Chapter 9

Tradeoffs and Limits of Performance

9.1 Introduction

Fundamental limits of feedback systems will be investigated in this chapter. We begin in Section 9.2 by discussing the basic feedback loop and typical requirements. This includes the ability to follow reference signals, effects of load disturbances and measurement noise and the effects of process variations. It turns out that these properties can be captured by a set of six transfer functions, called the Gang of Six. These transfer functions are introduced in Section 9.3. For systems where the feedback is restricted to operate on the error signal the properties are characterized by a subset of four transfer functions, called the Gang of Four. Properties of systems with error feedback and the more general feedback configuration with two degrees of freedom are also discussed in Section 9.3. It is shown that it is important to consider all transfer functions of the Gang of Six when evaluating a control system.

Another interesting observation is that for systems with two degrees of freedom the problem of response to load disturbances can be treated separately. This gives a natural separation of the design problem into a design of a feedback and a feedforward system. The feedback handles process uncertainties and disturbances and the feedforward gives the desired response to reference signals.

Attenuation of disturbances are discussed in Section 9.4 where it is demonstrated that process disturbances can be attenuated by feedback but that feedback also feeds measurement noise into the system. It turns out

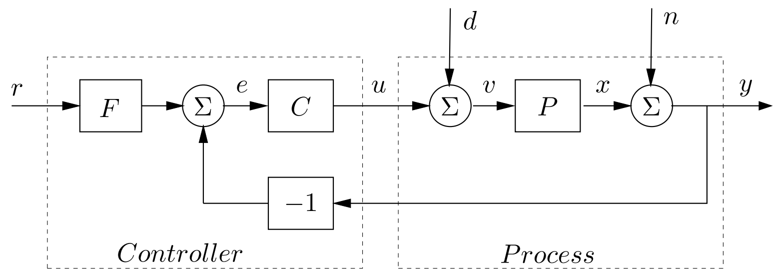

Figure 9.1: Block diagram of a basic feedback loop.

that the sensitivity function which belongs to the Gang of Four gives a nice characterization of disturbance attenuation. The effects of process variations are discussed in Section 9.5. It is shown that their effects are well described by the sensitivity function and the complementary sensitivity function. The analysis also gives a good explanation for the fact that control systems can be designed based on simplified models. When discussing process variations it is natural to investigate when two processes are similar from the point of view of control. This important nontrivial problem is discussed in Section **??**. Section 9.6 is devoted to a detailed treatment of the sensitivity functions. This leads to a deeper understanding of attenuation of disturbances and effects of process variations. A fundamental result of Bode which gives insight into fundamental limitations of feedback is also derived. This result shows that disturbances of some frequencies can be attenuated only if disturbances of other frequencies are amplified. Tracking of reference signals are investigated in Section **??**. Particular emphasis is given to precise tracking of low frequency signals. Because of the richness of control systems the emphasis on different issues varies from field to field. This is illustrated in Section **??** where we discuss the classical problem of design of feedback amplifiers.

9.2 The Basic Feedback Loop

A block diagram of a basic feedback loop is shown in Figure 9.1. The system loop is composed of two components, the process P 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 disturbance d and the measurement 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 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 9.1 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 9.1. This captures the essence and it can easily be modified if it is known precisely how disturbances enter the system.

More Abstract Representations

The block diagrams themselves are substantial abstractions but higher abstractions are sometimes useful. The system in Figure 9.1 can be represented by only two blocks as shown in Figure 9.2. 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 9.2 allows many control variables and many measured variables, but it shows less of the system structure than Figure 9.1. This representation can be used even when there are many input signals and many output signals. 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

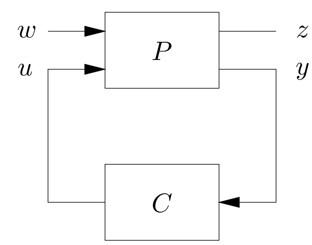

Figure 9.2: An abstract representation of the system in Figure 9.1. The input u represents the control signal and the input w represents the reference r, the load disturbance d and the measurement noise n. The output y is the measured variables and z are internal variables that are of interest.

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 very 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.

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. Basic requirements are

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

The possibility of instabilities is the primary 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. The focus on 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.

9.3 The Gang of Six

The feedback loop in Figure 9.1 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 9.1

$$\begin{aligned} X &= \frac{P}{1+PC}D - \frac{PC}{1+PC}N + \frac{PCF}{1+PC}R \\ Y &= \frac{P}{1+PC}D + \frac{1}{1+PC}N + \frac{PCF}{1+PC}R \\ U &= -\frac{PC}{1+PC}D - \frac{C}{1+PC}N + \frac{CF}{1+PC}R. \end{aligned} \tag{9.1}$$

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.

$$\begin{matrix} \dfrac{PCF}{1+PC} & \dfrac{PC}{1+PC} & \dfrac{P}{1+PC} \\ \dfrac{CF}{1+PC} & \dfrac{C}{1+PC} & \dfrac{1}{1+PC}, \end{matrix} \tag{9.2}$$

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 (9.2), i.e.

$$\begin{aligned} &\frac{PC}{1+PC}, && \text{the complementary sensitivity function} \\ &\frac{P}{1+PC}, && \text{the load disturbance sensitivity function} \\ &\frac{C}{1+PC}, && \text{the noise sensitivity function} \\ &\frac{1}{1+PC}, && \text{the sensitivity function} \end{aligned} \tag{9.3}$$

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 9.1 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 9.3 and 9.4. Figures 9.3 and 9.4 give useful insight into the properties of the closed loop system. The time responses in Figure 9.3 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 9.4 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 9.3 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 9.3 but the frequency response in 9.4 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 9.3 and 9.4. Consider for example the response of the output to load

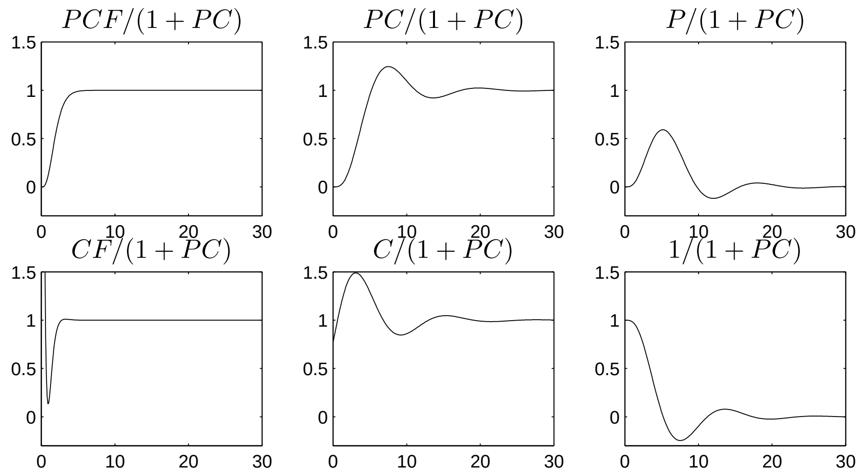

Figure 9.3: Step responses of the Gang of Six for PI control $k = 0.775$, $T_i = 2.05$ of the process $P(s) = (s+1)^{-4}$. The feedforward is designed to give the transfer function $(0.5s+1)^{-4}$ from reference r to output y.

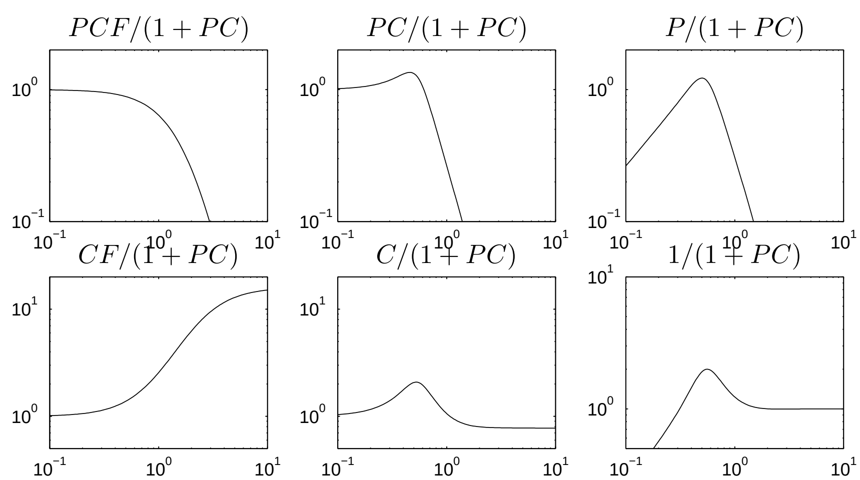

Figure 9.4: Gain curves of frequency responses of the Gang of Six for PI control $k = 0.775$, $T_i = 2.05$ of the process $P(s) = (s+1)^{-4}$ where the feedforward has been designed to give the transfer function $(0.5s+1)^{-4}$ from reference to output.

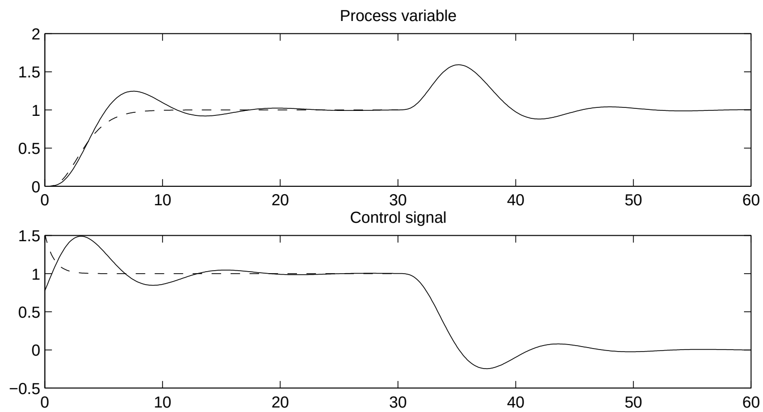

Figure 9.5: Representation of properties of a basic feedback loop by step responses in the reference at time 0, and at the process input at time 30. The dashed full lines show the response for a system with error feedback $F = 1$, and the dashed lines show responses for a system having two degrees of freedom.

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 $\omega_{max} t_{max} = 2.6$.

The step responses can also be represented by two simulations of the process. The complete system is first simulated with the full two-degree-of-freedom structure. The simulation begins with a step in the reference signal, when the system has settled to equilibrium a step in the load disturbance is then given. The process output and the control signals are recorded. The simulation is then repeated with a system without feedforward, i.e. $F = 1$. The response to the reference signal will be different but the response to the load disturbance will be the same as in the first simulation. The procedure is illustrated in Figure 9.5.

A Remark

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. To get

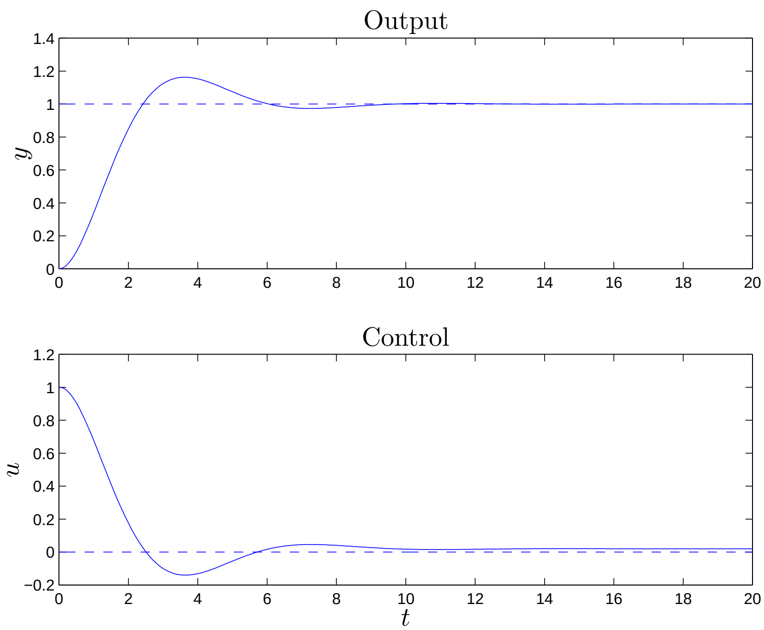

Figure 9.6: Response of output y and control u to a step in reference r.

a more complete representation of the system all six responses should be given. We illustrate the importance of this by an example.

Example 31 *(Assessment of a Control System).* A process with the transfer function

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

is controlled 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

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

Figure 9.6 shows that the responses to a reference signal look quite 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.

To explore the system further we will calculate the transfer functions of the Gang of Six, we have

$$
\begin{aligned}
\frac{PC}{1+PC} &= \frac{1}{s^2+s+1} & \frac{P}{1+PC} &= \frac{s}{(s+0.02)(s^2+s+1)}\\
\frac{C}{1+PC} &= \frac{(s+0.02)(s+1)}{s^2+s+1} & \frac{1}{1+PC} &= \frac{s(s+1)}{s^2+s+1}
\end{aligned}
$$

The responses of y and u to the reference r are given by

$$Y(s) = \frac{1}{s^2+s+1}R(s), \quad U(s) = \frac{(s+1)(s+0.02)}{s^2+s+1}R(s)$$

and the responses of y and u to the load disturbance d are given by

$$Y(s) = \frac{s}{(s+0.02)(s^2+s+1)}D(s), \quad U(s) = -\frac{1}{s^2+s+1}D(s)$$

Notice that the process pole $s = 0.02$ is cancelled 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 $e^{-0.02t}$ because the zero $s = -0.02$ will block the transmission of this signal. This is clearly seen in Figure 9.7, 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 9.7. 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.

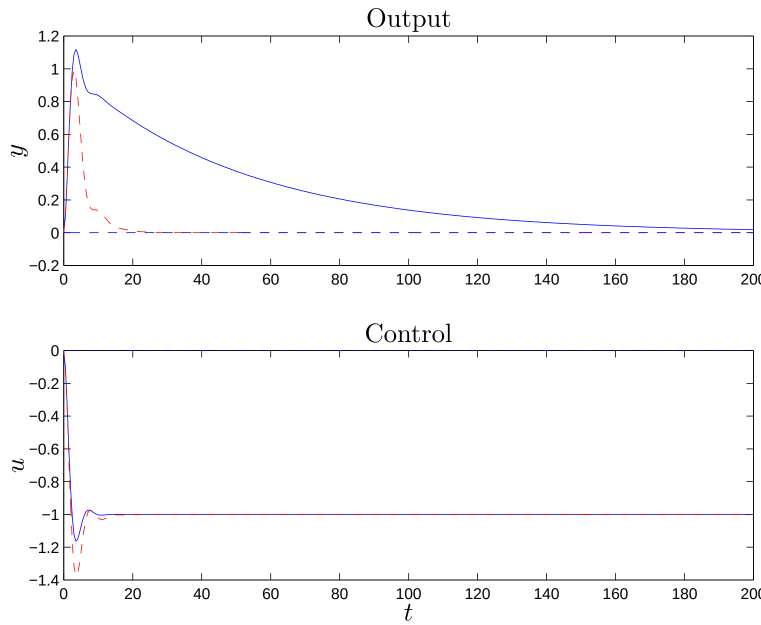

Figure 9.7: Response of output y and control u to a step in the load disturbance. Notice the very slow decay of the mode $e^{-0.02t}$. The control signal does not respond to this mode because the controller has a zero $s = -0.02$. The dashed curves show the results when the controller is modified to $C(s) = 1 + 0.2/s$.

The behavior illustrated in the example is typical when there are cancellations of poles and zeros in the transfer functions of the process and the controller. The canceled factors do not appear in the loop transfer function and the sensitivity functions. 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*. Another view of cancellations is given in Section **??**.

9.4 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

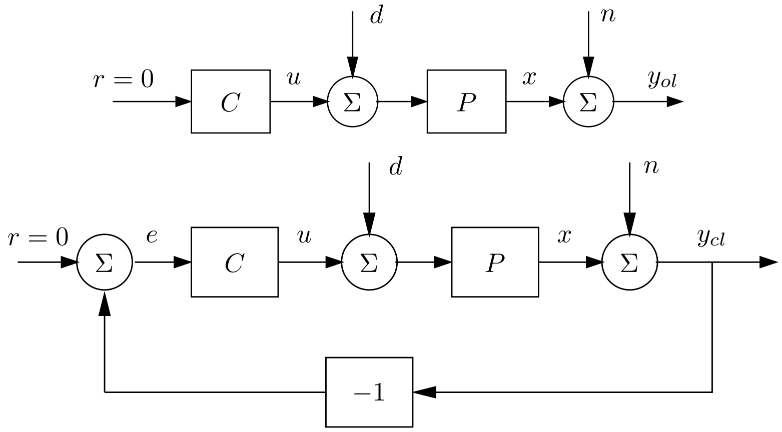

Figure 9.8: Open and closed loop systems subject to the same disturbances.

disturbances as is illustrated in Figure 9.8. Let the transfer function of the process be $P(s)$ and let the Laplace transforms of the load disturbance and the measurement noise be $D(s)$ and $N(s)$ respectively. The output of the open loop system is

$$Y_{ol} = P(s)D(s) + N(s) \tag{9.4}$$

and the output of the closed loop system is

$$Y_{cl} = \frac{P(s)D(s) + N(s)}{1 + P(s)C(s)} = S(s)(P(s)D(s) + N(s)) \tag{9.5}$$

where $S(s)$ is the sensitivity function, which belongs to the Gang of Four. We thus obtain the following interesting result

$$Y_{cl}(s) = S(s)Y_{ol}(s) \tag{9.6}$$

The sensitivity function will thus directly show the effect of feedback on the output. The disturbance attenuation can be visualized graphically by the gain curve of the Bode plot of $S(s)$. 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

$$M_s = \max_{\omega} |S(i\omega)| = \max_{\omega} \left| \frac{1}{1 + P(i\omega)C(i\omega)} \right| \tag{9.7}$$

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

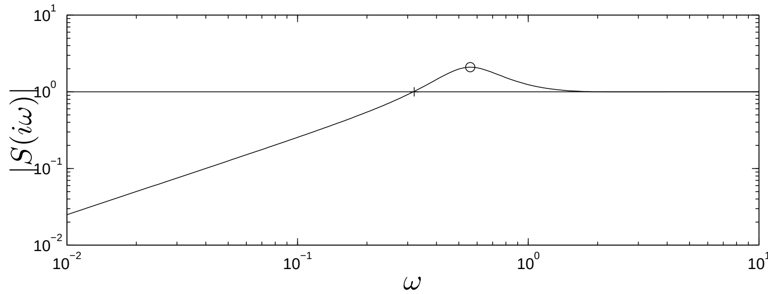

Figure 9.9: Gain curve of the sensitivity function for PI control ($k = 0.8$, $k_i = 0.4$) of process with the transfer function $P(s) = (s+1)^{-4}$. The sensitivity crossover frequency is indicated by + and the maximum sensitivity by o.

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 9.9. 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 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 $S(s)$. Figure 9.10 shows the output of the system with and without control. The sensitivity function can be written as

$$S(s) = \frac{1}{1 + P(s)C(s)} = \frac{1}{1 + L(s)}. \tag{9.8}$$

Since it 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 9.11. 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.

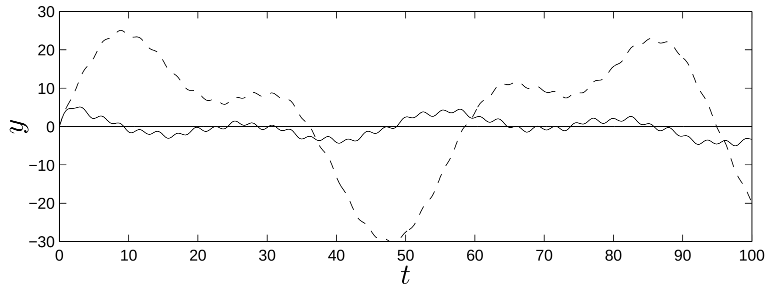

Figure 9.10: Outputs of process with control (full line) and without control (dashed line).

Variations in the Process Variable

So far we have discussed variations in the output. For the process variable we have

$$X(s) = \frac{P(s)}{1+P(s)C(s)}D(s) - \frac{P(s)C(s)}{1+P(s)C(s)}N(s) = S(s)D(s) - T(s)N(s)$$

The first term represents the effect of the load disturbance and the second term the effect of measurement noise. Load disturbance are thus attenuated but measurement noise is injected because of the feedback.

9.5 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.

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 9.11 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

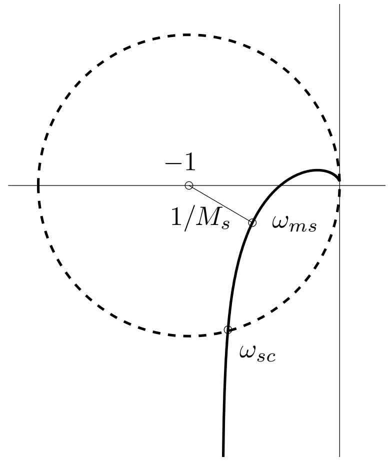

Figure 9.11: Nyquist curve of loop transfer function showing graphical interpretation of maximum sensitivity. The sensitivity crossover frequency ω_{sc} and the frequency ω_{ms} where the sensitivity has its largest value are indicated in the figure. All points inside the dashed circle have sensitivities greater than 1.

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 9.12. 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

$$|\Delta P| < \Big|\frac{1 + PC(i}{C}\Big| \tag{9.9}$$

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

$$\Big|\frac{\Delta P(i\omega)}{P(i\omega)}\Big| < \frac{1}{|T(i\omega)|} \tag{9.10}$$

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

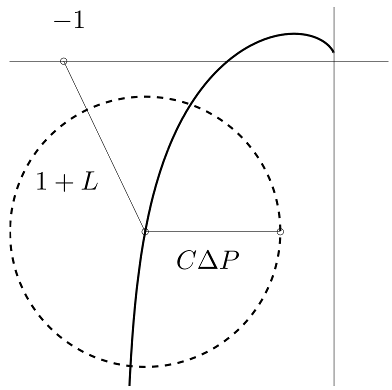

Figure 9.12: Nyquist curve of a nominal loop transfer function and its uncertainty caused by process variations ΔP.

that the condition (9.10) is conservative because it follows from Figure 9.12 that the critical perturbation is in the direction towards the critical point -1. Larger perturbations can be permitted in the other directions.

This formula (9.10) 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 (9.10) implies that the closed loop system will at least be stable for substantial variations in the process dynamics.

It follows from (9.10) 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

$$M_t = \max_{\omega} |T(i\omega)| = \max_{\omega}\left|\frac{P(i\omega)C(i\omega)}{1 + P(i\omega)C(i\omega)}\right| \tag{9.11}$$

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 32 *(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 9.13 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

- 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.

Small Gain Theorem

The robustness result given by Equation (9.10) can be given another interpretation. This is illustrated in Figure 9.14 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

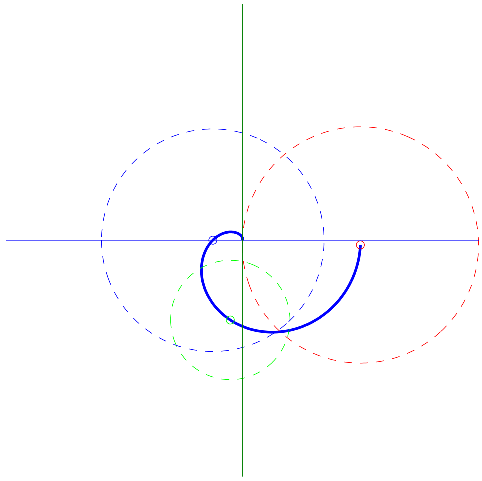

Figure 9.13: Nyquist curve of a nominal process transfer function $P(s) = (s+1)^{-4}$ shown in full lines. The circles show the uncertainty regions $|\Delta P| = 1/|T|$ obtained for a PI controller with $k = 0.775$ and $T_i = 2.05$ for $\omega = 0$, 0.46 and 1.

Figure 9.14B is

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

Equation 9.10 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. This result which holds under much more general assumptions is called the small gain theorem.

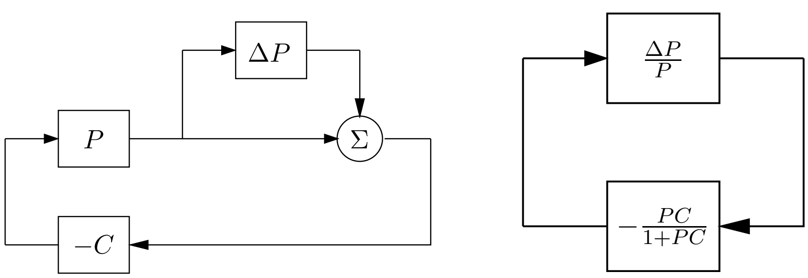

Figure 9.14: Illustration of robustness to process perturbations.

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 9.1. For simplicity we will assume that $F = 1$ and that the disturbances d and n are zero. The transfer function from reference to output is given by

$$\frac{Y}{R} = \frac{PC}{1+PC} = T \tag{9.12}$$

Compare with (9.2). The transfer function T which belongs to the Gang of Four is called the complementary sensitivity function. Differentiating (9.12) we get

$$\frac{dT}{dP} = \frac{C}{(1+PC)^2} = \frac{PC}{(1+PC)(1+PC)P} = S\frac{T}{P}$$

Hence

$$\frac{d\log T}{d\log P} = \frac{dT}{dP}\frac{P}{T} = S \tag{9.13}$$

This equation is the reason for calling S the sensitivity function. The relative error in the closed loop transfer function T will thus be small if 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 9.9 is a good way to determine the frequencies where model precision is essential.

Constraints on Design

Constraints on the maximum sensitivities M_s and M_t are important to ensure that closed loop system is insensitive to process variations. Typical constraints are that the sensitivities are in the range of 1.1 to 2. This has implications for design of control systems which are illustrated by an example.

Example 33 *(Sensitivities Constrain Closed Loop Poles).* PI control of a first order system was discussed in Section **??** where it was shown that the closed loop system was of second order and that the closed loop poles could be placed arbitrarily by proper choice of the controller parameters. The process and the controller are characterized by

$$
\begin{aligned}
Y(s) &= \frac{b}{s+a}U(s) \\
U(s) &= -kY(s) + \frac{k_i}{s}(R(s) - Y(s))
\end{aligned}
$$

where U, Y and R are the Laplace transforms of the process input, output and the reference signal. The closed loop characteristic polynomial is

$$s^2 + (a + bk)s + bk_i$$

requiring this to be equal to

$$s^2 + 2\zeta\omega_0 s + \omega_0^2 \tag{9.14}$$

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

$$
\begin{aligned}
k &= \frac{2\zeta\omega_0 - a}{b} \\
k_i &= \frac{\omega_0^2}{b}
\end{aligned}
$$

and there are no apparent constraints on the choice of parameters ζ and ω_0. Calculating the sensitivity functions we get

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

Figure 9.15 shows clearly that the sensitivities will be large if the parameter ω_0 is chosen smaller than a. The equation for controller gain also gives an indication that small values of ω_0 are not desirable because proportional gain then becomes negative which means that the feedback is positive.

We can thus conclude that if a closed loop characteristic polynomial of the form (9.14) with $\zeta \leq 1$ is chosen it is necessary to have $\omega_0 \geq a/(2\zeta)$ in order to have a system with reasonable robustness. The response time of the closed loop system thus must be sufficiently fast. It is however possible to obtain closed loop system with slower response time by choosing a closed loop characteristic polynomial with real roots. Let the characteristic polynomial be

$$(s + p_1)(s + p_2)$$

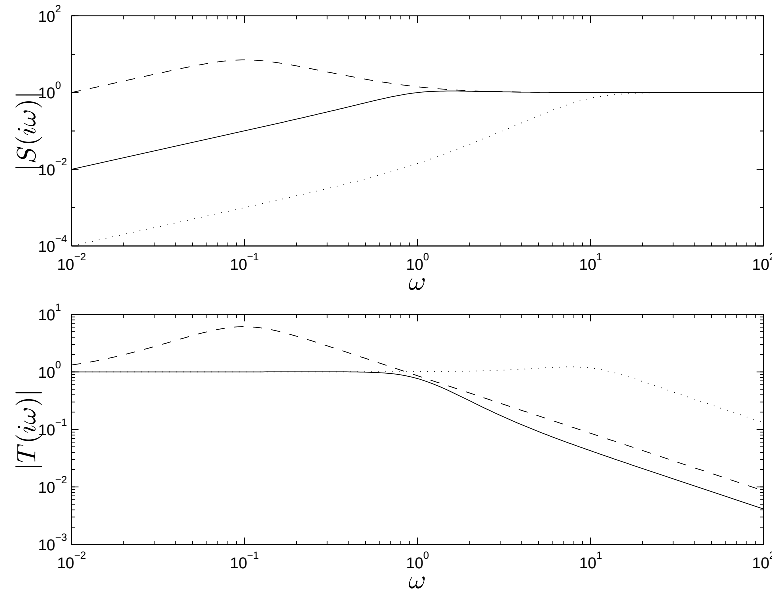

Figure 9.15: Magnitude curve for Bode plots of the sensitivity function S (above) and the complementary sensitivity function T (below) for $\zeta = 0.7$, $a = 1$ and $\omega_0/a = 0.1$ (dashed), 1 (solid) and 10 (dotted).

The controller parameters then becomes

$$
\begin{aligned}
k &= \frac{p_1 + p_2 - a}{b} \\
k_i &= \frac{p_1 p_2}{b}
\end{aligned}
$$

The is positive if one of the closed loop poles is chosen to be equal to or less than a. The sensitivity functions then becomes

$$
\begin{aligned}
S(s) &= \frac{s(s+a)}{(s+p_1)(s+p_2)} \\
T(s) &= \frac{(p_1+p_2)s + p_1 p_2}{(s+p_1)(s+p_2)}
\end{aligned}
$$

To guarantee that the sensitivity function is not too large one of the closed loop poles should be close to a. The complementary sensitivity function has a zero at

$$
z_1 = \frac{p_1 p_2}{p_1 + p_2}
$$

If $P_1 \leq p_2$ it follows that

$$0.5 < \frac{z_1}{p_1} = \frac{p_1}{p_1 + p_2} < 1$$

This implies that $|T(i\omega)|$¡2. We can thus conclude that if it is desired to have a closed loop system with slower response time one of the closed loop poles should be chosen close to the process pole a. This is another example of the fact that it is important to choose the closed loop poles carefully when using pole placement design.

Sensitivities and Relative Damping

For simple low order control systems we have based design criteria on the patterns of the poles and zeros of the complementary transfer function. To relate the general results on robustness to the analysis of the simple controllers it is of interest to find the relations between the sensitivities and relative damping. The complementary sensitivity function for a standard second order system is given by

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

This implies that the sensitivity function is given by

$$S(s) = 1 - T(s) = \frac{s(s + 2\zeta\omega_0)}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$

Straight forward but tedious calculations give

$$\begin{aligned}
M_s &= \sqrt{\frac{8\zeta^2 + 1 + (4\zeta^2 + 1)\sqrt{8\zeta^2 + 1}}{8\zeta^2 + 1 + (4\zeta^2 - 1)\sqrt{8\zeta^2 + 1}}} \\
w_{ms} &= \frac{1 + \sqrt{8\zeta^2 + 1}}{2}\omega_0 \\
M_t &= \begin{cases} 1/(2\zeta\sqrt{1 - \zeta^2}) & \text{if } \zeta \leq \sqrt{2}/2 \\ 1 & \text{if } \zeta > \sqrt{2}/2 \end{cases} \\
\omega_{mt} &= \begin{cases} \omega_0\sqrt{1 - 2\zeta^2} & \text{if } \zeta \leq \sqrt{2}/2 \\ 0 & \text{if } \zeta > \sqrt{2}/2 \end{cases}
\end{aligned} \tag{9.15}$$

The relation between the sensitivities and relative damping are shown in Figure 9.16. The values $\zeta = 0.3$, 0.5 and 0.7 correspond to the maximum sensitivities $M_s = 1.99$, 1.47 and 1.28 respectively.

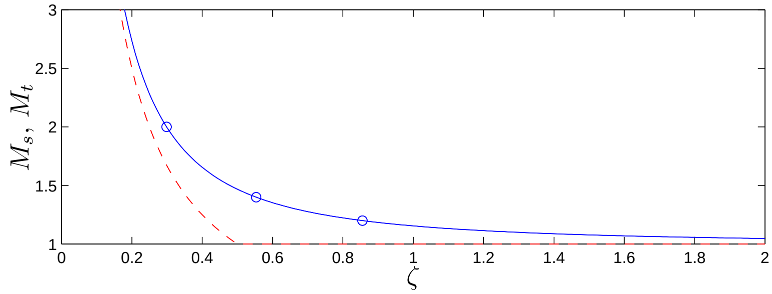

Figure 9.16: Maximum sensitivities M_s (full line) and M_t (dashed line) as functions of relative damping for $T(s) = \frac{\omega_0^2}{s^2+2\zeta\omega_0 s+\omega_0^2}$ and $S(s) = \frac{s(s+2\zeta\omega_0)}{s^2+2\zeta\omega_0 s+\omega_0^2}$.

9.6 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 (9.6) and (9.13) that it is advantageous to have a small value of the sensitivity function and it follows from (9.10) that a small value of the complementary sensitivity allows large process uncertainty. Since

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

it follows that

$$S(s) + T(s) = 1 \tag{9.16}$$

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 34 *(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 34 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 9.11 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

$$\int_0^\infty \log |S(i\omega)| d\omega = \int_0^\infty \log \frac{1}{|1+L(i\omega)|} d\omega = \pi \sum \mathrm{Re}\ p_k \tag{9.17}$$

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 has been been called the water bed effect.

Equation (9.17) 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 (9.17) 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 9.17 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.

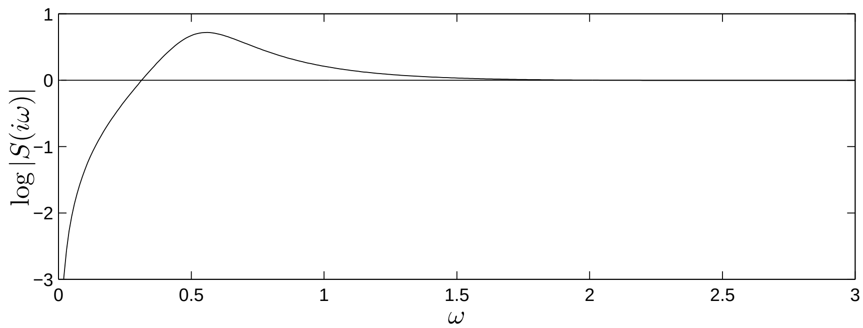

Figure 9.17: Geometric interpretation of Bode's integral formula (9.17).

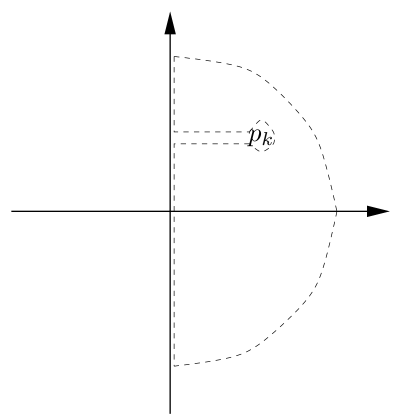

Figure 9.18: Contour used to prove Bode's theorem.

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 9.18. 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

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

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 9.18. 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

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

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 (9.17).

9.7 Fundamental Limitations*

It is important to be aware of fundamental limitations. In this section we will discuss these for the simple feedback loop. We will discuss how quickly a system can respond to changes in the reference signal. Some of the factors that limit the performance are

- Measurement noise
- Actuator saturation
- Process dynamics

Measurement Noise and Saturations

It seems intuitively reasonable that fast response requires a controller with high gain which gives a fast closed loop system. When the controller has high gain measurement noise is also amplified and fed into the system. This will result in variations in the control signal and in the process variable. It is essential that the fluctuations in the control signal are not so large that they cause the actuator to saturate. Since measurement noise typically has high frequencies the high frequency gain M_c of the controller is thus an important quantity. Measurement noise and actuator saturation thus gives a bound on the high frequency gain of the controller and therefore also on the response speed.

There are many sources of measurement noise, it can caused by the physics of the sensor, in can be electronic. In computer controlled systems it is also caused by the resolution of the analog to digital converter. Consider for example a computer controlled system with 12 bit AD and DA converters. Since 12 bits correspond to 4096 it follows that if the high frequency gain of the controller is $M_c = 4096$ one bit conversion error will make the control signal change over the full range. To have a reasonable system we may require that the fluctuations in the control signal due to measurement noise cannot be larger than 5% of the signal span. This means that the high frequency gain of the controller must be restricted to 200.

Dynamics Limitations

The limitations caused by noise and saturations seem quite obvious. It turns out that there may also be severe limitations due to the dynamical properties of the system. This means that there are systems that are inherently difficult or even impossible to control. Designers of any system should be aware of this. Since systems are often designed from static considerations the difficulties caused by dynamics do not show up. We have already encountered this in Section **??** where we found that it was necessary to choose the dominant pole so that ω_0 is larger than the fastest unstable pole but smaller than the slowest zero. It seems intuitively reasonable that a fast closed loop system is required to stabilize an unstable pole and that the response speed should be matched to the unstable pole. In the same way it seems reasonable that it is not possible to get a very rapid response for a system with a time delay. Since a right half plane zero $s = z$ is similar to a time delay $T = 1/2z$ it then follows that a right half plane zero limits the achievable response time.

To give quantitative results we will characterize the closed loop system by the gain crossover frequency ω_{gc}. This is the smallest frequency where the loop transfer function has unit magnitude, i.e. $|L(i\omega_{gc})|$. This parameter is approximately inversely proportional to the response time of a system. The dynamic elements that cause limitations are time delays and poles and zeros in the right half plane. The key observations are:

- A time delay T_d limits the response speed. A simple rule of thumb is

$$\omega_{gc} T_d < 0.7 \tag{9.18}$$

- A right half plane zero z_{rhp} limits the response speed. A simple rule of thumb is

$$\omega_{gc} < 0.5 z_{rhp} \tag{9.19}$$

  Slow RHP zeros are thus particularly bad. Notice that if a time delay is approximated by a zero in the right half plane we can apply the rule for right half plane zeros to get $\omega_{gc} T_d < 1$.

- A right half plane pole p_{rhp} requires high gain crossover frequency. A simple rule of thumb is

$$\omega_{gc} > 2 p_{rhp} \tag{9.20}$$

  Fast unstable poles require a high crossover frequency.

- Systems with a right half plane pole p and a right half plane zero z cannot be controlled unless the pole and the zero are well separated. A simple rule of thumb is

$$p_{rhp} > 6 z_{rhp} \tag{9.21}$$

- A system with a a right half plane pole and a time delay T_d cannot be controlled unless the product $p_{rhp}T_d$ is sufficiently small. A simple rule of thumb is
$$p_{rhp}T_d < 0.16 \tag{9.22}$$

A detailed discussion will be given in Chapter **??**. We illustrate with a few examples.

Example 35 *(Balancing an Inverted Pendulum).* Consider the situation when we attempt to balance a pole manually. An inverted pendulum is an example of an unstable system. With manual balancing there is a neural delay which is about $T_d = 0.04$ s. The transfer function from horizontal position of the pivot to the angle is

$$G(s) = \frac{s^2}{s^2 - \frac{g}{\ell}}$$

where $g = 9.8\ m/s^2$is the acceleration of gravity and ℓ is the length of the pendulum. The system has a pole $p = \sqrt{g/\ell}$. The inequality (9.22) gives

$$0.04\sqrt{g/\ell} = 0.16$$

Hence, $\ell = 0.6$ m. Investigate the shortest pole you can balance.

Example 36 *(Bicycle with rear wheel steering).* The dynamics of a bicycle was derived in Section **??**. To obtain the model for a bicycle with rear wheel steering we can simply change the sign of the velocity. It then follows from (**??**) that the transfer function from steering angle β to tilt angle θ is

$$P(s) = \frac{mV_0\ell}{b}\frac{Js^2 - mgl}{-as + V_0}$$

Notice that the transfer function depends strongly on the forward velocity of the bicycle. The system thus has a right half plane pole at $p = \sqrt{mg\ell/J}$ and a right half plane zero at $z = V_0/a$, and it can be suspected that the system is difficult to control. The location of the pole does not depend on velocity but the the position of the zero changes significantly with velocity. At low velocities the zero is at the origin. For $V_0 = a\sqrt{mg\ell/J}$ the pole and the zero are at the same location and for higher velocities the zero is to the right of the pole. To draw some quantitative conclusions we introduce the numerical values $m = 70$ kg, $\ell = 1.2$ m, $a = 0.7$, $J = 120$ kgm^2 and $V = 5$ m/s, give $z = V/a = 7.14$ rad/s and $p = \omega_0 = 2.6$ rad/s we find that $p = 2.6$. With $V_0 = 5$ m/s we get $z = 7.1$, and $p/z = 2.7$. To have a situation where the system can be controlled it follows from (9.21) that to have $z/p = 6$

the velocity must be increased to 11 m/s. We can thus conclude that if the speed of the bicycle can be increased to about 10 m/s so rapidly that we do not loose balance it can indeed be ridden.

The bicycle example illustrates clearly that it is useful to assess the fundamental dynamical limitations of a system at an early stage in the design. If this had been done the it could quickly have been concluded that the study of rear wheel steered motor bikes in **??** was not necessary.

Remedies

Having understood factors that cause fundamental limitations it is interesting to know how they should be overcome. Here are a few suggestions.

Problems with sensor noise are best approached by finding the roots of the noise and trying to eliminate them. Increasing the resolution of a converter is one example. Actuation problems can be dealt with in a similar manner. Limitations caused by rate saturation can be reduced by replacing the actuator.

Problems that are caused by time delays and RHP zeros can be approached by moving sensors to different places. It can also be beneficial to add sensors. Recall that the zeros depend on how inputs and outputs are coupled to the states of a system. A system where all states are measured has no zeros.

Poles are inherent properties of a system, they can only be modified by redesign of the system.

Redesign of the process is the final remedy. Since static analysis can never reveal the fundamental limitations it is very important to make an assessment of the dynamics of a system at an early stage of the design. This is one of the main reasons why all system designers should have a basic knowledge of control.

9.8 Summary

Having got insight into some fundamental properties of the feedback loop we are in a position to discuss how to formulate specifications on a control system. It was mentioned in Section 9.2 that requirements on a control system should include stability of the closed loop system, robustness to model uncertainty, attenuation of measurement noise, injection of measurement noise ability to follow reference signals. From the results given in this section we also know that these properties are captured by six transfer functions called the Gang of Six. The specifications can thus be expressed in terms of these transfer functions.

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

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

Load disturbance attenuation is described by the transfer function from load disturbances to process output

$$G_{yd} = \frac{P}{1+PC} = PS.$$

The effect of measurement noise is be captured by the transfer function

$$-G_{un} = \frac{C}{1+PC} = CS,$$

which describes how measurement noise influences the control signal. The response to set point changes is described by the transfer functions

$$G_{yr} = \frac{FPC}{1+PC} = FT, \qquad G_{ur} = \frac{FC}{1+PC} = FCS$$

Compare with (9.1). A significant advantage with controller structure with two degrees of freedom 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.

9.9 Further Reading