Chapter 9

Robustness and Performance

Quotation

Authors, citation.

This chapter treats robustness and performance. It begins with analysis of a simple controller. It is shown that seemingly reasonable design choices gives a closed loop system that is extremely sensitivity to parameter variations. New concepts, which give intuition and make it possible to characterize robustness and performance quantitatively, are introduced. Properties that fundamentally limits achievable performance are also discussed.

9.1 Introduction

Fundamental properties of feedback systems will be explored in this chapter. There are many requirements on a control system the ability to follow reference signals, to suppress external disturbances and effects if measurement noise and process variations. It is important to understand all these issues both to be able to analyze, and specify a system. Trade-offs between robustness and performance is a key issue for design.

Robustness is the ability of the closed loop system to be insensitive to component variations. It is one of the most useful properties of feedback. Robustness is also what make it possible to design feedback system based on strongly simplified models. It is therefore essential to have a good understanding of robustness and to have ways of expressing it quantitatively. One of the key questions is to describe variations in system dynamics. Using state space concepts this can be done by varying the parameters of a system. There are however many variations that are not captured by such an approach. For example, some dynamic phenomena many have been neglected.

If they were included the model would have had more states. There may also be small time delays that have been neglected in a model. For linear model variations are well captured by transfer functions. Using transfer functions it is possible make perturbations that correspond to additional dynamics as well as time delays.

It is also necessary to have quantitative ways to express how well a feedback system performs. Measures of performance and robustness are closely related. The sensitivity function introduced in Section **??** expresses both how well disturbances are affected by feedback and how sensitive the closed loop system is to small perturbations of the process dynamics. The fact that similar concepts are used makes it easy to make trade-offs between robustness and performance.

The structure of a feedback controller is another fundamental issue. There is a sharp distinction between two classes of systems. In *systems with error feedback* only the error signal is accessible through the sensors. A typical example is track following in a CD player where only the deviation from the track is measured. In other systems both the reference and the process output are available for measurement. It is then possible to completely separate command signal following from robustness and disturbance attenuation by using a *controller with two degrees of freedom.* The feedback is designed to deal with disturbances and robustness and feedforward is used to obtain the desired response to command signals. It is possible to have a complete decoupling of command signal following from disturbance attenuation and robustness. For systems with error feedback all issues must be deal with using feedback.

When specifying the performance of a control system it is common practice to relate it to how well the system is able to follow command signals. This is not sufficient. For linear systems where the controller has error feedback it is necessary to consider four transfer functions, the *Gang of Four*, to have a complete understanding of the behavior of the system. For a system with a controller having two degrees of freedom it is necessary to consider six transfer functions, the *Gang of Six*, to completely specify the behavior of the system. The specifications should also reflect this. It is interesting that in spite of all complications specifications for many problems can be captured by a few parameters.

Some systems are intrinsically difficult to control, typical examples are unstable systems with time delay. It is important to understand the underlying reasons and to have some feel for how they relate to fundamental system properties and to sensing and actuation. If the difficulties can be spotted at an early stage of the design they can be remedied by moving or adding sensors and actuators or by redesigning the system. This is one of the strong reasons for investigating dynamics and control at an early stage of a design. It is also a reason why design engineers should be aware of dynamics and control.

9.2 An Example

We will begin by a simple applications of controller design using state feedback and observers. Consider the system

$$\begin{aligned}\frac{dx}{dt} &= Ax + Bu = \begin{bmatrix}-1 & 0\\ 1 & 0\end{bmatrix} x + \begin{bmatrix}a-1\\ 1\end{bmatrix} u\\ y &= Cx = \begin{bmatrix}0 & 1\end{bmatrix} y.\end{aligned} \tag{9.1}$$

The system has the transfer function

$$G_P(s) = C[sI - A]^{-1}B = \frac{s+a}{s(s+1)} \tag{9.2}$$

State Feedback

We will begin by designing a state feedback assuming all states can be measured. The reachability matrix

$$W_r = \begin{bmatrix}a-1 & -a+1\\ 1 & a-1\end{bmatrix}$$

has the determinant $\det W_r = a(a-1)$. The system is thus reachable if a is neither 0 nor 1. A state feedback will first be designed assuming that all states are measured. The feedback

$$u = -l_1x_1 - l_2x_2 = -Lx \tag{9.3}$$

gives the closed loop system

$$\frac{dx}{dt} = (A - BL)x$$

where

$$A - BL = \begin{bmatrix}-1 & 0\\ 1 & 0\end{bmatrix} - \begin{bmatrix}a-1\\ 1\end{bmatrix}\begin{bmatrix}l_1 & l_2\end{bmatrix} = \begin{bmatrix}-1-(a-1)l_1 & -(a-1)l_2\\ 1-l_1 & -l_2\end{bmatrix}.$$

This matrix has the characteristic polynomial

$$\det \begin{bmatrix} s+1+(a-1)l_1 & (a-1)l_2 \\ l_1 - 1 & s+l_2 \end{bmatrix} = s^2 + (1+(a-1)l_1 + l_2)s + al_2.$$

Assume that a closed loop system with the characteristic polynomial

$$s^2 + 2\zeta_c\omega_c s + \omega_c^2$$

is desired. Equating coefficients of equal power of s gives

$$\begin{aligned} 1+(a-1)l_1 &= 2\zeta_c\omega_c \\ al_2 &= \omega_c^2. \end{aligned}$$

Solving this equation gives the following feedback gains

$$\begin{aligned} l_1 &= \frac{2a\zeta_c\omega_c - a - \omega_c^2}{a(a-1)} \\ l_2 &= \frac{\omega_c^2}{a} \end{aligned} \tag{9.4}$$

Notice that the gains become infinite for $a = 0$ and $a = 1$ when the systems looses reachability.

An Observer

Next we will design an observer for the system. The observability matrix is

$$W_o = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

This matrix has full rank and there are thus no restrictions in designing the observer. The observer is given by

$$\frac{d\hat{x}}{dt} = Ax + Bu + K(y - Cx) = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} \hat{x} + \begin{bmatrix} a-1 \\ 1 \end{bmatrix} u + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} (y - \begin{bmatrix} 0 & 1 \end{bmatrix} \hat{x}) \tag{9.5}$$

It follows from Equations (9.1) and (9.5) that the observer error is given by

$$\frac{d\tilde{x}}{dt} = (A - KC)\tilde{x} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix} \tilde{x} - \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} \tilde{x} = \begin{bmatrix} -1 & -k_1 \\ 1 & -k_2 \end{bmatrix} \tilde{x}$$

This system has the characteristic polynomial

$$\det \begin{bmatrix} s+1 & k_1 \\ -1 & s+k_2 \end{bmatrix} = s^2 + (1+k_2)s + k_1 + k_2.$$

Requiring the observer should have the characteristic polynomial

$$s^2 + 2\zeta_o\omega_o s + \omega_o^2,$$

we find that

$$\begin{aligned} 1 + k_2 &= 2\zeta_o\omega_o \\ k_1 + k_2 &= \omega_o^2 \end{aligned}$$

Solving this equation gives

$$\begin{aligned} k_1 &= \omega_o^2 - 2\zeta_o\omega_o + 1 \\ k_2 &= 2\zeta_o\omega_o - 1 \end{aligned} \tag{9.6}$$

Output Feedback

A controller with output feedback can now be obtained by combining the state feedback given by Equation (9.3) with the observer given by Equation (9.5). The controller then becomes

$$\begin{aligned} \frac{d\hat{x}}{dt} &= A\hat{x} + Bu + K(y - C\hat{x}) = (A - BL - KC) + Ky \\ u &= -L\hat{x} \end{aligned} \tag{9.7}$$

where the parameters are given by (9.4) and (9.6). The controller has the transfer function

$$\begin{aligned} C(s) &= L[sI - A + BL + KC]^{-1}K \\ &= \begin{bmatrix} l_1 & l_2 \end{bmatrix} \begin{bmatrix} s + 1 + (a-1)l_1 & (a-1)l_2 + k_1 \\ l_1 - 1 & s + l_2 + k_2 \end{bmatrix}^{-1} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \\ &= \frac{(k_1 l_1 + k_2 l_2)s + (k_1 + k_2)l_2}{s^2 + (1 + k_2 + (a-1)l_1 + l_2)s + (k_1 + k_2)(1 - l_1) + a(k_1 l_1 + l_2)} \end{aligned} \tag{9.8}$$

The calculations are somewhat tedious but they can be done very comfortably using Matlab in a straight forward manner. Since

$$C(s) = L[sI - A + BL + KC]^{-1}K = K^T[SI - A^T + C^T K^T + L^T B^T]^{-1}L^T$$

it follows that the controller transfer function does not depend on which closed loop poles are associated with state feedback and which are associated with the observer.

The closed loop system is described by the equations

$$\frac{d}{dt}\begin{bmatrix} x \\ \hat{x} \end{bmatrix} = \begin{bmatrix} A & -BL \\ KC & A - BL - KC\hat{x} \end{bmatrix}\begin{bmatrix} x \\ \hat{x} \end{bmatrix} = A_{cl}\begin{bmatrix} x \\ \hat{x} \end{bmatrix}.$$

The properties of the closed loop system will now be illustrated by a numerical example.

A Numerical Example

To illustrate some properties of the system introduce $a = 1.25$, the poles of the state feedback are chosen as $-3 \pm 4i$ ($\omega_c = 5$, $\zeta_c = 0.6$) and the observer poles as $-6 \pm 8i$ ($\omega_o = 10$, $\zeta_o = 0.6$). The feedback gain is

$$L = \begin{bmatrix} -60 & 20 \end{bmatrix},$$

the observer gain is

$$L = \begin{bmatrix} 89 & 11 \end{bmatrix}^T,$$

and the closed loop system has the dynamics matrix

$$A_{cl} = \begin{bmatrix} -1 & 0 & 15 & -5 \\ 1 & 0 & 60 & -20 \\ 0 & 89 & 14 & -94 \\ 0 & 11 & 61 & -31 \end{bmatrix}.$$

The matrix is changed to

$$A_{cl} = \begin{bmatrix} -1 & 0 & 15.3 & -5.1 \\ 1 & 0 & 61.2 & -20.4 \\ 0 & 89 & 14 & -94 \\ 0 & 11 & 61 & -31 \end{bmatrix},$$

the process gain is increased by 2%. Notice that only the elements in the upper left block are changed by 2%. This matrix has the eigenvalues $0.19 \pm 6.73i$, -3.88 and -14.50. We thus have the strange situation that a design that looks very reasonable with very well damped closed loop poles at $-3 \pm 4i$ and $-6 \pm 8i$ results in a closed loop system that is extremely sensitive to parameter variations. The reason is not poor reachability. The reachability matrix is

$$W_r = \begin{bmatrix} 0.25 & -0.25 \\ 1 & 0.25 \end{bmatrix}$$

which is well conditioned.

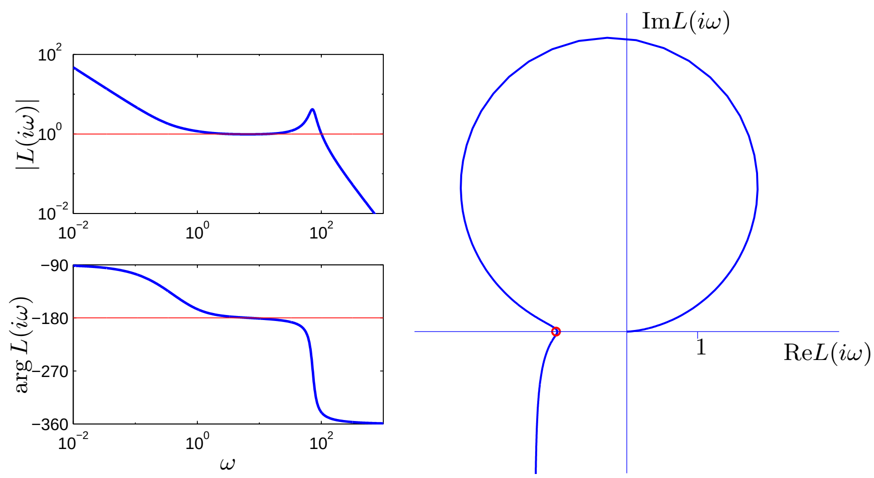

Figure 9.1: Bode (left) and Nyquist (right) plots of the loop transfer function (9.9).

Gain and Phase Margins

To get additional insight into the system we will calculate the gain and phase margins of the system. The controller transfer function for $a = 1.25$ is

$$C(s) = \frac{-5120s + 2000}{s^2 + 17s + 5300}.$$

Notice that the controller has a zero in the right half plane. The loop transfer function is

$$L(s) = P(s)C(s) = \frac{(s + 1.25)(-5120s + 2000)}{s(s + 1)(s^2 + 17s + 5300)} \tag{9.9}$$

The Bode and Nyquist plots of the loop transfer function is shown in Figure 9.1 The bode plot shows that the gain of the loop transfer functions is very close to one for frequencies in the range 1 to 20 rad/s. The gain crossover frequency is $\omega_{gc} = 3.14$ and the phase is very close to $-180°$ for frequencies around ω_{gc}. There is also a resonance peak at the frequency $\omega = \sqrt{5300} = 72.8$, which is caused by the poorly damped poles of the loop transfer function (9.9). The system has very poor stability margins, the gain margin is $g_m = 1.019$. This is the reason why the closed loop system becomes unstable when the loop gain is increased by 2%. The phase margin is $\varphi_m = 2.45°$. The poor stability margins are also clearly visible in the Nyquist plot which comes very close to the critical point. The circular bulb

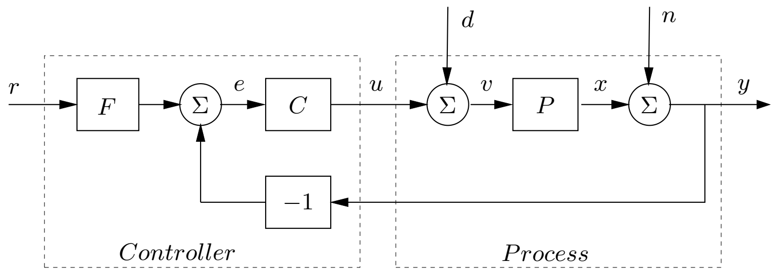

Figure 9.2: Block diagram of a basic feedback loop.

in the Nyquist plot is due to the poorly damped poles $s = -8.5 \pm 72.3i$ of the controller transfer function.

Summary

We thus have the seemingly contradictory situation that a controller design that seems quite reasonable gives a closed loop system with very poor robustness. One immediate conclusion is that it is not sufficient to require that a system is reachable and observable and that the robustness of a design based must always be checked a posteriori. It would however be useful to have more insight and in particular to understand what has to be done to obtain a closed loop system that is robust. This requires new concepts that will be developed in the next sections.

9.3 The Basic Feedback Loop

We will start by investigating some key properties of the feedback loop. A block diagram of a basic feedback loop is shown in Figure 9.2. The system loop is composed of two components, the process and the controller. The controller has two blocks the feedback block C and the feedforward block F. There are two disturbances acting on the process, the load $d$isturbance d and the $m$easurement noise n. The load disturbance represents disturbances that drive the process away from its desired behavior. The process variable x is the real physical variable that we want to control. Control is based on the measured signal y, where the measurements are corrupted by measurement noise n. Information about the process variable x is thus distorted by the measurement noise. The process is influenced by the controller via the control variable u. The process is thus a system with three inputs and

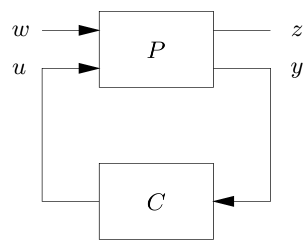

Figure 9.3: An abstract representation of the system in Figure 9.2. The input u represents the control signal and the input w represents the reference r, the load disturbance d and the measurement noise n. The output y is the measured variables and z are internal variables that are of interest.

one output. The inputs are: the control variable u, the load disturbance d and the measurement noise n. The output is the measured signal. The controller is a system with two inputs and one output. The inputs are the measured signal y and the reference signal r and the output is the control signal u. Note that the control signal u is an input to the process and the output of the controller and that the measured signal is the output of the process and an input to the controller. In Figure 9.2 the load disturbance was assumed to act on the process input. This is a simplification, in reality the disturbance can enter the process in many different ways. To avoid making the presentation unnecessarily complicated we will use the simple representation in Figure 9.2. This captures the essence and it can easily be modified if it is known precisely how disturbances enter the system.

More General Representation

The block diagrams themselves are substantial abstractions but higher abstractions are sometimes useful. The system in Figure 9.2 can be represented by only two blocks as shown in Figure 9.3. There are two types of inputs, the control u, which can be manipulated and the disturbances $w = (r,\ d,\ n)$, which represents external influences on the closed loop systems. The outputs are also of two types the measured signal y and other interesting signals $z = (e,\ v,\ x)$. The representation in Figure 9.3 allows many control variables and many measured variables, but it shows less of the system structure than Figure 9.2. This representation can be used even when there are many input signals and many output signals. It can also deal with the case when sensors and actuators have dynamics and when disturbances enter the system in other ways than shown in Figure 9.2. Representation with a higher level of abstraction are useful for the development of theory because they make it possible to focus on fundamentals and to solve general problems with a wide range of applications. Care must, however, be exercised to maintain the coupling to the real world control problems we intend to solve.

Disturbances

Attenuation of load disturbances is often a primary goal for control. This is particularly the case when controlling processes that run in steady state. Load disturbances are typically dominated by low frequencies. Consider for example the cruise control system for a car, where the disturbances are the gravity forces caused by changes of the slope of the road. These disturbances vary slowly because the slope changes slowly when you drive along a road. Step signals or ramp signals are commonly used as prototypes for load disturbances disturbances.

Measurement noise corrupts the information about the process variable that the sensors delivers. Measurement noise typically has high frequencies. The average value of the noise is typically zero. If this was not the case the sensor will give misleading information about the process and it would not be possible to control it well. There may also be dynamics in the sensor. Several sensors are often used. A common situation is that very accurate values may be obtained with sensors with slow dynamics and that rapid but less accurate information can be obtained from other sensors. The sensors may also have dynamics. Also notice that measurement noise only influences the process indirectly via the feedback.

Actuation

The process is influenced by actuators which typically are valves, motors, that are driven electrically, pneumatically, or hydraulically. There are often local feedback loops and the control signals can also be the reference variables for these loops. A typical case is a flow loop where a valve is controlled by measuring the flow. If the feedback loop for controlling the flow is fast we can consider the set point of this loop which is the flow as the control variable. In such cases the use of local feedback loops can thus simplify the system significantly. When the dynamics of the actuators is significant it is convenient to lump them with the dynamics of the process. There are cases where the dynamics of the actuator dominates process dynamics.

Design Issues

Many issues have to be considered in analysis and design of control systems. The basic requirements are

- Stability
- Ability to follow reference signals (performance)
- Reduction of effects of load disturbances (performance)
- Reduction of effects of measurement noise (performance)
- Reduction of effects of model uncertainties (robustness)

Instability is the major drawback of feedback. Avoiding instability is thus a primary goal. It is also desirable that the process variable follows the reference signal faithfully. The system should also be able to reduce the effect of load disturbances. Measurement noise is injected into the system by the feedback. This is unavoidable but it is essential that not too much noise is injected. It must also be considered that the models used to design the control systems are inaccurate. The properties of the process may also change. The control system should be able to cope with moderate changes, which is not a trivial task as is illustrated by the example in Section **??**. The relative importance of the different abilities vary with the application. In process control the major emphasis is often on attenuation of load disturbances, while the ability to follow reference signals is the primary concern in motion control systems. In other cases robustness may be the main requirement.

9.4 The Gangs of Four and Six

The feedback loop in Figure 9.2 is influenced by three external signals, the reference r, 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 R be the Laplace transforms of x, y, u, d, n r, respectively. The following relations are obtained from the block diagram in Figure 9.2

$$\begin{aligned} X &= \frac{P}{1+PC}D - \frac{PC}{1+PC}N + \frac{PCF}{1+PC}R \\ Y &= \frac{P}{1+PC}D + \frac{1}{1+PC}N + \frac{PCF}{1+PC}R \\ U &= -\frac{PC}{1+PC}D - \frac{C}{1+PC}N + \frac{CF}{1+PC}R. \end{aligned} \tag{9.10}$$

To simplify notations we have dropped the arguments of all Laplace transforms. 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 set of six transfer functions which we call the Gang of Six.

$$\begin{matrix} \dfrac{PCF}{1+PC} & \dfrac{PC}{1+PC} & \dfrac{P}{1+PC} \\ \dfrac{CF}{1+PC} & \dfrac{C}{1+PC} & \dfrac{1}{1+PC}, \end{matrix} \tag{9.11}$$

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 the transfer function $C/(1 + PC)$ gives the response of the control signal to measurement noise.

Notice that only four transfer functions are required to describe how the system reacts to load disturbance and the measurement noise and that 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 four transfer functions, namely the four rightmost transfer functions in (9.11), i.e.

$$\begin{aligned} &\frac{PC}{1+PC}, && \text{the complementary sensitivity function} \\ &\frac{P}{1+PC}, && \text{the load disturbance sensitivity function} \\ &\frac{C}{1+PC}, && \text{the noise sensitivity function} \\ &\frac{1}{1+PC}, && \text{the sensitivity function} \end{aligned} \tag{9.12}$$

These transfer functions and their equivalent systems are called the Gang of Four. The transfer functions have many interesting properties that will be discussed in the following. A good insight into these properties are essential for understanding feedback systems. The load disturbance sensitivity function is sometimes called the input sensitivity function and the noise sensitivity function is sometimes called the output sensitivity function.

Systems with Two Degrees of Freedom

The controller in Figure 9.2 is said to have two degrees of freedom because the controller has two blocks, the feedback block C which is part of the closed loop and the feedforward block F which is outside the loop. Using such a controller gives a very nice separation of the control problem because the feedback controller can be designed to deal with disturbances and process uncertainties and the feedforward will handle the response to reference signals. Design of the feedback only considers the gang of four and the feedforward deals with the two remaining transfer functions in the gang of six. For a system with error feedback it is necessary to make a compromise. The controller C thus has to deal with all aspects of the problem.

To describe the system properly it is thus necessary to show the response of all six transfer functions. The transfer functions can be represented in different ways, by their step responses and frequency responses, see Figures 9.4 and 9.5. Figures 9.4 and 9.5 give useful insight into the properties of the closed loop system. The time responses in Figure 9.4 show that the feedforward gives a substantial improvement of the response speed. The settling time is substantially shorter, 4 s versus 25 s, and there is no overshoot. This is also reflected in the frequency responses in Figure 9.5 which shows that the transfer function with feedforward has higher bandwidth and that it has no resonance peak.

The transfer functions $CF/(1 + PC)$ and $-C/(1 + PC)$ represent the signal transmission from reference to control and from measurement noise to control. The time responses in Figure 9.4 show that the reduction in response time by feedforward requires a substantial control effort. The initial value of the control signal is out of scale in Figure 9.4 but the frequency response in 9.5 shows that the high frequency gain of $PCF/(1 + PC)$ is 16, which can be compared with the value 0.78 for the transfer function $C/(1 + PC)$. The fast response thus requires significantly larger control signals.

There are many other interesting conclusions that can be drawn from Figures 9.4 and 9.5. Consider for example the response of the output to load

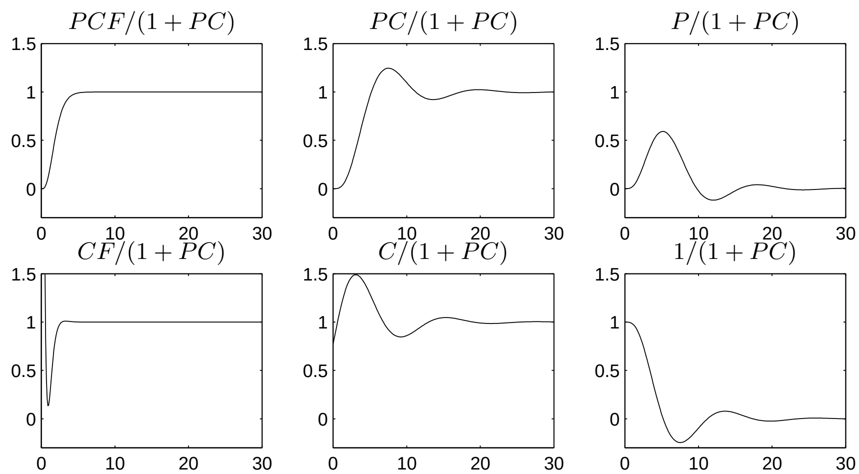

Figure 9.4: Step responses of the Gang of Six for PI control $k = 0.775$, $T_i = 2.05$ of the process $P(s) = (s + 1)^{-4}$. The feedforward is designed to give the transfer function $(0.5s + 1)^{-4}$ from reference r to output y.

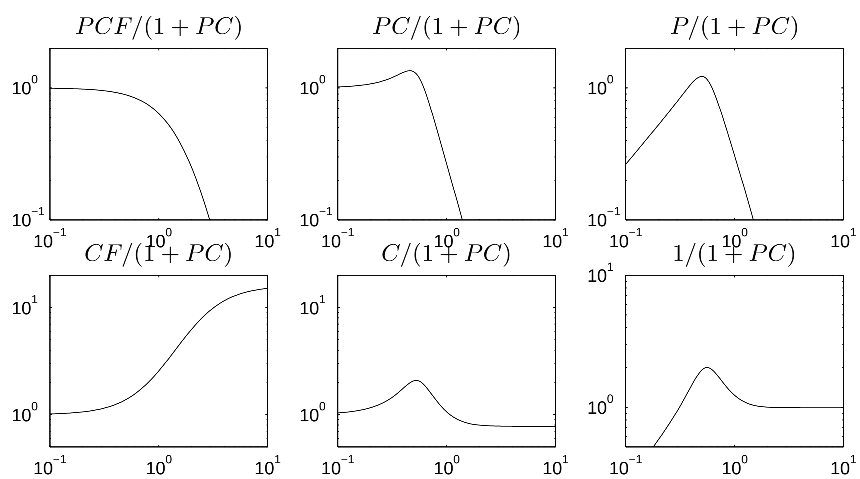

Figure 9.5: Gain curves of frequency responses of the Gang of Six for PI control $k = 0.775$, $T_i = 2.05$ of the process $P(s) = (s + 1)^{-4}$ where the feedforward has been designed to give the transfer function $(0.5s + 1)^{-4}$ from reference to output.

disturbances expressed by the transfer function $P/(1+PC)$. The frequency response has a pronounced peak 1.22 at $\omega_{max} = 0.5$ the corresponding time function has its maximum 0.59 at $t_{max} = 5.2$. Notice that the peaks are of the same magnitude and that the product of $\omega_{max} t_{max} = 2.6$. Similar relations hold for the other responses.

A Practical Consequence

The fact that 6 relations are required to capture properties of the basic feedback loop is often neglected in literature. 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 and can be strongly misleading. To get a more complete representation of the system all six responses should be given. Specifications on a control system should also reflect this. We illustrate the importance of this by an example.

Example 9.1 *(Assessment of a Control System).* Consider a process with the transfer function

$$P(s) = \frac{1}{(s+1)(s+0.02)}$$

with a PI controller using error feedback with a controller having the transfer function

$$C(s) = \frac{50s+1}{50s} = 1 + \frac{0.02}{s}.$$

The loop transfer function is

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

Notice that the process pole at $s = -0.02$ is canceled by a controller zero and that the factor $s + 0.02$ does not appear in the loop transfer function. Figure 9.6 shows that the step responses to a reference signal look very reasonable. Based on these responses we could be tempted to conclude that the closed loop system is well designed. The step response settles in about 10 s and the overshoot is moderate. In the right part of the figure we show the responses of y and u to a step in the load disturbance. Notice that the error decays very slowly.

To explore the system further we will calculate the transfer functions of

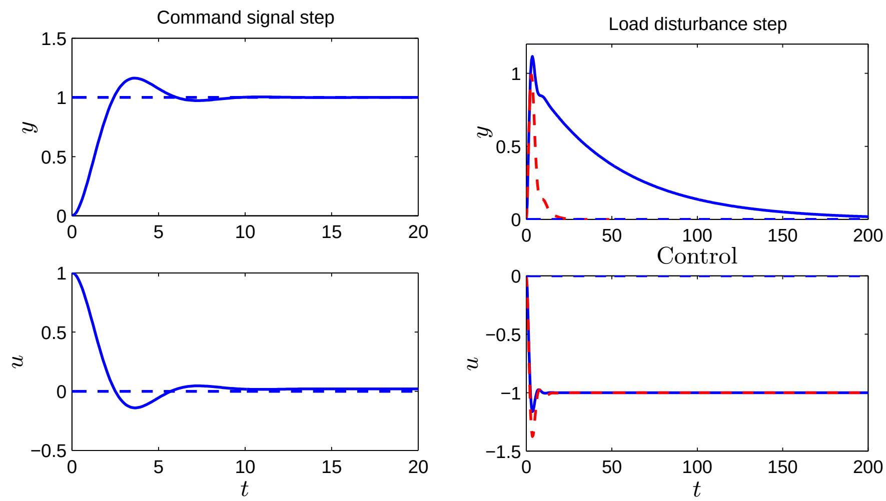

Figure 9.6: Response of output y and control u to a step in reference r (left) and to a step in the load disturbance (right). The dashed curves show the response with a controller having the transfer function $C(s) = 1 + 0.2/s$.

the Gang of Four. We have

$$\frac{PC}{1+PC} = \frac{1}{s^2+s+1} \qquad \frac{P}{1+PC} = \frac{s}{(s+0.02)(s^2+s+1)}$$
$$\frac{C}{1+PC} = \frac{(s+0.02)(s+1)}{s^2+s+1} \qquad \frac{1}{1+PC} = \frac{s(s+1)}{s^2+s+1}$$

The responses to the reference signal are given by the transfer function

$$G_{yr}(s) = \frac{1}{s^2+s+1}, \quad G_{ur}(s) = \frac{(s+1)(s+0.02)}{s^2+s+1}$$

and the responses to the load disturbance are given by

$$G_{yd}(s) = \frac{s}{(s+0.02)(s^2+s+1)}, \quad G_{ud}(s) = -\frac{1}{s^2+s+1}$$

Notice that the process pole $s = 0.02$ is canceled by a controller zero. This implies that the loop transfer function is of second order even if the closed loop system itself is of third order. The characteristic equation of the closed loop system is

$$(s+0.02)(s^2+s+1) = 0$$

where the the pole $s = -0.02$ corresponds the process pole that is canceled by the controller zero. The presence of the slow pole $s = -0.02$ which appears in the response to load disturbances implies that the output decays very slowly, at the rate of $e^{-0.02t}$. The controller will not respond to the signal $e^{-0.02t}$. Since the controller has a zero at $s = -0.02$ the transmission of the signal is blocked by the controller. This is clearly seen in Figure 9.19, which shows the response of the output and the control signals to a step change in the load disturbance. Notice that it takes about 200 s for the disturbance to settle. This can be compared with the step response in Figure 9.6 which settles in about 10s.

Having understood what happens it is straight forward to modify the controller. With the controller

$$C(s) = 1 + \frac{0.2}{s}$$

the response to a step in the load disturbance is as shown in the dashed curves in Figure 9.6. Notice that drastic improvements in the response to load disturbance are obtained with only moderate changes in the control signal. This is a nice illustration of the importance of timing to achieve good control.

The behavior illustrated in the example is typical when slow process poles are canceled. The canceled factors do not appear in the loop transfer function and the sensitivity functions S and T, they do however appear in the transfer function $P/(1 + PC)$. The canceled modes are not visible unless they are excited. The effects are even more drastic than shown in the example if the canceled modes are unstable. This has been known among control engineers for a long time and a good design rule that *cancellation of slow or unstable process poles by zeros in the controller give very poor attenuation of load disturbances.*

9.5 Disturbance Attenuation

The attenuation of disturbances will now be discussed. For that purpose we will compare an open loop system and a closed loop system subject to the same disturbances as is illustrated in Figure 9.7. Let the transfer function of the process be $P(s)$. The output of the open loop system is

$$y_{ol}(t) = x(t) + n(t).$$

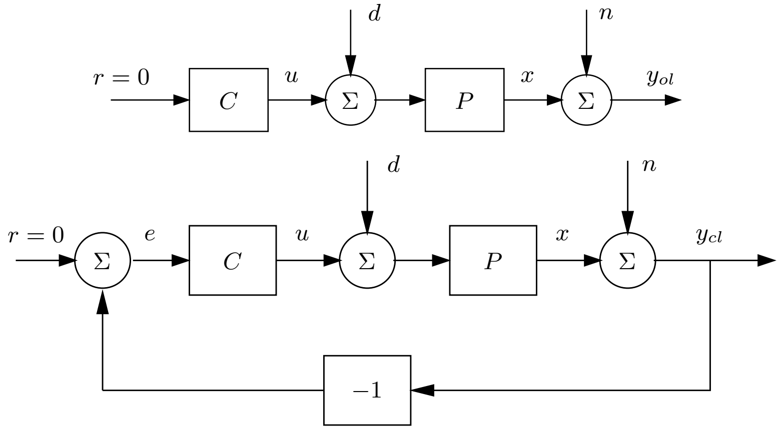

Figure 9.7: Open and closed loop systems subject to the same disturbances.

Tracing signals around the loop we find for exponential signals that the output is

$$y_{cl}(t) = \frac{1}{1+P(s)C(s)} y_{ol}(t) = S(s)y_{ol}(t)$$

where $S(s)$ is the sensitivity function, which belongs to the Gang of Four. We thus obtain the following interesting result: The output of a system with feedback can be obtained by sending the output of the open loop system through a dynamical system with the transfer function $S(s)$. The sensitivity function thus shows the effect of feedback. Disturbances are attenuated if their frequencies are such $|s(i\omega) < 1$, they are amplified if their frequencies are such that $|s(i\omega) > 1$. The lowest frequency where the sensitivity function has the magnitude 1 is called the sensitivity crossover frequency and denoted by ω_{sc}. The maximum sensitivity

$$M_s = \max_\omega |S(i\omega)| = \max_\omega \Big| \frac{1}{1+P(i\omega)C(i\omega)} \Big| \tag{9.13}$$

is an important variable which gives the largest amplification of the disturbances. The maximum occurs at the frequency ω_{ms}.

A quick overview of how disturbances are influenced by feedback is obtained from the gain curve of the Bode plot of the sensitivity function. An example is given in Figure 9.8. The figure shows that the sensitivity crossover frequency is 0.32 and that the maximum sensitivity 2.1 occurs at $\omega_{ms} = 0.56$. Feedback will thus reduce disturbances with frequencies less

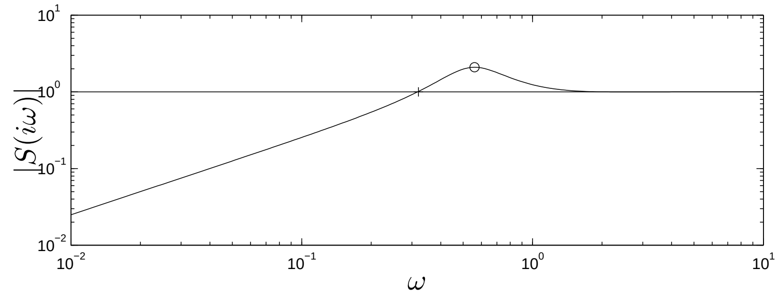

Figure 9.8: Gain curve of the sensitivity function for PI control ($k = 0.8$, $k_i = 0.4$) of process with the transfer function $P(s) = (s+1)^{-4}$. The sensitivity crossover frequency is indicated by $+$ and the maximum sensitivity by o.

than 0.32 rad/s, but it will amplify disturbances with higher frequencies. The largest amplification is 2.1.

If a record of the disturbance is available and a controller has been designed the output obtained under closed loop with the same disturbance can be visualized by sending the recorded output through a filter with the transfer function $S(s)$.

The sensitivity function can be written as

$$S(s) = \frac{1}{1 + P(s)C(s)} = \frac{1}{1 + L(s)}. \tag{9.14}$$

Since it only depends on the loop transfer function it can also be visualized graphically in the Nyquist plot of the loop transfer function. This is illustrated in Figure 9.9. The complex number $1 + L(i\omega)$ can be represented as the vector from the point -1 to the point $L(i\omega)$ on the Nyquist curve. The sensitivity is thus less than one for all points outside a circle with radius 1 and center at -1. Disturbances of these frequencies are attenuated by the feedback. If a control system has been designed based on a given model it is straight forward to estimated the potential disturbance reduction simply by recording a typical output and filtering it through the sensitivity function.

Random Disturbances

Process disturbances can often be described as stationary stochastic processes which can be characterized by their power spectral density $\phi(\omega)$.

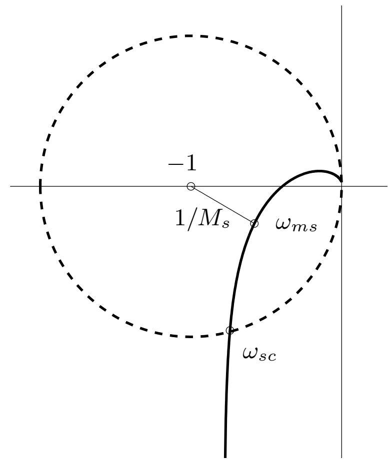

Figure 9.9: Nyquist curve of loop transfer function showing graphical interpretation of maximum sensitivity. The sensitivity crossover frequency ω_{sc} and the frequency ω_{ms} where the sensitivity has its largest value are indicated in the figure. All points inside the dashed circle have sensitivities greater than 1.

This physical meaning is that the energy of the signal in the frequency band $\omega_1 \le \omega \le \omega_2$ is $2\int_{\omega_1}^{\omega_2} \phi(\omega)d\omega$. If $\phi_{ol}(\omega)$ is the power spectral density of the output in open loop the power spectral density of the closed loop system is

$$\phi_{cl}(\omega) = |S(i\omega)|^2 \phi_{ol}(\omega)$$

and the ratio of the variances of the closed and open loop systems is

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

9.6 Robustness to Process Variations

Control systems are designed based on simplified models of the processes. Process dynamics will often change during operation. The sensitivity of a closed loop system to variations in process dynamics is therefore a fundamental issue.

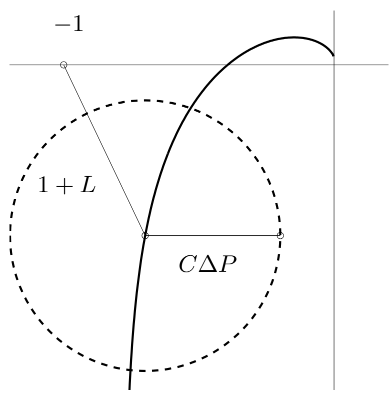

Figure 9.10: Nyquist curve of a nominal loop transfer function and its uncertainty caused by process variations ΔP.

Risk for Instability

Instability is the main drawback of feedback. It is therefore of interest to investigate if process variations can cause instability. The sensitivity functions give a useful insight. Figure 9.9 shows that the largest sensitivity is the inverse of the shortest distance from the point -1 to the Nyquist curve.

The complementary sensitivity function also gives insight into allowable process variations. Consider a feedback system with a process P and a controller C. We will investigate how much the process can be perturbed without causing instability. The Nyquist curve of the loop transfer function is shown in Figure 9.10. If the process is changed from P to $P + \Delta P$ the loop transfer function changes from PC to $PC + C\Delta P$ as illustrated in the figure. The distance from the critical point -1 to the point L is $|1+L|$. This means that the perturbed Nyquist curve will not reach the critical point -1 provided that

$$|C\Delta P| < |1 + L|$$

which implies

$$|\Delta P| < \left|\frac{1 + PC(i}{C}\right| \tag{9.15}$$

This condition must be valid for all points on the Nyquist curve, i.e pointwise for all frequencies. The condition for stability can be written as

$$\left|\frac{\Delta P(i\omega)}{P(i\omega)}\right| < \frac{1}{|T(i\omega)|} \tag{9.16}$$

A technical condition, namely that the perturbation ΔP is a stable transfer function, must also be required. If this does not hold the encirclement condition required by Nyquist's stability condition is not satisfied. Also notice that the condition (9.16) is conservative because it follows from Figure 9.10 that the critical perturbation is in the direction towards the critical point -1. Larger perturbations can be permitted in the other directions.

This formula (9.16) 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 (9.16) implies that the closed loop system will at least be stable for substantial variations in the process dynamics.

It follows from (9.16) that the variations can be large for those frequencies where T is small and that smaller variations are allowed for frequencies where T is large. A conservative estimate of permissible process variations that will not cause instability is given by

$$\left|\frac{\Delta P(i\omega)}{P(i\omega)}\right| < \frac{1}{M_t}$$

where M_t is the largest value of the complementary sensitivity

$$M_t = \max_{\omega} |T(i\omega)| = \max_{\omega}\left|\frac{P(i\omega)C(i\omega)}{1 + P(i\omega)C(i\omega)}\right| \tag{9.17}$$

The value of M_t is influenced by the design of the controller. For example if $M_t = 2$ pure gain variations of 50% or pure phase variations of 30° are permitted without making the closed loop system unstable. The fact that the closed loop system is robust to process variations is one of the reason why control has been so successful and that control systems for complex processes can indeed be designed using simple models. This is illustrated by an example.

Example 9.2 *(Model Uncertainty).* Consider a process with the transfer function

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

A PI controller with the parameters $k = 0.775$ and $T_i = 2.05$ gives a closed loop system with $M_s = 2.00$ and $M_t = 1.35$. The complementary sensitivity has its maximum for $\omega_{mt} = 0.46$. Figure 9.11 shows the Nyquist curve of the transfer function of the process and the uncertainty bounds $\Delta P = |P|/|T|$ for a few frequencies. The figure shows that

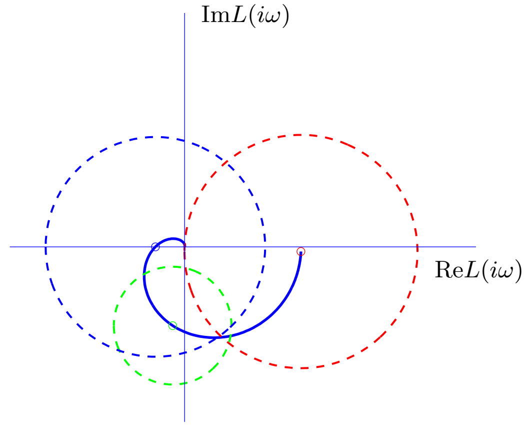

Figure 9.11: Illustration of the robustness of a feedback system. The Nyquist curve the process transfer function $P(s) = (s+1)^{-4}$ is shown in full lines. The circles show permissible perturbations uncertainty regions $|\Delta P| = 1/|T|$ for a PI control with $k = 0.8$ and $T_i = 0.4$ for the frequencies $\omega = 0$, 0.46 and 1.

- Large uncertainties are permitted for low frequencies, $T(0) = 1$.
- The smallest relative error $|\Delta P/P|$ occurs for $\omega = 0.46$.
- For $\omega = 1$ we have $|T(i\omega)| = 0.26$ which means that the stability requirement is $|\Delta P/P| < 3.8$
- For $\omega = 2$ we have $|T(i\omega)| = 0.032$ which means that the stability requirement is $|\Delta P/P| < 31$

The situation illustrated in the figure is typical for many processes, moderately small uncertainties are only required around the gain crossover frequencies, but large uncertainties can be permitted at higher and lower frequencies. A consequence of this is also that a simple model that describes the process dynamics well around the crossover frequency is sufficient for design. Systems with many resonance peaks are an exception to this rule because the process transfer function for such systems may have large gains also for higher frequencies.

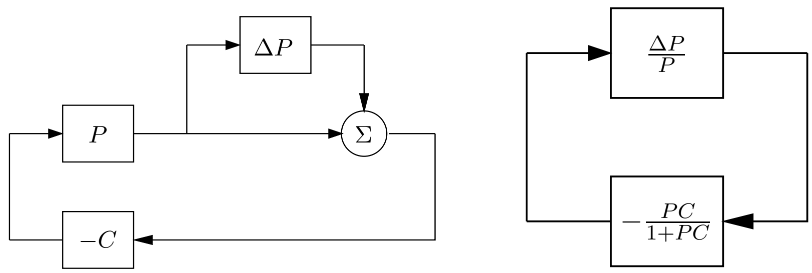

Figure 9.12: Illustration of robustness to process perturbations.

Small Gain Theorem

The robustness result given by Equation (9.16) can be given another interpretation. This is illustrated in Figure 9.12 which shows a block diagram of the closed loop system with the perturbed process in A. Another representation of the system is given in B. This representation is obtatined by combining two of the blocks. The loop transfer function of the system in Figure 9.12B is

$$L(s) = \frac{PC}{1+PC}\Delta P$$

Equation 9.16 thus simply implies that the largest loop gain is less than one. Since both blocks are stable it follows from Nyquists stability theorem that the closed loop is stable.

Variations in Closed Loop Transfer Function

So far we have investigated the risk for instability. The effects of small variation in process dynamics on the closed loop transfer function will now be investigated. To do this we will analyze the system in Figure 9.2. The transfer function from reference to output is given by

$$G_{yr} = \frac{PCF}{1+PC} = T \tag{9.18}$$

Compare with (9.11). The transfer function T which belongs to the Gang of Four is called the complementary sensitivity function. Differentiating (9.18) we get

$$\frac{dG_{yr}}{dP} = \frac{CF}{1+PC} - \frac{PCFC}{(1+PC)^2} = \frac{CF}{(1+PC)^2} = S\frac{G_{yr}}{P}$$

Hence

$$\frac{d\log G_{yr}}{d\log P} = \frac{dG_{yr}/G_{yr}}{dP/P} = S \tag{9.19}$$

The relative error in the closed loop transfer function thus equals the product of the sensitivity function and the relative error in the process. This equation is the reason for calling S the sensitivity function. The relative error in the closed loop transfer function is thus small when the sensitivity is small. This is one of the very useful properties of feedback. For example this property was exploited by Black at Bell labs to build the feedback amplifiers that made it possible to use telephones over large distances.

A small value of the sensitivity function thus means that disturbances are attenuated and that the effect of process perturbations also are negligible. A plot of the magnitude of the complementary sensitivity function as in Figure 9.8 is a good way to determine the frequencies where model precision is essential.

9.7 The Sensitivity Functions

We have seen that the sensitivity function S and the complementary sensitivity function T tell much about the feedback loop. We have also seen from Equations (9.5) and (9.19) that it is advantageous to have a small value of the sensitivity function and it follows from (9.16) that a small value of the complementary sensitivity allows large process uncertainty. Since

$$S(s) = \frac{1}{1 + P(s)C(s)} \text{ and } T(s) = \frac{P(s)C(s)}{1 + P(s)C(s)} \tag{9.20}$$

it follows that

$$S(s) + T(s) = 1 \tag{9.21}$$

This means that S and T cannot be made small simultaneously. The loop transfer function L is typically large for small values of s and it goes to zero as s goes to infinity. This means that S is typically small for small s and close to 1 for large. The complementary sensitivity function is close to 1 for small s and it goes to 0 as s goes to infinity.

A basic problem is to investigate if S can be made small over a large frequency range. We will start by investigating an example.

Example 9.3 *(System that Admits Small Sensitivities).* Consider a closed loop system consisting of a first order process and a proportional controller.

Let the loop transfer function

$$L(s) = P(s)C(s) = \frac{k}{s+1}$$

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

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

and we have

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

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

The system in Example 9.3 is unfortunately an exception. The key feature of the system is that the Nyquist curve of the process lies in the fourth quadrant. Systems whose Nyquist curves are in the first and fourth quadrant are called positive real. For such systems the Nyquist curve never enters the region shown in Figure 9.9 where the sensitivity is greater than one.

For typical control systems there are unfortunately severe constraints on the sensitivity function. Bode has shown that if the loop transfer has poles p_k in the right half plane and if it goes to zero faster than $1/s$ for large s the sensitivity function satisfies the following integral

$$\int_0^\infty \log |S(i\omega)| d\omega = \int_0^\infty \log \frac{1}{|1+L(i\omega)|} d\omega = \pi \sum \text{Re } p_k \qquad (9.22)$$

This equation shows that if the sensitivity function is made smaller for some frequencies it must increase at other frequencies. This means that if disturbance attenuation is improved in one frequency range it will be worse in other. This is called the *w*ater bed effect.

Equation (9.22) implies that there are fundamental limitations to what can be achieved by control and that control design can be viewed as a redistribution of disturbance attenuation over different frequencies.

For a loop transfer function without poles in the right half plane (9.22) reduces to

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

This formula can be given a nice geometric interpretation as shown in Figure 9.13 which shows $\log |S(i\omega)|$ as a function of ω. The area over the horizontal axis must be equal to the area under the axis.

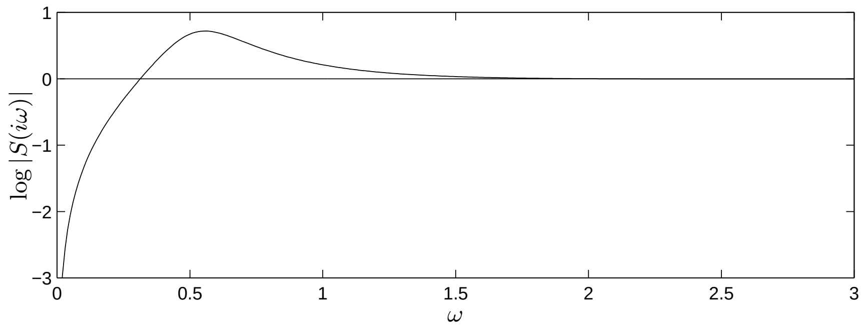

Figure 9.13: Geometric interpretation of Bode's integral formula (9.22).

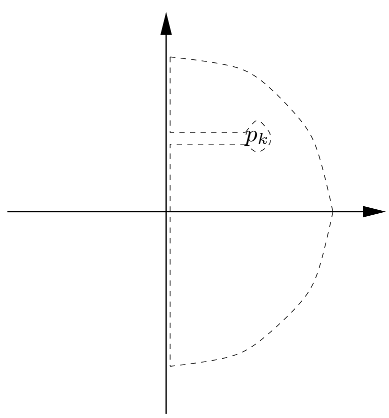

Figure 9.14: Contour used to prove Bode's theorem.

Derivation of Bode's Formula

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

Consider the integral of the logarithm of the sensitivity function $S(s) = 1/(1+L(s))$ over the contour shown in Figure 9.14. The contour encloses the right half plane except the points $s = p_k$ where the loop transfer function $L(s) = P(s)C(s)$ has poles and the sensitivity function $S(s)$ has zeros. The direction of the contour is counter clockwise.

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

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

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

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

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

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

Since $L(s)$ goes to zero faster than $1/s$ for large s the integral goes to zero when the radius of the circle goes to infinity. Next we consider the integral I_3, for this purpose we split the contour into three parts X_+, γ and X_- as indicated in Figure 9.14. We have

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

The contour γ is a small circle with radius r around the pole p_k. The magnitude of the integrand is of the order $\log r$ and the length of the path is $2\pi r$. The integral thus goes to zero as the radius r goes to zero. Furthermore we have

$$
\begin{aligned}
&\int_{X_+} \log S(s)ds + \int_{X_-} \log S(s)ds \\
&\qquad = \int_{X_+} \big(\log S(s) - \log S(s - 2\pi i)\big)ds = 2\pi p_k
\end{aligned}
$$

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

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

which is Bode's formula (9.22).

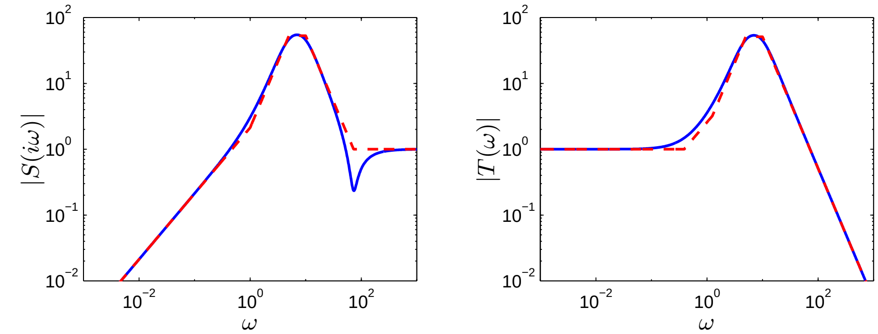

Figure 9.15: Gain curves of the sensitivity function (left) and the complementary sensitivity function (right). The straight line approximations are shown with dashed lines.

9.8 Implications for Design

We will now return to the system discussed in Section 9.2. The process and controller transfer functions are

$$P(s) = \frac{s+1.25}{s(s+1)}, \quad C(s) = \frac{-5120s+2000}{s^2+17s+5300}$$

The sensitivity functions are

$$S(s) = \frac{s(s+1.25)(s^2+17s+5300}{(s^2+6s+25)(s^2+12s+100)}$$
$$T(s) = \frac{(s+1)(-5120s+2000)}{(s^2+6s+25)(s^2+12s+100)}$$

The gain curves for the sensitivity functions are shown in Figure 9.15. The figures show that the maximum sensitivities are very large, $M_s = 55$ and $M_t = 54$.

The sensitivity curves also give insights. Consider the break points which are clearly seen from the straight line approximation. The large peak in the complementary sensitivity is caused by the controller zero at $s = -5300/2000$ and process zero at $s = -1.25$ the gain increases and it does not decay until the dominant process poles at $\omega = 5$ and $\omega = 10$. The peak can clearly be avoided by requiring that the closed loop system has a pole close to the process zero. Keeping the same state feedback and choosing observer poles at $s = -1.25$ and $s = -10$ instead of $s = -6 \pm 10i$ gives the

The loop transfer function is

$$L(s) = \frac{(s+1.25)(70s+250)}{s(s+1)(s^2+16.25s+18.74)}$$

Figure 9.16 shows the Bode and Nyquist plots of the loop transfer function. The gain and phase margins are $g_m = \infty$, $\varphi_m = 47.3$ and the maximum sensitivities are $M_s = 1.39$ and $M_t = 1.36$. The system thus has very good robustness properties.

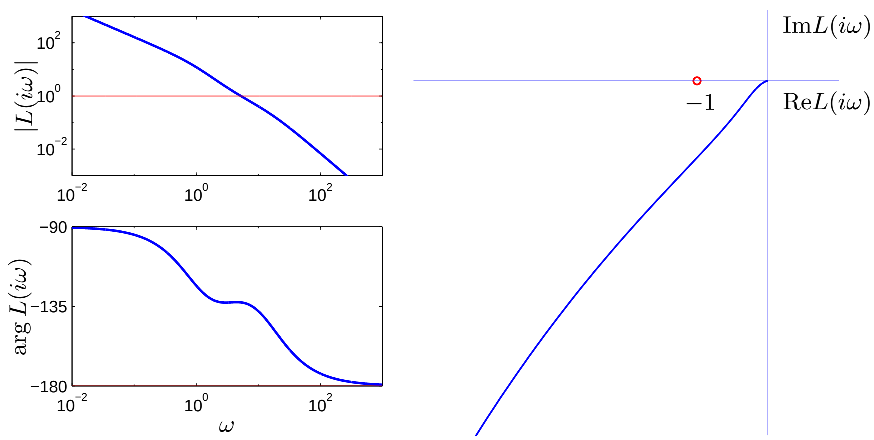

Figure 9.16: Bode (left) and Nyquist (right) plots of the loop transfer function (**??**).

observer gain the observer gain is

$$K = \begin{bmatrix}2.25 & 10.25\end{bmatrix}^T,$$

The controller transfer function is

$$C(s) = \frac{70s+250}{s^2+16.25s+18.74}$$

A comparison with Figure 9.16 shows that changing the observer poles from $-6\pm8i$ to -1.25 and -10 has a drastic impact of the robustness of the closed loop system. The important thing is to introduce a closed loop pole close to the slow open loop zero at $s = -1.25$. Another conclusion is that the open loop dynamics must be taken into account when choosing the closed loop poles. We will now use the ideas of the analysis to arrive at general design rules.

Design Rules

Let the transfer functions of the process and the controller be

$$P(s) = \frac{n_p(s)}{d_p(s)}$$
$$C(s) = \frac{n_c(s)}{d_c(s)}$$

where $n_p(s)$, $n_c(s)$, $d_p(s)$ and $d_c(s)$ are polynomials. The sensitivity functions then becomes

$$S(s) = \frac{d_p(s)d_c(s)}{d_p(s)d_c(s) + n_p(s)d_p(s)}$$
$$T(s) = \frac{n_p(s)n_c(s)}{d_p(s)d_c(s) + n_p(s)d_p(s)}$$

Let w_{gc} be the desired gain crossover frequency. Assume that the process has zeros which are slower than ω_{gc}. The complementary sensitivity function is one for low frequencies and it start to increase for frequencies close to the process zeros unless there is a closed loop pole in the neighborhood. To avoid large values of the complementary sensitivity function we find that the closed loop system should have poles close to the slow zeros.

Now consider process poles that are faster than the desired gain crossover frequency. The sensitivity function is one for high frequencies. Moving from high to low frequencies the sensitivity function increases at the fast process poles. Large peaks can be obtained unless there are process poles close to the closed loop poles. To avoid large peaks in the sensitivity the closed loop system should be have poles close that matches the fast process poles. We thus obtain the simple rules that $s$low process zeros should be matched slow closed loop poles and fast process poles should be matched by fast process poles. The rule are illustrated with an additional example.

Example 9.4 *(To Cancel or not to Cancel).* Consider PI control of a first order system where the process and the controller have the transfer functions

$$P(s) = \frac{b}{s+a}$$
$$C(s) = k + \frac{k_i}{s}$$

The loop transfer function is

$$L(s) = \frac{b(ks + k_i)}{s(s+a)}$$

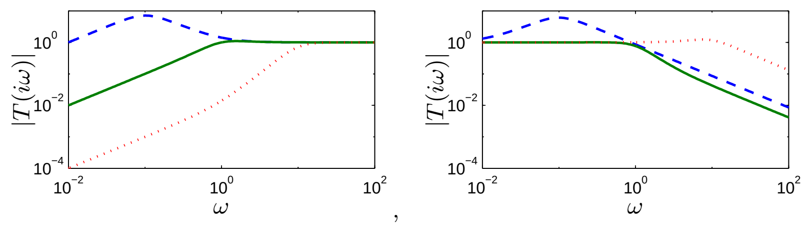

Figure 9.17: Magnitude curve for Bode plots of the sensitivity function S (above) and the complementary sensitivity function T (below) for $\zeta = 0.7$, $a = 1$ and $\omega_0/a = 0.1$ (dashed), 1 (solid) and 10 (dotted).

The closed loop characteristic polynomial is

$$s(s+a) + b(ks + k_i) = s62 + (a + bk)s + k_i$$

Let the desired closed loop characteristic polynomial be

$$s^2 + 2\zeta\omega_0 s + \omega_0^2 \tag{9.23}$$

Matching this with we find that the controller gains are

$$s^2 + 2\zeta\omega_0 s + \omega_0^2 \tag{9.24}$$

where $\zeta \leq 1$, we find that the controller parameters are given by

$$k = \frac{2\zeta\omega_0 - a}{b}$$
$$k_i = \frac{\omega_0^2}{b}$$

Notice that the controller has a zero in the right half plane if $2\zeta\omega_0 < a$, an indication that the system has bad properties. The sensitivity functions we get

$$S(s) = \frac{s(s+a)}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$
$$T(s) = \frac{(2\zeta\omega_0 - a)s + \omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$

Figure 9.17 shows clearly that the sensitivities are small for designs with $\omega_0 > a$, but high for designs with $\omega_0 > a$. If we desire a system with slower response the design rule above tells that the closed loop system should have a pole matching the fast process pole. This can be achieved by choosing the closed loop characteristic polynomial as

$$d_{cl} = (s + a)(s + \omega_0)$$

The controller gains then becomes

$$k = \frac{\omega_0}{b}$$
$$k_i = \frac{a\omega_0}{l}$$

and the loop transfer function becomes

$$L(s) = \frac{bk}{s}$$

Notice that the fast process pole is canceled. The sensitivity functions are

$$S(s) = \frac{s}{s + bk}$$
$$T(s) = \frac{bk}{s + bk}$$

The maximum sensitivities are less than one for all frequencies. Notice that this design is not sensible if $a < \omega_0$ because the controller then cancels a slow pole and the response to load disturbances will be poor as illustrated in Example 9.1.

9.9 When are Two Processes Similar?

A fundamental issue is to determine when two processes are close. This seemingly innocent problem is not as simple as it may appear. When discussing the effects of uncertainty of the process on stability in Section 9.6 we used the quantity

$$\delta(P_1, P_2) = \max_{\omega} |P_1(i\omega) - P_2(i\omega)| \tag{9.25}$$

as a measure of closeness of two processes. In addition the transfer functions P_1 and P_2 were assumed to be stable. This means conceptually that we compare the outputs of two systems subject to the same input. This may appear as a natural way to compare two systems but there are complications. Two systems that have similar open loop behaviors may have drastically different behavior in closed loop and systems with very different open loop behavior may have similar closed loop behavior. We illustrate this by two examples.

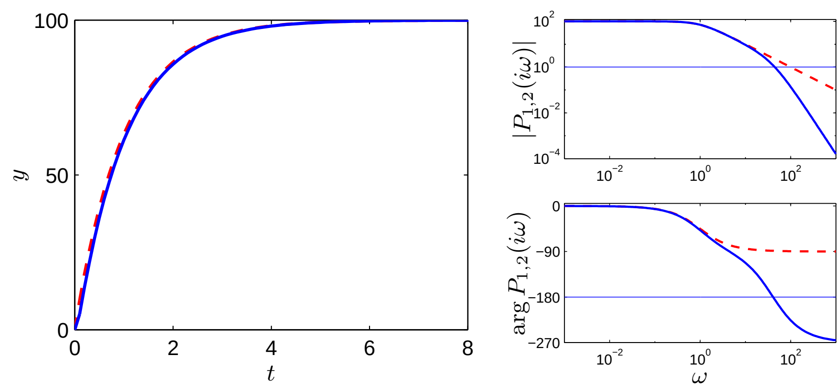

Figure 9.18: Step responses for systems with the transfer functions $P_1(s) = 100/(s+1)$ (dashed) and $P_2(s) = 160000/((s+1)(s+40)^2)$ (full).

Example 9.5 *(Similar in Open Loop but Different in Closed Loop).* Systems with the transfer functions

$$P_1(s) = \frac{100}{s+1}, \qquad P_2(s) = \frac{100a^2}{(s+1)(s+a)^2}$$

have very similar open loop responses for large values of a. This is illustrated in Figure 9.18 which shows the step responses of for $a = 40$. The differences between the step responses are barely noticeable in the figure. The transfer functions from reference values to output for closed loop systems obtained with error feedback with $C = 1$ are

$$T_1 = \frac{100}{s+101}, \qquad T_2 = \frac{161600}{(s+83.92)(s^2 - 2.9254s + 1925.5)}$$

The closed loop systems are very different because the system T_1 is stable and T_2 is unstable. Notice in Figure 9.18 that the Bode plots are very close for low frequencies but different at high frequencies.

Example 9.6 *(Different in Open Loop but Similar in Closed Loop).* Systems with the transfer functions

$$P_1(s) = \frac{100}{s+1}, \qquad P_2(s) = \frac{100}{s-1}$$

have very different open loop properties because one system is unstable and the other is stable. The transfer functions from reference values to output for closed loop systems obtained with error feedback with $C = 1$ are

$$T_1(s) = \frac{100}{s+101} \qquad T_2(s) = \frac{100}{s+99}$$

which are very close.

These examples show clearly that to compare two systems by investigating their open loop properties may be strongly misleading from the point of view of feedback control. Inspired by the examples we will instead compare the properties of the closed loop systems obtained when two processes P_1 and P_2 are controlled by the same controller C. To do this it will be assumed that the closed loop systems obtained are stable. The difference between the closed loop transfer functions is

$$\delta(P_1, P_2) = \left|\frac{P_1 C}{1+P_1 C} - \frac{P_2 C}{1+P_2 C}\right| = \left|\frac{(P_1 - P_2)C}{(1+P_1 C)(1+P_2 C)}\right| \tag{9.26}$$

This is a natural way to express the closeness of the systems P_1 and P_2, when they are controlled by C. It can be verified that δ is a proper norm in the mathematical sense. There is one difficulty from a practical point of view because the norm depends on the feedback C. The norm has some interesting properties.

Assume that the controller C has high gain at low frequencies. For low frequencies we have

$$\delta(P_1, P_2) \approx \frac{P_1 - P_2}{P_1 P_2 C}$$

If C is large it means that δ can be small even if the difference $P_1 - P_2$ is large. For frequencies where the maximum sensitivity is large we have

$$\delta(P_1, P_2) \approx M_{s1} M_{s2} |C(P_1 - P_2)|$$

For frequencies where P_1 and P_2 have small gains, typically for high frequencies, we have

$$\delta(P_1, P_2) \approx |C(P_1 - P_2)|$$

This equation shows clearly the disadvantage of having controllers with large gain at high frequencies. The sensitivity to modeling error for high frequencies can thus be reduced substantially by a controller whose gain goes to zero rapidly for high frequencies. This has been known empirically for a long time and it is called high frequency roll off.

9.10 Specifications

Having understood the fundamental properties of the basic feedback loop we will now quantify the requirements on a typical control system. Control problems are rich and there are many factors that have to be taken into account.

- Load disturbance attenuation
- Measurement noise response
- Robustness to process uncertainties
- Response to command signals

The emphasis on the different factors depend on the particular problem. Robustness is important for all applications. Command signal following is the major issue in motion control, where it is desired that the system follows commanded trajectories. . This is called the *s*ervo problem. The typical process control problem is to keep the process variable close to the reference signal, which is changed only when production is altered. This is called the *r*egulation problem. Attenuation of load disturbances is therefore 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. Control buffers is a typical example. The reason for using buffers is to smooth flow variations. A good strategy is to apply control only when there is a risk that buffers tend to be empty or full.

An advantage with a structure having two degrees of freedom, or setpoint weighting, is that the problem of setpoint 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 setpoints.

We will now discuss how specifications can be expressed in terms properties of the transfer functions (9.12).

The linear behavior of the system is completely determined by six transfer functions (9.11), the Gang of Six. Neglecting setpoint response it is

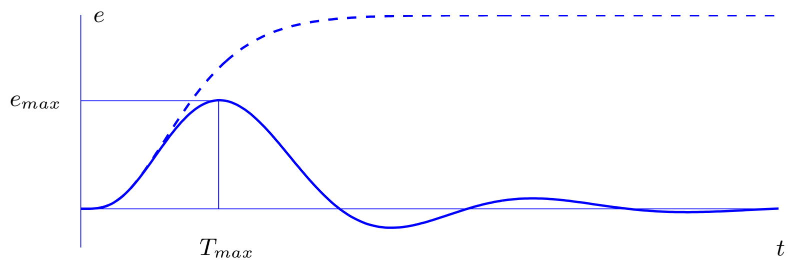

Figure 9.19: The error due to a unit step load disturbance some features used to characterize attenuation of load disturbances. The dashed curve show the open-loop error. The process transfer is $P(s) = (s+1)^{-4}$ is controlled by a PI controller having parameters $k = 1$ and $k_i = 0.4$ function is

sufficient to consider four transfer functions (9.12), the Gang of Four. Specifications can be expressed in terms of these transfer functions. It is common practice to characterize the transfer functions by a few features.

Features of Time Responses

Many criteria are related to time responses, for example the step response to setpoint 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_m is defined as

$$\begin{aligned} e_{max} &= \max_{0 \le t < \infty} |e(t)| \\ T_{max} &= \arg\max |e(t)|. \end{aligned} \tag{9.27}$$

The time T_{max} where the maximum occurs is a measure of the response time of the system. An example is given in Figure 9.19 which shows the output for a step in the load disturbance. The closed loop system has $\omega_{ms} = 0.559$, $e_{max} = 0.59$ and $T_{max} = 5.15$.

Other criteria are the integrated absolute error (IAE)

$$IAE = \int_0^\infty |e(t)|dt \tag{9.28}$$

the integrated error (IE)

$$IE = \int_0^\infty e(t)dt. \tag{9.29}$$

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.

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

$$ITNAE = \int_0^\infty t^n |e(t)| dt. \tag{9.30}$$

The integrated squared error (ISE) is defined as

$$ISE = \int_0^\infty e(t)^2 dt. \tag{9.31}$$

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

$$QE = \int_0^\infty (e^2(t) + \rho u^2(t)) dt. \tag{9.32}$$

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

Response to Command Signals

Classical specifications were strongly focused on the response of the output to step changes in the command signal. Many specifications were developed for that response, for example rise time, settling time, decay ratio, overshoot, and steady-state offset for step changes in setpoint. These quantities are defined as follows, see Figure 9.20.

- 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% to 90% of its steady state value.

- 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.

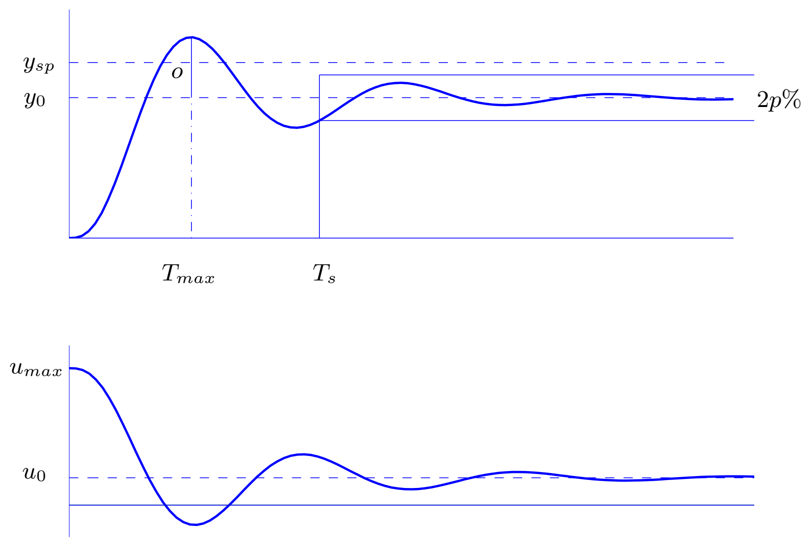

Figure 9.20: Specifications on command signal following based on the time response to a unit step in the setpoint. The upper curve shows the response of the output and the lower curve shows the corresponding control signal.

- The *decay ratio* d is the ratio between two consecutive maxima of the error for a step change in setpoint or load. 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 and the steady state value of the step response. It is often given in%. In industrial control applications it is common to specify an overshoot of 8%–10%. 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.

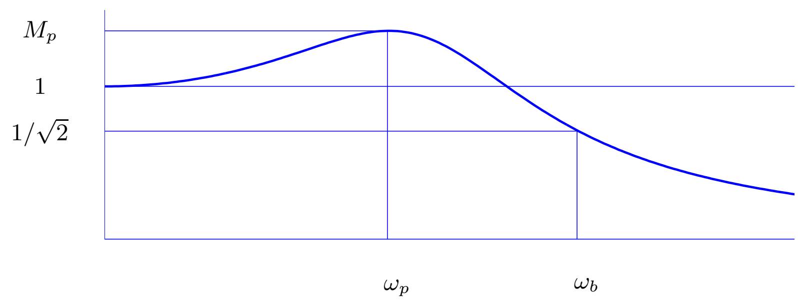

Figure 9.21: Gain curve for transfer function from setpoint to output.

Features of Frequency Responses

Specifications can also be related to frequency responses. Since specifications were originally focused on setpoint response it was natural to consider the transfer function from setpoint to output. A typical gain curve for this response is shown in Figure 9.21. It is natural to require that the steady state gain is one. Typical specifications are then.

- The *resonance peak* M_p is the largest value of the frequency response.
- The *peak frequency* ω_p is the frequency where the maximum occurs.
- The *bandwidth* ω_b is the frequency where the gain has decreased to $1/\sqrt{2}$.

For a system with error feedback the transfer function from setpoint 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 ω_{gc}.
- Gain margin g_m.
- Phase margin φ_m.
- Maximum sensitivity M_s.
- Frequency where the sensitivity function has its maximum ω_{ms}.
- Sensitivity crossover frequency ω_{sc}.

- Maximum complementary sensitivity M_t.
- Frequency where the complementary sensitivity function has its maximum ω_{ms}.

Another set of specifications is based on a Taylor series expansion of the transfer function $G_{ed}(s)$ from a disturbance to the error

$$G_{ed}(s) = e_0 + e_1 s + e_2 s^2 + \cdots .$$

The numbers e_i are called error coefficients. The first coefficient that is not zero is of particular interest. Assume for example that the first non-vanishing error coefficient is e_2. It means that constant disturbance and linearly increasing disturbances do not give any steady state errors and that the jerk disturbance $d(t) = kt^2$ gives a constant steady state error ke_2.

Relations between Time and Frequency Domain Features

There are approximate relations between specifications in the time and frequency domain. Let $G(s)$ be the transfer function from setpoint 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 ω_b, the gain crossover frequency ω_{gc}, the sensitivity frequency ω_{ms}. The product of bandwidth and rise time is approximately constant $T_r\omega_b \approx 2$. 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 lager oveshoot. Unfortunately there are no simple relation 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 a transfer functions that decay rapidly around the bandwidth. To avoid overshoots in systems with error feedback it is advisable to require that the maximum of the complementary sensitivity function is small, say $M_t = 1.1 - 1.2$.

Response to Load Disturbances

The sensitivity function (9.20) shows how feedback influences disturbances. Disturbances with frequencies that are lower than the sensitivity crossover frequency ω_{sc} are attenuated by feedback and those with $\omega > \omega_{sc}$ are amplified by feedback. The largest amplification is the maximum sensitivity M_s.

Consider the system in Figure 9.2. The transfer function from load disturbance d to process variable is

$$G_{xd} = \frac{P}{1 + PC} = PS = \frac{T}{C}. \tag{9.33}$$

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

$$G_{xd} = \frac{T}{C} \approx \frac{1}{C} \approx \frac{s}{k_i}. \tag{9.34}$$

Since load disturbances typically have low frequencies the integral gain k_i is a good measure of load disturbance rejection for systems where the controller has integral action. Figure 9.22 which gives the gain curve for a typical case shows that the approximation is very good for low frequencies. transfer functions. Measurement noise, which typically has high frequencies, generates rapid variations in the control variable which are detrimental because they cause wear in valves and motors and they can even saturate the actuator. It is important to keep the variations in the control signal at reasonable levels. A typical requirement is that the variations are only a fraction of the span of the control signal. The variations can be influenced by filtering and by proper design of the high frequency properties of the controller.

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

$$G_{un} = \frac{C}{1 + PC} = CS = \frac{T}{P}. \tag{9.35}$$

Figure 9.22 shows the gain curve of G_{un} for a typical system. For low frequencies the transfer function the sensitivity function equals 1 and (9.35) can be approximated by $1/P(s)$. For high frequencies is is approximated as $G_{un} \approx C(s)$.

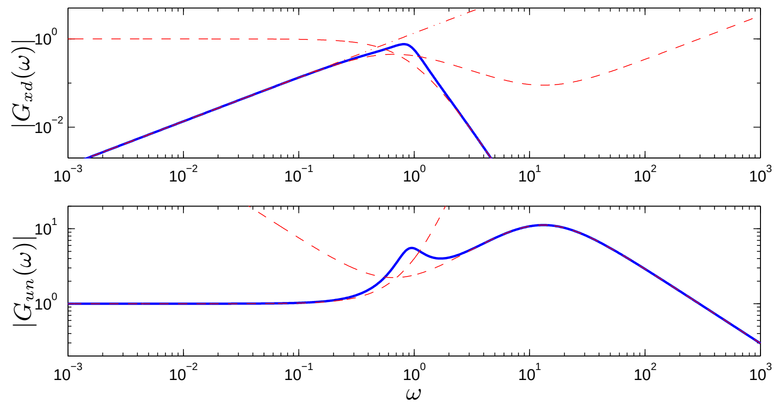

Figure 9.22: Gains of the transfer functions G_{xd} and G_{un} for PID control ($k = 2.235$, $T_i = 3.02$, $T_i = 0.756$ and $T_f = Td/5$) of the process $P = (s+1)^{-4}$. The the gain of the transfer functions $P(s)$, $C(s)$, $1/C(s)$ are shown with dashed lines and s/k_i with dash-dotted lines.

A simple measure of the effect of measurement noise is the high frequency gain of the transfer function G_{un}

$$M_{un} = \max_{\omega} |G_{un}(i\omega)|. \tag{9.36}$$

A more accurate measure is the standard deviation of the control signal. If the power spectrum of measurement noise is $\varphi_n(\omega)$, the standard deviation of the control signal is

$$\sigma_u^2 = \int_{-\infty}^{\infty} |G_{un}(i\omega)|^2 \phi_n(\omega) d\omega. \tag{9.37}$$

Tradeoffs

There are many tradeoffs in control design, one is between load disturbance rejection and measurement noise injection. This is illustrated in Figure 9.23 where a process with the transfer function $P(s) = 1/(s+1)^4$ is controlled with PI ($k = 0.5$ and $T_i = 2$) and PID ($k = 2.235$, $T_i = 3.02$, $T_i = 0.756$ and $T_f = Td/5$) controllers. The figure shows that the PID controller gives better attenuation of load disturbances $k_=0.74$ as compared with $k_i = 0.25$ for PI control. This is also illustrated in the time responses for load disturbances where the maximum error is much smaller $e_{max} = 0.38$ for PID control and

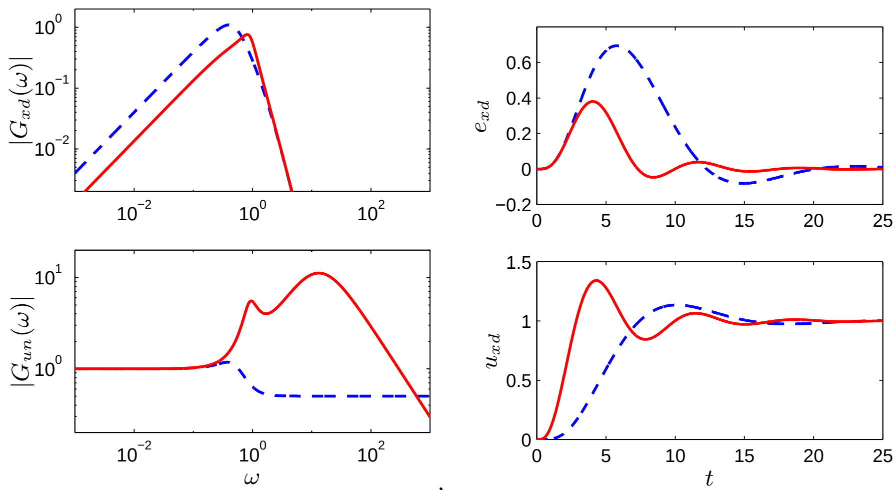

Figure 9.23: Illustrates trade off between attenuation of load disturbances and measurement noise injection. The figure on the left shows the gain curves of the transfer functions $G_{xd}(s)$ and $G_{un}(s)$ and the curves on the right shows the time responses to steps in the load disturbance. Results for PID control are shown in full lines and for PI with dashed lines.

$e_{max} = 0.69$ for PI control. Analyzing the control signals we find that the benefit by PID control is primarily due to the fact that the controller reacts faster to the disturbance. The penalty for the improved performance is that the largest gain of the transfer function G_{un} is $M_{un} = 11.2$ for PID control as compared to $M_{un} = 1.2$ for PI control.

Summary

Summarizing we find that the behavior of a closed loop system system can be characterized by the following four parameters:

- Load disturbance attenuation is described by integral gain k_i
- Measurement noise injection is described by the high frequency gain M_{un} of the transfer function from measurement noise to control signal.
- Robustness to process variations is described by the maximum sensitivities M_s and M_t

For systems permitting a controller with two degrees of freedom the desired response to command signal can be adjusted by feedforward. For systems with error feedback the overshoot of the response to load disturbances can be specified by M_t.

9.11 Further Reading

Bodes book, classical control, Doyle Francis Tannenbaum, Zhou and Doyle Vinnicombe.

9.12 Exercises