CHAPTER 8

THE DESIGN OF CONTROL SYSTEMS

8.1 INTRODUCTION

In this chapter techniques that can be used to design feedback control systems are examined. In the previous chapters techniques were developed to analyze the behavior of control systems. It is now time to apply the insight gained by the analysis of these systems to the design of the systems. A major tool for design is a combination of the Nyquist and Bode plots.

Let's examine what we have learned that may be useful in design. In Chap. 3 it was established that many of the desirable properties of a closed-loop controller come from creating a large return difference transfer function. More specifically, a large loop gain produces good disturbance rejection and good sensitivity reduction. The Bode plots of the loop gain that were developed in Chap. 5 graphically display the size of the loop gain as a function of frequency. In Chap. 6 it was shown how to establish stability of the closed-loop system from the Nyquist plot of the loop gain transfer function.

The transient response of the closed-loop system to inputs and disturbances can be estimated by combining the results of Chaps. 4, 5, and 6. The closed-loop frequency response is attainable from the M-circles of the Nyquist plot of the loop gain as shown in Sec. 6.5. The closed-loop transfer function can be estimated from

the frequency plot using the identification techniques of Sec. 5.5. Finally the transient response is estimated from the transfer function using the results of Chap. 4. This chain of analysis is not actually performed in detail for each design; rather, insights from general analyses create guidelines for design. The root locus techniques of Chap. 7 may be used to gain a different view of how transient response can be altered. Robustness results from Chap. 6 indicate that the loop gain must be allowed to become small at frequencies where modeling uncertainty becomes large.

These ideas are put together in Sec. 8.2 where an overall strategy for shaping the loop gain as a function of frequency is developed. In that section it is also shown how the loop gain is manipulated by adding elements to a series compensator. In Sec. 8.3 the process of introducing the commonly used building blocks of controller design is begun. The design of proportional compensators is examined in Sec. 8.3. The design of proportional-integral (PI) or, equivalently, lag compensators is examined in Sec. 8.4. The design of proportional-integral-derivative (PID) or, equivalently, lead-lag compensators is examined in Sec. 8.5. Design by canceling plant dynamics with notch filters is examined in Sec. 8.6. A controller design is demonstrated by examples of realistic controller designs of stable minimum phase plants in Sec. 8.7 and 8.8. In the design of Sec. 8.8 the question of active versus passive control of oscillatory dynamics is addressed. Techniques for stabilizing unstable open-loop plants are developed in Sec. 8.9. In Sec. 8.10 it is shown that plants containing right half-plane zeros contain natural limitations to achievable controller performance. In Sec. 8.11 it is shown that the mechanized approach of a pole placement technique can show how to attack difficult control problems but that well-placed closed-loop poles do not necessarily guarantee acceptable performance. Finally, in Sec. 8.12 the potential benefits of extra measurements in overcoming the limitations imposed by right half-plane zeros are displayed. First a state-variable approach is shown and then a technique for dealing with missing measurements is displayed. Detailed analysis of the state-variable design would lead into the analysis of multi-input, multi-output systems. This is left to another book and another course.[1]

8.1.1 CAD Notes

The CAD tools that are useful in design have been introduced during the previous chapters of the book. The student should use these tools wherever they help in understanding the design process. However, it is our suggestion that the student think through as much of the problem as possible and use the CAD system as a computational aid to eliminate dreary calculations. Almost always, the time spent to solve a problem can be minimized by thinking through the problem and gaining insight rather than making many disoriented computer iterations.

[1]The book by Doyle, Francis, and Tennenbaum is recommended as continuing the approach of this book. The book by Maciejowski is encyclopedic in delineating the varied approaches to the multi-input, multi-output systems. See the references listed at the end of this book.

8.2 GENERAL PRINCIPLES FOR DESIGNING SERIES COMPENSATORS USING FREQUENCY RESPONSE TECHNIQUES

In this chapter, we consider series compensators of the form given in Fig. 8.2-1 where $G_p(s)$ is a model of the plant and $G_C(s)$ is the compensator transfer function that is to be designed. The series compensator is easily translated into the now familiar G configuration by observing that

$$G(s) = G_C(s)G_p(s) \tag{8.2-1}$$

We have already spent a good deal of effort learning how to analyze control systems in the G configuration. In particular we have learned that, for these systems, much information is available concerning performance, stability, and robustness of the closed-loop system from the Bode and Nyquist plots of the loop gain $G(s)$. The series compensator is particularly convenient because we can start with the Bode plot of the plant model $G_P(s)$ and modify that plot as we add new pole and zero factors to $G(s)$ by including them in the compensator. As we learned in Chap. 5, new factors in $G(s)$ affect the Bode diagram of $G(s)$ by simply summing the values on the Bode diagram of the original $G(s)$ with the values on the Bode diagram of the new factors. In this way, we can build the Bode diagram of $G(s)$ into a desired shape. This process is sometimes referred to as the *loop shaping* method of design. We now demonstrate this process with an example. This example shows the thought process that goes into a controller design. You will learn the elements of this process in this chapter. For this example simply watch how a controller design evolves and notice the utility of the Bode plot and the ease with which this plot is modified by modifying the compensator.

Example 8.2-1. A plant model is given by

$$G_p(s) = \frac{1}{(s+1)(s+0.1)}$$

To obtain higher gain at low frequencies and a Type 1 system for total dc disturbance rejection and zero steady state tracking error, a design that includes an integrator in the loop is desirable. However, it is also desirable to negate the 90° phase lag of the integrator at higher frequencies. The designer sets the series compensator to be

$$G_{C1}(s) = \frac{s+0.1}{s}$$

It is seen in Sec. 8.4 that such a compensator is a standard building block called a *lag compensator* or a *PI compensator*. Here it is not important where $G_{C1}(s)$ comes from; it is only important to see how the addition of transfer functions in series affects the Bode plot of the evolving loop gain.

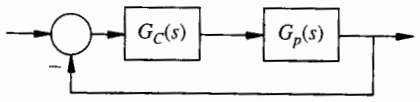


FIGURE 8.2-1
A series compensator.

The loop gain $G_1(s)$ is the product of the plant and the compensator.

$$G_1(s) = \frac{1}{s(s+1)}$$

Fig. 8.2-2 shows the Bode magnitude and phase plots of $G_P(s)$, $G_{C1}(s)$ and $G_1(s)$. Notice how the plots associated with $G_P(s)$, and $G_{C1}(s)$ simply add. (The logarithms or decibels in the magnitude plot add—not the magnitudes themselves).

If the designer decides to add to the overall gain of the loop gain transfer function, it can be done by multiplying the compensator transfer function by a constant, say 10.

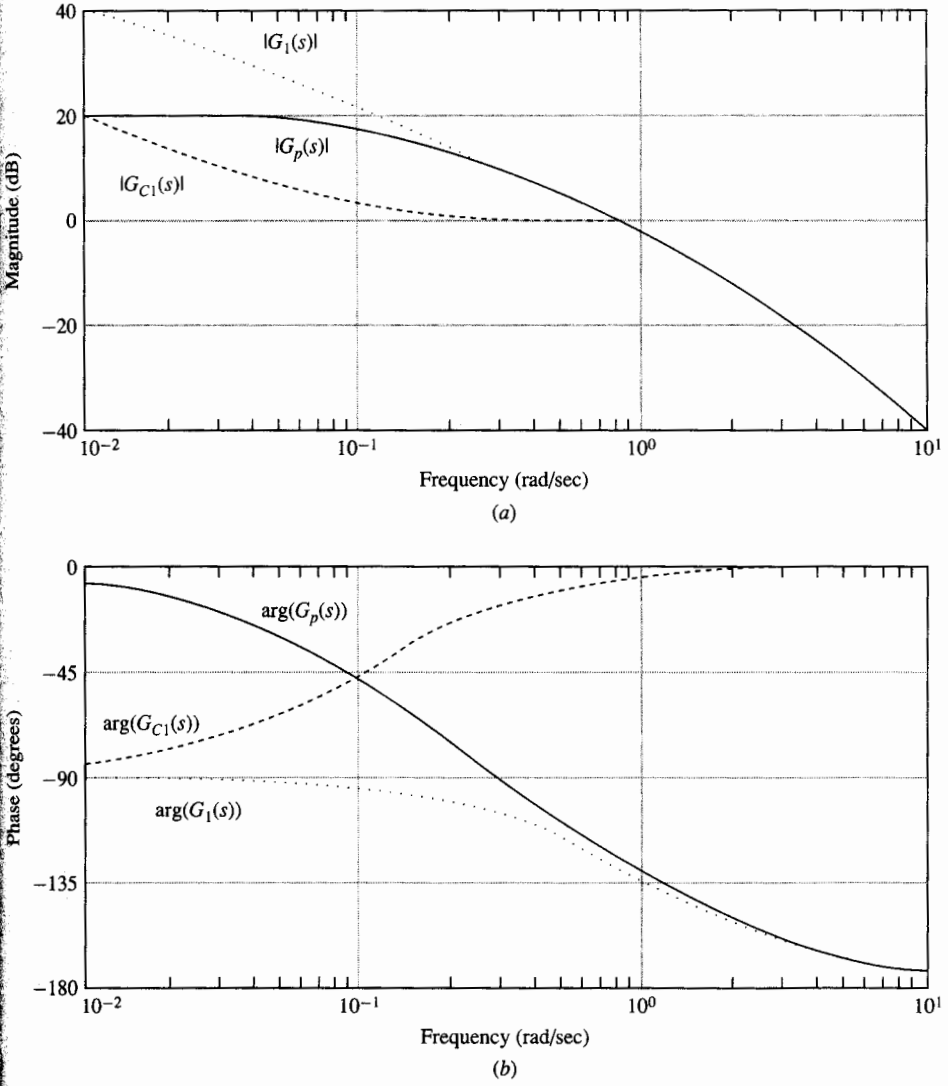


FIGURE 8.2-2
Bode plot for Example 8.2-1. (*a*) Magnitude; (*b*) phase.

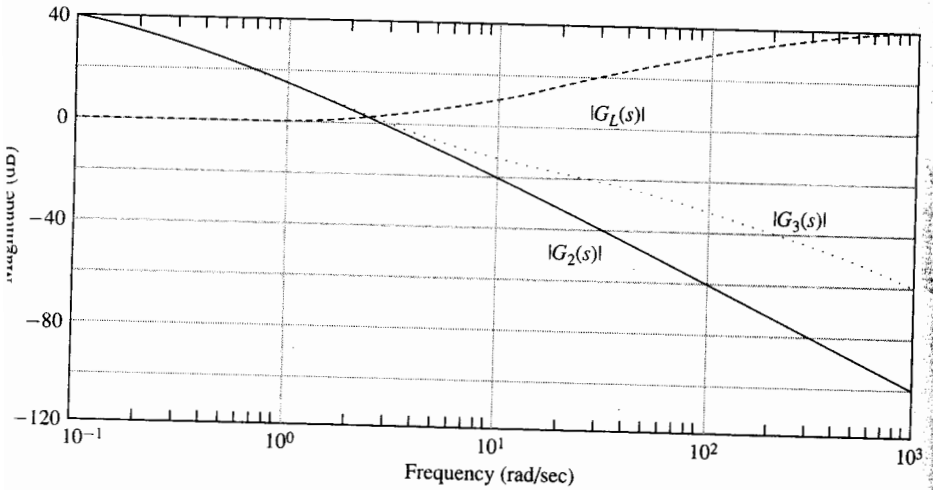
The series compensator $G_{C2}(s)$ is given by

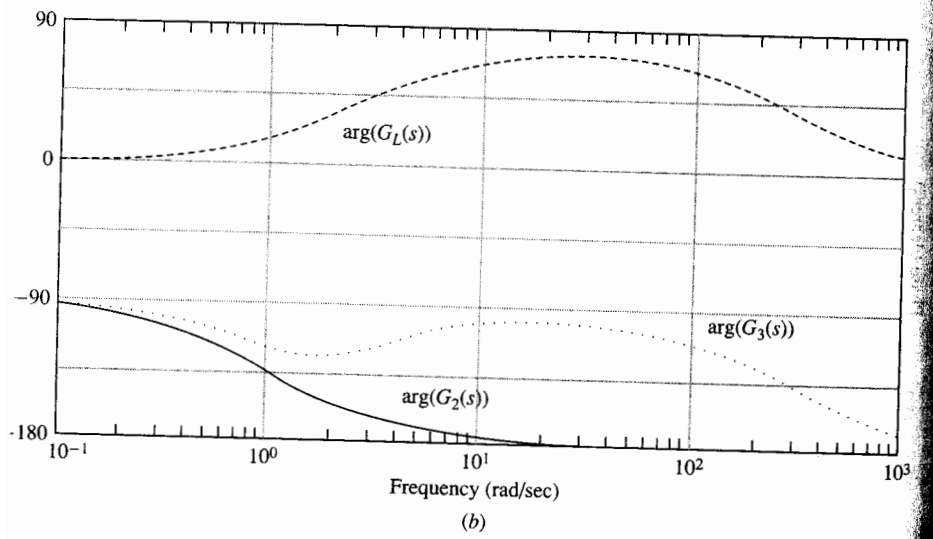
$$G_{C2}(s) = 10G_{C1}(s) = \frac{10(s+0.1)}{s}$$

The loop gain becomes

$$G_2(s) = G_{C2}(s)G_p(s) = 10G_1(s)$$

The Bode magnitude plot is shifted upward by a factor of 10, or equivalently, $20 \log 10 = 20$ dB, as shown in Fig. 8.2-3. The phase plot is unaltered. While the resulting design





URE 8.2-3
: plots for the revised design in Example 8.2-1. (*a*) Magnitude; (*b*) phase.

has adequate loop gain at low frequency, a look at the Nyquist plot of this loop gain indicates serious problems. If $G_{C2}(s)$ were used as a controller the resulting system would be highly oscillatory and would have so little robustness that unless the plant model were extremely precise the closed-loop system would likely be unstable. (Draw the Nyquist plot and verify these statements.)

The designer remedies this situation as you will learn to do in this chapter. Another pole and zero are added to the compensator in what we will come to call a lead compensator with transfer function

$$G_L(s) = \frac{100(s+3)}{s+300}$$

The new compensator $G_{C3}(s)$ becomes

$$G_{C3}(s) = G_L(s)G_{C2}(s) = \frac{1000(s+3)}{s(s+300)}$$

The new loop gain $G_3(s)$ becomes

$$G_3(s) = G_L(s)G_2(s)$$

Figure 8.2-3 displays the Bode magnitude and phase plots of $G_2(s)$, $G_L(s)$ and $G_3(s)$. Notice how the plots of $G_2(s)$ are changed into the plots of $G_3(s)$ by adding the plots of $G_2(s)$ and $G_L(s)$ point by point.

The Nyquist plot of $G_2(s)$ and $G_3(s)$ are given in Fig. 8.2-4. Notice how the compensator adds positive phase margin (phase lead) to the Nyquist plot around the critical point $s = -1$ to avoid approaching the critical point too closely. An analysis of the resulting loop gain $G_3(s)$ shows an adequate controller design. The designer has created an adequate design by modifying the loop gain one step at a time, adding new features with each new transfer function placed in series with the compensator of the previous step.

In previous chapters we learned how to analyze control systems. In Sec. 6.7 we saw that it is possible to define a function $p(j\omega)$ such that many aspects of good

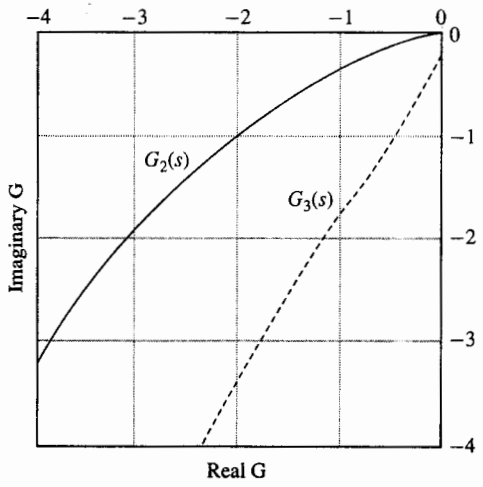


FIGURE 8.2-4
Nyquist plots for Example 8.2-1.

performance will be satisfied if Eq. (6.7-1) is satisfied. Equation (6.7-1) is repeated here as Eq. (8.2-2).

$$\left|1 + \tilde{G}(j\omega)\right| > p(j\omega) \tag{8.2-2}$$

The set of transfer functions represented by $\tilde{G}(j\omega)$ is the set of loop gains in the G configuration when there are perturbations present. Let a set of plant models be characterized by Eqs. (6.6-5) and (6.6-6), repeated here as Eqs. (8.2-3) and (8.2-4).

$$\tilde{G}_p(s) = G_p(s)(1 + L_m(s)) \tag{8.2-3}$$

$$|L_m(j\omega)| < l_m(j\omega) \tag{8.2-4}$$

Using a series compensator the set of loop gains is given by

$$\tilde{G}(s) = G_C(s)\tilde{G}_p(s) \tag{8.2-5}$$

The nominal loop gain is given by

$$G(s) = G_C(s)G_p(s) \tag{8.2-6}$$

and using Eqs. (8.2-3) through (8.2-6), we find that

$$\tilde{G}(s) = G(s)(1 + L_m(s)) \tag{8.2-7}$$

with Eq. (8.2-4) still satisfied. The controller is robustly stable and has the performance qualities specified by Eq. (8.2-2) if Eq. (6.7-18) is satisfied for all ω. Equation (6.7-18*b*) is repeated here as Eq. (8.2-8).

$$\left|\frac{1}{1 + G(j\omega)}\right| p(j\omega) + \left|\frac{G(j\omega)}{1 + G(j\omega)}\right| l_m(j\omega) < 1 \tag{8.2-8}$$

As discussed in Section 6.7, satisfying Eq. (8.2-8) requires a tradeoff between the magnitude of the nominal sensitivity function

$$S(j\omega) = (1 + G(j\omega))^{-1} \tag{8.2-9}$$

and the magnitude of the nominal complementary sensitivity function

$$T(j\omega) = G(j\omega)(1 + G(j\omega))^{-1} = 1 - S(j\omega) \tag{8.2-10}$$

The manner in which the tradeoff is made is determined by the functions $l_m(j\omega)$ and $p(j\omega)$. This tradeoff can be accomplished by manipulating the loop gain transfer function $G(j\omega)$ by adding poles and zeros to a series compensator as demonstrated in Example 8.2-1. We now develop guidelines for manipulating the loop gain transfer function so that Eq. (8.2-8) can be satisfied and an acceptable controller can result.

The first thing to notice from Eq. (8.2-8) is that if for some particular frequency ω_0, $l_m(j\omega_0) > 1$ and $p(j\omega_0) > 1$, then Eq. (8.2-8) cannot be satisfied. If $l_m(j\omega_0) > 1$

and $p(j\omega_0) > 1$, then

$$\left|\frac{G(j\omega_0)}{1 + G(j\omega_0)}\right| l_m(j\omega_0) + \left|\frac{1}{1 + G(j\omega_0)}\right| p(j\omega_0)$$
$$> \left|\frac{G(j\omega_0)}{1 + G(j\omega_0)}\right| + \left|\frac{1}{1 + G(j\omega_0)}\right|$$
$$\geq \left|\frac{G(j\omega_0)}{1 + G(j\omega_0)} + \frac{1}{1 + G(j\omega_0)}\right| = 1$$

This fact can be interpreted as showing that aggressive performance objectives cannot be met for frequencies where the system is not well modeled. In general $l_m(j\omega)$, which represents the modeling error, is small for low frequencies and grows large at high frequencies. To accommodate this, $p(j\omega)$, which represents performance requirements, is made acceptably low at high frequencies where the model is uncertain. A large performance constraint is required and attainable for low frequency inputs where the model is well known. A typical plot of $l_m^{-1}(j\omega)$ and $p(j\omega)$ is shown in Fig. 8.2-5. The frequency axis splits into three regions:

low frequencies, where $p(j\omega)$ is large and $l_m(j\omega)$ is small ($l_m^{-1}(j\omega)$ is large)
high frequencies, where $p(j\omega)$ is small and $l_m(j\omega)$ is large ($l_m^{-1}(j\omega)$ is small)
transition frequencies, where $p(j\omega)$ and $l_m(j\omega)$ are both fairly close to one

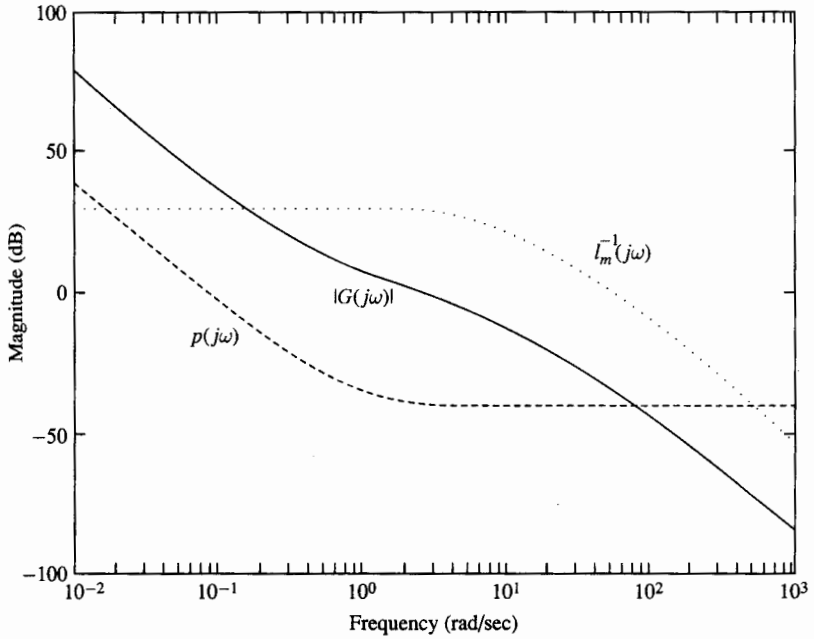


FIGURE 8.2-5
A typical acceptable design.

The requirements on $G(j\omega)$ so that Eq. (8.2-8) is satisfied can now be developed fairly easily for the high and low frequency sections while the requirements on $G(j\omega)$ for the transition frequency section is more difficult to determine.

At low frequencies $p(j\omega)$ is much larger than one. In this case Eq. (8.2-8) dictates that the magnitude of the sensitivity function $S(j\omega)$ must be much less than one. This situation can be accomplished if the magnitude of the loop gain $G(j\omega)$ is made much larger than one. When $|G(j\omega)|$ is much larger than one we can make approximations in Eq. (8.2-8). Specifically, if

$$|G(j\omega)| \gg 1 \tag{8.2-11a}$$

then

$$\left|\frac{1}{1+G(j\omega)}\right| \approx \frac{1}{|G(j\omega)|} \tag{8.2-11b}$$

and

$$\left|\frac{G(j\omega)}{1+G(j\omega)}\right| \approx 1 \tag{8.2-11c}$$

If the approximations are used in Eq. (8.2-8) we obtain the approximate requirement

$$\frac{1}{|G(j\omega)|}p(j\omega) < 1 - l_m(j\omega) \tag{8.2-11d}$$

Thus, the requirements of Eq. (8.2-8) are very close to being satisfied at low frequencies (where $p(j\omega) \gg 1$ and $l_m(j\omega) \ll 1$) if $|G(j\omega)|$ is made large enough, i.e., if the following inequality, obtained from Eq. (8.2-11d) is satisfied.

$$|G(j\omega)| > \frac{p(j\omega)}{1 - l_m(j\omega)} \tag{8.2-12}$$

Inequality (8.2-12) shows that at low frequencies, large loop gains are required to meet performance requirements. The nominal loop gain must be made slightly larger than the performance requirement dictates so that the performance is achieved in the face of the small amount of modeling error expected at low frequencies.

At high frequencies, $l_m(j\omega)$ is much larger than one. In this case, Eq. (8.2-8) dictates that the magnitude of the complementary sensitivity function be much less than one. However, for the magnitude of the complementary sensitivity function to be much less than one, the loop gain must be much less than one.

$$\left|\frac{G(j\omega)}{1+G(j\omega)}\right| \ll 1 \text{ implies } |G(j\omega)| \ll 1 \tag{8.2-13}$$

When

$$|G(j\omega)| \ll 1 \tag{8.2-14a}$$

then

$$\left|\frac{1}{1+G(j\omega)}\right| \approx 1 \tag{8.2-14b}$$

and

$$\left|\frac{G(j\omega)}{1+G(j\omega)}\right| \approx |G(j\omega)| \tag{8.2-14c}$$

Using these approximations in Eq. (8.2-8) leads to approximate requirement

$$|G(j\omega)|l_m(j\omega) + p(j\omega) < 1 \tag{8.2-14d}$$

The requirements of Eq. (8.2-8) are very nearly satisfied at high frequencies (where $p(j\omega) \ll 1$ and $l_m(j\omega) \gg 1$) if $|G(j\omega)|$ is made small enough, i.e., if the following inequality obtained from Eq. (8.2-14d) is satisfied.

$$|G(j\omega)| < \frac{1 - p(j\omega)}{l_m(j\omega)} \tag{8.2-15}$$

Comparing inequalities (8.2-15) and (6.6-11) shows that the main concern at high frequencies is to keep the loop gain small enough so that its magnitude cannot exceed one and cause instability in the face of large modeling errors and unknown phase characteristics. Thus $|G(j\omega)|$ should be less than $l_m^{-1}(j\omega)$ as required by Eq. (6.6-11) when $|G(j\omega)|$ is small. An additional reduction is dictated by the high frequency performance requirement. If, at some frequency, the perturbed loop gain is allowed to approach the -1 point on the Nyquist plot too closely, the system will demonstrate the unacceptable performance attributes of large oscillations, disturbance amplification, and high sensitivity at that frequency. Thus, we must allow for a safety margin and keep the magnitude of the *perturbed* loop gain below a threshold that is somewhat less than one. The nominal loop gain should be kept somewhat less than $l_m^{-1}(j\omega)$ to provide not only for robust stability but also for robust performance at high frequencies.

The strategy for manipulating the magnitude of the loop gain of a control design is now taking shape. Fig. 8.2-5 introduced previously shows a plot of $p(j\omega)$, $l_m^{-1}(j\omega)$ and $|G(j\omega)|$ for a possibly acceptable controller design. The loop gain is made larger than $p(j\omega)$ at low frequencies and smaller than $l_m^{-1}(j\omega)$ for high frequencies. There is a safety margin above $p(j\omega)$ at low frequencies to account for possible modeling errors and for the fact that approximations are used in attaining Eq. (8.2-12). There is a safety margin below $l_m^{-1}(j\omega)$ at high frequencies to ensure reasonable performance and to account for the approximations used in developing Eq. (8.2-15). Notice that the loop gain need not be above $p(j\omega)$ at *high* frequencies where $p(j\omega)$ is small or below $l_m^{-1}(j\omega)$ at *low* frequencies where $l_m^{-1}(j\omega)$ is small.

What qualities should $G(j\omega)$ have through the transition frequencies? This question is more difficult since the assumptions leading to the approximations Eq. (8.2-11) and Eq. (8.2-14) are not valid. We know that the $|G(j\omega)|$ must change from being large at low frequencies to being small at high frequencies. Somewhere in the transition band $|G(j\omega)|$ must pass through the point ω_c where

$$|G(j\omega_c)| = 1 \tag{8.2-16}$$

This frequency ω_c is defined in Sec. 6.5 as the crossover frequency of the control system. In Sec. 6.5, the phase margin ϕ_m of the control system is defined as the angle between the polar representation of $G(j\omega_c)$ and the negative real axis.

$$\phi_m = 180° + \arg\,(G\,(j\omega_c)) \tag{8.2-17}$$

Much of what we wish to learn about how we should make $G(j\omega)$ behave in the transition frequencies can be seen by studying the control system around ω_c.

It is desirable to create a loop gain with as large a phase margin as possible. A small phase margin means that the Nyquist plot of the loop gain passes very close to the -1 point, and indicates two problems. The first problem is that the closed-loop system responds to inputs and disturbances with poorly damped oscillations at the same frequency as the crossover frequency if the phase margin is small. This fact is shown in Sec. 6.5 where we find how to determine the closed-loop behavior from the Nyquist plot of the loop gain. The second problem is that if ϕ_m is small, a small modeling error in the plant at the crossover frequency causes the Nyquist plot of the loop gain to pass extremely close to the -1 point or actually encircle the -1 point, causing instability. Thus, there is a nominal performance problem and a robustness problem associated with a small phase margin.

The combination of a nominal performance and robustness problem can be seen by examining Eq. (8.2-8) at ω_c. Since

$$|G(j\omega_c)| = 1 \tag{8.2-18a}$$

then

$$\left|\frac{G\,(j\omega_c)}{1+G\,(j\omega_c)}\right| = \frac{1}{|1+G\,(j\omega_c)|} \tag{8.2-18b}$$

so that Eq. (8.2-8) becomes

$$\left|\frac{1}{1+G\,(j\omega_c)}\right| \quad (l_m\,(j\omega_c) + p\,(j\omega_c)) < 1 \tag{8.2-18c}$$

At ω_c the magnitude of the sensitivity function that relates to performance and the magnitude of complementary sensitivity function that relates to robustness are both made small by making the magnitude of the return difference large. In fact Eq. (8.2-18c) and also Eq. (8.2-8) will both be satisfied at the crossover frequency if the magnitude of the return difference is made larger than the sum of $l_m(j\omega_c)$ and $p(j\omega_c)$.

$$|1+G\,(j\omega_c)| > l_m\,(j\omega_c) + p\,(j\omega_c) \tag{8.2-19}$$

In Sec. 6.5 it is shown that, at the crossover frequency, the magnitude of the return difference can be expressed as a function of the phase margin.

$$|1+G\,(j\omega_c)| = 2\left|\sin\left(\frac{\phi_m}{2}\right)\right| \tag{8.2-20}$$

The largest the return difference can be at the crossover frequency is 2. This maximum occurs if the phase margin is 180°, i.e., if $\arg(G(j\omega_c)) = 0$. Any phase lag in the loop gain is associated with a decrease in the return difference and an associated drop in robust performance measure.

As is seen in the remainder of this chapter, attaining a large phase margin is difficult. In practice, a phase margin of 60° is usually considered to be large. From Eq. (8.2-20) a 60° phase margin corresponds to a return difference equal to one at the crossover frequency. At frequencies higher than the crossover frequency, the return difference is usually slightly smaller than one. As ω gets large the loop gain approaches zero so that the return difference approaches one.

In general, $l_m(j\omega)$ grows as ω grows. Equation (8.2-19) dictates that, in general, the crossover frequency be chosen below the frequency where $l_m(j\omega)$ first becomes larger than one. Since at the crossover frequency $|G(j\omega_c)| = 1$ Eq. (8.2-19) shows that the crossover frequency is absolutely required to be at a frequency where $l_m(j\omega) < 2$. The crossover frequency ω_c marks the boundary between frequencies where the loop gain is large, producing enhanced performance, and frequencies where the loop gain is small, producing less desirable performance. Thus, the crossover frequency is often used to describe the bandwidth of a control system. The bandwidth of a control system is limited by the frequencies where an accurate model of the plant can be attained. If perfect modeling were attainable so that $l_m(j\omega)$ were always zero then any bandwidth could be attained. Such ideal situations, however, exist only in textbooks.

We have found that unmodeled dynamics dictate that the complementary sensitivity function and thus that the loop gain be rolled off to small values at high frequencies. The desire to enhance performance requires that the return difference and thus the loop gain be large at low frequencies. These two desires are fairly easily accomplished. The skill of the control designer comes in making appropriate tradeoffs as the loop gain goes from large values to small values through the transition frequencies.

There are two conflicting desires in the transition region. We would like the cutoff between low frequencies and high frequencies to be as sharp as possible so that the low frequency region where performance is enhanced becomes as large as possible given the high frequency constraints imposed by modeling uncertainties. The second desire is to keep the amount of phase lag near crossover frequency as small as possible so that the loop gain avoids getting too close to the -1 point or creating extra encirclements.

The fact that these desires conflict can be seen by recalling from Chap. 5 that the phase plot and the slope of the magnitude plot of a transfer function are related. Assume that the loop gain transfer function of the system is minimum phase. (The situation is worse for nonminimum phase systems). Recall the discussion surrounding the Bode phase relationship of Eq. (5.4-1). If a system's magnitude plot has a slope of -40dB per decade for a decade on either side of the crossover frequency, then the phase shift at the crossover frequency is very close to $-180°$. Larger negative slopes in the magnitude plot correspond to greater phase lags. To achieve a reasonably large phase margin, the magnitude plot must slope gradually through the crossover frequency. If a phase margin of 60° is required the magnitude plot must have only

a −20dB per decade slope for about two-thirds of a decade on either side of the crossover frequency. Although we would like to transition quickly from high loop gain at low frequencies to low loop gain at high frequencies, the need to maintain a reasonable degree of robust stability and performance throughout the transition region dictates that the transition occur more slowly.

We have now seen the overall objective of shaping the frequency response of the loop gain transfer function to achieve an acceptable level of controller performance and robustness. If a feasible performance specification $p(j\omega)$ and modeling uncertainty function $l_m(j\omega)$ are specified, an acceptable controller is one in which the sensitivity function and complementary sensitivity function satisfy Eq. (8.2-8).

It should be noted that it is somewhat artificial to summarize various performance specifications in the function $p(j\omega)$. The lessons of loop shaping introduced here do not actually depend upon the ability to produce such a $p(j\omega)$. The role of $p(j\omega)$ is mainly to express mathematically the generally true concept that the performance of a control system is made better if the return difference is made larger.

In the next section we examine the building blocks that have been used in classical control theory for years to develop series compensators to shape the loop gain of a control system.

Exercises 8.2

8.2-1. Turn the following performance specifications into a guideline $p(j\omega)$ so that if

$$|1 + \tilde{G}(j\omega)| > p(j\omega) \qquad \text{for all } \omega$$

then the performance specifications are likely to be met.

1. If there is an output disturbance with a frequency below $\omega = 0.01$ rad/sec, its effect on the output must be less than 0.001 times its original magnitude.
2. The phase margin should be greater than 30°.
3. The closed-loop response to a unit step input should have zero steady-state error. It should have less than 25 percent overshoot in the transient and it should settle to within 2 percent of its final value in 10 secs.

8.2-2. Sketch the Bode magnitude straight line approximation plots for the following set of loop gain transfer functions on the same plot. Do the same for the phase plot. Notice how adding additional terms in series changes the plots.

$$G_1(s) = \frac{1}{s+1}$$

$$G_2(s) = \frac{10}{(s+1)}$$

$$G_3(s) = \frac{10}{s(s+1)}$$

$$G_4(s) = \frac{10(s+3)}{(s+30)}\left(\frac{10}{s(s+1)}\right)$$

8.3 SERIES COMPENSATOR BUILDING BLOCKS: PROPORTIONAL CONTROL

Having determined some goals for shaping the loop gain transfer function in the previous section, we use this section to begin an investigation of the elements of a series compensator that can be used to achieve the desired shape.

The simplest element for series compensation is a constant gain. On the Bode plot a positive gain leaves the phase plot unchanged. A gain greater than one raises the magnitude plot while a gain less than one lowers the magnitude plot. Raising the magnitude plot helps to achieve the large loop gain desired at low frequencies but, at the same time, creates the undesirable effect of raising the loop gain at high frequencies. Since the Bode magnitude plots of plants generally slope down to the right, raising the gain also has the effect of increasing the crossover frequency of the loop gain. In frequency response design methods, a first setting of the gain may be used to set the bandwidth of the system near the limit allowed by modeling uncertainty. A standard root locus analysis shows the effect of a change of gain on the closed-loop pole positions; the closed-loop poles tend to move away from the origin as the gain is increased, manifesting the increasing bandwidth of the system.

Generally, as the gain of a control system is increased, the system's response becomes fast, then oscillatory and finally unstable. Let's examine the effect of increasing the gain on a typical plant model. We restrict ourselves to a series compensator with a gain that is constant across all frequencies. The simple constant gain compensator produces a control signal that is proportional to the error signal between the desired reference input and the plant output. Such a control scheme is often referred to as a *proportional compensator*. Let the plant model $G_p(s)$ be

$$G_p(s) = \frac{10}{(s+1)(s+10)^2} \tag{8.3-1}$$

Our purpose here is to explore the effects of basic control building blocks on nominal plant models; consequently, we consider any perturbations to the model only in a qualitative, not in a quantitative sense.

Plants of the general form shown in Eq. (8.3-1) are very common; they arise in many applications such as process control and control of mechanical systems. An automobile speed control system is a plant that can be modeled with Eq. (8.3-1). Most of us have experienced trying to control such a system manually where the human operator functions as the controller. The operator's actions can be modeled as a proportional compensator as the driver compares the measured speed on the speedometer with the desired speed and changes the throttle position accordingly. If the difference in the two speeds is large, the driver makes a large change in the throttle position thus producing a proportional control law. The transfer function of Eq. (8.3-1) models the response of the car's speed to a change in the throttle position.

As given in Eq. (8.3-1), the plant's time response to a step change in the throttle position is that of a dominantly first-order system with a time constant of 1 sec, which is slowed or delayed slightly by the presence of additional poles. The dominant pole can be thought of as arising from the car's attempt to gain speed against the dissipative

force of aerodynamic drag while the additional poles arise from the dynamics required for the motor to increase its output in response to the change in throttle position.

The constant in the numerator is dependent on the units chosen to represent the input and output. It changes depending on whether the speed is in miles per hour or kilometers per hour and whether the throttle position is measured in inches from the floor or some angular deflection. Once the units are chosen, however, the constant is set and is part of the plant description. We assume that sensors and actuators are represented as part of the plant transfer function so that for an automated controller the input is really a voltage proportional to the throttle position and the output is really a voltage proportional to the speed. The sensor and actuator dynamics are assumed to be fast enough (of wide enough bandwidth) so that they are constant across frequencies of interest and add no dynamics. (They then become high frequency unmodeled dynamics and could be addressed by plant perturbations in a more refined design.)

The effect on the closed-loop system response that arises from increasing the gain of a proportional compensator can best be seen using a root locus. The root locus for the proportional control system of Fig. 8.3-1 using the plant model of Eq. (8.3-1) is given in Fig. 8.3-2. The scenario described at the beginning of this example indeed occurs as K is increased. First the dominant pole moves out along the negative real axis, indicating a faster response. Then complex conjugates poles break away from the real axis and move towards the imaginary axis, indicating oscillatory responses of increasing frequency and decreasing damping. Finally, the closed-loop poles move across the imaginary axis, indicating an unstable response, specifically, an oscillatory response with growing amplitude.

The results should match the intuition we have developed from driving a car. If we overreact to small desired changes in speed by making large swings in the the throttle pedal position the result will be a herky-jerky ride marked by sudden accelerations and decelerations. We have probably all ridden with stressed out drivers whose gain is too high, and the resulting oscillation was probably quite annoying.

While we can see how the gain affects the closed-loop pole position on the root locus, we do not get much information from the root locus about how much of the possible benefits of feedback control are achieved. Remember from Sec. 3.2 that, as well as a good transient response, a good feedback control design achieves sensitivity reduction and disturbance rejection. Remember also that these latter goals are achieved by creating as large a return difference transfer function over as large a bandwidth as possible.

Observe Fig. 8.3-3 where the Bode and Nyquist plots are drawn for the loop gain of Fig. 8.3-1 for three different values of K. Only the portion of the Nyquist plot corresponding to the positive imaginary axis portion of the Nyquist D-contour is drawn. This is the usual practice in design as it prevents clutter on the diagram. The

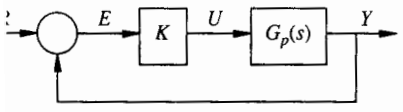


FIGURE 8.3-1
A proportional compensator system.

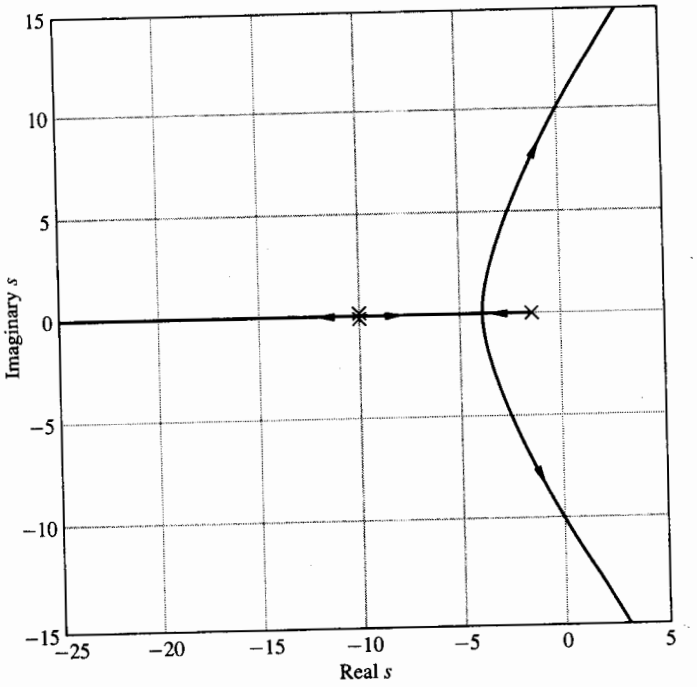
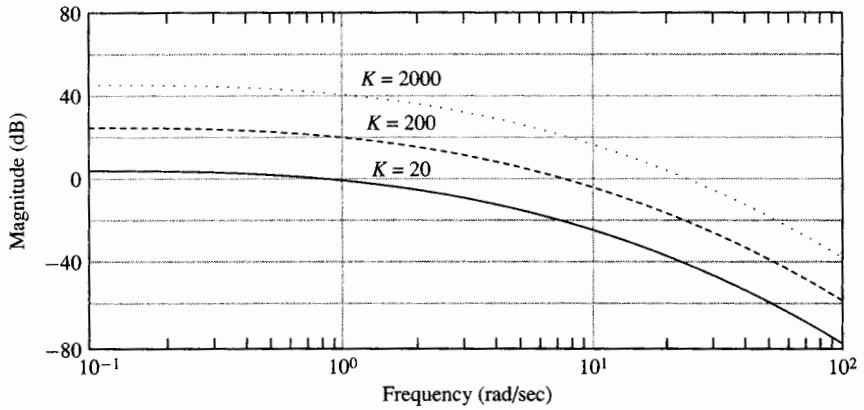


FIGURE 8.3-2
Root locus for proportional compensator.

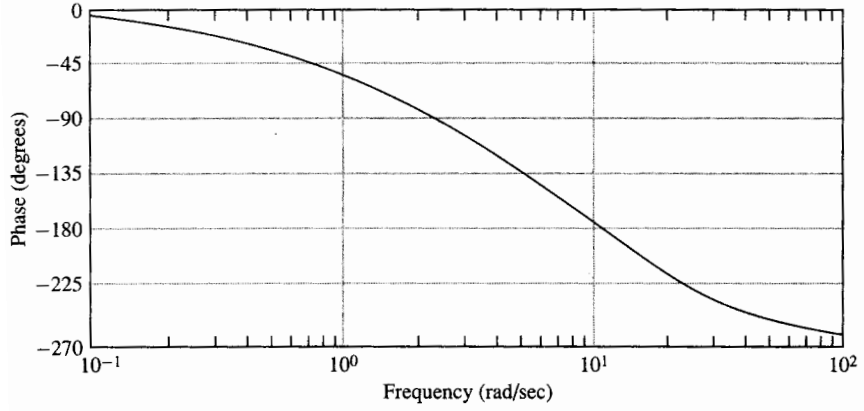
designer usually thinks of altering this portion of the Nyquist plot with the design. The rest of the plot is usually completed only within the designer's mind.

Qualitative information about the transient response can be read off Nyquist and Bode plots. For $K = 20$, the (completed) Nyquist plot shows no encirclements of the origin and a stable closed-loop system. In addition, the Nyquist plot avoids the -1 point well at all frequencies, indicating no peak in the closed-loop frequency response and a non-oscillatory closed-loop time response. The speed of the closed-loop system response can be approximated by finding the bandwidth of the closed-loop frequency response. The closed-loop frequency response has magnitude of approximately two-thirds at low frequencies where the loop gain has magnitude two and little phase lag. As the loop gain transitions from values larger than one to values smaller than one, the closed-loop frequency response falls off from near unity to smaller values approximately equal to the loop gain. The bandwidth of a control system is usually defined to be up to the frequency where the loop gain crosses one in magnitude; this frequency is the crossover frequency ω_c. It can be seen from the Bode plots of Fig. 8.3-3a that the crossover frequency for $K = 20$ is $\omega_c = 1.5$ This indicates the speed of the step response of the closed-loop system. The closed-loop system will respond with a time constant of $T = 1/1.5 = 2/3$ sec.

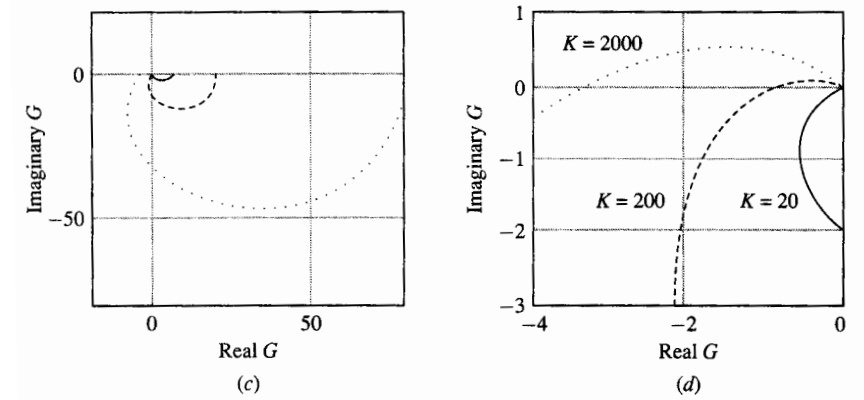
Now, notice the trend on the Nyquist plot of Fig. 8.3-3d as K is increased. As K is increased, the entire plot expands radially. The -1 point is first approached closely,



(a)



(b)



(c) (d)

FIGURE 8.3-3
Proportional control loop with 20, 200, and 2,000. (*a*) Bode magnitude plot; (*b*) Bode phase plot; (*c*) Nyquist plot (wide scale); (*d*) Nyquist plot (expanded scale).

indicating a peak in the closed-loop frequency response (a large M_p peak value) and an oscillatory closed-loop time response. As K is increased further the -1 point is reached and encircled, indicating an unstable closed-loop system. This is consistent with the root locus analysis.

Look at the Bode plot as K is increased. The basic change in the Bode plot is simple. The magnitude plot is shifted up, as K is increased while the phase plot is unaltered. However, there are also more subtle changes. As the Bode magnitude plot is shifted up, the crossover frequency ω_c is shifted out. The larger crossover frequency indicates a larger closed-loop system bandwidth. The larger closed-loop system bandwidth indicates faster closed-loop time responses in following changes in reference inputs and in rejecting disturbances.

Although the Bode phase plot doesn't change when the constant gain of the loop gain transfer function increases, the shifting out of the crossover frequency on the magnitude plot means that the system phase margin is read off the Bode phase plot at a higher frequency. Since the phase plot is sloping down, this means that, in this typical situation, increasing the gain decreases the phase margin. Remember from Sec. 6.5 that for smooth Nyquist plots where the Nyquist plot most closely approaches the -1 point near the crossover frequency, the phase margin provides a good indicator of the degree of oscillation of the closed-loop step response. As the gain increases, the crossover frequency becomes higher, the phase margin decreases, the Nyquist plot passes closer to the -1 point, and the closed-loop transient response becomes more oscillatory. The phase margin going to zero corresponds to an infinite closed-loop frequency response at crossover frequency and two closed-loop poles on the imaginary axis at the crossover frequency ω_c. Such a system is unstable. Thus the point on the imaginary axis of the root locus plot where the poles cross into the right half-plane can be found by examining the open-loop Bode phase plot for the frequency where the phase equals 180°. This frequency is often called the *phase crossover* frequency. In the example this frequency is near 10 rad/sec. The gain margin of the system can be read off the Bode magnitude plot at the phase crossover frequency. In the example, when the gain is 20, the gain margin is 20 dB or a factor of 10. Obviously, when the gain is 200 there is almost zero gain margin and the system is at the edge of instability.

From the Bode plot one can also see fairly directly how well the controller can handle disturbance rejection and sensitivity reduction. Remember that a loop gain that is much greater than one indicates good disturbance rejection and sensitivity reduction. We can examine the limits on the disturbance rejection and sensitivity reduction that are possible using a proportional compensator in the example. The largest gain that produces a stable controller is $K = 200$. When $K = 200$ the loop gain is greater than 11 (22 dB) for all frequencies less than 0.9 rad/sec. This implies that for these frequencies the sensitivity function satisfies

$$\left|\frac{1}{1+G(j\omega)}\right| \leq \frac{1}{-1+|G(j\omega)|} < 0.1 \text{ for } |G(j\omega)| > 11 \qquad (8.3\text{-}2)$$

Thus with $K = 200$ the steady-state magnitudes of disturbances at frequencies less than 0.9 rad/sec are attenuated to less than 10 percent of their original magnitudes.

Also, the sensitivity reduction means that changes in the plant transfer functions at frequencies below 0.9 rad/sec map into changes in the closed-loop transfer function of only one-tenth the magnitude of the change in the loop gain. Of course, an actual proportional compensator would have to be set with a lower gain due to the oscillatory nature of the time response when $K = 200$. Systems with lower gain would have less steady-state disturbance rejection and sensitivity reduction.

Finally, we note that as K increases, the robustness of the closed-loop system to perturbations in the plant decreases. This is due to two factors: first, the decrease in phase margin with increasing K indicates that the nominal loop gain approaches the -1 point more closely as K increases. (This is easily seen on the Nyquist plot of Fig. 8.3-3*d*). Secondly, since the crossover frequency increases as K increases and plant models are relatively less precise at higher frequencies, the multiplicative perturbation near the critical crossover frequency becomes larger as K increases. Again, this is not because the perturbation as a function of frequency changes with K but because the critical crossover frequency becomes higher as K increases.

The question of performance and robustness can be addressed more directly by referring to Fig. 8.3-4. In Fig. 8.3-4*a*, the magnitude of the sensitivity function $(1 + G(j\omega))^{-1}$ is plotted for various values of K that result in stable closed-loop systems. In Fig. 8.3-4*b*, the magnitude of the complementary sensitivity function $G(j\omega)(1 + G(j\omega))^{-1}$ is plotted for various values of K that produce stable closed-loop systems. Recall the robust performance specification, Eq. (8.2-8), repeated here as Eq. (8.3-3).

$$\left|\frac{1}{1+G(j\omega)}\right| p(j\omega) + \left|\frac{G(j\omega)}{1+G(j\omega)}\right| l_m(j\omega) < 1 \qquad (8.3\text{-}3)$$

The smaller the complementary sensitivity function is at any frequency, the larger the multiplicative perturbation that can be handled at that frequency. Fig. 8.3-4*b* shows that, as K increases, the peak in the complementary sensitivity function increases; also the peak moves to a higher frequency. This confirms the double effect of loss of robustness explained in the previous paragraph. The margin for plant mismodeling decreases and the worst case frequency increases to frequencies where more mismodeling is expected. Remember that the complementary sensitivity function is also the closed-loop frequency response function; thus the peaking that occurs with large K indicates oscillatory time responses as discussed previously.

The performance tradeoff displayed in Fig. 8.3-4*a* is even more interesting. As K increases, the sensitivity function decreases at most frequencies, indicating improved steady-state disturbance rejection and sensitivity reduction at these frequencies. At other frequencies (near the shifting crossover frequency) the sensitivity function peaks badly. This is due to the loop gain approaching -1 closely for larger values of K. At some point (upon which the designer and the customer must mutually decide) the disadvantages associated with the increase in sensitivity at some frequencies override the advantages associated with the decrease in sensitivity at most other frequencies.

With all the information available, how does one actually go about designing a proportional compensator, i.e., how does one choose K? When K is small, increasing K has only positive effects. The increasing loop gain provides decreasing sensitivity (and improving disturbance rejection) at almost all frequencies with very little increase in sensitivity at the few frequencies where it does increase. The bandwidth becomes larger and the transient response becomes faster and remains non-oscillatory. Obviously, we'd like to keep increasing the gain. What stops us?

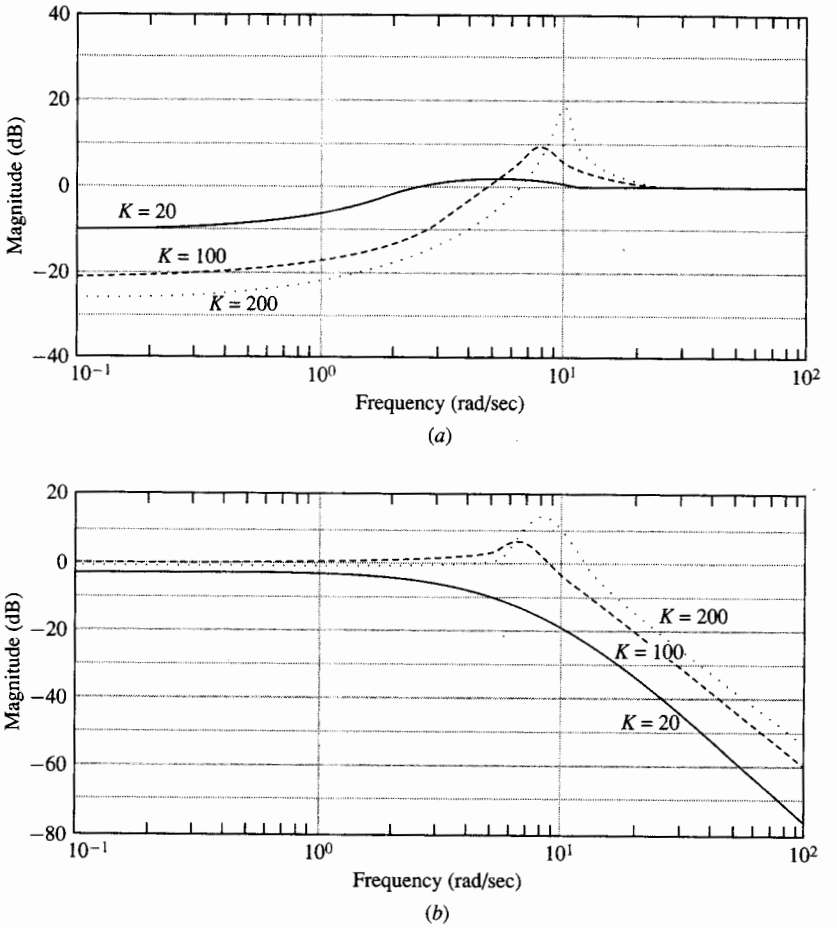


FIGURE 8.3-4
Sensitivity and complementary sensitivity functions for the proportional compensator with $K = 20$, 100, and 200. (*a*) Sensitivity; (*b*) complementary sensitivity.

Clearly we must stop increasing the gain before the closed-loop system becomes unstable. In actuality something else always stops us short of this point. If the system were perfectly modeled and the noise associated with the output measurement were small, the first negative effect of further increasing K would be the simultaneous peaking in the sensitivity and complementary sensitivity functions near the crossover frequency. The peaking occurs from the increase in bandwidth accompanied by the decrease in phase margin. This means the Nyquist plot approaches -1 too closely.

Along with the sharp increase in sensitivity near ω_c, the peak in the complementary sensitivity indicates an oscillatory transient response. Clearly we must set K below the value where this becomes unacceptable. (Let us note here that the oscillatory response to reference inputs can possibly be eliminated by the use of a prefilter but the peak in the sensitivity function indicates an oscillatory response to output disturbances that cannot be easily corrected.)

If there is sufficient uncertainty in the model of the plant we will have to stop increasing K sooner. The problem could arise from larger uncertainties at frequencies much higher than the initial crossover frequency. Uncertainties due to resonant modes may be large enough to require keeping the complementary sensitivity very small at high frequencies. Remember the robust performance condition of Eq. (8.3-3). If $l_m(j\omega)$ is large enough it will limit the size of the complementary sensitivity function and thus limit the value of K. The need for robust performance is even more likely to limit the value of K because of the effect near the crossover frequency. The requirement of good performance for a set of plants is more stringent than the requirement of good performance for the nominal plant model. While the Nyquist plot of the nominal model for a particular K may be kept far enough away from the -1 point to avoid large peaks in the sensitivity and complementary sensitivity function, it requires a smaller K to assure that every possible plant within the perturbed set remains far enough away from the -1 point. Stated another way using Eq. (8.3-3), if $l_m(j\omega)$ is significant at some ω_0, the complementary sensitivity function must be kept small at ω_0 so that the second term in Eq. (8.3-3) can remain small. It may be that the control system bandwidth needs to be kept lower than it would in the nominal case if the modeling uncertainty becomes significantly large at frequencies at or below the bandwidth that could be achieved if there were no modeling error.

Let us review the situation of designing the value K of the proportional compensator using the control design strategy outlined in Sec. 8.2. Assume that performance requirements have been established and have been translated into a performance bound $p(j\omega)$ on the possible perturbed sensitivity function in a manner similar to that presented in Example 6.7-1. Assume also that along with the plant model of Eq. (8.3-1) there has been established a bound $l_m(j\omega)$ on the possible multiplicative plant perturbations. One can use the approximations of Eqs. (8.2-12) and (8.2-15) as a guide for selecting the loop gain. Assume that the magnitudes of $p(j\omega)$ and $l_m^{-1}(j\omega)$ are as plotted in Fig. 8.3-5. From Eqs. (8.2-12) we find that we would like to keep the loop gain above $p(j\omega)$ for the frequencies where $p(j\omega) > 1$ and below $l_m^{-1}(j\omega)$ for frequencies where $l_m^{-1}(j\omega) < 1$. In this case these objectives can be accomplished by setting $K = 200$. The loop gain with $K = 200$ is also plotted in Fig. 8.3-5.

Unfortunately if we set $K = 200$ the magnitude crosses unity while sloping -20 dB per decade (one pole of slope) for a significant length before crossover frequency but begins sloping -60 dB per decade (three poles of slope) shortly after crossover frequency. For this minimum phase system, these slopes indicate a phase very close to 180° at crossover and little phase margin. We have already seen that when a Nyquist plot passes close to the -1 point as the Nyquist plot corresponding

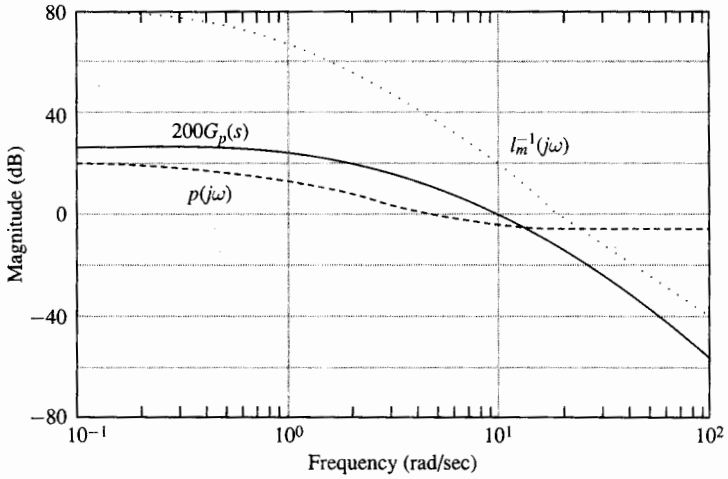


FIGURE 8.3-5
Loop gain for $K = 200$ with constraints.

to $K = 200$ does, the sensitivity function and the complementary sensitivity functio both peak to large values near the crossover frequency. While the approximation of Eqs. (8.2-12) and (8.2-15) are satisfied, the actual constraint of Eq. (8.2-8) i violated. This can be seen by comparing the plot of the sensitivity function and th complementary sensitivity functions for $K = 200$ in Fig. 8.3-4 with the plots of $p(j\omega$ and $l_m^{-1}(j\omega)$ in Fig. 8.3-5 at the crossover frequency. In this case, the factor limitin the loop gain is the behavior near crossover rather than the high frequency unmodele dynamics. The loop gain must be lowered until an acceptable transient response i obtained.

We now realize that we cannot meet all of our objectives with a simple pro portional compensator. Suppose that we decide on the following compromise. Le us find the largest value of K that will still produce a phase margin of 45°. Ir that way, an adequate transient response will likely be achieved. How do we finc this value of K? Refer to the Bode plots for the loop gain with $K = 20$ ir Figs. 8.3-3*a* and *b*. A 45° phase margin means that the loop gain's phase should be 135° at crossover. Remember that the phase plot of Fig. 8.3-3*b* does not change as K changes. The required crossover frequency of $\omega = 5$ rad/sec can be read from Fig. 8.3-3*b*. At $\omega = 5$ rad/sec, the magnitude plot for $K = 20$ passes through -10 dB. To make $\omega = 5$ rad/sec the new crossover frequency we must add 10 dB to the gain of $K = 20$, i.e., we must multiply the gain $K = 20$ by 3 to obtain $K = 60$.

The loop gain with $K = 60$ is plotted in Fig. 8.3-6 along with $p(j\omega)$ and $l_m^{-1}(j\omega)$. The crossover frequency is $\omega = 5$ rad/sec as predicted and, since the phase plot of Fig. 8.3-3*b* is still the loop gain phase plot, the phase margin is 45° as planned. Notice that for $K = 60$ the loop gain has a slope of -20 dB per decade (a one pole slope) for a decade preceding the crossover frequency and also for half a decade

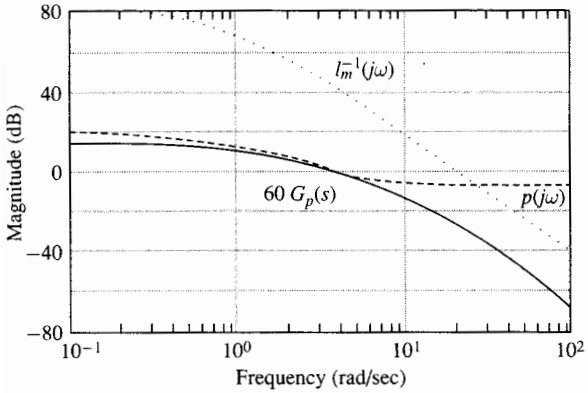


FIGURE 8.3-6
Loop gain for $K = 60$ with constraints.

after the crossover frequency before the slope of −60 dB per decade (the three pole slope) kicks in. This supports the known 45° phase margin. Evaluating the return difference for a 45° phase margin from Eq. (8.2-20), we see that the return difference has magnitude close to 0.75 at crossover so that the sensitivity function has a value of between 2 and 3 dB at crossover. Thus, the design violates the performance criteria near crossover frequency where the value of $p(j\omega)$ is 0 dB. Also, the low frequency performance specification is no longer met.

Sometimes a simple proportional compensator is adequate; usually, as in this case, it is not. In the next section we learn how to correct the low frequency part of the loop gain. Two sections hence, we learn how to correct the performance near crossover frequency. The use of these building blocks for creating more complex controllers is the subject of the remainder of this chapter.

It should be noted that formal specifications of $p(j\omega)$ and $l_m(j\omega)$ such as we used here are rarely available. However, the control designer must interface with the modeling team and higher level system designers to translate the information that is given into a form similar to $p(j\omega)$ and $l_m(j\omega)$ for design. The principles are generally true. The loop gain and therefore the controller bandwidth should be as large as possible to produce closed-loop advantages in speed of response, disturbance rejection, and sensitivity reduction. The size of the loop gain is limited either by the phase margin or crossover frequency indicating oscillatory responses, by the possible effects of model uncertainty at crossover and higher frequencies, or by the need to filter the effects of noisy sensors.

Exercises 8.3.

8.3-1. Find the closed-loop poles and the closed-loop step response for the control system using the plant given by Eq. (8.3-1) and a proportional compensator with $K = 60$. How well does the bandwidth obtained from the Bode plots predict the speed of response? Explain.

Answer: $s = -16.7 \quad s = -2.4 + j6 \quad s = -2.4 - j6$
$\omega_n = 6.5$, slightly higher than crossover of $\omega_c = 5.5$.
M_p occurs at ω_p, slightly higher than ω_c.

8.3-2. Given

$$G_p(s) = \frac{100}{(s+1)(s+10)^2}$$

$$l_m(s) = \frac{(s+1)^3}{10000}$$

If possible, find a proportional compensator so that the closed-loop system meets the following objectives:

1. The system is robustly stable to the perturbations encompassed by $l_m(s)$.
2. The magnitude of steady-state response to a disturbance at $\omega = 0.2$ is less than 20 percent of the value of the disturbance.
3. The transient response oscillates as little as possible given that the first two conditions are met.

Answer: $K = 5$

8.3-3. Let

$$G_p(s) = \frac{-(s-10)}{s(s+1)(s+10)}$$

(*a*) Find the value of K which produces a 45° phase margin. What is the crossover frequency?
(*b*) What is the crossover frequency and phase margin if K is doubled?
(*c*) What is the crossover frequency and phase margin if K is halved?

Answer: (*a*) $K = 0.8$, $\omega_c = 0.8$ rad/sec
(*b*) $\omega_c = 1.2$ $\phi_m = 30°$
(*c*) $\omega_c = 0.4$ $\phi_m = 60°$

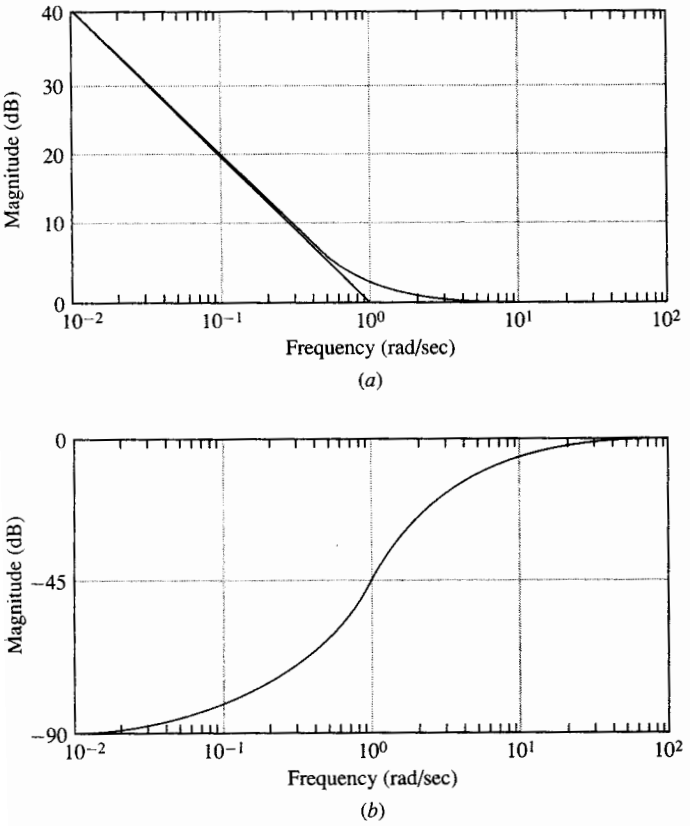
8.4 SERIES COMPENSATOR BUILDING BLOCKS: LAG COMPENSATORS, PI COMPENSATORS

A method of increasing the loop gain transfer function at low frequencies while producing minimal changes at transition frequencies and high frequencies is provided by a lag compensator. A lag compensator is characterized by a pole and a zero. The pole is located either at the origin or very close to the origin. The zero is located somewhat further out the negative real axis, but still at a frequency lower than the crossover frequency. A lag transfer function is given by the equation

$$G_{\text{lag}}(s) = \frac{s+a}{s+b} \qquad \text{with} \qquad a > b \geq 0 \tag{8.4-1}$$

Often, b is set to 0 to provide an integrator in the loop, which increases the System Type and provides improved steady-state behavior, as discussed in Section 4.7. The actual Bode plots and the straight line approximations of a typical lag compensator are shown in Fig. 8.4-1. The transfer function that corresponds to this Bode plot is given by

$$G_{\text{lag}}(s) = \frac{s+1}{s} \tag{8.4-2}$$



IGURE 8.4-1
ode plots of lag compensator. (*a*) Magnitude; (*b*) phase.

hich is clearly of the form of Eq. (8.4-1). The lag compensator has high gain t low frequencies with a gain close to unity for frequencies above the value of ie zero placement. When placed in series with a plant the lag compensator pro- uces the desired effect of raising the magnitude of the loop gain at low frequencies hile leaving the loop gain at the crossover frequency and beyond relatively un- ouched.

There is a price to be paid in achieving the benefits of a lag compensator. s its name suggests, the lag compensator introduces phase lag into the loop gain. ncreased phase lag around the crossover frequency usually moves the Nyquist plot f a loop gain closer to the -1 point with an attendant loss of performance and obustness. As can be seen from Fig. 8.4-1 the amount of phase lag introduced by ie lag compensator decreases as frequency increases beyond the frequency of the ero placement. To minimize the phase lag at crossover frequency, the zero of the g compensator is placed as close to the origin as the need for low frequency loop ain allows. If the zero is placed a decade below the crossover frequency, the lag

compensator reduces the phase margin by only 6°. A wider frequency gap results in a smaller impact on the phase margin.

The thought of adding an integrator in parallel with a proportional compensator provides another interpretation of a lag compensator. A lag compensator can be interpreted as a PI compensator, where *PI* represents Proportional-Integral. A diagram of a PI compensator is shown in Fig. 8.4-2. The effect of the integrator has an interesting intuitive explanation. The error between the reference input $r(t)$ and the plant output $y(t)$ is formed at the summing junction preceding the series compensator. If there were a constant steady-state error in response to a step input, a proportional compensator would result in a constant plant input signal and the error would remain. The output of the integrator, however, would grow in time if its input were constant. This would produce a large control signal, which then must reduce the error to zero if the plant input is to reach steady state. Such behavior is referred to as integral action and the PI compensator is said to have *reset capability* in that it can reset the output to the desired value in the presence of a constant offset disturbance.

The transfer function for a PI compensator is given as:

$$G_{PI}(s) = K_p + \frac{K_I}{s} = \frac{K_p s + K_I}{s} = \frac{K_p(s + K_I/K_p)}{s} \tag{8.4-3}$$

Comparing Eq. (8.4-3) with Eq. (8.4-2) we see that the PI compensator is the same as a lag compensator with $b = 0$ and $a = K_I/K_p$ in series with a proportional gain K_p. We have analyzed the effect of the integral action of a PI compensator from a different viewpoint when we discussed System Type and steady-state Error Constants in Sec. 4.7.

Let's examine the effect of a lag compensator placed in series with the proportional compensator designed for the automobile cruise control system in Sec. 8.3. Recall from Fig. 8.3-6 that a proportional compensator design with $K = 60$ achieves a phase margin of 45° at crossover frequency and acceptable high frequency attenuation, but fails to meet the desired performance specification at low frequency. Recall from the form of the lag compensator of Eq. (8.4-1) that its magnitude is one for frequencies higher than the zero position a. The lag compensator placed in series should raise the loop gain at low frequencies while leaving it relatively unaffected at crossover and high frequencies. Consider using the compensator

$$G_{\text{lag}}(s) = 60\frac{(s + a)}{s} \tag{8.4-4}$$

in series with the plant of Eq. (8.3-1).

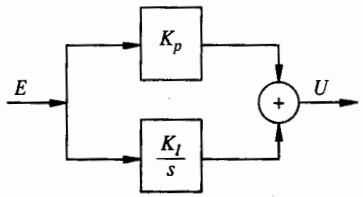


FIGURE 8.4-2
A PI series compensator.

The Bode plots for the resulting loop gains for $a = 0.5$, $a = 2.5$, and $a = 5.0$ are given in Fig. 8.4-3. The hypothetical performance specification $p(j\omega)$ and model uncertainty bound $l_m^{-1}(j\omega)$ from Fig. 8.3-6 are repeated in Fig. 8.4-3*a*. Notice that two phenomena occur as the design parameter a, the zero position, is increased. As a increases the loop gain at low frequencies increases and the range of frequencies affected also increases. At low frequencies the gain from the lag compensator is

$$\left|\frac{s+a}{s}\right|_{s=j\omega} \approx \frac{a}{\omega} \qquad \text{for} \qquad \omega \ll a \tag{8.4-5}$$

The second effect is that as a increases and the zero moves towards crossover frequency, the additional phase lag from the lag compensator at the crossover frequency erodes the phase margin of the proportional compensator.

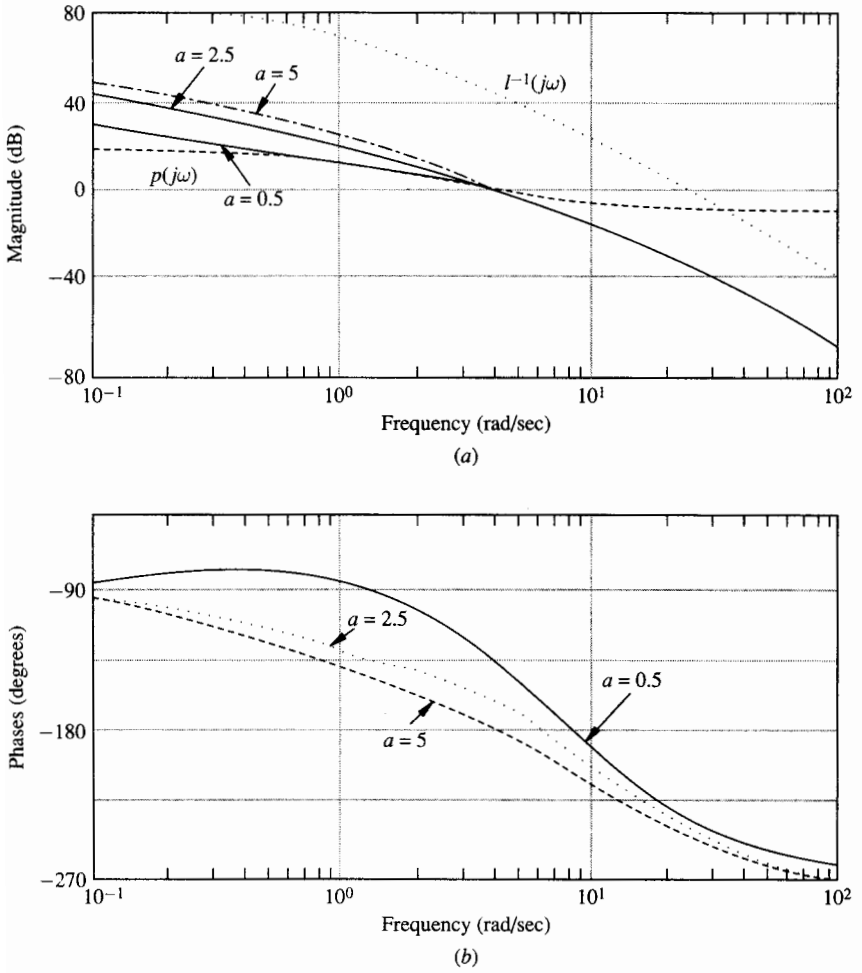


FIGURE 8.4-3
Bode plots for lag compensator control loops. (*a*) Magnitude; (*b*) phase.

When $a = 2.5 = 0.5\omega_c$ the phase margin has been reduced from 45° for the proportional compensator to 22° for the lag compensator. By the time $a = 5.0 = \omega_c$ almost all of the original 45° phase margin has been eroded by the lag compensator. The tradeoff in designing a lag compensator is to set the zero position for sufficient low frequency gain to achieve the performance required at low frequencies while leaving an adequate phase margin at crossover frequency. Note that, in the example, when $a = 0.5 = 0.1\omega_c$ the low frequency performance specification is just met while a phase margin of 39° is salvaged. (Remember that the performance specification curve plotted in Fig. 8.4-3 is only a guideline that the loop gain need exceed only when the magnitude of the loop gain is significantly greater than one.) We know that the resulting 39° phase margin is not adequate because the original 45° phase margin was inadequate. We will correct this in the next section. For now we need to check that we have indeed corrected the low frequency behavior.

Figure 8.4-4*a* shows the complementary sensitivity function for the control loop with $a = 0.5$. The complementary sensitivity function stays well below the inverse

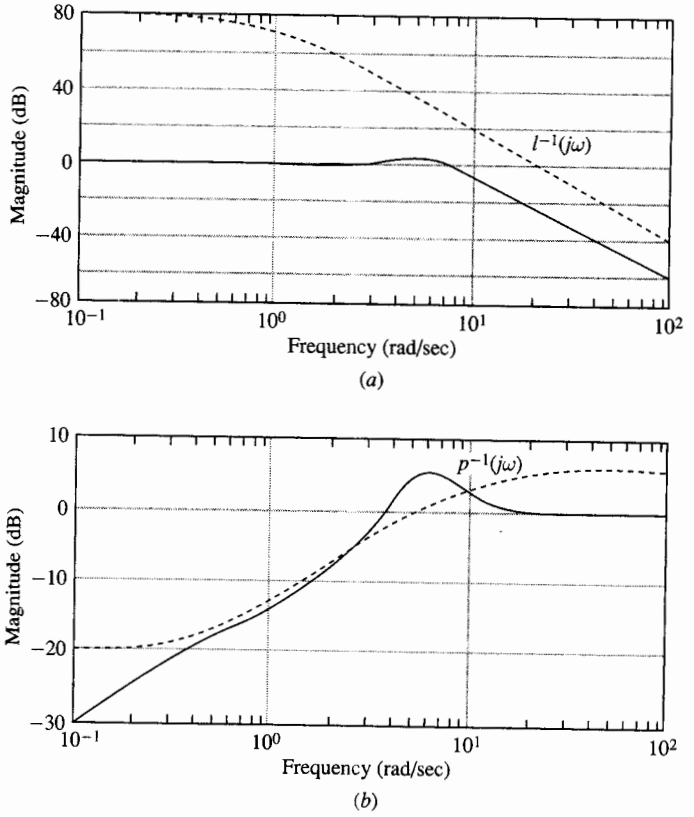


FIGURE 8.4-4
Sensitivity functions for lag control loop with $a = 0.5$. (*a*) Complementary sensitivity function; (*b*) sensitivity function.

of the multiplicative perturbation bound, rendering the design adequate in its robust stability property. Figure 8.4-4*b* shows the sensitivity function for the design. The sensitivity function is indeed sufficiently small at low frequencies. As expected, it rises above the inverse of the performance specification near the crossover frequency. Thus this design is unacceptable. At this point it is unclear how to fix this problem. If the gain is reduced or the zero position of the lag compensator is reduced the performance specification is violated at low frequencies. What is needed is a way of increasing the phase margin around crossover frequency. We learn how to achieve this in the next section, where we study lead compensators.

Exercises 8.4

8.4-1. Compare the closed-loop step responses of the proportional compensator $G_C(s) = 60$ and the lag compensator

$$G_{\text{lag}}(s) = \frac{60(s+0.5)}{s}$$

when used with the plant of Eq. (8.3-1). Next, compare the steady-state response of the two systems to an output disturbance $d(t) = 10 \sin 0.01t$.

8.4-2. Design a lag compensator series compensator for the plant

$$G_p(s) = \frac{100}{(s+1)(s+10)}$$

that does not effect the plant's high frequency magnitude response and has minimal effect on the phase margin. The lag compensator should make the plant-compensator combination into a Type 1 system so that it will follow a step input with no steady-state error. Sketch the Bode magnitude and phase plots of both the plant and combination plant-controller.

8.4-3. Let

$$l_m(s) = \frac{(s+0.1)^2}{100}$$

$$G_p(s) = \frac{10}{(s+0.1)(s+10)}$$

$$p(j\omega) = \frac{10}{(s+0.1)(s+1)}$$

Design a PI or lag compensator to meet the specifications of Eq. (8.2-8).

8.5 SERIES COMPENSATOR BUILDING BLOCKS: LEAD COMPENSATORS, PID COMPENSATORS

While lag compensators can be used to improve low frequency gain at the expense of slightly decreased phase margin, a series compensator element called a *lead compensator* can be used to increase a system's phase margin. A lead compensator provides a positive phase shift or phase lead over a limited frequency range. The equation for

a lead compensator with unity dc gain is

$$G_{\text{lead}}(s) = \frac{1+\frac{s}{c}}{1+\frac{s}{d}} = \frac{d}{c}\left(\frac{s+c}{s+d}\right) \qquad 0 < c < d \tag{8.5-1}$$

The lead compensator may also be written with unity high frequency gain as

$$G_{\text{lead}}(s) = \frac{s+c}{s+d} = \frac{c}{d}\frac{\left(1+\frac{s}{c}\right)}{\left(1+\frac{s}{d}\right)} \qquad 0 < c < d \tag{8.5-2}$$

The lead compensator consists of a zero at $s = -c$ and a pole at $s = -d$ on the negative real axis with the zero closer to the origin than the pole. The Bode plots and straight line approximations of a lead compensator are shown in Fig 8.5-1. The transfer function plotted in Fig. 8.5-1 is

$$G_{\text{lead}}(s) = \frac{1+\frac{s}{0.1}}{1+s} \tag{8.5-3}$$

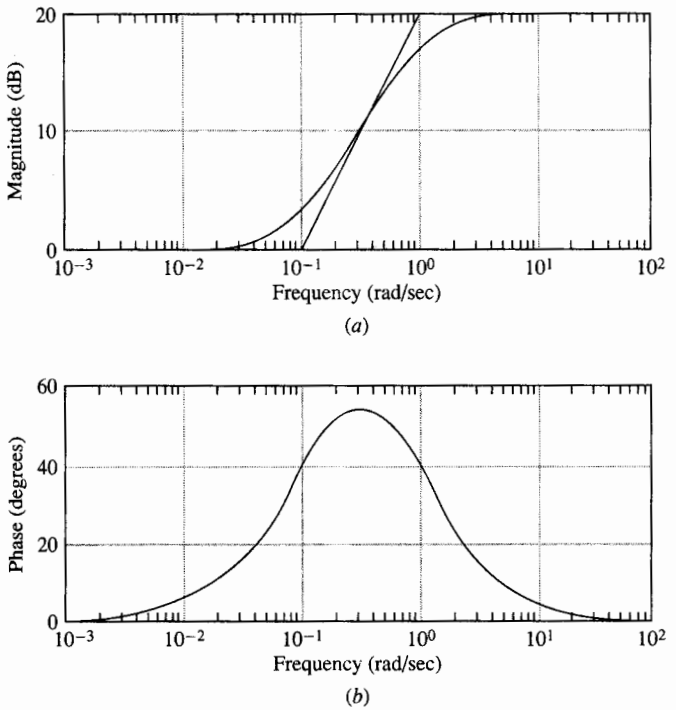


FIGURE 8.5-1
Bode plots for a lead compensator. (*a*) Magnitude; (*b*) phase.

The concept of the lead compensator is that the zero produces a positive slope in the Bode magnitude plot with its attendant phase lead. The pole is provided to create a proper and realizable transfer function; it returns the magnitude plot to its original slope. The need for this pole limits the frequency range over which phase lead is obtained and it also limits the maximum amount of phase lead obtained. In Eqs. (8.5-1) and (8.5-2), $d > c$ and we can define m as

$$m = \frac{d}{c}$$

In its usual application, a lead compensator is arranged so that its zero and pole straddle the crossover frequency of the resulting loop gain so that the maximum phase lead occurs at the resulting crossover frequency. This strategy provides for the greatest increase in phase margin, thereby increasing the return difference transfer function and improving performance through the transition frequencies. From Eq. (8.5-1) or Eq. (8.5-2), we can solve for the maximum amount of phase lead achieved, ϕ_l, and the frequency at which this maximum occurs, ω_l. The equation for ϕ_l is

$$\phi_l = \sin^{-1}\left(\frac{m-1}{m+1}\right), m = \frac{d}{c} \tag{8.5-4}$$

Solving for m in terms of ϕ_l, we find

$$m = \frac{1+\sin\phi_l}{1-\sin\phi_l} \tag{8.5-5}$$

The equation for ω_l is

$$\omega_l = \sqrt{cd} = c\sqrt{m} \tag{8.5-6}$$

Notice that

$$\log\omega_l = \frac{1}{2}(\log c + \log d) \tag{8.5-7}$$

so that the frequency of maximum phase shift occurs halfway between the frequency of the zero and the frequency of the pole on the log ω scale of the Bode plot. The relationship of Eq. (8.5-5) between the maximum phase attained and m can be seen in Table 8.5-1.

TABLE 8.5-1
The maximum phase lead of a lead compensator as a function of *m*

m	ϕ_l
3	30°
6	45°
10	55°
20	65°
100	78°

There is a side effect in using the lead compensator. As demonstrated in Fig. 8.5-1, the lead compensator of Eq. (8.5-1) produces a magnitude plot with gain that is $20\log m$ dB larger at high frequencies than at low frequencies. This gain increase tends to shift the crossover frequency out to a higher frequency. Conversely, if the lead compensator is implemented in the form of Eq. (8.5-2) with unity high frequency gain the low frequencies are attenuated by $20\log m$ dB and the crossover frequency is decreased.

In using the lead compensator, the effect of the lead compensator on the magnitude plot should be considered so that the maximum phase lead occurs at the new crossover frequency rather that the old crossover frequency. Here is the key observation. Since ω_l, the frequency of maximum phase lead, occurs halfway between the zero and pole positions on the Bode plot, the compensator achieves half of its gain (on a logarithmic basis) at this frequency. For the unity dc gain version of Eq. (8.5-1) the value of the magnitude of the lead compensator at ω_l is given as

$$20\log|G_{\text{lead}}(j\omega_l)| = 10\log m \text{ dB} \tag{8.5-8a}$$

or

$$|G_{\text{lead}}(j\omega_l)| = \sqrt{m} \tag{8.5-8b}$$

The new crossover frequency occurs at the frequency where the magnitude of the previous loop gain passes through $1/\sqrt{m}$ or $-10\log m$ dB.

For the unity high frequency gain version of Eq. (8.5-2) the loop gain is partially attenuated at ω_l

$$20\log|G_{\text{lead}}(j\omega_l)| = -10\log m \text{ dB} \tag{8.5-9a}$$

or

$$|G_{\text{lead}}(j\omega_l)| = 1/\sqrt{m} \tag{8.5-9b}$$

The new crossover frequency in this case occurs at the frequency where the magnitude of the previous loop gain passes through $\sqrt{m}$ or $+10\log m$ dB. These facts are used to properly place a lead compensator.

Let us examine the effects of a lead compensator in a loop gain transfer function and see how the previous observations can be used to properly place a lead compensator for maximal benefit, i.e., with maximal phase margin at the resulting crossover frequency. Let us first examine the use of the unity dc gain version of the lead compensator given by Eq. (8.5-1).

Assume a plant is given by

$$G_p(s) = \frac{1}{s(s+1)} \tag{8.5-10}$$

Such a plant is typical of position control systems where there is a first-order response between the input actuator and the velocity of the output. The velocity is integrated to produce the position output.

The Bode plots of the plant of Eq. (8.5-10) are given in Fig. 8.5-2. The closed-loop controller is nominally stable for all proportional compensators with negative feedback followed by positive gain. As the gain of the controller is increased the crossover frequency of the system is increased and the phase margin is eroded.

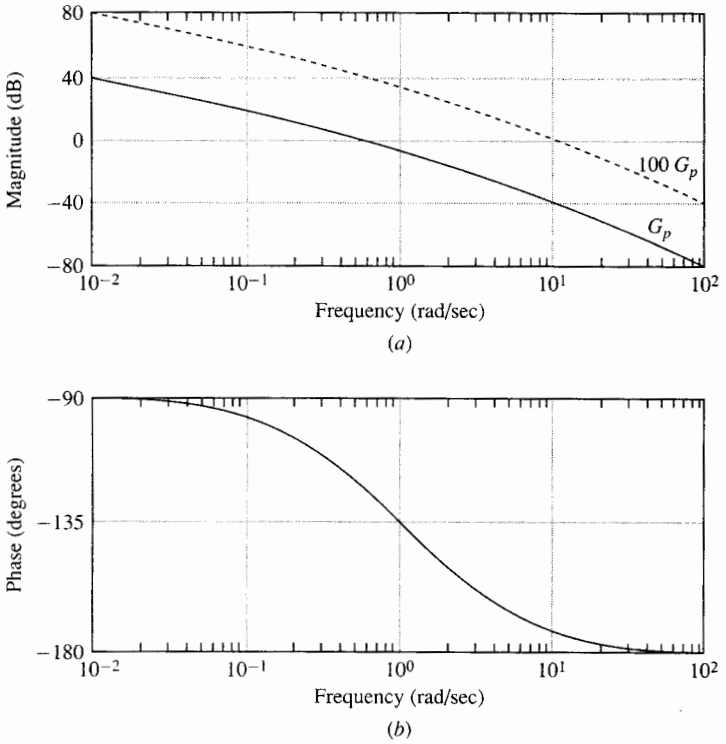


FIGURE 8.5-2
Bode plots of G_p and $100G_p$. (*a*) Magnitude; (*b*) phase.

Assume that a bandwidth of $\omega_c = 10$ is considered adequate speed. From the plant's Bode plots we can see that, at $\omega = 10$, the plant's magnitude is about 0.01 and its phase is 175°. Thus we conclude that, if a proportional compensator is used in the first controller design iteration, a gain of 100 is needed to set the crossover frequency at $\omega_c = 10$. The resulting loop gain Bode plot, also shown in Fig. 8.5-2, is large at low frequencies and small at high frequencies as generally desired, but it transitions from high to low frequencies too quickly. The large slope at crossover frequency around the Bode magnitude plot indicates a large amount of phase lag at crossover frequency and a small phase margin. The phase plot is unchanged by the proportional compensator so that the phase margin at the new crossover frequency of $\omega_c = 10$ is about 5°.

Now we attempt to modify the phase margin with the use of a lead compensator. Let us first examine adding a unity dc gain lead compensator of Eq. (8.5-1) to the loop gain $100G_P(s)$. Assume that we'd like to increase the phase at the new crossover frequency by 55°. Using either Table 8.5-1 or Eq. (8.5-5) we find that we must use $m = 10$. The only parameter left to choose is ω_l. After selecting m and ω_l, the c and d needed in Eq. (8.5-1) can be found using Eq. (8.5-6).

Choosing ω_l is somewhat tricky. Remember that the presence of the lead compensator changes the crossover frequency and we want the maximal phase lead which appears at ω_l to coincide with this new crossover frequency. Since the unity dc gain lead compensator adds $+10\log m$ dB of gain at ω_l we set ω_l equal to the frequency where the old loop gain passes through $-10\log m$ dB. The $10\log m$ dB gain produced by the lead compensator moves the gain at ω_l to 0 dB, i.e., ω_l becomes the new crossover frequency. In our case with m=10 we find the point where $100G_P(s)$ passes through -10 dB.

Figure 8.5-3 is a magnification of the magnitude plot of $100G_P(s)$ near the crossover frequency. From Fig. 8.5-3 we find that we should set ω_l around $\omega_l = 20$. (Closer calculation would put $\omega_l = 18$.) The resulting lead compensator with $m = 10$ and $\omega_l = 20$ is then calculated to be

$$G_{\text{lead1}}(s) = \frac{\frac{s}{6.3} + 1}{\frac{s}{63} + 1} = \frac{10s + 63}{s + 63} \tag{8.5-11}$$

The Bode plots of the proportionally compensated $100G_P(s)$ and the lead compensated loop gains are shown in Fig. 8.5-4. The crossover frequency and the maximal phase lead are indeed near $\omega_l = 20$ and the new phase margin is 58° (55° from the lead compensator and 3° from $100G_P(s)$ at $\omega_l = 20$). All the calculations work out as predicted.

The root locus for the loop gain with the lead compensator is a typical root locus for lead compensators and is shown in Fig. 8.5-5. The zero of the lead compensator at $s = -6.3$ "pulls" the closed-loop poles further into the left half-plane. The additional pole at $s = -63$ is chosen far enough out so that it has little effect on the dominant part of the root locus. The reason for this pole is to make the lead compensator physically realizable and to make the Bode magnitude plot of the compensated loop gain attenuate quickly enough. The chosen closed-loop poles are marked with stars.

From the root locus it appears that we could get better response by increasing the gain of the controller. Actually, for this plant, this observation is mathematically correct. In real applications, however, too large a loop gain would cause a loss of ro-

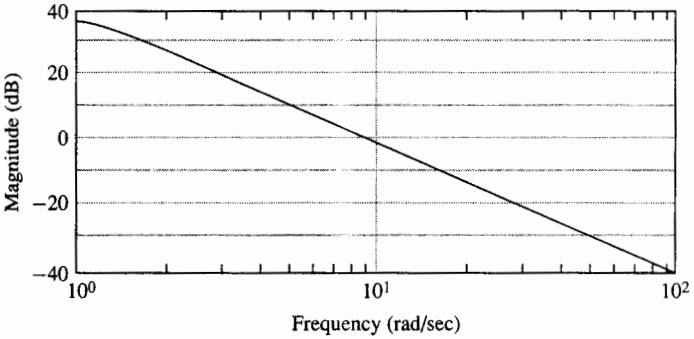


FIGURE 8.5-3
Magnification of the magnitude plot of Fig. 8.5-2 near crossover.

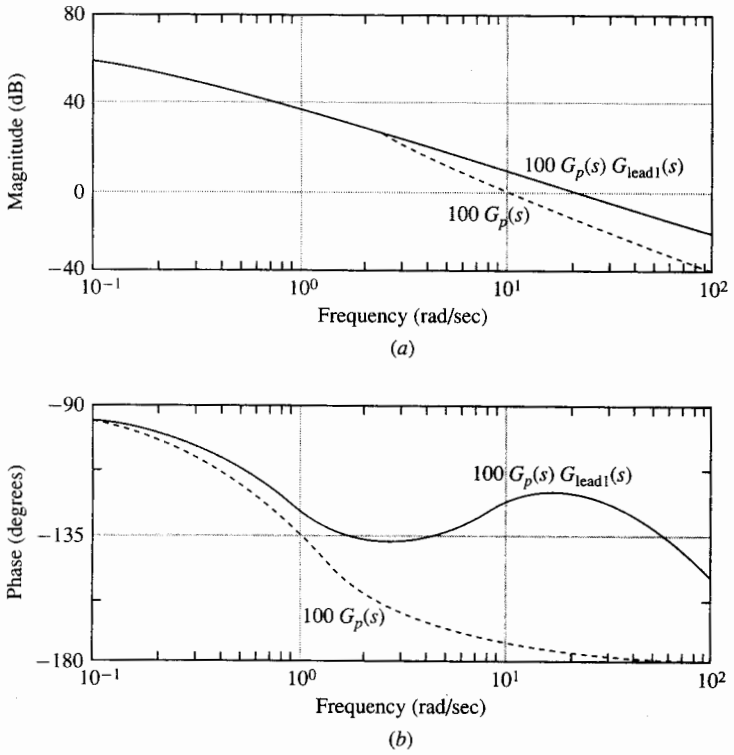


FIGURE 8.5-4
Bode plots of proportional and lead compensated loop gains. (*a*) Magnitude; (*b*) phase.

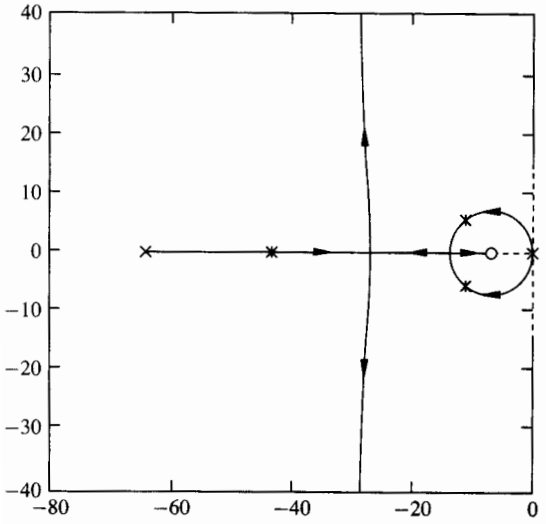


FIGURE 8.5-5
Root locus of lead compensated loop gain.

bustness properties. Since we wanted to concentrate on the lead design we chose not to complicate the problem statement by including a bound on unmodeled dynamics in the form a multiplicative perturbation. Instead we chose to state simply that the bandwidth of the system should be around 10 rad/sec. This is actually a more common way for model builders and control practitioners to describe model limititations. Because they know that bad things will happen if the system response is too fast, they express the model's limitations by putting a constraint on the bandwidth. It is the control designer's job to press the modeler on the exact nature of the uncertainties if the goal is to press the performance further. (It should be noted that often the control designer and the modeler are the same person. In that case, that person must develop a more refined understanding of the system and the requirements if the desire is to extend performance.)

Starting again from the loop gain, $100G_p(s)$, let us examine the effect of using a unity high frequency gain lead compensator given in Eq. (8.5-2). This time, the lead compensator attenuates the gain below and around the transition frequencies so that the unity high frequency gain lead compensator lowers the crossover frequency. For the unity high frequency gain lead compensator the gain at ω_l, the frequency of maximal phase lead, is $-10\log m$ dB (note the minus sign). To assure that ω_l coincides with the new crossover frequency, ω_l is chosen as the frequency where the previous loop gain magnitude crosses $+10\log m$ dB. (Notice the plus sign.) Suppose that this time, for a change, we intend to increase the phase margin by 45°. Now Eq. (8.5-5) dictates that $m = 6$. In this case we compute $10\log 6 = 7.8$ and from the Bode magnitude plot of Fig. 8.5-3, we select $\omega_l = 7$. The lead compensator is computed from Eq. (8.5-2) as

$$G_{\text{lead2}}(s) = \frac{s + 2.8}{s + 16.8} \tag{8.5-12}$$

The Bode plots of $100G_p(s)$ and $100G_p(s)G_{\text{lead2}}(s)$ are given in Fig. 8.5-6. Notice that the new crossover frequency is very near $\omega_l = 7$ as planned and that 45° phase lead has been added to the previous phase lead of 8° at this new crossover frequency. Notice also that the addition of the lead compensator reduces the system's bandwidth and lowers the low frequency gain but leaves the high frequency magnitude basically unaltered. The root locus of this lead compensator is similar to the root locus of the unity dc gain lead compensator since the root locus does not display the small difference in high frequency gain, low frequency gain and bandwidth as clearly as the Bode plot does.

8.5.1 PID and Lead-Lag Compensators

Very early in the design of control systems it was discovered that a great many systems can be adequately controlled with a simple series compensator. This is especially applicable to the control of chemical and manufacturing processes. This simple series compensator, called a *PID compensator*, consists of three terms

$$G_{\text{PID}}(s) = K_p + \frac{1}{T_I s} + \frac{T_D s}{\frac{s}{p} + 1} \tag{8.5-13}$$

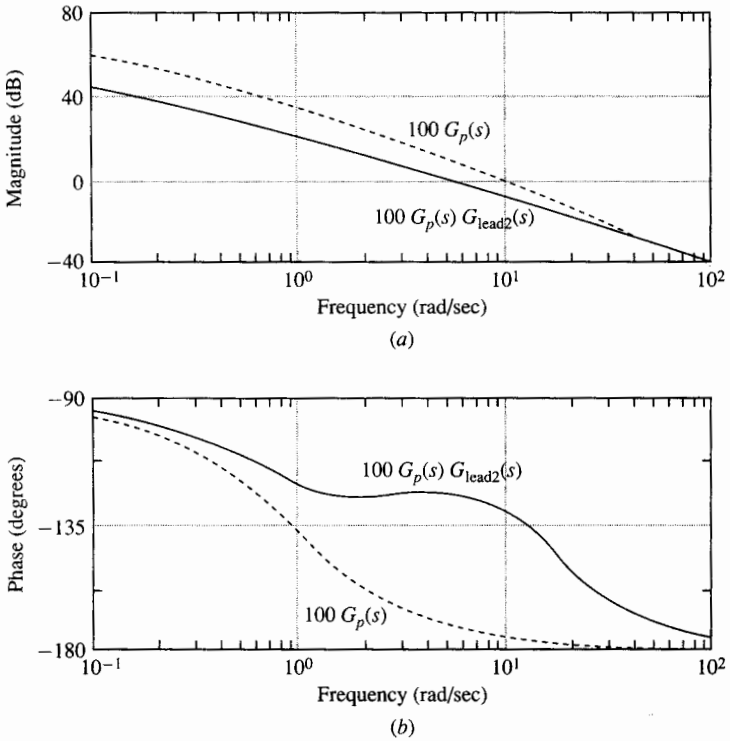


FIGURE 8.5-6
Bode plots for proportional and second lead compensated loop gains. (*a*) Magnitude; (*b*) phase.

The first two terms comprise the *PI* compensator, which was found in Sec. 8.4 to be equivalent to a lag compensator. The three parameters K_p, T_I, and T_D are adjusted to produce acceptable control performance. The first term is called the *proportional* term. The second term is called the *integral* or *reset* term. The third term is called the *derivative* term. The parameter p is generally chosen large. The pole controlled by p is present so that the compensator is realizable and the high frequency gain is limited. Intuitively, derivative action creates a control signal that is sensitive to changes in the error between the actual output and the desired output. Derivative action enables the system to react more quickly than a purely proportional compensator. By combining the terms in Eq. (8.5-13) we obtain

$$G_{\text{PID}}(s) = \frac{\left(T_D + \frac{K_p}{p}\right)s^2 + \left(K_p + \frac{1}{T_I P}\right)s + \frac{1}{T_I}}{s\left(1 + \frac{s}{p}\right)} \tag{8.5-14}$$

The PID series compensator contains two poles, one at the origin and one at a large negative value. The second-order numerator indicates that there are also two zeros. The zeros are usually located at intermediate frequencies between the two poles. By considering each zero as paired with a pole, we can see that the PID compensator is equivalent to the serial combination of a lag compensator and a lead compensator. The utility of this combination is fundamental to controller design. The overall gain is set to provide enough attenuation at high frequencies. The lead compensator, corresponding to the addition of the derivative term, provides for a non-oscillatory response by making the return difference larger at intermediary frequencies. The lag compensator, corresponding to the addition of the integral term, provides for adequate low frequency sensitivity reduction, disturbance rejection, and steady-state error properties by providing high loop gain at low frequencies.

Let us close this section by completing the design of the compensator for the automobile cruise control system discussed in Secs. 8.3 and 8.4.

Example 8.5-1. Recall from Fig. 8.4-4 that after placing a lag compensator (equivalently, a PI compensator) in front of the plant of Eq. (8.3-1) we are left with a design that meets the high frequency constraints on the complementary sensitivity function imposed by the multiplicative perturbation structure but does not meet the performance criterion. In particular, Fig. 8.4-4*b* shows that the sensitivity function stays just below its performance specification at low frequencies and that it exceeds its allowable values near crossover frequency. The problems can be solved by including a lead compensator in the design. Since we cannot afford to lower the low frequency gain much we choose the unity low frequency gain form of the lead compensator given by Eq. (8.5-1). This form increases the bandwidth of the system slightly and leaves the low frequency portion of the Bode plots almost unchanged.

We must be concerned with how much we increase the high frequency gain of the loop. Recall that the unity low frequency gain lead compensator increases the gain at high frequencies by $20 \log m$ dB. If the high frequency gain in Fig. 8.4-3*a* is increased by 20 dB the loop gain butts up against the model uncertainty bound, leaving no margin for robustness of performance. We choose $m = 8$, which raises the high frequency gain 18 dB. (This choice is not obvious; some educated guessing and some iteration of choices is part the design process. Of course, all such guesses are thoroughly tested through analysis and experimentation before implementation.)

The choice of $m = 8$ means that the gain of the system at the frequency of maximal phase lead is increased by $10 \log 8 = 9$ dB. Thus, ω_l is taken as the frequency where the working loop gain passes through -9 dB, in this case $\omega_l = 10$. The resulting lead compensator is given by

$$G_{L1}(s) = 8\frac{s + 3.5}{s + 28} \tag{8.5-15}$$

The loop gain without the lead compensator is shown in Fig. 8.5-7 by the solid line while the modified loop gain with the lead compensator is shown by the dashed line. The new crossover frequency is 10 rad/sec as expected. There is an increase of 51° of phase lead at this frequency as a properly placed lead compensator with $m = 8$ should attain.

The resulting design must be checked against the actual specifications. The results plotted in Fig 8.5-8*a* show that the complementary sensitivity function remains below its constraint as determined by the model bound. The results plotted in Fig. 8.5-8*b* show that the sensitivity function barely remains below its constraint as determined by the performance bound. Thus, assuming that the constraints are conservative enough,

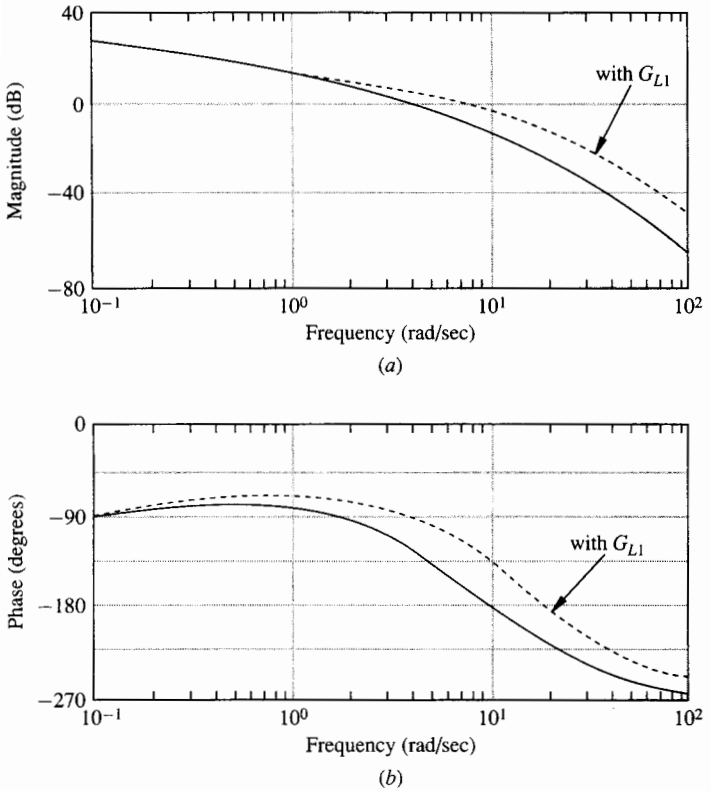


FIGURE 8.5-7
Bode plots of loop gains with and without G_{L1}. (*a*) Magnitude; (*b*) phase.

nominal performance and robust stability are achieved for this design. However, the small margin by which the nominal performance is met leads one to suspect that the robust performance condition of Eq. (8.2-8) is not met.

The results shown in Fig. 8.5-9 confirm that the robust performance condition is not met. Figure 8.5-9 contains a plot of 20 times the log of the quantity

$$\left|\frac{G(j\omega)}{1+G(j\omega)}\right| l_m(j\omega) + \left|\frac{1}{1+G(j\omega)}\right| p(j\omega) \tag{8.5-16}$$

To satisfy the robust performance condition of Eq. (8.2-8) this quantity should remain less that 0 dB for all frequencies. The current loop produces a peak of almost 2 dB around $\omega = 14$ or 15.

What do we do now? Unfortunately, we have run out of standard tricks to play. It is time to use some creativity in applying background knowledge, to use the computer to check out some guesses of what might work and to hope we get lucky. The following explanation is included as a reminder that, while the fundamentals of control design are well understood, any difficult design requires some creative effort. The creativity is greatly enhanced by a sound understanding of fundamental causes and effects and by logical reasoning. It is not clear that a novice control student should be able to adjust

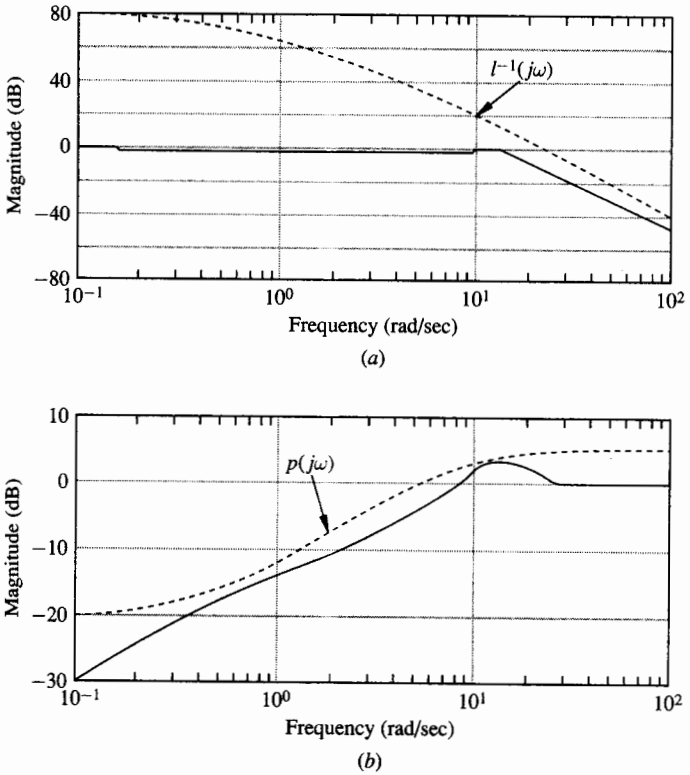


FIGURE 8.5-8
Sensitivity functions for loop gains with G_{L1}. (*a*) Complementary sensitivity functions; (*b*) sensitivity function.

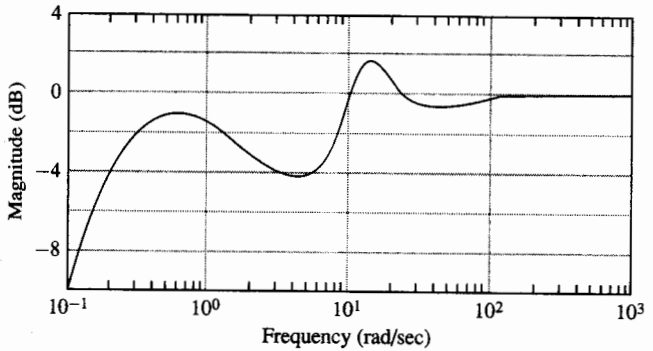


FIGURE 8.5-9
Robust performance measure using G_{L1}.

this design to make it work but a student should have the background by now to follow the reasoning in what follows. We believe that it should be instructive to see the kind of thinking required to fine tune a design in a logical manner.

The first helpful observation is that the problem area in Fig. 8.5-9 occurs not at the crossover frequency but slightly after the crossover frequency. It can be reasoned that the problem of an excessively small return difference might be corrected by increasing the phase lead at this critical frequency. The lead compensator that was tried first can be replaced by a redesigned lead compensator that achieves its maximal phase lead at the slightly higher frequency of $\omega = 14$. The value $m = 8$ is retained so that the very high and very low frequency performance is the same as it is for the first lead compensator. We emphasize that the new lead compensator is not just added to the loop but it *replaces* the original lead compensator. It represents a fine tuning away from the generally correct thought that maximal phase lead is desirable at crossover frequency to place the maximal phase lead where it is needed most *in this particular case*. The resulting lead compensator is computed to be

$$G_{L2}(s) = 8\frac{s+5}{s+40} \tag{8.5-17}$$

The resulting robust performance measure shown in Fig. 8.5-10 is encouraging. The peak has come down substantially. Now there is a real possibility of meeting the objective of keeping the robust performance measure below 0 dB. While the bound is still tight at high frequecies there is some room at low frequencies to trade off against the speed of response to further increase the phase margin near $\omega = 14$.

Consider the following. If the pole of the lead compensator is moved out along the negative real axis, more phase lead will be allowed to develop throughout the frequencies where the lead compensator has effect. The pole must be moved in such a way as to avoid changing the high frequency gain. The increased phase margin is offset by lowering the low frequency gains and increasing the sensitivity function at low frequency. Thus the pole should not be moved too far. Computer iterations show that this logic is sound.

As the parameter placing the pole is increased from 40, the peak in the robust performance measure of Fig. 8.5-10, which is located around $\omega = 14$, decreases while the local peak around $\omega = 0.7$ increases. After some trial and error the following redesigned lead compensator results.

$$G_{L3}(s) = 8\frac{s+5}{s+44} \tag{8.5-18}$$

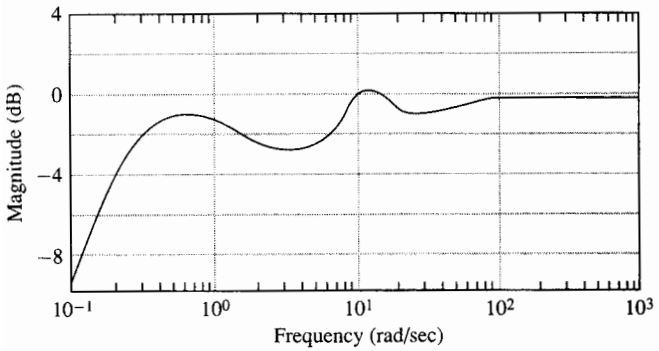


FIGURE 8.5-10
Robust performance measure using G_{L2}.

The plot of Fig. 8.5-11 shows the final robust performance measure for the plant of Eq. (8.3-1) in series with the lead-lag compensator given by

$$G_c(s) = 480\frac{(s+0.5)(s+5)}{s(s+44)} \tag{8.5-19}$$

The designed system meets the design criteria. It should be noted that in this example the design criteria were fairly difficult to meet, requiring some iteration in design beyond the straightforward application of standard building blocks. Often, the original requirements are easily met. Sometimes the original requirements cannot be met. If the original requirements cannot be met, the control designer must modify the design, while those who specify the requirements and those who build the models must also contribute their own modifications. If the requirements are loose it may be possible to take advantage of the additional margins available. If the requirements are unrealistic it must be determined where the expectations are too high, the bounding is too conservative, or the model is too inaccurate. Remember that the vast majority of effort in a control design is involved in creating a model, setting expectations, and validating the design. The actual design which lends itself to mathematical analysis and which is treated in this book is just one part of the overall design.

Exercises 8.5

8.5-1. Using the plant given by

$$G_p(s) = \frac{100}{(s+1)(s+10)}$$

design a lead compensator to increase the phase margin. Use a unity dc gain compensator with $m = d/c = 10$. Set ω_l equal to the original plant crossover frequency. Generate Bode magnitude and phase plots of the controller-plant combination. Then answer the following questions:

(*a*) Where in the frequency spectrum is performance lost due to the addition of the lead compensator?

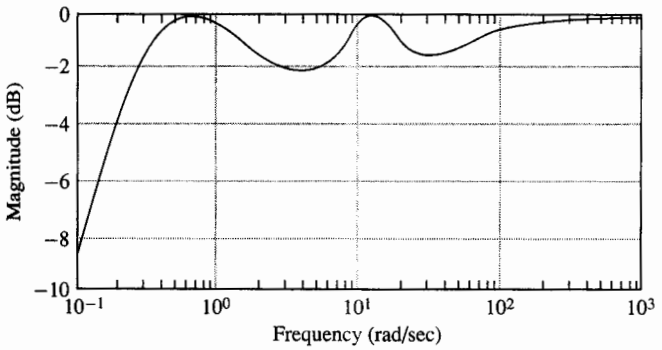


FIGURE 8.5-11
Robust performance measure for final design.

(*b*) Where does the crossover frequency move to when the lead compensator is added? Does this agree with the text?

(*c*) How much phase margin was actually added by the compensator? How does this compare to the maximum phase lead possible when $m = d/c = 10$?

8.5-2. Given the plant

$$G_p(s) = \frac{40}{(s+2)^2}$$

design a unity dc gain lead compensator so that the resulting crossover frequency has a phase margin of 50° and this new crossover frequency is located where the maximum phase lead of the lead compensator occurs.

8.6 SERIES COMPENSATOR BUILDING BLOCKS: HIGH FREQUENCY ROLL-OFF, NOTCH FILTERS, CANCELING PLANT DYNAMICS

Occasionally, additional control elements are needed to provide sufficient attenuation at high frequencies. This attenuation is often referred to as high frequency *roll-off*. It is most simply accomplished by adding poles in series with the previously designed compensator at a frequency higher than the crossover frequency. The poles cause the magnitude plot of the loop gain to roll off at a greater negative slope. However, care must be exercised as the additional poles also produce phase lag at frequencies less than the frequency position of the poles. Each additional pole produces 6° lag at a frequency one decade before the pole position and greater phase lag for higher frequencies. If poles for roll-off are placed too close to the crossover frequency, the system's phase margin decreases, performance suffers, and robustness at intermediate frequencies is adversely affected. This fact is in accord with the discussion in Sec. 8.2 where it was noted that the magnitude plot of the loop gain generally should be allowed to slope gradually through the area near the crossover frequency.

There is a series compensator element that produces a more sudden attenuation of the loop gain's magnitude plot than simple roll-off poles can produce. The element is called a *notch filter*. As its name suggests, the notch filter produces sharp attenuation over a small frequency range, creating a notch in the magnitude plot as seen in Fig. 8.6-1.

The notch filter is the most complicated of the series compensator elements that we address in this book. It consists of a pair of zeros in the left half-plane but very near the imaginary axis and a pair of fairly well-damped poles at the same frequency as the zeros. The zero pair provides the notch in the magnitude plot and the pole pair counteracts the long range increase in slope in the Bode magnitude plot caused by the zeros. A typical notch filter located at frequency ω_N has the transfer function

$$G_{\text{notch}}(s) = \frac{s^2 + 2\zeta_N\omega_N s + \omega_N^2}{s^2 + \omega_N s + \omega_N^2} \tag{8.6-1}$$

where ζ is a parameter that, as we shall see, controls the depth of the notch.

The Bode plots for a notch filter with $\omega_N = 1$ and $\zeta_N = 0.005$ are shown in Fig. 8.6-1. The straight line approximations of the magnitude plots for the numerator

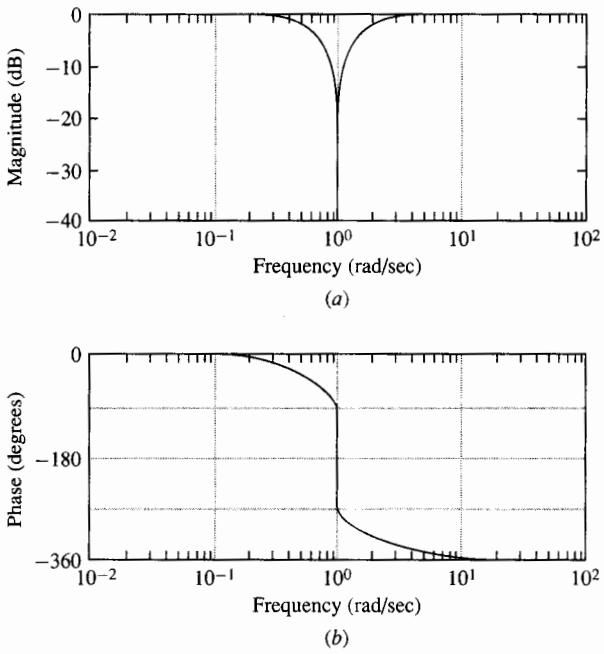


FIGURE 8.6-1
Bode plots of a notch filter. (*a*) Magnitude; (*b*) phase.

and the denominator exactly offset each other. The damping of 0.5 for the pole pair has its actual magnitude plot following the straight line approximation for that contributor closely but the actual Bode magnitude plot for the zero pair has a notch in it of depth $-20\log 2\zeta_N$. This creates the notch of the notch filter. With $\zeta_N = 0.005$ the depth of the notch is $20\log(0.01) = -40$ dB. The poorly damped zero pair produces a 180° phase shift at the notch frequency. Notice that the notch filter also produces phase lag at frequencies less than ω_N. This phase lag can be traded off against the amount of attenuation by manipulating the damping constant of the second-order poles and zeros.

The plots of Fig. 8.6-1 also provide a convenient demonstration of what can be expected from computer plots. The most obvious surprise in the plot is that the phase seems to shift by a negative 180° as the frequency passes the zero pair. Of course we think of the left half-plane zero pair as producing a sudden 180° positive phase shift. There is really no difference if the shifts are sudden enough since phase values are equal when 360° is added or subtracted. The computer computes the phase shift and connects the points. Some programs are better than others about correcting the plot to match the way control theorists think about the plots. Notice also that the depth of the notch is not obvious from the magnitude plot of Fig. 8.6-1. When things are changing quickly the accuracy of the computer plots depends on how closely to the bottom of the notch the computer takes a sample point. We must know enough about what we expect to see in order to properly interpret the computer's output. (In particular, when working with computer aids we must be able to quickly ascertain whether the data have been entered correctly and the computer is correctly programmed.)

Notch filters are often used in conjunction with a roll-off filter. The sharp attenuation of the notch is passed over to the more gradual attenuation of the roll-off filter. In this way the loop gain can be shaped around sharply cornered high frequency constraints. Such constraints appear in practice in mechanical systems when flexible bending modes or other poorly damped oscillatory modes are lumped into the unmodeled dynamics. The Bode plots of Fig. 8.6-2 show the frequency response of the transfer function

$$G_{NR}(s) = \frac{(s^2 + 0.1s + 1^2)}{(s^2 + s + 1^2)} \frac{2}{(s+2)} \tag{8.6-2}$$

When using the notch filter in conjunction with the roll-off filter the phase lag that occurs before the magnitude notch must be carefully considered lest the phase margin of the system near the crossover frequency be adversely affected.

There is a second possible use of notch filters. Since the zeros produce a narrow notch in frequency, a notch filter can be used to counteract a narrow peak in the plant's magnitude plot caused by poorly damped poles. This technique amounts to a near pole-zero cancellation near the stability boundary of the imaginary axis. When using this technique, great care must be taken that the position of the poles is well known so that near perfect cancellation is achieved. In addition, such a cancellation has no effect on oscillatory responses to input disturbances as closed-loop poles still exist near the imaginary axis and the canceling zeros do not appear in the resulting closed-loop transfer function from the input disturbance to the plant output.

A few words about canceling dynamics are in order. It has been mentioned previously that the cancellation of poles and zeros in the right half-plane is not allowed.

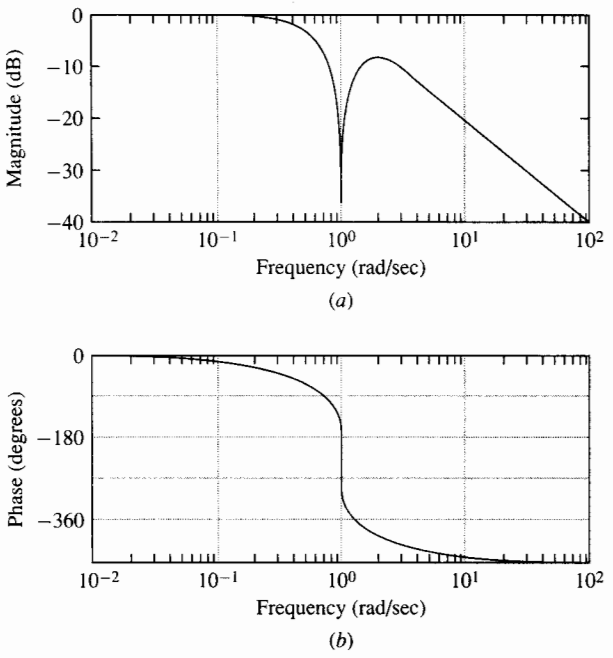


FIGURE 8.6-2
Bode plots of a notch filter with a roll-off pole. (*a*) Magnitude; (*b*) phase.

Let's examine this more closely. Look at the situation of Fig. 8.6-3, which shows an unstable plant pole at $s = 1$ canceled by a zero in a series compensator. (For this demonstration, we ignore the questions of improperness and physical unrealizability of the compensator.)

If the cancellation is perfect, the response to a reference input $R(s)$ is

$$\frac{Y(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1 + G_c(s)G_p(s)} = \frac{\dfrac{1}{s+1}}{1 + \dfrac{1}{s+1}} = \frac{1}{s+2} \tag{8.6-3}$$

The response to an output disturbance $D_o(s)$ is

$$\frac{Y(s)}{D_o(s)} = \frac{1}{1 + G_c(s)G_p(s)} = \frac{1}{1 + \dfrac{1}{s+1}} = \frac{s+1}{s+2} \tag{8.6-4}$$

The two transfer functions are stable. However, the response to a disturbance at the plant input is

$$\frac{Y(s)}{D_i(s)} = \frac{G_p(s)}{1 + G_c(s)G_p(s)} = \frac{\dfrac{1}{(s+1)(s-1)}}{1 + \dfrac{1}{(s+1)}} = \frac{1}{(s+2)(s-1)} \tag{8.6-5}$$

This transfer function is unstable even though we assumed a perfect pole-zero cancellation. If the unstable plant pole were perfectly modeled and the controller zero perfectly implemented, the controller can make sure it doesn't pass signals that excite the unstable dynamics but it can do nothing about countering the effects of other signals or initial conditions that might excite the unstable dynamics. Thus, canceling unstable plant poles is not allowed.

A slightly different situation occurs in attempting to cancel right half-plane plant zeros. In Fig. 8.6-4, the plant zero at $s = 1$ is canceled with a controller pole at $s = 1$.

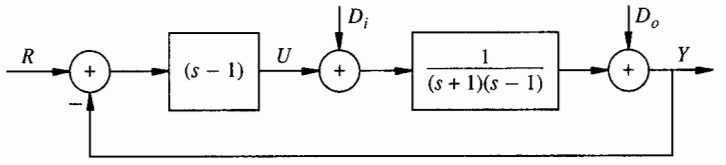


FIGURE 8.6-3
Canceling a right half-plane plant pole.

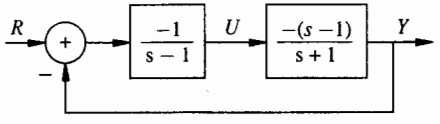


FIGURE 8.6-4
Canceling a right half-plane plant zero.

Again, the transfer function from R to Y is stable.

$$\frac{Y(s)}{R(s)} = \frac{G_c(s)G_p(s)}{1 + G_c(s)G_p(s)} = \frac{\dfrac{1}{s+1}}{1 + \dfrac{1}{s+1}} = \frac{1}{s+2}$$

Now the problem is that the transfer function from the reference input R to the plant input U is unstable.

$$\frac{U(s)}{R(s)} = \frac{G_c(s)}{1 + G_p(s)G_c(s)} = \frac{\dfrac{-1}{s-1}}{1 + \dfrac{1}{s+1}} = \frac{-1}{(s+1)(s-1)} \qquad (8.6\text{-}6)$$

The plant input $u(t)$ grows without bound and surely saturates the plant's actuators.

So we see that even if perfect pole-zero cancellation could be attained, cancellation of right half-plane zeros or poles cannot be allowed. Of course, perfect cancellation is impossible and as a result the transfer functions just considered are unstable in practical situations. Unstable poles in any closed-loop transfer function cannot be allowed. Systems that yield stable closed-loop transfer functions from the input of any block to the output of any block are called *internally stable*. For the series compensator, if the closed-loop system is stable and there are no pole-zero cancellations between $G_c(s)$ and $G_p(s)$ the system is internally stable.

What about canceling left half-plane poles and zeros? Any time that something like a cancellation is forbidden for right half-plane poles or zeros, a thinking engineer must take warning against the same actions when considering poles or zeros in the left half-plane but near the $j\omega$ axis. We have seen that canceled poles and zeros do not disappear completely. Repeating the argument used above, it can be seen that if highly oscillatory plant poles are canceled by a compensator then the oscillatory plant poles still appear in the transfer function between an input disturbance and the output. If the dynamic modes associated with these poles are excited they produce poorly damped oscillations in the output. By repeating the argument used in considering the cancellation of right half-plane zeros, it can be seen that if left half-plane zeros near the $j\omega$ axis are canceled by a compensator these zeros appear as poorly damped poles in the transfer function between the reference input and the plant input. In this position the dynamic modes associated with these poles produce large oscillations in the plant input, which may saturate or stress the input actuators.

While canceling dynamics (poles or zeros) that are in the right half-plane is forbidden and canceling dynamics that are near the imaginary axis is not recommenced unless absolutely necessary, canceling well-behaved plant dynamics is perfectly legitimate. Indeed, one easy way to shape a loop gain into a desirable position when the plant contains well-behaved dynamics is to invert the plant transfer function, canceling all plant dynamics with the controller and to add the desired loop shape in series. This strategy of canceling all well-behaved plant poles and zeros is legitimate although it most probably leads to a controller with more dynamic elements than are absolutely necessary.

We have now been exposed to a number of building blocks for series compensator design. In the next section, we explore a realistic example that shows how to combine these building blocks in our overall control design strategy to produce satisfactory control designs.

Exercises 8.6

8.6-1. Roll-off poles are used to drop the loop gain under high frequency constraints in much the same way that lag compensators are used to raise the loop gain above low frequency constraints. Let

$$l_m(s) = \frac{(s+0.1)^2}{100}$$

$$G_p(s) = \frac{3(s+0.3)}{s^2}$$

$$p(s) = \frac{(s+0.5)}{s^2}$$

Find the phase margin for a unity proportional compensator, i.e., $G_c(s) = 1$. Design a controller using only roll-off poles so that the robust performance measure of Eq. (8.2-8) is met. Find the phase margin of the resulting loop gain. Explain how and why the phase margin changes between the two controllers.

8.6-2. Let

$$G_p(s) = \frac{100}{(s+1)\left(s^2+s+100\right)}$$

and

$$G_c(s) = \frac{10^4 K\left(s^2+s+100\right)}{(s+1000)^2}$$

Sketch the Bode plots of the loop gain when $K = 1$ and when $K = 100$. Find the response to a step disturbance at the plant input when $K = 1$ and $K = 100$. Explain the difference between the two situations.

8.7 A REALISTIC DESIGN EXAMPLE USING A LEAD-LAG COMPENSATOR[1]

In this section, the process of designing a control system is demonstrated by the use of a simple but very real design problem. The design problem concerns attitude stabilization of an experimental communication satellite called ATS-6 (Applications Technology Satellite 6). This spacecraft was launched in May of 1974 to perform communication experiments at geosynchronous (stationary) altitude. It performed its mission virtually flawlessly with control laws very similar to those we develop here.

[1]Much of this material relies on material in Stein and Sandel, *Classical and Modern Methods for Control System Design*, MIT, 1979. It is used with permission.

Detailed information about ATS-6 and its mission can be found in the May 27, 1974 issue of *Aviation Week*. Here we introduce just enough of the mission description to explain the control design problem. Physically, the satellite is a large, 30-foot-diameter, plated nylon mesh parabolic antenna. At its focus is a box of equipment called the earth viewing module (EVM), which contains the communication equipment and all attitude control, thermal control, and telemetry hardware. Power is supplied by two large solar panel drums extended on booms above the antenna. A sketch of the satellite is presented in Fig. 8.7-1*a*.

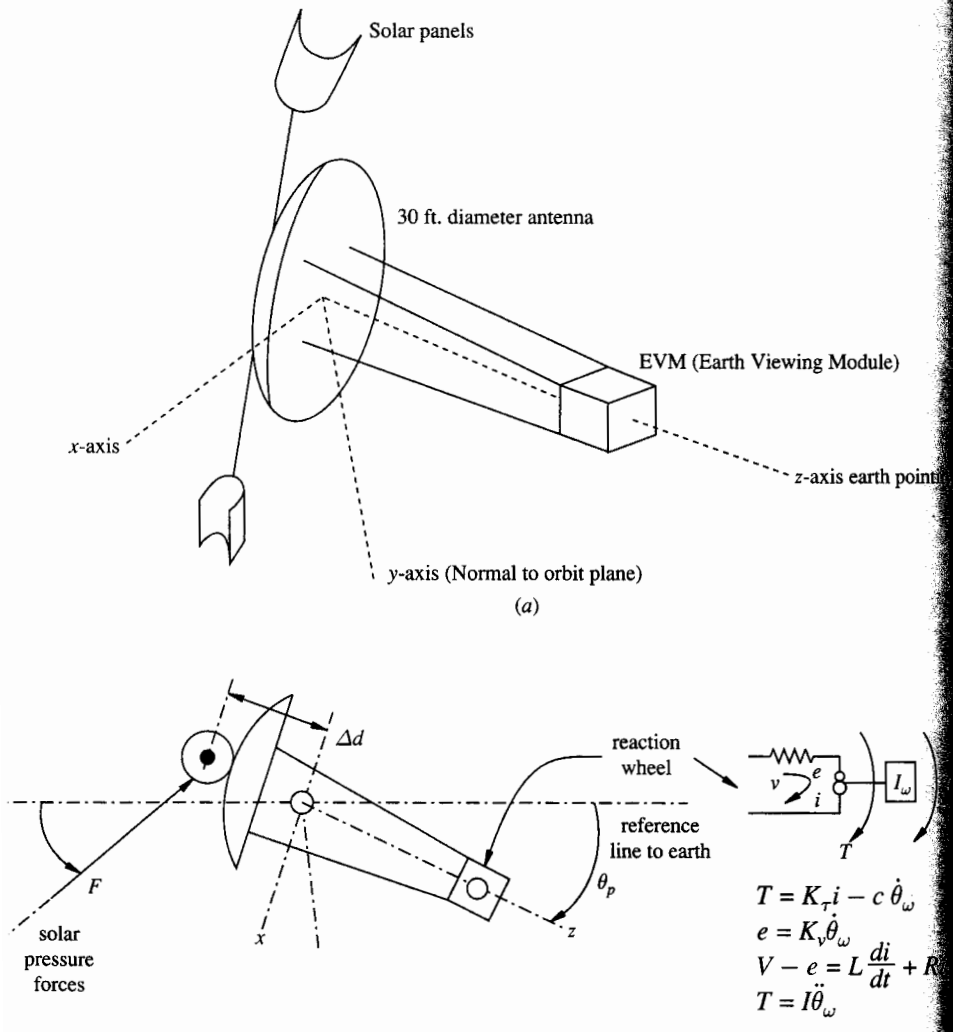


FIGURE 8.7-1
ATS Satellite. (*a*) General diagram, (*b*) variables and equations.

A satellite at geostationary altitude moves along its orbit at the same rate that the earth rotates about its axis. Hence the satellite is stationary with respect to points on earth. The controller's job is to point the communication antenna at selected receiving areas to within an absolute accuracy of 0.01 degree to 0.10 degree of arc, depending on the mission phase. The receiving areas include remote sites over Appalachia, the Rocky Mountains, and India. They also include special targets such as low-earth-orbit satellites (for data relay purposes such as the joint US-USSR Apollo-Soyuz space mission, where ATS-6 was the prime communications link), and aircraft within the earth's atmosphere (for air traffic control experiments). To sense where the antenna is pointing, the controller has two information sources. One is an infrared (IR) sensor on the EVM, which measures pointing angles with respect to the earth's infrared image. As seen from geostationary altitude, this image is a disc that the sensor attempts to keep centered in its field of view. The second information source is the communication system itself. Using wave interference techniques, the system can measure the antenna's pointing attitude directly to within approximately 0.01 degree of arc. The controller is expected to work equally well with either of these information sources. There is also a third sensor, which measures rotations about the line of sight to the earth by reference to the north star, Polaris. This sensor is not considered in this example.

We will consider the control design problem for one axis of motion for this satellite—the pitch axis or y axis. This axis is normal to the orbit plane and, hence, motion about it represents east-west pointing errors of the communication system. Such motions are uncoupled from other attitude motions if all motions are sufficiently small.

Referring to the sketch of Fig. 8.7-1*b* we see the variables of interest defined. The angle θ_p is the pointing angle to be controlled. The control is actuated by a reaction wheel, a device that produces torques by changing the speed of a rotating wheel through the use of a dc motor. The control variable V is the voltage driving the motor. As derived in Exercise 8.7-1, the transfer function between the control input V and the torque T produced by the motor is

$$\frac{T(s)}{V(s)} = \frac{s}{s + 0.01} \tag{8.7-1}$$

Since there is essentially no friction in space the transfer function between the torque T and the angular position Θ_p is a simple double integrator (with appropriate scaling). The relationship between the control input V and the angular position Θ_p is

$$\frac{\Theta_p(s)}{V(s)} = \frac{1}{10^3 s(s + 0.01)} \tag{8.7-2}$$

The angle of the satellite is not sensed instantaneously; there are dynamics associated with the sensors. The measured angle Θ_m is related to the actual angle Θ_p by the transfer function

$$\frac{\Theta_m(s)}{\Theta_p(s)} = \frac{10}{s + 10} \tag{8.7-3}$$

As the sensor follows the pointing angle much faster than the pointing angle is expected to change, we consider the actual and the measured angles to be practically identical and consider the measured angle Θ_m as our variable both to be fed back and to be controlled. The plant transfer function becomes

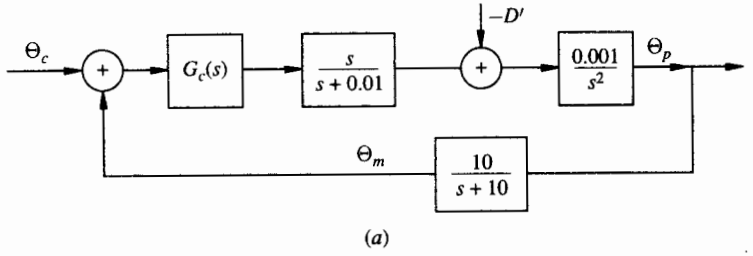
$$\frac{\Theta_m(s)}{V(s)} = \frac{10}{10^3 s(s+0.01)(s+10)} \tag{8.7-4}$$

The pointing control is to be kept accurate in the face of torque disturbances caused by the pressure of solar radiation on the solar collector panels. The torques are cyclical in nature with the period equal to the 24 hours it takes the earth to rotate; the frequency is then equal to 7×10^{-5} rad/sec. The maximum torque is 10^{-5} foot-pounds. The maximal disturbance $d'(t)$ is

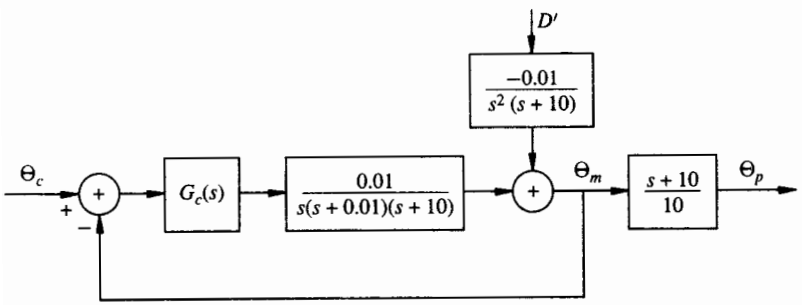
$$d'(t) = 10^{-5} \sin\left(7 \times 10^{-5} t\right) \tag{8.7-5}$$

As the disturbance is itself a torque on the spacecraft, it enters the plant at the same place as the control torque as shown in Fig. 8.7-2*a*. Since it is easier to treat disturbances as output disturbances, the disturbance is reflected to the output as in Fig. 8.7-2*b*. When there is no controller the transfer function from the disturbance $D'(s)$ to $\Theta_m(s)$ is

$$\frac{\Theta_m(s)}{D'(s)} = \frac{-(0.001)(10)}{s^2(s+10)} \tag{8.7-6}$$



(*a*)



(*b*)

FIGURE 8.7-2
Block diagram of the plant. (*a*) Original; (*b*) revised.

The block diagram manipulations between Fig. 8.7-2*a* and Fig. 8.7-2*b* also demonstrate the decision to consider $\Theta_m(s)$ as the output to be controlled since it is available for feedback. The resulting loop is in the G configuration with the inverse of the sensor dynamics appended in series. Again, this filter can be ignored since it passes with little change the signals of frequency less than 10 rad/sec, and the bandwidth of the control loop should be much less than that.

The model is fairly accurate for low frequencies when the satellite is moved slowly enough to avoid exciting any bending modes or flexible resonances. Since the satellite material must be light to be launched economically, the satellite's structural members are flexible and oscillate if the satellite is torqued hard or fast enough to create bending. The fundamental bending mode comes from the beams connecting the EVM to the dish antenna. The oscillations are poorly damped due to the lack of atmosphere in space. The resonant frequency occurs at $\omega = 1$ rad/sec and produces a resonant peak on the magnitude plot of up to 30 dB along with a sudden phase shift of 180°. Since there are also bending modes at harmonics of 1 rad/sec and nonlinearities that become more important at higher frequencies, all these model inaccuracies can be covered by a magnitude-bounded multiplicative perturbation. The perturbation model is given as

$$\tilde{G}_p(s) = G_p(s)\,(1 + L_m(s)) \tag{8.7-7a}$$

$$|L_m(j\omega)| < l_m(j\omega) = \left| \frac{\left(\dfrac{j\omega}{-0.03} + 1\right)^2}{30} \right| = \frac{1 + \dfrac{\omega^2}{0.0009}}{30} \tag{8.7-7b}$$

There is one specific performance requirement. In spite of the disturbance, the pointing angle must be held to within 0.01° or 1.7×10^{-4} rad. From the block diagram we see that

$$\Theta_m(s) = \frac{\dfrac{-0.01}{s^2(s+10)}}{1 + G_c(s)G_p(s)} D'(s) \tag{8.7-8}$$

Take the magnitude of both sides of Eq. (8.7-8). By substituting the maximum disturbance and evaluating the expression at the frequency of the disturbance ω_e, the bound on the size of the pointing error can be transformed into a point on the performance bound of the return difference function.

$$|\Theta_m(j\omega_e)|_{\max} = \frac{\dfrac{0.01}{\omega_e^2\left(\omega_e^2 + 100\right)^{1/2}} 10^{-5}}{\left|1 + G_c(j\omega_e)\,G_p(j\omega_e)\right|} < 1.7 \times 10^{-4} \tag{8.7-9}$$

To meet this criterion the return difference must be large enough at $\omega_e = 7 \times 10^{-5}$ rad/sec. Specifically,

$$\left|1 + G_c(j\omega_e)\,G_p(j\omega_e)\right| \geq \frac{0.01\left(10^{-5}\right)}{\left(7 \times 10^{-5}\right)^2 10\left(1.7 \times 10^{-4}\right)} = 12{,}000 \tag{8.7-10}$$

The other performance requirements are general in nature. The return difference should never be much smaller than one so that the sensitivity remains low and the responses to disturbances are not too oscillatory or amplifying at any frequency. The bandwidth should be as large as the model accuracy allows, so that the response to any disturbance occurs as quickly as possible. The response to a step change in command input θ_c should be smooth with little or no overshoot. These requirements can be translated into guidelines for the shape of the loop as discussed in Sec. 6.7. Rather than establish a complete performance bounding function before designing the system we try to achieve good performance and then examine the resulting worst case performance. The one specific requirement we have is from Eq. (8.7-10).

$$p(j\omega_e) = 12{,}000 \tag{8.7-11}$$

Otherwise we expect $p(j\omega)$ to slope down to a value slightly less than one near crossover and then return to approach one asymptotically.

The guiding equation for robust performance is Eq. (8.2-8), repeated here as Eq. (8.7-12).

$$\left|\frac{G(j\omega)}{1+G(j\omega)}\right| l_m(j\omega) + \left|\frac{1}{1+G(j\omega)}\right| p(j\omega) < 1 \tag{8.7-12}$$

In Sec. 8.2 we made approximations and turned these into conditions on the loop gain for low frequencies and high frequencies as in Eqs. (8.2-12) and (8.2-15). Equation (8.2-12), repeated here as Eq. (8.7-13), is accurate at low frequencies when $l_m(j\omega) \ll 1$ and $p(j\omega) \gg 1$.

$$|G(j\omega)| > \frac{p(j\omega)}{1 - l_m(j\omega)} \tag{8.7-13}$$

The performance requirement given by Eq. (8.7-11) occurs at $\omega_e = 7 \times 10^{-5}$ rad/sec. At this frequency, $l_m(j\omega_e) = 1/30$. We extend the requirement of Eq. (8.7-13) for lower frequencies and mark it as an objective for the loop gain on the Bode plot of Fig. 8.7-3.

Equation (8.2-15), repeated here as Eq. (8.7-14), is accurate at high frequencies when $l_m(j\omega) \gg 1$ and $p(j\omega) \ll 1$.

$$|G(j\omega)| < \frac{1 - p(j\omega)}{l_m(j\omega)} \tag{8.7-14}$$

The approximation is fairly accurate when $l_m(j\omega) > 10$. The curve of $l_m^{-1}(j\omega)$ is sketched on Fig. 8.7-3 as an objective for the loop gain magnitude. The area where $l_m^{-1}(j\omega) < 0.1$ is marked off in a distinct way as the loop gain must stay below this objective. Although there is no specific performance requirement here we need to leave a safety margin so that not only stability but reasonable performance is maintained in the face of perturbations.

Figure 8.7-3 contains the Bode plots of the plant $G_p(s)$ and $10G_p(s)$. The effect of a simple proportional compensator is seen here. With the additional gain both the low and high frequency objectives for the loop gain are met. There are, however,

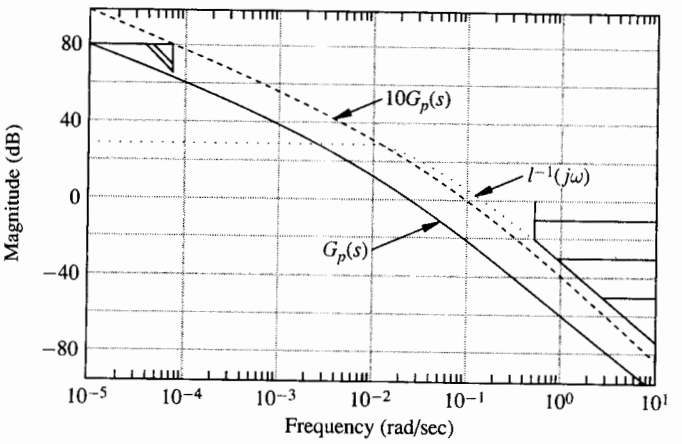


FIGURE 8.7-3
Bode plots of $G_p(s)$ and $10G_p(s)$. (*a*) Magnitude; (*b*) phase.

severe problems at intermediate frequencies. The crossover frequency is $\omega_c = 0.1$ and the phase margin is only 5°. This is unacceptably small since it implies

$$|1 + G(j\omega_c)| = 2\sin\frac{\phi_m}{2} = 0.09 \tag{8.7-15}$$

If the control design were stopped here, the closed-loop frequency response would contain a peak at ω_c with height $1/0.09 \approx 11$ or 21 dB. This would indicate closed-loop poles with a damping of approximately $\zeta = 0.05$. The step response of such a system would have overshoot and oscillation. What is worse is that the plant model could be inaccurate at $\omega_c = 0.1$ so that the actual Nyquist plot is closer to the -1 point than the nominal plot and the system would be even more oscillatory.

Notice, however, that the crossover frequency is low enough that, according to the multiplicative perturbation bound, the model should be quite accurate at this frequency and at least stability is assured. A phase margin this small at a higher frequency where the model is less accurate could result in instability if the controller were implemented on an actual satellite.

The phase margin can be improved by using a lead compensator. It is known from Sec. 8.5 that a lead compensator produces either high frequency amplification if Eq. (8.5-1) is used or low frequency attenuation if Eq. (8.5-2) is used. Referring to the loop gain plots of $10G_p(s)$ in Fig. 8.7-3 it can be seen that in either case the objectives are not met. We decide to use the unity high frequency gain form of Eq. (6.5-2) since the low frequency objective can be recovered by the later addition of a lag compensator. The fact that the high frequency constraint allows about three decades of frequency between the attainable crossover frequency and the low frequency constraint provides confidence that such a lag compensator can be built.

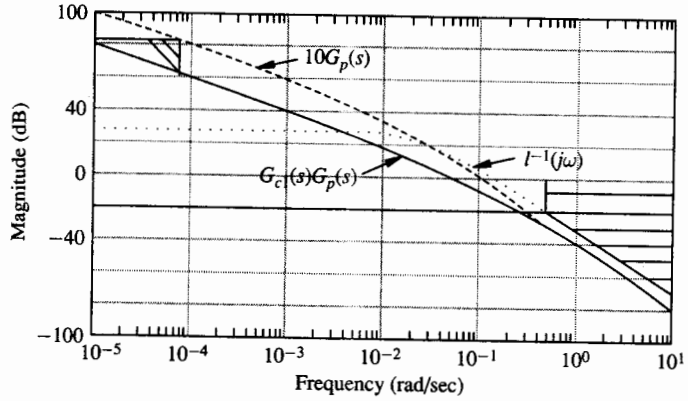
If 60° phase margin is to be achieved, the magnitude of the return difference at the crossover frequency must equal one. As this seems to be a reasonable objective, we try to get 55° of lead out of the lead compensator. Referring to Table 8.5-1, a value of $m = 10$ is needed for 55° phase lead at the frequency of maximum lead. The new crossover frequency coincides with ω_l if ω_l is chosen as the frequency where the old loop gain has a magnitude of $+10 \log m$ dB. From Fig. 8.7-3 we observe that the frequency where $10G_p(s)$ passes through $10 \log m = 10$ dB is $\omega = 0.06$. Therefore, we choose $m = 10$ and $\omega_l = 0.06$ for our lead compensator. The lead compensator is computed using Eqs. (8.5-6), and (8.5-2).

$$G_{\text{lead}}(s) = \frac{s + 0.019}{s + 0.19} \tag{8.7-16}$$

$$G_{c1}(s) = \frac{10(s + 0.019)}{s + 0.19} \tag{8.7-17}$$

The Bode plots of the old loop gain and the new loop gain are given in Fig. 8.7-4. As computed earlier, the new crossover frequency is $\omega_l = 0.06$ with 65° phase margin. The high frequency gain is unaffected while the response at low frequency has been attenuated. The next step is to append a lag compensator to the compensator to raise the low frequency gain while leaving the plots at and above crossover frequency nearly unaffected.

Before designing the needed lag compensator, let's examine the lead compensator from the root locus point of view. The root locus provides intuitively pleasing information about a system's transient response but little information about other performance and robustness qualities. The root locus of the plant with a lead compensator is shown in Fig. 8.7-5. The plant poles at $s = 0$ and $s = -0.01$ are shown along with the lead compensator's zero at $s = -0.019$ and pole at $s = -0.19$. The plant's pole at $s = -10$ is not shown as it moves very little. The basic concept of a lead compensator in the root locus setting is the placement of a zero in the left half-plane to the left of the dominant poles and the placement of a pole much further out in the left half-plane to provide realizability. The distant pole should have very little effect



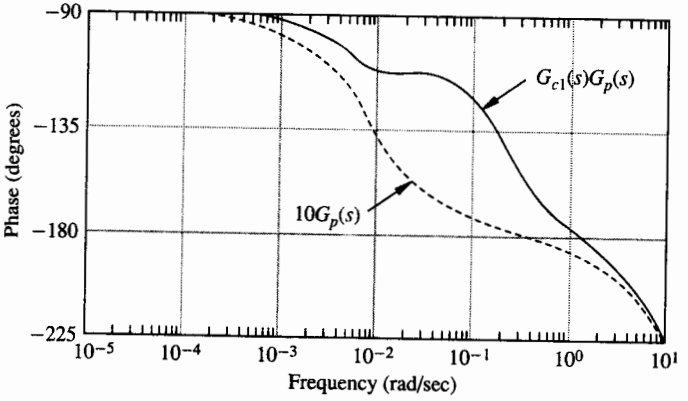


FIGURE 8.7-4
Bode plots of $10G_p(s)$ and $G_{c1}(s)G_p(s)$. (*a*) Magnitude; (*b*) phase.

on the root locus. The zero has the effect of pulling the dominant poles towards itself and further into the left half-plane. The result is a faster and less oscillatory response as the dominant poles are pulled away from the imaginary axis.

The closed-loop poles for the correct gain of the compensator of Eq. (6.7-17) are marked with stars on Fig. 8.7-5. Since the system is in the G configuration the root locus plane coincides with the closed-loop response plane. It is interesting to observe that there is a zero closer to the origin than the first closed-loop pole. We know from Chap. 4 that the step response of such a system will have some overshoot even though there are no complex pole pairs. This problem is easily corrected with a prefilter as will be seen later.

We now design the lag compensator to meet the low frequency performance constraint. The key part of the lag compensator design is to avoid causing problems at the crossover frequency. Since the phase margin of the lead compensated system is 65°, we can afford to place the zero of the lag compensator only a decade below the

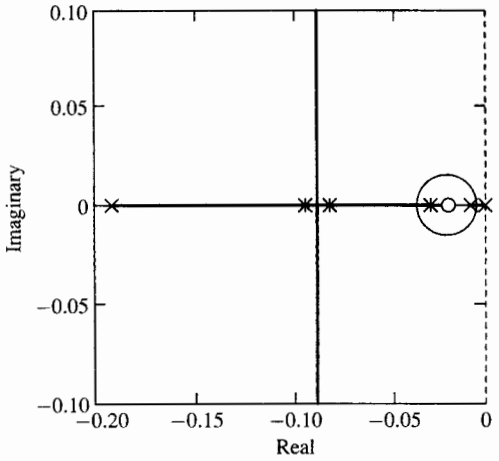


FIGURE 8.7-5
Root locus of $G_{c1}(s)G_p(s)$.

crossover frequency, creating 6° phase lag at the crossover frequency. The equation for the lag compensator is

$$G_{\text{lag}}(s) = \frac{s + 0.005}{s} \tag{8.7-18}$$

The compensator is now

$$G_{c2}(s) = \frac{10(s + 0.005)}{s} \frac{(s + 0.019)}{(s + 0.19)} \tag{8.7-19}$$

The Bode plots of the loop gains using compensators $G_{c1}(s)$ of Eq. (8.7-17) and $G_{c2}(s)$ of Eq. (8.7-19) are given in Fig. 8.7-6. It can be seen that the lag compensator raises the low frequency gain while leaving the frequencies higher than the crossover frequency unaffected.

The final loop gain using $G_{c2}(s)$ as a compensator appears to be acceptable. Both high and low frequency objectives are met with enough margin to expect adequately robust performance. The phase margin of 60° also indicates adequately robust performance.

It is interesting that, since the lag compensators are designed to have very little effect on a system's transient response, lag compensators have very little effect on closed-loop pole positions on a root locus plot. The root locus plot of the poles and zeros close to the origin for the plant with the lead-lag compensator $G_{c2}(s)$ is shown in Fig. 8.7-7. The insert in the figure gives a magnified view of the behavior of the poles and zeros that are very near the origin. The second pole at the origin and the low frequency zero from the lag compensator are so close to each other that they have very little effect on the root locus except for points very close to the origin. The resulting closed-loop pole near the low frequency zero from the lag compensator is effectively canceled in the closed-loop transfer function since the series compensator zeros appear in the closed-loop transfer function.

It is time now to visually check the sensitivity function and the complementary sensitivity function. In addition, the closed-loop step response should be checked.

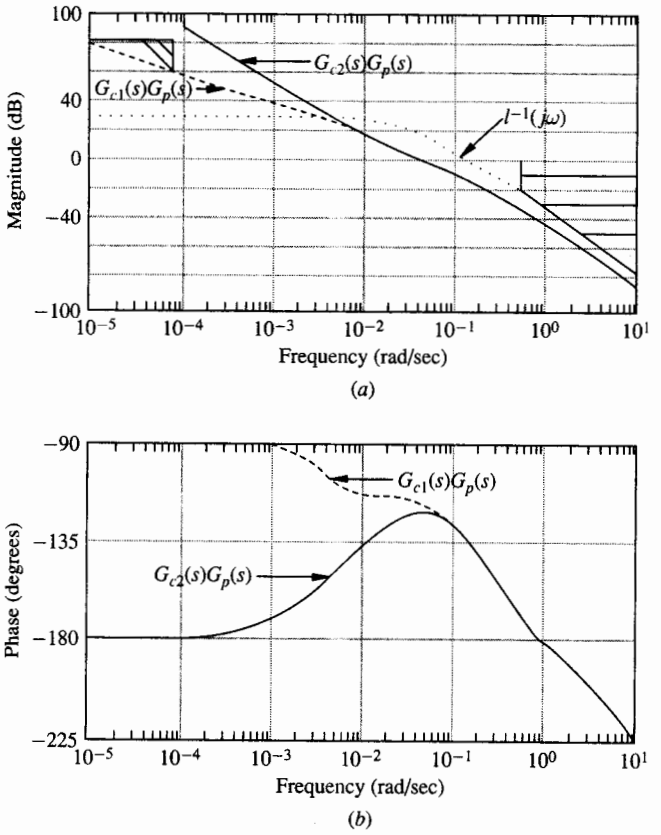


FIGURE 8.7-6
Bode plots of $G_{c2}(s)G_p(s)$. (*a*) Magnitude; (*b*) phase.

The plots of Fig. 8.7-8 show the sensitivity function and the complementary sensitivity function for the loop with the controller $G_{c2}(s)$. These plots can be used to check the robust performance condition of Eq. (8.2-8). First note that the specific low frequency disturbance rejection requirement is easily met as shown on the plot of the sensitivity function, which ascends from very small values at low frequency to a maximum value of about 2 dB at $\omega = 0.1$ and returns toward unity at high frequencies. The magnitude of the complementary sensitivity function is near unity for low frequencies, has a peak of about 1 dB at $\omega = 0.03$ and falls off rapidly, remaining at least 10 dB below the $l_m^{-1}(j\omega)$ curve. This 10 dB gap indicates that not only does the system remain stable in the face of all allowable perturbations but also maintains reasonable performance.

The sensitivity function that results when the system is confronted with the worst case allowable perturbation is obtained using Eq. (6.7-16). In Fig. 8.7-9 the nominal sensitivity function and the worst case sensitivity function are plotted. Even when the worst case perturbation is used the maximal sensitivity is about 4 dB = 1.6. This indicates that the system performs robustly for the mismodeling considered.

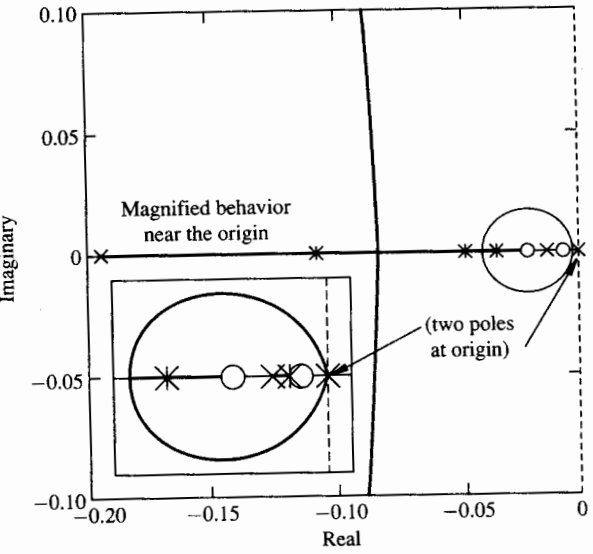


FIGURE 8.7-7
Root locus of $G_{c2}(s)G_p(s)$.

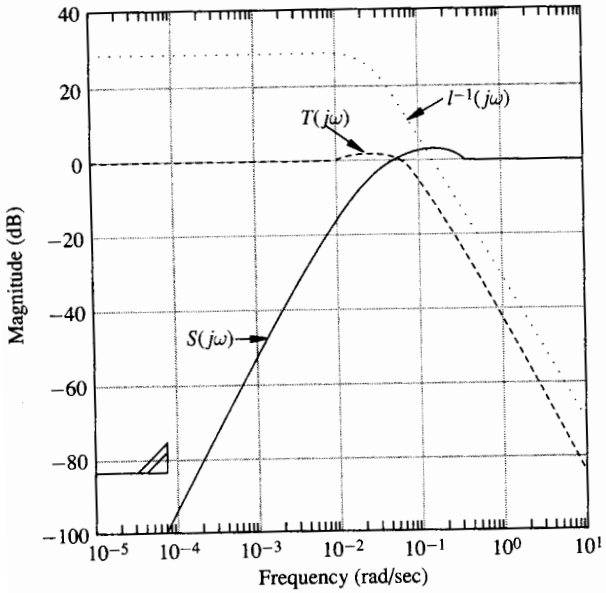


FIGURE 8.7-8
Sensitivity function $S(j\omega)$ and complementary sensitivity function $T(j\omega)$ plots for loop gain $G_{c2}(s)G_p(s)$.

The last measure of the design to check is the closed-loop response to a step change in command angle θ_c. This step response is plotted in Fig. 8.7-10. Despite no indications of a complex pole pair from the flat closed-loop frequency response curve (i.e., the complementary sensitivity function curve) of Fig. 8.7-9, the closed-loop step response displays some overshoot. The cause of the overshoot was discussed in examining the root locus of the lead compensator. The overshoot arises from the

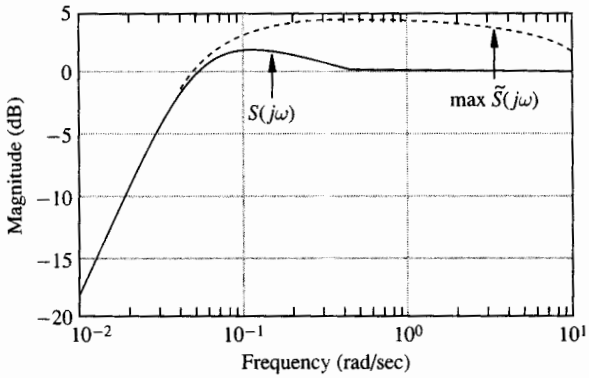


FIGURE 8.7-9
Nominal sensitivity function $S(j\omega)$ and worst case perturbed sensitivity function max $\tilde{S}(j\omega)$.

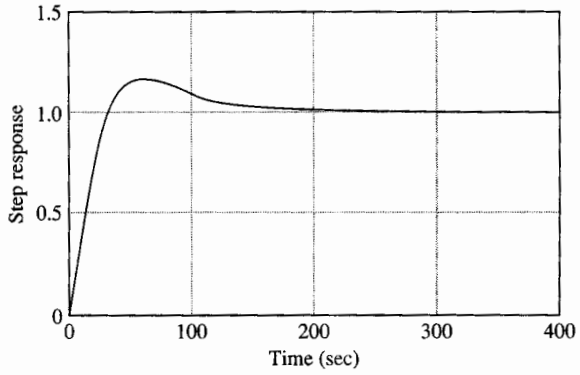


FIGURE 8.7-10
Step response of the closed-loop system.

fact that the zero of the lead compensator remains closer to the origin than does the nearest closed-loop pole. This often happens with lead compensators. The cause of the overshoot is more clearly shown by examining the closed-loop transfer function

$$M(s) = 0.1\frac{(s+0.005)(s+0.019)}{(s+0.00517)(s+0.033)(s+0.05)(s+0.11)(s+10)} \tag{8.7-20}$$

The system has a pole-zero pair from the lag compensator that essentially cancels at $s = -0.005$. The next element is a zero closer to the origin than is the first dominant pole. This is the cause of the overshoot. Since the zero is part of the compensator, it does not appear in the transfer function from either an input disturbance or an output disturbance. Since the overshoot affects only the closed-loop transient response, it is easily remedied with a prefilter that cancels the offending zero.

$$P(s) = \frac{0.019}{s+0.019}$$

The resulting control configuration, given in Fig. 8.7-11, uses block diagram manipulation to separate the sensor dynamics from the plant. The final step response, in Fig. 8.7-12, shows smooth response. The time constant of approximately 20 sec agrees

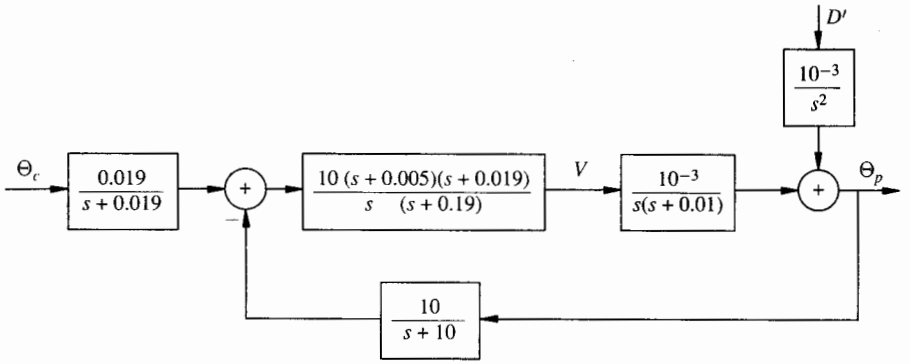


FIGURE 8.7-11
Final control configuration.

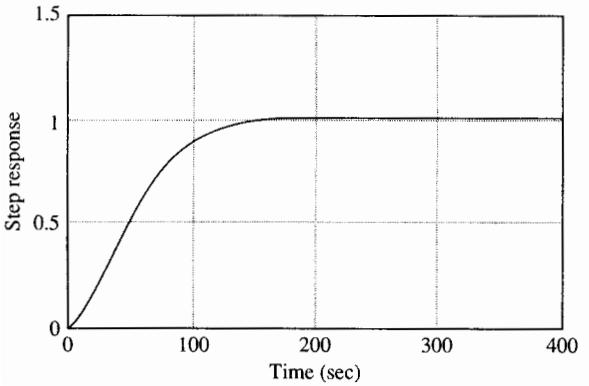


FIGURE 8.7-12
Step response of closed-loop system with prefilter.

with the loop gain bandwidth of 0.05 rad/sec, as shown in Fig. 8.7-6. The response can be made faster by further manipulation of the prefilter.

The final design of the ATS-6 was indeed quite similar to the design specified here. The control system was very successful in its mission.

Exercises 8.7

8.7-1. Using the equations given in Fig. 8.7-1*b* (*a*) Find the transfer function from V to T; (*b*) Find how this transfer function reduces for frequencies where $L\omega \ll R$. Hint: Use block diagrams.

Answers: (*a*) $\dfrac{K_\tau Is}{(Ls+R)(Is+C)+K_v K_\tau}$ (*b*) $\dfrac{K_\tau Is}{RIs+RC+K_v K_\tau}$

8.7-2. It is interesting to examine how the satellite control system might fare if it contained unmodeled dynamics that satisfy the magnitude bound $l_m(j\omega)$. According to the previous analysis, stability should be maintained and performance should be just partially degraded. Suppose that the ATS-6 satellite does indeed contain a lightly damped bending mode with a resonance at $\omega = 1$. Suppose that a better model for the plant is given

by

$$G_p'(s) = G_p(s)G_u(s)$$

with $G_p(s)$ given by the transfer function in Eq. (8.7-3) and

$$G_u(s) = \frac{(s+1)^2}{\left(s^2+0.04s+1\right)}$$

Solve for the particular $L_m(s)$ that results from this $G_u(s)$

$$L_m(s) = G_u(s) - 1$$

Plot the magnitude of $L_m(s)$ and show that this particular $L_m(s)$ satisfies the magnitude bound of Eq. (8.4-5*b*). Plot the Bode plot of the loop gain, $G_{c2}(s)G_p'(s)$ when the perturbed plant $G_p'(s)$ and the compensator $G_{c2}(s)$ are together in the loop. Plot the response to a step change in the reference input. Comment on the resulting frequency and time plots.

8.8 AN EXAMPLE USING ROLL-OFF AND A NOTCH FILTER TO CANCEL PLANT DYNAMICS

Often in the design of control systems for mechanical structures the plant model contains a high frequency, poorly damped pair of poles. If this pole pair can be modeled accurately enough it need not limit the bandwidth of the control system.

Let's create a hypothetical twist on the design problem of the previous section. Suppose that when the controller design was submitted, the customer felt that a larger bandwidth is needed so that the system could react more quickly to disturbances. (If the customer's only reason for wanting increased bandwidth were the desire to follow command inputs more closely, this problem could have been resolved by the use of a prefilter. However, prefilters have no effect on disturbance rejection properties.)

If we were using the plant model of Eq. (8.7-4) with the uncertainty of Eq. (8.7-7) we would have to tell our customer, that we could increase the bandwidth only by an insignificant amount. Referring to Fig. 8.7-6, we'd explain that the loop gain must be down significantly below −20 dB by $\omega = 0.7$ to account for modeling uncertainty. If we push the bandwidth out of its current position of $\omega = 0.06$ the gain would have to roll off more sharply. This increased slope would indicate additional phase lag, decreased phase margin and oscillatory transient responses to disturbances. If, however, the system could be modeled better so that the uncertainty wouldn't become large until higher frequencies were reached, then the bandwidth could be increased and still roll off soon enough to meet the high frequency uncertainty constraint.

Let's suppose that the customer responds by giving our modeling department a contract to determine a better model of the satellite. The modeling department returns the model

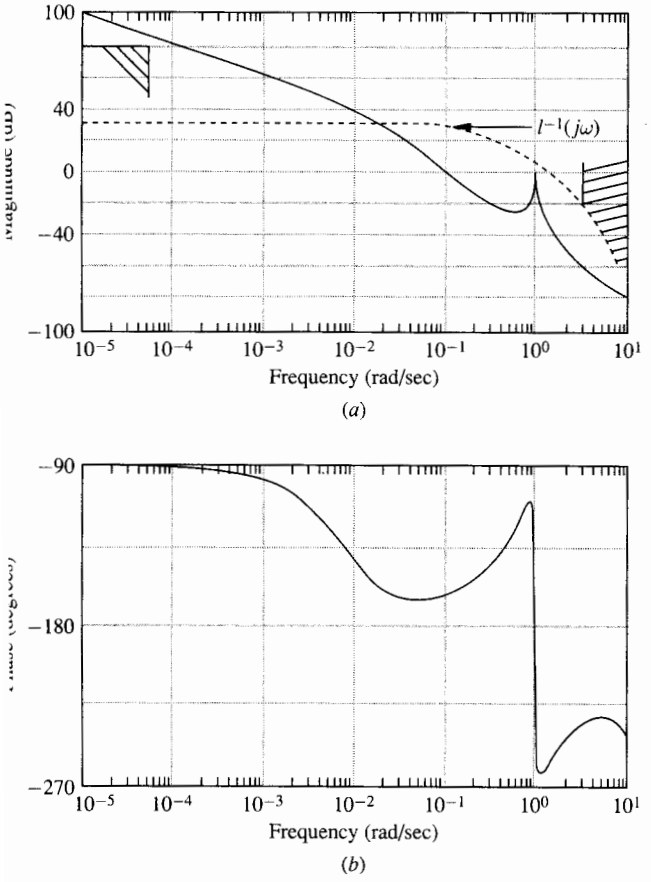
$$\tilde{G}_{p2}(s) = G_{p2}(s)(1 + L_m(s)) \qquad (8.8\text{-}1a)$$

$$G_{p2}(s) = \frac{10}{10^3 s(s+0.01)(s+10)} \qquad \frac{(s+1)^2}{s^2+0.04s+1} \qquad (8.8\text{-}1b)$$

with

$$|L_m(j\omega)| \le l_m(j\omega) = \frac{\left(1+\dfrac{\omega^2}{0.09}\right)(1+\omega^2)}{30\left(1+\dfrac{\omega^2}{100}\right)} \tag{8.8-1c}$$

Figure 8.8-1 contains the Bode plots of $10G_{p2}(s)$ and a magnitude plot of $l_m^{-1}(j\omega)$. The Bode plots show the loop gain meeting the high and low frequency constraints with a crossover frequency of $\omega_c = 0.1$ but a poor phase margin of 15°. If we try to add a lead compensator, either the low frequency constraint will be violated or the high frequency constraint will be approached too closely. We begin a set of new design iterations with the intent of meeting the low frequency constraint, keeping about 10 dB or more below the high frequency constraint, while achieving as high a crossover frequency as possible with a 60° phase margin.



IGURE 8.8-1
ode plot of $10G_{p2}(s)$. (*a*) Magnitude; (*b*) phase.

Look what happens if the loop gain is raised by another factor of 10. The Bode magnitude plot of Fig. 8.8-1 is raised by 20 dB. While the first 0 dB crossover point is extended to $\omega = 0.4$, the peak at $\omega = 1$ is driven above the 0 dB line. A sketch of the Nyquist plot of $100G_{p2}(s)$ is shown in Fig. 8.8-2. The main plot is the image of the entire Nyquist D-contour including positive frequencies, negative frequencies and the small semicircle that runs counterclockwise around the pole at the origin. The sketch is not drawn to scale. The insert shows the computer-generated polar plot of $100G_{p2}(j\omega)$ for positive ω only as ω goes from $\omega = 0.01$ to infinity. The loop gain of $100G_{p2}(s)$ gives rise to an unstable closed-loop system as there are two clockwise encirclements of the -1 point caused by the peak in the Bode plot at $\omega = 1$.

One strategy that can be used to avoid this unpleasant situation is to remove the peak with a notch filter. The notch is to be centered at $\omega = 1$ where the peak is located. It should have a depth of -30 dB. (Refer to Sec. 8.6 where the parameters of a notch filter are discussed.) The depth of notch is given by $2\zeta_N$. We choose

$$2\zeta_N = 0.0316 = -30 \text{ dB}$$

$$\zeta_N = 0.0158$$

The notch filter is given by

$$G_{\text{notch}}(s) = \frac{s^2 + 0.0316s + 1}{s^2 + s + 1} \tag{8.8-2}$$

The Bode plots of the loop gain $100G_{p2}(s)G_{\text{notch}}(s)$ are given in Fig. 8.8-3. The peak in magnitude is removed along with the associated phase shift. Note that the elimination of the peak is possible only when the parameters of the oscillatory pole pair are precisely modeled. If the frequency of the notch is slightly off from the frequency of

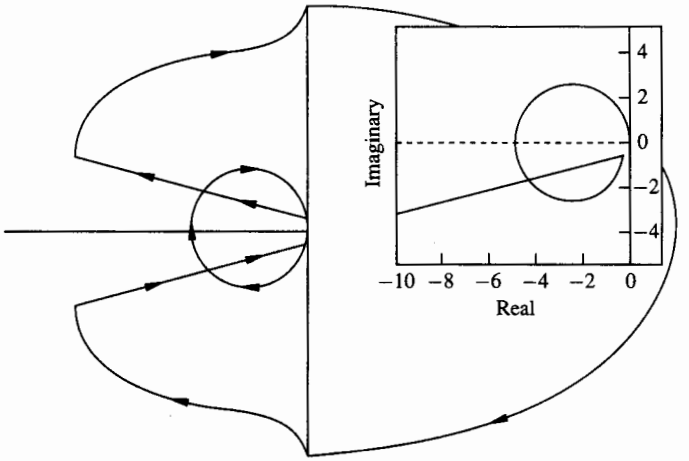
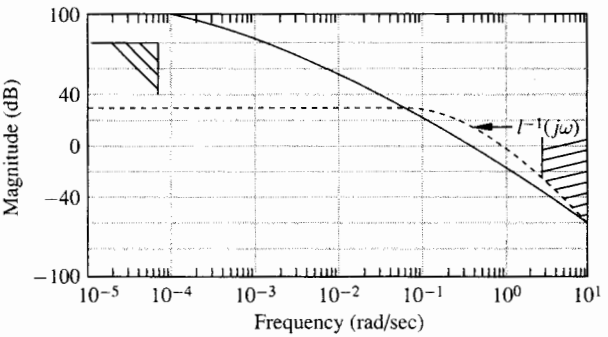
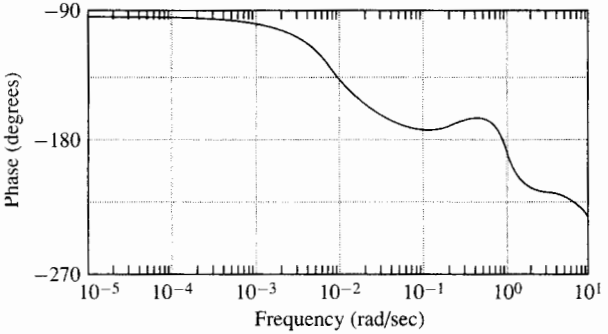


FIGURE 8.8-2
Nyquist of plot of $100G_{p2}(s)$.



(*a*)



(*b*)

FIGURE 8.8-3
Bode plots of $100G_{p2}(s)G_{\text{notch}}(s)$.
(*a*) Magnitude; (*b*) phase.

the oscillatory plant poles the resulting Bode plot would have a peak next to a notch and the closed-loop system would be unstable. The major effect of the notch filter is the near cancellation of the plant poles located at $s = -0.02 \pm j0.9996$ by the compensator zeros located at $s = -0.0158 \pm j0.9997$. The near cancellation makes the poles almost disappear from the loop gain. Of course, as discussed in Sec. 8.6, canceling oscillatory plant poles leaves the possibility of an oscillatory response to input disturbances in the closed-loop system.

Two problems remain in the loop gain of Fig. 8.8-3. High frequency attenuation is needed and the phase margin should be increased. The high frequency attenuation is addressed first using poles for roll-off. To provide a 10 dB pad on the high frequency constraint, the gain should be reduced by 20 dB at $\omega = 10$. This should be accomplished while producing as little phase lag near the crossover frequency as possible. The poles should be placed as far out the frequency axis as possible to minimize the phase lag at the crossover frequency; however, the poles must provide the required attenuation of 20 dB by the point $\omega = 10$. A pole at $\omega = 3$ produces only 6° lag near the crossover frequency which is a decade away near $\omega = 0.3$ and provides 10 dB of attenuation by $\omega = 10$. The needed roll-off is provided by two poles at $\omega = 3$.

$$G_{\text{roll}}(s) = \frac{9}{(s+3)^2} \tag{8.8-3}$$

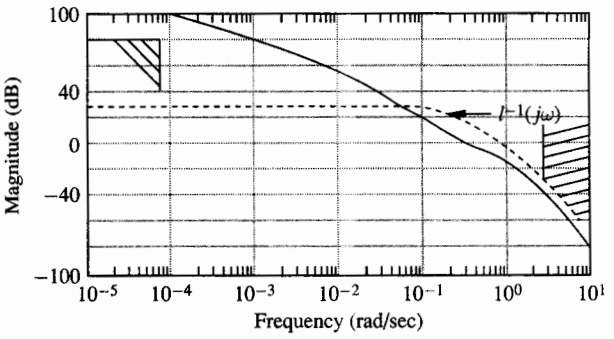
The Bode plots of the new loop gain, $100G_{p2}(s)G_{\text{notch}}(s)G_{\text{roll}}(s)$, are shown in Fig. 8.8-4. The loop gain now meets the high frequency constraint with a 10 dB margin as specified.

The last step in the design is to use a lead compensator to increase the phase margin. Since it is desirable to maintain the margin from the high frequency constraint, some of the bandwidth gained previously is sacrificed and a unity high frequency gain lead compensator is used. The loop gain of Fig. 8.8-4 has about 5° phase margin. A lead compensator that produces 55° maximum phase lead is used. The information given in Sec. 8.4 shows that $m = 10$ is appropriate. Since the gain from the lead compensator at the frequency of maximum phase lead is $-10 \log m = -10$ dB, ω_l is taken as the frequency where the loop gain crosses 10 dB. From Fig. 8.8-4, $\omega_l = 0.2$. The pole of the lead compensator needs to be at $s = -\omega_l\sqrt{m} = -0.2\sqrt{10} = -0.63$ and the zero needs to be at $s = -0.063$.

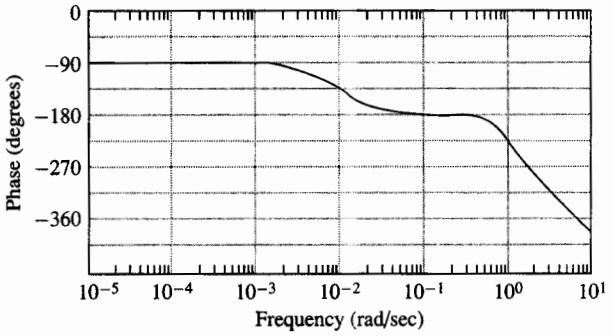
$$G_{\text{lead}}(s) = \frac{s + 0.063}{s + 0.63} \tag{8.8-4}$$

The total compensator is the product of the parts.

$$G_C(s) = \frac{900(s^2 + 0.0316s + 1)(s + 0.063)}{(s^2 + s + 1)(s + 0.63)(s + 3)^2} \tag{8.8-5}$$



(*a*)



(*b*)

FIGURE 8.8-4
Bode plots of $100G_{p2}(s)G_{\text{notch}}(s)G_{\text{roll}}(s)$.
(*a*) Magnitude; (*b*) phase.

The Bode plots of the total loop gain are shown in Fig. 8.8-5. The low frequency constraint is just met and the high frequency constraint is met with a 10 dB margin. The phase margin is 60° and the crossover frequency has been increased to $\omega_c = 0.2$ from the crossover frequency of $\omega_c = 0.06$, which was achieved in Sec. 8.7 with the inferior model.

Figure 8.8-6 displays the magnitude plots of the sensitivity function $S(j\omega)$ and the complementary sensitivity function $T(j\omega)$. The plot shows that the complementary sensitivity function stays at least 10 dB below $l_m^{-1}(j\omega)$ for all ω. This indicates that the system is stable for all plants within the set of plants described by Eq. (8.8-1). The 10 dB margin indicates that every plant in that set produces a loop gain that comes no closer than -10 dB $= 0.31$ to the -1 point on the Nyquist plot. Reasonable performance is thus maintained for all plants within the set. The sensitivity function shows that the low frequency performance specification is met. In addition the maximum nominal sensitivity is about 1.2. At frequencies lower than the crossover frequency the sensitivity function is much less than one and disturbances and the effects of changes in parameters are greatly attenuated. The design needs a prefilter to clean up the response to reference inputs just as the design of Sec. 8.7 did.

Can the design of the loop gain be improved? Yes, the low frequency gain could be raised with a lag compensator with little or no effect on the intermediate or high

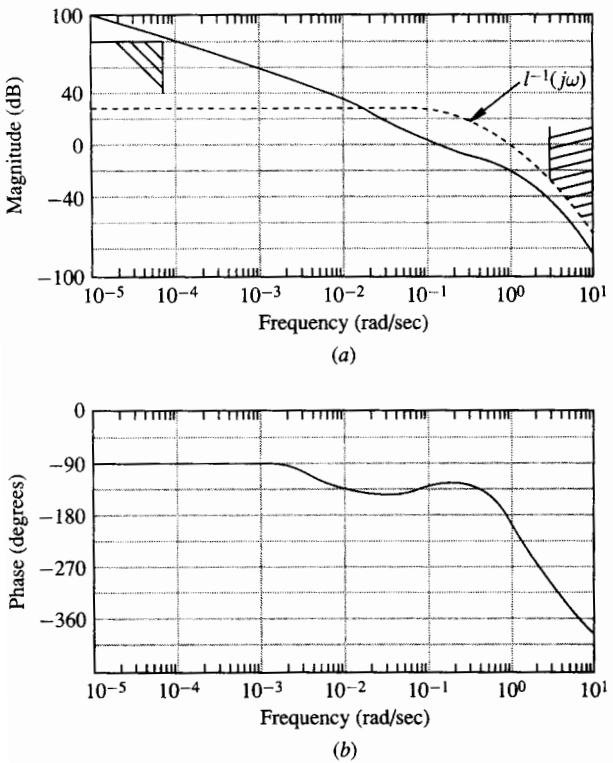


FIGURE 8.8-5
Bode plots of $G_c(s)G_{p2}(s)$. (*a*) Magnitude; (*b*) phase.

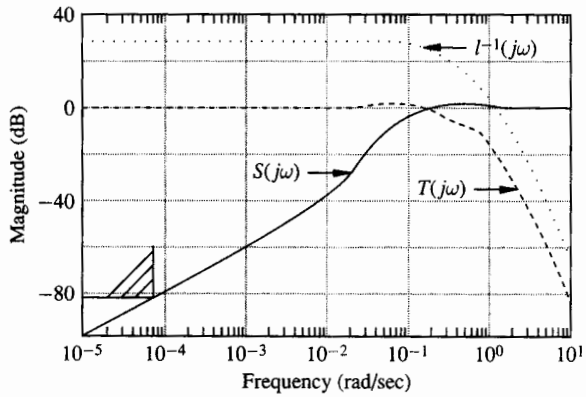


FIGURE 8.8-6
Plot of the sensitivity function $s(j\omega)$ and the complementary sensitivity function $T(j\omega)$.

frequency behavior. Can the bandwidth be further increased? The answer is probably yes. Almost any design measure can be improved somewhat with more careful and more complex designs, but a level of diminishing returns is always met. In this design, the bandwidth cannot be increased much. The loop gain must meet the high frequency constraint arising from the modeling uncertainty. If the bandwidth is pushed out much further the magnitude must slope down at a sharper angle. This indicates phase lag at lower frequencies and destroys the attainable phase margin. In general, the inability to model a plant well at high frequencies limits the bandwidth attainable in a control design. In a later section we see that delays in the plant dynamics may also limit the attainable bandwidth.

It is interesting to see how the new control design responds to disturbances. Let's compare the controller designed in this section with the controller designed in Sec. 8.4 when placed in a loop with the plant of Eq. 8.8-1*b* and excited with a step disturbance at the output, i.e., let $D_o(s)$ in Fig. 8.8-7 be a unit step. The two outputs are displayed in Fig. 8.8-8.

First, notice that the increased bandwidth of the design of this section has indeed resulted in a faster correction of the output from its disturbed value. Notice also that there is an oscillation at a rate of 1 rad/sec in the response of the Section 8.7 control system that doesn't exist in the response of the revised control system. The step disturbance contains some amount of signal of all frequencies including 1 rad/sec. The controller of Section 8.7 feeds back some of that frequency and excites the resonant mode in the plant. The zeros from the notch filter in the controller of this section block signals with a frequency of 1 rad/sec from reaching the plant and exciting the resonant mode. Another way to see this is to write the transfer function from the output disturbance $D_o(s)$ to the output $Y(s)$ for the controller of this section. The

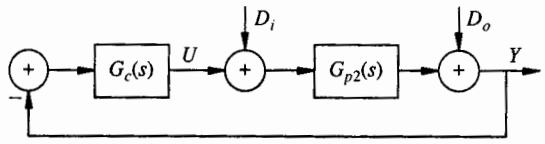


FIGURE 8.8-7
Loop with output disturbance.

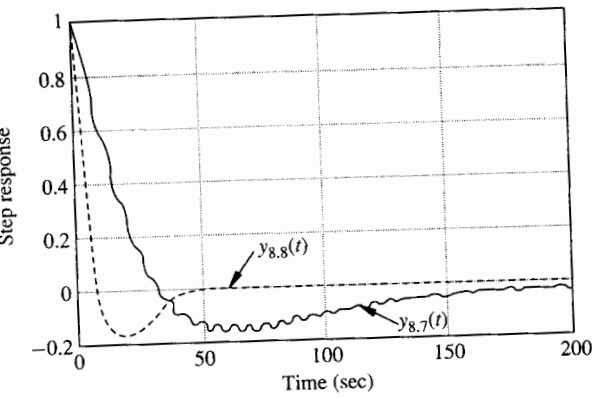


FIGURE 8.8-8
The responses of the control loop of Sec. 8.7, $y_{8.7}(t)$, and Sec. 8.8, $y_{8.8}(t)$, to a unit step output disturbance.

poles of this transfer function are the closed-loop poles, and the zeros of this transfer function are the poles of the loop gain transfer function.

$$\frac{Y(s)}{D_o(s)} = \frac{s(s+0.01)(s+10)(s^2+0.04s+1)}{\left((s+0.1)^2+(0.23)^2\right)(s+10)(s^2+0.0414s+1)} \cdot \frac{(s+3)^2(s+0.634)(s^2+s+1)}{(s^2+0.03s+9.09)(s+0.572)(s^2+0.82s+0.80)}$$

The dominant dynamics here are the zeros at $s = 0$ and $s = -0.01$ and the pole pair at $s = 0.1 + j0.23$. All other poles and zeros almost cancel, including the resonant poles represented by $s^2 + 0.0414s + 1$ and the zeros represented by $s^2 + 0.04s + 1$. Therefore no oscillation appears in the output.

We should remember, however, that for the satellite system under study, disturbances are more likely to enter the system at the input to the plant. The disturbance comes in as a force on the satellite. It excites the plant dynamics. The plot of Fig. 8.8-9 shows the output in response to an impulse disturbance at the plant input, i.e., an impulse in $D_i(s)$ as shown in Figure 8.8-7. The impulse disturbance excites the bending

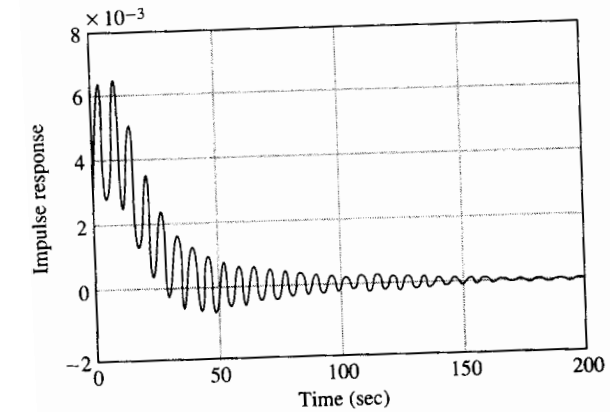


FIGURE 8.8-9
The response of the controller of Sec. 8.8 to a unit impulse input disturbance.

mode at $\omega = 1$. Since the loop gain does not have a large gain at this frequency, the controller cannot take much action to counteract the oscillations. They must play themselves out. The control that comes from canceling plant poles with zeros in the controller is a *passive* control of the oscillations. If a disturbance does not directly excite the oscillatory mode, the notch in the controller will ensure that the actions of the feedback control law also do not excite the oscillations. If the oscillations are directly excited, the controller does not allow enough signal into the plant input at the frequency of the oscillation to produce corrective action. This situation is in contrast to the *active* control that takes place with a large loop gain. At frequencies where the loop gain is large, if oscillations are excited, the oscillating signals are fed back around the loop and into the input correctly shifted in phase so that they can counteract the oscillations. In this way, the oscillation are eliminated very quickly. (The student may want to review the results of Exercise 8.6-2 in light of this discussion on active versus passive control.)

In the design in this section, large gain at the frequency of oscillation is not a feasible option since the uncertainty in the model limits the bandwidth. While canceling oscillatory plant dynamics is not something to be done carelessly, sometimes it is the best option available. In this case, the notch filter allows the bandwidth to be increased while restricting oscillations in response to reference inputs and output disturbances. No design can remove the oscillation from input disturbances since the loop gain at the resonant frequency $\omega = 1$ cannot be made larger while retaining the required robust stability properties.

Exercises 8.8

8.8-1. The goal of the roll-off poles of Eq. (8.8-3) is to provide 20 dB of attenuation by $\omega = 10$. It was decided to place two poles at $\omega = 3$. This produces 12° of phase lag at the original crossover frequency, $\omega = 0.3$.
(*a*) Where must a single roll-off pole be placed to provide 20 dB of attenuation at $\omega = 10$? How much phase lag is produced by this pole at $\omega = 0.3$?
(*b*) Where should a three pole roll-off be placed to produce 20 dB of attenuation at $\omega = 10$? How much phase lag do these poles produce at $\omega = 0.3$?

8.8-2. (*CAD Exercise*) Find the response of the controller designed in Sec. 8.7 when used with the plant of Eq. (8.8-1*b*) and excited with a unit impulse $D_i(s)$ at the input to the plant. Compare this response with that of Fig. 8.8-9 and explain the differences.

8.9 CONTROLLING UNSTABLE PLANTS

So far we have demonstrated control designs that improve the performance of systems containing only stable plants. One of the major advantages of designing closed-loop control systems is that an unstable plant can be stabilized and also meet performance criteria. In this section we consider control system designs starting with plants that contain poles in the right half-plane, i.e., unstable plants. It is interesting that the key technique used to stabilize loops with an unstable plant is a lead compensator—the same lead compensator used to achieve an improved margin of stability in loops with stable plants. Indeed, there is little difference between improving the performance

of a stable loop gain and creating acceptable performance with an initially unstable loop gain except for the urgency inherent in stabilizing an otherwise unstable system.

Let's compare a controller placed in series with two different plants, one stable and one unstable. Consider the stable plant

$$G_{p1}(s) = \frac{100}{(s+0.1)(s+1)(s+10)} \tag{8.9-1}$$

and the unstable plant

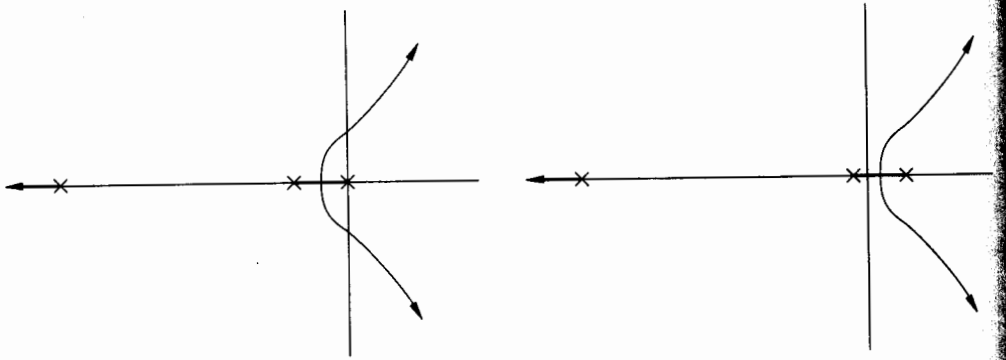
$$G_{p2}(s) = \frac{100}{(s+0.1)(s-1)(s+10)} \tag{8.9-2}$$

If these plants are placed in feedback loops with constant gain compensators the two root loci of Fig. 8.9-1 result. The two loci are very similar except for the location of the $j\omega$ axis, which indicates that the loop with the stable plant remains stable for small gains while the loop with the unstable plant is unstable for all constant gain compensators.

We know that the zero of a lead compensator *pulls* the poles of the plant of Eq. (8.9-1) further into the left half-plane and allows a higher gain while retaining stability. The same effect occurs when a lead compensator is used in series with the unstable plant of Eq. (8.9-2). Consider the lead compensator

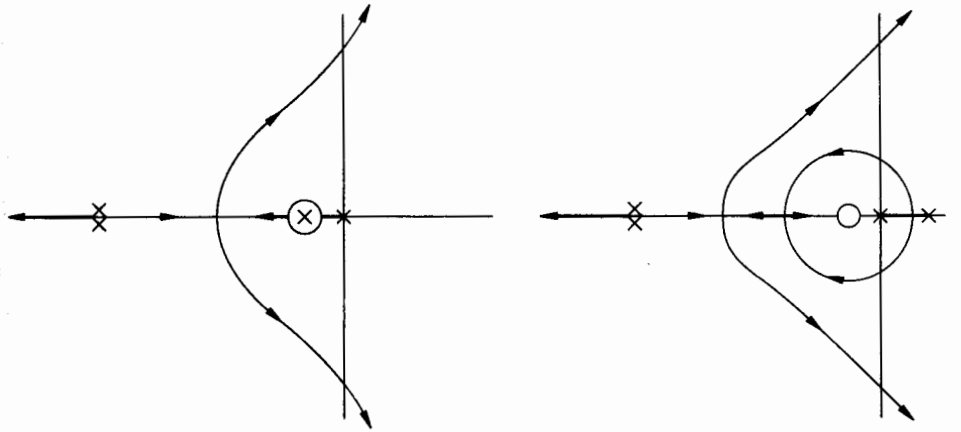
$$G_C(s) = \frac{K(s+1)}{s+10} \tag{8.9-3}$$

The root loci of the combination of this lead compensator in series with each of the plants is shown in Fig. 8.9-2. In Fig. 8.9-2*b*, it can be seen that for intermediate values of the gain K, the closed-loop system is stabilized.



(*a*) (*b*)

FIGURE 8.9-1
Root locus plot. (*a*) Stable plant; (*b*) unstable plant.



(*a*) (*b*)

FIGURE 8.9-2
Plant with lead compensator. (*a*) Stable plant; (*b*) unstable plant.

To further determine the behavior of the system with the unstable plant, the Bode and Nyquist plots of the system must be examined. In the compensator of Eq. (8.9-3), let $K = 3$. We examine the Bode plots of the plant $G_{p2}(s)$ given in Eq. (8.9-2) and the combination of the plant $G_{p2}(s)$ with the series compensator $G_C(s)$ given by Eq. (8.9-3).

$$G_C(s)G_{p2}(s) = \frac{300(s+1)}{(s+0.1)(s-1)(s+10)^2} \tag{8.9-4}$$

The Bode plots are given in Fig. 8.9-3. The usual effect of the lead compensator can be seen in these Bode plots. From the Bode plot, the Nyquist plots can be drawn. The entire Nyquist plot of the unstable plant $G_{p2}(s)$ is sketched (not to scale) in Fig. 8.9-4. The Nyquist plot contains one clockwise encirclement of the -1 point. Since the plant has one right half-plane pole, the Nyquist analysis indicates that a unity feedback compensator in a closed-loop system with this plant produces two unstable closed-loop poles. This observation is in agreement with the root locus of Fig. 8.9-1*b*.

The entire Nyquist plot of $G_C(s)G_{p2}(s)$ is sketched (not to scale) in Fig. 8.9-5. The lead compensator has pulled the portion of the Nyquist plot corresponding to intermediate positive frequency values below the negative real axis on the Nyquist plot. This produces a single counterclockwise encirclement of the -1 point and, because there is one right half-plane loop gain pole, the closed-loop system is stable. Thus, we see with frequency domain analysis that an unstable plant can be stabilized with the addition of a lead compensator. We should note that further compensation may be placed in series with $G_C(s)G_{p2}(s)$ to further improve the performance and robustness of the system. Once the Nyquist plot has the correct number of encirclements, a further lead compensator can pull the plot further away from the -1 point, giving a better phase margin and a less oscillatory response.

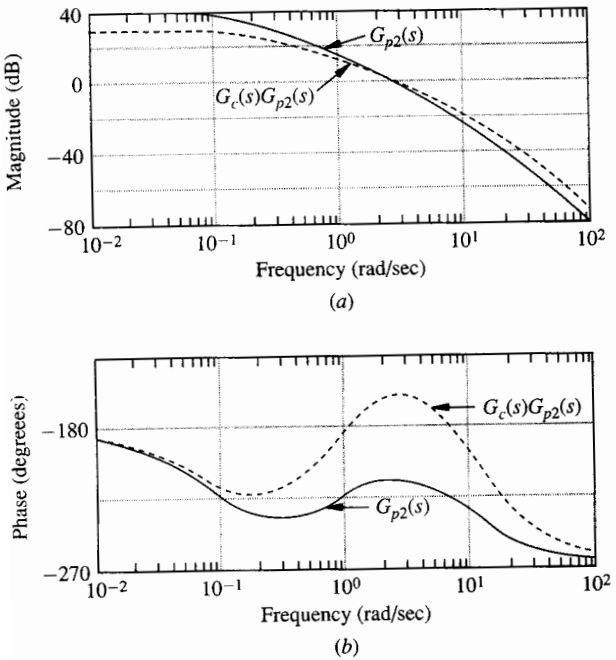


FIGURE 8.9-3
Bode plots $G_{p2}(s)$ and $G_C(s)G_{p2}(s)$. (*a*) Magnitude; (*b*) phase.

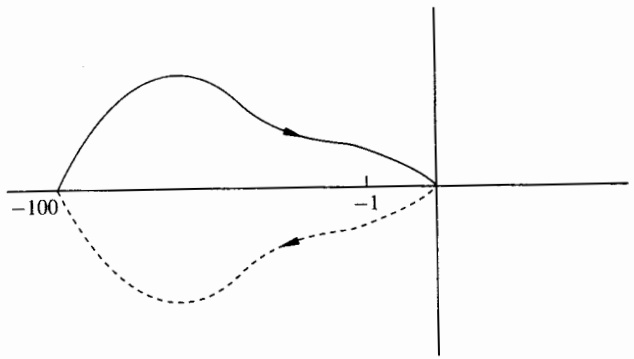


FIGURE 8.9-4
The Nyquist plot of $G_{p2}(s)$.

It is interesting to see the Nyquist plots before and after a plant with two unstable poles is stabilized. Let the plant be given by

$$G_{p3}(s) = \frac{10}{(s+0.1)(s-1)^2} \tag{8.9-5}$$

A sketch of the entire Nyquist plot (not to scale) of $G_{p3}(s)$ given by Eq. (8.9-5) is shown in Fig. 8.9-6. There are no encirclements of the -1 point, indicating two unstable closed-loop poles.

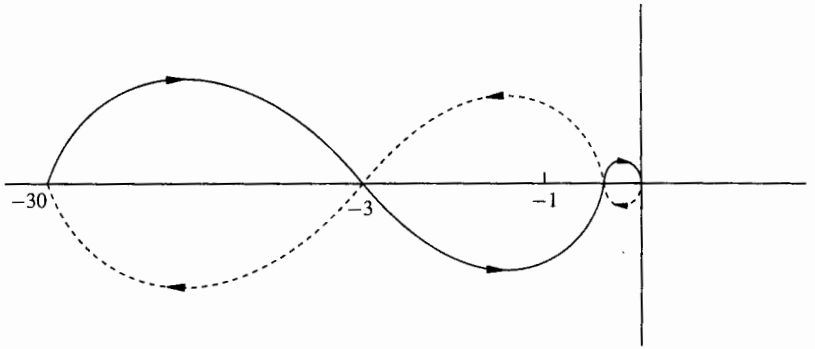


FIGURE 8.9-5
The Nyquist plot of $G_C(s)G_{p2}(s)$

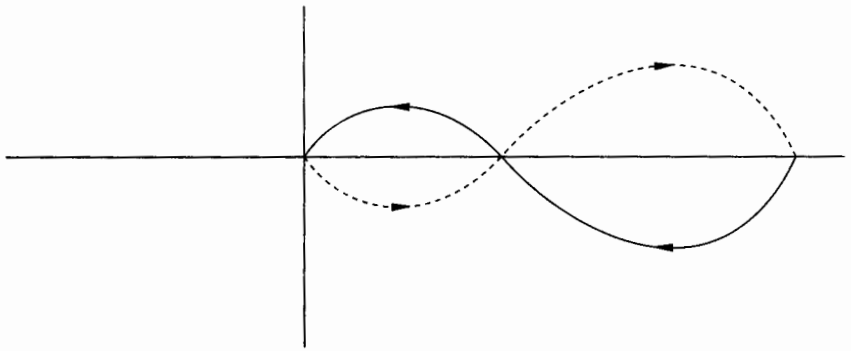


FIGURE 8.9-6
Nyquist plot of $G_{p3}(s)$.

It is not clear how one could possibly create two counterclockwise encirclements starting from this plot. However, if enough phase lead can be created, the Nyquist plot of Fig. 8.9-7 can result. This plot has the required two counterclockwise encirclements to create a stable closed-loop system. The transfer function used to sketch Fig. 8.9-7 is

$$G_{C2}(s)G_{p3}(s) = \frac{6(s+0.1)^2}{(s+10)^2}\,\frac{10}{(s+0.1)(s-1)^2} \tag{8.9-6}$$

The two lead compensators in series generate almost 180° phase lead near $\omega = 1$. This brings the Nyquist plot around the -1 point to create the counterclockwise encirclements.

We have seen that stable closed-loop systems can result from unstable plants by the introduction of phase lead in the loop gain. Two further observations can be made. First, since lead compensators produce only phase lead over a restricted frequency range, the gain of the system must be set so that the loop gain crosses over through unity magnitude when the phase lead is near its greatest value. In both

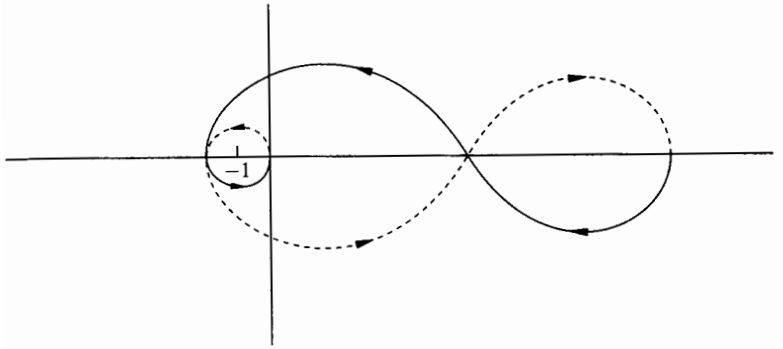


FIGURE 8.9-7
Nyquist plot of $G_{C2}(s)G_{p3}(s)$.

of the examples, $G_C(s)G_{p2}(s)$ and $G_{C2}(s)G_{p3}(s)$, the closed-loop system is unstable if too large or too small a gain is used. Such systems are sometimes referred to as *conditionally stable*.

Second, recall that on a Bode *phase* plot right half-plane poles produce the same effect as left half-plane zeros, i.e., they produce phase lead. In general, so much phase lead is needed to create counterclockwise encirclements on the Nyquist plot that both the lead of a series compensator and the lead produced by unstable poles themselves are needed near the crossover frequency.

There is a subtle indication here. Since the lead contribution from the unstable poles usually must be allowed to develop before the crossover frequency, it is a general rule of thumb that the bandwidth of control systems that stabilize unstable plants cannot be much less than the frequency associated with the highest frequency unstable pole. This relates to the concept of active control introduced in Sec. 8.8. Unstable plant poles must be actively, not passively, controlled. The range of frequencies where the return difference is greater than one is the area of active control. The area of active control should include the frequencies around any unstable pole.

The need for the bandwidth to be large enough to actively control the highest frequency unstable poles means that the frequency range where there is little modeling uncertainty must extend beyond the frequency of the highest frequency unstable pole. If such is the case the loop gain can still be rolled off to meet the magnitude constraint given by the unmodeled dynamics at high frequencies after attaining the necessary bandwidth to stabilize the loop.

Now that we have seen how right half-plane poles affect controller designs, it is natural to investigate how right half-plane zeros affect controller designs. This we do in the next section.

Exercise 8.9

8.9-1. Given the unstable plant

$$G_p(s) = \frac{100}{s(s-1)(s+10)}$$

design a stabilizing series compensator G_C which achieves 40° phase margin. Include along with the transfer function of your compensator the Bode plots of the plant and loop gain (i.e., combination plant and compensator) and answer the following questions:

(*a*) What is the bandwidth of the loop gain transfer function and how does it compare with the location of the unstable pole?

(*b*) What advantage is there in having the bandwidth near the point of maximum phase lead of a plant?

8.9-2. Using the plant of Exercise 8.9-1, design a stabilizing series compensator G_C such that the bandwidth of the loop gain is less than the unstable pole location (i.e., G_CG_p crosses through 0 dB well below 1 rad/sec).

In this design, you cannot use much of the phase lead available from the pole. How does this constraint change your compensator?

8.10 CONTROLLING PLANTS WITH RIGHT HALF-PLANE ZEROS

The presence of right half-plane zeros in a plant places some natural limitations upon what performance can be achieved with a control system. In this section we explore some of these limitations. A zero in the right half-plane means that the plant is a non-minimum phase plant as defined in Chap. 5. A right half-plane zero is often referred to as a *non-minimum phase zero*.

Picture the root locus of a loop gain transfer function that has a zero in the right half-plane. We know that as the gain constant of a root locus is increased, the closed-loop poles move either towards the loop gain zeros or towards infinity. If the loop gain has a right half-plane zero, a closed-loop pole must approach that zero and create an unstable closed-loop pole if the gain constant is made too large. Thus even in this simple analysis, one can see that a right half-plane zero limits the control gain that can be used.

Another way of exploring the limitations imposed by a non-minimum phase zero in the plant is to realize that a right half-plane zero is the one element in a plant model that cannot be effectively eliminated in the closed-loop design, even under ideal conditions. We learned in Chap. 3 that a designer can place closed-loop poles in desired locations using either state-variable feedback or solving a Diophantine equation to create a series or feedback compensator. The only restrictions on the pole to be moved is that it must not be canceled by a zero in the plant. Also, we have seen that left half-plane zeros can be canceled either directly in the loop gain or by placing a closed-loop pole on top of the left half-plane zero. However, a right half-plane zero cannot be moved or canceled. It is the one plant element that must appear in both the loop gain and the closed-loop transfer function.

Since the final loop gain must be zero at the location of a right half-plane zero, the type of return difference transfer function and sensitivity function that can be achieved are restricted by the fact that each of these functions must equal one when evaluated at a right half-plane plant zero. Arguments from the mathematical theory of complex variables can be used with this fact to show that, when the plant contains a right half-plane zero, the sensitivity function evaluated along the $j\omega$ axis must meet certain conditions. These conditions indicate that if the sensitivity is made

less than one at some frequencies it must become significantly larger than one at other frequencies. The closer the right half-plane zero is to the $j\omega$ axis, the more severe are the tradeoff restrictions. Also, if a right half-plane zero is close to a right half-plane pole, the tradeoff restrictions indicate that the sensitivity function must be much larger than one for some frequencies. Of course, many of the goals of designing closed-loop control systems can be translated into the objective of achieving a large return difference and thus a small sensitivity function. Right half-plane zeros are natural obstructions to this objective.

Let's examine how a non-minimum phase zero affects any attempts to design an appropriate loop gain for a control system. The objectives of the design effort were explained in Sec. 8.2. A designer would like a large loop gain at small frequencies with a sharp transition to a small loop gain at high frequencies. The sharpness of the transition from large loop gain to small loop gain is limited by the fact that large negative slopes on the magnitude plot are associated with large amounts of phase lag. Large amounts of phase lag cause problems with stability, performance and robustness near the crossover frequency. The control engineer is constantly trying to negate the effect of phase lag. We have seen the importance of phase lead compensators. In the previous section we saw that phase lead is even more important if unstable poles are present in the loop gain.

The effect of a non-minimum phase zero upon the Bode plot of a loop gain is a deadly combination of an increase in magnitude in conjunction with an increase of phase *lag*! Remember that in the transition frequency region, the control engineer is trying to sharply decrease the magnitude while minimizing the phase lag. The non-minimum phase zero works in direct opposition to the control designers objectives and it cannot be eliminated by cancellation. No wonder that plants with non-minimum phase zeros are the most difficult to control!

Notice that the effect of a non-minimum phase zero is minimal if that zero occurs at a frequency that is much higher than the crossover frequency. In the high frequency range, the magnitude of the loop gain is small enough to keep the sensitivity function near one, independent of the phase. Therefore, the extra phase lag from a non-minimum phase zero in this frequency range has little effect.

Let's look at an example which shows the difficulty presented by a non-minimum phase zero.

Example 8.10-1. Consider a plant described by the transfer function

$$G_{P1}(s) = G_{P0}(s)G_{AP}(s) \tag{8.10-1}$$

where

$$G_{P0}(s) = \frac{0.1}{s\left(\frac{s}{0.01}+1\right)\left(\frac{s}{10}+1\right)} \tag{8.10-2}$$

and

$$G_{AP}(s) = \frac{(-s+0.01)}{(s+0.01)} = \frac{\left(\frac{-s}{0.01}+1\right)}{\left(\frac{s}{0.01}+1\right)} \tag{8.10-3}$$

The transfer function $G_{P0}(s)$ is the same as the plant transfer function of the satellite given by Eq. (8.7-4), which was examined in Sec. 8.7. Let's examine $G_{AP}(s)$ and see how it affects the overall transfer function.

The transfer function $G_{AP}(s)$ contains a right half-plane zero with a pole located in the left half-plane at the mirror image of the zero. Notice that the pole and the zero in this arrangement have exactly equal and opposite effects on the Bode magnitude plot. Indeed, the magnitude plot of $G_{AP}(s)$ is equal to unity at all frequencies.

$$|G_{AP}(j\omega)|^2 = \left(\frac{0.01-j\omega}{0.01+j\omega}\right)\left(\frac{0.01+j\omega}{0.01-j\omega}\right) = 1 \tag{8.10-4}$$

It is said that $G_{AP}(s)$ is an *all-pass* transfer functions since, interpreted as a filter, it passes all frequencies equally. Since the magnitude of $G_{AP}(s)$ is unity for all frequencies, the magnitude plot of $G_{P1}(s)$ is the same as the magnitude plot of $G_{P0}(s)$ from Sec. 8.7.

In Fig. 8.10-1 the Bode plots of $10G_{P1}(s)$ and $10G_{P0}(s)$ are shown. Unlike the magnitude plot, the phase plot of $G_{P1}(s)$ is greatly affected by the all-pass section with its non-minimum phase zero. The left half-plane pole and the right half-plane zero of Eq. (8.10-3) each add about 45° of phase lag per decade to the phase plot in the area between $\omega = 0.001$ and $\omega = 0.1$. Thus, by the point $\omega = 0.1$ the plant $G_{P1}(s)$ has almost 180° more phase lag than the plant $G_{P0}(s)$. This can be seen in Fig. 8.10-1.

Recall from Sec. 8.7 that a major difficulty in the design of the satellite system was to achieve enough phase lead through the crossover frequency range so that a good phase

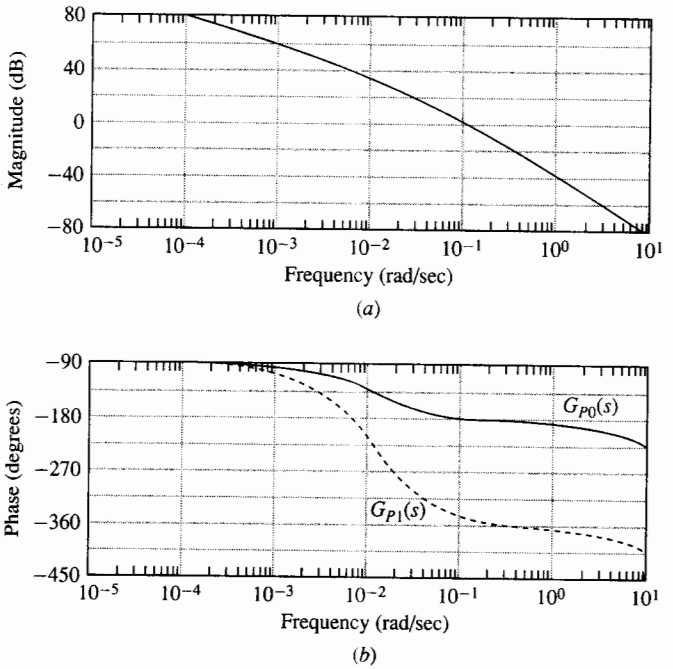


FIGURE 8.10-1
Bode plots of $10G_{P1}(s)$ and $10G_{P0}(s)$. (*a*) Magnitude; (*b*) phase.

margin and transient response result. With the plant $G_{P1}(s)$, this problem is exacerbated by the fact that almost 180° of additional phase lead is needed if the crossover frequency is to remain near $\omega = 0.1$.

Let's try to use a series compensator to negate the effects of the all-pass section of the plant. The pole at $s = -0.01$ can be directly canceled by a zero. Unfortunately, the zero at $s = 0.01$ cannot be canceled. However, its affect on the *phase* plot can be completely countered by adding another zero at $s = -0.01$. The magnitude plot is greatly affected. For the moment, let's ignore the fact that a compensator with two zeros and no poles is unrealizable. We can add more poles later. Let's examine the behavior of

$$G_{L1}(s) = G_{P1}(s)G_{C1}(s) \tag{8.10-5}$$

with

$$G_{C1}(s) = 0.01\left(\frac{s}{0.01} + 1\right)^2 \tag{8.10-6}$$

The Bode plots of $G_{L1}(s)$ and $10G_{P0}(s)$ are given in Fig. 8.10-2. The phase plot of $G_{P0}(s)$ has been recovered by $G_{L1}(s)$. Now, however, the magnitude plot is much flatter. (The difference between $G_{P0}(s)$ and $G_{L1}(s)$ is two zeros, the non-minimum phase zero at $s = +0.01$ and the zero at $s = -0.01$ which restores the phase plot.)

The magnitude of $G_{L1}(s)$ at $\omega = 0.1$ is 1.005 while the value of the magnitude of $G_{L1}(s)$ at $\omega = 10$ is 0.7. If this system is to meet the kind of performance constraints that the loop in Sec. 8.7 needed to meet, a number of poles need to be added well below

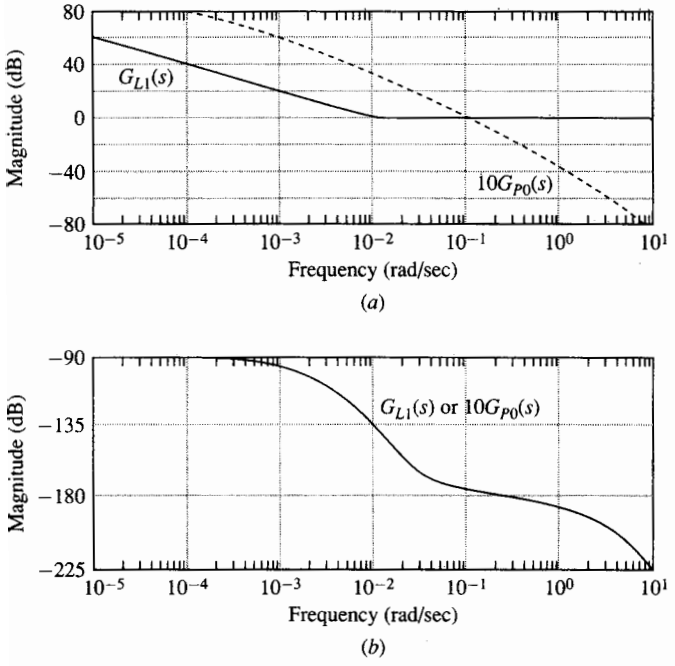


FIGURE 8.10-2
Bode plots of $G_{L1}(s)$ and $10G_{P0}(s)$. (*a*) Magnitude; (*b*) phase.

$\omega = 0.7$ in order to provide enough roll-off in the magnitude plot to meet the high frequency constraint. These poles will cause phase lag in the range $\omega = 0.01$ to $\omega = 0.1$. In addition, a multiple order phase lag is necessary to raise the magnitude over the low frequency performance constraint. This will further add to the phase lag near crossover. Consider the transfer function

$$G_{L2}(s) = G_{L1}(s)G_R(s)G_{lag}(s) \tag{8.10-7}$$

where

$$G_R(s) = \frac{1}{(10s + 1)^2} \tag{8.10-8}$$

is used for roll-off, and

$$G_{\text{lag}}(s) = \frac{(s + 0.001)^2(s + 0.0003)}{s^3} \tag{8.10-9}$$

is a third-order lag compensator.

The Bode plots of $G_{L2}(s)$ are given in Fig. 8.10-3. Notice that the system now meets the magnitude plot objectives (which can be found on Fig. 8.7-3) with some margin for added robustness. Notice, however, that the magnitude plot near crossover is very flat and that the phase lag is greater than 180° and has a large slope at crossover. The phase lag at the crossover frequency cannot be decreased because the increase in the slope of the magnitude plot that accompanies a phase lead compensator causes the crossover

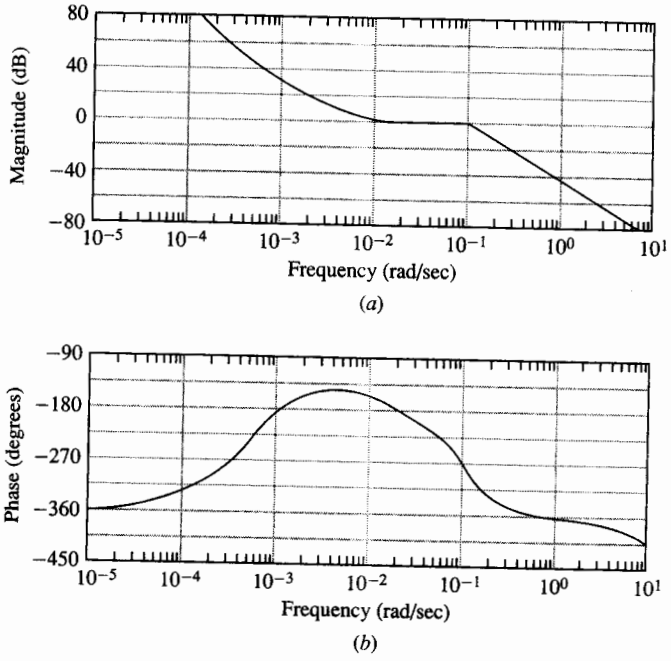


FIGURE 8.10-3
Bode plots of $G_{L2}(s)$. (*a*) Magnitude; (*b*) phase.

frequency to move out to the area where the phase lag is too large to be overcome. The bandwidth must be decreased. A reasonable phase margin can be achieved by decreasing the crossover frequency without the benefit of a lead compensator.

Simply allowing

$$G_{L3}(s) = \frac{G_{L2}(s)}{3} \tag{8.10-10}$$

lowers the magnitude plot about 10 dB. The final crossover frequency occurs at $\omega = 0.004$ with a phase margin of approximately 30°. A polar plot of $G_{L3}(s)$ for $\omega \geq 10^{-3}$ is shown in Fig. 8.10-4. (Use this plot and the transfer function $G_{L3}(j\omega)$ to draw the entire Nyquist plot of $G_{L3}(j\omega)$ and convince yourself that the closed-loop design is stable. Be careful how you treat the indentation around the *four* poles at the origin.) The plot of Fig. 8.10-4 indicates that the minimum value of the return difference for this design is 0.5 or −6 dB. This is somewhat more sensitive than is the design of Sec. 8.7, where the minimal return difference was −2 dB.

The key difference between the design here starting with a non-minimum phase plant and the design of Sec. 8.7 starting with a minimum phase plant is that the non-minimum phase zero in this design forces a reduction of bandwidth. Indeed, the bandwidth is reduced to $\omega_c = 0.004$. This is below the frequency of $\omega = 0.01$ associated with the non-minimum phase zero. It is fortunate that the low frequency performance specification in this problem is loose enough that it was possible to meet the specification even with the reduced bandwidth.

It is a general rule that non-minimum phase zeros restrict the bandwidth of a control system. The bandwidth is usually restricted to be less than the frequency of the lowest frequency non-minimum phase zero so that the phase lag from all the non-minimum phase zeros is absorbed when the magnitude is low. Couple this rule of thumb with the guideline of the last section which states that the bandwidth of a system is usually larger than the frequency associated with an unstable pole and you see that a right half-plane pole at a frequency higher than the right half-plane zero is a devastating combination. In such a situation something should be done to physically redesign the plant and eliminate either the pole or the zero.

It is important to understand how right half-plane zeros physically arise so that they can be eliminated or avoided if plant redesign is possible. There are two common physical situations that manifest themselves as right half-plane zeros in a

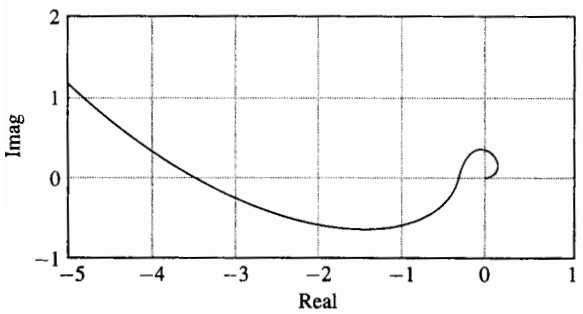


FIGURE 8.10-4
Polar plot of $G_{L3}(s)$.

system model. Right half-plane zeros arise due to time delays in a system and they arise when the measurement of important variables cannot or does not occur. We address time delays here and the importance of measurements in the final section of this chapter.

The impulse response for a system which consists of a time delay of T seconds is

$$h(t) = \delta(t - T) \tag{8.10-11}$$

The system function is given by

$$H(s) = e^{-sT} \tag{8.10-12}$$

Taking the magnitude and phase of the frequency response we get

$$|H(j\omega)| = 1 \qquad \text{for all } \omega$$
$$\arg(H(j\omega)) = -\omega T \tag{8.10-13}$$

A delay system is an all-pass system with a phase lag that increases linearly with frequency. A Bode phase plot of the system with a delay of one second is shown in Fig. 8.10-5. Notice that the phase lag, which increases linearly with ω, becomes an exponential curve on the semi-log Bode phase plot. One can see that it is extremely difficult to overcome this phase lag over a substantial frequency range as lead compensators produce a phase increase that is only linear on the Bode plot and this increase affects only a small frequency range.

The frequency response of the delay can be approximated over a finite frequency range by an all-pass network consisting of right half-plane zeros and mirror image left half-plane poles. The all-pass network matches the magnitude plot of a delay exactly since both systems have unity magnitude for all frequencies. The phase lag of the

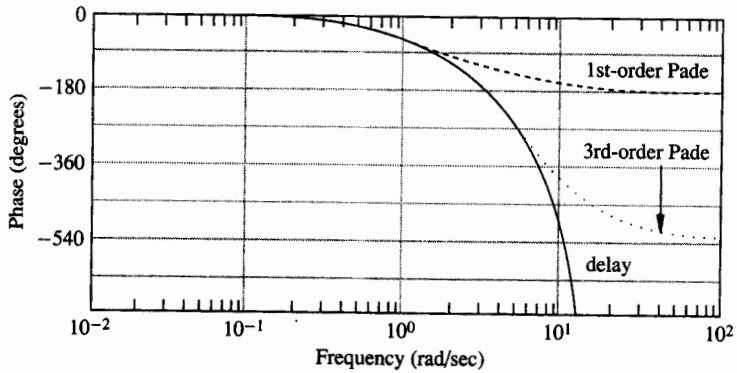


FIGURE 8.10-5
Bode phase plots of a time delay and approximations.

all-pass network using poles and zeros can match the phase lag of the delay over a limited frequency range. This frequency range can be extended by adding more poles and zeros. The total phase lag from each pole-zero combination is 180°. The phase lag from the delay increases without bound.

A reasonable position of the zeros and poles can be found by creating a series expansion for the two transfer functions and equating as many terms as possible. Such an approximation is called a *Padé approximation*. We assume that the number of zeros equals the number of poles and call the number of poles the order of the approximation. Let's demonstrate the technique on a first-order Padé approximation

$$e^{-sT} = 1 - sT + \frac{s^2T^2}{2} - \frac{s^3T^3}{6} + \cdots$$

$$\frac{a-s}{a+s} = 1 - \frac{2}{a}s + \frac{2}{a^2}s^2 + \cdots$$

Equating the first two terms gives $a = \frac{2}{T}$. Higher-order Padé approximations can match more terms and approximate the transfer function of a delay over a wider frequency range. The Bode phase plots of a first-order and a third-order Padé approximation of a one second time delay are shown along with the actual Bode phase plot of a one second time delay in Fig. 8.10-5. The first-order approximation is given by

$$G_{D1}(s) = \frac{-0.5s + 1}{0.5s + 1} \tag{8.10-14}$$

The third-order approximation is given by

$$G_{D3}(s) = \frac{-0.0083s^3 + 0.1s^2 - 0.5s + 1}{0.0083s^3 + 0.1s^2 + 0.5s + 1}$$

$$= \frac{-1\left((s-3.67)^2 + (3.5)^2\right)(s-4.64)}{\left((s+3.67)^2 + (3.5)^2\right)(s+4.64)} \tag{8.10-15}$$

We have seen that time delays cause phase lags that can be modeled with non-minimum phase zeros. We should note that the Bode plot and Nyquist plot frequency response analysis can be performed directly using the frequency response of the time delay transfer function e^{-sT} in series with the other elements of the loop gain. The phase lag limits the achievable bandwidth of the system. Time delays are to be avoided in the design of plants that must be controlled. In process control systems, this usually means placing valves on inputs very close to the reaction tank and using measurements that do not require much time to produce.

In Sec. 8.12 we see how non-minimum phase zeros can sometimes be avoided by taking extra measurements and using state-variable feedback.

8.10.1 CAD Notes

Bode plots of plants with time delays can be plotted directly with MATLAB b adding the phase shift associated with the time delay directly. The following exampl demonstrates this.

Example 8.10-2. Plot the Bode phase plot of the following plants.

$$G_{P1}(s) = \frac{1}{s+0.1}$$

$$G_{P2}(s) = \frac{e^{-2s}}{s+0.1}$$

```
% clear the graph
clg;
% set up the omega vector of frequencies
w=logspace(-2,1,100);
% enter the plant P1
ng=[1];dg=[1 .1];
[mg,pg]=bode(ng,dg,w);
% loop to compute the phase of P2 from the phase of P1
% the time delay causes an extra phase lag which is
% linear in w
% conversion from radians to degrees must be made
for ind=1:100
    pg2(ind)=pg(ind)-(w(ind)*2*(180/pi));
end
% place axis limits
v=[-2 1 -720 0];
axis(v);
% plot the two plots
p=[pg pg2'];
semilogx(w,p),xlabel('frequency (rad/sec)'),
ylabel('phase(degrees)');
% set up to place a grid on the plot
a=axis;
hold on
% plot horizontal line every 90 degrees
for m=a(3):90:a(4);
    x=[a(1) a(2)];
    y=[m,m];
    plot(x,y)
end
% plot vertical line every factor of ten
for x=a(1):1:a(2);
    x1=[x,x];
    y=[a(3) a(4)];
    plot(x1,y)
end
% turn hold off for next plot
hold off
```

The resulting plots appear in Fig. 8.10-6. Notice the grid lines placed in the program every 90° and the different scale used by MATLAB for tick marks.

Exercises 8.10

8.10-1. Given the non-minimum phase plant

$$G_P(s) = \frac{30(-s+1)}{(s+0.1)(s+10)}$$

(*a*) Design a series compensator G_C that achieves 20 degrees of phase margin. There are no other frequency domain requirements (i.e., bandwidth) so that the compensator can be as simple as possible while achieving as large a bandwidth as possible.

Include along with the transfer function of your compensator the Bode plots of the plant and loop gain (i.e., combination plant and compensator). What is the bandwidth of the loop gain transfer function and how does it compare with the location of the non-minimum phase zero?

(*b*) Try to design a controller with bandwidth larger than the right half-plane zero. Is it possible?

8.10-2. Consider the plant

$$G_P(s) = \frac{e^{-s}}{s}$$

and a set of possible controllers given by

$$G_C(s) = K\left(\frac{100(s+1)}{s+100}\right)^n$$

Sketch the Bode plots of the loop gains for $K = 1$ and $n = 1, 2, 3$. For each $n = 1, 2, 3$ set K for the maximum achievable bandwidth. What is the bandwidth achieved in each case? Are the diminishing returns mentioned in the text evident?

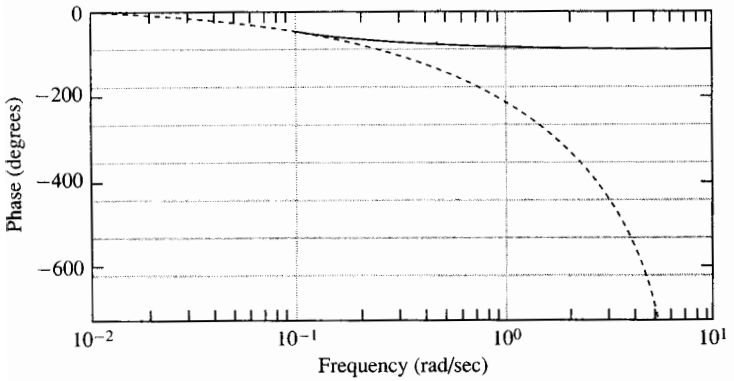


FIGURE 8.10-6
Phase plots with and without time delay.

8.11 POLE PLACEMENT CONTROL

In this section we examine the algorithms studied in Chap. 3 that allow arbitrary placement of the closed-loop poles. Such pole placement can be accomplished using transfer functions and the Diophantine equation or using state-variable feedback. We use a well-known example of a difficult control problem, the inverted pendulum on a cart. This problem is sometimes known as the broom balancing problem. In this section we examine a Diophantine equation approach and show that just because we can place closed-loop poles in a good position we don't necessarily have a good controller. However, a pole placement algorithm may provide a reasonable initial control design that can then be analyzed and improved using frequency domain techniques. In the next section we examine the state-variable approach.

Let's start by deriving a plant model for our example plant. We all have probably at some time tried to balance a broom upside down on our hands. Even with the fantastic processing power of the human brain, this is a difficult task. It becomes even more difficult if we attempt it blindfolded. The plant we examine describes a similar problem. Imagine a "broomstick" on a cart as depicted in Fig. 8.11-1. The dynamics are approximated by assuming that all the mass m of the broomstick is concentrated at the end of the broomstick located at a length l from the cart. The cart has mass M and we assume that, through a motor, we can control directly the force $u(t)$ on the cart. The variables of interest are the angle $\theta(t)$ that the broomstick makes with cart and the distance $y(t)$ that the cart has moved from some arbitrary zero position.

From the physics of the situation, we get the following equations. Applying $\boldsymbol{F} = m\boldsymbol{a}$ in the horizontal direction, we get

$$u(t) = M\frac{d^2y(t)}{dt^2} + m\frac{d^2(y(t) + l\sin\theta(t))}{dt^2} \tag{8.11-1}$$

The vertical force on the broomstick due to gravity is broken into two components. The component working along the shaft is opposed by the shaft and produces no net effect. The component perpendicular to the shaft is given by $mg\sin\theta(t)$. It produces a motion perpendicular to the shaft.

$$mg\sin\theta(t) = m\frac{d^2(y(t)\cos\theta(t) + l\theta(t))}{dt^2} \tag{8.11-2}$$

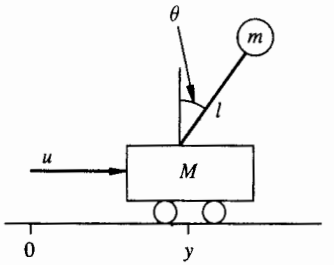


FIGURE 8.11-1
Broomstick on a cart.

The system of Eqs. (8.11-1) and (8.11-2) is nonlinear. These equations can be linearized for small $\theta(t)$ with the approximation

$$\sin\theta(t) \approx \theta(t) \qquad \text{and} \qquad \cos\theta(t) \approx 1 \tag{8.11-3}$$

The two plant equations become

$$u(t) = (M+m)\frac{d^2y(t)}{dt^2} + ml\frac{d^2\theta(t)}{dt^2} \tag{8.11-4}$$

$$mg\theta(t) = m\frac{d^2y(t)}{dt^2} + ml\frac{d^2\theta(t)}{dt^2} \tag{8.11-5}$$

Subtracting Eq. (8.11-5) from Eq. (8.11-4) to eliminate $d^2\theta(t)/dt^2$ produces

$$u(t) - mg\theta(t) = M\frac{d^2y(t)}{dt^2} \tag{8.11-6}$$

or

$$\frac{d^2y(t)}{dt^2} = \frac{-mg}{M}\theta(t) + \frac{1}{M}u(t) \tag{8.11-7}$$

Substituting Eq. (8.11-7) into Eq. (8.11-5) yields

$$mg\theta(t) = \frac{-m^2g}{M}\theta(t) + \frac{m}{M}u(t) + ml\frac{d^2\theta(t)}{dt^2} \tag{8.11-8}$$

or

$$\frac{d^2\theta(t)}{dt^2} = \left(\frac{g}{l} + \frac{mg}{Ml}\right)\theta(t) - \frac{1}{Ml}u(t) \tag{8.11-9}$$

Equations (8.11-7) and (8.11-9) can be turned into state-variable equations by defining a state-variable vector

$$\boldsymbol{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ x_4(t) \end{bmatrix} = \begin{bmatrix} \dot\theta(t) \\ \theta(t) \\ \dot y(t) \\ y(t) \end{bmatrix} \tag{8.11-10}$$

where $\dot\theta(t) = d\theta(t)/dt$ and $\dot y(t) = dy(t)/dt$.

We then achieve the physical state-variable description

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t) + \boldsymbol{b}u(t) \tag{Ab}$$

$$y(t) = \boldsymbol{c}^T\boldsymbol{x}(t) \tag{c}$$

$$\boldsymbol{A} = \begin{bmatrix} 0 & \frac{g}{l} + \frac{mg}{Ml} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{-mg}{M} & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} \frac{-1}{Ml} \\ 0 \\ \frac{1}{M} \\ 0 \end{bmatrix} \tag{8.11-11}$$

$$\boldsymbol{c}^T = [0\ 0\ 0\ 1]$$

Inserting hypothetical values for the constants we arrive at our example plant model

$$\boldsymbol{A} = \begin{bmatrix} 0 & 11 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{8.11-12}$$

$$\boldsymbol{c}^T = [0\ 0\ 0\ 1]$$

We have considered the output variable to be the position of the cart. We wish to move the cart from one position to another with the pendulum coming to rest in the vertical position. Notice that the output does not include a measurement of the angle $\theta(t)$. This is similar to balancing a broomstick blindfolded. The angle must be inferred from the forces on the cart. This is expected to be a difficult if not impossible task.

Let's find the resolvent matrix of our system and determine the plant poles.

$$(s\boldsymbol{I} - \boldsymbol{A})^{-1} = \frac{1}{s^2(s^2-11)}\begin{bmatrix} s^3 & 11s^2 & 0 & 0 \\ s^2 & s^3 & 0 & 0 \\ -s & -s^2 & s(s^2-11) & 0 \\ -1 & -s & s^2-11 & s(s^2-11) \end{bmatrix} \tag{8.11-13}$$

There are four plant poles; two poles are at the origin, one is at $s = -\sqrt{11}$ and another is an unstable pole at $s = +\sqrt{11}$. It should not be surprising that the plant is unstable. If the state-variable vector is perturbed from its zero state and no further control action is supplied, the pendulum swings down. In the linear model, $\theta(t)$ grows until the linear model assumption of small $\theta(t)$ becomes invalid. However, if a stable closed-loop system can be created then any initial offset in the position of the cart or the angle of the pendulum is returned to zero and the cart returns to the zero position with the pendulum vertical. A step input then produces a constant offset in the position $y(t)$ and a vertical pendulum represented by $\theta(t) = 0$. Let's proceed by deriving the plant transfer function.

$$(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b} = \frac{1}{s^2(s^2-11)}\begin{bmatrix} -s^3 \\ -s^2 \\ s(s^2-10) \\ s^2-10 \end{bmatrix} \tag{8.11-14}$$

$$\frac{Y(s)}{U(s)} = \boldsymbol{c}^T(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b} = \frac{s^2-10}{s^2(s^2-11)} \tag{8.11-15}$$

The expectations based on physical intuition that this plant would be very difficult to control are verified in the transfer function of the plant. It contains a non-minimum phase zero at a lower frequency than an unstable pole. This is the difficult combination mentioned in the previous section. It is especially difficult here because the zero is very near the pole. Remember, if the zero were to cancel the pole exactly, it would be impossible to move the pole. In a root locus analysis a zero very near the unstable pole means a large gain is required to wrench that pole away from the zero.

Let's proceed with a pole placement design using the transfer function and Diophantine equation approach. Ideally, the bandwidth of the system should be greater

han that of the unstable pole at $s = \sqrt{11} \simeq 3.32$ but less that of the non-minimum hase zero at $s = \sqrt{10} \simeq 3.16$. Of course, this is not possible so we pick our closed-oop poles to have a compromise bandwidth of $\omega = 3.25$.

Recall from Sec. 3.2 that to have a realizable controller, the degree of the desired closed-loop polynomial $D_M(s)$ must equal $n + m$ when n is the degree of the plant denominator and $m \geq n - 1$.

$$\deg D_M(s) \geq n + n - 1 = 2(4) - 1 = 7 \tag{8.11-16}$$

The desired closed-loop polynomial must be at least seventh-order. We choose to use an eighth-order polynomial for convenience. $D_M(s)$ is chosen rather arbitrarily to have well-damped poles with a bandwidth of $\omega = 3.25$ and other poles at larger frequencies. Let

$$\begin{aligned} D_M s &= (s + 2.3 + j2.3)(s + 2.3 - j2.3)(s + 5 + j5)(s + 5 - j5)(s + 10)^4 \\ &= s^8 + 54.6s^7 + 1290.58s^6 + 17359s^5 + 146309s^4 \\ &\quad + 794960s^3 + 2726400s^2 + 5474000s + 5290000 \end{aligned} \tag{8.11-17}$$

Let

$$G_c(s) = \frac{b_3 s^3 + b_2 s^2 + b_1 s + b_0}{s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0} \tag{8.11-18}$$

Then

$$\begin{aligned} &D_c(s)D_p(s) + N_c(s)N_p(s) = \\ &\quad s^8 + a_3 s^7 + (a_2 - 11)s^6 + (a_1 - 11a_3 + b_3)s^5 + (a_0 - 11a_2 + b_2)s^4 \\ &\quad + (-11a_1 + b_1 - 10b_3)s^3 + (-11a_0 + b_1 - 10b_3)s^3 + (-11a_0 + b_0 - 10b_2)s^2 \\ &\quad + (-10b_1)s + (-10b_0) \end{aligned} \tag{8.11-19}$$

Equating coefficients in Eq. (8.11-17) and Eq. (8.11-19) we can solve for $G_c(s)$

$$\begin{aligned} G_c(s) &= \frac{1539925.6s^3 + 5022290.18s^2 - 547400s - 529000}{s^4 + 54.6s^3 + 1301.58s^2 - 1521956s - 4861663.8} \\ &= \frac{1539926(s + 3.337)(s - 0.3609)(s + 0.2852)}{\left[(s + 74.44)^2 + (100.5)^2\right](s - 97.48)(s + 3.187)} \end{aligned} \tag{8.11-20}$$

The second equality in Eq. (8.11-20) is only approximate because in computing the roots some rounding occurs.

It is interesting to note that the series compensator of Eq. (8.11-20) is itself unstable, having a pole at $s = 97.48$. However, the closed-loop system is stable. The closed-loop transfer function is computed to be

$$\begin{aligned} M_c(s) &= \frac{1539926(s + 3.162)(s - 3.162)(s + 3.337)(s - 0.361)(s + 0.285)}{\left[(s + 2.30)^2 + (2.29)^2\right]\left[(s + 4.8)^2 + (4.91)^2\right]\left[(s + 9.78)^2 + (4.28)^2\right]} \\ &\quad \times \frac{1}{(s + 4.8)(s + 6.54)} \end{aligned} \tag{8.11-21}$$

Notice that the closed-loop poles are close to the desired positions but not exactly on the desired positions. The reason for this is not a problem with the theory of the pole placement algorithm but rather it is due to numerical inaccuracies in the computation. Such movement in the closed-loop poles due to small numerical inaccuracies may alert us to the fact that, in this case, even with feedback, the closed-loop system is highly sensitive to parameter inaccuracies. While the closed-loop poles of Eq. (8.11-21) are not exactly where we desire them, they are the poles of an apparently well-behaved closed-loop transfer function. Notice, however, that the closed-loop transfer function must contain all the zeros of both the plant, $G_p(s)$, and the series compensator, $G_c(s)$. In this case, there are two zeros much closer to the origin than the closest poles and the effects of these zeros greatly alter the dynamics. The transient response of this system is shown in Fig. 8.11-2. The response is very different from what we had in mind when we chose the closed-loop poles. It is unlikely that such a wild response would be acceptable. Indeed, except for all but the smallest inputs, the assumptions made about the linearity of the model and the abilities of a motor to provide the required force would be invalid.

Let's examine the frequency domain plots of the loop gain of this controller. In Fig. 8.11-3, the Bode magnitude and phase plots are given. The magnitude of the loop gain is near unity for almost two decades from $\omega = 1$ until $\omega = 100$. Also, the phase is close to 180° for these same frequencies. This clearly indicates trouble as the loop gain is near the critical point, $s = -1$, for all these frequencies. The phase lag of nearly 180° near the almost zero magnitude slope is due to the extra 180° additional phase lag that the non-minimum phase zeros provide over the phase lag of a minimum phase system with the same magnitude plot.

The Bode plots are magnified in Fig. 8.11-4 to show only the range $\omega = 1$ to $\omega = 100$. Notice the scale of the magnitude and phase axes. The magnitude plot crosses 0 dB three times while the phase plot crosses 180° three times. The polar plot for frequencies in the same range is shown in Fig. 8.11-5. Again, notice the scale of the plot. The crossings of the Bode plots occur in such a manner as to create a counterclockwise encirclement of the critical point on the polar plot. A second counterclockwise encirclement occurs when the negative frequencies are traced in the Nyquist D-contour. Two counterclockwise encirclements are needed for closed-loop stability since there are two right half-plane poles in the loop gain—one from the

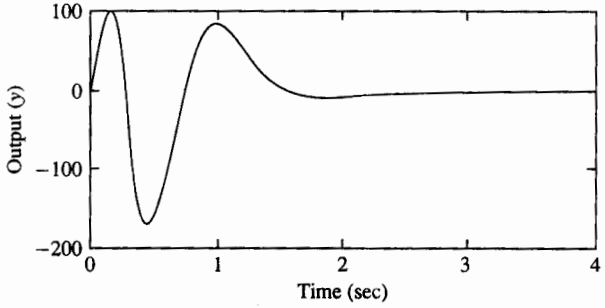


FIGURE 8.11-2
Unit step response of the closed-loop system.

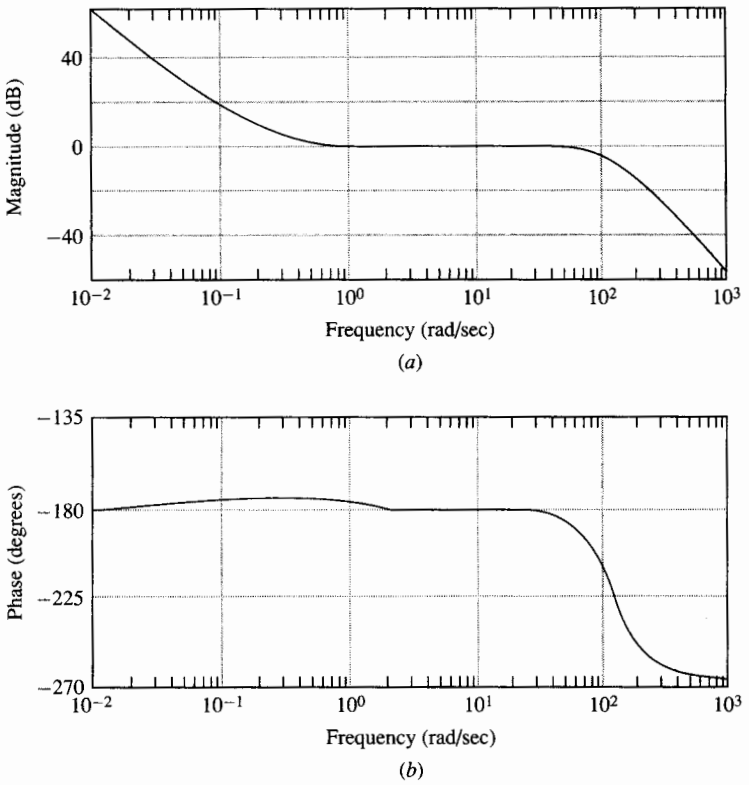


FIGURE 8.11-3
Bode plots of the loop gain. (*a*) Magnitude; (*b*) phase.

plant and one from the controller. The reader should verify that the Nyquist theorem for this loop gain does indeed show closed-loop stability.

Here is a case in which we have stability for the mathematical models involved but the loop gain passes so close to the critical point on all sides that if the mathematical model is less than perfect the system will be unstable. It has almost no robustness. One would not attempt to use this controller on a physical system. It could not be used despite the fact that mathematically the closed-loop poles are placed well into the left half-plane. Apparently, good closed-loop pole positions do not necessarily mean a well-behaved control system. We must check other loop properties by checking the frequency domain plots of the loop gain or sensitivity transfer functions.

The fact that the pole placement technique does not produce an acceptable controller in this case does *not* mean that the pole placement technique is a poor approach to controller design. In this case we started with a plant that is extremely difficult to control. We know of no controller that adequately handles a similar system using only the measurement we have used here. Indeed, one would be hard pressed to even stabilize the system with the iterative approach using frequency domain plots. It is an interesting exercise to try this after seeing how the pole placement controller

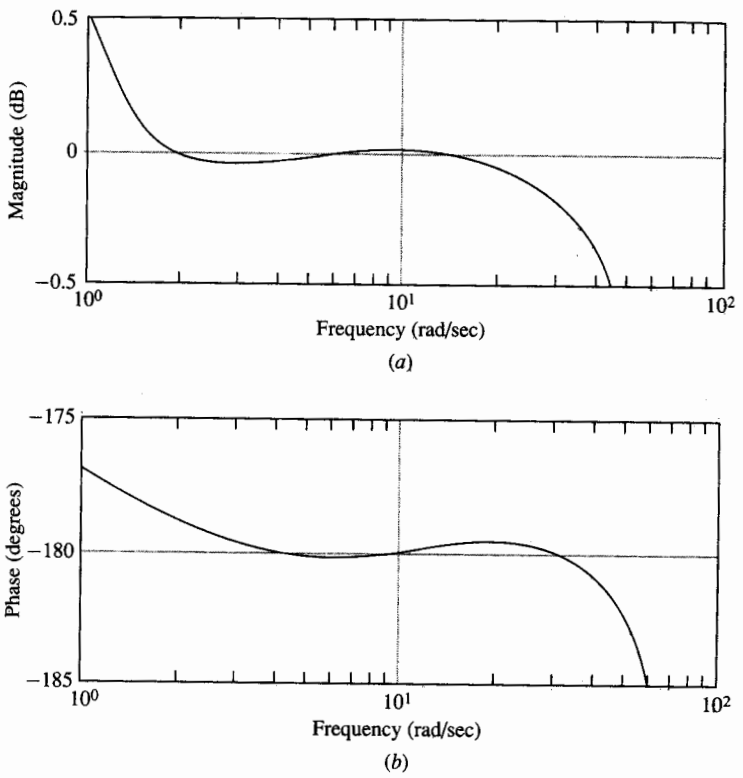


FIGURE 8.11-4
Magnification of Bode plots of the loop gain. (*a*) Magnitude; (*b*) phase.

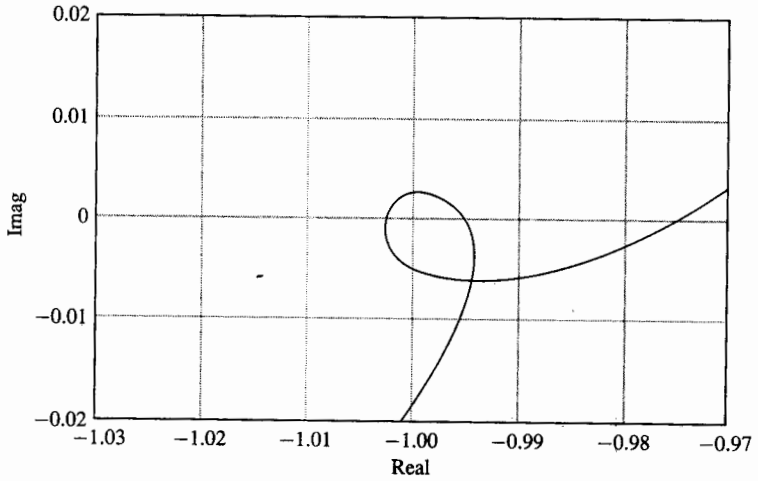


FIGURE 8.11-5
Polar plot of the loop gain.

achieves a stabilizing gain. One can also note that there is no way to use frequency domain techniques to improve the existing design as any lead or lag around the intermediate range of frequencies causes a change in the number of encirclements and an unstable closed-loop system. It would also require a clever designer to stabilize this system using the root locus technique. The root locus technique also suffers the same problem as the pole placement technique in that the method gives little indication of the properties of the loop gain, specifically robustness, sensitivity, and frequency response.

The pole placement technique using a series compensator is often a good way to get an initial design on difficult control problem. That initial design should then be analyzed using the frequency domain methods with Bode and Nyquist plots. If the design is inadequate it can possibly be improved by using series compensator techniques such as lead or lag compensators. If the initial design cannot be sufficiently improved, it may be necessary to alter the design of the plant itself by reducing time delays or adding measurements or control variables. In the next section we approach the broom balancing problem from a state-variable pole placement approach. We will see that if certain state-variables can be measured, this system can be adequately controlled.

Exercises 8.11

8.11-1. There is nothing in the theory of pole placement indicating that particular placements of poles won't work. For example, nothing in the theory prevents choosing closed-loop system poles at a higher frequency than right half-plane zeros even though we know from Bode-Nyquist ideas that such a design does not generally produce an acceptable loop gain.

Consider the plant from Exercise 8.10-1.

$$G_p(s) = \frac{30(-s+1)}{(s+0.1)(s+10)}$$

Design a pole placement controller with all the closed-loop poles placed at a frequency $\omega = 10$ or greater. Examine the Bode and Nyquist plots of the resulting loop gain and critique the design.

8.11-2. Now redesign a controller for the plant of Exercise 8.11-1, placing the dominant closed-loop pole at $s = -0.5$ and the rest of the closed-loop poles at a frequency greater than $\omega = 10$. Examine the Bode and Nyquist plots of the resulting loop gain and critique the design.

8.12 STATE-VARIABLE FEEDBACK—THE ADVANTAGE OF EXTRA MEASUREMENTS

In this section we examine how much better the broom balancing plant of Sec. 8.11 can be controlled when certain variables besides the plant output are measured. First, we show that if we can measure all state-variables, we can avoid the problems associated with non-minimum phase zeros and achieve a good controller. Then, we show how to adjust a state-variable controller if not all states are measurable. Of course, as less states are measurable, we lose the benefits of state-variable control. When we get to the point where only the plant output is measurable, the state-variable controller is equivalent to an output feedback controller. The final conclusion is that the problem associated with non-minimum phase zeros can be alleviated if certain specific additional measurements can be made.

Let's consider the same plant as in Sec. 8.11 but this time let's assume all the state-variables are available for measurement so that pole placement can be achieved by state-variable feedback. The plant is represented by the state Eqs. (***Ab***) and (***c***) with $\boldsymbol{A}$, $\boldsymbol{b}$, and $\mathbf{c}$ given by Eq. (8.11-12). With the state-variable approach we need only place the four poles associated with the eigenvalues of the state matrix $\boldsymbol{A}$. The closed-loop system has four poles given by the eigenvalues of $\boldsymbol{A} - \boldsymbol{b}\boldsymbol{k}^T$. The appropriate feedback vector $\boldsymbol{k}$ is determined using the methods of Sec. 3.5. Let's place the closed-loop poles so that the closed-loop pole polynomial is given by

$$\begin{aligned} D_H(s) &= (s+2.3+j2.3)(s+2.3-j2.3)(s+5+j5)(s+5-j5) \\ &= s^4 + 14.6s^3 + 128.58s^2 + 335.8s + 529 \end{aligned} \tag{8.12-1}$$

The closed-loop poles chosen are the same as the four poles closest to the origin of the eight desired poles chosen in Sec. 8.11. The appropriate feedback gains in $\boldsymbol{k}$ are determined according to Eq. (3.5-21), repeated here as Eq. (8.12-2).

$$\boldsymbol{k}^T = \frac{\boldsymbol{d}^T - \boldsymbol{a}^T}{K} \boldsymbol{C}_{A_p b_p} \boldsymbol{C}_{Ab}{}^{-1} \tag{8.12-2}$$

In Equation (8.12-2), $\boldsymbol{C}_{Ab}$ is the controllability matrix associated with the physical state-variable representation of the system.

In this case

$$\boldsymbol{C}_{Ab} = \begin{bmatrix} -1 & 0 & 11 & 0 \\ 0 & -1 & 0 & -11 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{8.12-3}$$

This matrix is invertible, indicating that the system is controllable and arbitrary pole placement is achievable.

The inverse is given by

$$\boldsymbol{C}_{Ab}{}^{-1} = \begin{bmatrix} 0.1 & 0 & 1.1 & 0 \\ 0 & 0.1 & 0 & 1.1 \\ -0.1 & 0 & -0.1 & 0 \\ 0 & -0.1 & 0 & -0.1 \end{bmatrix} \tag{8.12-4}$$

We also need the phase variable description of the system. The phase variable description of the system can be obtained directly from the plant transfer function

Eq. (8.11-15), repeated here as Eq. (8.12-5)

$$G_p(s) = \frac{s^2 - 10}{s^2(s^2 - 11)} = \frac{s^2 - 10}{s^4 - 11s^2} \tag{8.12-5}$$

$$\boldsymbol{A_p} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 11 & 0 \end{bmatrix} \qquad \boldsymbol{b_p} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{c_p}^T = [-10\ 0\ 0\ 0] \tag{8.12-6}$$

The vector $\boldsymbol{a}^T$ in Equation (8.12-2) is the negative of last row in $\boldsymbol{A}_p$

$$\boldsymbol{a}^T = [0\ 0\ -11\ 0] \tag{8.12-7}$$

The matrix $\boldsymbol{A_P b_P}$ in Eq. (8.12-2) is the controllability matrix associated with the phase variable description of Eq. (8.12-6)

$$\boldsymbol{C}_{A_p b_p} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 11 \\ 1 & 0 & 11 & 0 \end{bmatrix} \tag{8.12-8}$$

The vector $\boldsymbol{d}^T$ is composed of the coefficients of the desired pole polynomial of Eq. (8.12-1)

$$\boldsymbol{d}^T = [529 \quad 335.8 \quad 128.58 \quad 14.6] \tag{8.12-9}$$

Feeding all the appropriate information into Eq. (8.12-2), and letting the extra parameter $K = 1$, the feedback gains $\boldsymbol{k}^T$ are computed as

$$\boldsymbol{k}^T = [-48.18 \quad -192.48 \quad -33.58 \quad -52.9] \tag{8.12-10}$$

One can check the eigenvalues of $\boldsymbol{A} - \boldsymbol{bk}^T$ to ascertain that the closed-loop poles are placed as desired.

Now let us analyze the resulting controller. Look first at the elementary block diagram of the system given by Fig. 8.12-1. An output disturbance has been added to this diagram for later analysis. The controller is realized by measuring the four state-variables and feeding each through a constant gain. There is no problem with pole-zero cancellation or trying to realize an improper transfer function. The closed-loop transfer function is given by

$$\boldsymbol{M_c}(s) = \boldsymbol{c}^T(s\boldsymbol{I} - \boldsymbol{A} + \boldsymbol{bk}^T)^{-1}\mathbf{b} = \frac{s^2 - 11}{\left((s+2.3)^2 + (2.3)^2\right)\left((s+5)^2 + (5)^2\right)} \tag{8.12-11}$$

The plant zeros remain in the closed-loop transfer function as they must. The poles have been moved to the desired location. The system is not very sensitive to changes in the coefficients of the controller as was the output feedback system of the last section. The closed-loop transfer function can also be solved by finding $H_{eq}(s)$ and using

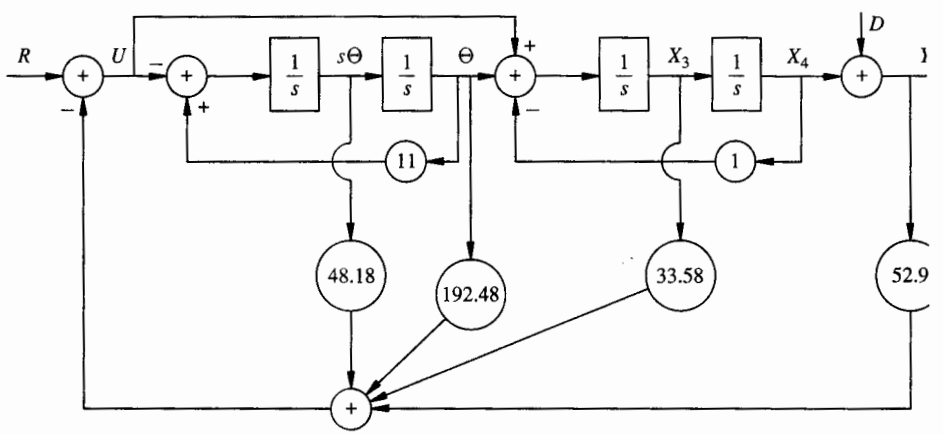


FIGURE 8.12-1
State-variable feedback controller.

$$M_c(s) = \frac{G_p(s)}{1 + G_p(s)H_{eq}(s)} \tag{8.12-12}$$

The transfer function $H_{eq}(s)$ is

$$H_{eq}(s) = \frac{\boldsymbol{k}^T(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b}}{\mathbf{c}^T(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{b}} = \frac{14.6\left((s+1.6)^2 + (2.2)^2\right)(s+4.85)}{\left(s + \sqrt{10}\right)\left(s - \sqrt{10}\right)} \tag{8.12-13}$$

Of course, we could not implement a controller by building a transfer function equal to $H_{eq}(s)$ and using it in the feedback path. The transfer $H_{eq}(s)$ is improper and, more seriously, has a right half-plane pole at the same location as a right half-plane zero of the plant. Implementing $H_{eq}(s)$ would create an illegal right half-plane pole-zero cancellation. The resulting system would not be internally stable and small reference inputs would result in unbounded plant inputs. The state-variable feedback of Fig. 8.12-1 is implementable and mathematically produces the same closed-loop transfer function as an output feedback controller using $H_{eq}(s)$ would. The $H_{eq}(s)$ is a mathematical abstraction that allows a different way of thinking about and analyzing the state-variable controller.

8.12.1 Disturbance Rejection

One must be careful in analyzing the reaction of a state-variable control system to a disturbance. The problem is that, depending on where the disturbance enters, only part of the controller affects the disturbance by feeding it back. The state-variable feedback system does not respond to an output disturbance in the same way that the $H_{eq}(s)$ abstraction does because $H_{eq}(s)$ contains parts of the controller and plant that do not directly act upon the disturbance.

Let's analyze the response of the system of Fig. 8.12-1 to an output disturbance. First note the placement of the output disturbance on Fig. 8.12-1. The disturbance is not part of the plant's internal feedback of $X_4(s)$ but it is part of the controller's feedback of $Y(s)$ to $U(s)$. First we express $Y(s)$ as

$$\begin{aligned} Y(s) &= \boldsymbol{c}^T \boldsymbol{X}(s) + D(s) \\ &= \boldsymbol{c}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b} U(s) + D(s) \end{aligned} \tag{8.12-14}$$

The tricky part is to correctly factor $D(s)$ into the equation that describes how the plant input $U(s)$ is formed.

$$\begin{aligned} U(s) &= -\boldsymbol{k}^T \boldsymbol{X}(s) - 52.9\ D(s) \\ &= -\mathbf{k}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b} U(s) - 52.9\ D(s) \end{aligned} \tag{8.12-15}$$

Notice that only the term $k_4 = 52.9$ multiplies the $D(s)$ term in the feedback. The other feedback gains come before the disturbance. Equation (8.12-15) can be rewritten as

$$U(s) = \frac{-52.9}{1 + \boldsymbol{k}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}} D(s) \tag{8.12-16}$$

Substituting Eq. (8.12-16) into Eq. (8.12-14), we get

$$Y(s) = \left(1 - \frac{52.9 \boldsymbol{c}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}}{1 + \boldsymbol{k}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}} \right) D(s) \tag{8.12-17}$$

Recalling

$$G_p(s) = \boldsymbol{c}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}$$

and

$$G_p(s) H_{\text{eq}}(s) = \boldsymbol{k}^T (s\boldsymbol{I} - \boldsymbol{A})^{-1} \boldsymbol{b}$$

We can rewrite Eq. (8.12-17) in terms of $G_p(s)$ and $H_{\text{eq}}(s)$ as

$$\begin{aligned} Y(s) &= \left(1 - \frac{52.9 G_p(s)}{1 + G_p(s) H_{\text{eq}}(s)} \right) D(s) \\ &= \frac{1 + G_p(s)(H_{\text{eq}}(s) - 52.9)}{1 + G_p(s) H_{\text{eq}}(s)} \end{aligned} \tag{8.12-18}$$

This expression is more complicated than the usual sensitivity function. When more than one variable is measured and fed back, the plant is no longer a single-input, single-output system but it must be considered as a single-input, multi-output system. The analysis of such systems with respect to sensitivity and robustness is somewhat more complicated and beyond the scope of this book. The principle of large loop gain being good for sensitivity reduction still holds but must be interpreted properly. In Eq. (8.12-18), if the loop gain is large and if most of its size comes from the last loop that works on the disturbance, the reaction to the output disturbance is small.

Notice that, in Eq. (8.12-18), the amount of disturbance rejection depends on how closely $H_{\text{eq}}(s)$ matches the outside loop feedback constant, 52.9. If $H_{\text{eq}}(s)$ is near

52.9 then the response to a disturbance is approximately

$$\frac{D(s)}{1 + G_p(s) H_{\text{eq}}(s)}$$

If 52.9 is small compared to $H_{\text{eq}}(s)$, the response to a disturbance is approximate $D(s)$ itself, i.e., there is almost no disturbance attenuation. Thus we see that wh using state-variable feedback the attenuation of a disturbance depends not as much the size of the overall loop gain but more on the size of the gain that surrounds t disturbance.

We have seen that, if we are able to measure all state-variables for feedbac then the problems of right half-plane zeros can be avoided. While poles arise from t internal dynamics of the plant, zeros arise from the interaction of the dynamics wi the plant inputs and outputs. If we can create more measurements or control inputs w can eliminate bothersome plant zeros. This allows the successful control of unusuall difficult plants that cannot be adequately controlled using only output feedback.

8.12.2 Inaccessible State Variables

Can some of the advantages of state-variable feedback still be achieved without havin to measure *all* of the state-variables? The answer is yes, if certain needed variable are measurable. Let's look at the system of Fig. 8.12-1 and assume that y and are measurable but their derivatives $\dot{y}$ and $\dot{\theta}$ are not. By a simple block-diagra manipulation the system of Fig. 8.12-1 can be turned into the system of Fig. 8.12- with

$$H_{01}(s) = 4.81s + 17.04 \tag{8.12-19}$$

$$H_{02}(s) = 33.5s + 52.9 \tag{8.12-20}$$

These two transfer functions are improper and thus unrealizable. They do no however, try to cancel right half-plane poles or zeros. The transfer functions ca

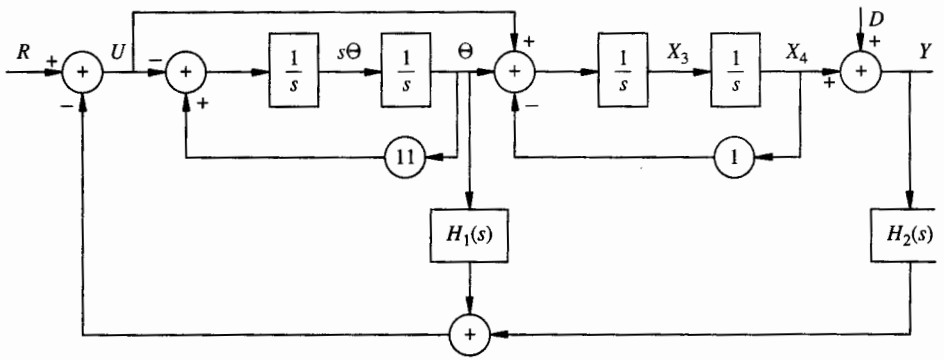


FIGURE 8.12-2
Controller with $Y(s)$ and $\theta(s)$ measured.

be made realizable by adding a denominator with a pole in the left half-plane at a frequency high enough that the extra pole does not slow the system down even after it is moved to its closed-loop position. If the poles are fast enough, the systems will react almost exactly like those of Eqs. (8.12-19) and (8.12-20) for the slower signals fed into them. Since the eventual bandwidth of the closed loop system is about $\omega = 3$, a reasonable placement for the extra poles is $\omega = 30$. The transfer functions

$$H_1(s) = \frac{4.81s + 17.04}{\frac{s}{30} + 1} \tag{8.12-21}$$

$$H_2(s) = \frac{33.5s + 52.9}{\frac{s}{30} + 1} \tag{8.12-22}$$

are realizable and produce almost the same effect as the state-variable controller.

The key extra measurement in controlling this plant is θ, the angle, of the pendulum. It makes sense physically that measuring this variable facilitates control. It supplies the measurement provided visually in the broom balancing problem. Being able to measure θ is like removing the blindfold. If we try to perform block-diagram manipulation to eliminate the feedback from θ, thus turning the system into an output feedback control system, the feedback transfer function which results is the $H_{eq}(s)$ from Eq. (8.12-13), which cannot be used because of right half-plane pole-zero cancellations. Thus, the measurement of θ is indeed the key to controlling this system.

We have seen that additional measurements can make plants easier to control. Plants can always be controlled with reasonable sensitivity if all state-variables are measured, but often all variables are not needed. If the key variables are identified the added expense of achieving extra measurements can be applied only when needed. Whenever extra measurements are used the system can no longer be simply analyzed for sensitivity and robustness properties using the single-input, single-output techniques of this book. Research on multi-input, multi-output techniques is ongoing. Much has been learned about these systems and many schools teach graduate courses in multivariable control systems.

Exercise 8.12

8.12-1. The objective is to design a state-space controller for the plant given by Eqs. (***Ab***) and (***c***) with

$$\boldsymbol{A} = \begin{bmatrix} -0.1 & 0 \\ 0 & -10 \end{bmatrix} \qquad \boldsymbol{b} = \begin{bmatrix} 30 \\ 30 \end{bmatrix} \qquad \boldsymbol{c} = \begin{bmatrix} 1/9 \\ -10/9 \end{bmatrix}$$

Calculate the elements in the $\boldsymbol{k}$ vector and the constant gain K that, in combination with this plant, gives a closed-loop transfer function

$$\frac{Y(s)}{R(s)} = \frac{75(-s+1)}{(s+4)(s+20)}$$

(i.e., place the closed loop poles at -4 and -20). For this control design, find the transfer function H_{eq}. Is H_{eq} realizable? Why? Generate the Bode plot of the loop gain, $G_p H_{eq}$. What is the bandwidth and phase margin of this design? How does H_{eq} appear to increase the bandwidth while maintaining some phase margin?

8.13 CONCLUSIONS

In this chapter we have completed our study of the basic principles of control theory by applying the insights we have gained in modeling and analyzing control systems to the problem of designing control systems. First a strategy for controller design by manipulation of the loop gain transfer function with series compensators was developed. After stability of the nominal control system is assured, performance objectives are pursued.

At low frequencies a large loop gain corresponding to a large return difference and small sensitivity function is sought to provide low sensitivity to plant perturbations, good low frequency disturbance rejection and reasonable reference input following for low frequency inputs. At high frequencies a small loop gain corresponding to a small complementary sensitivity function or closed-loop transfer function is necessary for the design to be robust in the face of high frequency model uncertainty. A small loop gain is also advantageous for minimizing the effect of sensor noise and limiting the control action at high frequencies. In the transition frequencies where the loop gain goes from large to small, a large return difference is desired so that the system has an adequate transient response and enjoys a large degree of robustness. The crossover frequency should be as large as high frequency constraints allow and the phase margin should be as large as a fast transition from high loop gain to small loop gain allows.

There are a number of building blocks for shaping the loop gain of a system and the more fundamental of these were explored. It was found that unstable plants can be stabilized using the same techniques as are used to improve the performance of stable plants. In particular, the control system designer is constantly trying to achieve as much phase lead and as large a bandwidth as conditions allow. The design is often limited by model uncertainty, which limits the bandwidth, and often by right half-plane or non-minimum phase zeros, which create extra phase lag. Non-minimum phase zeros can arise from time delays within the plant or from insufficient measurements.

Pole placement techniques can produce reasonable starting points for controller designs involving difficult plants. They do not, however, overcome problems associated with non-minimum phase zeros or modeling uncertainty. The loop gain, sensitivity function, and complementary sensitivity function associated with a pole placement design must be analyzed to assess the design. Sometimes problems with non-minimum phase zeros can be overcome with the use of extra plant measurements. The extreme is to measure all state-variables and perform state-variable feedback. If not all state-variables can be measured, then some can be accounted for with block-diagram manipulations. At some point, however, a set of required measurements is reached.

In attempting to analyze the sensitivity, disturbance rejection properties and robustness properties of a state feedback controller, one comes to the realization that the techniques developed in this book do not directly apply. Since the controller's position in the loop is physically distributed it is difficult to define a single loop gain.

Indeed, we have relied on the assumption that the systems we have been analyzing deal with a single input and a single output. A controller with more than one measurement must be analyzed as a system with more than one output. The techniques developed in this book can be generalized to include multi-input and multi-output systems (often called *multivariable* systems). We leave this generalization to the next course of study. Let us remark, however, that the major difference is that, instead of dealing with transfer functions, multivariable systems can be described with matrices of transfer functions. Many problems of multivariable control arise from two difficult properties of these matrices—matrix multiplication does not commute and matrices can have both large and small elements at the same frequency. After the details concerning these problems with matrices are worked out the principles explored in this book remain.

There are many other topics of interest in a continuing study of control theory. Optimal control techniques have been applied in both the time and frequency domain. Researchers are just beginning to understand the control of nonlinear plants and the use of nonlinear controllers. On line plant identification and adaptive control techniques that allow modeling and control to occur simultaneously are in need of more study. We hope that the principles presented in this book provide a solid foundation for those who go on to design working controllers and those who will continue forging new results in the theory of control systems.

PROBLEMS

8.1. (*a*) Why does a control system designer want the loop gain high at low frequencies? Why does the designer want the loop gain low at high frequencies? Write a short, one sentence answer for each reason. You should mention at least six reasons total.

(*b*) Using ideas from Bode plots, explain why increasing the loop gain of a controller generally makes the closed-loop system respond faster.

8.2. For a design problem a disturbance exists at the input to the plant as shown in Fig. P8.2. The maximum magnitude of the disturbance is $A = 100$. The design can not tolerate a sinusoidal output from this disturbance with a magnitude greater than 0.01. If at $\omega = 0.1$, $|G_p| = 10$ find the constraint on $|1 + G_cG_p|$ at $\omega = 0.1$. What is the constraint on $|G_cG_p|$ at $\omega = 0.1$?

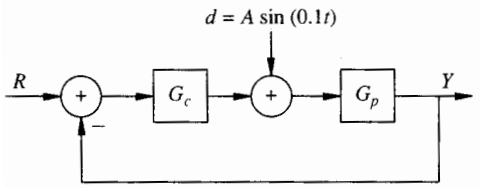


FIGURE P8.2
Problem 8.2.

8.3. A plant model set is described by

$$\tilde{G}_p(s) = G_p(s)\,(1 + L_m(s))$$

and

$$|L_m(j\omega)| < l_m(j\omega)$$

where

$$G_p(s) = \frac{100}{s^2}$$

and

$$l_m(s) = \left|\frac{(s + 10^2)}{3000}\right|$$

Design a controller $G_c(s)$ so that the closed-loop system is stable for all plants in $\tilde{G}_p(s)$ and has reasonable performance for all plants in $\tilde{G}_p(s)$. This loop should also achieve 60° of phase margin at crossover frequency and a loop gain of greater than 90 dB for $\omega < 0.1$ when the nominal plant $G_p(s)$ is used. The crossover frequency is not specified.

8.4. The following parts of the problem refer to Fig. P8.4.

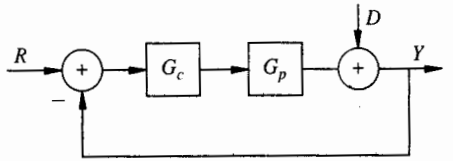


FIGURE P8.4
Problem 8.4.

(*a*) Sketch the Bode plots of the plant

$$G_p(s) = \frac{0.1(s + 1)}{s^2 + 0.1s + 1}$$

(*b*) Assume that $\tilde{G}(s) = G_p(s)\,(1 + L_m(s))$.
with $|L_m(j\omega)| < |0.003(j\omega + 1)^2|$
and that it is required that $\left|\frac{Y(j\omega)}{D(j\omega)}\right| < 0.1$ for $\omega < 1$.
where $D(j\omega)$ is an output disturbance.
Design a cascade compensator $G_c(s)$ so that

(i) The closed-loop system is stable.
(ii) The closed-loop system remains stable for all $L_m(j\omega)$ satisfying the stated bound.
(iii) The nominal system has at least a 30° phase margin.
(iv) The disturbance rejection requirement is met.

8.5. A control system is to be designed with reference to Fig. P8.5.

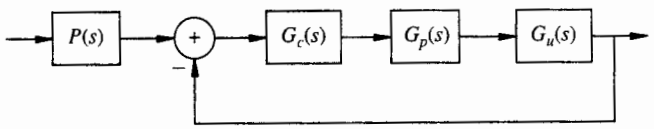


FIGURE P8.5
Problem 8.5.

where $G(s)$ is the nominal plant model.

$G_u(s)$ represents dynamics that are missing from the nominal plant model.

$G_c(s)$ and $P(s)$ are the controller and prefilter you are to design. (You may not need $P(s)$.)

Let

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

Design a controller to meet the following specifications:

1. If there is an output disturbance with a frequency below $\omega = 0.01$ rad/sec, its effect on the output must be less than 0.001 times its original magnitude.
2. The nominal phase margin should be greater than 30°.
3. The closed-loop response of the nominal system to a unit step input should have zero steady-state error. It should have less than 25 percent overshoot in the transient and it should settle to within 5 percent of its final value in 10 sec.
4. The closed-loop system should be stable in the presence of either of the following representations of unmodeled dynamics.

$$G_{u1}(s) = \frac{100}{s^2 + 0.1s + 100}$$

$$G_{u2}(s) = \frac{10{,}000}{s^2 + 0.01s + 10{,}000}$$

8.6. Unstable poles and non-minimum phase zeros limit the ability to attain good control performance even when there is no modeling error in the plant.

(*a*) Consider the unstable plant whose transfer function is

$$G_{p1}(s) = \frac{100}{s - 10}$$

Sketch the Bode phase and magnitude plot and the Nyquist plot for $G_{p1}(s)$. Is there a simple controller that stabilizes this plant? If there is, describe one such controller. If there is not, explain why?

(*b*) Now assume that there is a non-minimum phase zero in the plant so that

$$G_{p2}(s) = \frac{100\left(\frac{-s}{z} + 1\right)}{(s-5)(s+20)}$$

If z is large, $G_{p2}(s)$ has a right half-plane zero far from the origin. How will the job of stabilizing $G_{p2}(s)$ compare to the job of stabilizing $G_{p1}(s)$ if z is greater than 1000? Will it be much harder, harder, the same, easier or much easier? Explain your answer.

(*c*) If z is small, the zero is now close to the origin. Sketch the Bode magnitude and phase plots for $G_{p2}(s)$ when $z = 1$. Add these plots to those of (*a*). Sketch the Nyquist plot for $G_{p2}(s)$ with $z = 1$. Can this plant be stabilized with a simple controller? Explain how or why not. (Don't design the controller specifically.) Compare this situation to that of (*a*) and write conclusions.

8.7. We know from Chap. 3 that the poles of $H_{eq}(s)$ are zeros of the plant. We also know that it is acceptable to cancel plant zeros if the zeros are in the left half-plane and away from the $j\omega$ axis. This implies that a state-variable pole placement controller can be turned into a realizable transfer function feedback controller if it is acceptable to cancel all th zeros of the plant.

Given the plant in state-variable form with

$$A = \begin{bmatrix} 1 & -1 & 2 \\ -52 & 12 & -26 \\ -53 & 13 & -27 \end{bmatrix} \quad b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \quad c = \begin{bmatrix} 9 \\ 16 \\ -15 \end{bmatrix}$$

(*a*) Find the state-variable feedback controller that places the closed-loop poles at $s =$ -10, $s = -15 + j15$ and $s = -15 - j15$.

(*b*) Find the $H_{eq}(s)$ associated with this controller.

(*c*) Add enough distant poles to $H_{eq}(s)$ to make it into $H(s)$, a realizable transfer functio that reacts similarly to $H_{eq}(s)$.

(*d*) Find the closed-loop poles using the new $H(s)$ in feedback with the plant.

8.8. (*CAD Problem*) Ziegler-Nichols tuning laws for PID compensators

It is quite fortuitous that the beneficial properties of feedback control are ofte obtained when a favorable transient response is obtained. The following conditions lea to a favorable transient response in a simple series compensator; they also are favorabl properties to produce a loop gain with low sensitivity:

1. An integrator in the loop indicates zero steady-state error for constant referenc inputs. (An integrator in the loop indicates an infinite loop gain at dc, which indicate complete steady-state rejection of constant disturbances.)
2. A fast transient response indicates a large bandwidth. (A large bandwidth indicate that the sensitivity function is kept small for a large range of frequencies.)
3. A non-oscillatory transient response indicates large phase margin at crossover. (A large phase margin indicates that the sensitivity function does not have a large peak near the crossover frequency.)
4. A first order step response with zero steady-state error indicates that the closed-loop system's magnitude response is near unity for low frequencies, which indicates a large loop gain at low frequencies. (A large loop gain at low frequencies indicates low sensitivity at low frequencies.)

A large number of control loops operating today, especially in the process control industry, are not designed by mathematical manipulations on a model. Instead, a simple PID compensator is implemented and the three parameters associated with the PID compensator are *tuned* to provide an acceptable transient response. The disturbance rejection and sensitivity reduction benefits are achieved from the structure of the controller that includes an integrator. The results listed above indicate that, for a simple controller, a good transient response leads to good loop properties.

In 1942 two control practitioners, Ziegler and Nichols, produced a prescription for tuning P, PI, and PID compensators. These tuning rules bear their names today. The prescription that Ziegler and Nichols produced is summarized here.

Set up a proportional compensator, i.e., let $G_c(s) = K_p$. Starting from very small K_p increase K_p until the output of the system in response to a unit step input becomes a sinusoid whose amplitude does not grow with time but also does not decay with time. Call the value of K that produces this condition of borderline instability K_{max}. Measure the period of the oscillation and call this value T_p. Set up a P, PI, or PID compensator according to the following equation and Table P8.1.

$$G_c(s) = K\left(1 + \frac{1}{T_I s} + \frac{T_D s}{1 + (s/10T_D)}\right)$$

The pole at $10T_D$ in the derivative compensator term is there only to make the compensator realizable. It has little effect and is largely ignored in this analysis.

TABLE P8.1
Ziegler-Nichols PID tuning rules

Controller type	K	T_I	T_D
P	$0.5K_{\max}$	∞	0
PI	$0.45K_{\max}$	$Tp/1.2$	0
PID	$0.6K_{\max}$	$Tp/2$	$Tp/8$

The Ziegler-Nichols rules work well on plants whose step response closely matches that of a simple first-order system with a delay. Since time can be scaled, the important parameter of such a system is the speed of the pole in relation to the length of the delay. Therefore, we can analyze an entire group of plants of this type by assuming a one time unit delay and moving the pole relative to the delay length. Let

$$G_p(s) = \frac{a}{(s+a)}e^{-s}$$

(*a*) A PID compensator is equivalent to a lead-lag compensator. Without regard to the Ziegler-Nichols rules, design a lead-lag compensator for the plant, $G_p(s)$, with $a = 0.01$. Make the bandwidth as large as you can without letting the sensitivity function become greater than 1.5.

Now let's analyze the Ziegler-Nichols proportional control laws for this plant.

(*b*) Plot the Bode plots for $G_p(s)$ with $a = 0.01$, 0.1, 1,and 10. From the Bode plots determine the $K_{\max}$ and T_p that result from the Ziegler-Nichols experiment for each value of a. For each value of a, find the phase margin that results from the Ziegler-Nichols proportional compensator. What can you say about the expected transient responses for the Ziegler-Nichols proportional compensators for various values of a? If the dc gain of $G_p(s)$ were to double while the pole position and the delay remained the same, what happens to the gains K of the Ziegler-Nichols *proportional* compensators? What happens to the loop gains that result? What is the relation between T_p and the crossover frequency ω_c when $a = 0.01$ and $a = 0.1$? Can you argue that there is a general approximate relationship between T_p and ω_c for the Ziegler-Nichols *proportional* compensators when a is small compared to the delay?

In (*b*) you should have convinced yourself that the Ziegler-Nichols rules result in a reasonable loop gain transfer function as long as the plant is well modeled by the $G_p(s)$ given above. This is true for all values of the pole position, the delay and the dc gain of the plant. The Ziegler-Nichols PI compensators behave similarly to the proportional compensators near crossover frequency but add much low frequency gain with the integral action.

(*c*) Using the values of $K_{\max}$ and T_p found in (*b*), plot the Bode plots of prescribed Ziegler-Nichols PI compensators for each of the four values of the pole position parameter a. The zero of the PI compensator is set based upon T_p. For each value of a find the relationship between the zero of the PI compensator and the crossover frequency? When a is small how does this result relate to the relationship between T_p and ω_c found in (*b*)? For each a how much phase lag does the addition of the integral term cause at the ω_c associated with the proportional compensator? How is this additional phase lag counteracted?

(*d*) Using the values of $K_{\max}$ and T_p found in (*b*), plot the Bode plots of prescribed Ziegler-Nichols PID compensators for each of the four values of the pole position parameter a. Why can K be increased for the PID compensators over its value for the proportional compensators? How does the Ziegler-Nichols PID compensator for $a = 0.01$ compare with your lead-lag design of (*a*)? If a CAD tool is available plot the unit step responses for each of the Ziegler-Nichols PID compensators. To do this more easily replace the unit delay by the third-order Padé approximation for a unit delay given in Eq. (8.10-15).

8.9. (*CAD Problem*) Control of airplane phugoid mode

The transfer function between the differential stick position which commands the elevator position $\Delta_e(s)$, and the differential pitch angle $\Theta(s)$ is developed in the appendix to Chap. 2 and Problem 4.11. We consider this to be our plant.

$$G_p(s) = \frac{\Theta(s)}{\Delta_e(s)} = \frac{-0.5632s^2 - 0.3123s - 0.0102}{s^4 + 1.0480s^3 + 0.7862s^2 + 0.0193s + 0.0120}$$

Since aircraft must be designed to be light the dynamics are subject to various bending modes. These dynamics are too complicated to model precisely but their effect can be summarized with a bound on the magnitude of all possible multiplicative perturbations. Let

$$l_m(j\omega) = 0.01(s+1)^2$$

Design a series compensator for this plant that meets the following design criteria:

1. The nominal loop gain stays at least 20 dB below $l_m^{-1}(j\omega)$ for all values of ω where $l_m^{-1}(j\omega) < 0.1$.
2. Output disturbances of frequencies less than $\omega = 0.1$ are rejected to less than 10 percent of their original magnitude in steady-state. Output disturbances of frequencies less than $\omega = 0.01$ are rejected to less than 0.1 percent of their original magnitude in steady-state. Both these objectives are still met when the worst case multiplicative perturbation is present.
3. When the worst case multiplicative perturbation is present, the perturbed sensitivity function remains less than 1.3 for all frequencies.
4. The bandwidth of the system is as large as possible given the other requirements are met.

Observe and comment upon the transient response of your design to a negative step change in elevator position. Design a prefilter to remove any overshoot in the transient response while maintaining a similar speed of response. Observe and comment upon the transient response of your design to a unit step disturbance at the plant input. Comment upon the utility of the prefilter in this situation. Is the oscillatory pole pair of the phugoid actively or passively controlled? Compare this with the control of the oscillatory pair in Sec. 8.8.

(Optional) To observe the effect of your feedback loop in suppressing the phugoid oscillations during a climbing maneuver repeat Problem 4.12 using the controller designed during this problem.

8.10. (*CAD Problem*) Control of an exothermic chemical reactor

In Example 2.7-3 and Problem 3.11 the transfer function for the linearized dynamics of an exothermic chemical reactor was derived. The transfer function relates the percentage change in the flow of liquid in the cooling jacket to the percentage change in temperature inside the reaction tank. The transfer function is given by

$$\frac{Y(s)}{U(s)} = \frac{\dfrac{\delta T(s)}{\overline{T}}}{\dfrac{\delta F_J(s)}{\overline{F}_J}} = \frac{-3.36s - 5.7}{s^3 + 18.3s^2 - 43.78s - 119.8}$$

Because the time unit used in deriving the transfer function is hours, all results must be interpreted in units of radians per hour rather than the usual radians per second. Assume that the most significant dynamics that are not captured by the transfer function are the time delays involved in mixing in the tank, measurement of the temperature, and the change in the flow of the liquids after the valve positions are changed. Assume that the total of the time delays in the transfer functions is less than 20 sec or 1/180 of an hour.

(*a*) Use Eq. (5.6-21) to establish a bound on the multiplicative perturbation arising from a delay of 1/180 of an hour.

(*b*) Design a series compensator transfer function to work in this setting. Your compensator should contain integral action and as large a bandwidth as possible. It should be robustly stable with respect to the multiplicative perturbations bounded in (*a*) and the maximum over frequency of the worst case perturbed sensitivity function should be as small as possible. The transient response should be as fast and as smooth as possible with little or no oscillations or overshoot.

(*c*) After you have settled on a design, discuss its strengths and weaknesses. Explain how some aspects of the design might be improved and at what expenses to other aspects these improvements might occur.

APPENDI A

TH LAPLAC TRANSFORM— SUMMAR

This appendix is designed to serve as a brief review of the properties of the Lapla transform methods that are needed in this book. As such, it does not contain a comple treatment of all phases of the Laplace transform nor does it serve as an introducti to that subject. The reader who is not familiar with Laplace transforms is directed one of the many available textbooks for an introductory treatment.[1]

Definitions. The Laplace transform of a time function $f(t)$, written as $F(s) = \mathcal{L}[f(t)$ is defined by the following integral operation:

$$F(s) = \mathcal{L}[f(t)] = \int_0^\infty f(t)e^{-st}dt \qquad \text{(A-}$$

[1]See, for example, the books by Oppenheim, Willsky, and Young, the book by Papoulis or the book b McGillem and Cooper in the references.