# 4

# Controller Design

## 4.1 Introduction

Control system design is a very rich field. There have been substantial advances over the past 50 years that have resulted in much insight and understanding as well as specific design methods. This development has been augmented by the advances in computing and the development of computer-based design tools. Broadly speaking, PID controllers have been designed using two different approaches; model-based control and direct tuning. The model based approaches start with a simple mathematical model of the process. Very simple models have been used, typically a first-order system with a time delay. In direct tuning a controller is applied to the process, and some simple experiments are performed to arrive at the controller parameters. Because of the simplicity of models and the controller special methods have been developed for PID control. From 1990 there has been a significant increase in the interest in design of PID controllers, partially motivated by the needs of automatic tuning devices for such controllers.

To develop design methods it is necessary to realize that there is a very wide range of different types of control problems even if the controller is restricted to PID. Some typical examples are:

- Design of a simple controller for a non-critical application.
- Design of a controller for a special process that minimizes fluctuations in important control variables.
- Development of a design technique that can be used in a universal autotuner for PID control.

There are also a number of important non-technical issues that should be considered: What is the time and effort required to apply the method? What is the knowledge level required of the user? A solution to the design problem should also give an understanding of when it is beneficial to add derivative action to a PI controller and when even more complex controllers should be considered.

This chapter gives an overview of ideas and concepts that are relevant for

PID control. It is attempted to bring design of PID controllers more into the mainstream of control design.

## 4.2 A Rich Variety of Control Problems

Before discussing specific tuning methods it is useful to realize that there is a wide range of control problems with very diverse goals. Some examples are: steady-state regulation, set-point tracking and path following, and control of buffers and surge tanks.

The goal of steady-state regulation is to keep process variables close to desired values. The key problems are caused by load disturbances, measurement noise, and process variations. Steady-state regulation is very common in process control.

In set-point tracking it is attempted to make process variables follow a given time function or a given curve. These problems typically occur in motion control and robotics. In some cases, for example, machine tool control or robotics, the demand on tracking precision is very severe. In other cases, for example, moving robots, the requirements are less demanding. There is a significant difference between tracking a given time curve and path following, which typically involves control of several variables.

Buffers are common in the industrial production. They are used to smooth variations between different production processes, both in process control and in discrete manufacturing. In process control they are often called *surge tanks*. Buffers are also common in computing systems. They are used in servers to smooth variations in demand of clients, and they are used in computer networks to smooth variations in the load. Buffers are also key elements in supply chains where effective buffer control has a major impact on profitability. The buffer levels should fluctuate; otherwise the buffer does not function. Ideally, no control should be applied unless there is a risk of over- or underflow. An integrating controller with low gain and a scheduling that gives higher gains at the buffer limits are commonly used.

The key issues in many of the control problems are attenuation of load disturbances, injection of measurement noise, robustness to process variations, and set-point following. The relative importance of these factors and the requirements vary from application to application, but all factors must be considered.

## 4.3 Feedback Fundamentals

A block diagram of a basic feedback loop with a controller having two degrees of freedom is shown in Figure 4.1. The process is represented by the block $P$. The controller is represented by the feedback block $C$ and the feedforward

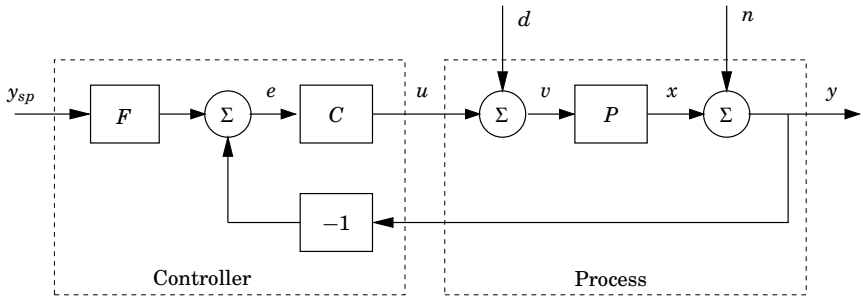


**Figure 4.1** Block diagram of a basic feedback loop having two degrees of freedom.

part $F$. For an ideal PID controller with set-point weighting we have

$$\begin{aligned} C(s) &= K\left(1 + \frac{1}{sT_i} + sT_d\right) \\ F(s) &= \frac{b + \frac{1}{sT_i} + csT_d}{1 + \frac{1}{sT_i} + sT_d}. \end{aligned} \tag{4.1}$$

Compare with (3.7) and (3.20). The signal $u$ is the control signal, and the signal $x$ is the real process variable. Information about $x$ is obtained from the sensor signal $y$, which is corrupted by measurement noise $n$. The signal $d$ represents load disturbances that drive the system away from its desired state. This signal can enter the process in different ways; in Figure 4.1 it is assumed that it acts on the process input.

The goal of control design is to determine the transfer functions $C$ and $F$ so that the process variable $x$ is close to the set point $y_{sp}$ in spite of load disturbances, measurement noise, and process uncertainties. The feedback can reduce the effect of load disturbances. Because of the feedback measurement noise is fed back into the system. It is essential to make sure that this does not cause large variations in the process variable. Since the process model is never accurate it is essential that the behavior of the closed-loop system is insensitive to variations in the process. The feedforward transfer function $F$ is designed to give the desired response to set-point changes.

### Fundamental Relations

The feedback loop is influenced by three external signals, the set point $y_{sp}$, the load disturbance $d$, and the measurement noise $n$. There are at least three signals $x$, $y$, and $u$ that are of great interest for control. This means that there are nine relations between the input and the output signals. Since the system is linear these relations can be expressed in terms of the transfer functions. Let $X$, $Y$, $U$, $D$, $N$, and $Y_{sp}$ be the Laplace transforms of $x$, $y$, $u$, $d$, $n$, and $y_{sp}$, respectively. The following relations are obtained from the block diagram

in Figure 4.1:

$$\begin{aligned} X &= \frac{PCF}{1+PC}Y_{sp} + \frac{P}{1+PC}D - \frac{PC}{1+PC}N \\ Y &= \frac{PCF}{1+PC}Y_{sp} + \frac{P}{1+PC}D + \frac{1}{1+PC}N \\ U &= \frac{CF}{1+PC}Y_{sp} - \frac{PC}{1+PC}D - \frac{C}{1+PC}N. \end{aligned} \tag{4.2}$$

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

$$\begin{matrix} \frac{PCF}{1+PC} & \frac{PC}{1+PC} & \frac{P}{1+PC} \\ \frac{CF}{1+PC} & \frac{C}{1+PC} & \frac{1}{1+PC}. \end{matrix} \tag{4.3}$$

The transfer functions in the first column give the response of process variable and control signal to the set point. The second column gives the same signals in the case of pure error feedback when $F = 1$. The transfer function $P/(1+PC)$ in the third column tells how the process variable reacts to load disturbances, and the transfer function $C/(1+PC)$ gives the response of the control signal to measurement noise.

Notice that only four transfer functions,

$$\begin{matrix} \frac{PC}{1+PC} & \frac{P}{1+PC} \\ \frac{C}{1+PC} & \frac{1}{1+PC}, \end{matrix} \tag{4.4}$$

are required to describe how the system reacts to load disturbance and the measurement noise. These transfer functions are called the *Gang of Four*. They also capture robustness, as will be discussed in Section 4.6. Two additional transfer functions are required to describe how the system responds to set-point changes.

The special case when $F = 1$ is called a system with (pure) error feedback. In this case, all control actions are based on feedback from the error only. In this case, the system is completely characterized by the Gang of Four (4.4).

We are often interested in the magnitude of the transfer functions given by Equation 4.4. It is important to be aware that the transfer functions $PC/(1+PC)$ and $1/(1+PC)$ are dimension free, but the transfer functions $P/(1+PC)$ and $C/(1+PC)$ are not. For practical purposes it is therefore important to normalize the signals, for example, by scaling process inputs and outputs to the interval 0 to 1 or $-1$ to 1.

### A Practical Consequence

The fact that six relations are required to capture properties of the basic feedback loop is often neglected in literature, particularly in the papers on PID control. To describe the system properly it is thus necessary to show the response of all six transfer functions. The transfer functions can be represented

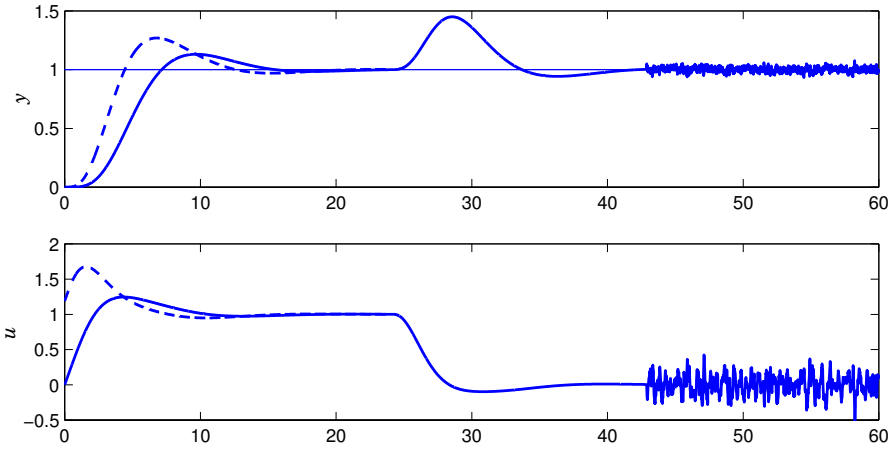


**Figure 4.2** Representation of the properties of a basic feedback system by responses to a step in the reference, a step in the load disturbance, and measurement noise. The full lines are for set-point weight $b = 0$ and the dashed line is for set-point weight $b = 1$.

in different ways, by their step responses or by their frequency responses. Most papers on control only show the response of the process variable to set-point changes. Such a curve gives only partial information about the behavior of the system. To get a more complete representation of the system all six responses should be given, for example, as shown in Figure 4.2. This figure shows the responses in process variable and control signal to an experiment with a step change in set point followed by a step in the load disturbance, and measurement noise. The solid lines show the response when $F = 1$ and the dashed lines show the response when feedforward is used. Figure 4.2 thus gives a complete characterization of all six transfer functions in Equation 4.3.

### Many Variations

The system shown in Figure 4.1 is a prototype problem. There are many variations of this problem. In Figure 4.1 the load disturbances act on the process input. In practice the disturbances can appear in many other places in the system. The measurement noise also acts at the process output. There may also be dynamics in the sensor, and the measured signal is often filtered. All these variations can be studied with minor modifications of the analysis based on Figure 4.1. As an illustration we will investigate the effects of a sensor filter. Figure 4.3 shows a block diagram of such a system. A typical example is a PID controller with set-point weighting and a second-order measurement filter. The transfer functions $F(s)$ and $C(s)$ in Figure 4.3 are given by (4.1) and the filter transfer function $G_f(s)$ is

$$G_f(s) = \frac{1}{1 + sT_f + s^2T_f^2/2}. \tag{4.5}$$

The relations between the input signals and output signals in Figure 4.3 are

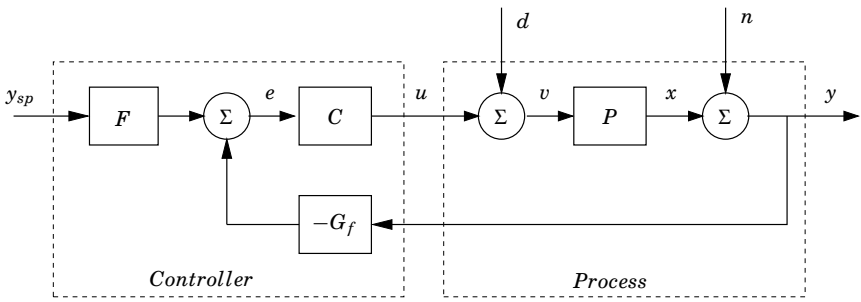


**Figure 4.3** Block diagram of a basic feedback loop having two degrees of freedom and filtering of the measurement.

given by

$$\begin{aligned} X &= \frac{PCF}{1+PCG_f}Y_{sp} + \frac{P}{1+PCG_f}D - \frac{PCG_f}{1+PCG_f}N \\ Y &= \frac{PCF}{1+PCG_f}Y_{sp} + \frac{P}{1+PCG_f}D + \frac{1}{1+PCG_f}N \\ U &= \frac{CF}{1+PCG_f}Y_{sp} - \frac{PCG_f}{1+PCG_f}D - \frac{CG_f}{1+PCG_f}N. \end{aligned} \tag{4.6}$$

Equation (4.6) is identical to (4.2) if the transfer function $C(s)$ and $F(s)$ are replaced by

$$\bar{C}(s) = C(s)G_f(s), \quad \bar{F}(s) = \frac{F(s)}{G_f(s)}, \tag{4.7}$$

The modifications required to deal with filtering are thus minor, and it suffices to develop the theory for the configuration in Figure 4.1.

### Separation of Responses to Disturbances and Set Points

In early work on PID control it was a tradition to have two tuning rules, one for good set-point response and another for efficient attenuation of load disturbances. This practice still continues. A strong advantage of a controller with two degrees of freedom is that the responses to disturbances and set point can be designed separately. This follows from (4.2), which shows that the response to load disturbances and measurement noise is given by the $C(s)$, or from (4.6) by $\bar{C}(s) = C(s)G_f(s)$ when the measurement is filtered. A good design procedure is thus to determine $C(s)$ to account for robustness and disturbances. The feedforward transfer function $F(s)$ can then be chosen to give the desired set-point response. In general, this requires that the feedforward transfer function can be chosen freely. Simply choosing the set-point weights often give satisfactory results. Notice that there are some situations where only the error signal is available. The decoupling of the design problem then is not possible, and the design of the feedback then has to consider a trade-off between disturbances, robustness, and set-point response.

### Fundamental Limitations

In any design problem it is important to be aware of the fundamental limitations. Typical sources of limitations are

- Process dynamics
- Nonlinearities
- Disturbances
- Process uncertainty

Process dynamics is often the limiting factor. Time delays and poles and zeros in the right half plane are relevant factors. It is important to be aware of these limitations. Time delays are the most common factor for PID control. It seems intuitively reasonable that it is impossible to have tight control of a system with a time delay. It can be shown that for a process with a time delay $L$ the achievable gain crossover frequency $\omega_{gc}$, which is defined in Section 4.4, is limited by

$$\omega_{gc} L < 1. \tag{4.8}$$

Since

$$e^{-sL} \approx \frac{1 - sL/2}{1 + sL/2},$$

it also seems reasonable that right-half plane zeros also limit the achievable performance. It can be shown that a right-half plane zero at $s = b$ limits the gain crossover frequency to

$$\omega_{gc} < 0.5b. \tag{4.9}$$

A right-half plane pole $s = a$ in the process limits the achievable gain crossover frequency $\omega_{gc}$ to

$$\omega_{gc} > 2a. \tag{4.10}$$

Notice that time delays and right-half plane zeros give an upper bound to the achievable gain crossover frequency while right-half plane poles give a lower bound.

Nonlinearities, saturation, and rate saturation are very common; they impose limitations on how much and how fast the process variables can change. Saturations combined with unstable process dynamics are particularly serious because they may lead to situations where it is not possible to recover stable operating conditions. Such situations are fortunately not common in process control.

Load disturbances and measurement noise limit how accurately a process variable can be controlled. The limitations often interact. The allowable controller gain is, for example, limited by a combination of measurement noise and actuator saturation. The effect of load disturbances depends critically on the achievable bandwidth.

Process models used for control are always approximations. Process dynamics may also change during operation. Insensitivity to model uncertainty is one of the essential properties of feedback. There is, however, a limit to the

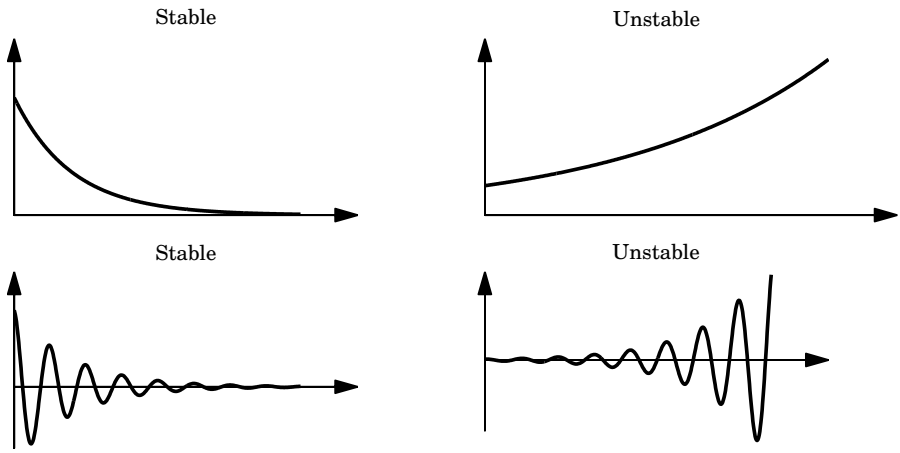


**Figure 4.4** Illustration of different system behaviors used to define stability.

uncertainty that can be dealt with. Feedback cannot be active in frequency ranges where the uncertainty in the phase of the process is larger than $\pm 90^\circ$. To have reasonable control performance the uncertainty should be less than about $\pm 15^\circ$. If the process variations correlate well with some measured quantity it is possible to compensate for the uncertainties by changing the controller parameters. This technique, which is called gain scheduling, will be discussed in Section 9.3.

## 4.4 Stability

Feedback has many useful properties. The main drawback is that feedback may cause instability. It is therefore essential to have a good understanding of stability and the mechanisms that cause instability.

### Stability Concepts

The notion of stability is intuitively very simple. It tells how a system behaves after a perturbation. Already in 1868 Maxwell classified the behavior as follows:

U1: The variable increases continuously

S1: The variable decreases continuously

U2: The variable increases in an oscillatory manner

S2: The variable decreases in an oscillatory manner

These behaviors are illustrated in Figure 4.4. Maxwell called the behaviors labeled S stable and the ones labeled U unstable. He also found that for linear time-invariant systems stability was related to properties of the roots of an algebraic equation.

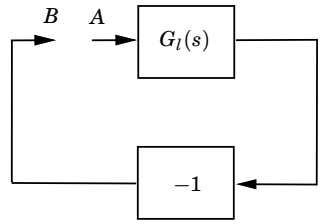


**Figure 4.5** Block diagram of a simple feedback system.

Consider a system with the transfer function

$$G(s) = \frac{b(s)}{a(s)}, \tag{4.11}$$

where $a(s)$ and $b(s)$ are polynomials. Recall that the roots of the polynomial $a(s)$ are called the poles of the system. Since a pole $s_i$ corresponds to a time function $e^{s_i t}$ the following relations are obtained between the behaviors and the roots of an algebraic equation:

U1: Corresponds to real poles with positive real part

S1: Corresponds to real poles with negative real part

U2: Corresponds to complex poles with positive real part

S2: Corresponds to complex poles with negative real part

The system (4.11) is stable if it has no poles in the right half plane. The equation

$$a(s) = 0 \tag{4.12}$$

is called the characteristic equation. A system is stable if the characteristic equation does not have any roots with positive real parts. It is common practice to label poles on the imaginary axis as unstable.

### Nyquist's Stability Criterion

The algebraic definition of stability based on the roots of the characteristic equation is useful, but it also has some drawbacks. Consider, for example, the feedback system in Figure 4.5 where the transfer functions of the process and the controller have been combined into one block with the transfer function $G_l = PC$. The characteristic equation for this system is

$$1 + G_l(s) = 0. \tag{4.13}$$

The transfer function, which is the product of the transfer functions of the process and the controller, describes how signals propagate around the feedback loop and is called the *loop transfer function*. It is not easy to see how the roots of (4.13) are influenced by the transfer functions of the process and the controller. This can, however, be done by using a totally different view of stability, which was developed by Nyquist. He started by investigating the conditions

for maintaining an oscillation in the system shown in Figure 4.5. Assume that the feedback loop is broken as is indicated in the figure and that the signal $u_A(t) = \sin \omega_0 t$ is injected at point A. After a transient the output at point B is then given by

$$u_B(t) = -|G_l(i\omega_0)| \sin\big(\omega_0 t + \arg G_l(i\omega)\big).$$

The signals $u_A(t)$ and $u_B(t)$ are identical if

$$G_l(i\omega_0) = -1, \tag{4.14}$$

and an oscillation will be maintained if the loop is closed by joining points A and B. Equation 4.14 thus gives the condition for oscillations in the system. It follows from (4.13) and (4.14) that the condition for oscillation implies that the characteristic equation of the system has a root $s = i\omega_0$. The frequencies where the system can maintain an oscillation can be determined by solving (4.14) for $\omega_0$.

Nyquist developed a stability criterion based on the idea of how sinusoids propagate around the feedback loop. Nyquist argues as follows. He first investigated frequencies where the signals $u_A$ and $u_B$ are in phase, i.e., when $\arg G_l(i\omega_0) = \pi$. Intuitively it seems reasonable that the system is stable if $|G_l(i\omega_0)| < 1$ because the amplitude is then decreased when the signal traverses the loop. The situation is actually a little more complicated because the system may be stable even if $|G_l(i\omega_0)| > 1$. The precise result can be expressed in terms of the Nyquist curve introduced in Section 2.3. Recall that the Nyquist curve is a plot of $(\mathrm{Re} G_l(i\omega), \mathrm{Im} G_l(i\omega))$ for $0 \le \omega \le \infty$. When the loop transfer function does not have poles in the right half plane the condition for stability is that the critical point $-1$ is to the left of the Nyquist curve when it is traversed for increasing $\omega$.

A nice property of the Nyquist's criterion is that it indicates how a system should be changed in order to move the Nyquist curve away from the critical point. Figure 6.4 shows that derivative action, which introduces phase lead, bends the curve away from the critical point. Integral action, which introduces phase lag, bends the curve towards the critical point. The idea is to modify the controller so that the curve is bent away from the critical point. This has led to a whole class of design methods called loop shaping.

### Stability Margins

In practice it is not enough to require that the system is stable. There must also be some margins of stability. This means that the Nyquist curve should not be too close to the critical point. This is illustrated in Figure 4.6, which shows several stability margins. The gain margin $g_m$ tells how much controller gain can be increased before reaching the stability limit. Let *phase crossover frequency* $\omega_{180}$ be the smallest frequency where the phase lag of the loop transfer function $G_l(s)$ is 180° and the gain margin be defined as

$$g_m = \frac{1}{|G_l(i\omega_{180})|}. \tag{4.15}$$

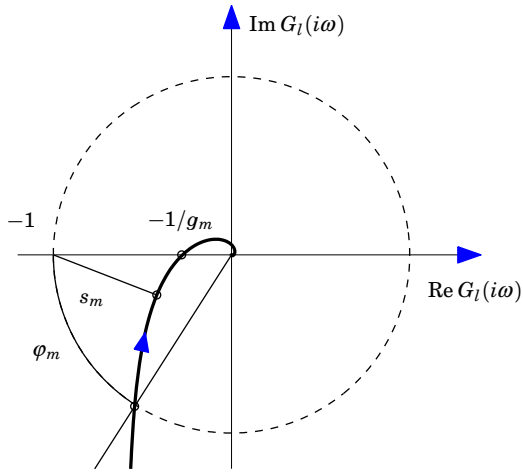


**Figure 4.6** Nyquist plot of the loop transfer function $G_l$ with gain margin $g_m$, phase margin $\varphi_m$ and stability margin $s_m$.

The point where the Nyquist curve intersects the unit circle is another interesting point. This point can be characterized by the angle $\varphi_m$. This angle called *the phase margin* is also a measure of how close the Nyquist curve is to the critical point. The angle $\varphi_m$ is the amount of phase lag required to reach the stability limit. The *gain crossover frequency* $\omega_{gc}$ is the lowest frequency where the loop transfer function $G_l(s)$ has unit magnitude. The phase margin is formally defined as

$$\varphi_m = \pi + \arg G_l(i\omega_{gc}). \tag{4.16}$$

Both gain and phase margin are classical measures of degrees of stability. Both values must be specified in order to ensure that the Nyquist curve is far from the critical point. They can be replaced by a single number, the shortest distance from the Nyquist curve to the critical point $-1$, which is called the *stability margin* $s_m$.

Reasonable values of the margins are phase margin $\varphi_m = 30^\circ - 60^\circ$, gain margin $g_m = 2 - 5$, stability margin $s_m = 0.5 - 0.8$.

The gain and phase margins were originally conceived for the case when the Nyquist curve only intersects the unit circle and the negative real axis once. For more complicated systems there may be many intersections, and it is then necessary to consider the intersections that are closest to the critical point. For more complicated systems there is also another number that is highly relevant, namely, *the delay margin*. The delay margin is defined as the smallest time delay required to make the system unstable. For loop transfer functions that decay quickly the delay margin is closely related to the phase margin, but for systems where the amplitude ratio of the loop transfer function has several peaks at high frequencies the delay margin is a much more relevant measure. This is particularly relevant for the Smith predictor that will be discussed in Chapter 8.

### Internal Stability

So far we have only discussed the simple feedback system in Figure 4.5. For the more general system in Figure 4.1 which is characterized by six transfer functions, it is necessary to require that all four transfer functions,

$$\begin{matrix} \frac{PC}{1+PC} & \frac{P}{1+PC} \\ \frac{C}{1+PC} & \frac{1}{1+PC} \end{matrix}, \tag{4.17}$$

are stable; compare with (4.3). This is called internal stability. Notice that there may be cancellations of poles and zeros in the product $PC$.

### Stability Regions

A primary requirement for a PID controller is that the parameters of the controller are chosen in such a way that the closed-loop system is stable. A PID controller of the form

$$C(s) = k + \frac{k_i}{s} + k_d s \tag{4.18}$$

has three parameters only, and the stability region can be represented by a volume in three dimensions. To describe this volume the process transfer function is represented as

$$P(i\omega) = r(\omega)e^{i\phi(\omega)} = r(\omega)(\cos(\omega) + i\sin(\omega)),$$

and the condition for oscillation (4.14) then becomes

$$P(i\omega)C(i\omega) = r(\omega)(\cos(\omega) + i\sin(\omega))\big(k - i\frac{k_i}{\omega} + ik_d\omega\big) = -1.$$

Separating the real and imaginary parts we find that the boundary of the stability region can be represented parametrically as

$$\begin{aligned} k &= -\frac{\cos\phi(\omega)}{r(\omega)} \\ k_i &= \omega^2 k_d - \frac{\omega\sin\phi(\omega)}{r(\omega)}. \end{aligned} \tag{4.19}$$

It is thus straightforward to determine the stability region for a constant value of $k_d$. Repeating the calculations for a set of $k_d$-values gives the stability region for the PID controller.

EXAMPLE 4.1—STABILITY REGION FOR $P(s) = 1/(s+1)^4$

Figure 4.7 shows the stability region for a process with the transfer function $P(s) = 1/(s+1)^4$. The value $k_d = 0$ corresponds to PI control. Integral gain $k_i$ may be increased by adding derivative action. The integral gain has its maximum $k_i = 36$ at the boundary of the stability region for $k = 8$ and $k_d = 20$. The system is unstable for all values of $k$ and $k_i$ if $k_d > 20$. □

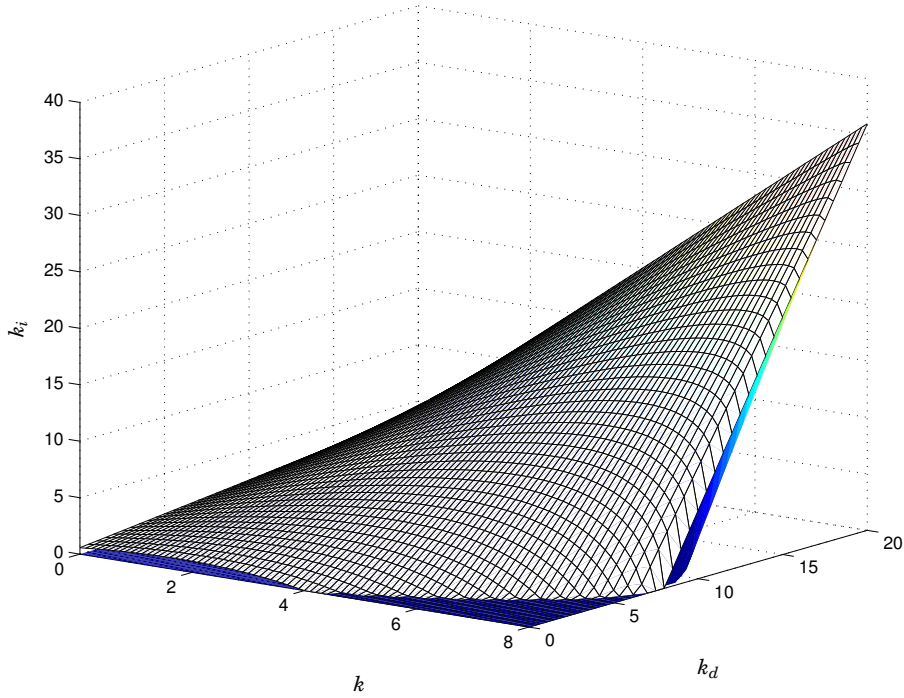


**Figure 4.7** Stability region for the system $P(s) = (1+s)^{-4}$.

Some interesting conclusions can be drawn from Example 4.1. To have good disturbance rejection it is desirable to have a large value of $k_i$. This is shown in Section 4.9. With PI control, the largest value of $k_i$ for a stable system is $k_i = 1$. Figure 4.7 shows that the value of $k_i$ can be increased substantially by introducing derivative action. The highest value of $k_i$ that can be obtained with a stable system is $k_i = 36$. This will, however, be a very fragile controller because the system can be made unstable by arbitrarily small changes in controller gains. For large values of $k_d$ the curves have sharp corners at the point of maximum integral gain. This property of derivative action is one reason why tuning of controllers with derivative action is difficult. It will be discussed further in Chapters 6 and 7.

### Constant Proportional Gain

The region of parameters where the system is stable is a subset of $R^3$. The calculations performed give the two-dimensional intersections with constant derivative gain. Additional insight can be obtained from another representation of the stability regions. To investigate the stability we will use the Nyquist criterion and plot the locus of the loop transfer function $G_l(s)$. With proportional control we have $G_l = kP$. For a fixed value of the proportional gain $k > 0$ we determine the frequency $\omega_n$ where the Nyquist curve of $kP(i\omega)$ intersects the circle with the line segment $(-1, 0)$ as a diameter; see Figure 4.8. We will first consider the case when the intersection of the Nyquist curve and the circle

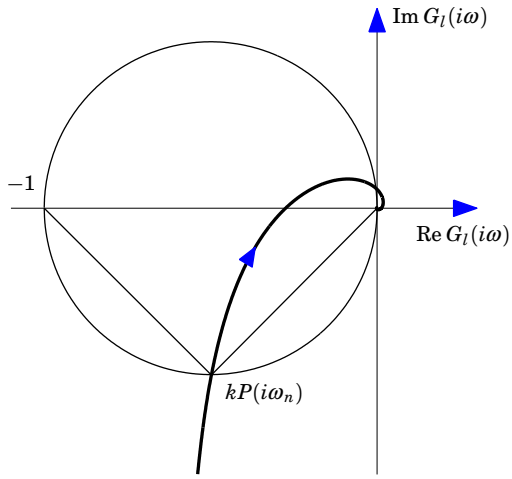


**Figure 4.8** Nyquist curve of the loop transfer function $G_l(s) = kP(s)$.

occurs in the lower half plane as shown in Figure 4.8. The controller transfer function is

$$C(i\omega) = k + i\Big(-\frac{k_i}{\omega} + k_d\omega\Big) = k - i\Big(\frac{k_i}{\omega} - k_d\omega\Big),$$

hence,

$$G_l(i\omega_n) = P(i\omega_n)C(i\omega_n) = kP(i\omega_n) - i\Big(\frac{k_i}{\omega_n} - k_d\omega_n\Big)kP(i\omega_n).$$

If proportional gain $k$ is fixed the point $kP(i\omega_n)$ moves to $G_l(i\omega_n)$ when proportional and integral gains are different from zero. To avoid reaching the critical point it must be required that

$$\Big(\frac{k_i}{\omega_n} - k_d\omega_n\Big)|P(i\omega_n)| < |1 + P(i\omega_n)|.$$

The same analysis can be made when the intersection of the Nyquist curve and the circle occurs in the upper half plane. Combining the inequalities we find that the stability regions are given by the conditions

$$\begin{aligned} &k_i > 0 \\ &k_i < \omega_n^2 k_d + \omega_n \frac{|1 + kP(i\omega_n)|}{|P(i\omega_n)|}, \text{ for Im } P(i\omega_n) < 0 \\ &k_i > \omega_n^2 k_d - \omega_n \frac{|1 + kP(i\omega_n)|}{|P(i\omega_n)|}, \text{ for Im } P(i\omega_n) > 0 \end{aligned} \tag{4.20}$$

which should hold for for all $\omega_n$ such that

$$\Big|kP(i\omega_n) + \frac{1}{2}\Big| = \frac{1}{2}. \tag{4.21}$$

We can thus conclude that for constant proportional gain the stability region is represented by several convex polygons in the $k_i$-$k_d$ plane. In general, there may be several polygons, and each may have many surfaces. The number of surfaces of the polygons is determined by the number of roots of the Equation 4.21. In many cases, the polygons are also very simple, as is illustrated with the following example.

EXAMPLE 4.2—FOUR EQUAL POLES
To illustrate the results we consider a process with the transfer function

$$P(s) = \frac{1}{(s+1)^4} = \frac{1}{s^4 + 6s^2 + 1 + 4s(s^2+1)}.$$

In this case, Equation 4.21 becomes

$$\omega^4 - 6\omega^2 + 1 + k = 0.$$

This equation has only two positive solutions,

$$\omega^2 = 3 \pm \sqrt{8-k},$$

and it follows from (4.20) that the stability region is given by the inequalities

$$\begin{aligned} k_i &> 0 \\ k_i &< (3 - \sqrt{8-k})k_d + 4k - 56 + 20\sqrt{8-k} \\ k_i &> (3 + \sqrt{8-k})k_d + 4k - 56 - 20\sqrt{8-k}. \end{aligned} \tag{4.22}$$

The stability region is shown in Figure 4.7. The integral gain has its maximum $k_i = 36$ at the boundary of the stability region for $k = 8$ and $k_d = 20$. □

## 4.5 Closed-Loop Poles and Zeros

Many properties of a feedback system can be obtained from the closed-loop poles and zeros. For PID control the behavior is often characterized by a few dominant poles, typically those closest to the origin. Many properties of the closed-loop system can be deduced from the poles and the zeros of complementary sensitivity function

$$T(s) = \frac{PC(s)}{1 + PC(s)}.$$

With error feedback, $F = 1$ in Figure 4.1, the closed-loop zeros are the same as the zeros of loop transfer function $G_l(s)$, and the closed-loop poles are the roots of the equation

$$1 + G_l(s) = 0.$$

The pole-zero configurations of closed-loop systems may vary considerably. Many simple feedback loops, however, will have a configuration of the type

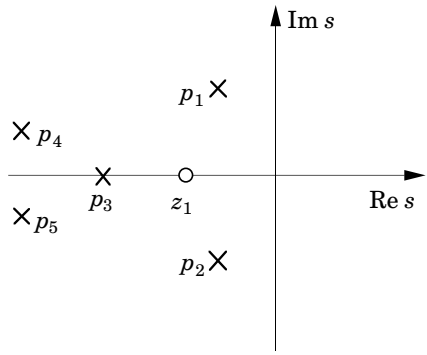


**Figure 4.9** Pole-zero configuration of the transfer function from set point to output for a simple feedback system.

shown in Figure 4.9, where the principal characteristics of the response are given by a complex pair of poles, $p_1$ and $p_2$, called the *dominant poles*. The response is also influenced by real poles and zeros $p_3$ and $z_1$ close to the origin. The position of $p_3$ and $z_1$ may be reversed. There may also be more poles and zeros far from the origin, which typically are of less influence. Poles and zeros to the left of the dominant poles have little influence on the transient response if they are sufficiently far away from the dominant poles. The influence of a pole diminishes if there is a zero close to it.

Complex poles can be characterized in terms of their frequency $\omega_0$, which is the distance from the origin, and their relative damping $\zeta$. A first approximation of the response is obtained from the equivalent second-order system. The response is modified if there are poles and zeros close to the dominating poles. Classical control was very much concerned with closed-loop systems having the pole-zero configuration shown in Figure 4.9.

Even if many closed-loop systems have a pole-zero configuration similar to the one shown in Figure 4.9, there are, however, exceptions. For instance, systems with mechanical resonances, which may have poles and zeros close to the imaginary axis, are generic examples of systems that do not fit the pole-zero pattern of the figure. Another example is processes with a long dead time.

Design of PID controllers are typically based on low-order models, which gives closed-loop systems with a small number of poles and zeros.

### Dominant Poles from the Loop Transfer Function

A simple method for approximate determination of the dominant poles from knowledge of the Nyquist curve of the loop transfer function will now be given. Consider the loop transfer function $G_l(s)$ as a mapping from the $s$-plane to the $G_l$-plane. The map of the imaginary axis in the $s$-plane is the Nyquist curve $G_l(i\omega)$, which is indicated in Figure 4.10. The closed-loop poles are the roots of the characteristic equation

$$1 + G_l(s) = 0.$$

The map of a straight vertical line through the dominant closed-loop poles in

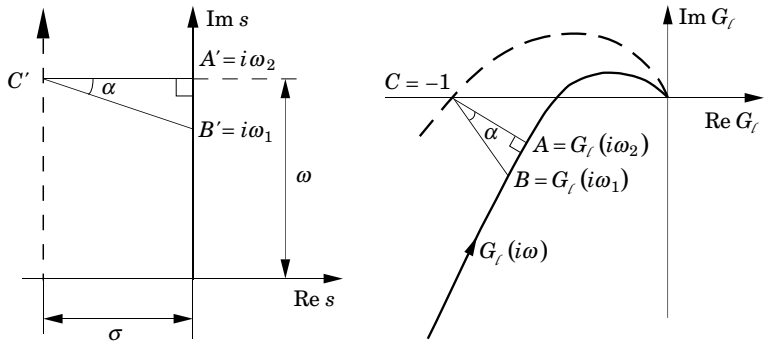


**Figure 4.10** Representation of the loop transfer function $G_l(i\omega)$ as a map of complex planes.

the $s$-plane is thus a curve through the critical point $G_l = -1$ in the $G_l$-plane. This curve is shown by a dashed line in Figure 4.10. Since the map is conform, the straight line A′C′ is mapped on the curve AC, which intersects the Nyquist curve orthogonally. The triangle ABC is also mapped conformally to A′B′C′. If ABC can be approximated by a triangle, we have

$$\frac{G_l(i\omega_2) - G_l(i\omega_1)}{i\omega_2 - i\omega_1} \approx \frac{1 + G_l(i\omega_2)}{\sigma}.$$

When $\omega_1$ is close to $\omega_2$ this becomes

$$\sigma = (1 + G_l(i\omega_2)) \frac{i\omega_2 - i\omega_1}{G_l(i\omega_2) - G_l(i\omega_1)} \approx \frac{1 + G_l(i\omega_2)}{G_l'(i\omega_2)}, \tag{4.23}$$

where $G_l'(s) = dG_l(s)/ds$. To determine the dominant poles we first determine the point A on the Nyquist curve that is closest to the critical point $-1$. This point is characterized by the frequency $\omega_2$. Then determine the derivative of the loop transfer function at $\omega_2$. The dominant poles are then given by $s = -\sigma \pm i\omega_2$, where $\sigma$ is given by Equation 4.23.

## 4.6 The Sensitivity Functions

Two of the transfer functions (4.3) are of particular interest, the sensitivity function $S$ and the complementary sensitivity function $T$. These functions are defined by

$$S = \frac{1}{1 + PC} = \frac{1}{1 + G_l}, \qquad T = \frac{PC}{1 + PC} = \frac{G_l}{1 + G_l}. \tag{4.24}$$

The sensitivity functions are uniquely given by the loop transfer function $G_l(s) = P(s)C(s)$ and have the property $S + T = 1$. The transfer functions reflect many interesting properties of the closed-loop system, particularly robustness to process variations.

### Small Process Variations—The Sensitivity Function

We will start by investigating how sensitive the response to set-point changes is to small process variations. It follows from (4.2) that the transfer function from set point to process variable is

$$G_{xy_{sp}} = G_{yy_{sp}} = \frac{PCF}{1+PC}.$$

Consider $G_{xy_{sp}}$ as a function of the process transfer function $P$. Differentiating with respect to $P$ gives

$$\frac{dG_{xy_{sp}}}{dP} = \frac{CF}{1+PC} - \frac{PC^2F}{(1+PC)^2} = \frac{CF}{(1+PC)^2} = \frac{1}{1+PC}\frac{CF}{1+PC}.$$

Hence,

$$\frac{dG_{xy_{sp}}}{G_{xy_{sp}}} = \frac{1}{1+PC}\frac{dP}{P} = S\frac{dP}{P}. \tag{4.25}$$

Notice that the quantity $dG/G$ can be interpreted as the relative variation in the transfer function $G$. Equation 4.25 thus implies that the relative error in the closed-loop transfer function $G_{yy_{sp}}$ is equal to the product of the sensitivity function and the relative error in the process. For frequencies where the sensitivity function is small it thus follows that the closed-loop system is very insensitive to variations in the process. This is actually one of the key reasons for using feedback. The formula (4.25) is one of the reasons why $S$ is called the sensitivity function. The sensitivity function also has other interesting properties.

### Disturbance Attenuation

A very fundamental question is how much the fluctuations in the process variable are influenced by feedback. Consider the situation shown in Figure 4.11 where the same load disturbance acts on a process $P$ in open loop and on the process $P$ in a closed loop with the controller $C$. Let $y_{ol}$ be the output of the open-loop system and $y_{cl}$ the output of the closed-loop system. We have the following relation between the Laplace transforms of the signals,

$$\frac{Y_{cl}(s)}{Y_{ol}(s)} = \frac{1}{1+P(s)C(s)} = S(s). \tag{4.26}$$

Disturbances with frequencies $\omega$ such that $|S(i\omega)| < 1$ are thus attenuated by feedback, but disturbances such that $|S(i\omega)| > 1$ are amplified by the feedback. A plot of the amplitude ratio of $S$ thus immediately tells the effect of feedback.

Since the sensitivity only depends on the loop transfer function it can be visualized graphically in the Nyquist plot of the loop transfer function. This is illustrated in Figure 4.12. The complex number $1 + G_l(i\omega)$ can be represented as the vector from the point $-1$ to the point $G_l(i\omega)$ on the Nyquist curve. The sensitivity is thus less than one for all points outside a circle with radius 1 and center at $-1$. Disturbances of these frequencies are attenuated by the feedback.

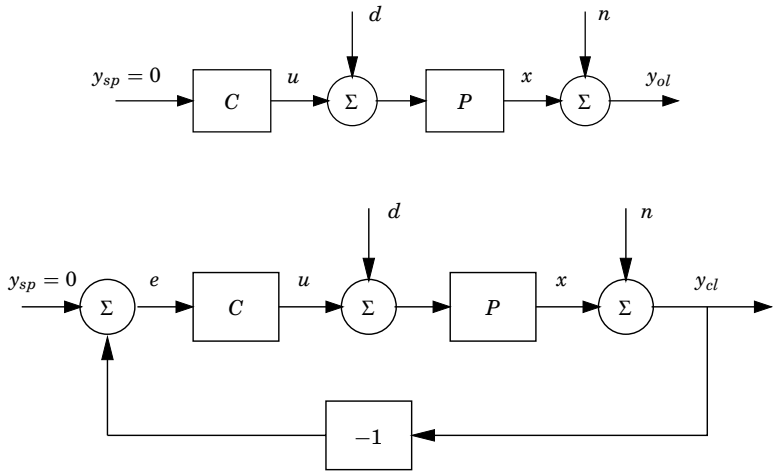


**Figure 4.11** Block diagrams of open- and closed-loop systems subject to the same disturbances.

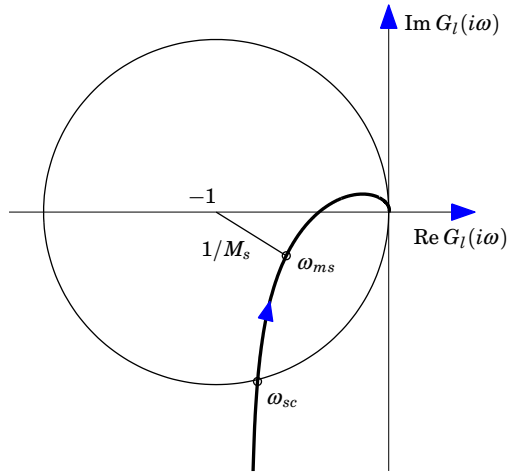


**Figure 4.12** Nyquist curve of loop transfer function showing graphical interpretation of maximum sensitivity. The sensitivity crossover frequency $\omega_{sc}$, and the frequency $\omega_{ms}$ where the sensitivity has its largest value are indicated in the figure. All points inside the circle with center at the $-1$ have sensitivities greater than 1.

The lowest frequency where the sensitivity function has magnitude 1 is called the *sensitivity crossover frequency* $\omega_{sc}$. The value

$$M_s = \max_{\omega} |S(i\omega)| = \max_{\omega} \left| \frac{1}{1 + P(i\omega)C(i\omega)} \right| = \max_{\omega} \left| \frac{1}{1 + G_l(i\omega)} \right|, \qquad (4.27)$$

which is called the maximum sensitivity, tells the worst-case amplification of the disturbances.

The sensitivity cannot be made arbitrarily small. The following relation

holds under reasonably general conditions for stable systems

$$\int_0^\infty \log|S(i\omega)|d\omega = 0. \tag{4.28}$$

This very important relation is called Bode's integral. It says that if the sensitivity is reduced for one frequency it increases at another frequency. Feedback can thus redistribute the attenuation of disturbances for different frequencies, but it cannot reduce the effect of disturbances for all frequencies.

In Section 2.6 it was mentioned that random fluctuations can be modeled by a power spectral density. If the spectral density is $\phi(\omega)$ for a system without control it becomes $|S(i\omega)|^2\phi(\omega)$ for a system with control. The rations of the variances under open and closed loop are thus

$$\frac{\sigma_{cl}^2}{\sigma_{ol}^2} = \frac{\int_{-\infty}^{\infty}|S(i\omega)|^2\phi(\omega)d\omega}{\int_{-\infty}^{\infty}\phi(\omega)d\omega}. \tag{4.29}$$

### Stability Margins and Maximum Sensitivity

Notice that $|1+G_l(i\omega)|$ is the distance from a point on the Nyquist curve of the loop transfer function to the point $-1$. See Figure 4.12. The shortest distance from the Nyquist curve of the loop transfer function to the critical point $-1$ is thus $1/M_s$, which is equal to the stability margin $s_m$. Compare Figures 4.12 and 4.6. The maximum sensitivity can thus also serve as a stability margin. A requirement on $M_s$ gives the following bounds for gain and phase margins

$$g_m \geq \frac{M_s}{M_s-1}$$
$$\varphi_m \geq 2\arcsin\left(\frac{1}{2M_s}\right).$$

The requirement $M_s = 2$ implies that $g_m \geq 2$ and $\varphi_m \geq 29^\circ$ and $M_s = 1.4$ implies that $g_m \geq 3.5$ and $\varphi_m \geq 41^\circ$.

### Nonlinearities in the Loop

The condition that the Nyquist curve of the loop transfer function is outside a circle at the critical point with radius $1/M_s$ has strong implications. It follows from Nyquist's stability criterion that the system remains stable even if the gain is increased by the factor $M_s/(M_s-1)$ or if it is decreased by the factor $M_s/(M_s+1)$. More surprising is that the closed loop is stable even if a static nonlinearity $f$ is inserted in the loop, provided that

$$\frac{M_s}{M_s+1} < \frac{f(x)}{x} < \frac{M_s}{M_s-1}. \tag{4.30}$$

A small value of $M_s$ thus ensures that the system will remain stable in spite of nonlinear actuator characteristics. With $M_s = 2$ the function lies in a sector limited by straight lines through the origin with slopes $2/3$ and 2. With $M_s = 1.4$ the slopes are between 0.28 and 3.5.

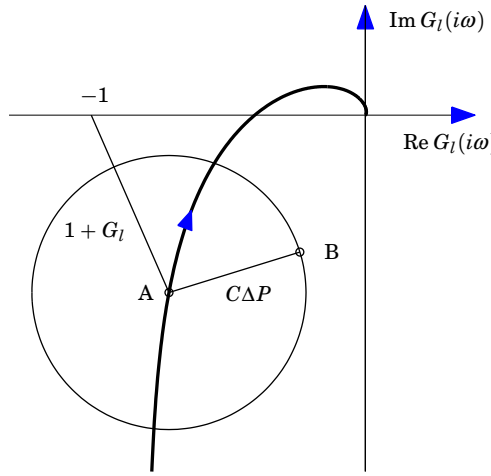


**Figure 4.13** Nyquist curve of a nominal loop transfer function and its uncertainty caused by process variations $\Delta P$.

### Large Variations

We will now investigate conditions for the system to remain stable under large variations in the process transfer function. Assume that the process transfer function changes from $P$ to $P + \Delta P$, where $\Delta P$ is a stable transfer function. Consider a point A on the the Nyquist curve of the loop transfer function; see Figure 4.13. This point then changes from A to B in the figure. The distance from the critical point $-1$ to the point $A$ is $|1 + G_l|$. This means that the perturbed Nyquist curve will not reach the critical point $-1$ provided that

$$|C\Delta P| < |1 + G_l|,$$

which implies

$$|\Delta P| < \Big|\frac{1 + G_l}{C}\Big|. \tag{4.31}$$

Notice that the condition is conservative because it follows from Figure 4.13 that much larger changes can be made in directions from the critical point. The condition (4.31) must be valid for all points on the Nyquist curve, i.e, pointwise for all frequencies. The condition (4.31) for stability can then be written as

$$\Big|\frac{\Delta P(i\omega)}{P(i\omega)}\Big| < \frac{1}{|T(i\omega)|}, \tag{4.32}$$

where $T$ is the complementary transfer function. The inequality (4.32) tells that large relative perturbations are permitted as long as $T$ is small. A simple conservative estimate of the permissible relative error in the process transfer function is $1/M_t$ where

$$M_t = \max_\omega |T(i\omega)| = \max_\omega \Big|\frac{P(i\omega)C(i\omega)}{1 + P(i\omega)C(i\omega)}\Big| = \max_\omega \Big|\frac{G_l(i\omega)}{1 + G_l(i\omega)}\Big|, \tag{4.33}$$

is the largest magnitude of $|T|$. Notice that $M_t$ is also the largest gain of the transfer function from set point to output for a system with error feedback.

Equation 4.32 can also be written as

$$|\Delta P(i\omega)| < \frac{|P(i\omega)|}{|T(i\omega)|}. \tag{4.34}$$

It follows from this equation that the magnitude of the permissible error $|\Delta P(i\omega)|$ is small when $|P(i\omega)|$ is less than $|T(i\omega)|$. High model precision is thus required for frequencies where the gain of the closed-loop system is larger than the gain of the open-loop system.

### Graphical Interpretation of Constraint on Sensitivities

The requirements that the sensitivities are less than given values have nice geometric interpretations in the Nyquist plot. Since the sensitivity is defined by

$$S(i\omega) = \frac{1}{1 + G_l(i\omega)},$$

it follows that the sensitivity has constant magnitude on circles with center at the critical point $-1$. The condition that the largest sensitivity is less than $M_s$ is equivalent to the condition that the Nyquist curve of the loop transfer function is outside a circle with center at $-1$ and radius $1/M_s$.

There is a similar interpretation of the complementary sensitivity

$$T = \frac{G_l(i\omega)}{1 + G_l(i\omega)}.$$

Introducing

$$G_l(i\omega) = \mathrm{Re}G_l(i\omega) + i\mathrm{Im}G_l(i\omega) = x + iy,$$

we find that the magnitude of $T$ is given by

$$|T| = \frac{\sqrt{x^2 + y^2}}{\sqrt{(1+x)^2 + y^2}}.$$

The magnitude of the complementary sensitivity function is constant if

$$x^2 + y^2 = M_t^2((1+x)^2 + y^2) = M_t^2(1 + 2x + x^2 + y^2),$$

or

$$x^2\frac{M_t^2 - 1}{M_t^2} + 2x + y^2\frac{M_t^2 - 1}{M_t^2} + 1 = 0.$$

This condition can be written as

$$\begin{aligned} x^2 + 2\frac{M_t^2}{M_t^2 - 1}x + y^2 + \frac{M_t^2}{M_t^2 - 1} \\ = \left(x + \frac{M_t^2}{M_t^2 - 1}\right)^2 + y^2 + \frac{M_t^2}{M_t^2 - 1} - \left(\frac{M_t^2}{M_t^2 - 1}\right)^2 \\ = \left(x + \frac{M_t^2}{M_t^2 - 1}\right)^2 + y^2 - \frac{M_t^2}{(M_t^2 - 1)^2} = 0. \end{aligned}$$

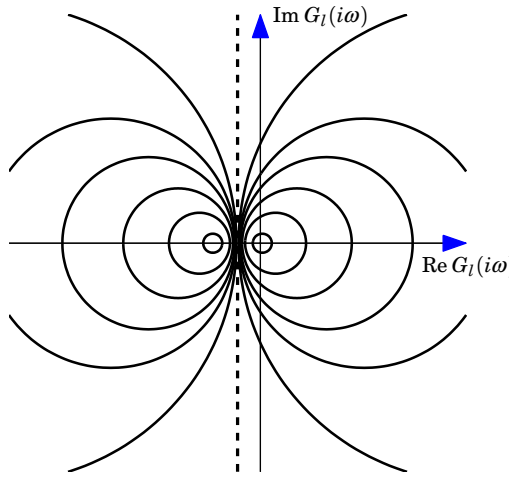


**Figure 4.14** Loci where the complementary sensitivity function has constant magnitude. The solid lines show points where the magnitude of the sensitivity function is $M_t = 1.1$, 1.2, 1.4, 1.5, 2, and 5 and the inverses of these values. The dashed line corresponds to $M_t = 1$.

This is a circle with center at $x = -M_t^2/(M_t^2 - 1)$ and $y = 0$, and with radius $r = M_t/(M_t^2 - 1)$. For $M_t = 1$ the circle degenerates to the straight line with $x = -0.5$. The requirement that the complementary sensitivity function is less than $M_t$ thus implies that the Nyquist curve is outside the corresponding circle. The loci of constant gain of the complementary sensitivity function $G_l$ are shown in Figure 4.14. Notice that the circles enclose the critical point $-1$. Notice also that the closed-loop transfer function is insensitive to variations at frequencies where the loop transfer function is far from the origin, particularly if the Nyquist curve is close to the straight line $\mathrm{Re} G_l(i\omega) = -0.5$. This implies that controllers with the property

$$T_i \approx T_{ar} \frac{2KK_p}{1 + 2KK_p} \tag{4.35}$$

are very robust. Compare with Section 6.3.

### Combined Sensitivities

The requirements that the maximum sensitivity is less than $M_s$ and the complementary sensitivity is less than $M_t$ imply that the Nyquist curve should be outside the corresponding circles. It is possible to find a slightly more conservative condition by determining a circle that encloses both circles as is illustrated in Figure 4.15. The radii and the centers of the circles are given in Table 4.1. In that table we have also given the circles that guarantee that both $M_t$ and $M_s$ are smaller than specified values. A particular simple criteria is obtained if it is required that $M_s = M_t$.

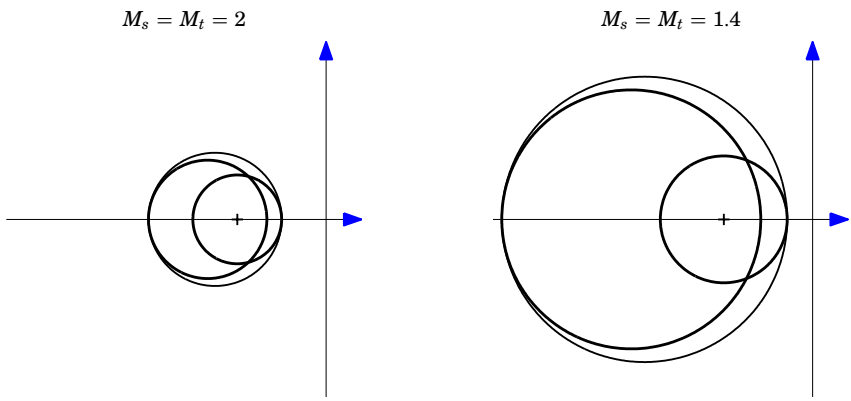


**Figure 4.15** Curves for constant sensitivity, constant complementary sensitivity, and constant combined sensitivity.

**Table 4.1** Center and radius of circles defining locus for constant sensitivity $M_s$, constant complementary sensitivity $M_t$, constant mixed sensitivity, and equal sensitivities $M = M_s = M_t$.

| Contour | Center | Radius |
|---|---|---|
| $M_s$ | $-1$ | $1/M_s$ |
| $M_t$ | $-\dfrac{M_t^2}{M_t^2-1}$ | $\dfrac{M_t}{M_t^2-1}$ |
| $M_s, M_t$ | $-\dfrac{M_s(2M_t-1)-M_t+1}{2M_s(M_t-1)}$ | $\dfrac{M_s+M_t-1}{2M_s(M_t-1)}$ |
| $M = M_s = M_t$ | $-\dfrac{2M^2-2M+1}{2M(M-1)}$ | $\dfrac{2M-1}{2M(M-1)}$ |

## 4.7 Robustness to Process Variations

Robustness to process variations is a key issue in control systems design. Process parameters can change for many reasons; they typically depend on operating conditions. Time delays and time constants often change with production levels. Parameters can also change because of aging of equipment. One of the key reasons for using feedback is that it is possible to obtain closed-loop systems that are insensitive to variations in the process.

The analysis of the sensitivity functions in Section 4.6 gives insight into the effects of process variations. Equation 4.25 shows the effect of small process variations on the closed-loop system. In particular it tells that a closed-loop system is insensitive to small process variations for frequencies where the sensitivity function is small.

The robustness inequality given by (4.32) tells that a closed-loop system will remain stable when the process is perturbed from $P(s)$ to $P(s) + \Delta P(s)$,

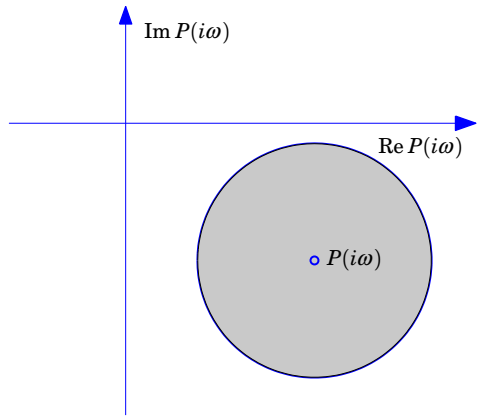


**Figure 4.16** Shaded circle shows permissible values of $P(i\omega) + \Delta P(i\omega)$ given by the inequality (4.32). The circle is drawn for $M_t = 2$.

where $\Delta P(i\omega)$ is a stable transfer function, if the perturbations are bounded by

$$\frac{|\Delta P(i\omega)|}{|P(i\omega)|} < \frac{1}{|T(i\omega)|}.$$

This equation is one of the reasons why feedback systems work so well in practice. The mathematical models used to design control system are often strongly simplified. There may be model errors and the properties of a process may change during operation.

Equation (4.32) implies that the closed-loop system will be stable for substantial variations in the process dynamics. The closed-loop system is stable if, for all $\omega$, the perturbed process transfer function $P(i\omega) + \Delta P(i\omega)$ lies in a circle with center at $P(i\omega)$ and radius $1/|T(i\omega)|$, see Figure 4.16. For a system designed with $M_t = 2$ it is possible to change the process gain by factors in the range 0.5 to 1.5 and the phase can be changed by 60°. For a system with $M_t = 1.414$ the gain can be changed by factors in the range 0.3 to 1.7, and the phase can be changed by 45°.

### The Cancellation Problem

The sensitivities depend on the loop transfer function $G_l = PC$. Robustness criteria based on sensitivities can give misleading results when there are factors in the process and controller transfer functions that cancel each other. We will illustrate what happens with an example.

EXAMPLE 4.3—CANCELLATIONS
Consider a process with the transfer function

$$P(s) = \frac{1}{s^2 + 2\zeta as + a^2},$$

and a controller with the transfer function

$$C(s) = \frac{50(s^2 + 2\zeta as + a^2)}{s(s^2 + 10s + 50)}.$$

This controller is a combination of a PID controller with a filter to provide high-frequency roll-off and a notch filter to reduce the excitation of the low-frequency oscillatory mode. The loop transfer function is

$$G_l(s) = \frac{50}{s(s^2 + 10s + 50)}.$$

Notice that the oscillatory modes vanish because the same factor appears both in the controller and the process. The sensitivity functions are

$$S(s) = \frac{s(s+5)^2}{s^3 + 10s^2 + 50s + 50}$$
$$T(s) = \frac{1}{s^3 + 10s^2 + 50s + 50}.$$

With the numerical values $a = 0.5$ and $\zeta = 0.02$ we get $M_s = 1.2$ and $M_t = 1$. A casual application of the robustness inequality (4.32) may lead us to believe that the closed-loop system is robust. However, if a controller is designed based on the nominal value $a = 0.5$ and if the process parameter is changed by 5 percent to $a = 0.4775$ the system becomes unstable. The reason is that if we interpret the parameter variation as an additive disturbance in the process model the small perturbation in the process parameter $a$ translates as a much larger additive disturbance because it is associated with a resonant mode with a very small relative damping. □

The controller in the example is not a good design because it is bad practice to cancel slow process poles.

### Other Robustness Measures

There are other robustness results that permit more realistic process variations than the stable additive perturbation used in the robustness inequality (4.32). One result represents the process transfer function as

$$P(s) = \frac{N(s)}{D(s)}$$

where $N(s)$ and $D(s)$ are stable transfer functions. The results state that the system is stable for variations $\Delta N$ and $\Delta D$ such that

$$\max(|N(i\omega)|, |D(i\omega)|) = \bar{\sigma} \begin{pmatrix} \frac{1}{1+P(i\omega)C(i\omega)} & \frac{P(i\omega)}{1+P(i\omega)C(i\omega)} \\ \frac{C(i\omega)}{1+P(i\omega)C(i\omega)} & \frac{P(i\omega)C(i\omega)}{1+P(i\omega)C(i\omega)} \end{pmatrix} \tag{4.36}$$
$$= \frac{\sqrt{(1+|P(i\omega)|^2)(1+|C(i\omega)|^2)}}{|1+P(i\omega)C(i\omega)|} = \Sigma(\omega),$$

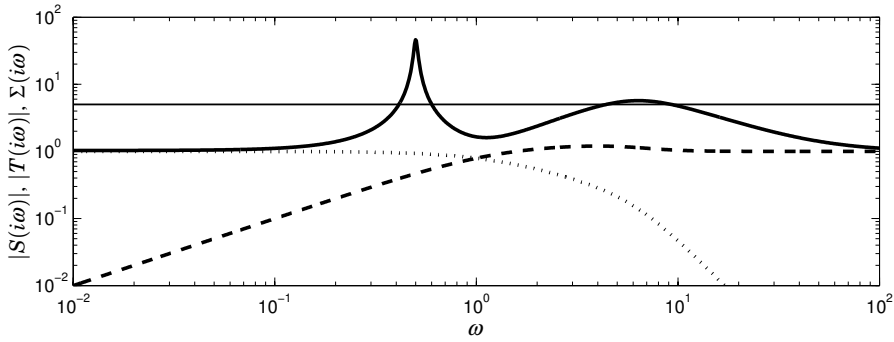


**Figure 4.17** The magnitudes of the sensitivity function $|S(i\omega)|$ (dotted), the complementary sensitivity function $|S(i\omega)|$ (dashed) and the largest singular value $\Sigma(i\omega)$ (solid) for the system in Example 4.3.

where $\bar{\sigma}$ is the largest singular value. The parameter

$$M_\sigma = max_\omega \Sigma(\omega)$$

is a robustness measure. The robustness condition (4.32) requires that the process perturbation $\Delta P(s)$ is a stable transfer function. Criteria based on $M_\sigma$ do not have this limitation because it permits more general perturbations of the process, for example, changing a small stable pole, an integrator, or an unstable pole. It also covers the situation when there are cancellations of poles and zeros. To have good robustness the parameter $M_\sigma$ should be less than 3 to 5. Notice that $M_\sigma$ is larger than both $M_s$ and $M_t$.

To illustrate the effectiveness of $M_\sigma$ we apply it to Example 4.3. Figure 4.17 shows $|S(i\omega)|$, $|T(i\omega)|$, and $\Sigma(\omega)$ for the nominal system in Example 4.3. We have $M_\sigma = 46$; since this is much larger than 5 it follows that the closed-loop system has very poor robustness.

Another way of investigating robustness is to explore variations in process parameters required to make the closed-loop system unstable. Changes in gain and time constants can be captured by replacing $P(s)$ by $\kappa P(\alpha s)$. Process variations that make the system unstable are given by

$$\kappa P(i\alpha\omega)C(i\omega) + 1 = 0.$$

Solving for $\alpha$ and $\kappa$ for all $\omega$ gives the functions $\kappa(\omega)$ and $\alpha(\omega)$. Peter Hansen has suggested the following robustness index

$$R_{ph} = \min_\omega(\log|\kappa(\omega)| + \log|\alpha(\omega)|). \tag{4.37}$$

This measure is a generalization of gain margin and delay margin.

The largest singular value $M_\sigma$ and the robustness measure $R_{ph}$ are more complicated than $M_s$ and $M_t$, and we will therefore mostly use $M_s$ and $M_t$. It should, however, be kept in mind that evaluating robustness requires some care, particularly when there are cancellations and when the loop transfer

function has high peaks above the gain crossover frequency. This is typically the cases for motion control with systems having mechanical resonances and for predictive controllers investigated in Chapter 8.

## 4.8 Quantifying the Requirements

Having understood the fundamental properties of the basic feedback loop we will now quantify the requirements on a typical control system. To do this it is necessary to have a clear understanding of the primary goal of control. Control problems are very rich as was discussed in Section 4.2. In general, we have to consider

- Load disturbance attenuation
- Measurement noise response
- Robustness to process uncertainties
- Set-point response

The emphasis on the different factors depends on the particular problem. Robustness is important for all applications. Set-point following is the major issue in motion control, where it is desired that the system follows commanded trajectories. In process control, the set point is normally kept constant most of the time; changes are typically made only when production is altered. Rejection of load disturbances is instead the key issue in process control. There are also situations where the purpose of control is not to keep the process variables at specified values. Level control in buffer tanks is a typical example. The reason for using a buffer tank is to smooth flow variations. In such a case the tank level should fluctuate within some limits. A good strategy is to take no control actions as long as the tank level is within certain limits and only apply control when the level is close to the limits. This is called averaging control or surge tank control. There are special strategies developed for dealing with such problems, techniques such as gain scheduling have also been applied. This is discussed in Section 9.3.

The linear behavior of the system is completely determined by the *Gang of Six* (4.3). Neglecting set-point response it is sufficient to consider the *Gang of Four* (4.4). Specifications can be expressed in terms of these transfer functions.

A significant advantage with a structure having two degrees of freedom, or set-point weighting, is that the problem of set-point response can be decoupled from the response to load disturbances and measurement noise. The design procedure can then be divided into two independent steps.

- First design the feedback controller $C$ that reduces the effects of load disturbances and the sensitivity to process variations without introducing too much measurement noise into the system.
- Then design the feedforward $F$ to give the desired response to set points.

We will now discuss how specifications can be expressed in terms properties of the transfer functions (4.4).

### Response to Load Disturbances

An estimate of the effectiveness of a control system to reject disturbances is given by (4.26), which compares the outputs of a closed- and an open-loop system when the disturbances are the same. The analysis shows that disturbances with frequencies less than the sensitivity crossover frequency $\omega_{sc}$ are attenuated by feedback and that the largest amplification of disturbances is the maximum sensitivity $M_s$.

We will now turn specifically to load disturbances which are disturbances that drive the process variables away from their desired values. Attenuation of load disturbances is a primary concern for process control. This is particularly the case for regulation problems where the processes are running in steady state with constant set point. Load disturbances are often dominated by low frequencies. Step signals are therefore used as prototype disturbances. The disturbances may enter the system in many different ways. If nothing else is known, it is often assumed that the disturbances enter at the process input. The response of the process variable is then given by the transfer function

$$G_{xd} = \frac{P}{1+PC} = PS = \frac{T}{C}. \tag{4.38}$$

Since load disturbances typically have low frequencies it is natural that the criterion emphasizes the behavior of the transfer function at low frequencies. Filtering of the measurement signal has only marginal effect on the attenuation of load disturbances because the filter only attenuates high frequencies. For a system with $P(0) \neq 0$ and a controller with integral action control the controller gain goes to infinity for small frequencies, and we have the following approximation for small $s$;

$$G_{xd} = \frac{T}{C} \approx \frac{1}{C} \approx \frac{s}{k_i}. \tag{4.39}$$

Since load disturbances typically have low frequencies this equation implies that integral gain $k_i$ is a good measure of load disturbance rejection.

EXAMPLE 4.4—LOAD DISTURBANCE ATTENUATION

Consider a process with the transfer function $P = (s+1)^{-4}$ and a PI controller with $k = 0.5$ and $k_i = 0.25$. The system has $M_s = 1.56$ and $\omega_{ms} = 0.494$. Figure 4.18 shows the magnitude curve of the transfer function (4.38). The figure shows clearly that feedback reduces the low-frequency gain significantly compared with the open-loop system. The dashed-dotted line in the figure shows the gain curve for the transfer function $s/k_i$. The figure shows clearly that this is a very good approximation of $G_{xd}$ for low frequencies, approximately up to $\omega_{ms}$. Integral gain $k_i$ is a good measure of load frequency disturbance attenuation. For high frequencies the load disturbance rejection is given by the process dynamics; feedback has no influence. The sensitivity crossover frequency is $\omega_{sc} = 0.25$, which is close to $k_i$.

Attenuation of load disturbances can also be characterized in the time domain by showing the time response due to a representative disturbance. This

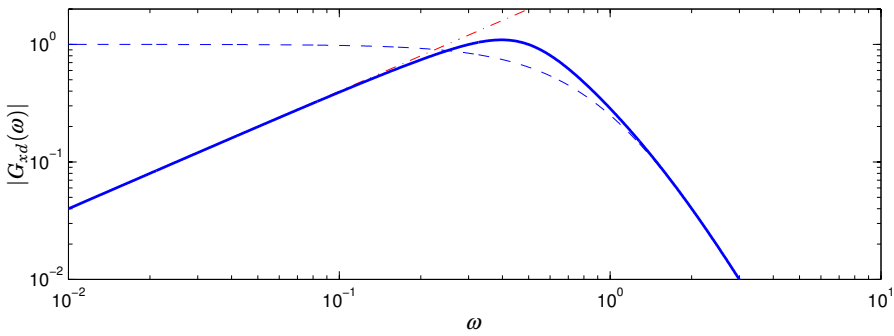


**Figure 4.18** The gain of the transfer function $G_{xd}$ from load disturbance to process variable for PI control ($k = 0.5$, $T_i = 2.0$) of the process $P = (s+1)^{-4}$. The dashed dotted curve shows the gain of $s/k_i$, and the dashed curve shows gain of the process transfer function $P$.

is illustrated in Figure 4.19, which shows the response of the process output to a unit step disturbance at the process input. The output has its maximum $y_{max} = 0.66$ for $t_{max} = 5.62$. Furthermore, $t_{max}\omega_{ms} = 2.76$, integrated error $IE = 4.00$ and integrated absolute error $IAE = 4.26$. □

The steady-state error caused by a unit step load disturbance for proportional control is

$$e_{ss} = \frac{P(0)}{1 + kP(0)}, \tag{4.40}$$

where $k$ is the proportional gain of the controller. As indicated in Figure 4.19, the steady-state error for proportional control can be used as an approximation of the largest error for PID control. For the system in Example 4.4 we have $P(0) = 1$ and $k = 0.5$ and (4.40) gives the estimate $e_{max} \approx e_{ss} = 1/1.5 = 0.67$ which is close to the correct value 0.66.

### Response to Measurement Noise

An inevitable consequence of using feedback is that measurement noise is fed back into the system. Measurement noise, which typically has high frequencies, generates undesirable control actions and variations in the process variable. Rapid variations in the control variable are detrimental because they cause wear in valves and motors and they even saturate the actuator. It is important to keep these variations at a reasonable level. A typical requirement is that the variations are only a fraction of the span of the control signal. The variations can be influenced by filtering and by proper design of the high-frequency properties of the controller.

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

$$G_{un} = \frac{C}{1 + PC} = CS = \frac{T}{P}. \tag{4.41}$$

For low frequencies (small $s$) the transfer function approaches $1/P(0)$ and for

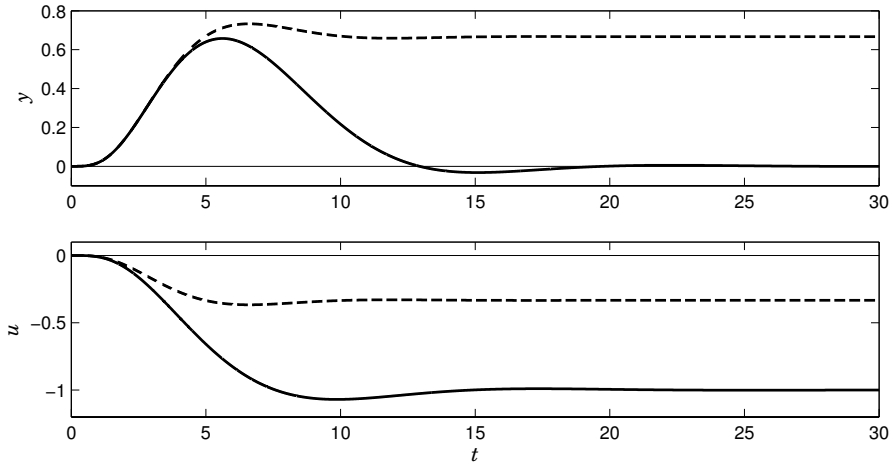


**Figure 4.19** Response to a load disturbance in the form of a unit step with a PI controller having parameters $k = 0.5$ and $k_i = 0.25$ and the process $P = (s+1)^{-4}$. The dashed curve shows the response to a proportional controller with gain $k = 0.5$.

high frequencies (large $s$) we have approximately

$$G_{un} \approx C.$$

For an ideal PID controller the transfer function $G_{un}$ becomes infinite for large $s$ which clearly indicates the necessity to filter the derivative, as discussed in Section 3.3. We illustrate with an example.

EXAMPLE 4.5—EFFECT OF FILTERING
Figure 4.20 shows the gain curve of the transfer function (4.41) for PID control of the process $P = (s+1)^{-4}$. The dashed line is for a controller with a first-order filter of the derivative and the full line for a controller with a second-order filter of the measured signal. The significant differences in the transfer functions for high frequencies is a good motivation for preferring the controller with filtering of the measurement signal. For low frequencies (small $s$) the transfer function approaches $1/P(0)$. □

A simple measure of the effect of measurement noise is the largest gain of the transfer function $G_{un}$,

$$M_{un} = \max_{\omega} |G_{un}(i\omega)|. \tag{4.42}$$

For PI control the gain of the transfer function $G_{un}$ has a peak close to the peak of the sensitivity function and we have approximately

$$M_{un} \approx M_s K. \tag{4.43}$$

For PID control the gain of the transfer function $G_{un}$ typically has two local maxima, one is close to the maximum of the sensitivity function. The other peak is larger

$$M_{un} \approx k_d/T_d, \tag{4.44}$$

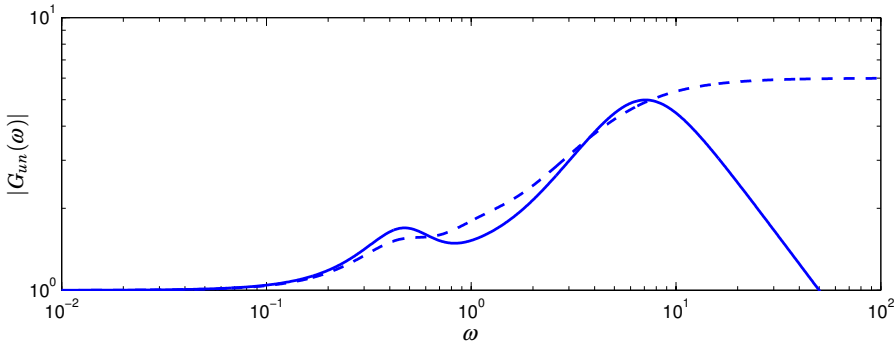


**Figure 4.20** The magnitude of the transfer function $G_{un} = CS$ for PID control ($k = 1$, $T_i = 2$, $T_d = 1$, $T_f = 0.2$) of the process $P = (s + 1)^{-4}$. The solid line represents a controller with a second-order noise filter of the measured signal (3.16) and the dashed line a controller with a first-order filter of the derivative (3.15).

and it occurs close to the frequency $1/T_f$.

If the standard deviation of the measurement noise is $\sigma_n$, a crude estimate of the variations in the control signal is $M_{un}\sigma_n$. More accurate assessment can be made if the power spectrum $\phi_n$ of the measurement noise is known. The standard deviation of the control signal is then given by

$$\sigma_u^2 = \int_{-\infty}^{\infty} |G_{un}(i\omega)|^2 \phi_n(\omega) d\omega. \tag{4.45}$$

However, it is rare that such detailed information is rarely available for typical applications.

### Robustness to Process Variations

The inverse of the maximum sensitivity is the shortest distance from the critical point $-1$ to the Nyquist curve of the loop transfer function.

The sensitivity to small variations in process dynamics is captured by the sensitivity function. We have

$$\frac{dT}{T} = S\frac{dP}{P}.$$

Variations in process dynamics thus have small influence on the closed-loop system for frequencies where the sensitivity function is small.

Variations in process dynamics may also lead to instability. The condition

$$\frac{|\Delta P(i\omega)|}{|P(i\omega)|} < \frac{1}{|T(i\omega)|}$$

guarantees that a variation $\Delta P(i\omega)$ in the process transfer function does not make the system unstable. Robustness to process variations is thus captured by the sensitivity and the complementary sensitivity functions. Simple measures are the maximum sensitivity $M_s$, the maximum of the complementary

sensitivity $M_t$, or the largest combined sensitivity $M$. Typical values of the sensitivities are in the range of $1.2 - 2.0$.

Other measures are the gain margin $g_m$ (typically 2 to 8), the phase margin $\varphi_m$ (typically 30° to 60°), or the stability margin $s_m = 1/M_s$ (typically 0.5 to 0.8). Compare with Section 4.4.

### Trade-offs

Load disturbance attenuation is captured by integral gain $k_i$. It follows from (4.39) that attenuation of low-frequency disturbances is approximately inversely proportional to $k_i$. Injection of measurement noise is captured by the noise gain $M_{un}$. It follows from (4.42) that $M_{un}$ gives the gain from measurement noise to control variable. The trade-off between load disturbance attenuation and injection of measurement noise can thus be achieved by balancing $k_i$ and $M_{un}$.

### Set-Point Response

By using a controller with two degrees of freedom it is possible to obtain any desired response to set-point changes. This will be discussed further in Chapter 5. The limitations are given by the permissible magnitude of the control signal. In some cases only the control error is measured. A controller with two degrees of freedom then cannot be used and the response to set points has to be handled by proper choosing of the controller transfer function. Large overshoots can be avoided by requiring low values of $M_t$.

### Summary

Summarizing we find that the behavior of the system can be characterized in the following way. The transfer function from load disturbance to process variable is

$$G_{yd} = \frac{P}{1 + PC} = PS \approx \frac{s}{k_i}, \tag{4.46}$$

where the approximation holds for low frequencies.

The effect of measurement noise can be captured by the noise gain

$$M_{un} = \max_{\omega} |G_{un}(i\omega)| \approx \begin{cases} kM_s & \text{for PI control} \\ k_d/T_f & \text{for the PID controller (3.16),} \end{cases} \tag{4.47}$$

which strongly depends on the filtering of measurement noise.

Stability and robustness to process uncertainties can be expressed by the sensitivity function and the complementary sensitivity function

$$S = \frac{1}{1 + PC}, \qquad T = \frac{PC}{1 + PC},$$

where the largest values of the sensitivity functions $M_s$ and $M_t$ are good quantitative measures. The parameter $1/M_s$ is the shortest distance from the critical point to the Nyquist curve of the loop transfer function.

Essential features of load disturbance attenuation, measurement noise injection, and robustness can thus be captured by four parameters $k_i$, $M_{un}$, $M_t$,

and $M_s$. An attractive feature of this choice of parameters is that $k_i$ and $M_{un}$ are directly related to the controller parameters and that there are good design methods that can guarantee given $M_s$ and $M_t$.

## 4.9 Classical Specifications

The specifications we have given have the advantage that they capture robustness as well as the responses to load disturbances, measurement noise, and set points with only four parameters. Unfortunately, it has been the tradition in PID to judge a system based on one response only, typically the response of the output to a step change in the set point. This can be highly misleading as we have discussed previously. A large number of different parameters have also been used to characterize the responses. For completeness and to connect with classical literature on PID control some of the classical specifications will be summarized in this section.

### Criteria Based on Time Responses

Many criteria are related to time responses, for example, the step response to set-point changes or the step response to load disturbances. It is common to use some feature of the error typically extrema, asymptotes, areas, etc.

The maximum error $e_{\max}$ is defined as

$$\begin{aligned} e_{\max} &= \max_{0 \le t < \infty} |e(t)| \\ T_{\max} &= \arg\max |e(t)|. \end{aligned} \tag{4.48}$$

The time $T_{\max}$ where the maximum occurs is a measure of the response time of the system. The integrated absolute error (IAE) is defined as

$$IAE = \int_0^\infty |e(t)| dt. \tag{4.49}$$

A related error is integrated error (IE), defined as

$$IE = \int_0^\infty e(t) dt. \tag{4.50}$$

The criteria IE and IAE are the same if the error does not change sign. Notice that IE can be very small even if the error is not. For IE to be relevant it is necessary to add conditions that ensure that the error is not too oscillatory. The criterion IE is a natural choice for control of quality variables for a process where the product is sent to a mixing tank. The criterion may be strongly misleading, however, in other situations. It will be zero for an oscillatory system with no damping. It will also be zero for a control loop with two integrators.

There are many other criteria, for example, the time multiplied absolute error, defined by

$$ITNAE = \int_0^\infty t^n |e(t)| dt. \tag{4.51}$$

The integrated squared error (ISE) is defined as

$$ISE = \int_0^\infty e(t)^2 dt. \tag{4.52}$$

There are other criteria that take account of both input and output signals, for example, the quadratic criterion

$$QE = \int_0^\infty (e^2(t) + \rho u^2(t)) dt, \tag{4.53}$$

where $\rho$ is a weighting factor. The criteria IE and QE can easily be computed analytically, simulations are, however, required to determine IAE.

One reason for using IE is that its value is directly related to the parameter $k_i$ of the PID controller, as is illustrated by the following example.

EXAMPLE 4.6—INTEGRAL GAIN AND IE FOR LOAD DISTURBANCES
Consider the control law

$$u(t) = ke(t) + k_i \int_0^t e(t)dt - k_d \frac{dy}{dt}.$$

Assume that this controller gives a stable closed-loop system. Furthermore, assume that the error is zero initially and that a unit step load disturbance is applied at the process input. Since the closed-loop system is stable and has integral action the control error will go to zero. We thus find

$$u(\infty) - u(0) = k_i \int_0^\infty e(t)dt.$$

Since the disturbance is applied at the process input, the change in control signal is equal to the change of the disturbance. Hence, $u(\infty) - u(0) = 1$ and we get

$$IE = \int_0^\infty e(t)dt = \frac{1}{k_i} = \frac{T_i}{K}. \tag{4.54}$$

□

Integral gain $k_i$ is thus inversely proportional to the integrated error caused by a unit step load disturbance applied to the process input.

### Set-Point Response

Specifications on set-point following are typically expressed in the time domain. They may include requirements on rise time, settling time, decay ratio, overshoot, and steady-state offset for step changes in set point. These quantities are defined as follows, see Figure 4.21.

- The *rise time* $T_r$ is defined either as the inverse of the largest slope of the step response or the time it takes for the step response to change from 10 percent to 90 percent of its steady-state value.

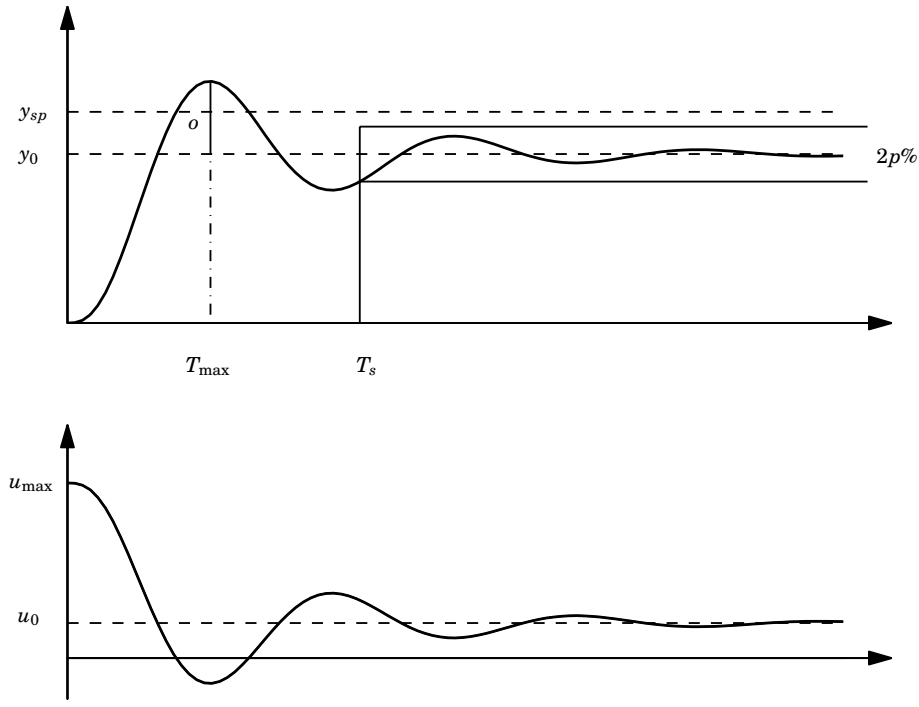


**Figure 4.21** Specifications on set-point following based on the time response to a unit step in the set point. The upper curve shows the response of the output, and the lower curve shows the corresponding control signal.

- The *settling time* $T_s$ is the time it takes before the step response remains within $p$ percent of its steady-state value. The values $p = 1, 2$, and 5 percent of the steady-state value are commonly used.
- The *decay ratio* $d$ is the ratio between two consecutive maxima of the error for a step change in set point or load; see Figure 2.35. The value $d = 1/4$, which is called quarter amplitude damping, has been used traditionally. This value is, however, normally too high, as will be shown later.
- The *overshoot* $o$ is the ratio between the difference between the first peak and the steady-state value of the step response. It is often given in percent. In industrial control applications it is common to specify a maximum overshoot of 8 to 10 percent. In many situations it is desirable, however, to have an over-damped response with no overshoot.
- The *steady-state error* $e_{ss} = y_{sp} - y_0$ is the steady-state control error $e$. This is always zero for a controller with integral action.

Actuators may have rate limitations, which means that step changes in the control signal will not appear instantaneously. In motion control systems it is often more relevant to consider responses to ramp signals instead of step signals.

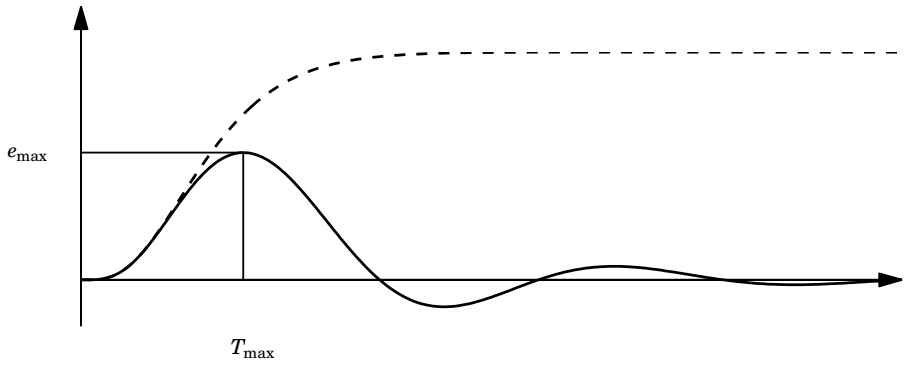


**Figure 4.22** The error due to a unit step load disturbance at the process input and some features used to characterize attenuation of load disturbances. The dashed curve show the open-loop error.

### Response to Load Disturbances

The response to load disturbances is of primary importance in process control. Figure 4.22 shows the output for a step disturbance in a load applied at the process input and some features that are used to characterize the response. The figure shows the *maximum error* $e_{\max}$, the time it takes to reach the maximum $T_{\max}$, and the settling time $T_s$. In addition to these numbers the integrated error (IE) or the integrated absolute error (IAE) are also commonly used to characterize the load disturbance response. The maximum error for a unit step and the time where this is reached can be approximated by

$$\begin{aligned} e_{\max} &= \frac{1}{1+kP(0)} \\ T_{\max} &\approx \frac{3}{\omega_{ms}}. \end{aligned} \tag{4.55}$$

We illustrate these estimates by an example.

EXAMPLE 4.7—ESTIMATING THE MAXIMUM ERROR
When a process with the transfer function $P(s) = (s+1)^{-4}$ is controlled by a PI controller having parameters $k = 0.78$ and $k_i = 0.38$, we have $\omega_{ms} = 0.559$, $e_{\max} = 0.59$, and $T_{\max} = 5.15$. The estimates above give $e_{\max} \approx 0.56$, $T_{\max} = 5.6$. □

### Criteria Based on Frequency Responses

Specifications can also be related to frequency responses. Since specifications were originally focused on set-point response it was natural to consider the transfer function from set point to output. A typical gain curve for this response is shown in Figure 4.23. It is natural to require that the steady-state gain is one. Typical specifications are then as follows:

- The *resonance peak* $M_p$ is the largest value of the frequency response.

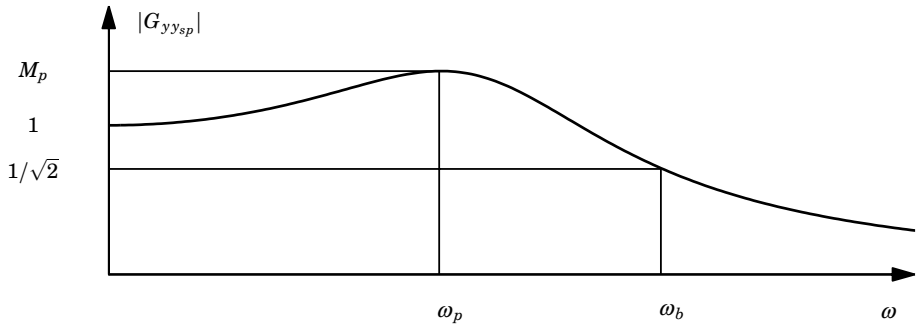


**Figure 4.23** Gain curve for transfer function from set point to output.

- The *peak frequency* $\omega_p$ is the frequency where the maximum occurs.
- The *bandwidth* $\omega_b$ is the frequency where the gain has decreased to $1/\sqrt{2}$.

For a system with error feedback the transfer function from set point to output is equal to the complementary transfer function, and we have $M_p = M_t$.

Specifications can also be related to the loop transfer function. Useful features that have been discussed previously are

- Gain crossover frequency $\omega_{gc}$.
- Gain margin $g_m$.
- Phase margin $\varphi_m$.
- Maximum sensitivity $M_s$.
- Frequency where the sensitivity function has its maximum $\omega_{ms}$.
- Sensitivity crossover frequency $\omega_{sc}$.
- Maximum complementary sensitivity $M_t$.
- Frequency where the complementary sensitivity function has its maximum $\omega_{mt}$.

### Relations between Time and Frequency Domain Specifications

There are approximate relations between specifications in the time and frequency domain. Let $G(s)$ be the transfer function from set point to output. In the time domain the response speed can be characterized by the rise time $T_r$, the average residence time $T_{ar}$, or the settling time $T_s$. In the frequency domain the response time can be characterized by the closed-loop bandwidth $\omega_b$, the gain crossover frequency $\omega_{gc}$, and the sensitivity frequency $\omega_{ms}$. The product of bandwidth and rise time is approximately constant, and we have

$$T_r \omega_b \approx 2. \tag{4.56}$$

It has previously been shown that

$$T_{ar} = -\frac{G'(0)}{G(0)};$$

see (2.16).

The overshoot of the step response $o$ is related to the peak $M_p$ of the frequency response in the sense that a larger peak normally implies a larger overshoot. Unfortunately, there are no simple relations because the overshoot also depends on how quickly the frequency response decays. For $M_p < 1.2$ the overshoot $o$ in the step response is often close to $M_p - 1$. For larger values of $M_p$ the overshoot is typically less than $M_p - 1$. These relations do not hold for all systems; there are systems with $M_p = 1$ that have a positive overshoot. These systems have transfer functions that decay rapidly around the bandwidth.

To avoid overshoots in systems with error feedback it is therefore advisable to require that the maximum of the complementary sensitivity function is small, say, $M_t = 1.1 - 1.2$ in order to avoid too large overshoot in the step response to command signals.

### Performance Assessment

Before designing a controller it is useful to make a preliminary assessment of achievable performance. It is interesting to know if a PID controller is sufficient or if the performance can be increased substantially by using a more complex controller. It is also interesting to know if a PI controller is sufficient or if derivative action gives significant improvements. To make the assessment we need some measure of performance. In this section we will use the gain crossover frequency $\omega_{gc}$ as a yardstick. When the phase margin is $60^\circ$ this frequency is equal to the sensitivity crossover frequency $\omega_{sc}$. Recall from Section 4.6 that disturbances with frequencies lower than $\omega_{sc}$ are reduced by feedback. For phase margins lower than $60^\circ$ we have $\omega_{sc} < \omega_{gc}$, and for larger phase margins we have $\omega_{sc} > \omega_{gc}$. Attenuation of load disturbances is thus improved with increasing gain crossover frequencies.

Process dynamics with non-minimum phase properties like a time delay imposes fundamental limitations on the achievable performance which can be expressed by the inequality

$$\omega_{gc} L < a, \tag{4.57}$$

where $a$ is a number less than 1. Since the true time delay $L$ is rarely known it can be approximated by the apparent time delay $L_a$. Figure 4.24 shows the product $\omega_{gc} L_a$ for a large batch of systems under robust PID control. The circles which represents FOTD systems show that the product is 0.5 for FOTD systems. For high order systems with lag dominated dynamics the product $\omega_{gc} L_a s$ is larger than 0.5 because the apparent time delay of the approximating FOTD model is larger than the true time delay of the system.

Consider a closed-loop system with a process having transfer function $P(s)$ and a controller with transfer function $C(s)$. The gain crossover frequency is defined by

$$\arg P(i\omega_{gc}) + \arg C(i\omega_{gc}) = -\pi + \varphi_m. \tag{4.58}$$

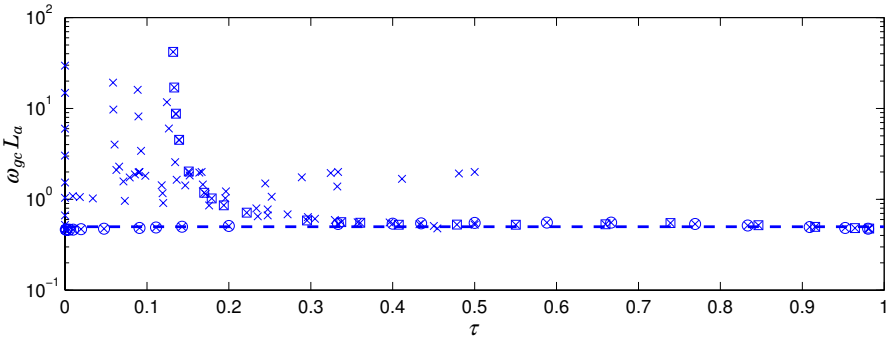


**Figure 4.24** The product of gain crossover frequency $\omega_{gc}$ and apparent time delay $L_a$ as a function of normalized dead time, for a large set of systems under PID control. The circles show results for FOTD systems and the squares for SOTD systems. The PID controllers are designed to give a combined sensitivity $M = 1.4$. All systems are described in Section 7.2. The dashed line gives the relation $\omega_{gc} La = 0.5$.

Notice that the units radians is used in this equation. A PD controller has a maximum phase lead of about 60° ($\pi/3$ rad), a proportional controller has zero phase lag, a PI controller has a phase lag of about 45° ($\pi/4$ rad), and a PID controller can have a phase lead of about 45°. If a phase margin of 45° is desired it follows from Equation 4.58 that crossover frequencies for PI, PID, and PD control are the frequencies where the process has phase-lags of 90°, 135°, and 195°, respectively. These frequencies are denoted as $\omega_{90}$, $\omega_{135}$, and $\omega_{195}$. An estimate of the controller gains required can be obtained by computing the process gains at the corresponding frequencies. Notice that this assessment only requires the process transfer function. We illustrate by two examples.

EXAMPLE 4.8—MULTI-LAG PROCESS
Consider a process with the transfer function

$$P(s) = \frac{1}{(s+1)^4}.$$

We have $\omega_{90} = 0.41$ and $K_{90} = 0.73$, where $K_{90}$ denotes the process gain at $\omega_{90}$. Furthermore, we have $\omega_{135} = 0.67$, $K_{135} = 0.48$, and $\omega_{195} = 1.14$, $K_{195} = 0.19$. We can thus expect that disturbances with frequencies lower than 0.4 rad/s can be reduced by PI control. Since $\omega_{135}$ is moderately larger than $\omega_{90}$ we can expect that a PI controller can be improved somewhat by introducing derivative action. The gain of a PID controller can be expected to be about twice as large as for PI control. Also notice that the apparent time delay is $L = 2.14$ and that $\omega_{gc} = 0.47$ which is in good agreement with (4.24). □

EXAMPLE 4.9—A LAG-DOMINATED PROCESS
Consider a process with the transfer function

$$P(s) = \frac{1}{(s+1)(0.1s+1)(0.01s+1)(0.001s+1)}.$$

**Table 4.2** Parameters of PI controllers for the process $P(s) = (s+1)^{-3}$ designed with different $M_s$.

| $M_s$ | $k$ | $k_i$ | $M_{un}$ | $b$ | $\omega_{ms}$ | $IAE$ | $T_s$ | $M_t$ |
|---|---|---|---|---|---|---|---|---|
| 1.2 | 0.355 | 0.171 | 0.426 | 1.00 | 0.671 | | | 1.00 |
| 1.4 | 0.633 | 0.325 | 0.866 | 1.00 | 0.74 | 3.07 | 10.3 | 1.00 |
| 1.6 | 0.862 | 0.461 | 1.379 | 0.93 | 0.79 | 2.28 | 7.87 | 1.05 |
| 1.8 | 1.056 | 0.580 | 1.901 | 0.70 | 0.83 | 2.00 | 6.77 | 1.24 |
| 2.0 | 1.222 | 0.685 | 2.444 | 0.50 | 0.86 | 1.89 | 6.27 | 1.45 |

We have $\omega_{90} = 3.0$, $K_{90} = 0.3$, $\omega_{135} = 9.9$, $K_{135} = 0.07$, and $\omega_{195} = 47.5$, $K_{195} = 0.004$. We can thus expect that disturbances with frequencies lower than 3 rad/s can be reduced by PI control. With a PID controller disturbances with frequencies up to 9.9 rad/s can be reduced. In this case, there are significant performance benefits from using derivative action. Since $\omega_{195}$ is much larger than $\omega_{135}$ there may be substantial benefits by using more complex controllers. Since the process gain $K_{135}$ is so low the improved benefits require controllers with high gain, and the benefits may be not be realizable unless sensor noise if very low. □

### Design Parameters

In control system design and implementation it is convenient to have a parameter that can be changed to influence the key trade-offs in a design problem. Performance expressed by fast response time and good attenuation of load disturbances can be obtained, but large control signals may be required. Stricter requirements on robustness may lead to poorer performance.

The trade-off between performance and robustness varies between different control problems. Therefore, it is desirable to have a design parameter to change the properties of the closed-loop system. Ideally, the parameter should be directly related to the performance or the robustness of the system; it should not be process oriented. There should be good default values so a user is not forced to select some value. This is of special importance when the design procedure is used for automatic tuning. The design parameter should also have a good physical interpretation and natural limits to simplify its adjustment.

The behavior of a system can often be characterized by a few dominant poles that are close to the origin. When there is one real dominant pole this pole can be used as a design parameter. This is used, for example, in the design method Lambda Tuning, which will be discussed in Section 6.5. When the dominant poles are complex the distance from the origin of the poles $\omega_0$ and their relative damping $\zeta$ are good design parameters. This applies to controllers based on pole placement design, which will be discussed in Section 6.4. The maximum sensitivity $M_s$ or the combined sensitivity $M$ are good design variables for regulation problems. This is illustrated in Figure 4.25, which shows the effect of $M_s$ on time and frequency responses for a PI controller, and in Table 4.2,

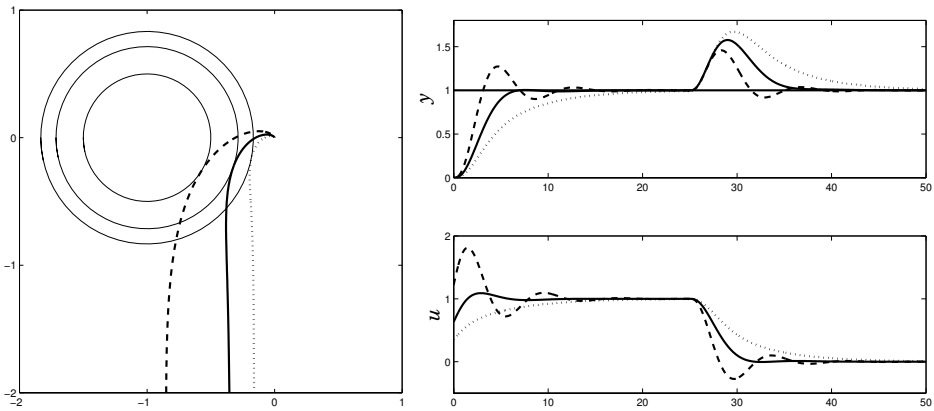


**Figure 4.25** Illustrates the effects of using $M_s$ as a design parameter. The left curves show the Nyquist plots of the loop transfer functions together with the circles of constant $M_s$=1.2 (dotted), 1.4, 2.0 (dashed). The curves on the right show process outputs and control signals for the different values of the design parameters.

which gives numerical values of controller parameters and various criteria. The step response for $M_s = 1.2$ has no overshoot and relatively long settling time. The settling time decreases and the overshoot increases as the value of $M_s$ is increased. Notice that set-point weighting is used for larger values of $M_s$ to reduce the overshoot. The performance also increases with increasing $M_s$. The values of IAE decrease with about a factor of 2. Apart from use in design it is also possible to implement systems where the user can adjust the design parameters on line.

## 4.10 Summary

In this section we have summarized some important issues for the design of control systems, with particular emphasis for PID control. A discussion of the basic feedback loop showed that it is necessary to consider six transfer functions (the Gang of Six) to determine the properties of a feedback loop. This is severely neglected in most elementary texts in control and in the literature of PID controllers. The notion of stability was then discussed. This is important because the risk for instability is the main disadvantage of feedback. Stability criteria and stability margins were also introduced. The stability criteria also made it possible to obtain the parameter regions which give a stable closed-loop system under PID control. Characterization of a closed-loop system by its poles and zeros give very valuable insight, and it is also closely related to many design methods. The sensitivity function and the complementary sensitivity functions, which are useful to express the robustness to parameter variations, were also introduced. The problem of controller design was then discussed, and a number of criteria used to give specifications on a control system were also introduced. The key factors are load disturbances, measurement noise, robustness, and set-point response. A nice result is that for systems having

two degrees of freedom it is possible to design for disturbances and robustness. The desired set-point response can then be obtained by using feedforward. For PID control set-point weighting is a special form of controller with two degrees of freedom that often is sufficient. It is also shown that the key requirements can be parameterized in a simple way. Load disturbance response is captured by integral gain of the controller $k_i$. Effects of measurement noise are captured by the noise gain $M_{un}$, which has a simple relation to controller parameters. Robustness is captured by the maximum sensitivities $M_s$ and $M_t$.

## 4.11 Notes and References

Control system design is complicated because many factors have to be considered and trade-offs have to be made. It is therefore natural that it took time before a good understanding was developed. Early work on control design was based on the differential equations describing the closed-loop system. A typical approach was to adjust the controller parameters so that the dominant closed-loop poles had desired properties. Systematic methods for control system design appeared in the 1940s when the field of control emerged. The design methods were based on frequency response, computations were based on graphics, and modeling was often done experimentally by perturbing the system with sinusoidal signals; see [Bode, 1945; James *et al.*, 1947; Brown and Campbell, 1948; Chestnut and Mayer, 1959]. It is noteworthy that particular emphasis was given to robustness to process variations. The insightful book [Horowitz, 1963] gives a mature account. This book also emphasizes the important concept of controllers that have two degrees of freedom. Such controllers admit a decoupling of the responses to set points and load disturbances.

There was a paradigm shift in the 1960s when differential equations reappeared in the name of state-space systems; see [Zadeh and Desoer, 1963]. This coincided with the appearance of digital computers, which permitted efficient numerical computations. The important ideas of optimal control and Kalman filtering are key contributions; see [Bellman, 1957; Kalman, 1960; Kalman and Bucy, 1961; Kalman, 1961; Pontryagin *et al.*, 1962; Athans and Falb, 1966; Bryson and Ho, 1969].

There was a very dynamic development of theory, many design methods and efficient computational techniques were also developed; see [Boyd and Barratt, 1991].

The robustness issue was unfortunately neglected for a long period. This was remedied with the emergence of the so called $\mathcal{H}_\infty$ theory, which led to a reconciliation with the classical frequency response methods. The books [Doyle *et al.*, 1992; Zhou *et al.*, 1996; Skogestad and Postlethwaite, 1996] give a balanced perspective. The robustness criteria $M_s$, $M_t$, and $M_\sigma$ are results of robust control theory. An interesting novel robustness criterion which focuses on variations in the process parameters has been suggested in [Hansen, 2000] and [Hansen, 2003]. The question of fundamental limitations is closely related to robustness as is discussed by [Åström, 2000]. For process control the true time delay is a key limiting factor. Notice that the true time delay can be different from the apparent time delay obtained when fitting FOTD models.

Many practitioners of control have been fully aware of the importance of the compromise between performance and robustness; see [Shinskey, 1990], and it is now pleasing to see that robust control theory has made it possible to merge theory and practice; see [Panagopoulos and Åström, 2000].

In the literature on PID control there has been a long discussion, whether tuning should be based on response to set-point changes or load disturbances. It is surprising that so many papers just show the response of process output to a step change in the set point. Since steady-state regulation is the essential problem in process control, load-disturbance responses are more important than responses to set points as has been emphasized many times by Shinskey; see for example [Shinskey, 1996]. One of the useful conclusions of robust control theory is that six responses are required to get a complete understanding of a closed loop system,

Another lesson from robust control theory is that high-frequency roll-off improves robustness. This is a good reason to use effective filtering in PID control.