Chapter 6

Performance Specifications and Limitations

In this chapter, we consider further the feedback system properties and discuss how to achieve desired performance using feedback control. We also consider the mathematical formulations of optimal $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. A key step in the optimal control design is the selection of weighting functions. We shall give some guidelines to such selection process using some SISO examples. We shall also discuss in some detail the design limitations imposed by bandwidth constraints, the open-loop right-half plane zeros, and the open-loop right-half plane poles using Bode's gain and phase relation, Bode's sensitivity integral relation, and the Poisson integral formula.

6.1 Feedback Properties

In this section, we discuss the properties of a feedback system. In particular, we consider the benefit of the feedback structure and the concept of design tradeoffs for conflicting objectives — namely, how to achieve the benefits of feedback in the face of uncertainties.

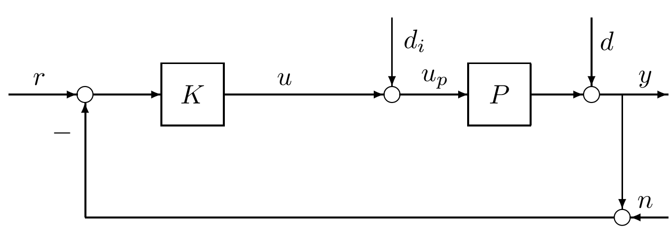

Figure 6.1: Standard feedback configuration

Consider again the feedback system shown in Figure 5.1. For convenience, the system diagram is shown again in Figure 6.1. For further discussion, it is convenient to define the *input loop transfer matrix*, L_i, and *output loop transfer matrix*, L_o, as

$$L_i = KP, \quad L_o = PK,$$

respectively, where L_i is obtained from breaking the loop at the input (u) of the plant while L_o is obtained from breaking the loop at the output (y) of the plant. The *input sensitivity* matrix is defined as the transfer matrix from d_i to u_p:

$$S_i = (I + L_i)^{-1}, \quad u_p = S_i d_i.$$

The *output sensitivity* matrix is defined as the transfer matrix from d to y:

$$S_o = (I + L_o)^{-1}, \quad y = S_o d.$$

The *input* and *output complementary sensitivity* matrices are defined as

$$T_i = I - S_i = L_i(I + L_i)^{-1}$$
$$T_o = I - S_o = L_o(I + L_o)^{-1},$$

respectively. (The word *complementary* is used to signify the fact that T is the complement of S, $T = I - S$.) The matrix $I + L_i$ is called the *input return difference matrix* and $I + L_o$ is called the *output return difference matrix*.

It is easy to see that the closed-loop system, if it is internally stable, satisfies the following equations:

$$y = T_o(r - n) + S_o P d_i + S_o d \tag{6.1}$$
$$r - y = S_o(r - d) + T_o n - S_o P d_i \tag{6.2}$$
$$u = K S_o(r - n) - K S_o d - T_i d_i \tag{6.3}$$
$$u_p = K S_o(r - n) - K S_o d + S_i d_i. \tag{6.4}$$

These four equations show the fundamental benefits and design objectives inherent in feedback loops. For example, equation (6.1) shows that the effects of disturbance d on the plant output can be made "small" by making the output sensitivity function S_o small. Similarly, equation (6.4) shows that the effects of disturbance d_i on the plant input can be made small by making the input sensitivity function S_i small. The notion of smallness for a transfer matrix in a certain range of frequencies can be made explicit using frequency-dependent singular values, for example, $\overline{\sigma}(S_o) < 1$ over a frequency range would mean that the effects of disturbance d at the plant output are effectively desensitized over that frequency range.

Hence, good disturbance rejection at the plant output (y) would require that

$$\overline{\sigma}(S_o) = \overline{\sigma}\left((I + PK)^{-1}\right) = \frac{1}{\underline{\sigma}(I + PK)} \quad \text{(for disturbance at plant output, } d\text{)},$$
$$\overline{\sigma}(S_o P) = \overline{\sigma}\left((I + PK)^{-1} P\right) = \overline{\sigma}(P S_i) \quad \text{(for disturbance at plant input, } d_i\text{)}$$

be made small and good disturbance rejection at the plant input (u_p) would require that

$$\begin{aligned}
\overline{\sigma}(S_i) &= \overline{\sigma}\left((I+KP)^{-1}\right) = \frac{1}{\underline{\sigma}(I+KP)} \quad \text{(for disturbance at plant input, } d_i\text{)}, \\
\overline{\sigma}(S_iK) &= \overline{\sigma}\left(K(I+PK)^{-1}\right) = \overline{\sigma}(KS_o) \quad \text{(for disturbance at plant output, } d\text{)}
\end{aligned}$$

be made small, particularly in the low-frequency range where d and d_i are usually significant.

Note that

$$\begin{aligned}
\underline{\sigma}(PK) - 1 &\leq \underline{\sigma}(I+PK) \leq \underline{\sigma}(PK) + 1 \\
\underline{\sigma}(KP) - 1 &\leq \underline{\sigma}(I+KP) \leq \underline{\sigma}(KP) + 1
\end{aligned}$$

then

$$\begin{aligned}
\frac{1}{\underline{\sigma}(PK)+1} \leq \overline{\sigma}(S_o) \leq \frac{1}{\underline{\sigma}(PK)-1}, &\quad \text{if } \underline{\sigma}(PK) > 1 \\
\frac{1}{\underline{\sigma}(KP)+1} \leq \overline{\sigma}(S_i) \leq \frac{1}{\underline{\sigma}(KP)-1}, &\quad \text{if } \underline{\sigma}(KP) > 1
\end{aligned}$$

These equations imply that

$$\begin{aligned}
\overline{\sigma}(S_o) \ll 1 &\iff \underline{\sigma}(PK) \gg 1 \\
\overline{\sigma}(S_i) \ll 1 &\iff \underline{\sigma}(KP) \gg 1.
\end{aligned}$$

Now suppose P and K are invertible; then

$$\begin{aligned}
\underline{\sigma}(PK) \gg 1 \text{ or } \underline{\sigma}(KP) \gg 1 &\iff \overline{\sigma}(S_oP) = \overline{\sigma}\left((I+PK)^{-1}P\right) \approx \overline{\sigma}(K^{-1}) = \frac{1}{\underline{\sigma}(K)} \\
\underline{\sigma}(PK) \gg 1 \text{ or } \underline{\sigma}(KP) \gg 1 &\iff \overline{\sigma}(KS_o) = \overline{\sigma}\left(K(I+PK)^{-1}\right) \approx \overline{\sigma}(P^{-1}) = \frac{1}{\underline{\sigma}(P)}
\end{aligned}$$

Hence good performance at plant output (y) requires, in general, large output loop gain $\underline{\sigma}(L_o) = \underline{\sigma}(PK) \gg 1$ in the frequency range where d is significant for desensitizing d and large enough controller gain $\underline{\sigma}(K) \gg 1$ in the frequency range where d_i is significant for desensitizing d_i. Similarly, good performance at plant input (u_p) requires, in general, large input loop gain $\underline{\sigma}(L_i) = \underline{\sigma}(KP) \gg 1$ in the frequency range where d_i is significant for desensitizing d_i and large enough plant gain $\underline{\sigma}(P) \gg 1$ in the frequency range where d is significant, which cannot be changed by controller design, for desensitizing d. [In general, $S_o \neq S_i$ unless K and P are square and diagonal, which is true if P is a scalar system. Hence, small $\overline{\sigma}(S_o)$ does not necessarily imply small $\overline{\sigma}(S_i)$; in other words, good disturbance rejection at the output does not necessarily mean good disturbance rejection at the plant input.]

Hence, *good multivariable feedback loop design boils down to achieving high loop (and possibly controller) gains in the necessary frequency range.*

Despite the simplicity of this statement, feedback design is by no means trivial. This is true because loop gains cannot be made arbitrarily high over arbitrarily large frequency ranges. Rather, they must satisfy certain performance tradeoff and design limitations. A major performance tradeoff, for example, concerns commands and disturbance error reduction versus stability under the model uncertainty. Assume that the plant model is perturbed to $(I+\Delta)P$ with Δ stable, and assume that the system is nominally stable (i.e., the closed-loop system with $\Delta = 0$ is stable). Now the perturbed closed-loop system is stable if

$$\det\left(I+(I+\Delta)PK\right) = \det(I+PK)\det(I+\Delta T_o)$$

has no right-half plane zero. This would, in general, amount to requiring that $\|\Delta T_o\|$ be small or that $\bar{\sigma}(T_o)$ be small at those frequencies where Δ is significant, typically at high-frequency range, which, in turn, implies that the loop gain, $\overline{\sigma}(L_o)$, should be small at those frequencies.

Still another tradeoff is with the sensor noise error reduction. The conflict between the disturbance rejection and the sensor noise reduction is evident in equation (6.1). Large $\underline{\sigma}(L_o(j\omega))$ values over a large frequency range make errors due to d small. However, they also make errors due to n large because this noise is "passed through" over the same frequency range, that is,

$$y = T_o(r-n) + S_oPd_i + S_od \approx (r-n)$$

Note that n is typically significant in the high-frequency range. Worst still, large loop gains outside of the bandwidth of P — that is, $\underline{\sigma}(L_o(j\omega)) \gg 1$ or $\underline{\sigma}(L_i(j\omega)) \gg 1$ while $\overline{\sigma}(P(j\omega)) \ll 1$ — can make the control activity (u) quite unacceptable, which may cause the saturation of actuators. This follows from

$$u = KS_o(r-n-d) - T_id_i = S_iK(r-n-d) - T_id_i \approx P^{-1}(r-n-d) - d_i$$

Here, we have assumed P to be square and invertible for convenience. The resulting equation shows that disturbances and sensor noise are actually amplified at u whenever the frequency range significantly exceeds the bandwidth of P, since for ω such that $\overline{\sigma}(P(j\omega)) \ll 1$ we have

$$\underline{\sigma}[P^{-1}(j\omega)] = \frac{1}{\overline{\sigma}[P(j\omega)]} \gg 1$$

Similarly, the controller gain, $\overline{\sigma}(K)$, should also be kept not too large in the frequency range where the loop gain is small in order not to saturate the actuators. This is because for small loop gain $\overline{\sigma}(L_o(j\omega)) \ll 1$ or $\overline{\sigma}(L_i(j\omega)) \ll 1$

$$u = KS_o(r-n-d) - T_id_i \approx K(r-n-d)$$

Therefore, it is desirable to keep $\overline{\sigma}(K)$ not too large when the loop gain is small.

To summarize the above discussion, we note that good performance requires in some frequency range, typically some low-frequency range $(0, \omega_l)$,

$$\underline{\sigma}(PK) \gg 1, \quad \underline{\sigma}(KP) \gg 1, \quad \underline{\sigma}(K) \gg 1$$

and good robustness and good sensor noise rejection require in some frequency range, typically some high-frequency range (ω_h, ∞),

$$\overline{\sigma}(PK) \ll 1, \quad \overline{\sigma}(KP) \ll 1, \quad \overline{\sigma}(K) \leq M$$

where M is not too large. These design requirements are shown graphically in Figure 6.2. The specific frequencies ω_l and ω_h depend on the specific applications and the knowledge one has of the disturbance characteristics, the modeling uncertainties, and the sensor noise levels.

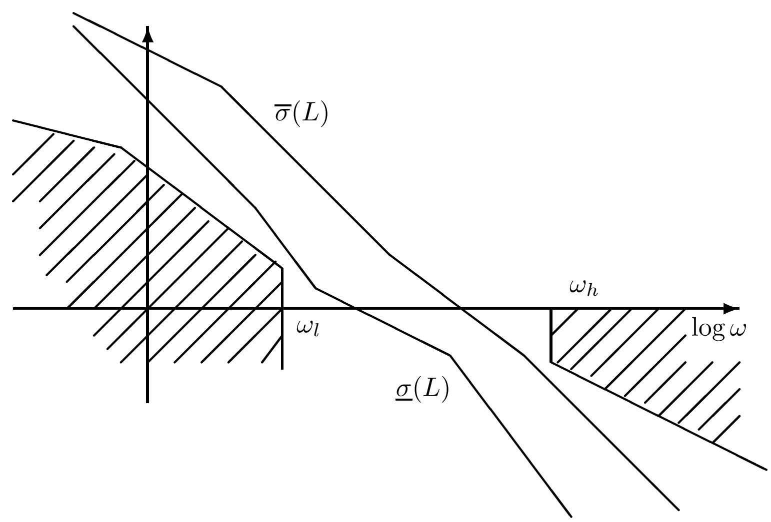

Figure 6.2: Desired loop gain

6.2 Weighted $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Performance

In this section, we consider how to formulate some performance objectives into mathematically tractable problems. As shown in Section 6.1, the performance objectives of a feedback system can usually be specified in terms of requirements on the sensitivity functions and/or complementary sensitivity functions or in terms of some other closed-loop transfer functions. For instance, the performance criteria for a scalar system may be specified as requiring

$$\begin{cases} |S(j\omega)| \leq \varepsilon, & \forall \omega \leq \omega_0, \\ |S(j\omega)| \leq M, & \forall \omega > \omega_0 \end{cases}$$

where $S(j\omega) = 1/(1 + P(j\omega)K(j\omega))$. However, it is much more convenient to reflect the system performance objectives by choosing appropriate weighting functions. For example, the preceding performance objective can be written as

$$|W_e(j\omega)S(j\omega)| \le 1, \quad \forall\omega$$

with

$$|W_e(j\omega)| = \begin{cases} 1/\varepsilon, & \forall\omega \le \omega_0 \\ 1/M, & \forall\omega > \omega_0 \end{cases}$$

To use W_e in control design, a rational transfer function $W_e(s)$ is usually used to approximate the foregoing frequency response.

The advantage of using weighted performance specifications is obvious in multivariable system design. First, some components of a vector signal are usually more important than others. Second, each component of the signal may not be measured in the same units; for example, some components of the output error signal may be measured in terms of length, and others may be measured in terms of voltage. Therefore, weighting functions are essential to make these components comparable. Also, we might be primarily interested in rejecting errors in a certain frequency range (for example, low frequencies); hence some frequency-dependent weights must be chosen.

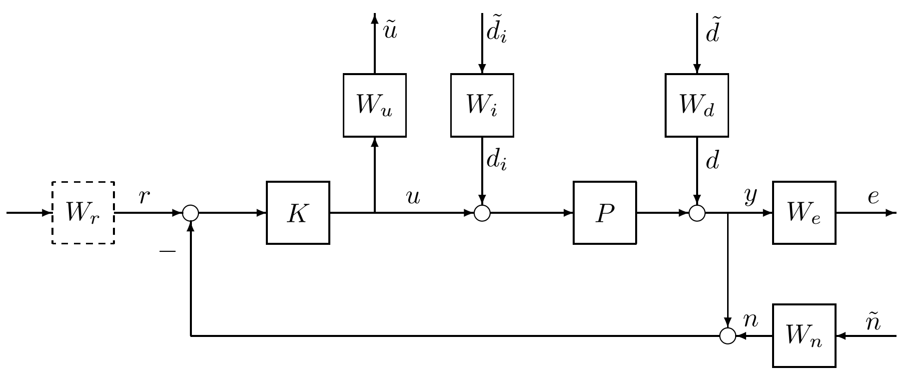

Figure 6.3: Standard feedback configuration with weights

In general, we shall modify the standard feedback diagram in Figure 6.1 into Figure 6.3. The weighting functions in Figure 6.3 are chosen to reflect the design objectives and knowledge of the disturbances and sensor noise. For example, W_d and W_i may be chosen to reflect the frequency contents of the disturbances d and d_i or they may be used to model the disturbance power spectrum depending on the nature of signals involved in the practical systems. The weighting matrix W_n is used to model the frequency contents of the sensor noise while W_e may be used to reflect the requirements on the shape of certain closed-loop transfer functions (for example, the shape of the output sensitivity function). Similarly, W_u may be used to reflect some restrictions on the control or actuator signals, and the dashed precompensator W_r is an optional element used to achieve deliberate command shaping or to represent a nonunity feedback system in equivalent unity feedback form.

It is, in fact, essential that some appropriate weighting matrices be used in order to utilize the optimal control theory discussed in this book (i.e., $\mathcal{H}_2$ and $\mathcal{H}_\infty$ theory). So a very important step in the controller design process is to choose the appropriate weights, W_e, W_d, W_u, and possibly W_n, W_i, W_r. The appropriate choice of weights for a particular practical problem is not trivial. In many occasions, as in the scalar case, the weights are chosen purely as a design parameter without any physical bases, so these weights may be treated as tuning parameters that are chosen by the designer to achieve the best compromise between the conflicting objectives. The selection of the weighting matrices should be guided by the expected system inputs and the relative importance of the outputs.

Hence, control design may be regarded as a process of choosing a controller K such that certain weighted signals are made small in some sense. There are many different ways to define the smallness of a signal or transfer matrix, as we have discussed in the last chapter. Different definitions lead to different control synthesis methods, and some are much harder than others. A control engineer should make a judgment of the mathematical complexity versus engineering requirements.

Next, we introduce two classes of performance formulations: $\mathcal{H}_2$ and $\mathcal{H}_\infty$ criteria. For the simplicity of presentation, we shall assume that $d_i = 0$ and $n = 0$.

$\mathcal{H}_2$ Performance

Assume, for example, that the disturbance $\tilde{d}$ can be approximately modeled as an impulse with random input direction; that is,

$$\tilde{d}(t) = \eta\delta(t)$$

and

$$E(\eta\eta^*) = I$$

where E denotes the expectation. We may choose to minimize the expected energy of the error e due to the disturbance $\tilde{d}$:

$$E\left\{\|e\|_2^2\right\} = E\left\{\int_0^\infty \|e\|^2 \, dt\right\} = \|W_e S_o W_d\|_2^2$$

In general, a controller minimizing only the above criterion can lead to a very large control signal u that could cause saturation of the actuators as well as many other undesirable problems. Hence, for a realistic controller design, it is necessary to include the control signal u in the cost function. Thus, our design criterion would usually be something like this:

$$E\left\{\|e\|_2^2 + \rho^2 \|\tilde{u}\|_2^2\right\} = \left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_u K S_o W_d \end{bmatrix} \right\|_2^2$$

with some appropriate choice of weighting matrix W_u and scalar ρ. The parameter ρ clearly defines the tradeoff we discussed earlier between good disturbance rejection at the output and control effort (or disturbance and sensor noise rejection at the actuators). Note that ρ can be set to $\rho = 1$ by an appropriate choice of W_u. This problem can be viewed as minimizing the *energy* consumed by the system in order to reject the disturbance d.

This type of problem was the dominant paradigm in the 1960s and 1970s and is usually referred to as linear quadratic Gaussian control, or simply as LQG. (Such problems will also be referred to as $\mathcal{H}_2$ mixed-sensitivity problems for consistency with the $\mathcal{H}_\infty$ problems discussed next.) The development of this paradigm stimulated extensive research efforts and is responsible for important technological innovation, particularly in the area of estimation. The theoretical contributions include a deeper understanding of linear systems and improved computational methods for complex systems through state-space techniques. The major limitation of this theory is the lack of formal treatment of uncertainty in the plant itself. By allowing only additive noise for uncertainty, the stochastic theory ignored this important practical issue. Plant uncertainty is particularly critical in feedback systems. (See Paganini [1995,1996] for some recent results on robust $\mathcal{H}_2$ control theory.)

$\mathcal{H}_\infty$ Performance

Although the $\mathcal{H}_2$ norm (or $\mathcal{L}_2$ norm) may be a meaningful performance measure and although LQG theory can give efficient design compromises under certain disturbance and plant assumptions, the $\mathcal{H}_2$ norm suffers a major deficiency. This deficiency is due to the fact that the tradeoff between disturbance error reduction and sensor noise error reduction is not the only constraint on feedback design. The problem is that these performance tradeoffs are often overshadowed by a second limitation on high loop gains — namely, the requirement for tolerance to uncertainties. Though a controller may be designed using FDLTI models, the design must be implemented and operated with a real physical plant. The properties of physical systems (in particular, the ways in which they deviate from finite-dimensional linear models) put strict limitations on the frequency range over which the loop gains may be large.

A solution to this problem would be to put explicit constraints on the loop gain in the cost function. For instance, one may chose to minimize

$$\sup_{\|\tilde{d}\|_2 \le 1} \|e\|_2 = \|W_e S_o W_d\|_\infty$$

subject to some restrictions on the control energy or control bandwidth:

$$\sup_{\|\tilde{d}\|_2 \le 1} \|\tilde{u}\|_2 = \|W_u K S_o W_d\|_\infty$$

Or, more frequently, one may introduce a parameter ρ and a mixed criterion

$$\sup_{\|\tilde{d}\|_2 \le 1} \left\{ \|e\|_2^2 + \rho^2 \|\tilde{u}\|_2^2 \right\} = \left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_u K S_o W_d \end{bmatrix} \right\|_\infty^2$$

Alternatively, if the system robust stability margin is the major concern, the weighted complementary sensitivity has to be limited. Thus the whole cost function may be

$$\left\| \begin{bmatrix} W_e S_o W_d \\ \rho W_1 T_o W_2 \end{bmatrix} \right\|_\infty$$

where W_1 and W_2 are the frequency-dependent uncertainty scaling matrices. These design problems are usually called $\mathcal{H}_\infty$ mixed-sensitivity problems. For a scalar system, an $\mathcal{H}_\infty$ norm minimization problem can also be viewed as minimizing the maximum magnitude of the system's steady-state response with respect to the worst-case sinusoidal inputs.

6.3 Selection of Weighting Functions

The selection of weighting functions for a specific design problem often involves ad hoc fixing, many iterations, and fine tuning. It is very hard to give a general formula for the weighting functions that will work in every case. Nevertheless, we shall try to give some guidelines in this section by looking at a typical SISO problem.

Consider an SISO feedback system shown in Figure 6.1. Then the tracking error is $e = r - y = S(r - d) + Tn - SPd_i$. So, as we have discussed earlier, we must keep $|S|$ small over a range of frequencies, typically low frequencies where r and d are significant. To motivate the choice of our performance weighting function W_e, let $L = PK$ be a standard second-order system

$$L = \frac{\omega_n^2}{s(s + 2\xi\omega_n)}$$

It is well-known from the classical control theory that the quality of the (step) time response can be quantified by rise time t_r, settling time t_s, and percent overshoot $100M_p\%$. Furthermore, these performance indices can be approximately calculated as

$$t_r \approx \frac{0.6 + 2.16\xi}{\omega_n},\ 0.3 \le \xi \le 0.8;\ t_s \approx \frac{4}{\xi\omega_n};\ M_p = e^{-\frac{\pi\xi}{\sqrt{1-\xi^2}}},\ 0 < \xi < 1$$

The key points to note are that (1) the speed of the system response is proportional to ω_n and (2) the overshoot of the system response is determined only by the damping ratio ξ. It is well known that the frequency ω_n and the damping ratio ξ can be essentially captured in the frequency domain by the open-loop crossover frequency and the phase margin or the bandwidth and the resonant peak of the closed-loop complementary sensitivity function T.

Since our performance objectives are closely related to the sensitivity function, we shall consider in some detail how these time domain indices or, equivalently, ω_n and ξ are related to the frequency response of the sensitivity function

$$S = \frac{1}{1 + L} = \frac{s(s + 2\xi\omega_n)}{s^2 + 2\xi\omega_n s + \omega_n^2}$$

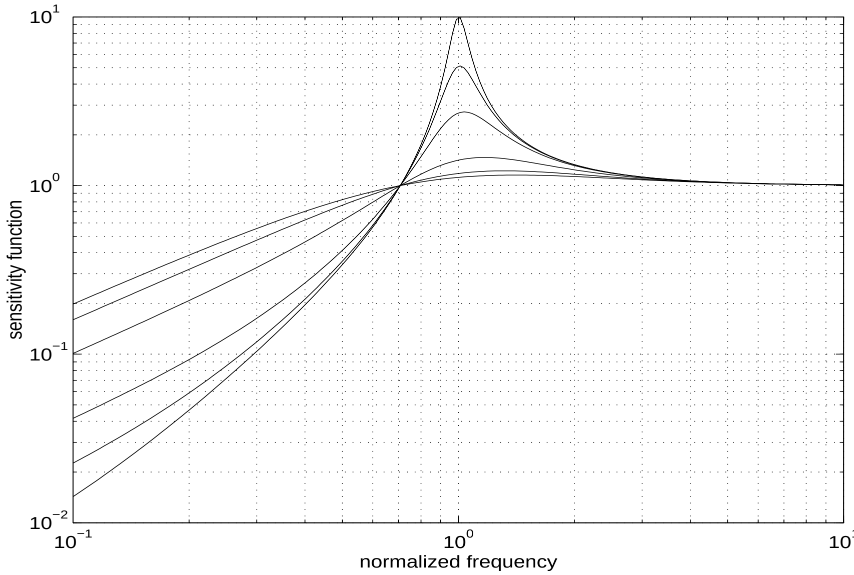

Figure 6.4: Sensitivity function S for $\xi = 0.05, 0.1, 0.2, 0.5, 0.8$, and 1 with normalized frequency (ω/ω_n)

The frequency response of the sensitivity function S is shown in Figure 6.4. Note that $|S(j\omega_n/\sqrt{2})| = 1$. We can regard the closed-loop bandwidth $\omega_b \approx \omega_n/\sqrt{2}$, since beyond this frequency the closed-loop system will not be able to track the reference and the disturbance will actually be amplified.

Next, note that

$$M_s := \|S\|_\infty = |S(j\omega_{\max})| = \frac{\alpha\sqrt{\alpha^2 + 4\xi^2}}{\sqrt{(1-\alpha^2)^2 + 4\xi^2\alpha^2}}$$

where $\alpha = \sqrt{0.5 + 0.5\sqrt{1+8\xi^2}}$ and $\omega_{\max} = \alpha\omega_n$. For example, $M_s = 5.123$ when $\xi = 0.1$. The relationship between ξ and M_s is shown in Figure 6.5. It is clear that the overshoot can be excessive if M_s is large. Hence a good control design should not have a very large M_s.

Now suppose we are given the time domain performance specifications then we can determine the corresponding requirements in frequency domain in terms of the bandwidth ω_b and the peak sensitivity M_s. Hence a good control design should result in a sensitivity function S satisfying both the bandwidth ω_b and the peak sensitivity M_s requirements, as shown in Figure 6.6. These requirements can be approximately represented as

$$|S(s)| \le \left|\frac{s}{s/M_s + \omega_b}\right|, \ s = j\omega, \ \forall\ \omega$$

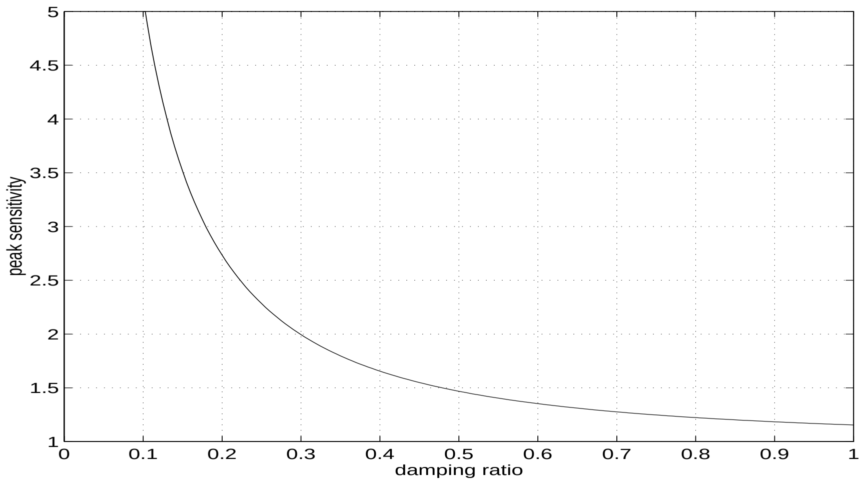

Figure 6.5: Peak sensitivity M_s versus damping ratio ξ

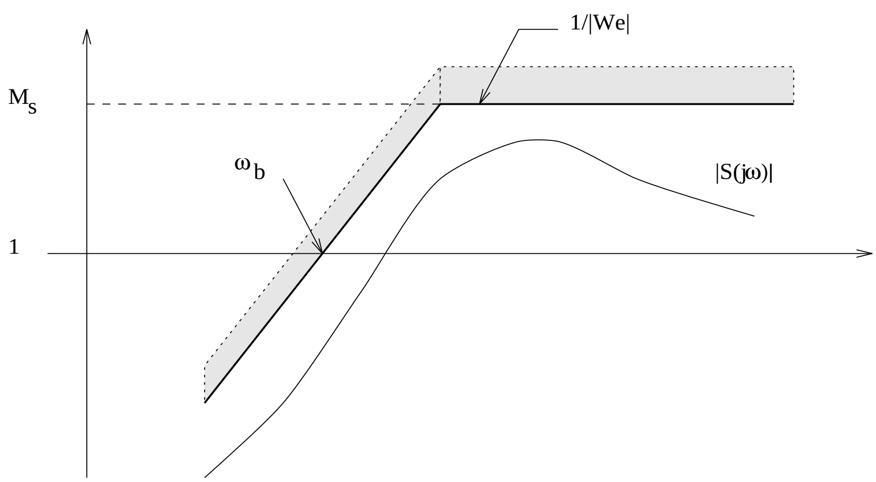

Figure 6.6: Performance weight W_e and desired S

Or, equivalently, $|W_e S| \leq 1$ with

$$W_e = \frac{s/M_s + \omega_b}{s} \tag{6.5}$$

The preceding discussion applies in principle to most control design and hence the preceding weighting function can, in principle, be used as a candidate weighting function in an initial design. Since the steady-state error with respect to a step input is given by $|S(0)|$, it is clear that $|S(0)| = 0$ if the closed-loop system is stable and $\|W_e S\|_\infty < \infty$.

Unfortunately, the optimal control techniques described in this book cannot be used *directly* for problems with such weighting functions since these techniques assume that all unstable poles of the system (including plant and all performance and control weighting functions) are stabilizable by the control and detectable from the measurement outputs, which is clearly not satisfied if W_e has an imaginary axis pole since W_e is not detectable from the measurement. We shall discuss in Chapter 14 how such problems can be reformulated so that the techniques described in this book can be applied. A theory dealing directly with such problems is available but is much more complicated both theoretically and computationally and does not seem to offer much advantage.

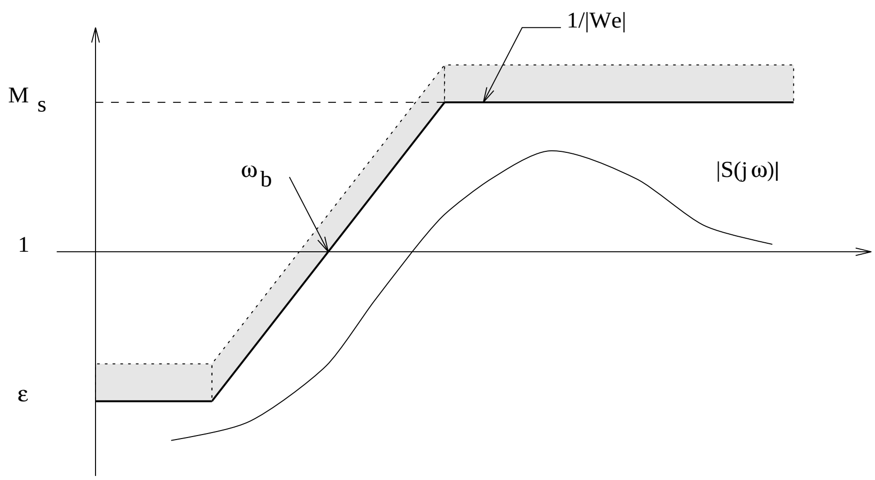

Figure 6.7: Practical performance weight W_e and desired S

Now instead of perfect tracking for step input, suppose we only need the steady-state error with respect to a step input to be no greater than ϵ (i.e., $|S(0)| \leq \epsilon$); then it is sufficient to choose a weighting function W_e satisfying $|W_e(0)| \geq 1/\varepsilon$ so that $\|W_e S\|_\infty \leq 1$ can be achieved. A possible choice of W_e can be obtained by modifying the weighting function in equation (6.5):

$$W_e = \frac{s/M_s + \omega_b}{s + \omega_b \varepsilon} \tag{6.6}$$

Hence, for practical purpose, one can usually choose a suitable ε, as shown in Figure 6.7, to satisfy the performance specifications. If a steeper transition between low-frequency and high-frequency is desired, the weight W_e can be modified as follows:

$$W_e = \left(\frac{s/\sqrt[k]{M_s} + \omega_b}{s + \omega_b \sqrt[k]{\varepsilon}} \right)^k \tag{6.7}$$

for some integer $k \geq 1$.

The selection of control weighting function W_u follows similarly from the preceding discussion by considering the control signal equation

$$u = KS(r - n - d) - Td_i$$

The magnitude of $|KS|$ in the low-frequency range is essentially limited by the allowable cost of control effort and saturation limit of the actuators; hence, in general, the maximum gain M_u of KS can be fairly large, while the high-frequency gain is essentially limited by the controller bandwidth (ω_{bc}) and the (sensor) noise frequencies. Ideally, one would like to roll off as fast as possible beyond the desired control bandwidth so that the high-frequency noises are attenuated as much as possible. Hence a candidate weight W_u would be

$$W_u = \frac{s + \omega_{bc}/M_u}{\omega_{bc}} \tag{6.8}$$

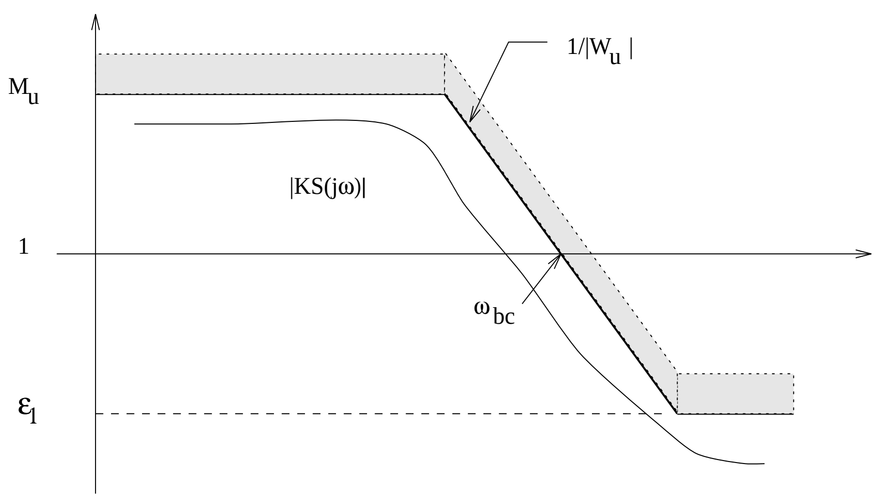

Figure 6.8: Control weight W_u and desired KS

However, again the optimal control design techniques developed in this book cannot be applied directly to a problem with an improper control weighting function. Hence we shall introduce a far away pole to make W_u proper:

$$W_u = \frac{s + \omega_{bc}/M_u}{\varepsilon_1 s + \omega_{bc}} \tag{6.9}$$

for a small $\varepsilon_1 > 0$, as shown in Figure 6.8. Similarly, if a faster rolloff is desired, one may choose

$$W_u = \left(\frac{s + \omega_{bc}/\sqrt[k]{M_u}}{\sqrt[k]{\varepsilon_1} s + \omega_{bc}} \right)^k \tag{6.10}$$

for some integer $k \geq 1$.

The weights for MIMO problems can be initially chosen as diagonal matrices with each diagonal term chosen in the foregoing form.

6.4 Bode's Gain and Phase Relation

One important problem that arises frequently is concerned with the level of performance that can be achieved in feedback design. It has been shown in Section 6.1 that the feedback design goals are inherently conflicting, and a tradeoff must be performed among different design objectives. It is also known that the fundamental requirements, such as stability and robustness, impose inherent limitations on the feedback properties irrespective of design methods, and the design limitations become more severe in the presence of right-half plane zeros and poles in the open-loop transfer function.

In the classical feedback theory, Bode's gain-phase integral relation (see Bode [1945]) has been used as an important tool to express design constraints in scalar systems. This integral relation says that the phase of a stable and minimum phase transfer function is determined uniquely by the magnitude of the transfer function. More precisely, let $L(s)$ be a stable and minimum phase transfer function: then

$$\angle L(j\omega_0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{d \ln |L|}{d\nu} \ln \coth \frac{|\nu|}{2} d\nu \tag{6.11}$$

where $\nu := \ln(\omega/\omega_0)$. The function $\ln \coth \frac{|\nu|}{2} = \ln \frac{e^{|\nu|/2} + e^{-|\nu|/2}}{e^{|\nu|/2} - e^{-|\nu|/2}}$ is plotted in Figure 6.9.

Note that $\ln \coth \frac{|\nu|}{2}$ decreases rapidly as ω deviates from ω_0 and hence the integral depends mostly on the behavior of $\frac{d \ln |L(j\omega)|}{d\nu}$ near the frequency ω_0. This is clear from the following integration:

$$\frac{1}{\pi} \int_{-\alpha}^{\alpha} \ln \coth \frac{|\nu|}{2} d\nu = \begin{cases} 1.1406 \text{ (rad)}, & \alpha = \ln 3 \\ 1.3146 \text{ (rad)}, & \alpha = \ln 5 \\ 1.443 \text{ (rad)}, & \alpha = \ln 10 \end{cases} = \begin{cases} 65.3^{\mathrm{o}}, & \alpha = \ln 3 \\ 75.3^{\mathrm{o}}, & \alpha = \ln 5 \\ 82.7^{\mathrm{o}}, & \alpha = \ln 10. \end{cases}$$

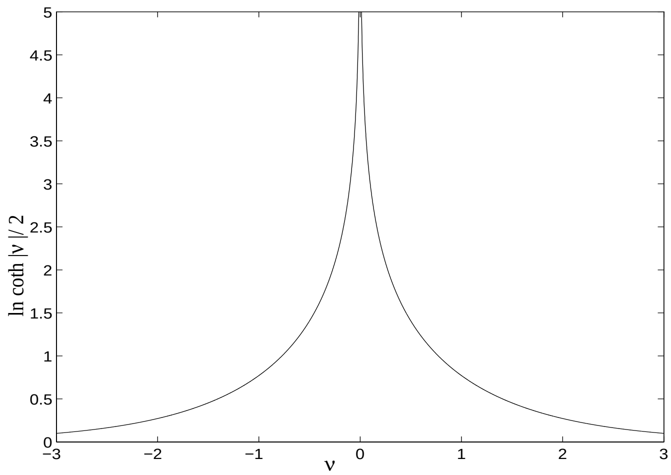

Figure 6.9: The function $\ln\coth\dfrac{|\nu|}{2}$ vs ν

Note that $\dfrac{d\ln|L(j\omega)|}{d\nu}$ is the slope of the Bode plot, which is generally negative for almost all frequencies. It follows that $\angle L(j\omega_0)$ will be large if the gain L attenuates slowly near ω_0 and small if it attenuates rapidly near ω_0. For example, suppose the slope $\dfrac{d\ln|L(j\omega)|}{d\nu} = -\ell$; that is, $(-20\ell$ dB per decade), in the neighborhood of ω_0; then it is reasonable to expect

$$\angle L(j\omega_0) < \begin{cases} -\ell\times 65.3^{\mathrm{o}}, & \text{if the slope of } L=-\ell \text{ for } \frac{1}{3}\le\frac{\omega}{\omega_0}\le 3 \\ -\ell\times 75.3^{\mathrm{o}}, & \text{if the slope of } L=-\ell \text{ for } \frac{1}{5}\le\frac{\omega}{\omega_0}\le 5 \\ -\ell\times 82.7^{\mathrm{o}}, & \text{if the slope of } L=-\ell \text{ for } \frac{1}{10}\le\frac{\omega}{\omega_0}\le 10. \end{cases}$$

The behavior of $\angle L(j\omega)$ is particularly important near the crossover frequency ω_c, where $|L(j\omega_c)|=1$ since $\pi+\angle L(j\omega_c)$ is the phase margin of the feedback system. Further, the return difference is given by

$$|1+L(j\omega_c)| = |1+L^{-1}(j\omega_c)| = 2\left|\sin\frac{\pi+\angle L(j\omega_c)}{2}\right|,$$

which must not be too small for good stability robustness. If $\pi+\angle L(j\omega_c)$ is forced to be very small by rapid gain attenuation, the feedback system will amplify disturbances and exhibit little uncertainty tolerance at and near ω_c. Since it is generally required that the loop transfer function L roll off as fast as possible in the high-frequency range, it is reasonable to expect that $\angle L(j\omega_c)$ is at most $-\ell\times 90^{\mathrm{o}}$ if the slope of $L(j\omega)$ is $-\ell$ near ω_c. Thus it is important to keep the slope of L near ω_c not much smaller than -1 for a reasonably wide range of frequencies in order to guarantee some reasonable performance. The conflict between attenuation rate and loop quality near crossover is thus clearly evident.

Bode's gain and phase relation can be extended to stable and nonminimum phase transfer functions easily. Let $z_1, z_2, \ldots, z_k$ be the right-half plane zeros of $L(s)$, then L can be factorized as

$$L(s) = \frac{-s+z_1}{s+z_1}\frac{-s+z_2}{s+z_2}\cdots\frac{-s+z_k}{s+z_k}L_{\mathrm{mp}}(s)$$

where L_{mp} is stable and minimum phase and $|L(j\omega)| = |L_{\mathrm{mp}}(j\omega)|$. Hence

$$\begin{aligned}\angle L(j\omega_0) &= \angle L_{\mathrm{mp}}(j\omega_0) + \angle\prod_{i=1}^{k}\frac{-j\omega_0+z_i}{j\omega_0+z_i}\\ &= \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d\ln|L_{\mathrm{mp}}|}{d\nu}\ln\coth\frac{|\nu|}{2}d\nu + \sum_{i=1}^{k}\angle\frac{-j\omega_0+z_i}{j\omega_0+z_i},\end{aligned}$$

which gives

$$\angle L(j\omega_0) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d\ln|L|}{d\nu}\ln\coth\frac{|\nu|}{2}d\nu + \sum_{i=1}^{k}\angle\frac{-j\omega_0+z_i}{j\omega_0+z_i}. \tag{6.12}$$

Since $\angle\frac{-j\omega_0+z_i}{j\omega_0+z_i} \leq 0$ for each i, a nonminimum phase zero contributes an additional phase lag and imposes limitations on the rolloff rate of the open-loop gain. For example, suppose L has a zero at $z > 0$; then

$$\phi_1(\omega_0/z) := \angle\frac{-j\omega_0+z}{j\omega_0+z}\bigg|_{\omega_0=z,z/2,z/4} = -90^{\mathrm{o}}, -53.13^{\mathrm{o}}, -28^{\mathrm{o}},$$

as shown in Figure 6.10. Since the slope of $|L|$ near the crossover frequency is, in general, no greater than -1, which means that the phase due to the minimum phase part, L_{mp}, of L will, in general, be no greater than -90^{o}, the crossover frequency (or the closed-loop bandwidth) must satisfy

$$\omega_c < z/2 \tag{6.13}$$

in order to guarantee the closed-loop stability and some reasonable closed-loop performance.

Next suppose L has a pair of complex right-half zeros at $z = x \pm jy$ with $x > 0$; then

$$\begin{aligned}\phi_2(\omega_0/|z|) &:= \angle\frac{-j\omega_0+z}{j\omega_0+z}\,\frac{-j\omega_0+\bar{z}}{j\omega_0+\bar{z}}\bigg|_{\omega_0=|z|,|z|/2,|z|/3,|z|/4}\\ &\approx \begin{cases} -180^{\mathrm{o}}, & -106.26^{\mathrm{o}}, & -73.7^{\mathrm{o}}, & -56^{\mathrm{o}}, & \mathrm{Re}(z) \gg \Im(z)\\ -180^{\mathrm{o}}, & -86.7^{\mathrm{o}}, & -55.9^{\mathrm{o}}, & -41.3^{\mathrm{o}}, & \mathrm{Re}(z) \approx \Im(z)\\ -360^{\mathrm{o}}, & 0^{\mathrm{o}}, & 0^{\mathrm{o}}, & 0^{\mathrm{o}}, & \mathrm{Re}(z) \ll \Im(z)\end{cases}\end{aligned}$$

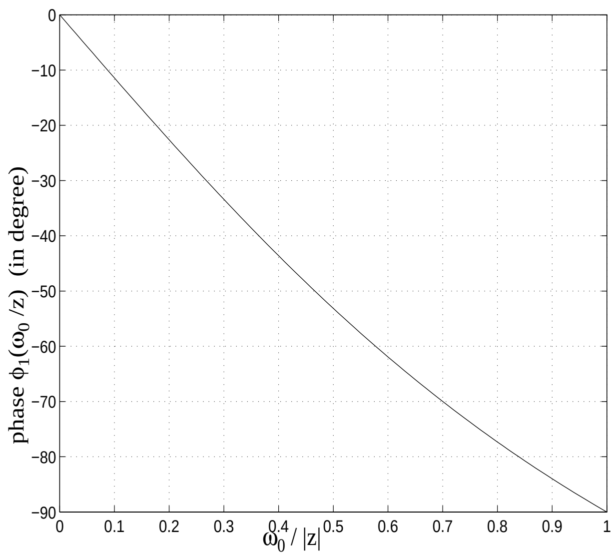

Figure 6.10: Phase $\phi_1(\omega_0/z)$ due to a real zero $z > 0$

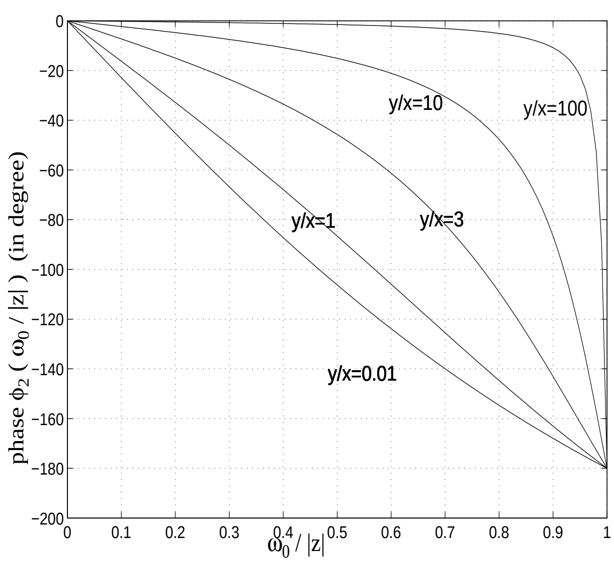

Figure 6.11: Phase $\phi_2(\omega_0/|z|)$ due to a pair of complex zeros: $z = x \pm jy$ and $x > 0$

as shown in Figure 6.11. In this case we conclude that the crossover frequency must satisfy

$$\omega_c < \begin{cases} |z|/4, & \mathrm{Re}(z) \gg \Im(z) \\ |z|/3, & \mathrm{Re}(z) \approx \Im(z) \\ |z|, & \mathrm{Re}(z) \ll \Im(z) \end{cases} \tag{6.14}$$

in order to guarantee the closed-loop stability and some reasonable closed-loop performance.

6.5 Bode's Sensitivity Integral

In this section, we consider the design limitations imposed by the bandwidth constraints and the right-half plane poles and zeros using Bode's sensitivity integral and Poisson integral. Let L be the open-loop transfer function with at least two more poles than zeros and let $p_1, p_2, \ldots, p_m$ be the open right-half plane poles of L. Then the following Bode's sensitivity integral holds:

$$\int_0^\infty \ln |S(j\omega)| d\omega = \pi \sum_{i=1}^{m} \mathrm{Re}(p_i) \tag{6.15}$$

In the case where L is stable, the integral simplifies to

$$\int_0^\infty \ln |S(j\omega)| d\omega = 0 \tag{6.16}$$

These integrals show that there will exist a frequency range over which the magnitude of the sensitivity function exceeds one if it is to be kept below one at other frequencies, as illustrated in Figure 6.12. This is the so-called water bed effect.

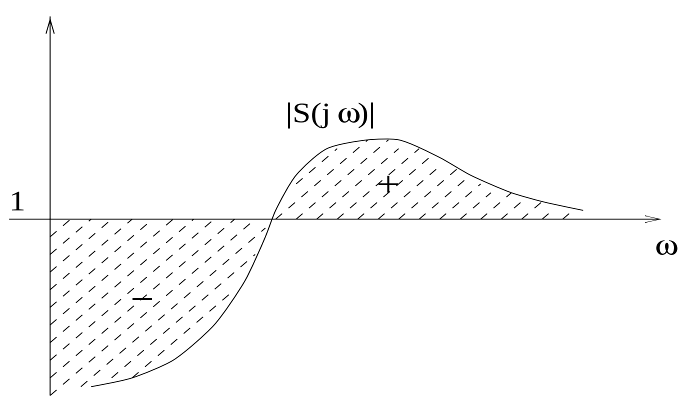

Figure 6.12: Water bed effect of sensitivity function

Suppose that the feedback system is designed such that the level of sensitivity reduction is given by

$$|S(j\omega)| \le \epsilon < 1, \quad \forall \omega \in [0,\ \omega_l]$$

where $\epsilon > 0$ is a given constant.

Bandwidth constraints in feedback design typically require that the open-loop transfer function be small above a specified frequency, and that it roll off at a rate of more than one pole-zero excess above that frequency. These constraints are commonly needed to ensure stability robustness despite the presence of modeling uncertainty in the plant model, particularly at high frequencies. One way of quantifying such bandwidth constraints is by requiring the open-loop transfer function to satisfy

$$|L(j\omega)| \le \frac{M_h}{\omega^{1+\beta}} \le \tilde{\epsilon} < 1, \quad \forall \omega \in [\omega_h,\ \infty)$$

where $\omega_h > \omega_l$, and $M_h > 0$, $\beta > 0$ are some given constants.

Note that for $\omega \ge \omega_h$,

$$|S(j\omega)| \le \frac{1}{1 - |L(j\omega)|} \le \frac{1}{1 - \frac{M_h}{\omega^{1+\beta}}}$$

and

$$\begin{aligned}
-\int_{\omega_h}^{\infty} \ln\left(1 - \frac{M_h}{\omega^{1+\beta}}\right) d\omega &= \sum_{i=1}^{\infty} \int_{\omega_h}^{\infty} \frac{1}{i}\left(\frac{M_h}{\omega^{1+\beta}}\right)^i d\omega \\
&= \sum_{i=1}^{\infty} \frac{1}{i}\frac{\omega_h}{i(1+\beta)-1}\left(\frac{M_h}{\omega_h^{1+\beta}}\right)^i \\
&\le \frac{\omega_h}{\beta}\sum_{i=1}^{\infty}\frac{1}{i}\left(\frac{M_h}{\omega_h^{1+\beta}}\right)^i = -\frac{\omega_h}{\beta}\ln\left(1 - \frac{M_h}{\omega_h^{1+\beta}}\right) \\
&\le -\frac{\omega_h}{\beta}\ln(1-\tilde{\epsilon}).
\end{aligned}$$

Then

$$\pi \sum_{i=1}^{m} \mathrm{Re}(p_i) = \int_0^{\infty} \ln|S(j\omega)| d\omega$$

$$\begin{aligned}
&= \int_0^{\omega_l} \ln|S(j\omega)|d\omega + \int_{\omega_l}^{\omega_h} \ln|S(j\omega)|d\omega + \int_{\omega_h}^{\infty} \ln|S(j\omega)|d\omega \\
&\le \omega_l \ln\epsilon + (\omega_h - \omega_l)\max_{\omega\in[\omega_l,\omega_h]} \ln|S(j\omega)| - \int_{\omega_h}^{\infty}\ln\left(1 - \frac{M_h}{\omega^{1+\beta}}\right)d\omega \\
&\le \omega_l \ln\epsilon + (\omega_h - \omega_l)\max_{\omega\in[\omega_l,\omega_h]} \ln|S(j\omega)| - \frac{\omega_h}{\beta}\ln(1-\tilde{\epsilon}),
\end{aligned}$$

which gives

$$\max_{\omega\in[\omega_l,\omega_h]} |S(j\omega)| \geq e^{\alpha}\left(\frac{1}{\epsilon}\right)^{\frac{\omega_l}{\omega_h-\omega_l}}(1-\tilde{\epsilon})^{\frac{\omega_h}{\beta(\omega_h-\omega_l)}}$$

where

$$\alpha = \frac{\pi\sum_{i=1}^m \mathrm{Re}(p_i)}{\omega_h-\omega_l}.$$

The above lower bound shows that the sensitivity can be very significant in the transition band.

Next, using the Poisson integral relation, we investigate the design constraints on sensitivity properties imposed by open-loop nonminimum phase zeros. Suppose L has at least one more poles than zeros and suppose $z = x_0 + jy_0$ with $x_0 > 0$ is a right-half plane zero of L. Then

$$\int_{-\infty}^{\infty} \ln|S(j\omega)|\frac{x_0}{x_0^2+(\omega-y_0)^2}d\omega = \pi\ln\prod_{i=1}^m\left|\frac{z+p_i}{z-p_i}\right| \tag{6.17}$$

This integral implies that the sensitivity reduction ability of the system may be severely limited by the open-loop unstable poles and nonminimum phase zeros, especially when these poles and zeros are close to each other.

Define

$$\theta(z) := \int_{-\omega_l}^{\omega_l}\frac{x_0}{x_0^2+(\omega-y_0)^2}d\omega$$

Then

$$\begin{aligned}\pi\ln\prod_{i=1}^m\left|\frac{z+p_i}{z-p_i}\right| &= \int_{-\infty}^{\infty}\ln|S(j\omega)|\frac{x_0}{x_0^2+(\omega-y_0)^2}d\omega \\ &\leq (\pi-\theta(z))\ln\|S(j\omega)\|_\infty + \theta(z)\ln(\epsilon),\end{aligned}$$

which gives

$$\|S(s)\|_\infty \geq \left(\frac{1}{\epsilon}\right)^{\frac{\theta(z)}{\pi-\theta(z)}}\left(\prod_{i=1}^m\left|\frac{z+p_i}{z-p_i}\right|\right)^{\frac{\pi}{\pi-\theta(z)}}$$

This lower bound on the maximum sensitivity shows that for a nonminimum phase system, its sensitivity must increase significantly beyond one at certain frequencies if the sensitivity reduction is to be achieved at other frequencies.

6.6 Analyticity Constraints

Let $p_1, p_2, \ldots, p_m$ and $z_1, z_2, \ldots, z_k$ be the open right-half plane poles and zeros of L, respectively. Suppose that the closed-loop system is stable. Then

$$S(p_i) = 0,\quad T(p_i) = 1,\ i = 1, 2, \ldots, m$$

and

$$S(z_j) = 1, \quad T(z_j) = 0, \; j = 1, 2, \ldots, k$$

The internal stability of the feedback system is guaranteed by satisfying these analyticity (or interpolation) conditions. On the other hand, these conditions also impose severe limitations on the achievable performance of the feedback system.

Suppose $S = (I + L)^{-1}$ and $T = L(I + L)^{-1}$ are stable. Then $p_1, p_2, \ldots, p_m$ are the right-half plane zeros of S and $z_1, z_2, \ldots, z_k$ are the right-half plane zeros of T. Let

$$B_p(s) = \prod_{i=1}^{m} \frac{s - p_i}{s + p_i}, \quad B_z(s) = \prod_{j=1}^{k} \frac{s - z_j}{s + z_j}$$

Then $|B_p(j\omega)| = 1$ and $|B_z(j\omega)| = 1$ for all frequencies and, moreover,

$$B_p^{-1}(s)S(s) \in \mathcal{H}_\infty, \quad B_z^{-1}(s)T(s) \in \mathcal{H}_\infty.$$

Hence, by the maximum modulus theorem, we have

$$\|S(s)\|_\infty = \left\|B_p^{-1}(s)S(s)\right\|_\infty \geq |B_p^{-1}(z)S(z)|$$

for any z with $\mathrm{Re}(z) > 0$. Let z be a right-half plane zero of L; then

$$\|S(s)\|_\infty \geq |B_p^{-1}(z)| = \prod_{i=1}^{m} \left| \frac{z + p_i}{z - p_i} \right|$$

Similarly, one can obtain

$$\|T(s)\|_\infty \geq |B_z^{-1}(p)| = \prod_{j=1}^{k} \left| \frac{p + z_j}{p - z_j} \right|$$

where p is a right-half plane pole of L.

The weighted problem can be considered in the same fashion. Let W_e be a weight such that $W_e S$ is stable. Then

$$\|W_e(s)S(s)\|_\infty \geq |W_e(z)| \; \prod_{i=1}^{m} \left| \frac{z + p_i}{z - p_i} \right|$$

Now suppose $W_e(s) = \dfrac{s/M_s + \omega_b}{s + \omega_b \epsilon}$, $\|W_e S\|_\infty \leq 1$, and z is a real right-half plane zero. Then

$$\frac{z/M_s + \omega_b}{z + \omega_b \epsilon} \leq \prod_{i=1}^{m} \left| \frac{z - p_i}{z + p_i} \right| =: \alpha,$$

which gives

$$\omega_b \leq \frac{z}{1 - \alpha\epsilon}(\alpha - \frac{1}{M_s}) \approx z(\alpha - \frac{1}{M_s})$$

where $\alpha = 1$ if L has no right-half plane poles. This shows that the bandwidth of the closed-loop must be much smaller than the right-half plane zero. Similar conclusions can be arrived at for complex right-half plane zeros.

6.7 Notes and References

The loop-shaping design is well-known for SISO systems in the classical control theory. The idea was extended to MIMO systems by Doyle and Stein [1981] using the LQG design technique. The limitations of the loop-shaping design are discussed in detail in Stein and Doyle [1991]. Chapter 16 presents another loop-shaping method using $\mathcal{H}_\infty$ control theory, which has the potential to overcome the limitations of the LQG/LTR method. Some additional discussions on the choice of weighting functions can be found in Skogestad and Postlethwaite [1996]. The design tradeoffs and limitations for SISO systems are discussed in detail in Bode [1945], Horowitz [1963], and Doyle, Francis, and Tannenbaum [1992]. The monograph by Freudenberg and Looze [1988] contains many multivariable generalizations. The multivariable generalization of Bode's integral relation can be found in Chen [1995], on which Section 6.5 is based. Some related results can be found in Boyd and Desoer [1985]. Additional related results can be found in a recent book by Seron, Braslavsky, and Goodwin [1997].

6.8 Problems

Problem 6.1 Let P be an open-loop plant. It is desired to design a controller so that the overshoot $\leq 10\%$ and settling time ≤ 10 sec. Estimate the allowable peak sensitivity M_s and the closed-loop bandwidth.

Problem 6.2 Let $L_1 = \dfrac{1}{s(s+1)^2}$ be an open-loop transfer function of a unity feedback system. Find the phase margin, overshoot, settling time, and the corresponding M_s.

Problem 6.3 Repeat Problem 6.2 with

$$L_2 = \frac{100(s+10)}{(s+1)(s+2)(s+20)}.$$

Problem 6.4 Let $P = \dfrac{10(1-s)}{s(s+10)}$. Use classical loop-shaping method to design a first-order lead or lag controller so that the system has at least 30^o phase margin and as large a crossover frequency as possible.

Problem 6.5 Use the root locus method to show that a nonminimum phase system cannot be stabilized by a very high-gain controller.

Problem 6.6 Let $P = \dfrac{5}{(1-s)(s+2)}$. Design a lead or lag controller so that the system has at least 30^o phase margin with loop gain ≥ 2 for any frequency $\omega \leq 0.1$ and the smallest possible bandwidth (or crossover frequency).

Problem 6.7 Use the root locus method to show that an unstable system cannot be stabilized by a very low gain controller.

Problem 6.8 Consider the unity-feedback loop with proper controller $K(s)$ and strictly proper plant $P(s)$, both assumed square. Assume internal stability.

1. Let $w(s)$ be a scalar weighting function, assumed in $\mathcal{RH}_\infty$. Define

$$\epsilon = \|w(I+PK)^{-1}\|_\infty, \quad \delta = \|K(I+PK)^{-1}\|_\infty$$

so ϵ measures, say, disturbance attenuation and δ measures, say, control effort. Derive the following inequality, which shows that ϵ and δ cannot both be small simultaneously in general. For every Re $s_0 \geq 0$

$$|w(s_0)| \leq \epsilon + |w(s_0)|\sigma_{\min}[P(s_0)]\delta.$$

2. If we want very good disturbance attenuation at a particular frequency, you might guess that we need high controller gain at that frequency. Fix ω with $j\omega$ not a pole of $P(s)$, and suppose

$$\epsilon := \sigma_{\max}[(I+PK)^{-1}(j\omega)] < 1.$$

Derive a lower bound for $\sigma_{\min}[K(j\omega)]$. This lower bound should blow up as $\epsilon \to 0$.

Problem 6.9 Suppose that P is proper and has one right half plane zero at $s = z > 0$. Suppose that $y = \frac{w}{1+PK}$, where w is a unit step at time $t = 0$, and that our performance specification is

$$|y(t)| \leq \begin{cases} \alpha, & \text{if } \; 0 \leq t \leq T \; ; \\ \beta, & \text{if } \; T < t \end{cases}$$

for some $\alpha > 1 > \beta > 0$. Show that for a proper, internally stabilizing, LTI controller K to exist that meets the specification, we must have that

$$\ln\left(\frac{\alpha-\beta}{\alpha-1}\right) \leq zT.$$

What tradeoffs does this imply?

Problem 6.10 Let K be a stabilizing controller for the plant

$$P = \frac{s-\alpha}{(s-\beta)(s+\gamma)}$$

$\alpha > 0, \beta > 0, \gamma \geq 0$. Suppose $|S(j\omega)| \leq \delta < 1, \ \forall\omega \in [-\omega_0, \omega_0]$ where

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

Find a lower bound for $\|S\|_\infty$ and calculate the lower bound for $\alpha = 1$, $\beta = 2$, $\gamma = 10$, $\delta = 0.2$, and $\omega_0 = 1$.