Feedback Systems

An Introduction for Scientists and Engineers
SECOND EDITION

Karl Johan Åström
Richard M. Murray

Version v3.0h (16 Nov 2016)

This is the electronic edition of *Feedback Systems* and is available from http://www.cds.caltech.edu/~murray/FBS. Hardcover editions may be purchased from Princeton University Press, http://press.princeton.edu/titles/8701.html.

PRINCETON UNIVERSITY PRESS
PRINCETON AND OXFORD

Chapter Twelve
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, 1989 [Ste03].

In this chapter we continue to explore the use of frequency domain techniques with a focus on the 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 time delays and right half-plane poles and zeros.

12.1 Sensitivity Functions

In the previous chapter, we considered the use of proportional-integral-derivative (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 12.1. The system loop is composed of two components: the process and the controller. The controller itself has two blocks: the feedback block C and the feedforward block F. There are two disturbances acting on the process, the load disturbance v and the measurement noise w. The load disturbance represents disturbances that drive the process away from its desired behavior, while the measurement noise represents disturbances that corrupt information about the process given by the sensors. For example in

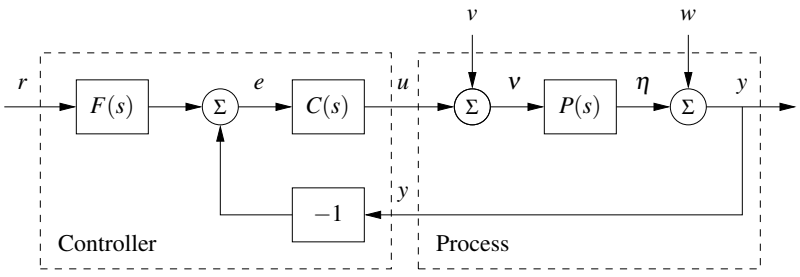


Figure 12.1: Block diagram of a basic feedback loop with two degrees of freedom. The controller has a feedback block C and a feedforward block F. The external signals are the reference signal r, the load disturbance v and the measurement noise w. The process output is η, and the control signal is u.

cruise control the major load disturbance is changes in the slope of the road and measurement noise is caused by the electronics that convert pulses measured on a rotating shaft to a velocity signal. The load disturbances typically have low frequencies, lower than the servo bandwidth, while measurement noise typically has higher frequencies. In the figure, 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 w. 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 v and the measurement noise w—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.

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 12.1:

$$\begin{pmatrix} y \\ \eta \\ v \\ u \\ e \end{pmatrix} = \begin{pmatrix} \frac{PCF}{1+PC} & \frac{P}{1+PC} & \frac{1}{1+PC} \\ \frac{PCF}{1+PC} & \frac{P}{1+PC} & \frac{-PC}{1+PC} \\ \frac{CF}{1+PC} & \frac{1}{1+PC} & \frac{-C}{1+PC} \\ \frac{CF}{1+PC} & \frac{-PC}{1+PC} & \frac{-C}{1+PC} \\ \frac{F}{1+PC} & \frac{-P}{1+PC} & \frac{-1}{1+PC} \end{pmatrix} \begin{pmatrix} r \\ v \\ w \end{pmatrix}. \tag{12.1}$$

In addition, we can write the transfer function for the error between the reference r and the output η, which satisfies

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

Notice that many of the transfer functions in (12.1) are the same. Only four transfer functions are required to describe how the system reacts to load disturbances and measurement noise, three additional transfer functions are required to describe how the system responds to reference signals.

The special case of $F = 1$ is called a system with (pure) *error feedback* because all control actions are based on feedback from the error. In this case the system (12.1) is completely characterized by four transfer functions, which are called the *sensitivity functions*:

$$\begin{aligned} S &= \frac{1}{1+PC} && \text{sensitivity function} & \qquad PS &= \frac{P}{1+PC} && \text{load sensitivity function} \\ T &= \frac{PC}{1+PC} && \text{complementary sensitivity function} & \qquad CS &= \frac{C}{1+PC} && \text{noise sensitivity function} \end{aligned} \tag{12.2}$$

These transfer functions and their equivalent systems are called the *Gang of Four*. The load 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 in understanding the performance of feedback systems for the purposes of both analysis and design. The transfer functions are also used to formulate requirements on the closed loop system.

When $F \neq 1$ there are three additional transfer functions in (12.1).

$$PS = \frac{P}{1+PC}, \quad CSF = \frac{PCT}{1+PC}, \quad TF = \frac{PCF}{1+PC} \tag{12.3}$$

To summarize, we find that the closed loop system can be characterized by seven transfer functions given by (12.2) and (12.3), which we call the *Gang of Seven*.

The closed loop system is stable if all transfer functions (12.2) are stable, which is called *internal stability*. If there are no pole-zero cancellations in the product PC, the closed loop system is stable if the rational function $1 + PC$ has all its zeros in the left half plane.

Analyzing the Gang of Seven, we find that the feedback controller C influences the effects of load disturbances and measurement noise. Notice that measurement noise enters the process via the feedback. In Section 13.2 it will be shown that the controller influences the sensitivity of the closed loop to process variations. The feedforward part F of the controller influences only the response to command signals.

In Chapter 10 we focused on the loop transfer function, and we found that its

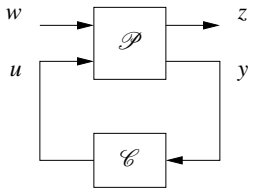


Figure 12.2: A more general representation of a feedback system. The process input u represents the control signal, which can be manipulated, and the process input w represents other signals that influence the process. The process output y is the vector of measured variables and z are other signals of interest.

properties gave useful insights into the properties of a system. To make a proper assessment of a feedback system it is necessary to consider the properties of all the transfer functions (12.2) and (12.3) in the Gang of Seven (or the Gang of Four, if there is no feedforward controller), as illustrated in the following example.

Example 12.1 The loop transfer function gives only 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 the 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 v can lead to an unbounded output, which is clearly not desirable. ∇

The system in Figure 12.1 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 the 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 12.2, which has only 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 12.1 can be captured by choosing $w = (r,d,n)$ and $z = (e,u,\nu,\eta,\varepsilon)$. The process transfer function $\mathscr{P}$ from w to z is a 5×3 matrix, and the controller transfer function $\mathscr{C}$ from y,r to u is a 2×1 matrix; compare with Exercise 12.3.

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

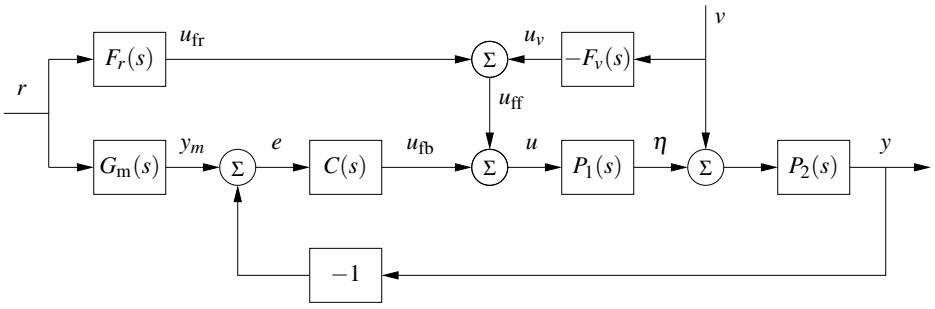


Figure 12.3: Block diagram of a system with feedforward compensation for improved response to reference signals and measured disturbances (2 DOF system). Three feedforward elements are present: $G_{\text{m}}(s)$ sets the desired output value, $F_r(s)$ generates the feedforward command u_{fr} to improve command signal response and $F_v(s)$ generates the feedforward signal u_{fv} that reduces the effect of the measured disturbance v.

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.

12.2 Feedforward Design

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. Design of feedforward for controllers based on state feedback and observers was developed in Section 8.5 (see Figure 8.11). Section 11.5 presented set point weighting as simple form of feedforward for PID controllers; see equation (11.17). In this section we will use transfer functions to develop more advanced methods for feedforward design.

Figure 12.3 shows a block diagram of a system with feedback and feedforward control. The process transfer function $P(s)$ is separated in two blocks $P_1(s)$ and $P_2(s)$, where the measured disturbance v enters at the input of the block P_2. The transfer function G_{m} represents the desired response to reference signals. There are two feedforward blocks with the transfer functions F_r and F_v to deal with the reference signal r and the measured disturbances v.

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 transfer function C can be designed to give good robustness and effective disturbance attenuation, and the feedforward transfer functions F_r and F_v can be designed independently to give the dessired responses to reference signals and to reduce effects of measured disturbances.

We will first explore the response to command signals. The transfer function

$F_{yr}(s)$ from reference input r to process output y is

$$\begin{aligned} G_{yr}(s) &= \frac{P(CG_{\mathrm{m}}+F_r)}{1+PC} = G_{\mathrm{m}} + \frac{PF_r - G_{\mathrm{m}}}{1+PC} \\ &= G_{\mathrm{m}} + S(PF_r - G_{\mathrm{m}}) = TG_{\mathrm{m}} + SPF_r, \end{aligned} \tag{12.4}$$

where S is the sensitivity function and T the complementary sensitivity function, see (12.2). We can make G_{yr} close to the desired transfer function G_{m} in two different ways, by choosing the feedforward transfer function F_r so that $PF_r - G_{\mathrm{m}}$ is small, or by choosing the feedback transfer function C so that the sensitivity $S = 1/(1+PC)$ is small. Perfect feedforward compensation is obtained for

$$F_r = \frac{G_{\mathrm{m}}}{P}, \tag{12.5}$$

which gives $G_{yr} = G_{\mathrm{m}}$. Notice that the feedforward compensator F_r contains an inverse model of the process dynamics.

Next we will consider attenuation of disturbances that can be measured. The transfer function from load disturbance v to process output y is given by

$$G_{yv} = \frac{P_2(1-P_1F_v)}{1+PC} = P_2S(1-P_1F_v). \tag{12.6}$$

Recall that S is the sensitivity function (12.3). The transfer function G_{yv} can be made small in two different ways, by choosing the feedforward transfer function F_r so that $PF_r - G_{\mathrm{m}}$ is small, or by choosing the feedback transfer function C so that the sensitivity $S = 1/(1+PC)$ is small. Perfect compensation is obtained by choosing

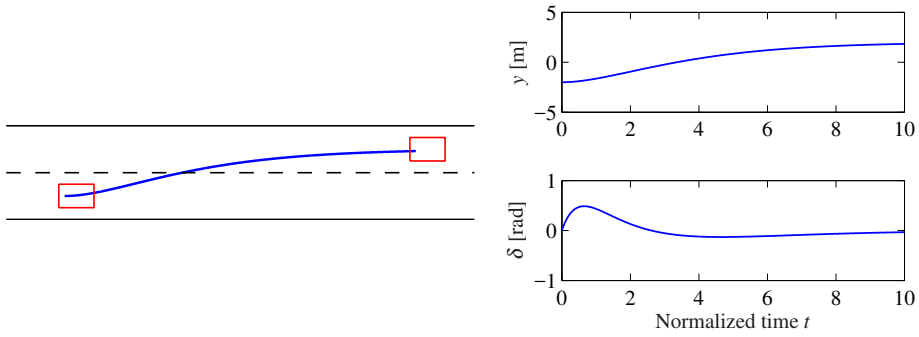
$$F_v = P_1^{-1}. \tag{12.7}$$

Design of feedforward to improve responses to command signals and disturbances using transfer functions is thus a simple task, but it requires inversion of process models. We illustrate with an example.

Example 12.2 Vehicle steering
A linearized model for vehicle steering was given in Example 7.4. The normalized transfer function from steering angle δ to lateral deviation y is $P(s) = (\gamma s+1)/s^2$. For a lane transfer system we would like to have a nice response without overshoot, and we therefore choose the desired response as $G_m(s) = a^2/(s+a)^2$, where the response speed or aggressiveness of the steering is governed by the parameter a. Equation (12.5) gives

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

which is a stable transfer function as long as $\gamma > 0$. Figure 12.4 shows the responses of the system for $a = 0.5$. The figure shows that a lane change is accomplished in about 10 vehicle lengths with smooth steering angles. The largest steering angle is slightly larger than 0.1 rad (6°). Using the scaled variables, the curve showing



(a) Overhead view (b) Position and steering

Figure 12.4: Feedforward control for vehicle steering. The plot on the left shows the trajectory generated by the controller for changing lanes. The plots on the right show the lateral deviation y (top) and the steering angle δ (bottom) for a smooth lane change control using feedforward (based on the linearized model).

lateral deviations (y as a function of t) can also be interpreted as the vehicle path (y as a function of x) with the vehicle length as the length unit. ∇

Sensitivity to Process Variations

Control design is based on knowledge of the transfer function of the process. Since there are always modeling errors we will investigate the effects of small modeling errors. We will first investigate the response to reference signals. The transfer function from reference signals to process output G_{yr} is given by equation (12.4). Taking the logarithms of this equation gives

$$\log G_{yr} = \log P + \log(F_{\mathrm{r}} + CG_{\mathrm{m}}) - \log(1 + PC).$$

Differentiating with respect to P gives the following expression for the sensitivity of the closed loop transfer function:

$$\frac{dG_{yr}}{G_{yr}} = \frac{dP}{P} - \frac{CdP}{1+PC} = \frac{1}{1+PC}\frac{dP}{P} = S\frac{dP}{P}. \tag{12.8}$$

With pure feedforward we have $C = 0$ and $S = 1$, and the relative error in G_{yr} is the same as the relative error in the process transfer function. With feedback the relative model error dP/P is multiplied by the sensitivity function S. Notice in particular that with a controller having integral action we have $S(0) = 0$. Feedback thus gives a significant reduction of the effects of modeling error at low frequencies.

Next we will investigate the transfer function G_{yv} from load disturbances v to process output, which is given by

$$G_{yv} = \frac{P_2(1 - P_1F_v)}{1+PC}. \tag{12.9}$$

	FF	FB	FB and FF
G_{yr}	G_{m}	TG_{m}	G_{m}
$\frac{dG_{yr}}{G_{yr}}$	$\frac{dP}{P}$	$S\frac{dP}{P}$	$S\frac{dP}{P}$
G_{yv}	0	SP_2	0
$\frac{dG_{yv}}{P_2}$	$-\frac{dP_1}{P_1}$	$-S\frac{dP_1}{P_1}$	$-S\frac{dP_1}{P_1}$

Table 12.1: Properties of command signal following G_{yr} and disturbance attenuation G_{yv} of closed loop systems based on feedback (FB), feedforward (FF) and combined feedback and feedforward.

For small variations dP of the process transfer function P we have

$$dG_{yv} = \frac{(1+PC)\big(dP_2(1-P_1F_v) - P_2dP_1F_v\big) - P_2(1-P_1F_v)dPC}{(1+PC)^2}$$
$$= -\frac{P_2dP_1F_v}{1+PC} + (1-P_1F_v)\frac{dP_2(1+PC) - P_2dPC}{(1+PC)^2}$$

Using the expression (12.7) for the feedforward transfer function F_r we find

$$\frac{dG_{yv}}{P_2} = -\frac{dP_1}{(1+PC)P_1} = -S\frac{dP_1}{P_1}. \tag{12.10}$$

Notice that the transfer function P_1 is the critical element; variations in P_2 are not as serious.

The properties of feedback and feedforward are summarized in Table 12.1. With feedforward control we have $G_{yr} = G_{\mathrm{m}}$ and $G_{yv} = 0$, while with feedback control we have instead $G_{yr} = TG_{\mathrm{m}}$ and $G_{yv} = SP_2$. The sensitivity to modeling errors with pure feedforward is always equal to one; feedback reduces it significantly.

Feedback and feedforward have complementary properties. Feedforward acts by matching two transfer functions, requiring precise knowledge of the process dynamics, while feedback attempts to make the error small by multiplication with the sensitivity function S. For a controller having integral action, the sensitivity function is small for low frequencies, and it is thus sufficient that the condition for ideal feedforward holds at higher frequencies. This is easier than trying to satisfy the conditions (12.5) and (12.7) for all frequencies.

Difficulties with Feedforward

The feedforward compensators in Figure 12.3 are given by

$$F_r = \frac{G_{\mathrm{m}}}{P}, \qquad F_v = \frac{1}{P_1}. \tag{12.11}$$

Both transfer functions require inversion of process transfer functions. There are problems with inversion if the process transfer function has time delays, right half plane zeros or high pole excess. Inversion of time delays require prediction, which cannot be done perfectly except in the situation when the command signal is known in advance. The inverse process transfer function is unstable if the process transfer functions has zeros in the right half plane. The inverse requires differentiation if the pole excess of the process transfer funtion is greater than zero. We must then require that the reference signal is sufficiently smooth and there may also be problems with noise. Approximate inverses have to be used when the inverse is unstable.

There is an extra freedom when finding the transfer function F_r because it also contains the transfer function G_{m}, which specifies the ideal behavior. A stable feedforward transfer function can be obtained if G_{m} has the same time delays and right half plane zeros as the process. We illustrate with an example.

Example 12.3 Process with a right half plane zero
Let the process and the desired response have the transfer functions

$$P(s) = \frac{1-s}{(s+1)^2}, \qquad G_{\mathrm{m}}(s) = \frac{\omega_m^2(1-s)}{s^2+2\zeta_c\omega_m s+\omega_m^2}.$$

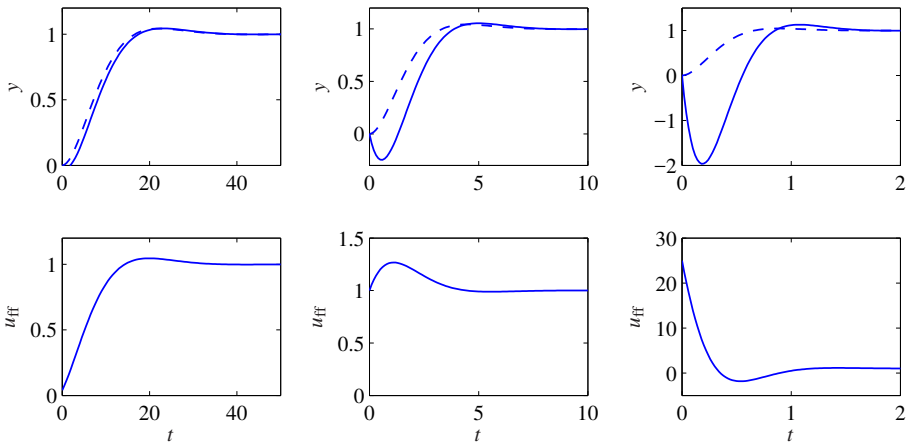
Since the process has a right half plane zero at $s = 1$ the desired transfer function $G_{\mathrm{m}}(s)$ must have the same zeros to avoid having an unstable feedforward transfer function F_{r}. Equation (12.5) gives the feedforward transfer function

$$F_{\mathrm{r}}(s) = \frac{\omega_m^2(s+1)^2}{s^2+2\zeta_c\omega_m s+\omega_m^2}. \tag{12.12}$$

Figure 12.5 shows the outputs y and the feedforward signals u_{ff} for different values of ω_m. The response to the command signal goes in the wrong direction initially because of the right half plane zero at $s = 1$. This effect, called *inverse response*, is barely noticeable if the response is slow ($\omega_m = 1$) but increases with increasing response speed. For $\omega_m = 5$ the undershoot is more than 200%. A reasonable choice of ω_m is in the range 0.2 to 0.5. Notice that the feedforward transfer function (12.12) is also obtained if the process and the desired model have the transfer functions

$$P(s) = \frac{1}{(s+1)^2}, \qquad G_{\mathrm{m}}(s) = \frac{\omega_m^2}{s^2+2\zeta_c\omega_m s+\omega_m^2}.$$

The corresponding responses are shown as dashed lines in Figure 12.5. When there is no right half plane zero it is thus possible to get nice fast responses. A comparison of the full and solid lines in Figure 12.5 shows that the presence of a right half plane zero in the process transfer function is a severe limitation. The control signals for different values of ω_m differ significantly. To understand what happens we note that $F_{\mathrm{r}}(0) = 1$ and that $F_{\mathrm{r}}(\infty) = \omega_m^2$. For a unit step reference signal $r = 1$, the initial value of the control signal is $u_{\mathrm{ff}}(0) = F_{\mathrm{r}}(\infty) = \omega_m^2$ and the final value is $u_{\mathrm{ff}}(\infty) = F_{\mathrm{f}}(0) = 1$. For $\omega_m = 0.2$ the control signal grows from 0.04 to the final value 1 with a small overshoot. For $\omega_m = 1$ the control signal starts from 1, has



(a) $\omega_m = 0.2$ (b) $\omega_m = 1$ (c) $\omega_m = 5$

Figure 12.5: Outputs y (top plots) and feedforward signals u_{ff} (lower plots) for a unit step command signal. The design parameter has the values $\omega_m = 0.2$ (left), 1 (center) and 0.5 (right) for a unit step command in the reference signal. The dashed curve shows the response that could be achieved if the process did not have the right half plane zero.

an overshoot and then settles on the final value 1. For $\omega_m = 5$ the control signals starts at 25 and decays towards the final value 1 with an undershoot. ∇

Approximate Inverses

Processes with right half plane zeros do not have stable inverses. To design feedforward compensators for such systems we need approximate inverses that are stable. The following theorem, which is presented without proof, is useful.

Theorem 12.1 (Approximate inverse)**.** *Let the rational transfer function $G(s)$ have all its poles in the left half plane and no zeros on the imaginary axis. Factor the transfer function as $G(s) = G^+(s)G^-(s)$, where $G^+(s)$ has all its zeros in the left half plane and $G^-(s)$ has all its zeros in the right half plane. An approximate stable inverse of $G(s)$ that minimizes the mean square error for a step input is*

$$G^\dagger(s) = \frac{1}{G^+(s)G^-(-s)}. \tag{12.13}$$

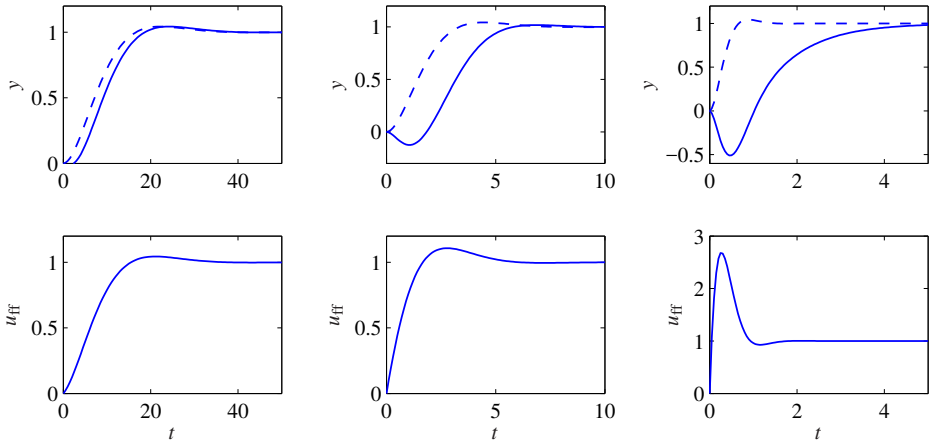
We illustrate the theorem with an example.

Example 12.4 Approximate inverses
Let the transfer functions of the process and the model be

$$P(s) = \frac{1-s}{(s+1)^2}, \qquad G_{\text{m}}(s) = \frac{\omega_m^2}{s^2 + 2\zeta_c\omega_m s + \omega_m^2}$$

(compare with Example 12.3). The process transfer function can be factored as

$$P^-(s) = 1-s, \quad P^+(s) = \frac{1}{s+1)^2}.$$



(a) $\omega_m = 0.2$ (b) $\omega_m = 1$ (c) $\omega_m = 5$

Figure 12.6: Feedforward design based on an approximate inverse. Outputs y (top plots) and feedforward signals u_{ff} (lower plots) for a unit step command signal. The design parameter has the values $\omega_m = 0.2$ (left) 1 (center) and 0.5 (right) for a unit step command in the reference signal. The dashed curve shows the response that could be achieved if the process did not have the right half plane zero.

Theorem 12.1 then gives the following approximate inverse

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

The feedforward transfer function is then

$$F_r(s) = \frac{G_{\text{m}}(s)}{P^{\dagger}(s)} = \frac{\omega_m^2(s+1)}{s^2+2\zeta_c\omega_m s+\omega_m^2},$$

compared with equation (12.12). The transfer function from reference r to output y is then

$$G_{yr}(s) = P(s)F_r(s) = \frac{1-s}{(s^2+2\zeta_c\omega_m s+\omega_m^2)(s+1)}.$$

Figure 12.6 iillustrates the step responses for different values of ω_m.

Comparing Figures 12.5 and 12.6 we find that there are small differences for $\omega_m = 0.2$. and very large differences for $\omega_m = 5$. Notice in particular the differences between the feedforward signal u_{ff}. The design based on the approximate inverse has smaller undershoot but the response has larger lags. ∇

In summary, feedforward can be used to improve the responses to command signals and to reduce the effects of load disturbances that can be measured. Feedforward compensation has inheriently unit sensitivity to model uncertainty, a combination with feedback can reduce the effects of model uncertainty significantly. A feedback controller with integral action reduces the effects of model errors at low frequencies. Design of feedforward controllers is a simple manipulation of trans-

fer functions. There are limitations if the process has time delays, right half plane zeros and high pole excess.

12.3 Performance Specifications

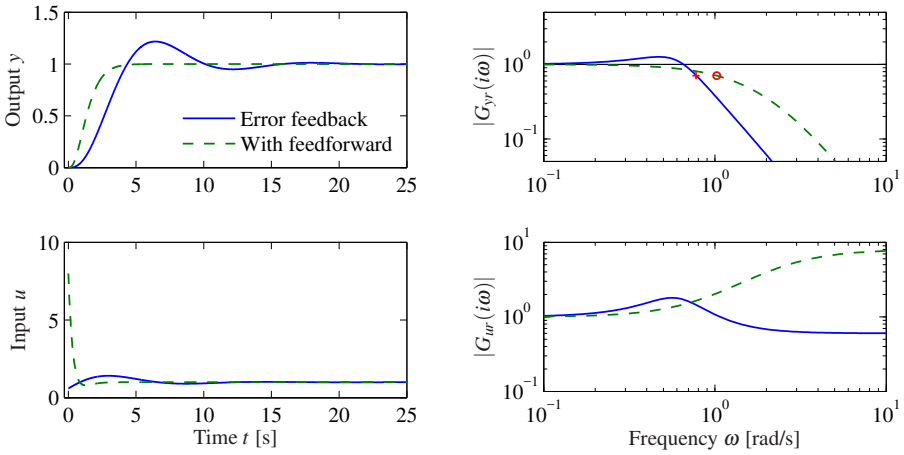
A key element of the control design process is how we specify the desired performance of the system. It is also important for users to understand performance specifications so that they know what to ask for and how to test a system. Specifications are often given in terms of robustness to process variations and responses to reference signals and disturbances. They can be given in terms of both time and frequency responses. Specifications for the step response to reference signals were given in Figure 6.9 in Section 6.3 and in Section 7.3. Robustness specifications based on frequency domain concepts were provided in Section 10.3 and will be considered further in Chapter 13. The specifications discussed previously were based on the loop transfer function. Since we found in Section 12.1 that a single transfer function 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 Seven.

The transfer function gives a good characterization of the linear behavior of a system. To provide specifications it is desirable to capture the characteristic properties of a system with a few parameters. Common features of step responses are overshoot, rise time and settling time, as shown in Figure 6.9. Common features of frequency responses are resonant peak, peak frequency, gain crossover frequency and bandwidth. A *resonant peak* is a maximum of the gain, and the peak frequency is the corresponding frequency. The *gain crossover frequency* is the smallest frequency where the open loop gain is equal one. The *bandwidth* is defined as the frequency range where the closed loop gain is $1/\sqrt{2}$ of the low-frequency gain (low-pass), mid-frequency gain (band-pass) or high-frequency gain (high-pass). The crossover frequency and the bandwidth are only well defined if the curves are monotone; if this is not the case the bandwidth is typically defined as the lowest frequency.

There are interesting relations between specifications in the time and frequency domains. Roughly speaking, the behavior of time responses for short times is related to the behavior of frequency responses at high frequencies, and vice versa. The precise relations are given by the Laplace transform.

Response to Reference Signals

Consider the basic feedback loop in Figure 12.1. 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.



(a) Step responses (b) Frequency responses

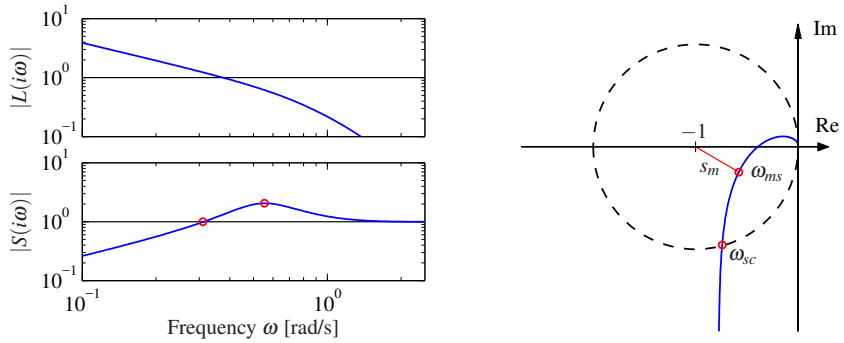
Figure 12.7: Reference signal responses. The responses in process output y and control signal u to a unit step in the reference signal r are shown in (a), and the gain curves of G_{yr} and G_{ur} are shown in (b). Results with PI control with error feedback are shown by solid lines, and the dashed lines show results for a controller with a feedforward compensator.

Example 12.5 Third-order system

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 12.7. The solid lines show results for a proportional-integral (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 ∘) and no resonant 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 12.1 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 12.7). 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 frequency ω_{ms}, is thus a measure of the largest amplification of the disturbances. The maximum



(a) Gain curves (b) Nyquist plot

Figure 12.8: Graphical interpretation of the sensitivity function. Gain curves of the loop transfer function and the sensitivity function (a) can be used to calculate 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 sensitivity plot. The Nyquist plot (b) shows the same information in a different form. All points inside the dashed circle have sensitivities greater than 1.

magnitude of $1/(1+L)$ corresponds to the minimum of $|1+L|$, which is precisely the stability margin s_m defined in Section 10.3, so that $M_s = 1/s_m$. 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 the disturbance attenuation. Since the sensitivity function depends only 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 12.8. 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 1 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_{yv} from load disturbance v to process output y for the system in Figure 12.1 is

$$G_{yv} = \frac{P}{1+PC} = PS = \frac{T}{C}. \tag{12.14}$$

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

$$G_{yv} = \frac{T}{C} \approx \frac{1}{C} \approx \frac{s}{k_i}, \tag{12.15}$$

The process transfer function P typically goes to zero for large s as does a well

designed controller C. The loop transfer function PC then also goes to zero and the sensitivity function S goes to 1 for large s. We then have the approximation $G_{yv} \approx P$ for large s.

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 variations 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 the measurement noise to the control signal,

$$-G_{uw} = \frac{C}{1+PC} = CS = \frac{T}{P}. \tag{12.16}$$

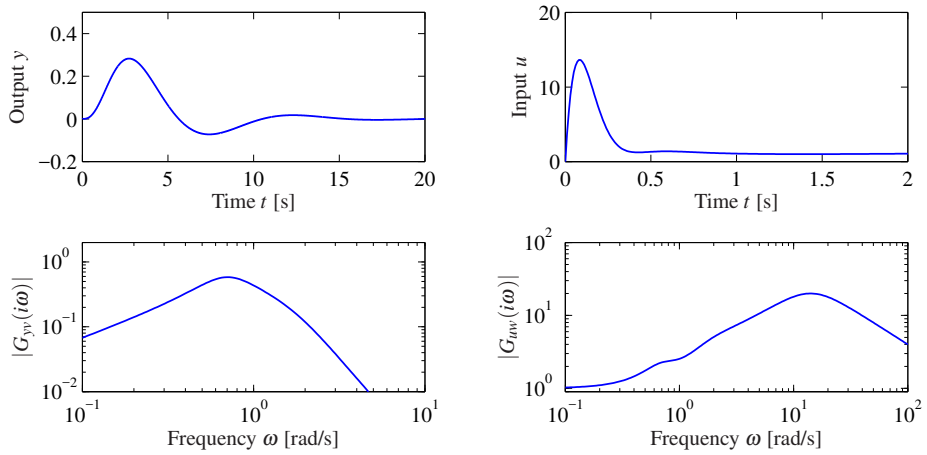
For controllers with integral action, the complementary sensitivity function is close to 1 for low frequencies ($\omega < \omega_{\text{gc}}$), and G_{uw} can be approximated by $-1/P$. Similarly, the sensitivity function is close to 1 for high frequencies ($\omega > \omega_{\text{gc}}$), and G_{uw} can be approximated by $-C$.

Example 12.6 Third-order system
Consider a process with the transfer function $P(s) = (s+1)^{-3}$ and a proportional-integral-derivative (PID) controller with gains $k_p = 0.6$, $k_i = 0.5$ and $k_d = 2.0$. We augment the controller using a second-order noise filter with damping ratio $1/\sqrt{2}$ and $T_f = 0.1$. The controller transfer function then becomes

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

The closed loop system responses are illustrated in Figure 12.9. The closed loop response of the output to a step in the load disturbance in the top part of Figure 12.9a has a peak of 0.28 at time $t = 2.73$ s. The frequency response in Figure 12.9a shows that the gain has a maximum of 0.58 at $\omega = 0.7$ rad/s.

The closed loop response of the control signal to a step in measurement noise is shown in Figure 12.9b. The high-frequency roll-off of the transfer function $G_{uw}(i\omega)$ is due to filtering; without it the gain curve in Figure 12.9b would continue to rise after 20 rad/s. The step response has a peak of 13 at $t = 0.08$ s. The frequency response has its peak 20 at $\omega = 14$ rad/s. Notice that the peak occurs at a frequency far above the peak of the response to load disturbances and far above the gain crossover frequency $\omega_{\text{gc}} = 0.78$ rad/s. An approximation derived in Exercise 12.9 gives $\max |CS(i\omega)| \approx k_d/T_f = 20$, which occurs at $\omega = \sqrt{2}/T_d = 14.1$ rad/s. ∇



(a) Output load response (b) Input noise response

Figure 12.9: Closed loop disturbance responses for Example 12.6. The closed loop time and frequency responses of process output y to load disturbance v are shown in (a) and the responses of control signal u to measurement noise w are shown in (b).

12.4 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*. These 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 $C = L/P$. This approach may lead to controllers of high order and there are limitaions if the process transfer function has poles and zeros in the right half plane as will be discussed in Section 12.6. Another possibility 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 for the loop transfer function that gives good performance and good stability margins. Figure 12.10 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 attenuation

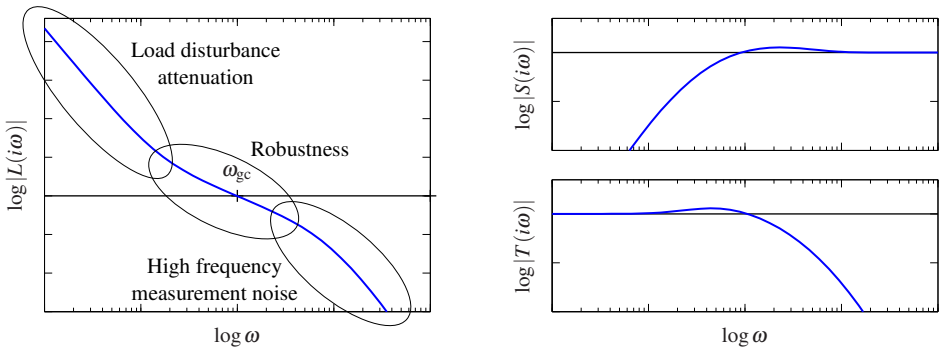


Figure 12.10: Gain curve and sensitivity functions for a typical loop transfer function. The plot on the left shows the gain curve and the plots on the right show the sensitivity function and complementary sensitivity function. The crossover frequency ω_{gc} determines the attenuation of load disturbances, bandwidth and response time of the closed loop system. The slope n_{gc} determines the robustness of the closed loop systems, see (12.17). At low frequencies, a large magnitude of L provides good load disturbance rejection and reference tracking, while at high frequencies a small loop gain avoids injecting too much measurement noise.

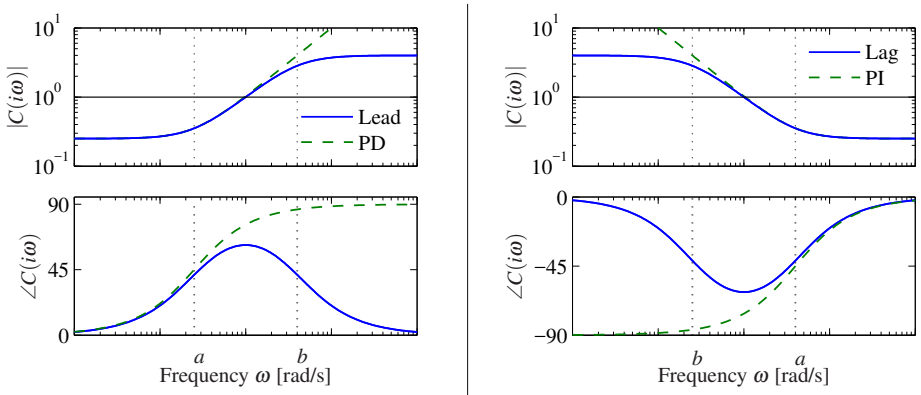
of low-frequency disturbances. Since $S = 1/(1+L)$, it follows that for frequencies where $|L| > 101$ disturbances will be attenuated by a factor of 100 and the tracking error is less than 1%. It is therefore desirable to have a large crossover frequency and a steep (negative) slope of the gain curve. The choice of gain crossover frequency ω_{gc} is a compromise among attenuation of load disturbances, injection of measurement noise and robustness. The controller 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, the controller transfer function should have low gain at high frequencies, a property called *high-frequency roll-off*.

Bode's relations (see Section 10.4) impose restrictions on the shape of the loop transfer function. Equation (10.8) implies that the slope of the gain curve at gain crossover cannot be too steep. If the gain curve has a constant slope, we have the following relation between slope n_{gc} and phase margin φ_m:

$$n_{gc} = -2 + \frac{2\varphi_m}{\pi}. \tag{12.17}$$

This formula is a reasonable approximation when the gain curve does not deviate too much from a straight line. It follows from equation (12.17) 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 the controller gain and adding poles and zeros to the controller transfer function. 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



(a) Lead compensation, $a < b$

(b) Lag compensation, $b < a$

Figure 12.11: Frequency response for lead and lag compensators $C(s) = k(s+a)/(s+b)$. Lead compensation (a) occurs when $a < b$ and provides phase lead between $\omega = a$ and $\omega = b$. Lag compensation (b) 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 compensation. PI/PD frequency responses are shown by dashed curves.

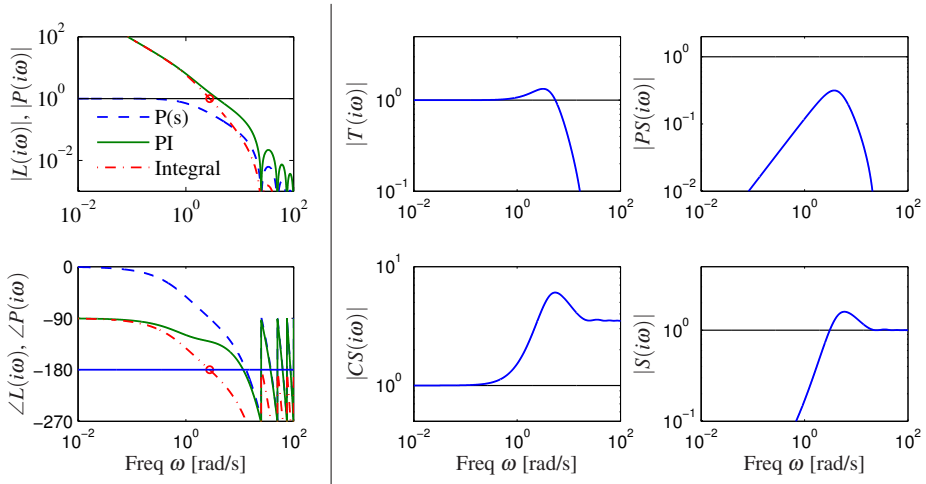
single-input, single-output systems. It can also be applied to systems with one input and many outputs by closing the loops one at a time. The only limitation for minimum phase systems is that large phase leads and high controller gains may be required to obtain closed loop systems with a fast response. Many specific procedures are available: they all require experience, but they also give 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 add simple compensators with the transfer function

$$C(s) = k\frac{s+a}{s+b}, \quad a > 0, b > 0. \tag{12.18}$$

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 a 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 12.11. Lag compensation, which increases the gain at low frequencies, is typically used to improve tracking performance and disturbance attenuation at low frequencies. Lead and lag compensators can also be combined to form a lead-lag compensator (see Exercise 12.11). Compensators that are tailored to specific disturbances can be also designed, as shown in Exercise 12.10. Lead compensation is typically used to improve phase margin. The following examples gives an illustration of the use of lag compensation (via PI control).



(a) Loop shaping (b) Gang of Four

Figure 12.12: Loop-shaping design of a controller for an atomic force microscope in tapping mode. (a) Bode plots of the process (dashed), the loop transfer function for an integral controller with critical gain (dash-dotted) and a PI controller (solid) adjusted to give reasonable robustness. (b) Gain curves for the Gang of Four for the system.

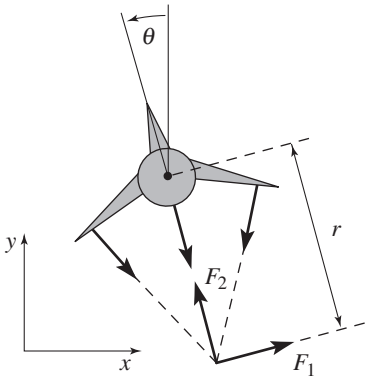
Example 12.7 Atomic force microscope in tapping mode

A simple model of the dynamics of the vertical motion of an atomic force microscope in tapping mode was given in Exercise 10.2. The transfer function for the system dynamics is

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

and the parameters $a = \zeta\omega_0$, $\tau = 2\pi n/\omega_0$ are explained in Example 11.2. The gain has been normalized to 1. A Bode plot of this transfer function for the parameters $a = 1$ and $\tau = 0.25$ is shown using dashed curves in Figure 12.12a. To improve the attenuation of load disturbances we increase the low-frequency gain by introducing an integral controller. The loop transfer function then becomes $L = k_i P(s)/s$, and we start by adjusting the gain k_i so that the closed loop system is marginally stable, giving $k_i = 8.3$. The Bode plot is shown by the dash-dotted line in Figure 12.12a, where the critical point is indicated by $\circ$. Notice the increase of the gain at low frequencies. To obtain a reasonable 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 maximum sensitivity $M_s = 1.6$ and maximum complementary sensitivity $M_t = 1.3$. The loop transfer function is shown in solid lines in Figure 12.12a. Notice the significant increase of the phase margin compared with the purely integral controller (dash-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 12.12b. The peaks of the sensitivity curves are reasonable, and the plot of PS shows that the largest value of PS is 0.3,



(a) Simplified model

Symbol	Description	Value
m	Vehicle mass	4.0 kg
J	Vehicle inertia, φ_3 axis	0.0475 kg m^2
r	Force moment arm	25.0 cm
c	Damping coefficient	0.05 kg m/s
g	Gravitational constant	9.8 m/s^2

(b) Parameter values

Figure 12.13: Roll control of a vectored thrust aircraft. (a) The roll angle θ is controlled by applying maneuvering thrusters, resulting in a moment generated by F_1. (b) The table lists the parameter values for a laboratory version of the system.

which implies that the load disturbances are well attenuated. The plot of CS shows that the largest noise gain $|C(i\omega)S(i\omega)|$ is 6. The controller has a gain $k_p = 3.5$ at high frequencies, and hence we may consider adding high-frequency roll-off to make CS smaller at high frequencies. ∇

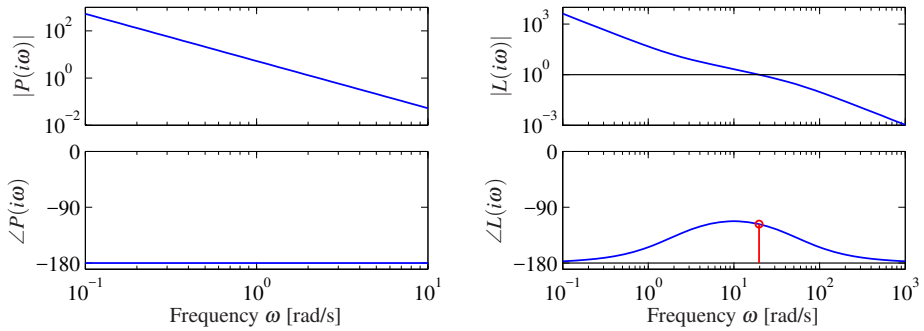
A common problem in the design of feedback systems is that the phase margin is too small, and phase *lead* must then be added to the system. If we set $a < b$ in equation (12.18), we add phase lead in the frequency range between the pole/zero pair (and extending approximately $10\times$ in frequency in each direction). By appropriately choosing the location of this phase lead, we 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. In Exercise 12.15 it is shown that the gain increases exponentially with the amount of phase lead. 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 12.8 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 12.13. Following Exercise 9.11, we model the system with a second-order transfer function of the form

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

with the parameters given in Figure 12.13b. 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/s.



(a) Process dynamics (b) Lead compensator

Figure 12.14: Control design for a vectored thrust aircraft using lead compensation. The Bode plot for the open loop process P is shown in (a) and for the loop transfer function $L = PC$ using a lead compensator C in (b). Note the phase lead in the crossover region near $\omega = 100$ rad/s.

The open loop transfer function from F_1 to θ is shown in Figure 12.14a. To achieve our performance specification, we would like to have a gain of at least 10 at a frequency of 10 rad/s, 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 (12.18) 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 12.14b. We see that this system has a gain of greater than 10 at all frequencies up to 10 rad/s and that it has more than 60° of phase margin. ∇

The action of a lead compensator is essentially the same as that of the derivative portion of a PID controller. As described in Section 11.5, 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 (12.18) is a first-order compensator and can provide up to 90° of phase lead. Larger phase lead can be obtained by using a higher-order lead compensator (Exercise 12.15):

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

12.5 The Root-Locus Method

In design methods such as eigenvalue assignment, discussed in Sections 7.2 and 8.3, we designed controllers that give desired closed loop poles. The controllers were

sufficiently complex so that all closed loop poles could be specified. The complexity of the controller is thus related to the complexity of the process. In practice we may have to use a simple controller for a complex process and it is then not possible to find a controller that gives all closed poles their desired values. It is interesting to explore what can be done with a controller having restricted complexity as was the case for loop shaping in Section 12.4. The simplest case with only one selectable controller parameter can be investigated with the *root locus method*. The *root locus* is a graph of the roots of the characteristic equation as a function of the parameter. The method gives insight into the effects of the controller parameter. It is straightforward to obtain the root locus by finding the roots of of the closed-loop characteristic polynomial for different values of the parameter. There are also good computer tools for generating the root locus. Of greater interest is the fact that the general shape of the root locus can be obtained with very little effort, and often gives considerable insight.

Proportional Control

To illustrate the root locus method we consider a process with the transfer function

$$P(s) = \frac{b(s)}{a(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots b_m}{s^n + a_1 s^{n-1} + \cdots a_n} = b_0 \frac{(s-z_1)(s-z_2)\ldots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)}.$$

The polynomial $a(s)$ has degree n and the polynomial $b(s)$ has degree m. We assume that the integer $n_{pe} = n - m$, which is called the *pole excess*, is positive or zero. The controller is assumed to be a proportional controller with the transfer function $C(s) = k$. We will explore the poles of the closed loop system when the gain k of the proportional controller ranges from 0 to ∞.

The closed loop characteristic polynomial is

$$a_{cl}(s) = a(s) + kb(s) \tag{12.19}$$

and the closed loop poles are the roots of $a_{cl}(s)$. The root locus is a graph of the roots of $a_{cl}(s)$ as the gain k is varies from 0 to ∞. Since the polynomial $a_{cl}(s)$ has degree n the plot will have n branches.

The branches start from the open loop poles. When the gain is k zero we have $a_{cl}(s) = a(s)$ and the closed loop poles are equal to the open loop poles. When there are open loop poles with multiplicity n^* the characteristic equation can be written as

$$(s-p_l)^{n^*}\tilde{a}(s) + kb(s) \approx (s-p_1)^{n^*}\tilde{a}(p_l) + kb(p_l) = 0.$$

For small values of k the roots are $s = p_l + \sqrt[n^*]{-kb(p_l)/\tilde{a}(p_l)}$. The root locus has a star pattern with n^* branches emanating from the open loop pole $s = p_l$. The angle between two neighboring brances is $2\pi/n^*$.

To explore what happens for large gain we approximate the characteristic equa-

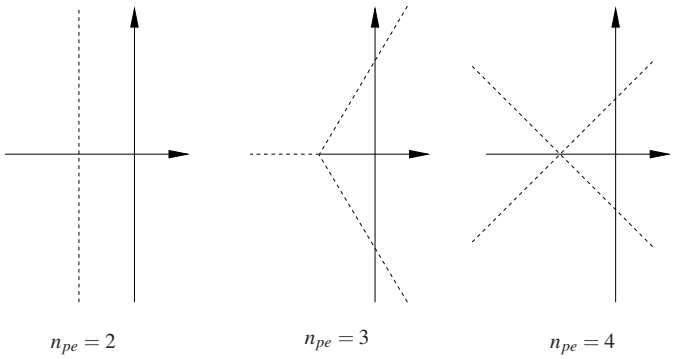


Figure 12.15: Asymptotes of of root locus for systems with pole excess $n_{pe} = 2$, 3 and 4. There are n_{pe} asymptotes radiate from the point given by equation (12.21), and the angles between the asymptotes are $2\pi/n_{pe}$.

tion (12.19) for large s and k, hence

$$a_{cl}(s) = b(s)\Big(\frac{a(s)}{b(s)} + k\Big) \approx b(s)\Big(\frac{s^{n_{pe}}}{b_0} + k\Big). \tag{12.20}$$

For large k the closed loop poles are approximately the roots of $b(s)$ and $\sqrt[n_{pe}]{-b_0 k}$. A better approximation of equation (12.20) is

$$s = s_0 + \sqrt[n_{pe}]{-kb_0}, \qquad s_0 = \frac{1}{n_{pe}}\left(\sum_{k=1}^{n} p_k - \sum_{k=1}^{m} z_k\right) \tag{12.21}$$

(Exercise 12.12). The asymptotes are thus n_{pe} lines that radiate from $s = s_0$, the center of mass of poles and zeros. When $b_0 k > 0$ the lines have the angles $(\pi + 2l\pi)/n_{pe}, l = 1, \cdots, n_{pe}$ with respect to the real line. Figure 12.15 shows the asymptotes of the root locus for large gain for different values of the pole excess n_{pe}.

Summarizing, we find that the root locus thus has n branches that start at the open loop poles and end either at the open loop zeros or at infinity. The branches that ends at infinity have the star-patterned asymptotes given by (12.21). An immediate consequence is that open loop systems with right half plane zeros or a pole excess larger than 2 will always be unstable for sufficiently large gains.

There are simple rules for sketching root loci. We describe here a few of them. The root locus has locally the symmetric star pattern at points where there are multiple roots; the number of branches depend on the multiplicity of the roots. For systems with $kb_0 > 0$ the root locus has segments on the real line where there are odd numbers of real poles and zeros to the right of the segment (Exercise 12.13). It is also straightforward to find direction where a branch of the rool locus leaves a pole, see Exercise 12.14.

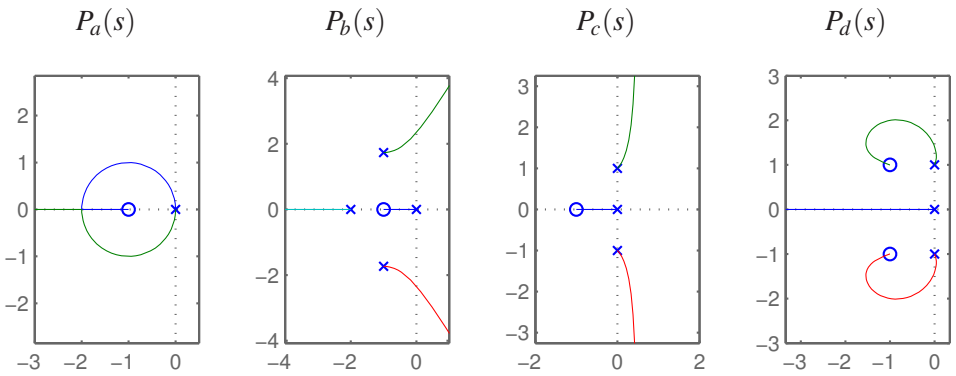


Figure 12.16: Examples of root loci for processes with the transfer functions $P_a(s)$, $P_b(s)$, $P_c(s)$ and $P_d(s)$ given by equation (12.22).

Figure 12.16 show root loci for systems with $k > 0$ and the transfer functions

$$P_a(s) = k\frac{s+1}{s^2}, \qquad P_b(s) = k\frac{s+1}{s(s+2)(s^2+2s+4)},$$
$$P_c(s) = k\frac{s+2}{s(s^2+1)}, \qquad P_d(s) = k\frac{s^2+2s+2}{s(s^2+1)}. \tag{12.22}$$

The locus of $P_a(s)$ in Figure 12.16a starts with two roots at the origin and the pattern locally has the star configuration with $n^* = 2$. As the gain increases the locus bends because of the attraction of the zero. In this particular case the locus is actually a circle around the zero $s = -1$. Two roots meet at the real axis, and there is the typical star pattern. One root goes towards the zero and the other one goes to infinity along the negative real axis as the gain k increases. The root locus has the segment $(-\infty, -1]$ on the real axis. The locus in Figure 12.16b starts at the open loop poles $s = -2, 0$ and $-1 \pm i\sqrt{3}$. The pole excess is $n_{pe} = 3$ and the asymptotes which originate from $s_0 = -1$, have the corresponding pattern. The locus in Figure 12.16c has vertical asymptotes since $n_{pe} = 2$ (see Figure 12.15). The asymptotes originate from $s_0 = 0.5$. The root locus has the segment $[-1, 0)$ on the real line. The locus in Figure 12.16d has three branches, one is the segment $(-\infty, 0]$ on the real line the other segment originates on the complex open loop poles and ends at the open loop zeros.

The root locus can also be used for design. Consider for example the system in Figure 12.16c, which can represent PI control of a system with the transfer function

$$P(s) = \frac{1}{s^2+1}, \qquad C(s) = k\frac{s+2}{s}.$$

The root locus in Figure 12.16c shows that the system is unstable for all values of the controller gain and we can immediately conclude that the process cannot be stabilized with a PI controller. To obtain a stable closed loop system we can attempt to choose a PID controller with zeros to the left of the undamped poles,

for example

$$C(s) = k\,\frac{s^2+2s+2}{s}.$$

The root locus obtained with this controller is shown in Figure 12.16d.

We have illustrated the root locus with a closed loop system with a proportional controller where the parameter is the gain. The root locus can also be used to find the effects of other parameters, as was illustrated in Example 5.16.

12.6 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 because of difficult dynamics; additional limitations related to 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

$$P(s) = P_{mp}(s)P_{ap}(s), \tag{12.23}$$

where P_{mp} is the minimum phase part and P_{ap} is the nonminimum phase part, we require that P_{mp} has all its zeros in the open left half plane. 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. We have for example

$$P(s) = \frac{s-2}{(s+1)(s-1)} = \frac{s+2}{(s+1)^2}\,\frac{(s-2)(s+1)}{(s+2)(s-1)}. \tag{12.24}$$

The transfer function P_{ap} does not influence the gain curve in the Bode plot but it does influence the phase curve. Requiring that the phase margin be φ_m, we get

$$\arg L(i\omega_{\rm gc}) = \arg P_{ap}(i\omega_{\rm gc}) + \arg P_{mp}(i\omega_{\rm gc}) + \arg C(i\omega_{\rm gc}) \geq -\pi + \varphi_m, \tag{12.25}$$

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

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

If the slope $n_{\rm gc}$ is negative, it has to be larger than -2 for the closed loop system

to be stable. It follows from Bode's relations, equation (10.8), that

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

Combining this with equation (12.25) gives the following inequality for the allowable phase lag of the all-pass part at the gain crossover frequency:

$$-\arg P_{ap}(i\omega_{\text{gc}}) \leq \pi - \varphi_m + n_{\text{gc}} \frac{\pi}{2} =: \varphi_l. \tag{12.26}$$

In addition we must require that the encirclement condition of the Nyquist theorem is satisfied.

The condition (12.26), which we call the *gain crossover frequency inequality*, shows that the gain crossover frequency must be chosen so that the phase lag of the nonminimum phase component is not too large. For systems with high robustness requirements we may choose a phase margin of 60° ($\varphi_m = \pi/3$). It follows from the approximate version of Bode's relations (10.8) that the slope n_{gc} cannot be steeper than $n_{\text{gc}} = -1$. To have a reasonable flexibility in choosing the gain crossover frequency we choose $n_{\text{gc}} = -1$, which gives an admissible phase lag $\varphi_l = \pi/6 = 0.52$ rad (30°) of the allpass component. For systems where we can accept a lower robustness we may choose a phase margin of 45° ($\varphi_m = \pi/4$) and the slope $n_{\text{gc}} = -1/2$, which gives an admissible phase lag $\varphi_l = \pi/2 = 1.57$ rad (90°).

The crossover frequency inequality (12.26), shows that nonminimum phase components impose severe restrictions on possible crossover frequencies. It also means that there are systems that cannot be controlled with sufficient stability margins. We illustrate the limitations in a number of commonly encountered situations.

Example 12.9 Zero in the right half-plane
The nonminimum 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 $z > 0$. Notice that we have $z-s$ in the numerator instead of $s-z$ to satisfy the condition that P_{ap} should have negative phase. The phase lag of the nonminimum phase part is

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

Since the phase lag of P_{ap} increases with frequency, the inequality (12.26) gives the following bound on the crossover frequency:

$$\omega_{\text{gc}} < z\tan(\varphi_l/2). \tag{12.27}$$

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

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

$$e^{-s\tau} \approx \frac{1-0.5s\tau}{1+0.5s\tau} = \frac{2/\tau - s}{2/\tau + s}.$$

A long time delay is thus approximately equivalent to a slow right half-plane zero $z = 2/\tau$.

Example 12.10 Pole in the right half-plane
The nonminimum 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 sign of P_{ap} is dictated by the condition that it should have negative phase. The phase lag of the nonminimum phase part is

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

and the crossover frequency inequality becomes

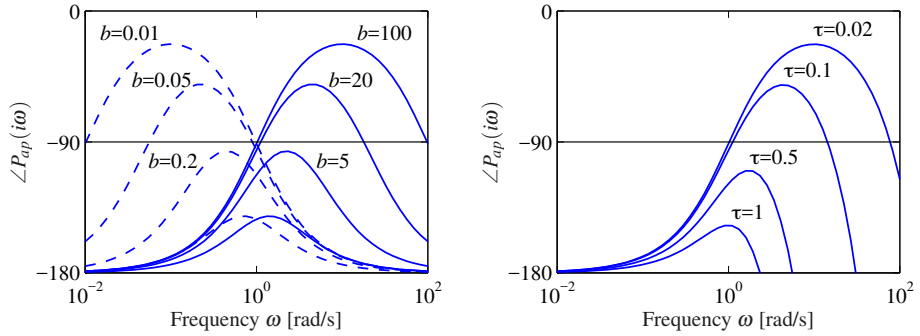
$$\omega_{gc} > \frac{p}{\tan(\varphi_l/2)}. \tag{12.28}$$

Right half-plane poles thus require that the closed loop system have a sufficiently high gain crossover frequency, and a consequence is that the actuators must be fast. With $\varphi_l = \pi/3$ we get $\omega_{gc} > 1.7p$. Fast right half-plane poles (p large) therefore require a larger gain crossover frequency than slower right half-plane poles. The control of unstable systems imposes minimum bandwidth requirements for process actuators and sensors. ∇

We will now consider systems with a right half-plane zero z and a right half-plane pole p. If $p = z$, there will be cancellation of an unstable system mode and the system cannot be stabilized. A cancellation of an unstable pole means that the system has an unstable mode that is not reachable and observable; see Section 8.3. We can therefore expect that the system is difficult to control if the right half-plane pole and zero are close. A straightforward way to use the crossover frequency inequality is to plot the phase of the nonminimum phase factor P_{ap} of the process transfer function. Such a plot, which can be incorporated in an ordinary Bode plot, will immediately show the permissible gain crossover frequencies. An illustration is given in Figure 12.17, which shows the phase of the transfer functions for systems with a right half-plane pole/zero pair and systems with a right half-plane pole and a time delay. The transfer functions of the systems are

$$P_{ap}(s) = \frac{(bs-1)(s+1)}{(bs+1)(s-1)}, \qquad P_{ap}(s) = \frac{s+1}{s-1}e^{-\tau s} \tag{12.29}$$

To illustrate the limitations we will introduce numerical values. If we require that the phase lag φ_l of the nonminimum phase factor be less than 90°, we must require



(a) RHP pole/zero pair (b) RHP pole and time delay

Figure 12.17: Example limitations due to the gain crossover frequency inequality. The figures show the phase lag of the all-pass factor P_{ap} as a function of frequency for the systems (12.29). Since the phase lag of P_{ap} at the gain crossover frequency cannot be too large, it is necessary to choose the gain crossover frequency properly. All systems have a right half-plane pole at $s = 1$. The system in (a) has zeros at $s = 2$, 5, 20 and 100 (solid lines) and at $s = 0.5$, 0.2, 0.05 and 0.01 (dashed lines). The system in (b) has time delays $\tau = 0.02$ 0.1, 0.5 and 1.

that the ratio z/p be larger than 6 or smaller than 1/6 for systems with right half-plane poles and zeros and that the product $p\tau$ be less than 0.3 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 12.16). Also notice that possible values of the gain crossover frequency ω_{gc} are quite restricted.

Using the theory of functions of complex variables, it can be shown that for systems with a right half-plane pole p and a right half-plane zero z (or a time delay τ), any stabilizing controller gives sensitivity functions with the property

$$\sup_{\omega} |S(i\omega)| \geq \frac{p+z}{|p-z|}, \qquad \sup_{\omega} |T(i\omega)| \geq e^{p\tau}. \tag{12.30}$$

This result is proven in Exercise 12.17.

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 how the 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 the sensors and actuators or by adding sensors and actuators.

Example 12.11 Balance system
As an example of a system with both right half-plane poles and zeros, consider the balance system with zero damping shown in Example 3.1, whose dynamics are given by

$$H_{\theta F} = \frac{ml}{-(M_t J_t - m^2 l^2)s^2 + mglM_t},$$
$$H_{pF} = \frac{-J_t s^2 + mgl}{s^2\left(-(M_t J_t - m^2 l^2)s^2 + mglM_t\right)}.$$

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 $\{0, 0, \pm\sqrt{mglM_t/(M_t J_t - m^2 l^2)}\}$ and zeros $\{\pm\sqrt{mgl/J_t}\}$. Using the parameters in Example 7.7, the right half-plane pole is at $p = 2.68$ and the zero is at $z = 2.09$. Equation (12.30) then gives $|S(i\omega)| \geq 8$, which shows that it is not possible to control the system robustly.

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

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

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. The gain crossover frequency inequality is then based just on the right half-plane pole (Example 12.10). If our admissible phase lag for the nonminimum phase part is $\varphi_l = 45^\circ$, then our gain crossover must satisfy

$$\omega_{gc} > \frac{p}{\tan(\varphi_l/2)} = 6.48 \text{ rad/s}.$$

If the actuators have sufficiently high bandwidth, e.g., a factor of 10 above ω_{gc} or roughly 10 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 the disturbance attenuation and also relates the tracking error e to the reference signal r: 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 12.12 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 arbitrarily small for any finite frequency by making k sufficiently large. ∇

The system in Example 12.12 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 *passive*, their transfer functions are *positive real* and their physics is associated with energy dissipation. For typical control systems there are severe constraints on the sensitivity function. The following theorem, due to Bode, provides insights into the limits of performance under feedback.

Theorem 12.2 (Bode's integral formula)**.** *Let $S(s)$ be the sensitivity function of an internally stable system with loop transfer function $L(s)$. Assume that the loop transfer function $L(s)$ is such that $sL(s)$ goes to zero as $s \to \infty$, then the sensitivity function satisfies the following integral:*

$$\int_0^\infty \log|S(i\omega)|\,d\omega = \int_0^\infty \log\frac{1}{|1+L(i\omega)|}\,d\omega = \pi\sum p_k. \tag{12.31}$$

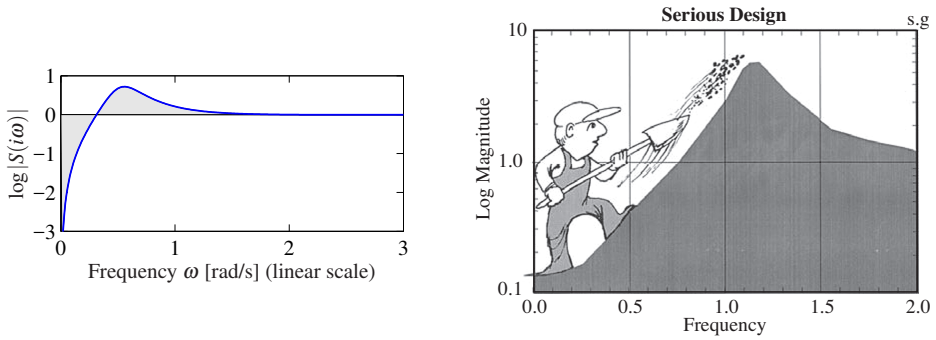
where the sum is over the right half plane poles p_k of $L(s)$.

Equation (12.31) 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 (12.31) 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 12.18, which shows $\log|S(i\omega)|$ as a function of ω. The area over the horizontal



(a) Bode integral formula (b) Control design process

Figure 12.18: 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 (12.31), 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 [Ste03]).

axis must be equal to the area under the axis when the 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. Notice that the system in Example 12.12 violates the condition that $\lim_{s\to\infty} sL(s) = 0$ and hence the integral formula does not apply.

There is a result analogous to equation (12.31) for the complementary sensitivity function:

$$\int_0^\infty \frac{\log|T(i\omega)|}{\omega^2}\,d\omega = \pi\sum\frac{1}{z_i}, \tag{12.32}$$

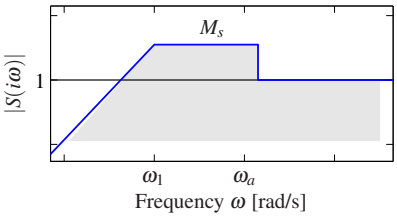
where the summation is over all right half-plane zeros. Notice that slow right half-plane zeros are worse than fast ones and that fast right half-plane poles are worse than slow ones.

Example 12.13 X-29 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 12.19a), 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" [Ste03], 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 inverted pendulum dynamics (Exercise 9.3) 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

(a) X-29 aircraft



(b) Sensitivity analysis

Figure 12.19: 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 (a). The desired sensitivity for the closed loop system is shown in (b). 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.

4.3, we may expect that it may be difficult to achieve the specifications.

To evaluate the achievable performance, we search for a control law such that the sensitivity function is small up to the desired bandwidth and not greater than M_s beyond that frequency. Because of the Bode integral formula, we know that M_s must be greater than 1 at high frequencies to balance the small sensitivity at low frequency. We thus ask if we can find a controller that has the shape shown in Figure 12.19b with the smallest value of M_s. 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 (12.31) must hold. We will assume that the sensitivity function is given by

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

corresponding to Figure 12.19b. If we further assume that $|L(s)| \le \delta/\omega^2$ for frequencies larger than the actuator bandwidth, Bode's integral becomes

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

Evaluation of the integral gives $-\omega_1 + \omega_a \log M_s = \pi p$ or

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

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 find that

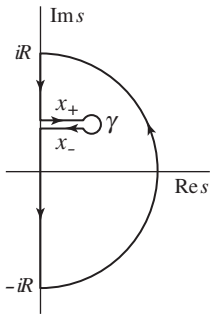


Figure 12.20: Contour used to prove Bode's theorem. For each right half-plane pole we create a path from the imaginary axis that encircles the pole as shown. To avoid clutter we have shown only one of the paths that enclose one right half-plane.

$M_s = 1.75$, 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 12.18, the maximum achievable phase margin for this system is approximately 35°, which is below the usual design limit of 45° in aerospace systems. The zero at $s = 26$ limits the maximum gain crossover that can be achieved. ∇

Derivation of Bode's Formula

We now derive Bode's integral formula (Theorem 12.2). This is a technical section that 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 12.20. The contour encloses the right half-plane except for 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 counterclockwise.

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 analytic inside the contour. We have

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

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 12.20. 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. Since $S(s) \approx k/(s-p_k)$ close to the pole, the argument of $S(s)$ decreases by 2π as the contour encircles the pole. On the contours X_+ and X_- we therefore have

$$|S_{X_+}| = |S_{X_-}|, \qquad \arg S_{X_-} = \arg S_{X_+} - 2\pi.$$

Hence

$$\log(S_{X_+}) - \log(S_{X_-}) = 2\pi i,$$

and we get

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

Repeating the argument for all poles p_k in the right half plane, letting the small circles go to zero gives

$$I_1 + I_2 + I_3 = -2i\int_0^\infty \log|S(i\omega)|d\omega + i\sum_k 2\pi\ \mathrm{Re}\, p_k = 0.$$

Since complex poles appear as complex conjugate pairs, $\sum_k \mathrm{Re}\, p_k = \sum_k p_k$, which gives Bode's formula (12.31).

12.7 Design Example

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

Example 12.14 Lateral control of a vectored thrust aircraft
The problem of controlling the motion of a vertical takeoff and landing (VTOL) aircraft was introduced in Example 3.11 and in Example 12.8, where we designed a controller for the roll dynamics. We now wish to control the position of the

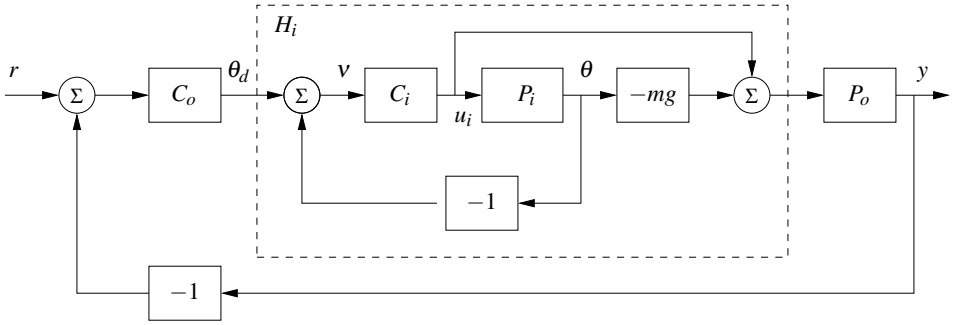


Figure 12.21: Inner/outer control design for a vectored thrust aircraft. The inner loop H_i controls the roll angle of the aircraft using the vectored thrust. The outer loop controller C_o commands the roll angle to regulate the lateral position. The process dynamics are decomposed into inner loop (P_i) and outer loop (P_o) dynamics, which combine to form the full dynamics for the aircraft.

.

aircraft, a problem that requires stabilization of the attitude. The system thus has two control loops.

To control the lateral dynamics of the vectored thrust aircraft, we make use of a "inner/outer" loop design methodology, as illustrated in Figure 12.21. 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 lateral position dynamics and controller. This decomposition follows the block diagram representation of the dynamics given in Exercise 9.11.

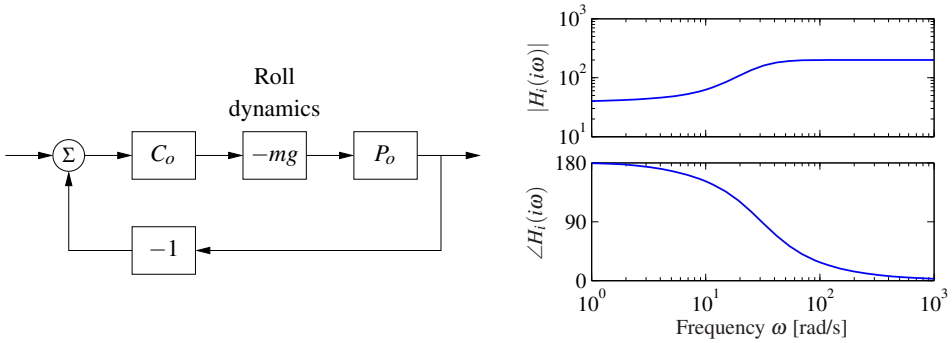
The approach that we take is to design a controller C_i for the inner loop so that the resulting closed loop system H_i assures that the roll angle θ follows its reference θ_r fast and accurately. We then design a controller for the lateral position y that uses the approximation that we can directly control the roll angle as an input θ 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

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

We choose the desired bandwidth to be 10 rad/s (10 times that of the outer loop) and the low-frequency error to be no more than 5%. This specification is satisfied using the lead compensator of Example 12.8 designed previously, so we choose

$$C_i(s) = k\frac{s+a}{s+b}, \qquad a = 2, \quad b = 50, \quad k = 1.$$



(a) Outer loop approximation (b) Actual roll dynamics

Figure 12.22: Outer loop control design for a vectored thrust aircraft. (a) The outer loop approximates the roll dynamics as a state gain $-mg$. (b) The Bode plot for the roll dynamics, indicating that this approximation is accurate up to approximately 10 rad/s.

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 12.22, and we see that $H_i \approx -mg = -39.2$ is a good approximation up to 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 9.11, the outer loop dynamics can be written as

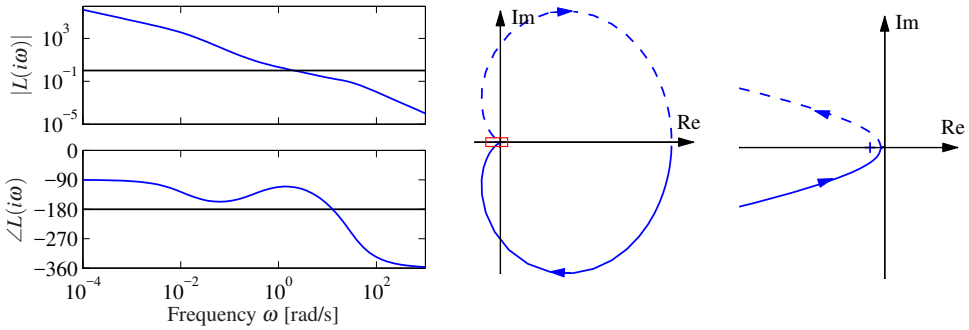
$$P(s) = H_i(0)P_o(s) = \frac{H_i(0)}{ms^2+cs},$$

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 y for a step input and has a bandwidth of 1 rad/s. The outer loop process dynamics are given by a double 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 high frequency 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 $b_o/10$. We desire phase lead at crossover, and we desire the crossover at $\omega_{gc} = 1$ rad/s, so this gives $b_o = 10$. To ensure 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



(a) Bode plot (b) Nyquist plot

Figure 12.23: Inner/outer loop controller for a vectored thrust aircraft. Bode plot (a) and Nyquist plot (b) for the loop transfer function cut at θ_d, for the complete system. The system has a phase margin of 68° and a gain margin of 6.2.

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 1. A simple calculation shows that $k_o = 2$ satisfies this objective. Thus, the final outer loop controller becomes

$$C_o(s) = -2\frac{s+0.3}{s+10}.$$

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 12.21 with inner and outer loop controllers are shown in Figure 12.23, and we see that the specifications are satisfied. In addition, we show the Gang of Four in Figure 12.24, and we see that the transfer functions between all inputs and outputs are reasonable. The sensitivity to load disturbances PS is large at low frequency because the controller does not have integral action.

The approach of splitting the dynamics into an inner and an 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 y to the input u_1. The use of the additional measurement of θ greatly simplifies the design because it can be broken up into simpler pieces. ∇

12.8 Further Reading

Design by loop shaping was a key element in the early development of control, and systematic design methods were developed; see James, Nichols and Phillips [JNP47], Chestnut and Mayer [CM51], Truxal [Tru55] and Thaler [Tha89]. Loop shaping is also treated in standard textbooks such as Franklin, Powell and Emami-Naeini [FPEN05], Dorf and Bishop [DB04], Kuo and Golnaraghi [KG02] and Ogata [Oga01]. Systems with two degrees of freedom were developed by Horowitz [Hor63],

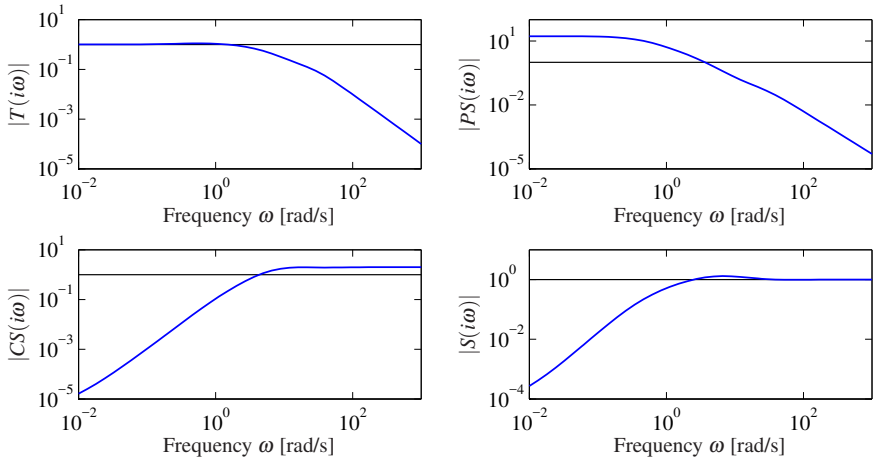


Figure 12.24: Gang of Four for vectored thrust aircraft system.

who also discussed the limitations of poles and zeros in the right half-plane. Fundamental results on limitations are given in Bode [Bod45]; more recent presentations are found in Goodwin, Graebe and Salgado [GGS01]. The treatment in Section 12.6 is based on [Åst00]. 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 that resulted in H_∞ design methods. A compact presentation is given in the texts by Doyle, Francis and Tannenbaum [DFT92] and Zhou, Doyle and Glover [ZDG96]. Loop shaping was integrated with the robust control theory in McFarlane and Glover [MG90] and Vinnicombe [Vin01]. Comprehensive treatments of control system design are given in Maciejowski [Mac89] and Goodwin, Graebe and Salgado [GGS01].

Exercises

12.1 Consider the system in Figure 12.1. Give all signal pairs that are related by the transfer functions $1/(1+PC)$, $P/(1+PC)$, $C/(1+PC)$ and $PC/(1+PC)$.

12.2 Consider the system in Example 12.1, where the process and controller transfer fucctions are given by

$$P(s) = 1/(s-a), \qquad C(s) = k(s-a)/s.$$

Choose the parameters $a = -1$ and compute the time (step) and frequency responses for all the transfer functions in the Gang of Four for controllers with $k = 0.2$ and $k = 5$.

12.3 (Equivalence of Figures 12.1 and 12.2) Consider the system in Figure 12.1 and let the outputs of interest be $z = (\eta, \nu)$ and the major disturbances be $w = (n, d)$. Show that the system can be represented by Figure 12.2 and give the matrix

transfer functions $\mathscr{P}$ and $\mathscr{C}$. Verify that the elements of the closed loop transfer function H_{zw} are the Gang of Four.

12.4 Consider the spring–mass system given by (3.17), which has the transfer function

$$P(s) = \frac{1}{ms^2 + cs + k}.$$

Design a feedforward compensator that gives a response with critical damping ($\zeta = 1$).

12.5 (Sensitivity of feedback and feedforward) Consider the system in Figure 12.1 and let G_{yr} be the transfer function relating the measured signal y to the reference r. Show that the sensitivities of G_{yr} with respect to the feedforward and feedback transfer functions F and C are given by $dG_{yr}/dF = CP/(1+PC)$ and $dG_{yr}/dC = FP/(1+PC)^2 = G_{yr}S/C$.

12.6 (Equivalence of controllers with two degrees of freedom) Show that the systems in Figures 12.1 and 12.3 give the same responses to command signals if $F_m C + F_u = CF$.

12.7 (Disturbance attenuation) Consider the feedback system shown in Figure 12.1. 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.

12.8 (Disturbance reduction through feedback) Consider a problem in which an output variable has been measured to estimate the potential for disturbance attenuation by feedback. Suppose an analysis shows that it is possible to design a closed loop system with the sensitivity function

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

Estimate the possible disturbance reduction when the measured disturbance is

$$y(t) = 5\sin(0.1t) + 3\sin(0.17t) + 0.5\cos(0.9t) + 0.1t.$$

12.9 Show that the effect of high frequency measurement noise on the control signal for the system in Example 12.6 can be approximated by

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

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

12.10 (Attenuation of low-frequency sinusoidal disturbances) Integral action eliminates constant disturbances and reduces low-frequency disturbances because the controller gain is infinite at zero frequency. A similar idea can be used to reduce the effects of sinusoidal disturbances of known frequency ω_0 by using the controller

$$C(s) = k_p + \frac{k_s s}{s^2 + 2\zeta\omega_0 s + \omega_0^2}.$$

This controller has the gain $C_s(i\omega_0) = k_p + k_s/(2\zeta)$ for the frequency ω_0, which can be large by choosing a small value of ζ. Assume that the process has the transfer function $P(s) = 1/s$. Determine the Bode plot of the loop transfer function and simulate the system. Compare the results with PI control.

12.11 (Lead-lag compensation) Lead and lag compensators can be combined into a lead-lag compensator which has the transfer function

$$C(s) = k\frac{(s+a_1)(s+a_2)}{(s+b_1)(s+b_2)}$$

Show that the controller reduces to a PID controller with special choice of parameters and give the relations between the parameters.

12.12 (Asymptotes of root locus) Consider proportional control of a system with the transfer function

$$P(s) = \frac{b(s)}{a(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots b_m}{s^n + a_1 s^{n-1} + \cdots a_n} = b_0 \frac{(s-z_1)(s-z_2)\ldots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)}.$$

Show that the root locus has asymptotes that are straight line that emerge from the point

$$s_0 = \frac{1}{n_e}\Big(\sum_{k=1}^{n} p_k - \sum_{k=1}^{m} z_k\Big),$$

where $n_e = n - m$ is the pole excess of the transfer function.

12.13 (Real line segments of root locus) Consider proportional control a process with a rational transfer function. Assume that $b_o k > 0$, show that the root locus has segments on the real line there are an odd number of real poles and zeros to the right the segment.

12.14 (Initial direction of root locus) Consider proportional control of a system with the transfer function

$$P(s) = \frac{b(s)}{a(s)} = \frac{b_0 s^m + b_1 s^{m-1} + \cdots b_m}{s^n + a_1 s^{n-1} + \cdots a_n} = b_0 \frac{(s-z_1)(s-z_2)\ldots(s-z_m)}{(s-p_1)(s-p_2)\cdots(s-p_n)}.$$

Let p_j be an isolated pole and assume that $kb_p > 0$. Show that the root locus starting at p_j has the initial direction.

$$\angle(s-p_j) = \pi + \Sigma_{k=1}^{m}\angle(p_j - s_k) - \Sigma_{k\neq j}\angle(p_j - p_k).$$

Give a geometric interpretation of the result.

12.15 Consider a lead compensator with the transfer function

$$C_n(s) = \Big(\frac{s\sqrt[n]{k} + a}{s+a}\Big)^n,$$

which has zero frequency gain $C(0) = 1$ and high-frequency gain $C(\infty) = k$. Show that the gain required to give a given phase lead φ is

$$k = \Big(1 + 2\tan^2(\varphi/n) + 2\tan(\varphi/n)\sqrt{1+\tan^2(\varphi/n)}\Big)^n,$$

and that $\lim_{n\to\infty} k = e^{2\varphi}$.

12.16 Consider a process with the loop transfer function

$$L(s) = k\frac{z-s}{s-p},$$

with positive z and p. Show that the system is stable if $p/z < k < 1$ or $1 < k < p/z$, and that the largest stability margin is $s_m = |p-z|/(p+z)$ is obtained for $k = 2p/(p+z)$. Determine the pole/zero ratios that gives the stability margin $s_m = 2/3$.

12.17 Prove the inequalities given by equation (12.30). (Hint: Use the maximum modulus theorem.)

12.18 (Phase margin formulas) Show that the relationship between the phase margin and the values of the sensitivity functions at gain crossover is given by

$$|S(i\omega_{gc})| = |T(i\omega_{gc})| = \frac{1}{2\sin(\varphi_m/2)}.$$

12.19 (Stabilization of an inverted pendulum with visual feedback) Consider stabilization of an inverted pendulum based on visual feedback using a video camera with a 50-Hz frame rate. Let the effective pendulum length be l. Assume that we want the loop transfer function to have a slope of $n_{gc} = -1/2$ at the crossover frequency. Use the gain crossover frequency inequality to determine the minimum length of the pendulum that can be stabilized if we desire a phase margin of 45°.

12.20 (Rear-steered bicycle) Consider the simple model of a bicycle in Equation (4.5), which has one pole in the right half-plane. The model is also valid for a bicycle with rear wheel steering, but the sign of the velocity is then reversed and the system also has a zero in the right half-plane. Use the results of Exercise 12.16 to give a condition on the physical parameters that admits a controller with the stability margin s_m.

12.21 Prove the formula (12.32) for the complementary sensitivity.