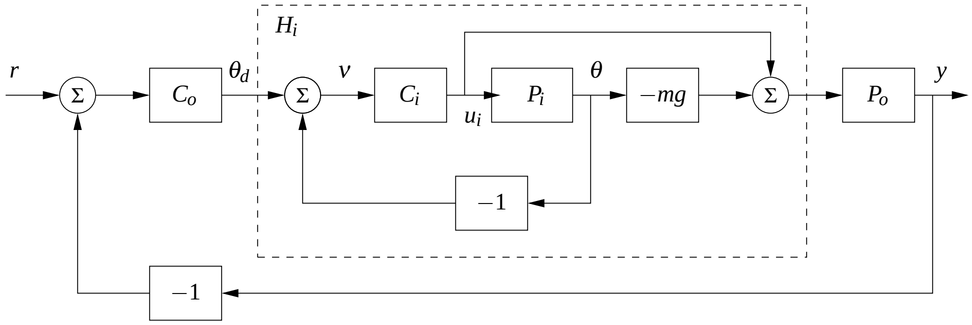


Figure 11.17: Inner/outer control design for a vectored thrust aircraft. The inner loop H_i controls the roll angle of the aircraft using the vectored thrust. The outer loop controller C_o commands the roll angle to regulate the lateral position. The process dynamics are decomposed into inner loop (P_i) and outer loop (P_o) dynamics, which combine to form the full dynamics for the aircraft.

.

lateral position dynamics and controller. This decomposition follows the block diagram representation of the dynamics given in Exercise 8.11.

The approach that we take is to design a controller C_i for the inner loop so that the resulting closed loop system H_i provides fast and accurate control of the roll angle for the aircraft. We then design a controller for the lateral position that uses the approximation that we can directly control the roll angle as an input to the dynamics controlling the position. Under the assumption that the dynamics of the roll controller are fast relative to the desired bandwidth of the lateral position control, we can then combine the inner and outer loop controllers to get a single controller for the entire system. As a performance specification for the entire system, we would like to have zero steady state error in the lateral position, a bandwidth of approximately 1 rad/s and a phase margin of 45°.

For the inner loop, we choose our design specification to provide the outer loop with accurate and fast control of the roll. The inner loop dynamics are given by

$$P_i = H_{\theta u_1} = \frac{r}{Js^2 + cs}.$$

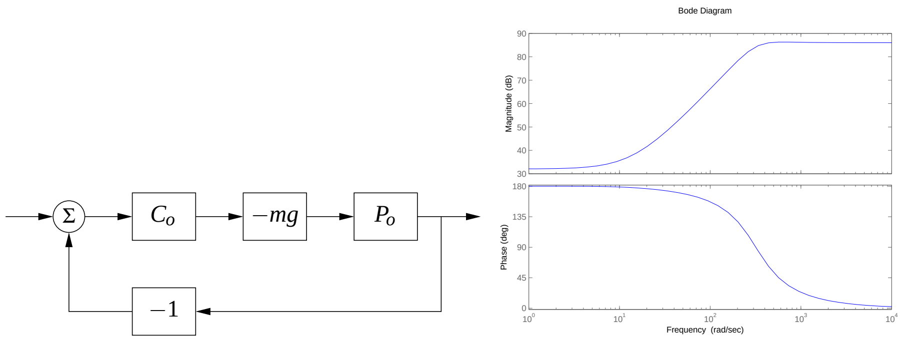
We choose the desired bandwidth to be 10 rad/s (10 times the outer loop) and the low frequency error to be no more than 5%. This specification is satisfied using the lead compensator of Example 11.6 designed previously, so we choose

$$C_i(s) = k\frac{s+a}{s+b} \qquad a = 2, \qquad b = 25, \qquad k = 1.$$

The closed loop dynamics for the system satisfy

$$H_i = \frac{C_i}{1+C_iP_i} - mg\frac{C_iP_i}{1+C_iP_i} = \frac{C_i(1-mgP_i)}{1+C_iP_i}.$$

A plot of the magnitude of this transfer function is shown in Figure 11.18 and we see that it is a good approximation up to 10 rad/s.



(a) **(b)**

Figure 11.18: Outer loop control design for a vectored thrust aircraft. The outer loop approximates the roll dynamics as a state gain $-mg$. The Bode plot for the roll dynamics are shown on the right, indicating that this approximation is accurate up to approximately 10 rad/s.

To design the outer loop controller, we assume the inner loop roll control is perfect, so that we can take θ_d as the input to our lateral dynamics. Following the diagram shown in Exercise 8.11, the outer loop dynamics can be written as

$$P(s) = H_i(0)P_o(s) = \frac{H_i(0)}{ms^2},$$

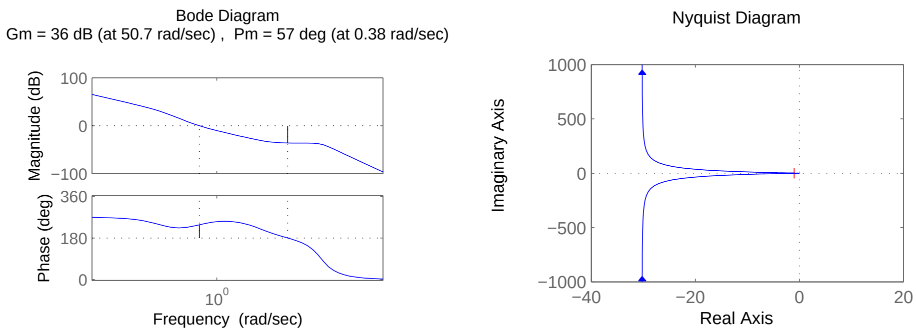
where we replace $H_i(s)$ with $H_i(0)$ to reflect our approximation that the inner loop will eventually track our commanded input. Of course, this approximation may not be valid and so we must verify this when we complete our design.

Our control goal is now to design a controller that gives zero steady state error in x and has a bandwidth of 1 rad/s. The outer loop process dynamics are given by a second order integrator and we can again use a simple lead compensator to satisfy the specifications. We also choose the design such that the loop transfer function for the outer loop has $|L_o| < 0.1$ for $\omega > 10$ rad/s so that the H_i dynamics can be neglected. We choose the controller to be of the form

$$C_o(s) = -k_o\frac{s+a_o}{s+b_o},$$

with the negative sign to cancel the negative sign in the process dynamics. To find the location of the poles, we note that the phase lead flattens out at approximately $b/10$. We desire phase lead at crossover and we desire the crossover at $\omega_{gc} =$ 1 rad/s, so this gives $b_o = 10$. To insure that we have adequate phase lead, we must choose a_o such that $b_o/10 < 10a_o < b_o$, which implies that a_o should be between 0.1 and 1. We choose $a_o = 0.3$. Finally, we need to set the gain of the system such that at crossover the loop gain has magnitude one. A simple calculation shows that $k_o = 0.8$ satisfies this objective. Thus, the final outer loop controller becomes

$$C_o(s) = 0.8\frac{s+0.3}{s+10}.$$



(a) **(b)**

Figure 11.19: Inner/outer loop controller for a vectored thrust aircraft. The Bode and Nyquist plots for the transfer function for the combined inner and outer loop transfer functions are show. The system has a phase margin of 68° and a gain margin of 6.2.

Finally, we can combine the inner and outer loop controllers and verify that the system has the desired closed loop performance. The Bode and Nyquist plots corresponding to Figure 11.17 with the inner and outer loop controllers is shown in Figure 11.19 and we see the specifications are satisfied. In addition, we show the Gang of Four in Figure 11.20 and we see that the transfer functions between all inputs and outputs are reasonable.

The approach of splitting the dynamics into an inner and outer loop is common in many control applications and can lead to simpler designs for complex systems. Indeed, for the aircraft dynamics studied in this example, it is very challenging to directly design a controller from the lateral position x to the input u_1. The use of the additional measurement of θ greatly simplifies the system requirements allows the design to be broken up into simpler pieces.

∇

11.7 FURTHER READING

Design by loop shaping was a key element of the early development of control and systematic design methods were developed, see James, Nichols and Philips [108],, Chestnut and Mayer [51], Truxal [190], and Thaler [187]. Loop shaping is also treated in standard textbooks such as Franklin, Powell and Emami-Naeini [80], Dorf and Bishop [60], Kuo and Golnaraghi [130] and Ogata [158]. Systems with two degrees of freedom were developed by Horowitz [102], who also discussed limitations of poles and zeros in the right half plane. Fundamental results on limitations are given in Bode [41]; more recent presentations are found in Goodwin, Graebe and Salgado [88]. The treatment in Section 11.5 is based on [16]. Much of the early work was based on the loop transfer function; the importance of the sensitivity functions appeared in connection with the development in the 1980s

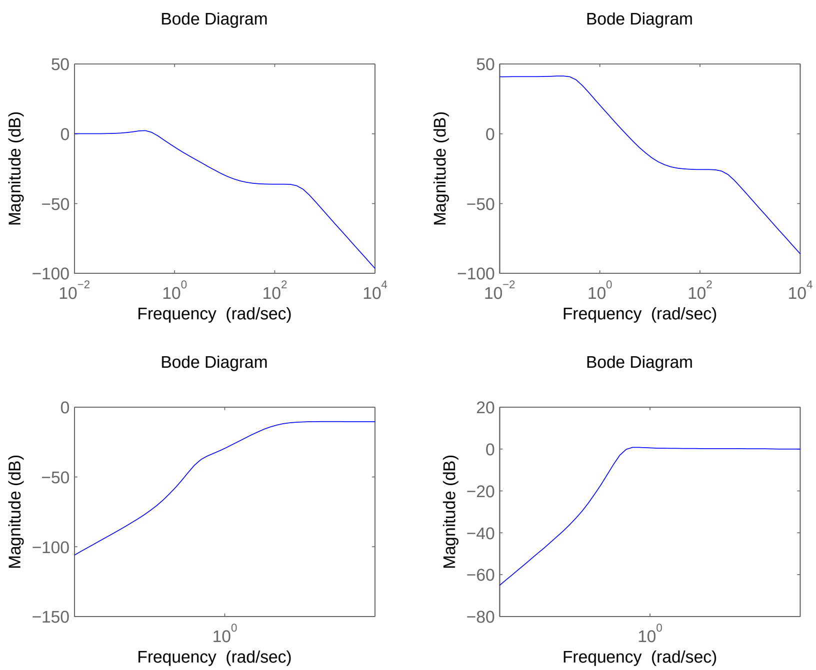


Figure 11.20: Gang of Four for vector thrust aircraft system.

which resulted in the so called H_∞ design methods. A compact presentation is given in the text by Doyle, Frances and Tannenbaum [63] and Zhou, Doyle and Glover [203]. Loop shaping was integrated with the robust control theory in MacFarlane and Glover [137] and Vinnicombe [192]. Comprehensive treatments of control system design are given in Maciejowski [138] and Goodwin, Graebe and Salgado [88].

EXERCISES

11.1 Consider the system in Figure 11.1 give all signal pairs which are related by the transfer functions $1/(1+PC)$, $P/(1+PC)$, $C/(1+PC)$ and $PC/(1+PC)$.

11.2 (Cancellation of unstable process pole) Consider the system in Example 11.1. Choose the parameters $a=-1$ compute time and frequency responses for all transfer functions in the Gang of Four for controllers with $k=0.2$ and $k=5$.

11.3 (Equivalence of Figure 11.1 and 11.2) Show that the system in Figure 11.1 can be represented by 11.2 by proper choise of the matrices $\mathscr{P}$ and $\mathscr{C}$.

11.4 (Sensitivity of feedback and feedforward) Consider the system in Figure 11.1, let G_{yr} be the transfer function relating measured signal y to reference r. Compute the sensitivities of G_{yr} with respect to the feedforward and feedback transfer functions F and C ($\partial G_{yr}/\partial F$ and $\partial G_{yr}/\partial C$).

11.5 (Equivalence of controllers with two degrees of freedom) Show that the systems in Figure 11.1 and Figure 11.3 give the same responses to command signals if $F_m C + F_u = CF$.

11.6 (Rise-time-bandwidth product)Prove Consider a stable system with the transfer function $G(s)$, where $G(0) = 1$. Define the rise time T_r as the inverse of the largest slope of the step response and the bandwidth as $\omega_B = (1/2)\int_{\infty}^{\infty} |G(i\omega)| d\omega$. Show that $\omega_B T_r \geq \pi$.

11.7 Regenerate the controller for the system in Example 11.6 and use the frequency responses for the Gang of Four to show that the performance specification is met.

11.8 Let $f(t) = f(0) + t f'(0) + t^2/2 f''(0) + \cdots$ be a Taylor series expansion of the time function f, show that

$$F(s) = \frac{1}{s} f(0) + \frac{1}{s^2} f'(0) + \frac{1}{2s^3} f^{''}(0) + \cdots$$

11.9 Exercise 11.8 shows that the behavior of a time function for small t is related to the Laplace transform for large s. Show that the behavior of a time function for large t is related to the Laplace transform for small s.

11.10 Consider the feedback system shown in Figure 11.1. Assume that the reference signal is constant. Let y_{ol} be the measured output when there is no feedback and y_{cl} be the output with feedback. Show that

$$Y_{cl}(s) = S(s) Y_{ol}(s)$$

where S is the sensitivity function.

11.11 (Approximate expression for noise sensitivity) Show that the effect of noise on the control signal for the system in Exercise 11.4 can be approximated by

$$CS \approx C \approx \frac{k_d s}{(sT_d)^2/2 + sT_d + 1}$$

Show that using this approximation the largest value of $|CS(i\omega)|$ is k_d/T_f and that it occurs for $\omega = \sqrt{2}/T_f$.

11.12 Show that the nonminimum phase part of the transfer function $P = e^{-sT}$ for a time delay has the the phase lag ωT which implies that the gain crossover frequency must be chosen so that $\omega_{gc} T < \varphi_l$. Also use the approximation $e^{-sT} \approx \frac{1-sT/2}{1+sT/2}$ so show that a time delay is similar to a system with a right half plane zero at $s = 2/T$. A slow zero thus corresponds to a long time delay.

11.13 (The pole zero ratio) Consider a process with the transfer function

$$P(s) = k\frac{a-s}{s-b}$$

with positive a and b. Show that the the closed loop system with unit feedback is either $b/a < k < 1$ or $1 < k < b/a$.

11.14 (Pole in the right half plane and time delay) The non-minimum phase part of the transfer function for a system with one pole in the right half plane and a time delay T is

$$P_{nmp}(s) = \frac{s+p}{s-p} e^{-sT}. \tag{11.21}$$

Using the gain crossover inequality, compute the limits on the achievable bandwidth of the system.

11.15 (Integral formula for complementary sensitivity) Prove the formula (11.20) for the complementary sensitivity.

11.16 (Phase margin formulas) Show that if the relationship between the phase margin and the magnitude of the sensitivity function at crossover is given by

$$S_{\min} = \frac{1}{2\sin(\varphi_m/2)}$$

11.17 (Limitations on achievable phase lag) Derive analytical formulas corresponding to the plots in Figure 11.13.

11.18 (Design of a PI controller) Consider a system with process transfer function

$$P(s) = \frac{1}{(s+1)^4}, \tag{11.22}$$

and a PI controller with the transfer function

$$C(s) = k_p + \frac{k_i}{s} = k\frac{1+sT_i}{sT_i}.$$

The controller has high gain at low frequencies and its phase lag is negative for all parameter choices. To achieve good performance it is desirable to have large gain at low frequencies and a high crossover frequency.

11.19 (Stabilization of inverted pendulum with visual feedback) Consider stabilization of an inverted pendulum based on visual feedback using a video camera with 50 Hz frame rate. Let the effective pendulum length be l. Use the gain crossover inequality to determine the minimum length of the pendulum that can be stabilized if we desire a phase lag φ_l of no more than $90°$.