6

PID Design

6.1 Introduction

This chapter describes methods for finding parameters of a PID controller, which is a special case of the problem of control system design that was discussed in Chapter 4. Design of PID controllers differs from the general design problem because the controller complexity is restricted. The general design methods give a controller with a complexity that matches the process model. To obtain a controller with restricted complexity we can either simplify the process models so that the design gives a PID controller, or we can design a controller for a complex model and approximate it with a PID controller. Another reason why special design methods for PID controllers emerged is the desire to have simple design methods that can be used by persons with poor knowledge of control. The situation has changed substantially with the advent of tuning tools and automatic tuners, which have made it possible to improve the process knowledge and permitted the use of more extensive calculations. This has brought design of PID controllers closer to the mainstream of control systems design.

In this chapter it has been attempted to strike a balance by providing both a historical perspective and to present powerful methods. Section 6.2 describes the methods developed by Ziegler and Nichols, which have had a major impact on the practice of PID control even if they do not result in good tuning. Some extensions of the Ziegler-Nichols methods are also discussed.

It is often necessary to complement the design methods with manual fine-tuning to obtain the desired goals of the closed-loop dynamics. These manual tuning rules are discussed in Section 6.3.

Section 6.4 presents the pole placement method, which is one of the main stream methods in control system design. To apply this method it is necessary to approximate process dynamics by a first order model for PI control and a second order model for PID control. Instead of attempting to position all closed-loop poles, it can be attempted to assign only a few dominating poles. Such methods are discussed in Section 6.4. The most common dominant pole placement design method is the lambda tuning method, presented in Section 6.5.

In Section 6.6, algebraic tuning methods are presented. In these methods,

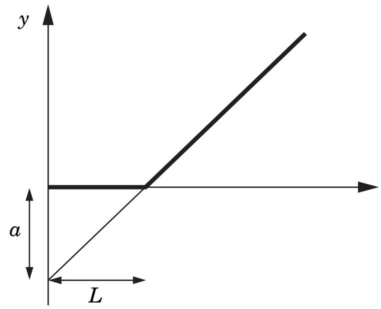

Figure 6.1 Characterization of a step response in the Ziegler-Nichols step response method.

the controller parameters are obtained from the specifications by a direct algebraic calculation. In these techniques it is also necessary to approximate process dynamics by low order models.

Many techniques for control system design are based on optimization. This gives a very flexible way of balancing conflicting design criteria. It is also possible to apply directly to controllers having restricted complexity. A number of uses of optimization for PID control are discussed in Section 6.7.

Loop shaping is another well-known technique for control system design. In Section 6.8 it is shown how this can be used for PID control. This gives a very flexible design method, which allows a nice trade-off between performance and robustness. An analysis of the method also gives useful insight into the difficulties with derivative action.

Conclusions and references are given in Sections 6.9 and 6.10.

6.2 Ziegler-Nichols and Related Methods

Two classical methods for determining the parameters of PID controllers were presented by Ziegler and Nichols in 1942. These methods are still widely used, either in their original form or in some modification. They often form the basis for tuning procedures used by controller manufacturers and the process industry. The methods are based on determination of some features of process dynamics. The controller parameters are then expressed in terms of the features by simple formulas. It is surprising that the methods are so widely referenced because they give moderately good tuning only in restricted situations. Plausible explanations may be the simplicity of the methods and the fact that they can be used for simple student exercises in basic control courses.

The Step Response Method

The first design method presented by Ziegler and Nichols is based on process information in the form of the open-loop step response. This method can be viewed as a traditional method based on modeling and control where a very simple process model is used. The step response is characterized by only two parameters a and L, as shown in Figure 6.1. Compare also with Figure 2.32.

Table 6.1 Controller parameters for the Ziegler-Nichols step response method.

Controller	aK	T_i/L	T_d/L	T_p/L
P	1			4
PI	0.9	3		5.7
PID	1.2	2	L/2	3.4

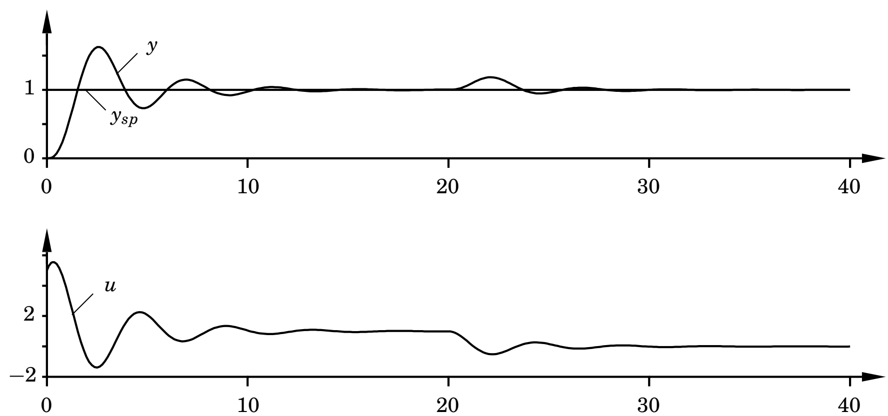

Figure 6.2 Set-point and load disturbance response of a process with transfer function $1/(s+1)^3$ controlled by a PID controller tuned with the Ziegler-Nichols step response method. The diagrams show set point y_{sp}, process output y, and control signal u.

The point where the slope of the step response has its maximum is first determined, and the tangent at this point is drawn. The intersections between the tangent and the coordinate axes give the parameters a and L. In Chapter 2, a model of the process to be controlled was derived from these parameters. This corresponds to modeling a process by an integrator and a time delay. Ziegler and Nichols have given PID parameters directly as functions of a and L. These are given in Table 6.1. An estimate of the period T_p of the closed-loop system is also given in the table.

EXAMPLE 6.1—ZIEGLER-NICHOLS STEP RESPONSE METHOD
Ziegler-Nichols' method will be applied to a process with the transfer function

$$P(s) = \frac{1}{(s+1)^3}. \tag{6.1}$$

Measurements on the step response give the parameters $a = 0.218$ and $L = 0.806$. The controller parameters can now be determined from Table 6.1. The parameters of a PID controller are $K = 5.50$, $T_i = 1.61$, and $T_d = 0.403$. The response of the closed-loop systems to a step change in set point followed by a step change in the load is shown in Figure 6.2. The behavior of the controller is as can be expected. The decay ratio for the step response is close to one quarter.

Table 6.2 Controller parameters for the Ziegler-Nichols frequency response method.

Controller	K/K_u	T_i/T_u	T_d/T_u	T_p/T_u
P	0.5			1.0
PI	0.4	0.8		1.4
PID	0.6	0.5	0.125	0.85

It is smaller for the load disturbance. The overshoot in the set-point response is too large. This can be improved by the set-point weighting b. Compare with Section 3.4. □

The Frequency Response Method

This method is also based on a simple characterization of the process dynamics. The design is based on knowledge of the point on the Nyquist curve of the process transfer function $P(s)$ where the Nyquist curve intersects the negative real axis. In Section 2.4 this point was characterized by K_{180} and ω_{180}. For historical reasons the point has been referred to as the ultimate point and characterized by the parameters $K_u = 1/K_{180}$ and $T_u = 2\pi/\omega_{180}$, which are called the *ultimate gain* and the *ultimate period*. These parameters can be determined in the following way. Connect a controller to the process, and set the parameters so that control action is proportional, i.e., $T_i = \infty$ and $T_d = 0$. Increase the gain slowly until the process starts to oscillate. The gain when this occurs is K_u, and the period of the oscillation is T_u. We have $K_u = 1/K_{180}$ and $T_u = 2\pi/\omega_u$. The parameters can also be determined approximately by relay feedback as is discussed in Section 2.7.

Ziegler-Nichols have given simple formulas for the parameters of the controller in terms of the ultimate gain and the ultimate period shown in Table 6.2. An estimate of the period T_p of the dominant dynamics of the closed-loop system is also given in the table.

The frequency response methods can also be viewed as an empirical tuning procedure where the controller parameters are obtained by direct experiments on the process combined with some simple rules. For a proportional controller the rule is simply to increase the gain until the process oscillates and then to reduce the gain by 50 percent.

We illustrate the design procedure with an example.

EXAMPLE 6.2—THE ZIEGLER-NICHOLS FREQUENCY RESPONSE METHOD
Consider the same process as in Example 6.1. The process given by (6.1) has the ultimate gain $K_u = 8$ and the ultimate period $T_u = 2\pi/\sqrt{3} = 3.63$. Table 6.2 gives the parameters $K = 4.8$, $T_i = 1.81$, and $T_d = 0.44$ for a PID controller. The closed-loop set-point and load disturbance responses when the controller is applied to the process given by (6.1) are shown in Figure 6.3.

The parameters and the performance of the controllers obtained with the frequency response method are close to those obtained by the step response method. The responses are slightly better damped. □

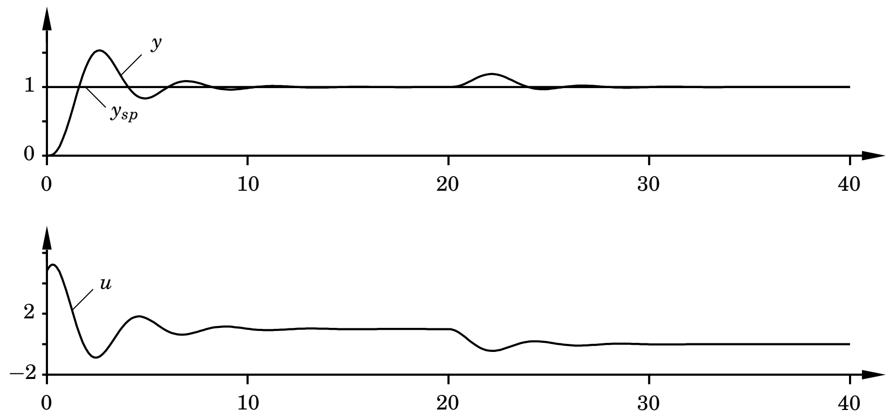

Figure 6.3 Set-point and load disturbance response of a process with the transfer function $1/(s+1)^3$ controlled by a PID controller that is tuned with the Ziegler-Nichols frequency response method. The diagrams show set point y_{sp}, process output y, and control signal u.

The Ziegler-Nichols tuning rules were originally designed to give systems with good responses to load disturbances. They were obtained by extensive simulations of many different systems with manual assessment of the results. The design criterion was quarter amplitude decay ratio, which is often too large, as is seen in the examples. For this reason the Ziegler-Nichols method often requires modification or re-tuning. Since the primary design objective was to reduce load disturbances, it is often necessary to choose set-point weighting carefully in order to obtain a satisfactory set-point response.

An Interpretation of the Frequency Response Method

The frequency response method can be interpreted as a method where one point of the Nyquist curve is positioned. With PI or PID control, it is possible to move a given point on the Nyquist curve of the process transfer function to an arbitrary position in the complex plane, as indicated in Figure 6.4. By changing the gain, a point on the Nyquist curve is moved radially from the origin. The point can be moved in the orthogonal direction by changing integral or derivative gain. Notice that with positive controller parameters the point can be moved to a quarter plane with PI or PD control and to a half plane with PID control. From this point of view the Ziegler-Nichols method can be interpreted as a primitive loop-shaping method where one point of the loop transfer function is moved to a desired point.

The frequency response method starts with determination of the point $(-1/K_u, 0)$ where the Nyquist curve of the open-loop transfer function intersects the negative real axis.

Let us now investigate how the ultimate point is changed by the controller. For a PI controller with Ziegler-Nichols tuning we have $K = 0.4K_u$ and $\omega_u T_i = (2\pi/T_u)0.8T_u = 5.02$. Therefore, the transfer function of the PI controller at

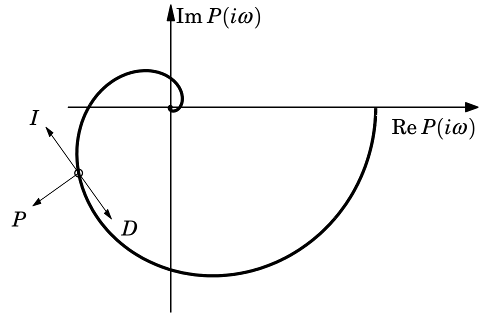

Figure 6.4 Illustrates that a point on the Nyquist curve of the process transfer function may be moved to another position by PID control. The point marked with a circle may be moved in the directions $P(i\omega)$, $-iP(i\omega)$, and $iP(i\omega)$ by changing the proportional, integral, and derivative gain, respectively.

the ultimate frequency is

$$C(i\omega_u) = K\left(1 + \frac{1}{i\omega_u T_i}\right) = 0.4K_u(1 - i/5.02) = K_u(0.4 - 0.08i).$$

The ultimate point is thus moved to $-0.4 + 0.08i$. This means that a lag of $11.2°$ is introduced at the ultimate frequency.

For a PID controller we have $K = 0.6K_u$, $\omega_u T_i = \pi$, and $\omega_u T_d = \pi/4$. The frequency response of the controller at frequency ω_u is

$$C(i\omega_u) = K\left(1 + i\left(\omega_u T_d - \frac{1}{\omega_u T_i}\right)\right) \approx 0.6K_u(1 + 0.467i).$$

This controller gives a phase advance of $25°$ at the ultimate frequency. The loop transfer function is

$$G_\ell(i\omega_u) = P(i\omega_u)C(i\omega_u) = -0.6(1 + 0.467i) = -0.6 - 0.28i.$$

The Ziegler-Nichols frequency response method for a PID controller thus moves the ultimate point $(-1/K_u, 0)$ to the point $-0.6 - 0.28i$. The distance from this point to the critical point is 0.5. This means that the method gives a sensitivity that is always greater than 2.

It has been suggested by Pessen to move the ultimate point to $-0.2 - 0.36i$ or $-0.2 - 0.21i$. Suda used approximations to obtain $M_t = 1.3$ by moving the critical point to $-0.628 - 0.483i$.

Design of PI Controller with a Given Phase Margin

Using the idea that the PI controller can be interpreted as moving a point on the loop transfer function it is easy to develop a design method that gives a closed-loop system with a given phase margin. Let the process transfer function be

$$P(i\omega) = \alpha(\omega) + i\beta(\omega) = \rho(\omega)e^{i\psi(\omega)}.$$

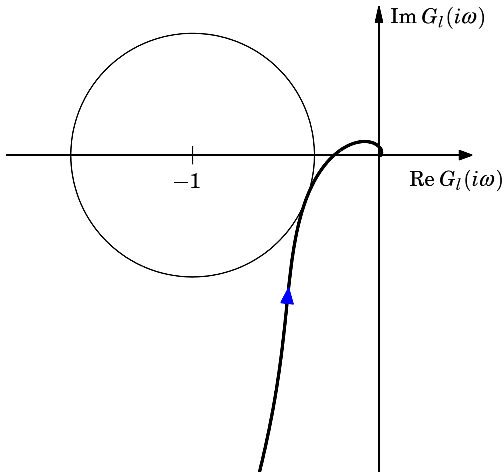

Figure 6.5 Nyquist plot for the loop transfer function G_l for PI control of the process $P(s) = e^{-\sqrt{s}}$. The controller was designed to give the phase margin of 60°.

With PI control the loop transfer function becomes

$$G_l(i\omega) = \Big(k - i\frac{k_i}{\omega}\Big)\big(\alpha(\omega) + i\beta(\omega)\big) = \alpha(\omega)k + \frac{\beta(\omega)k_i}{\omega} + i\Big(\beta(\omega)k - \frac{\alpha(\omega)k_i}{\omega}\Big).$$

Let ω_{gc} be the gain crossover frequency; requiring that the system has a phase margin φ_m it follows that

$$G_l(i\omega_{gc}) = -\cos(\varphi_m) - i\sin(\varphi_m),$$

which implies that

$$\begin{aligned}\alpha(\omega_{gc})k + \frac{\beta(\omega_{gc})k_i}{\omega_{gc}} &= -\cos(\varphi_m)\\ \beta(\omega_{gc})k - \frac{\alpha(\omega_{gc})k_i}{\omega_{gc}} &= -\sin(\varphi_m).\end{aligned}$$

Solving this equation for k and k_i gives

$$\begin{aligned}k &= -\frac{\alpha(\omega_{gc})\cos\varphi_m + \beta(\omega_{gc})\sin\varphi_m}{\alpha^2(\omega_{gc}) + \beta^2(\omega_{gc})} = -\frac{1}{\rho(\omega_{gc})}\cos\left(\varphi_m - \psi(\omega_{gc})\right)\\ k_i &= \omega_{gc}\frac{\alpha(\omega_{gc})\sin\varphi_m - \beta(\omega_{gc})\cos\varphi_m}{\alpha^2(\omega_{gc}) + \beta^2(\omega_{gc})} = \frac{\omega_{gc}}{\rho(\omega_{gc})}\sin\left(\varphi_m - \psi(\omega_{gc})\right).\end{aligned} \tag{6.2}$$

It is thus straightforward to compute the controller gains when the gain crossover frequency is given. Reasonable values of the gain crossover frequency are in the range $\omega_{90} \le \omega_{gc} \le \omega_{180-\varphi_m}$. The method can be improved by sweeping over ω_{gc} to maximize integral gain. Applying the method to design a PI controller for the process $P(s) = e^{-\sqrt{s}}$ with a phase margin of 60° gives $\omega_{gc} = 5.527$ $K = 4.79$ and $T_i = 0.392$ and $M_s = 1.53$. The Nyquist plot of the loop transfer function is shown in Figure 6.5.

Relations Between the Ziegler-Nichols Tuning Methods

The step response method and the frequency response method do not give the same values of the controller parameters. Comparing Examples 6.1 and 6.2 we find that the controller gains are 5.5 and 4.8 and that the integral times are 1.61 and 1.81. The step response method will in general give larger gains and smaller integral times. This is further illustrated in the following example.

EXAMPLE 6.3—PROCESS WITH INTEGRATION AND DELAY
Consider a process with the transfer function

$$P(s) = \frac{K_v}{s} e^{-sL},$$

which is the model originally used by Ziegler and Nichols to derive their tuning rules for the step response method. For this process we have $a = K_v L$. The ultimate frequency is $\omega_u = \pi/2L$, which gives the ultimate period $T_u = 4L$, and the ultimate gain is $K_u = \pi/2K_v L$.

For PI control the step response method gives the following parameters:

$$K = \frac{0.9}{K_v L}, \quad T_i = 3L.$$

This can be compared with the parameters

$$K = \frac{0.63}{K_v L}, \quad T_i = 3.2L$$

obtained for the frequency response method. Notice that the integral times are within 10 percent, but that the step response method gives a gain that is about 40 percent higher.

The PID parameters obtained from the step response method are

$$K = \frac{1.2}{bL}, \quad T_i = 2L \quad \text{and} \quad T_d = \frac{L}{2},$$

and those given by the frequency response methods are

$$K = \frac{0.94}{bL}, \quad T_i = 2L \quad \text{and} \quad T_d = \frac{L}{2}.$$

Both methods give the same values of integral and derivative times, but the step response method gives a gain that is about 25 percent higher than the frequency response method. □

EXAMPLE 6.4—PROCESS WITH PURE DELAY
Consider a process with the transfer

$$P(s) = K_p e^{-sL}.$$

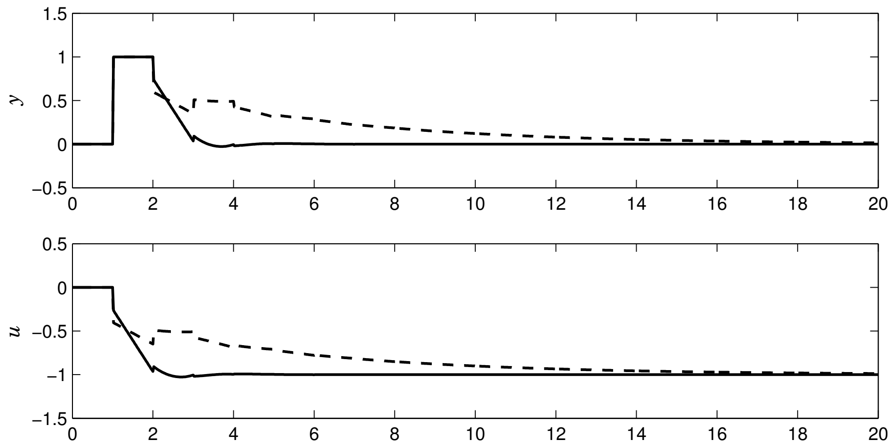

Figure 6.6 Responses to a load disturbance for a process with pure delay ($L = 1$) with PI controllers tuned by Ziegler-Nichols frequency response method (dashed) and a proper method (solid).

In this case we find that $a = \infty$! The step response method thus gives zero controller gain for PI and PID control.

The ultimate period is $T_u = 2L$, and the ultimate gain is $K_u = 1/K_p$. Using the frequency response method it follows from Table 6.2 that $KK_p = 0.4$ and $T_i = 1.6L$ for PI control. The PI controller gives a very poor result as is illustrated in Figure 6.6. The integral action is too small, which implies that it takes a very long time for the error to approach zero. For comparison we also show the response with a PI controller having $KK_p = 0.25$ and $T_i = 0.35$. This controller has a much better response to load disturbances.

For PID control the frequency response method gives $KK_p = 0.6$, $T_i = L$ and $T_d = 0.25$, which results in an unstable closed-loop system. □

These examples show that there can be considerable differences between the controller parameters obtained by the step response and the frequency response methods.

The Chien, Hrones, and Reswick Method

There have been many suggestions for modifications of the Ziegler-Nichols methods. There are methods that use the same information about the process as the Ziegler-Nichols methods, but the coefficients in Tables 6.1 and 6.2 are modified. Many methods of this type are used by controller manufacturers. There are also other methods that use more process data. Many methods are based on the idea that the process is approximated with the FOTD model

$$P(s) = \frac{K_p}{1 + sT} e^{-sL}.$$

As an illustration we will describe a method developed by Chien, Hrones, and Reswick (CHR). Their method gives closed-loop systems with slightly better

Table 6.3 Controller parameters obtained from the Chien, Hrones and Reswick load disturbance response method.

	No overshoot			20% overshoot		
Controller	aK	T_i/L	T_d/L	aK	T_i/L	T_d/L
P	0.3			0.7		
PI	0.6	4.0		0.7	2.3	
PID	0.95	2.4	0.42	1.2	2.0	0.42

Table 6.4 Controller parameters obtained from the Chien, Hrones and Reswick set-point response method.

	No overshoot			20% overshoot		
Controller	aK	T_i/L	T_d/L	aK	T_i/L	T_d/L
P	0.3			0.7		
PI	0.35	1.2		0.6	1.0	
PID	0.6	1.0	0.5	0.95	1.4	0.47

robustness than the Ziegler-Nichols method. The design criteria used were "quickest response without overshoot" or "quickest response with 20 percent overshoot." They proposed different tuning rules for load disturbances and set-point response.

To tune the controller according to the CHR method, the parameters a and L of the process model are first determined in the same way as for the Ziegler-Nichols step response method. The controller parameters are then given as functions of these two parameters. The tuning rule for load disturbance response are given in Table 6.3. The tuning rules in Table 6.3 have in general lower gains than the corresponding Ziegler-Nichols rule in Table 6.1.

Chien, Hrones, and Reswick found that tuning for set-point response was different than tuning for load disturbances. At that time the advantages of set-point weighting and systems with two degrees of freedom were not known. An additional parameter, time constant T, was required, and the controller gains were in general lower; see Table 6.4.

The Cohen-Coon Method

The Cohen-Coon method is also based on the FOTD process model

$$P(s) = \frac{K_p}{1+sT}\, e^{-sL}.$$

The main design criterion is rejection of load disturbances. It attempts to position dominant poles that give a quarter amplitude decay ratio. For P and PD

Table 6.5 Controller parameters from the Cohen-Coon method.

Controller	aK	T_i/L	T_d/L
P	$1+\dfrac{0.35\tau}{1-\tau}$		
PI	$0.9\left(1+\dfrac{0.092\tau}{1-\tau}\right)$	$\dfrac{3.3-3.0\tau}{1+1.2\tau}$	
PD	$1.24\left(1+\dfrac{0.13\tau}{1-\tau}\right)$		$\dfrac{0.27-0.36\tau}{1-0.87\tau}$
PID	$1.35\left(1+\dfrac{0.18\tau}{1-\tau}\right)$	$\dfrac{2.5-2.0\tau}{1-0.39\tau}$	$\dfrac{0.37-0.37\tau}{1-0.81\tau}$

controllers the poles are adjusted to give maximum controller gain, subject to the constraint on the decay ratio. This minimizes the steady-state error due to load disturbances. For PI and PID control the integral gain $k_i = K/T_i$ is maximized. This corresponds to minimization of IE, the integral error due to a unit step load disturbance. For PID controllers three closed-loop poles are assigned; two poles are complex, and the third real pole is positioned at the same distance from the origin as the other poles. The pole pattern is adjusted to give quarter amplitude decay ratio, and the distance of the poles to the origin are adjusted to minimize IE.

Since the process is characterized by three parameters (K_p, L, and T), it is possible to give tuning formulas where controller parameters are expressed in terms of these parameters. Such formulas were derived by Cohen and Coon based on analytical and numerical computations. The formulas are given in Table 6.5. The parameters $a = K_pL/T$ and $\tau = L/(L+T)$ are used in the table to facilitate comparisons with Ziegler-Nichols tuning. A comparison with Table 6.1 shows that the controller parameters are close to those obtained by the Ziegler-Nichols step response method for small τ. Also notice that the integral time decreases for increasing τ, which is desirable as was found in Section 6.2. A peculiarity is that the gains go to infinity when τ goes to 1, which is not correct. The method does also suffer from the decay ratio being too large, which means that the closed-loop systems obtained have poor damping and high sensitivity.

Commentary

The Ziegler-Nichols tuning rules are simple and intuitive. They require little process knowledge, and they can be applied with modest effort. The process is characterized by two parameters that can be determined by simple experiments. The frequency response method has the advantage that parameters K_u and T_u are easier to determine accurately than the parameters a and L, which are used by the step response method.

The methods are still widely used even if they give closed-loop systems that are not robust. The rules are often combined with manual tuning, which will be discussed in Section 6.3. The main drawbacks with the methods are that too little process information is used and the design criterion quarter amplitude damping gives closed-loop systems with poor robustness. It is not clear why this design criterion was used. The load disturbance responses look quite reasonable, but without analysis or sensitivity studies it is not obvious that the closed-loop systems are not robust. The simulations shown in Figure 6.2 and Figure 6.3 indicate that the methods give reasonable control. Repeated simulations with perturbations in controller parameters reveal very clearly that the closed-loop system is not robust. Systems like the ones shown in Examples 6.3 and 6.4 also illustrate that it is not sufficient to characterize the process by two parameters only.

A very large number of variations of the Ziegler-Nichols methods have been proposed. Here we have chosen to discuss two methods. The modifications of the Chien-Hrones-Reswick method give systems with somewhat better robustness, but it still uses too little process information. The Cohen-Coon method uses three parameters to characterize the process, but it still uses quarter amplitude damping as a design criterion.

In Chapter 7 we will develop new methods that address the major shortcomings of the Ziegler-Nichols methods while retaining their simplicity.

6.3 Rule-Based Empirical Tuning

Since the Ziegler-Nichols methods only give "ball-park" values, it is necessary to complement the methods by manual tuning to obtain reasonable closed-loop properties. Manual tuning is typically performed by experiments on the process in closed loop. A perturbation is introduced either as a set-point change or as a change in the control variable. The closed-loop response is observed, and the controller parameters are adjusted. The adjustments are based on simple rules, which give guidelines for changing the parameters. The rules were developed by extensive experimentation. The following is a simple set of rules:

- Increasing proportional gain decreases stability
- Error decays more rapidly if integration time is decreased
- Decreasing integration time decreases stability
- Increasing derivative time improves stability

Lately, the tuning rules have also been formalized in various types of formal rule-based systems such as expert systems or fuzzy logic.

Tuning maps are one way to express the tuning rules. The purpose of these maps is to provide intuition about how changes in controller parameters influence the behavior of the closed-loop system. The tuning maps are simply arrays of transient or frequency responses corresponding to systematic variations in controller parameters. An example of a tuning map is given in Figure 6.7.

The figure illustrates how the load disturbance response is influenced by changes in gain and integral time. The process model

$$P(s) = \frac{1}{(s+1)^8}$$

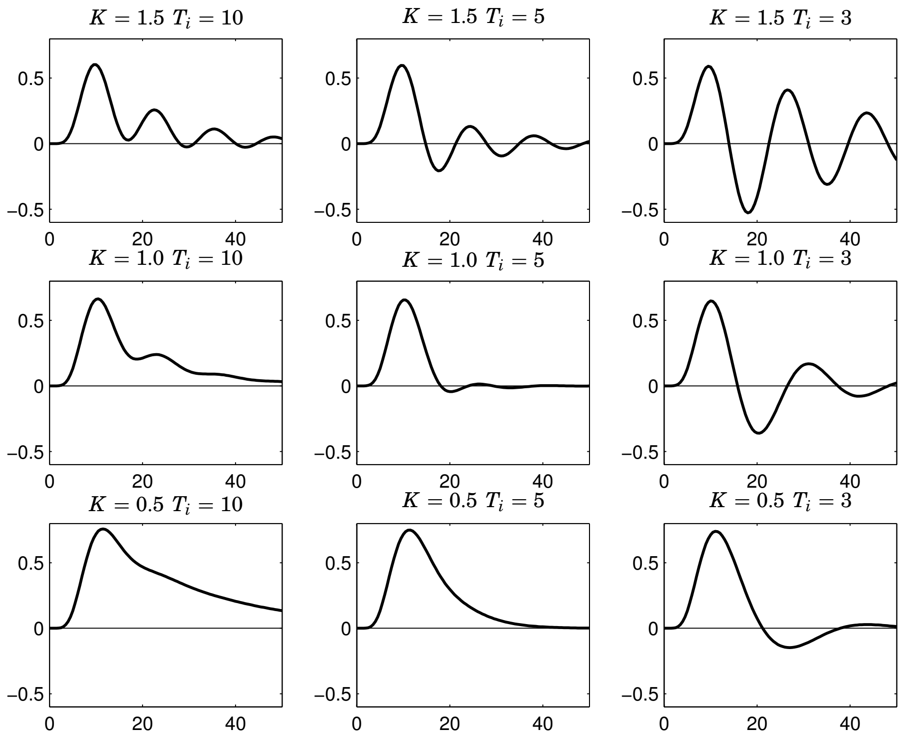

Figure 6.7 Tuning map for PID control of a process with the transfer function $P(s) = (s+1)^{-8}$. The figure shows the responses to a unit step disturbance at the process input. Parameter T_d has the value 1.9.

has been used in the example. The Ziegler-Nichols frequency response method gives the controller parameters $K = 1.13$, $T_i = 7.58$, and $T_d = 1.9$. The figure shows clearly the benefits of having a smaller value of T_i. Judging from the figure, the values $K = 1$ and $T_i = 5.0$ appear reasonable. The figure also shows that the choice of T_i is fairly critical. Also notice that controllers with $T_i < 7.6$ cannot be implemented on series form (compare with Section 3.4).

A different type of tuning map is shown in Figure 6.8, which shows the Nyquist curves of the loop transfer function. The figure shows that several of the Nyquist curves bend over too much to the right at low frequencies; see the figures in the left positions with $T_i = 10$. This means that the controller introduces too much phase lead. This is reduced by reducing parameter T_i.

A comparative study of curves like Figure 6.7 and Figure 6.8 is a good way to develop intuition for the relations between the time and frequency responses. An even better way is to use the interactive software that is now emerging.

Counter-Intuitive Behavior

Common rules for manual tuning says that the system becomes less oscillatory if the gain is reduced, if the integral time is increased, and if the derivative time is increased. Compare with Figure 6.4. These rules hold for the system

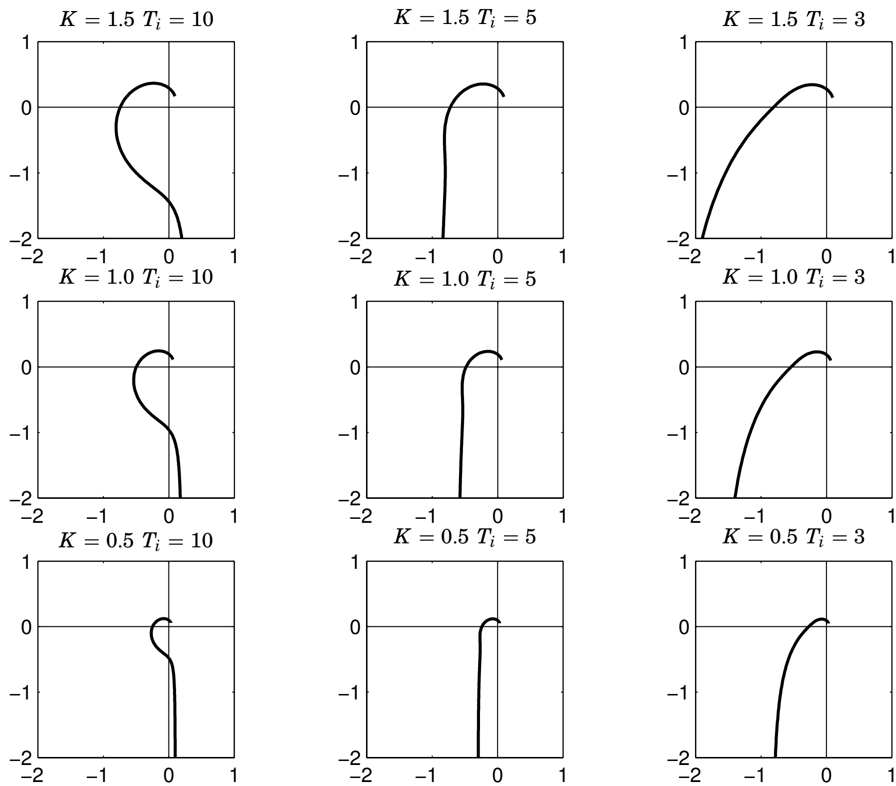

Figure 6.8 Tuning map for PID control of a process with the transfer function $P(s) = (s+1)^{-8}$. The figure shows the Nyquist plots of the loop transfer functions. Parameter T_d has the value 1.9.

shown in Figure 6.7 and Figure 6.8. There are, however, situations where these rules do not hold. The following is a simple common example.

Example 6.5—PI Control of an Integrator
Consider a process with the transfer function

$$P(s) = \frac{1}{s},$$

and a PI controller with the transfer function

$$C(s) = K(1 + \frac{1}{sT_i}).$$

The loop transfer function is

$$G_l(s) = P(s)C(s) = K\frac{1 + sT_i}{s^2 T_i};$$

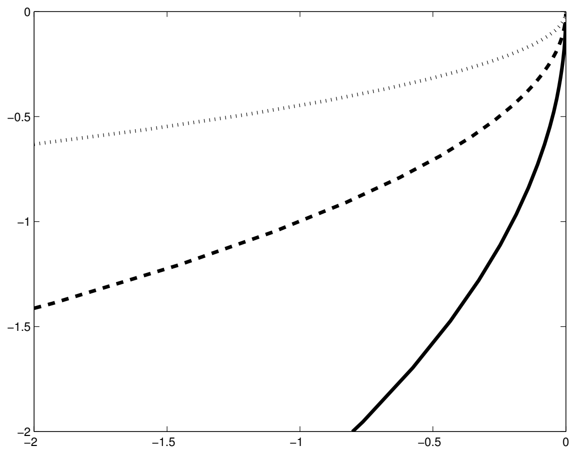

Figure 6.9 Nyquist curves for the loop transfer functions for an integrator with PI control. Integration time T_i is constant, and the gain has the values $K = 0.2$ (dotted), 1 (dashed), and 5 (solid). Notice the counterintuitive behavior that phase margin increases with increasing controller gain.

and the characteristic equation is

$$s^2 + Ks + \frac{K}{T_i} = 0.$$

Identifying this with a standard second-order system $s^2 + 2\zeta\omega_0 + \omega_0^2$ we find that

$$\zeta = \sqrt{\frac{K}{2T_i}}.$$

It follows from this equation that the damping increases when the controller gain is increased contrary to the intuition developed for the simple systems. This is also illustrated by the Nyquist curves in Figure 6.9. Notice that the Nyquist curve moves away from the critical point -1 as the gain increases. The reason for this is that the Nyquist curve is very close to the negative imaginary axis for large ω. Notice that a small time delay or a small lag will destroy this property. □

Situations like this make it difficult to form efficient rules that cover a wide range of conditions.

An Inequality for the Integration Time

It is useful to have a simple way to judge if the integral action of a controller is too weak, as in the three left and the lower middle examples in Figure 6.7 and Figure 6.8. Such a criterion can be based on a calculation of the asymptotic behavior of the loop transfer function for low frequencies. For a process with transfer function P and a PI controller with transfer function C we have

$$G_\ell(s) = P(s)C(s) \approx (P(0) + sP'(0))\, K\left(1 + \frac{1}{sT_i}\right)$$
$$= \frac{KP(0)}{sT_i} + KP(0) + \frac{KP'(0)}{T_i} + KP'(0)s.$$

Thus, for low frequencies the asymptote of the Nyquist curve is parallel to the imaginary axis with the real part equal to

$$KP(0) + \frac{KP'(0)}{T_i} = KK_p\left(1 - \frac{T_{ar}}{T_i}\right),$$

where $K_p = G(0)$ is the static process gain, and T_{ar} is the average residence time. It is reasonable to require that the real part of the asymptote be less than -0.5. This gives

$$T_i < T_{ar}\,\frac{2KK_p}{1 + 2KK_p} < T_{ar}. \tag{6.3}$$

For the system in Figure 6.7 and Figure 6.8, we get the requirement $T_i < 6.0$ for the systems in the upper row, $T_i < 5.3$ for the systems in the middle row, and $T_i < 4.0$ for the systems in the lower row. This means that condition (6.3) excludes the three left and the lower middle examples in Figure 6.7 and Figure 6.8.

The inequality for the integration time given by (6.3) can be used to give insight into the limitations of the Ziegler-Nichols rules for systems with large time delays. Consider a process with the transfer function

$$P(s) = K_p\,\frac{e^{-sL}}{1 + sT}.$$

For this system we have $T_{ar} = L + T$. Consider a PI controller tuned by the Ziegler-Nichols step response method. It follows from Table 6.1 that $KK_p = 0.9T/L$ and $T_i = 3L$. Equation 6.3 then gives

$$3L < (L + T)\frac{1.8T}{L + 1.8T},$$

which implies that $L < 0.38T$. This means that the Ziegler-Nichols step response method for PI control will not give good control unless the time delay is sufficiently small. Compare with Example 6.4.

Commentary

Manual tuning was used before any systematic tuning methods were available. It became a necessary complement to the Ziegler-Nichols method. It is essential for all practitioners of control to gain experience in judging the properties of closed-loop systems and to change controller parameters to modify the behavior. The assessment can be based on simple bump tests where set points or controller output is perturbed or by more elaborate frequency response measurements of the transfer function. It is necessary to be aware of the counter-intuitive behavior of processes with integral action illustrated in Example 6.5. The rule-based systems have been formalized when automatic tuners based on expert systems and fuzzy logic were developed. In Section 6.7 we will present systematic methods for improving the tuning based on optimization.

6.4 Pole Placement

Many properties of a closed-loop system are expressed by its poles. The idea with pole placement is to design a controller that gives a closed-loop system with desired closed-loop poles. The method requires a complete model of the process. Subject to some technical conditions it is possible to find a controller that gives the desired closed-loop poles, provided that the controller is sufficiently complex. To use the method for PID control it is necessary to restrict the complexity of the model by various approximation methods. The selected poles must then be chosen with care in order to ensure that the approximated model is valid for frequencies that correspond to the chosen poles.

A refinement of the procedure is to consider also the zeros of the transfer functions. This is particularly relevant for the set-point response. The zeros of the transfer function originating from the controller can be influenced by set-point weighting.

EXAMPLE 6.6—PI CONTROL OF A FIRST-ORDER SYSTEM
Suppose that the process can be described by the following first-order model

$$P(s) = \frac{K_p}{1+sT},$$

which has only two parameters, process gain K_p and time constant T. Let the process be controlled by a standard PI controller with set-point weighting,

$$\begin{aligned} C(s) &= K\big(1 + \frac{1}{sT_i}\big) \\ C_{ff}(s) &= K\big(b + \frac{1}{sT_i}\big). \end{aligned}$$

The closed-loop system is of second order. The loop transfer function is

$$G_\ell(s) = P(s)C(s) = \frac{K_pK(1+sT_i)}{sT_i(1+sT)} = \frac{K_pK(s+1/T_i)}{T(s+1/T)},$$

and the characteristic polynomial

$$s^2 + \frac{1+K_pK}{T}s + \frac{K_pK}{TT_i}. \tag{6.4}$$

The closed-loop system has two poles that can be given arbitrary values by a suitable choice of gain K and integral time T_i of the controller. Now suppose that the desired closed-loop poles are characterized by their relative damping ζ and their frequency ω_0. The desired characteristic polynomial then becomes

$$s^2 + 2\zeta\omega_0 s + \omega_0^2. \tag{6.5}$$

Identifying coefficients of equal powers of s in (6.4) and (6.5) we get

$$\begin{aligned} K &= \frac{2\zeta\omega_0 T - 1}{K_p} \\ T_i &= \frac{2\zeta\omega_0 T - 1}{\omega_0^2 T} \\ k_i &= \frac{K}{T_i} = \frac{\omega_0^2 T}{K_p}. \end{aligned} \tag{6.6}$$

It is convenient to use the parameters ω_0 and ζ as design parameters; ω_0 determines the response speed and ζ determines the shape of the response.

With controller parameters given by (6.6) the closed-loop system is characterized by the *Gang of six*, see Equation (4.2).

$$\begin{aligned} \frac{PC}{1+PC} &= \frac{(2\zeta\omega_0 - 1/T)s + \omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} & \frac{C}{1+PC} &= \frac{K(s+1/T_i)(s+1/T)}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \\ \frac{P}{1+PC} &= \frac{K_p s/T}{s^2 + 2\zeta\omega_0 s + \omega_0^2} & \frac{1}{1+PC} &= \frac{s(s+1/T)}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \\ \frac{PC_{ff}}{1+PC} &= \frac{b(2\zeta\omega_0 - 1/T)s + \omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} & \frac{C_{ff}}{1+PC} &= \frac{K(bs+1/T_i)(s+1/T)}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \end{aligned} \tag{6.7}$$

The largest value of the transfer function from a load disturbance at the process input to the process output is

$$\max_\omega |G_{xd}(i\omega)| = \max_\omega \left| \frac{P(i\omega)}{1 + P(i\omega)C(i\omega)} \right| = \frac{K_p}{\omega_0 T \min(1, \zeta)}.$$

To have good rejection of load disturbances it is thus desirable to choose ω_0 as large as possible. The largest value of ω_0 is limited by the magnitude of the control signals and the validity of the process model. The transfer function from measurement noise to the control signal has the magnitude K for high frequencies. If $K_{\max}$ is the largest permissible value of the controller gain it follows from (6.6) that

$$\omega_0 T < \frac{1 + K_p K_{\max}}{2\zeta}.$$

Let T_e be the sum of neglected time constants or time delays and using the rule of thumb that the phase error should be less than $\pm 15^\circ$ we find that ω_0 must be chosen so that $\omega_0 T_e < 0.25$. Compare with Section 2.8.

The frequency ω_0 chosen should not be too small. An indication of this is given by Equation 6.6, which shows that the proportional gain is negative if $2\zeta\omega_0 T < 1$. Further evidence is given in Figure 6.10, which shows Bode plots

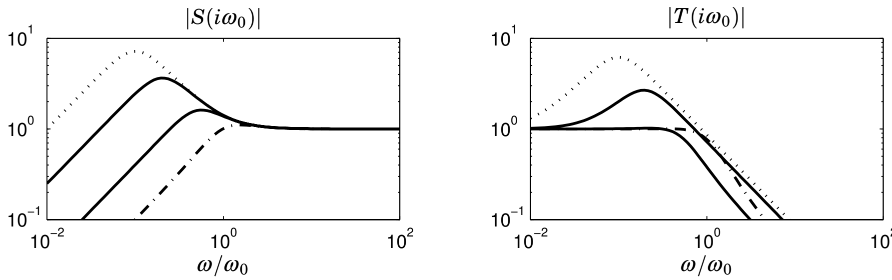

Figure 6.10 Gain curves of the sensitivity functions for $\zeta = 0.7$ and $\omega_0 T = 0.1, 0.2, 0.5,$ and 1. The dotted curve corresponds to $\omega_0 L = 0.1$ and the dash-dotted curve to $\omega_0 L = 1$.

of the gain curves of the sensitivity functions for different values of $\omega_0 T$. The figure shows that the sensitivities are large when $\omega_0 T$ is small. The maximum of the sensitivity function is approximately $M_s = 1/(2\zeta\omega_0 T)$. A reasonable choice of the parameter ω_0 is thus

$$\frac{1}{2\zeta} \le \omega_0 T < \min\Big(\frac{0.25}{T_e}, \frac{1 + K_p K_{max}}{2\zeta}\Big). \tag{6.8}$$

The lower limit corresponds to pure integral control; see (6.6).

It follows from (6.7) that the transfer function from set point to process output has a zero at $s = -1/(bT_i)$. To avoid excessive overshoot in the set-point response, parameter b should be chosen so that the zero is to the left of the dominant closed-loop poles. A reasonable value is $b = 1/(\omega_0 T_i)$, which places the zero at $s = -\omega_0$. This gives

$$b = \frac{1}{2\zeta - 1/(\omega_0 T)}.$$

It is particularly important to use a small value of b when $\omega_0 T$ is small and for unstable systems where T is negative. A response to set-point changes that does not have an overshoot is obtained by choosing $b = 0$ and $\zeta \ge 1$.

The reason why the sensitivities are large for small values of $\omega_0 T$ is that the characteristic polynomial (6.5) is a poor choice for designs where the closed-loop system is slower than the open-loop system. In such cases it is better to make a design that cancels the process pole and gives a closed-loop system with a time constant T_0. Such a controller has the parameters

$$\begin{aligned} K &= \frac{T}{K_p T_0} \\ T_i &= T, \end{aligned} \tag{6.9}$$

and it gives a closed-loop system with $M_s = M_t = 1$. The controller is not suitable when $\omega_0 T_0 > 1$ because it follows from (6.7) that the transfer function from load disturbances to process output is

$$\frac{P}{1 + PC} = \frac{sK_p T_0}{(1 + sT)(1 + sT_0)}.$$

The attenuation of load disturbances is thus poor for large values of T_0/T. □

EXAMPLE 6.7—PI CONTROL OF PROCESS WITH TWO REAL POLES
Assume that the process is characterized by the second-order model

$$P(s) = \frac{K_p}{(1+sT_1)(1+sT_2)},$$

and that a PI controller is used. The loop transfer function becomes

$$G_\ell(s) = P(s)C(s) = \frac{K_pK(1+sT_i)}{sT_i(1+sT_1)(1+sT_2)} = \frac{K_pK(s+1/T_i)}{T_1T_2(s+1/T_1)(s+1/T_2)},$$

and the characteristic polynomial becomes

$$s^3 + \left(\frac{1}{T_1} + \frac{1}{T_2}\right)s^2 + \frac{1+K_pK}{T_1T_2}s + \frac{K_pK}{T_1T_2T_i}. \tag{6.10}$$

The zeros of this third-order polynomial cannot be assigned arbitrary values since the controller only has two parameters. In particular, we find that the coefficient of s^2 is given by the time constants of the process. However, if we also consider the frequency ω_0 as a parameter it is possible to match the polynomial (6.10) to

$$(s+\alpha\omega_0)(s^2 + 2\zeta\omega_0 s + \omega_0^3).$$

Matching coefficients of equal powers of s we get

$$\begin{aligned}
\omega_0 &= \frac{T_1+T_2}{(\alpha+2\zeta)T_1T_2}\\
K &= \frac{(1+2\alpha\zeta)\omega_0^2T_1T_2 - 1}{K_p}\\
T_i &= \frac{K_pK}{\alpha\omega_0^3T_1T_2}\\
k_i &= \frac{\alpha\omega_0^3T_1T_2}{K_p}.
\end{aligned}$$

It is thus possible to obtain a design that gives a prescribed configuration of the poles with PI control, i.e., specified α and ζ. The parameter ω_0 is a scale factor that is determined by the process dynamics. □

EXAMPLE 6.8—PID CONTROL OF PROCESS WITH TWO REAL POLES
Suppose that the process is characterized by the second-order model

$$P(s) = \frac{K_p}{(1+sT_1)(1+sT_2)}.$$

This model has three parameters. By using a PID controller, which also has three parameters, it is possible to arbitrarily place the three poles of the closed-loop system. The transfer function of the PID controller can be written as

$$C(s) = \frac{K(1 + sT_i + s^2 T_i T_d)}{sT_i}.$$

The characteristic polynomial of the closed-loop system is

$$s^3 + s^2 \left(\frac{1}{T_i} + \frac{1}{T_2} + \frac{K_p K T_d}{T_1 T_2} \right) + s \left(\frac{1}{T_1 T_2} + \frac{K_p K}{T_1 T_2} \right) + \frac{K_p K}{T_1 T_2 T_i}. \tag{6.11}$$

A suitable closed-loop characteristic polynomial for a third-order system is

$$(s + \alpha\omega_0)(s^2 + 2\zeta\omega_0 s + \omega_0^2), \tag{6.12}$$

which contains two dominant poles with relative damping ζ and frequency ω_0, and a real pole located in $-\alpha\omega_0$. Identifying the coefficients of equal powers of s in Equations 6.11 and 6.12 gives

$$\begin{aligned} \frac{1}{T_i} + \frac{1}{T_2} + \frac{K_p K T_d}{T_1 T_2} &= \omega_0(\alpha + 2\zeta) \\ \frac{1}{T_1 T_2} + \frac{K_p K}{T_1 T_2} &= \omega_0^2(1 + 2\zeta\omega_0) \\ \frac{K_p K}{T_1 T_2 T_i} &= \alpha\omega_0^3. \end{aligned}$$

Solving these equations gives the following controller parameters:

$$\begin{aligned} K &= \frac{T_1 T_2 \omega_0^2 (1 + 2\alpha\zeta) - 1}{K_p} \\ T_i &= \frac{T_1 T_2 \omega_0^2 (1 + 2\alpha\zeta) - 1}{T_1 T_2 \alpha \omega_0^3} \\ T_d &= \frac{T_1 T_2 \omega_0 (\alpha + 2\zeta) - T_1 - T_2}{T_1 T_2 \omega_0^2 (1 + 2\alpha\zeta) - 1} \\ k_i &= \frac{\alpha\omega_0^3 T_1 T_2}{K_p}. \end{aligned}$$

Provided that $c = 0$, the transfer function from set point to process output has one zero at $s = -1/(bT_i)$. To avoid excessive overshoot in the set-point response, parameter b can be chosen so that this zero cancels the pole at $s = -\alpha\omega_0$. This gives

$$b = \frac{1}{\alpha\omega_0 T_i} = \frac{\omega_0^2 T_1 T_2}{\omega_0^2 T_1 T_2 (1 + 2\alpha\zeta) - 1}.$$

Also, notice that pure PI control is obtained for

$$\omega_0 = \omega_c = \frac{T_1 + T_2}{(\alpha + 2\zeta) T_1 T_2}.$$

The choice of ω_0 may be critical. The derivative time is negative for $\omega_0 < \omega_c$. Thus, frequency ω_c gives a lower bound to the bandwidth. The gain increases rapidly with ω_0. The upper bound to the bandwidth is given by the validity of the model. □

The General Case

Since there is a relation between the complexity of the model and the complexity of the controller it is natural to ask what is the most general model that will give PI and PID controllers. A PI controller has two parameters which are sufficient to characterize a second-order equation; this permits a process model of first order. The system in Example 6.6 is thus the most general system where pole placement will give a PI controller.

Since a PID controller has three parameters, it is possible to determine all parameters of a third-order equation. With PID control it is thus possible to use pole placement for a second-order system. The most general second-order system is not the one in Example 6.8, but the one in the next example.

If only a pattern of the pole is specified a PI controller suffices for a second-order system and a PID controller for a third-order system.

EXAMPLE 6.9—GENERAL SECOND-ORDER SYSTEM
Suppose that the process is characterized by the second-order model

$$P(s) = \frac{b_1 s + b_2}{s^2 + a_1 s + a_2}. \tag{6.13}$$

This model has four parameters. It has two poles that may be real or complex, and it has one zero. This model captures many processes, oscillatory systems, and systems with right half-plane zeros. The right half-plane zero can also be used as an approximation of a time delay. We assume that the process is controlled by a PID controller parameterized as

$$\begin{aligned} C(s) &= k + \frac{k_i}{s} + k_d s \\ C_{ff}(s) &= bk + \frac{k_i}{s} + ck_d s. \end{aligned}$$

The closed-loop system is of third order, and the characteristic polynomial is

$$s(s^2 + a_1 s + a_2) + (b_1 s + b_2)(k_d s^2 + ks + k_i).$$

A suitable closed-loop characteristic equation of a third-order system is

$$(s + \alpha\omega_0)(s^2 + 2\zeta\omega_0 s + \omega_0^2).$$

Equating coefficients of equal power in s in these equations gives the following equations:

$$\begin{aligned} a_1 + b_2 k_d + b_1 k &= (\alpha\omega_0 + 2\zeta\omega_0)(1 + b_1 k_d) \\ a_2 + b_2 k + b_1 k_i &= (1 + 2\alpha\zeta)\omega_0^2(1 + b_1 k_d) \\ b_2 k_i &= \alpha\omega_0^3(1 + b_1 k_d). \end{aligned}$$

This is a set of linear equations in the controller parameters. The solution is straightforward but tedious and is given by

$$\begin{aligned}
k &= \frac{a_2b_2^2 - a_2b_1b_2(\alpha+2\zeta)\omega_0 - (b_2-a_1b_1)(b_2(1+2\alpha\zeta)\omega_0^2+\alpha b_1\omega_0^3)}{b_2^3 - b_1b_2^2(\alpha+2\zeta)\omega_0 + b_1^2b_2(1+2\alpha\zeta)\omega_0^2 - \alpha b_1^3\omega_0^3}\\
k_i &= \frac{(-a_1b_1b_2 + a_2b_1^2 + b_2^2)\alpha\omega_0^3}{b_2^3 - b_1b_2^2(\alpha+2\zeta)\omega_0 + b_1^2b_2(1+2\alpha\zeta)\omega_0^2 - \alpha b_1^3\omega_0^3} \qquad (6.14)\\
k_d &= \frac{-a_1b_2^2 + a_2b_1b_2 + b_2^2(\alpha+2\zeta)\omega_0 - b_1b_2\omega_0^2(1+2\alpha\zeta) + b_1^2\alpha\omega_0^3}{b_2^3 - b_1b_2^2(\alpha+2\zeta)\omega_0 + b_1^2b_2(1+2\alpha\zeta)\omega_0^2 - \alpha b_1^3\omega_0^3}.
\end{aligned}$$

These formulas are quite useful because many processes can be approximately described by the transfer function given by (6.13).

The transfer function from set point to process output is

$$G_{yy_{sp}}(s) = \frac{(b_1s+b_2)(ck_ds^2+bks+k_i)}{(s+\alpha\omega_0)(s^2+2\zeta\omega_0s+\omega_0^2)}.$$

The parameters b and c have a strong influence on the response of this transfer function. □

The formulas given in Example 6.9 are particularly useful in cases when we are "stretching" the PID controller to extreme situations. The standard tuning rules will typically not work in these cases. Typical examples are systems with zeros in the right half-plane and systems with poorly damped oscillatory modes. To illustrate this we will consider an example.

EXAMPLE 6.10—OSCILLATORY SYSTEM WITH RHP ZERO

Consider a system with the transfer function

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

This system has one right half-plane zero and two undamped complex poles. The process is difficult to control. To provide damping for the undamped poles at $s = \pm i$ it is necessary to have a reasonable control gain at $\omega = 1$. This is difficult because the right-half plane zero at $s = 1$ implies that the gain crossover frequencies should be less than 0.5 in order to have a reasonably robust closed-loop system. None of the standard methods for tuning PID controllers work well for this system. To apply the pole placement method we specify that the closed-loop system has the characteristic polynomial

$$s^3 + 2s^2 + 2s + 1.$$

The formulas in Example 6.9 give a controller with the parameters $k = 0$, $k_i = 1/3$, and $k_d = 2/3$. This can also be verified with a simple calculation. Notice that the proportional gain is zero and that the controller has two complex zeros at $\pm i\sqrt{2}$. Such a controller can only be implemented with a PID controller having the non-interacting form. Compare with Section 3.2. □

Using Approximate Models

Since pole placement will only give PID controllers if the process model is of second order or less it is necessary to develop approximate models in order to use pole placement. Different approximation methods were discussed in Section 2.8. In this section we will illustrate the method with a few examples.

Consider a process described by the transfer function

$$P(s) = \frac{1}{(1+s)(1+0.2s)(1+0.05s)(1+0.01s)}. \tag{6.15}$$

This process has four lags with time constants 1, 0.2, 0.05, and 0.01. The approximations can be done in several different ways.

EXAMPLE 6.11—APPROXIMATION WITH A FIRST-ORDER SYSTEM

If the control requirements are not too severe, we can attempt to approximate the transfer function by

$$P(s) = \frac{1}{1+1.26s},$$

where the time constant is the average residence time of the system. As discussed in Section 2.8, this approximation is good at low frequencies. The sum of the neglected time constants is $T_e = 0.26$. The phase error is less than 15° for frequencies below 1 rad/s. Designing a PI controller with the pole placement method with $\zeta = 0.5$, the following controller parameters are obtained,

$$\begin{aligned} K &= 1.26\omega_0 - 1 \\ T_i &= \frac{1.26\omega_0 - 1}{1.26\omega_0^2} \\ b &= \frac{1.26\omega_0}{1.26\omega_0 - 1}. \end{aligned}$$

where b is chosen so that the zero becomes $s = -\omega_0$. If the process model would be correct, the phase margin with $\zeta = 0.5$ would be 50°. Because of the approximations made, the phase margin will be less. It will decrease with ω_0. For $\omega_0 = 1$ the phase margin is $\varphi_m = 42°$. The closed-loop poles for the system are $-100, -20, -4.99, -0.46 \pm 1.02i$. The closed-loop poles obtained when the controller is applied to the simplified model are $-0.5 \pm 0.87i$. Because of the approximation the dominant poles differ from the design values. The difference increases with increasing ω_0. The system becomes unstable for $\omega_0 = 3.8255$.

Figure 6.11 shows the sensitivity functions for the approximate and the exact system. The maximum sensitivities are $M_t = 1.35$ and $M_s = 1.66$, respectively. This indicates that the closed-loop poles must be chosen with care when using pole placement. □

The next example shows what happens when the system is approximated with a second-order model.

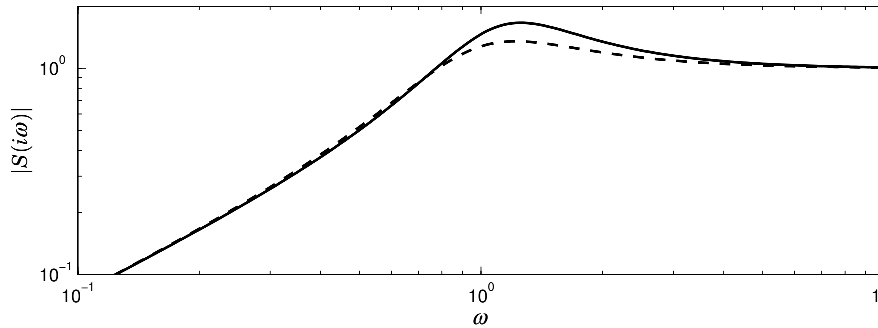

Figure 6.11 Sensitivity functions for the approximate system (dashed) and the true system in Example 6.11.

EXAMPLE 6.12—APPROXIMATION WITH A SECOND-ORDER MODEL
Consider the system given by (6.15). Approximate the transfer function by

$$P(s) = \frac{1}{(1+s)(1+0.26s)}.$$

It is obtained by keeping the longest time constant and approximating the three shorter time constants with their sum. The sum of the neglected time constants is $T_e = 0.06$. The phase error is less than 15° for frequencies below 4.4 rad/s. By making an approximation of the process model that is valid for higher frequencies than in the previous example, we can thus design a faster controller. If $\zeta = 0.5$ and $\alpha = 1$ are chosen in (6.16), the design calculations in Example 6.12 give the following PID parameters:

$$\begin{aligned} K &= 0.52\omega_0^2 - 1 \\ T_i &= \frac{0.52\omega_0^2 - 1}{0.26\omega_0^3} \\ T_d &= \frac{0.52\omega_0 - 1.26}{0.52\omega_0^2 - 1} \\ b &= \frac{0.26\omega_0^2}{0.52\omega_0^2 - 1}. \end{aligned} \tag{6.16}$$

In this case, pure PI control is obtained for $\omega_0 = 2.4$. The derivative gain becomes negative for lower bandwidths. The approximation neglects the time constant 0.05. If the neglected dynamics are required to give a phase error of, at most, 0.3 rad (17 deg) at the bandwidth, $\omega_0 < 6$ rad/s can be obtained. In Figure 6.12, the behavior of the control is demonstrated for $\omega_0 = 4, 5$, and 6.

The specification of the desired closed-loop bandwidth is crucial, since the controller gain increases rapidly with the specified bandwidth. It is also crucial to know the frequency range where the model is valid. Alternatively, an upper bound to the controller gain can be used to limit the bandwidth. Notice the effect of changing the design frequency ω_0. The system with $\omega_0 = 6$ responds

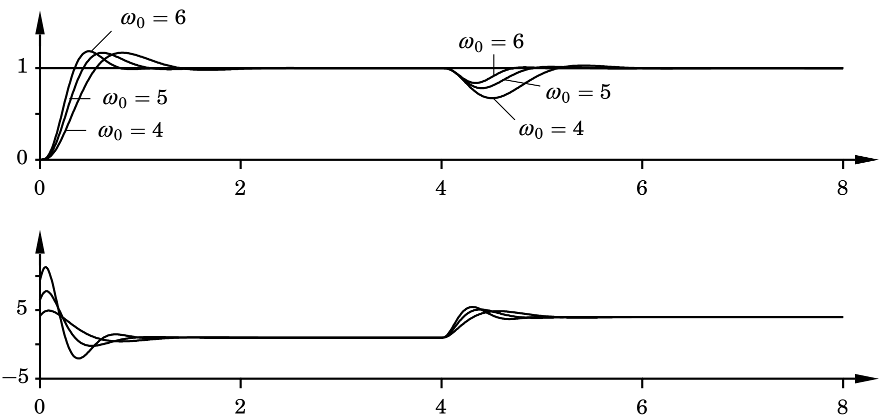

Figure 6.12 Set-point and load disturbance responses of the process with two poles controlled by a PID controller tuned according to Example 6.12. The responses for $\omega_0 = 4, 5$, and 6 are shown. The upper diagram shows set point $y_{sp} = 1$ and process output y, and the lower diagram shows control signal u.

faster and has a smaller error when subjected to load disturbances. The design will not work well when ω_0 is increased above 8. □

Dominant Pole Design

In pole placement design it is attempted to assign all closed-loop poles. One drawback with the method is that it is difficult to specify many closed-loop poles. In Section 4.5 it was mentioned that the behavior of a system can often be characterized with a few dominant poles. It can therefore be attempted to place a few dominant poles. We will illustrate this with a few examples.

EXAMPLE 6.13—AN INTEGRATING CONTROLLER
Consider a process with the transfer function $P(s)$ and an integrating controller

$$C(s) = \frac{k_i}{s}.$$

The closed-loop poles are given by

$$1 + k_i \frac{P(s)}{s} = 0.$$

Since the controller has one adjustable parameter, it is possible to assign one pole. To obtain a pole at $s = -a$ the controller parameter should be chosen as

$$k_i = \frac{a}{P(-a)}. \tag{6.17}$$

To obtain a good attenuation of load disturbances we will choose the closed-loop pole so that the integral gain k_i is as large as possible. For example, if

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

we get

$$k_i = a(-a+1)^2 = a^3 - 2a^2 + a,$$

which has its maximum $4/27$ for $a = 1/3$. □

EXAMPLE 6.14—PI CONTROL

A PI controller has two parameters. Consequently, it is necessary to assign two poles. Consider a process with transfer function $P(s)$, and let the controller be parameterized as

$$C(s) = k + \frac{k_i}{s}.$$

The closed-loop characteristic equation is

$$1 + \left(k + \frac{k_i}{s}\right) P(s) = 0.$$

Require that this equation has roots at

$$p_{1,2} = \omega_0 \left(-\zeta_0 \pm i\sqrt{1-\zeta_0^2}\right) = \omega_0 e^{i(\pi \pm \gamma)} = \omega_0(-\cos\gamma \pm i \sin\gamma),$$

where $\gamma = \arccos \zeta_0$. The condition that the closed-loop system has a pole p_1 is thus

$$1 + \left(k + \frac{k_i}{p_1}\right) P(p_1) = 0. \tag{6.18}$$

This is a linear equation in complex variables with two unknown variables. To solve it we introduce $a(\omega_0)$ and $\phi(\omega_0)$, defined as

$$P\left(\omega_0 e^{i(\pi-\gamma)}\right) = a(\omega_0) e^{i\phi(\omega_0)}.$$

Notice that $P\left(\omega_0 e^{i(\pi-\gamma)}\right)$ represents the values of the transfer function on the ray $e^{i(\pi-\gamma)}$. When $\gamma = \pi/2$, then $P\left(\omega_0 e^{i(\pi-\gamma)}\right) = P(i\omega_0)$, which is the normal frequency response.

Equation 6.18 can be written as

$$1 + \left(k + \frac{k_i}{\omega_0 e^{i(\pi-\gamma)}}\right) a(\omega_0) e^{i\phi(\omega_0)} = 0.$$

This equation, which is linear in k and k_i, has the solution

$$\begin{aligned} k &= -\frac{\sin(\phi(\omega_0)+\gamma)}{a(\omega_0)\sin\gamma} \\ k_i &= -\frac{\omega_0 \sin\phi(\omega_0)}{a(\omega_0)\sin\gamma}. \end{aligned} \tag{6.19}$$

Notice that $\phi(\omega_0)$ is zero for $\omega_0 = 0$ and typically negative as ω_0 increases. This implies that the proportional gain is negative and the integral gain positive but small for small ω_0. When ω_0 increases both k and k_i will increase initially. For larger values of ω_0 both parameters will decrease. Requiring that both parameters are positive, we find that ω_0 must be selected so that

$$\gamma < -\phi(\omega_0) < \pi.$$

The integral time of the controller is

$$T_i = \frac{k}{k_i} = \frac{\sin(\phi(\omega_0) + \gamma)}{\omega_0 \sin \phi(\omega_0)}.$$

Notice that T_i is independent of $a(\omega_0)$. □

EXAMPLE 6.15—A PURE DEAD-TIME PROCESS
Consider a process with the transfer function

$$P(s) = e^{-sL}.$$

Using pure integral control, it follows from Equation 6.17 that $k_i = ae^{-aL}$. The gain has its largest value $k_i = e^{-1}/L$ for $a = 1/L$. The loop transfer function for the system is then

$$G_l(s) = P(s)C(s) = \frac{1}{esL}\, e^{-sL}.$$

The sensitivity of the system is $M_s = 1.39$, which is a reasonable value.

With PI control it follows from Equation 6.19 that

$$\begin{aligned} k &= \frac{\sin(\omega_0 L \sin\gamma - \gamma)}{\sin\gamma}\, e^{-\omega_0 L \cos\gamma} \\ k_i &= \omega_0 \frac{\sin(\omega_0 L \sin\gamma)}{\sin\gamma}\, e^{-\omega_0 L \cos\gamma}. \end{aligned}$$

To minimize IE, we determine the value of ω_0 that maximizes k_i. The results are given in Table 6.6. This table also gives the M_s values and the IAE. The IAE has minimum at $\zeta \approx 0.6$. Notice that there are significant variations in the gain but that the values of integration time are fairly constant for all values of the design parameter ζ. The value of IAE should be small to give good rejection of load disturbances, and M_s should be small to give good robustness. The table illustrates the trade-offs between these goals. To obtain a reasonable robustness of $M_s < 2$, the relative damping should be greater than 0.5.

Notice that for $\zeta = 1$ we get $k = e^{-2}$ and $k_i = 4e^{-2}/L$. This can be compared with $k_i = e^{-1}L$ for pure I control. With PI control the integral gain can thus be increased by a factor of 1.5 compared with an I controller. For a well-damped system ($\zeta = 0.707$) the gain is about 0.2 and the integral time is $T_i = 0.28L$. This can be compared with the values 0.45 and 2L obtained with the Ziegler-Nichols frequency response method. The dominant pole design thus gives a controller with much stronger integral action than the Ziegler-Nichols method. In Example 6.4 we found that this was highly desirable.

Table 6.6 Controller parameters for dominant pole design of a PI controller for a pure time delay process.

ζ	k	$k_i L$	T_i/L	$\omega_0 L$	M_s	IAE/L
0.1	0.388	1.50	0.258	1.97	6.34	4.03
0.2	0.343	1.27	0.270	1.93	3.60	2.42
0.5	0.244	0.847	0.288	1.86	1.99	1.56
0.707	0.195	0.688	0.284	1.88	1.69	1.54
1.0	0.135	0.541	0.250	2.00	1.49	1.85

In summary, we find that a process with a pure delay dynamics can be controlled quite well with a PI controller. □

Dominant pole design is a special case of pole placement where it is only attempted to place a few dominant poles. For pure P, I, or D controllers one pole can be placed. For PI and PD controllers there are two dominant poles, which can conveniently be parameterized with the relative damping ζ. The method becomes more complicated for PID control. After the design it is necessary to check that the closed-loop poles obtained are actually dominating. It is also necessary to evaluate the robustness of the closed-loop system.

Commentary

Pole placement is a standard method for control system design. The specifications are given in terms of all poles of the closed-loop system or possibly only the pole pattern. Good judgment is required to choose the poles properly. When using pole placement the complexity of the controller is determined by the complexity of the process model. To obtain a PID controller it is required that the model is of low order or that the model is approximated by a low-order model. Time delay are often approximated when using pole placement. There is no natural way to introduce a robustness constraint in pole placement. The resulting closed-loop system must be analyzed to ensure that it is sufficiently robust.

6.5 Lambda Tuning

Lambda tuning is a special case of pole placement that is commonly used in the process industry. The process is modeled by the FOTD model

$$P(s) = \frac{K_p}{1+sT} e^{-sL}.$$

Different approximations of time delay L result in both PI and PID controllers.

PI Control

If a PI controller with the transfer function

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

is used with integral time T_i chosen equal to the time constant T of the process, the loop transfer function becomes

$$G_l(s) = P(s)C(s) = \frac{K_pK}{sT}e^{-sL} \approx \frac{K_pK(1 - sL)}{sT},$$

where the exponential function has been approximated using a Taylor series expansion. The characteristic equation of the closed-loop system is

$$s(T - K_pKL) + K_pK = 0.$$

Requiring that the closed-loop pole is $s = -1/T_{cl}$, where T_{cl} is the desired closed-loop time constant, we find

$$K_pK = \frac{T}{L + T_{cl}},$$

which gives the following simple tuning rule

$$\begin{aligned} K &= \frac{1}{K_p}\frac{T}{L + T_{cl}} \\ T_i &= T. \end{aligned} \tag{6.20}$$

The closed loop response time T_{cl} is the design parameter. In the original work by Dahlin [Dahlin, 1968] it was denoted as $T_{cl} = \lambda$, which explains the name lambda tuning.

The choice of T_{cl} is critical. A common rule of thumb is to choose $T_{cl} = 3T$ for a robust controller and $T_{cl} = T$ for aggressive tuning when the process parameters are well determined. Both choices lead to controllers with zero gain and zero integral time for pure time delay systems. For delay-dominated processes it is therefore sometimes recommended to choose T_i as the largest of the values T and $3L$.

A drawback with lambda tuning is that the process pole is canceled. This is not serious if for delay dominated processes. The integral gain is

$$k_i = \frac{K}{T_i} = \frac{1}{K_p(L + T_{cl})}.$$

When T_{cl} is proportional to T integral gain is thus small for large T. The response to load disturbances is thus very poor for lag-dominated processes.

For lag-dominated processes it is therefore useful to make a design that does not cancel the process pole. When the FOTD process is controlled with a PI controller the loop transfer function is

$$G_l(s) = P(s)C(s) = \frac{K_pK(1 + sT_i)e^{-sL}}{sT_i(1 + sT)} \approx \frac{K_pK(1 + sT_i)(1 - sL)}{sT_i(1 + sT)},$$

where the exponential function has been approximated by a Taylor series expansion. The characteristic equation is of second order:

$$s^2\Big(\frac{T_iT}{K_pK} - T_iL\Big) + s\Big(T_i + \frac{T_i}{K_pK} - L\Big) + 1 = 0.$$

Comparing this with the desired characteristic equation,

$$s^2T_{cl}^2 + 2\zeta T_{cl}s + 1 = 0,$$

gives the controller parameters

$$\begin{aligned} K &= \frac{L + 2\zeta T_{cl}}{T_{cl}^2 + T_{cl}^2/(K_pK) + 2\zeta T_{cl}L + L^2} \\ T_i &= \frac{K_pK(L + 2\zeta T_{cl})}{1 + K_pK}. \end{aligned} \tag{6.21}$$

These tuning rules can also be applied to integrating process provided that T_{cl} is chosen properly. For lag-dominated processes it is reasonable to choose T_{cl} proportional to L.

PID Control

For the derivation of the PID design, the interacting form of the PID controller (3.8) is used:

$$C'(s) = K'\frac{(1 + sT_i')(1 + sT_d')}{sT_i'}.$$

The time delay is approximated using (2.59), which gives the process transfer function

$$P(s) = \frac{K_p}{1 + sT}e^{-sL} \approx \frac{K_p(1 - sL/2)}{(1 + sT)(1 + sL/2)}.$$

The integral time is chosen to $T_i' = T$ and the derivative time to $T_d' = L/2$. The zeros of the controller will then cancel the poles of the process, and the loop transfer function becomes

$$G_l(s) = P(s)C'(s) \approx \frac{K_pK'(1 - sL/2)}{sT}.$$

The characteristic equation is

$$s(T - K_pK'L/2) + K_pK' = 0.$$

Requiring that the closed-loop pole is $s = -1/T_{cl}$ we find

$$K_pK' = \frac{T}{L/2 + T_{cl}},$$

which gives the following simple tuning rules:

$$
\begin{aligned}
K' &= \frac{1}{K_p}\frac{T}{L/2 + T_{cl}}\\
T_i' &= T\\
T_d' &= \frac{L}{2}.
\end{aligned}
$$

Using (3.9), the corresponding parameters for the noninteracting PID controller becomes

$$
\begin{aligned}
K &= \frac{1}{K_p}\frac{L/2 + T}{L/2 + T_{cl}}\\
T_i &= T + L/2\\
T_d &= \frac{TL}{L + 2T}.
\end{aligned}
\tag{6.22}
$$

Notice that there is no derivative action for pure delay processes ($T = 0$).

Commentary

Lambda tuning is a special case of pole placement. It is a simple method that can give good results in certain circumstances provided that the design parameter is chosen properly. The basic method cancels a process pole which will lead to poor response to load disturbances for lag-dominated processes. Various ad hoc fixes can be made, but this requires insight.

6.6 Algebraic Design

There are several algebraic tuning methods where the controller transfer function is obtained from the specifications by a direct algebraic calculation. The methods are closely related to pole placement.

Standard Forms

A fundamental question is to determine transfer functions that give suitable responses to set-point changes. This can be done by starting with a transfer function of a given form and determining the parameters so that some error criterion such as IAE, ISE, or ITAE is minimized.

Typical examples are

$$
\begin{aligned}
G_1 &= \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 + \omega_0^2}\\
G_2 &= \frac{\alpha\omega_0^3}{(s^2 + 2\zeta\omega_0 s + \omega_0^2)(s + \alpha\omega_0)}\\
G_3 &= \frac{\omega_0(s + \beta\omega_0)}{\beta(s^2 + 2\zeta\omega_0 s + \omega_0^2)}\\
G_4 &= \frac{\alpha\omega_0^2(s + \beta\omega_0)}{\beta(s^2 + 2\zeta\omega_0 s + \omega_0^2)(s + \alpha\omega_0)}.
\end{aligned}
\tag{6.23}
$$

The parameter ω_0 is a scale factor that determines the response speed. Parameters α, β, and ζ determine the shape of the transfer functions. Relative damping ζ is typically in the range of 0.5 to 1. The parameters α and β have a significant influence if they are less than one. Decreasing α makes the response slower and reduces the overshoot. Decreasing β makes the response faster and increases the overshoot. There have been many efforts to find parameters that optimize various criteria. Consider a system where the process has transfer function $P(s)$ and the controller transfer functions are

$$C(s) = K\left(1 + \frac{1}{sT_i} + sT_d\right)$$
$$C_{ff}(s) = K\left(b + \frac{1}{sT_i} + scT_d\right).$$

The closed-loop transfer function from set point to process output is then

$$G_{yy_{sp}} = \frac{PC_{ff}}{1+PC}.$$

The controller parameters K, T_i, and T_d are first chosen to match the denominator of the specified transfer function, and the set-point weights b and c are then chosen to match the numerator of the specified transfer function. Since the simple controllers only have a few parameters it is necessary that the chosen transfer functions be sufficiently simple.

For systems with error feedback where $C(s) = C_{ff}(s)$ it is possible to give an explicit expression for the controller transfer function:

$$C = \frac{1}{P} \cdot \frac{G_{yy_{sp}}}{1 - G_{yy_{sp}}}. \tag{6.24}$$

To make sure that the controller obtained is a PID controller it is necessary to make approximations or cancellations as was discussed in Section 2.8.

It follows from (6.24) that all process poles and zeros are canceled by the controller unless $G_{yy_{sp}}$ has corresponding poles and zeros. This means that error feedback cannot be applied when the process has poorly damped poles and zeros. The method will also give a poor load disturbance response when slow process poles are canceled.

There are many different versions of algebraic design methods. Let it suffice to present a few cases.

Haalman's Method

For systems with a time delay L, Haalman has suggested choosing the loop transfer function

$$G_l(s) = P(s)C(s) = \frac{2}{3Ls}\, e^{-sL}.$$

The value $2/3$ was found by minimizing the mean square error for a step change in the set point. This choice gives a sensitivity $M_s = 1.9$, which is a reasonable value. Notice that it is only the dead time of the process that influences the loop transfer function. All other process poles and zeros are canceled, which may lead to difficulties.

Applying Haalman's method to a process with the transfer function

$$P(s) = \frac{K_p}{1+sT}\, e^{-sL}$$

gives the controller

$$C(s) = \frac{2(1+sT)}{3K_pLs} = \frac{2T}{3K_pL}\left(1+\frac{1}{sT}\right),$$

which is a PI controller with $K = 2T/3K_pL$ and $T_i = T$. These parameters can be compared with the values $K = 0.9T/L$ and $T_i = 3L$ obtained by the Ziegler-Nichols step response method.

Comparing Haalman's method with lambda tuning we find that the integral times are the same and that the gains are the same if we choose $T_{cl} = L/2$. Since lambda tuning is based on approximations of the time delay it appears more reasonable to use Haalman's method when the time delay L is large.

Applying Haalman's method to a process with the transfer function

$$P(s) = \frac{K_p}{(1+sT_1)(1+sT_2)}\, e^{-sL}$$

gives a PID controller with parameters $K = 2(T_1+T_2)/3K_pL$, $T_i = T_1+T_2$, and $T_d = T_1T_2/(T_1+T_2)$. For more complex processes it is necessary to approximate the processes to obtain a transfer function of the desired form as was discussed in Section 2.8.

Figure 6.13 shows a simulation of Haalman's method for a system with normalized dead time $\tau = 0.5$. The figure shows that the responses are good.

Dangers of Cancellation of Slow Process Poles

A key feature of Haalman's method is that process poles and zeros are canceled by poles and zeros in the controller. When poles and zeros are canceled, there will be uncontrollable modes in the closed-loop system. This may lead to poor performance if the modes are excited. The problem is particularly severe if the canceled modes are slow or unstable. We use an example to illustrate what may happen.

EXAMPLE 6.16—LOSS OF CONTROLLABILITY DUE TO CANCELLATION

Consider a closed-loop system where a process with the transfer function

$$P(s) = \frac{1}{1+sT}\, e^{-sL}$$

is controlled with a PI controller whose parameters are chosen so that the process pole is canceled. The transfer function of the controller is then

$$C(s) = K\left(1+\frac{1}{sT}\right) = K\,\frac{1+sT}{sT}.$$

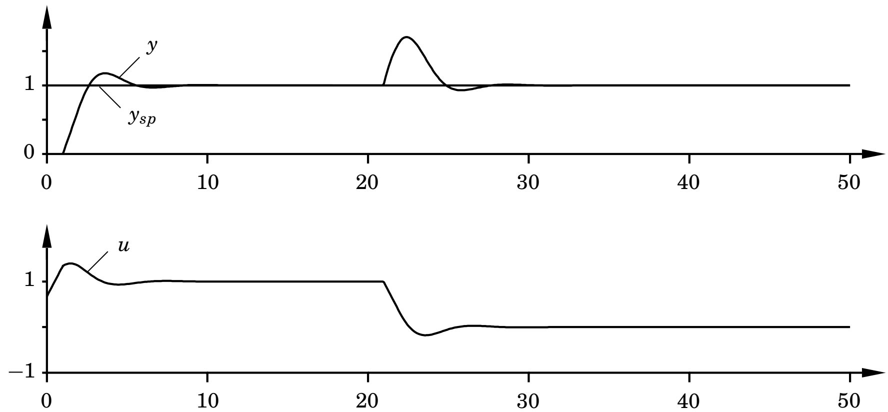

Figure 6.13 Simulation of a closed-loop system obtained by Haalman's method. The plant transfer function is $P(s) = e^{-s}/(s+1)$. The diagrams show set point y_{sp}, process output y, and control signal u.

The process can be represented by the equation

$$\frac{dy(t)}{dt} = \frac{1}{T}\left(u(t-L) - y(t)\right), \tag{6.25}$$

and the controller can be described by

$$\frac{du(t)}{dt} = -K\left(\frac{dy(t)}{dt} + \frac{y(t)}{T}\right). \tag{6.26}$$

Consider the behavior of the closed-loop system when the initial conditions are chosen as $y(0) = 1$ and $u(t) = 0$ for $-L < t < 0$. Without feedback the output is given by

$$y_{ol}(t) = e^{-t/T}.$$

To compute the output for the closed-loop system we first eliminate $y(t)$ between (6.25) and (6.26). This gives

$$\frac{du(t)}{dt} = -\frac{K}{T}\,u(t-L).$$

It thus follows that $u(t) = 0$, and (6.25) then implies that

$$y_{cl}(t) = e^{-t/T} = y_{ol}(t).$$

The trajectories of the closed-loop system and the open-loop system thus are the same. The control signal is zero, which means that the controller does not attempt to reduce the control error. □

The example clearly indicates that there are drawbacks with cancellation of process poles. Another illustration of the phenomenon is given in Figure 6.14,

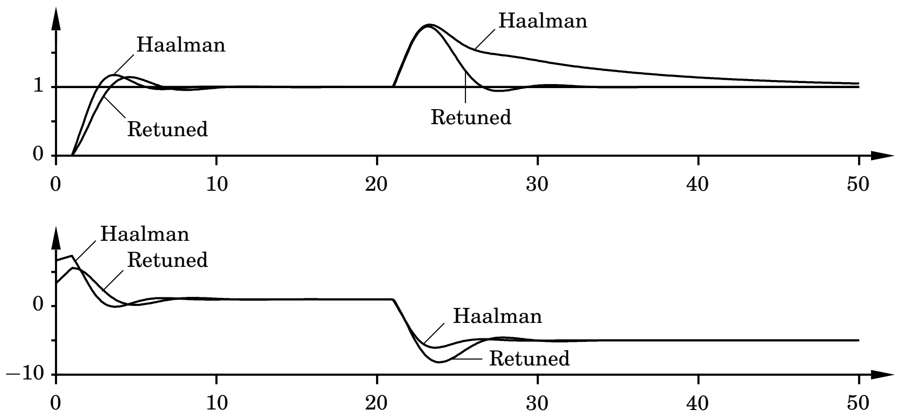

Figure 6.14 Simulation of a closed-loop system obtained by Haalman's method. The process transfer function is $P(s) = e^{-s}/(10s + 1)$, and the controller parameters are $K = 6.67$ and $T_i = 10$. The upper diagram shows set point $y_{sp} = 1$ and process output y, and the lower diagram shows control signal u. The figure also shows the responses to a re-tuned controller with $K = 6.67$, $T_i = 3$, and $b = 0.5$.

which is a simulation of a closed-loop system where the controller is designed by Haalman's method. This simulation is identical to the simulation in Figure 6.13, but the process time constant is now 10 instead of 1 for the simulation in Figure 6.14.

In this case we find that the set-point response is excellent but that the response to load disturbances is very poor. The reason for this is that the controller cancels the pole $s = -0.1$ by having a controller zero at $s = -0.1$. Notice that the process output after a load disturbance decays with the time constant $T = 10$ but that the control signal is practically constant due to the cancellation. The attenuation of load disturbances is improved considerably by reducing the integral time of the controller as shown in Figure 6.14.

We have thus shown that cancellation of process poles may give systems with poor rejection of load disturbances. Notice that this does not show up in simulations unless the process is excited. For example, it will not be noticed in a simulation of a step change in the set point. We may also ask why there is such a big difference in the simulations in Figures 6.13 and 6.14. The reason is that the canceled pole in Figure 6.14 is slow in comparison with the closed-loop poles, but it is of the same magnitude as the closed-loop poles in Figure 6.13.

We can thus conclude that pole cancellation can be done for systems that are dead-time dominated but not for systems that are lag dominated.

Internal Model Control (IMC)

The internal model principle is a general method for design of control systems that can be applied to PID control. A block diagram of such a system is shown in Figure 6.15. In the diagram it is assumed that all disturbances acting on the process are reduced to an equivalent disturbance d at the process output. In the figure $\hat{P}$ denotes a model of the process, $\hat{P}^{\dagger}$ is an approximate inverse

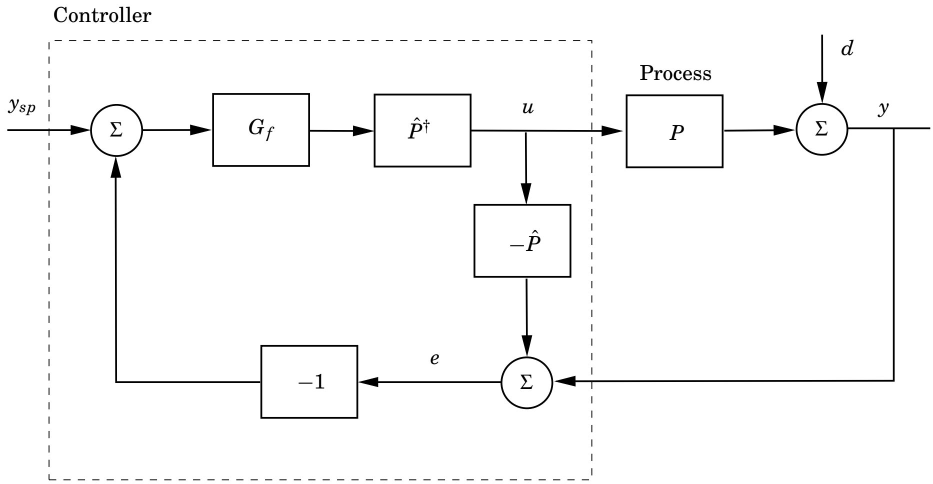

Figure 6.15 Block diagram of a closed-loop system with a controller based on the internal model principle.

of $\hat{P}$, and G_f is a low-pass filter. The name *internal model controller* derives from the fact that the controller contains a model of the process internally. This model is connected in parallel with the process.

If the model matches the process, i.e., $\hat{P} = P$, the signal e is equal to the disturbance d for all control signals u. If $G_f = 1$ and $\hat{P}^\dagger$ is an exact inverse of the process, then the disturbance d will be canceled perfectly. The filter G_f is introduced to obtain a system that is less sensitive to modeling errors. A common choice is $G_f(s) = 1/(1 + sT_f)$, where T_f is a design parameter.

The controller obtained by the internal model principle can be represented as an ordinary series controller with the transfer function

$$C = \frac{G_f \hat{P}^\dagger}{1 - G_f \hat{P}^\dagger \hat{P}}. \tag{6.27}$$

From this expression it follows that controllers of this type cancel process poles and zeros.

The internal model principle will typically give controllers of high order. By making special assumptions it is, however, possible to obtain PI or PID controllers from the principle. To see this consider a process with the transfer function

$$P(s) = \frac{K_p}{1 + sT} e^{-sL}. \tag{6.28}$$

An approximate inverse is given by

$$\hat{P}^\dagger(s) = \frac{1 + sT}{K_p}.$$

Notice that it is not attempted to find an inverse of the time delay. Choosing the filter

$$G_f(s) = \frac{1}{1 + sT_f},$$

and approximating the time delay by

$$e^{-sL} \approx 1 - sL,$$

Equation 6.27 now gives

$$C(s) = \frac{1 + sT}{K_p s(L + T_f)},$$

which is a PI controller. Notice that this controller is identical to the one obtained by lambda tuning if $T_f = T_{cl}$; see Equation 6.20.

If the time delay is approximated instead by a first-order Padé approximation,

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

Equation 6.27 gives instead the PID controller

$$C(s) = \frac{(1 + sL/2)(1 + sT)}{K_p s(L + T_f + sT_f L/2)} \approx \frac{(1 + sL/2)(1 + sT)}{K_p s(L + T_f)}.$$

For processes described by Equation 6.28, we thus find that the internal model principle will give PI or PID controllers. Approximations like the ones discussed in Section 2.8 can be used in the usual manner to obtain PI and PID controllers for more complex processes.

An interesting feature of the internal model controller is that robustness is considered explicitly in the design. Robustness can be adjusted by selecting the filter G_f properly. A trade-off between performance and robustness can be made by using the filter constant as a design parameter. The IMC can be designed to give excellent response to set-point changes. Since the design method inherently implies that poles and zeros of the plant are canceled, the response to load disturbances may be poor if the canceled poles are slow in comparison with the dominant poles. Compare with the responses in Figure 6.14. The IMC controller can also be viewed as an extension of the Smith predictor; see Chapter 8.

Skogestad's Internal Model Controller (SIMC)

Skogestad has developed a version of internal model control tuning method for PID control that avoids some of the drawbacks mentioned above. The method starts with a FOTD model for PI control or a SOTD model for PID control. It is required that the closed-loop system should have the transfer function

$$G_{yysp} = \frac{1}{1 + sT_{cl}} e^{-sL}$$

For an FOTD system it then follows from Equation 6.24 that the controller transfer function is

$$C(s) = \frac{1 + sT}{K_p(1 + sT_{cl} - e^{-sL})} \approx \frac{1 + sT}{sK_p(T_{cl} + L)}$$

where the exponential function is approximated using a Taylor series expansion. In contrast with the recommendation for IMC the closed-loop time constant is proportional to the time delay L. The choice $T_{cl} = L$ is recommended. The integral term is also modified for lag-dominated processes. The tuning rule for PI control is

$$\begin{aligned} K &= \frac{T}{2K_p L} \\ T_i &= \min(T, 8L) \end{aligned} \tag{6.29}$$

The same parameters are used for a PID controller in series form, and the derivative time is chosen as the shortest time constant.

Commentary

The analytical design methods are very closely related to pole placement. The main difference is that the complete transfer function is specified instead of just the closed-loop poles. A nice feature is that the calculations required are very simple. A drawback is that process poles are canceled. This is particularly serious for lag-dominated systems.

6.7 Optimization Methods

Optimization is a powerful tool for design of controllers. The method is conceptually simple. A controller structure with a few parameters is specified. Specifications are expressed as inequalities of functions of the parameters. The specification that is most important is chosen as the function to optimize. The method is well suited for PID controllers where the controller structure and the parameterization are given. There are several pitfalls when using optimization. Care must be exercised when formulating criteria and constraints; otherwise, a criterion will indeed be optimal, but the controller may still be unsuitable because of a neglected constraint. Another difficulty is that the loss function may have many local minima. A third is that the computations required may easily be excessive. Numerical problems may also arise. Nevertheless, optimization is a good tool that has successfully been used to design PID controllers. In this section we discuss some of these methods.

A Warning

Optimization algorithms are very powerful. They will solve whatever criterion is formulated. It is therefore very important to formulate the problems correctly and to introduce all relevant constraints. For PID control it is particularly important to introduce robustness constraints. This has frequently been disregarded in much work the on use of optimization for PID control. The following example illustrates what can happen.

EXAMPLE 6.17—A PI CONTROLLER OPTIMIZED FOR IAE
Consider a process with the transfer function

$$P(s) = \frac{1}{s+1}\, e^{-sL}.$$

Table 6.7 Controller parameters obtained from minimization of integrated absolute error, IAE. K_{hf} is the high-frequency gain of the controller.

L	IAE	M_s	K_{hf}	aK	T_i/L
0.0	0		∞		
0.2	0.14	3.3	4.7	0.94	2.9
0.5	0.60	3.0	2.0	1.0	2.2
1.0	1.5	2.4	1.0	1.0	1.4
2.0	3.2	2.1	0.60	1.2	1.0
5.0	7.7	2.0	0.42	2.1	0.6
10.0	15	1.9	0.37	3.7	0.53

Table 6.7 gives controller parameters that minimize IAE for load disturbances. Some of the other criteria are also given in the table. The table shows that the integrated absolute error increases with L, as can be expected. The table shows that the maximum sensitivity is large for practically all systems, particularly those with small L. The table also shows that the high-frequency gain of the controller is high for small values of L. For example, if we require $M_s < 1.8$, which is a fairly modest robustness requirement, none of the systems is acceptable. □

The example illustrates the necessity of considering all aspects of a problem when formulating the problem. Unfortunately, this was not observed in much of the early work on controller tuning.

Tuning Formulas Based on Optimization

Many studies have been devoted to development of tuning rules based on optimization. Very often a process described by

$$P = \frac{K_p}{1 + sT}\, e^{-sL}$$

has been considered. The loss functions obtained for unit step changes in set point and process input have been computed and formulas of the type

$$p = a \left(\frac{L}{T}\right)^b,$$

where p is a controller parameter and a and b are constants, have been fitted to the numerical values obtained. In many cases, the criterion is IAE for load disturbances, which often gives systems with low damping and poor sensitivity. The formulas given often only hold for a small range of normalized dead times, e.g., $0.2 < \tau < 0.6$. It should also be observed that criteria based on set-point changes can often be misleading because it is often not observed that the set-point changes are drastically influenced by different set-point weightings.

Modulus and Symmetrical Optimum

Modulus Optimum (BO) and Symmetrical Optimum (SO) are two methods for selecting and tuning controllers that also can be viewed as analytical designs where the desired transfer functions given by Equations 6.23 are obtained by optimization. The acronyms BO and SO are derived from the German words Betrags Optimum and Symmetrische Optimum. The methods were developed for motor drives where the response to set-point changes is particularly important. The basic idea is to find a controller that makes the frequency response from set point to plant output as close to one as possible for low frequencies. If $G(s)$ is the transfer function from the set point to the output, the controller is determined in such a way that $G(0) = 1$ and that $d^n|G(i\omega)|/d\omega^n = 0$ at $\omega = 0$ for as many n as possible. An interesting property is that the design method takes account of unmodeled dynamics explicitly. We illustrate the idea with a few examples.

EXAMPLE 6.18—SECOND-ORDER SYSTEM
Consider the transfer function

$$G(s) = \frac{a_2}{s^2 + a_1 s + a_2},$$

which has been chosen so that $G(0) = 1$. Let us first consider how the parameters should be chosen in order to get a maximally flat frequency response. We have

$$|G(i\omega)|^2 = \frac{a_2^2}{a_1^2\omega^2 + (a_2 - \omega^2)^2} = \frac{a_2^2}{a_2^2 + \omega^2(a_1^2 - 2a_2) + \omega^4}.$$

By choosing $a_1 = \sqrt{2a_2}$ we find

$$|G(i\omega)|^2 = \frac{a_2^2}{a_2^2 + \omega^4}.$$

The first three derivatives of $|G(i\omega)|$ will vanish at the origin. The transfer function then has the form

$$G(s) = \frac{\omega_0^2}{s^2 + \omega_0 s\sqrt{2} + \omega_0^2}.$$

The step response of a system with this transfer function has an overshoot $o = 4\%$. The settling time to 2% of the steady-state value is $T_s = 6/\omega_0$. □

If the transfer function G in the example is obtained by error feedback of a system with the loop transfer function G_{BO}, the loop transfer function is

$$G_{\mathrm{BO}}(s) = \frac{G(s)}{1 - G(s)} = \frac{\omega_0^2}{s(s + \sqrt{2}\omega_0)}, \tag{6.30}$$

which is the desired loop transfer function for the method called modulus optimum.

The calculation in Example 6.18 can be performed for higher-order systems with more effort. We illustrate by another example.

EXAMPLE 6.19—THIRD-ORDER SYSTEM WITH NO ZEROS
Consider the transfer function

$$G(s) = \frac{a_3}{s^3 + a_1 s^2 + a_2 s + a_3}.$$

After some calculations we get

$$|G(i\omega)| = \frac{a_3}{\sqrt{a_3^2 + (a_2^2 - 2a_1 a_3)\omega^2 + (a_1^2 - 2a_2)\omega^4 + \omega^6}}.$$

Five derivatives of $|G(i\omega)|$ will vanish at $\omega = 0$, if the parameters are such that $a_1^2 = 2a_2$ and $a_2^2 = 2a_1 a_3$. The transfer function then becomes

$$G(s) = \frac{\omega_0^3}{s^3 + 2\omega_0 s^2 + 2\omega_0^2 s + \omega_0^3} = \frac{\omega_0^3}{(s + \omega_0)(s^2 + \omega_0 s + \omega_0^2)}. \tag{6.31}$$

The step response of a system with this transfer function has an overshoot $o = 8.1\%$. The settling time to 2% of the steady state value is $9.4/\omega_0$. A system with this closed-loop transfer function can be obtained with a system having error feedback and the loop transfer function

$$G_l(s) = P(s)C(s) = \frac{\omega_0^3}{s(s^2 + 2\omega_0 s + 2\omega_0^2)}.$$

The closed-loop transfer function (6.31) can also be obtained from other loop transfer functions if the controller has set-point weighting. For example, if a process with the transfer function

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

is controlled by a PI controller having parameters $K = 2$, $T_i = 2/\omega_0$, and $b = 0$, the loop transfer function becomes

$$G_{SO} = \frac{\omega_0^2(2s + \omega_0)}{s^2(s + 2\omega_0)}. \tag{6.32}$$

The symmetric optimum aims at obtaining the loop transfer function given by Equation 6.32. Notice that the Bode plot of this transfer function is symmetrical around the frequency $\omega = \omega_0$. This is the motivation for the name symmetrical optimum.

If a PI controller with $b = 1$ is used, the transfer function from set point to process output becomes

$$G(s) = \frac{G_{SO}(s)}{1 + G_{SO}(s)} = \frac{(2s + \omega_0)\omega_0^2}{(s + \omega_0)(s^2 + \omega_0 s + \omega_0^2)}.$$

This transfer function is not maximally flat because of the zero in the numerator. This zero will also give a set-point response with a large overshoot, about 43 percent. □

The methods BO and SO can be called loop-shaping methods since both methods try to obtain a specific loop transfer function. The design methods can be described as follows. It is first established which of the transfer functions, G_{BO} or G_{SO}, is most appropriate. The transfer function of the controller $C(s)$ is then chosen so that the loop transfer $G_l(s) = P(s)C(s)$ meet specifications. We illustrate the methods with the following examples.

EXAMPLE 6.20—BO CONTROL
Consider a process with the transfer function

$$P(s) = \frac{K_p}{s(1+sT)}. \tag{6.33}$$

With a proportional controller the loop transfer function becomes

$$G_\ell(s) = \frac{KK_p}{s(1+sT)}.$$

To make this transfer function equal to G_{BO} given by Equation 6.30 it must be required that

$$\omega_0 = \frac{\sqrt{2}}{2T}.$$

The controller gain should be chosen as

$$K = \frac{\omega_0\sqrt{2}}{2K_p} = \frac{1}{2K_pT}.$$

□

EXAMPLE 6.21—SO CONTROL
Consider a process with the same transfer function as in the previous example (Equation 6.33). With a PI controller having the transfer function

$$C(s) = \frac{K(1+sT_i)}{sT_i},$$

we obtain the loop transfer function

$$G_l(s) = P(s)C(s) = \frac{K_pK(1+sT_i)}{s^2T_i(1+sT)}.$$

This is identical to G_{SO} if we choose

$$K = \frac{1}{2K_pT}$$
$$T_i = 4T.$$

To obtain the transfer function given by Equation 6.31 the set-point weight b should be zero. □

A Design Procedure A systematic design procedure can be based on the methods BO and SO. The design method consists of two steps. In the first step the process transfer function is simplified to one of the following forms:

$$
\begin{aligned}
P_1(s) &= \frac{K_p}{1+sT} \\
P_2(s) &= \frac{K_p}{(1+sT_1)(1+sT_2)}, \quad T_1 > T_2 \\
P_3(s) &= \frac{K_p}{(1+sT_1)(1+sT_2)(1+sT_3)}, \quad T_1 > T_2 > T_3 \\
P_4(s) &= \frac{K_p}{s(1+sT)} \\
P_5(s) &= \frac{K_p}{s(1+sT_1)(1+sT_2)}, \quad T_1 > T_2.
\end{aligned}
$$

Process poles may be canceled by controller zeros to obtain the desired loop transfer function. A slow pole may be approximated by an integrator; fast poles may be lumped together as discussed in Section 2.8. The rule of thumb given in the original papers on the method is that time constants T such that $\omega_0 T < 0.25$ can be regarded as integrators.

The controller is derived in the same way as in Examples 6.20 and 6.21 by choosing parameters so that the loop transfer function matches either G_{BO} or G_{SO}. By doing this we obtain the results summarized in Table 6.8. Notice, for example, that Examples 6.20 and 6.21 correspond to the entries Process G_4 in the table. It is natural to view the smallest time constant as an approximation of neglected dynamics in the process. It is interesting to observe that it is this time constant that determines the bandwidth of the closed-loop system.

The set-point response for the BO method is excellent. Notice that it is necessary to use a controller with a two-degree-of-freedom structure or a pre-filter to avoid a high overshoot for the SO method. Notice also that process poles are canceled in the cases marked C1 or C2 in Table 6.8. The response to load disturbances will be poor if the canceled pole is slow compared to the closed-loop dynamics, which is characterized by ω_0 in Table 6.8.

These design principles can be extended to processes other than those listed in the table.

EXAMPLE 6.22—APPLICATION OF BO AND SO

Consider a process with the transfer function

$$P(s) = \frac{1}{(1+s)(1+0.2s)(1+0.05s)(1+0.01s)}. \tag{6.34}$$

Since this transfer function is of fourth order, the design procedure cannot be applied directly. We show how different controllers are obtained depending on the approximations made. The performance of the closed-loop system depends on the approximation. We use parameter ω_0 as a crude measure of performance.

If a controller with low performance is acceptable, the process (6.34) can be approximated with

$$P(s) = \frac{1}{1+1.26s}. \tag{6.35}$$

Table 6.8 Controller parameters obtained with the BO and SO methods. Entry P gives the process transfer function, entry C gives the controller structure, and entry M tells whether the BO or SO method is used. In the entry Remark, A1 means that $1 + sT_1$ is approximated by sT_1, and Ci means that the time constant T_i is canceled.

P	C	M	Remark	KK_p	T_i	T_d	ω_0	b	c
P_1	I	BO		0.5	T		$\frac{0.7}{T}$		
P_2	P	BO	A1	$\frac{T_1}{2T_2}$			$\frac{0.7}{T_2}$	1	
P_2	PI	BO	C1	$\frac{T_1}{2T_2}$	T_1		$\frac{0.7}{T_2}$	1	
P_2	PI	SO	A1	$\frac{T_1}{2T_2}$	$4T_2$		$\frac{0.5}{T_2}$	0	
P_3	PD	BO	A1, C2	$\frac{T_1}{2T_3}$		T_2	$\frac{0.7}{T_3}$	1	1
P_3	PID	BO	C1, C2	$\frac{T_1 + T_2}{2T_3}$	$T_1 + T_2$	$\frac{T_1 T_2}{T_1 + T_2}$	$\frac{0.7}{T_3}$	1	1
P_3	PID	SO	A1, C2	$\frac{T_1(T_2 + 4T_3)}{8T_3^2}$	$T_2 + 4T_3$	$\frac{4T_2 T_3}{T_2 + 4T_3}$	$\frac{0.5}{T_3}$	$\frac{T_2}{T_2 + 4T_3}$	0
P_4	P	BO		$\frac{1}{2T}$			$\frac{0.7}{T}$	1	
P_4	PI	SO		$\frac{1}{2T}$	4T		$\frac{0.5}{T}$	0	
P_5	PD	BO	C1	$\frac{1}{2T_2}$		T_1	$\frac{0.7}{T_2}$	1	1
P_5	PD	SO	A1	$\frac{T_1}{8T_2^2}$		$4T_2$	$\frac{0.5}{T_2}$	1	0
P_5	PID	SO	C1	$\frac{T_1 + 4T_2}{8T_2^2}$	$T_1 + 4T_2$	$\frac{4T_1 T_2}{T_1 + 4T_2}$	$\frac{0.5}{T_2}$	$\frac{T_1}{T_1 + 4T_2}$	0

Table 6.9 Results obtained with different controllers designed by the BO and SO methods in Example 6.22. The frequency ω_m defines the upper limit when the phase error is less than 10%.

Controller	K	T_i	T_d	k_i	b	c	ω_0	ω_m	IAE
1				0.4			0.55	1.12	2.7
2	1.92	1		0.52	1		2.7	5.15	0.52
3	10	1.2	0.17	8.3	1	1	11.7	26.6	0.12
4	15.3	0.44	0.11	35	0.45	0	8.3	26.6	0.029

The approximation has a phase error less than 10° for $\omega \leq 1.12$. It follows from Table 6.8 that the system (6.35) can be controlled with an integrating controller with

$$k_i = \frac{K}{T_i} = \frac{0.5}{1.26} = 0.4.$$

This gives a closed-loop system with $\omega_0 = 0.55$.

A closed-loop system with better performance is obtained if the transfer function (6.34) is approximated with

$$P(s) = \frac{1}{(1+s)(1+0.26s)}. \tag{6.36}$$

The slowest time constant is thus kept, and the remaining time constants are approximated by lumping them together. The approximation has a phase error less than 10° for $\omega \leq 5.15$. A PI controller can be designed using the BO method. The parameters $K = 1.92$ and $T_i = 1$ are obtained from Table 6.8. The closed-loop system has $\omega_0 = 2.7$.

If the transfer function is approximated as

$$P(s) = \frac{1}{(1+s)(1+0.2s)(1+0.06s)}, \tag{6.37}$$

the approximation has a phase error less than 10° for $\omega \leq 26.6$. The BO method can be used also in this case. Table 6.8 gives the controller parameters $K = 10$, $T_i = 1.2$, and $T_d = 0.17$. The controller structure is defined by the parameters $b = 1$ and $c = 1$. This controller gives a closed-loop system with $\omega_0 = 11.7$.

The method SO can also be applied to the system (6.37). Table 6.8 gives the controller parameters $K = 15.3$, $T_i = 0.44$, $T_d = 0.11$, and $b = 0.45$. For these parameters we get $\omega_0 = 8.3$.

Controllers with different properties can be obtained by approximating the transfer function in different ways. A summary of the properties of the closed-loop systems obtained is given in Table 6.9, where IAE refers to the load disturbance response. Notice that Controller 2 cancels a process pole with time constant 1 s and that Controller 3 cancels process poles with time constants 1 s and 0.25 s. This explains why the IAE drops drastically for Controller 4, which

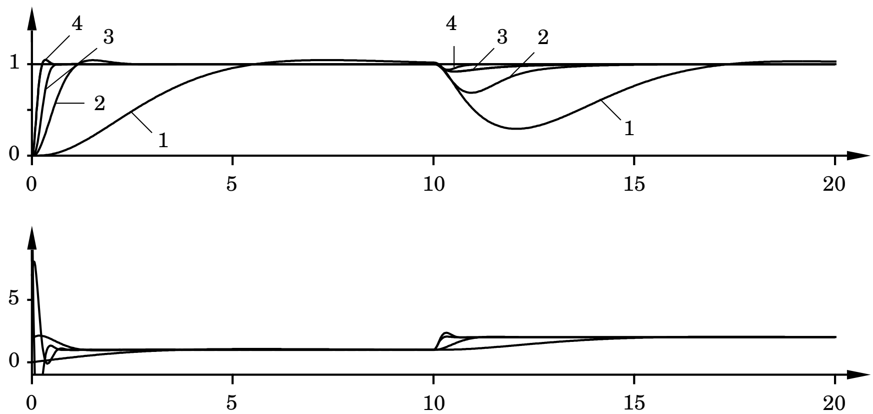

Figure 6.16 Simulation of the closed-loop system obtained with different controllers designed by the BO and SO methods given in Table 6.9. The upper diagram shows set point y_{sp} and process output y, and the lower diagram shows control signal u.

does not cancel any process poles. Controller 4 actually has a lower bandwidth ω_0 than Controller 3.

A simulation of the different controllers is shown in Figure 6.16. The simulation and the data shown in Table 6.9 clearly illustrate the benefits of improved modeling and more complicated controllers. □

Design for Disturbance Rejection

The design methods discussed so far have been based on a characterization of process dynamics. The properties of the disturbances have only influenced the design indirectly. A load disturbance in the form of a step was used, and in some cases a loss function based on the error due to a load disturbance was minimized. Measurement noise was also incorporated by limiting the high-frequency gain of the controller.

In this section, we briefly discuss design methods that directly attempt to make a trade-off between attenuation of load disturbances and amplification of measurement noise due to feedback.

Consider the system shown in Figure 6.17. Notice that the measurement signal is filtered before it is fed to the controller. Let D and N be the Laplace transforms of the load disturbance and the measurement noise, respectively. The process output and the control signal are then given by

$$\begin{aligned} X &= \frac{P}{1+G_\ell} D - \frac{G_\ell}{1+G_\ell} N \\ U &= -\frac{G_\ell}{1+G_\ell} D - \frac{CG_f}{1+G_\ell} N, \end{aligned} \tag{6.38}$$

where $G_\ell = PCG_f$ is the loop transfer function. Different assumptions about the disturbances and different design criteria can now be given. We illustrate by an example.

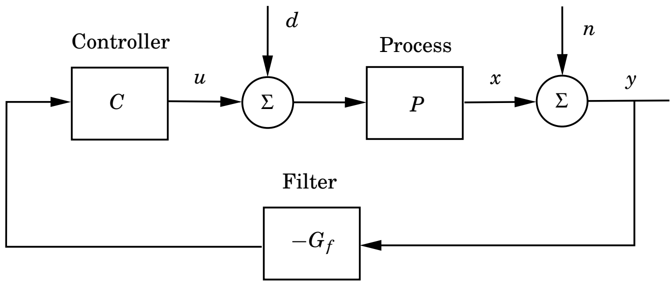

Figure 6.17 Block diagram of a closed-loop system.

EXAMPLE 6.23—DESIGN FOR DISTURBANCE REJECTION
Assume that the transfer functions in Figure 6.17 are given by

$$P = \frac{1}{s}, \qquad G_f = 1, \qquad C = k + \frac{k_i}{s}.$$

Furthermore, assume that n is stationary noise with spectral density ϕ_n and that d is obtained by sending stationary noise with the spectrum ϕ_d through an integrator. This is one way to model the situation that the load disturbance is drifting and the measurement noise has high frequency.

With the given assumptions, Equation 6.38 then becomes

$$\begin{aligned} X &= \frac{s}{s^2+ks+k_i}\,\frac{1}{s}\,D_1 - \frac{sk+k_i}{s^2+ks+k_i}\,N \\ U &= -\frac{sk+k_i}{s^2+ks+k_i}\,\frac{1}{s}\,D_1 - \frac{s^2k+k_is}{s^2+ks+k_i}\,N, \end{aligned}$$

where we have assumed

$$D(s) = \frac{1}{s}\,D_1(s).$$

If n and d_1 are white noises, it follows that the variance of x is given by

$$J = Ex^2 = \frac{1}{2kk_i}\,\phi_d + \frac{1}{2}\left(k + \frac{k_i}{k}\right)\phi_n.$$

This equation clearly indicates the compromise in designing the controller. The first term of the right-hand side is the contribution to the variance due to the load disturbance. The second term represents the contribution due to the measurement noise. Notice that the attenuation of the load disturbances increases with increasing k and k_i, but that large values of k and k_i also increase the contribution of measurement noise.

We can attempt to find values of k and k_i that minimize J. A straightforward calculation gives

$$\begin{aligned} k &= \sqrt{2}\left(\frac{\phi_d}{\phi_n}\right)^{1/4} \\ k_i &= \sqrt{\frac{\phi_d}{\phi_n}}. \end{aligned}$$

This means that the controller parameters are uniquely given by the ratio of the intensities of the process noise and the measurement noise. Also notice that with these parameters the closed-loop characteristic polynomial becomes

$$s^2 + \omega_0 s\sqrt{2} + \omega_0^2,$$

with $\omega_0 = \sqrt{\phi_v/\phi_e}$. The optimal system thus has a relative damping $\zeta = 0.707$ and a bandwidth that is given by the ratio of the intensities of load disturbance and measurement noise. □

Commentary

Optimization techniques are very powerful. When using them it is essential to include all relevant aspects of the problem in the formulation; otherwise, the so-called optimal controller may have very bad properties. In this section we have covered a few optimization methods that have been used for PID control.

The methods BO and SO are widely used for drive systems. The optimization is to find a transfer function from set point to process output that is maximally flat. The methods are primarily intended for systems without dead time. Small dead times can be dealt with by approximation.

An interesting feature of the procedure is the use of approximations; fast poles and slow time constants are neglected, and slow dynamics are approximated by integrators. Model uncertainty also appears explicitly in the design because the achievable bandwidth is determined by slowest neglected time constants.

The methods can be interpreted as pole placement where the desired closed-loop characteristic polynomial is

$$A_{\mathrm{BO}}(s) = s^2 + \omega_0 s\sqrt{2} + \omega_0^2$$

for the modulus optimum and

$$A_{\mathrm{SO}}(s) = (s + \omega_0)(s^2 + \omega_0 s + \omega_0^2)$$

for the symmetrical optimum. There are possibilities for combining the approaches. A drawback with all design methods of this type is that process poles may be canceled. This may lead to poor attenuation of load disturbances if the canceled poles are excited by disturbances and if they are slow compared to the dominant closed-loop poles.

6.8 Robust Loop Shaping

The design methods discussed so far all have the property that the robustness to process variations has to be checked after a design. One of the major advances in control theory in the end of the last century was the emergence of design methods with guaranteed robustness (the so called $\mathcal{H}^\infty$ theory). In this section we will present a method for design of PID controllers in the same spirit. In Section 4.6 it was shown that robustness conditions can be expressed in terms of circular discs that are forbidden regions for the Nyquist curve of the loop transfer function. For PID control these conditions give a set of admissible values of the controller parameters, called the robustness region. Attenuation of low-frequency load disturbances is inversely proportional to integral gain k_i. Measurement noise injection is captured by controller gain k for P and PI control or derivative gain k_d for PD and PID control. The design method is to maximize integral gain k_i subject to constraints on robustness and noise injection. Good set-point response is then obtained by set-point weighting or feedforward as discussed in Section 5.3. This design method brings design of PID controllers into the mainstream of control system design.

The Robustness Region

In Section 4.6 it was shown that robustness to process variations can be expressed by the maximum sensitivity M_s, the maximum complementary sensitivity M_t, or with the joint sensitivity M. All these conditions say that the Nyquist curve of the loop transfer function should avoid circles enclosing the critical point. For PID control of a process with given transfer function the robustness constraint translates into constraints on the controller parameters, called the robustness region. To determine the robustness region we consider a process with the transfer function $P(s)$ and an ideal PID controller with the transfer function $C(s)$. The loop transfer function is $G_l(s)$, and the square of the distance from a point on the Nyquist curve of the loop transfer function to the point $-c$ is

$$f(k, k_i, k_d, \omega) = |c + G_l(i\omega)|^2 = |c + \big(k + i(k_d\omega - k_i/\omega)\big)P(i\omega)|^2;$$

and the robustness constraint becomes

$$f(k, k_i, k_d, \omega) \geq r^2. \tag{6.39}$$

Introduce

$$P(i\omega) = \alpha(\omega) + i\beta(\omega) = \rho(\omega)e^{i\varphi(\omega)}, \tag{6.40}$$

where

$$\begin{aligned}\alpha(\omega) &= \rho(\omega)\cos\varphi(\omega),\\ \beta(\omega) &= \rho(\omega)\sin\varphi(\omega).\end{aligned}$$

The following straightforward but tedious calculation shows that the function f can be written as

$$\begin{aligned}f(k, k_i, k_d, \omega) &= \Big|c + \big(k + i(k_d\omega - k_i/\omega)\big)\big(\alpha(\omega) + i\beta(\omega)\big)\Big|^2\\ &= \big|c + \alpha k + \beta(k_d\omega - k_i/\omega) + i\big(\beta k + \alpha(k_d\omega - k_i/\omega)\big)\big|^2\\ &= c^2 + \rho^2k^2 + 2c\alpha k + \rho^2(k_d\omega - k_i/\omega)^2 - 2\beta c(k_d\omega - k_i/\omega)\\ &= \rho^2\Big(k + \frac{\alpha c}{\rho^2}\Big)^2 + \frac{\rho^2}{\omega^2}\Big(k_i + \frac{\omega\beta c}{\rho^2} - k_d\omega^2\Big)^2 \geq r^2,\end{aligned}$$

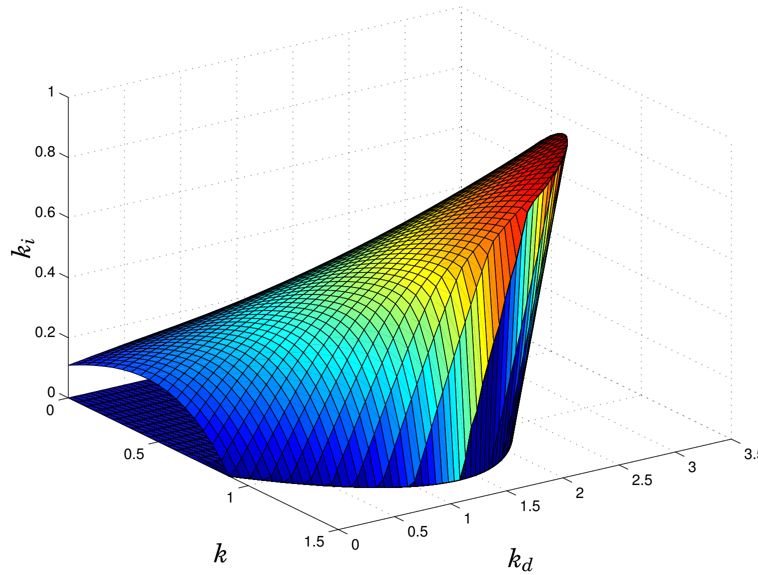

Figure 6.18 Robustness region for a process with the transfer function $P(s) = 1/(s+1)^4$ and the robustness criterion $M_s \leq 1.4$.

where the argument ω in the functions α and β have been dropped to simplify the writing. Inserting the arguments the robustness condition can be written as

$$\Big(\frac{\rho(\omega)}{r}\Big)^2\Big(k+\frac{\alpha(\omega)c}{\rho(\omega)^2}\Big)^2+\Big(\frac{\rho(\omega)}{\omega r}\Big)^2\Big(k_i+\frac{\omega\beta(\omega)c}{\rho(\omega)^2}-\omega^2 k_d\Big)^2\leq 1. \qquad (6.41)$$

To have a stable closed-loop system there is also an encirclement condition required by Nyquist's stability theorem. The robustness constraint thus implies that the controller parameters must belong to a region called the *robustness region*; see Figure 6.18. Design of PID controllers can thus be formulated as the following semi-infinite programming problem: maximize k_i subject to the robustness constraint (6.41) and constraints on k and k_d.

Figure 6.18 gives good insight into the design problem. The PI controller, which maximizes integral gain, can be found from the intersection of the robustness region with the plane $k_d = 0$. The best PI controller has $k = 0.4$ and $k_i = 0.2$. Five times larger values of the integral gain can be obtained by using derivative action.

The optimization problem is not straightforward since the constraint (6.41) must be satisfied for all ω, and the set of parameters that satisfy the constraint is not necessarily convex. Before solving the optimization problem we will therefore investigate the constraint set.

A Geometric Interpretation

The robustness constraint (6.41) has a nice interpretation. For fixed ω and k_d it represents the exterior of an ellipse in the k-k_i plane; see Figure 6.19.

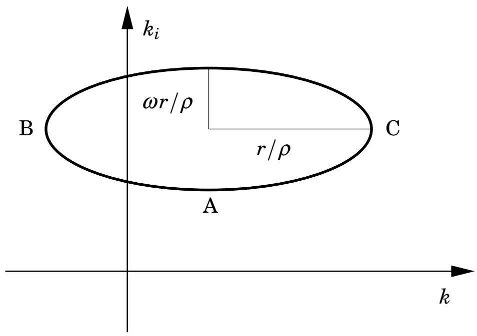

Figure 6.19 Graphical illustration of the sensitivity constraint (6.41).

The ellipse has its center in $k = \alpha c/\rho^2$ and $k_i = \omega\beta c/\rho^2$, and its axes are parallel to the coordinate axes. The horizontal half axis has length r/ρ, and the vertical half axis has length $\omega r/\rho$. The center of the ellipse lies on the stability boundary.

When ω ranges from 0 to ∞ the ellipses have an envelope

$$\begin{aligned} f(k, k_i, k_d, \omega) &= r^2, \\ \frac{\partial f}{\partial \omega}(k, k_i, k_d, \omega) &= 0, \end{aligned} \tag{6.42}$$

which defines one boundary of the sensitivity constraint. Assuming that the process has positive gain the other boundary is given by the $k - k_d$ plane. Since the function f is quadratic in k_i the envelope has two branches; only one branch corresponds to stable closed-loop systems.

Having understood the nature of the constraints it is conceptually easy to solve the optimization problem by finding the largest value of k_i on the envelope. There may be local minima and the envelope may have edges. This is illustrated in Figure 6.20, which shows the envelopes and the locus of the lowest vertex of the ellipse in two cases. The figure on the left has a smooth envelope, and the locus of the lowest vertex coincides with the envelope at the maximum. The figure on the right has an envelope with an edge at the maximum value of k_i. Since it is quite time-consuming to generate the envelope, it is desirable to find algorithms that can give a more effective solution. It is also of interest to characterize the cases where there is only one local minimum.

Smooth Envelope

We will first consider the case where the envelope is smooth and does not have corners near the maximum. The largest value of k_i for fixed k_d then occurs at a tangency with the lower vertex of the ellipse; see Figure 6.19. The locus of the lower vertical vertex is given by

$$\begin{aligned} k(\omega) &= -\frac{\alpha c}{\rho^2} = -\frac{c}{\rho(\omega)}\cos\varphi(\omega), \\ k_i(\omega) &= -\frac{\omega\beta c}{\rho^2} - \frac{\omega r}{\rho} + \omega^2 k_d = -\frac{\omega}{\rho(\omega)}(r + c\sin\varphi(\omega)) + \omega^2 k_d. \end{aligned} \tag{6.43}$$

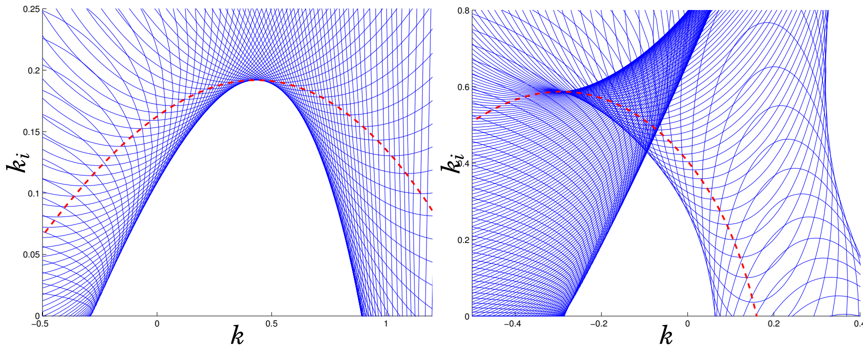

Figure 6.20 Geometrical illustration of the ellipses generated by the sensitivity constraint (6.39) and the envelope generated by it. The curves on the left are generated for a system with the transfer function $P(s) = (s+1)^{-4}$ with the constraint $M_s = 1.4$. The curves on the right are generated for a system with the transfer function $P(s) = 1/(s+1)(s^2+0.2s+9)$ with the constraint $M_s = 1.4$.

It is shown as a dashed line in the Figure 6.20. The largest value of k_i can thus be found by maximizing k_i on the locus of the lowest vertex. Differentiating the expression for k_i in (6.43) gives

$$\begin{aligned}\frac{dk_i}{d\omega} &= -\frac{d}{d\omega}\Big(\frac{\omega(r+c\sin\varphi)}{\rho}\Big) + 2\omega k_d \\ &= (r+c\sin\varphi)\Big(\frac{\omega\rho'}{\rho^2} - \frac{1}{\rho}\Big) - \frac{\omega\varphi' c\cos\varphi}{\rho} + 2\omega k_d = 0.\end{aligned}$$

To simplify the writing we have dropped the argument ω of the functions α, β, and φ. Dividing the above equation with ω and multiplying it with ρ, the condition for extremum becomes

$$h_{PID}(\omega) = (r+c\sin\varphi)\Big(\frac{\rho'}{\rho} - \frac{1}{\omega}\Big) - c\varphi'\cos\varphi + 2\rho k_d = 0. \tag{6.44}$$

To find the optimum we thus have to find the solution ω^*_{PID} of this equation; the controller parameters are then obtained from Equation 6.43.

Equation 6.44 is satisfied for a minimum, a maximum, or saddle point. To ensure that there is a maximum it must be required that

$$\frac{d^2 f}{d\omega^2}(\omega^*) > 0. \tag{6.45}$$

To guarantee that the constraint (6.41) is satisfied globally we have to evaluate it for all ω. This can be done by the Nyquist plot of the loop transfer function.

Equation 6.44 can be solved iteratively by bisection or with the Newton-Raphson method, both methods converge very fast, but they require appropriate initial conditions. Notice, however, that in general, there may be several solutions that can be found by starting the iteration from different initial conditions.

For special classes of systems it is possible to give good initial conditions. Consider systems where the transfer function $P(s)$ has positive low-frequency gain and

$$\begin{aligned} \frac{d \arg P(i\omega)}{d\omega} &< 0, \\ \frac{d \log_{10} |P(i\omega)|}{d \log_{10} \omega} &< 1. \end{aligned} \tag{6.46}$$

These conditions imply that the quantity $\rho'/\rho - 1/\omega$ is negative. For PI control, when $k_d = 0$, it follows from (6.44) and (6.46) that $h_{PI}(\omega_{90}) > 0$ and that $h_{PI}(\omega_{180-\arcsin r/c}) < 0$. Equation 6.44 then has a root in the interval

$$\omega_{90} < \omega^*_{PI} \leq \omega_{180-\arcsin (r/c)}. \tag{6.47}$$

The monotonicity condition (6.46) thus only has to be valid in the interval (6.47). If condition (6.46) holds it follows from Equation 6.43 and 6.47 that both k and k_i are positive. Many processes satisfy this condition.

PD Control

For PD control it is natural to maximize proportional gain subject to the robustness constraint. Working out the details for the case of smooth envelopes we find that the problem can be solved as follows: Find a value of ω such that

$$h_{PD}(\omega) = (r + c \cos \varphi) \frac{\rho'}{\rho} + c\varphi' \sin \varphi = 0. \tag{6.48}$$

Then compute the controller gains from the equations

$$\begin{aligned} k(\omega) &= -\frac{\alpha c}{\rho^2} - \frac{r}{\rho} = -\frac{r + c \cos \varphi}{\rho}, \\ k_d(\omega) &= \frac{\beta c}{\omega \rho^2} = \frac{c \sin \varphi}{\omega \rho}. \end{aligned} \tag{6.49}$$

If ρ'/ρ is negative (6.48) always has a solution ω^*_{PD} in the interval

$$\omega_{180} < \omega^*_{PD} < \omega_{270-\arcsin (r/c)} = \omega_{180+\arccos (r/c)}. \tag{6.50}$$

The formula and the code for design of PD controllers can also be used simply by making the observation that designing a PD controller for the system $P(s)$ is the same as designing a PI controller for the system $sP(s)$.

A Design Algorithm

We obtained the following algorithm for solving the design problem in the case of smooth envelopes.

ALGORITHM 6.1—CONTROLLER DESIGN FOR SMOOTH ENVELOPE

1. Design a PD controller by solving (6.48) by bisection starting with the interval $(\omega_{180}, \omega_{180+\arccos(r/c)})$. The solution gives the frequency ω^*_{PD}.
2. Design a PI controller by solving (6.44) with $k_d = 0$ by bisection starting with the interval $(\omega_{90}, \omega_{180-\arcsin(r/c)})$. The solution gives the frequency ω^*_{PI}.
3. Design a PID controller for fixed k_d by solving (6.44) by bisection starting with the interval $(\omega^*_{PI}, \omega^*_{PD})$. Increase k_d to the largest value for which the robustness constraint is satisfied.
4. Verify that there is a smooth envelope by computing (6.45) or by the Nyquist plot of the loop transfer function.

□

If the envelope is not smooth the solution obtained by iteration corresponds to a local maximum of the distance from the critical point to the Nyquist curve. The Nyquist curve then enters the constraint region for points around the maximum.

Envelope with Corners

The largest value of k_i may also occur at a point where the envelope has an edge. This is illustrated in Figure 6.21. The vertices B and C of the ellipse in Figure 6.19 are given by

$$\begin{aligned} k(\omega) &= -\frac{\alpha c}{\rho^2} \pm \frac{r}{\rho} = -\frac{\alpha(\omega) c \cos\phi(\omega)}{\rho(\omega)} \pm \frac{r}{\rho(\omega)}, \\ k_i(\omega) &= -\frac{\omega\beta c}{\rho^2} + \omega^2 k_d = -\frac{\omega c \sin\phi(\omega)}{\rho(\omega)} + \omega^2 k_d, \end{aligned} \tag{6.51}$$

where the left vertex corresponds to a minus sign and the right vertex to a plus sign. The loci of these vertices are shown in thin dotted lines, and the loci of the center of the ellipses are shown in thin dashed lines. The envelope is shown as a thick solid line, and the locus of the lowest vertex of the ellipse by thick dashed lines. Notice that the maximum occurs at the intersection of ellipses corresponding to two different frequencies, ω_1 and ω_2; see Figure 6.21. The envelope condition (6.42) is then satisfied for both frequencies. This gives the condition

$$\begin{aligned} f(k, k_i, k_d, \omega_1) &= R^2, \\ \frac{\partial f}{\partial \omega}(k, k_i, k_d, \omega_1) &= 0, \\ f(k, k_i, k_d, \omega_2) &= R^2, \\ \frac{\partial f}{\partial \omega}(k, k_i, k_d, \omega_2) &= 0. \end{aligned} \tag{6.52}$$

In the Nyquist plot this corresponds to the case when the loop transfer function is tangent to the M circle at two points.

It is thus possible to characterize the point where k_i has its largest value by algebraic equations. This means that the design problem is reduced to solving

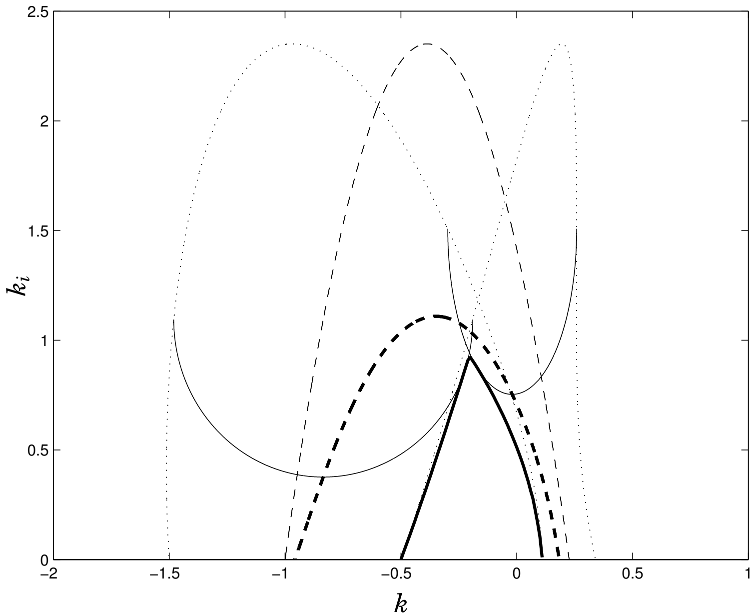

Figure 6.21 Geometrical illustration of the sensitivity constraint (6.39) and the envelope generated by it. The envelope is shown by the thick solid line; the locus of the lower vertex by the thick dashed line. One half ellipse is shown as a thin solid line. The locus of the center of the ellipses is shown as a thin dashed line, and the loci of the vertical vertices by dotted lines. The curves are generated for a system with the transfer function $P(s) = 1/(s+1)(s^2+0.2s+9)$ with the constraint $M_s = 2.0$.

algebraic equations, (6.52), and that elaborate search procedures are avoided. The equation can be solved using the Newton-Raphson method.

Good initial values essential for the Newton-Raphson iteration can be obtained by approximating the envelope by the loci of the right horizontal locus and the locus of the lowest vertex of the ellipse; see Figure 6.21. We illustrate the case of envelopes with corners with an example.

EXAMPLE 6.24—AN OSCILLATORY SYSTEM
Consider the process with the transfer function

$$P(s) = \frac{9}{(s+1)(s^2+as+9)}.$$

This is an interesting process from two points of view. First, the system has two oscillatory poles with relative damping $\zeta = a/6$. When parameter a is decreased it becomes more and more difficult to control the process. Second, depending on the value of parameter a the envelope may have a continuous derivative, $a \geq 1.0653$, or a corner, $a < 1.0653$.

For the case when the envelope has a corner, a PI controller was designed for $M_s = 2.0$. In Figure 6.22 the Nyquist curves and the time responses are shown for the cases $a = 0.2,\ 0.5$, and 1.0. The controller behaves reasonably well in spite of the poorly damped poles. In Table 6.10 the controller parameters and

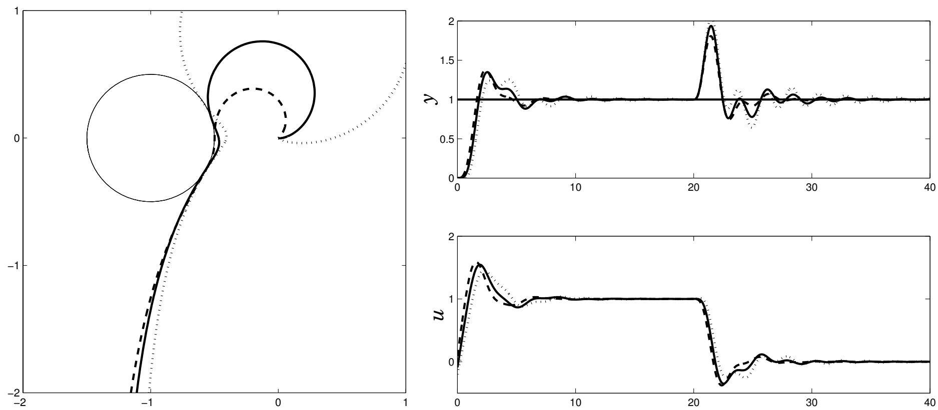

Figure 6.22 Nyquist curves of the loop transfer function and time responses for Example 6.24 with $a = 0.2$ (dotted), 0.5, and 1.0 (dashed), when designing for $M_s = 2.0$.

Table 6.10 Interesting parameters when designing a controller for $M_s = 2.0$ and different values of a in Example 6.24.

a	K	k_i	ω_1	ω_2
0.0	−0.29	0.68	0.97	2.75
0.1	−0.25	0.82	1.08	2.71
0.2	−0.20	0.93	1.16	2.67
0.5	−0.09	1.17	1.37	2.55
1.0	0.09	1.38	1.65	2.30
2.0	0.48	1.54	2.79	2.79

the frequencies at which the loop transfer function is tangent to the M_s-circle are shown. Notice in Table 6.10 how the proportional gain is negative for small values of a. This is the only way to increase the damping of the oscillatory poles with a PI controller.

Finally, we illustrate how our design method will provide a reasonable PI controller for the extreme case $a = 0$. With the design parameter $M_s = 1.4$ we obtain the controller parameters $K = -0.183$, $k_i = 0.251$, and $b = 0$. The time responses are shown in Figure 6.23. We observe that the set-point response is quite reasonable, even if there is a trace of poorly damped modes. The load disturbance will, however, excite the oscillatory modes. The fact that the PI controller is unable to provide damping of these modes is clearly noticeable in the figure. □

The Derivative Cliff

Smooth envelopes are frequently encountered for PI control of systems with essentially monotone frequency responses, and for PID control with moderate

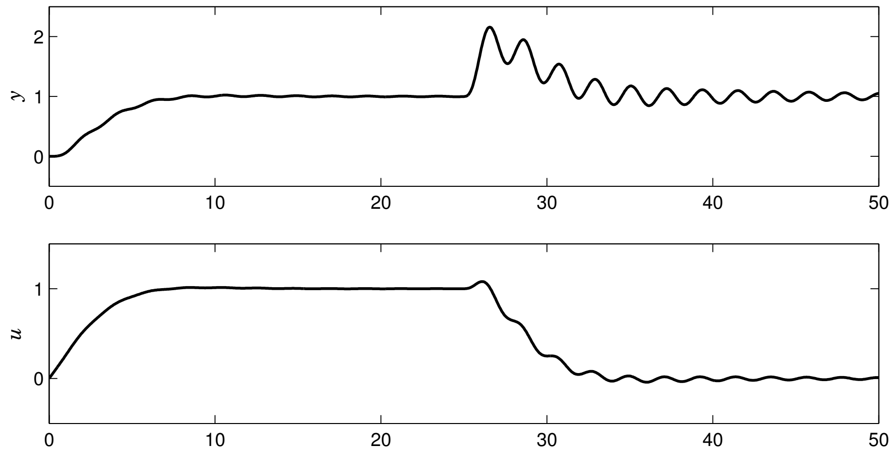

Figure 6.23 Time response of the closed-loop system of Example 6.24 obtained for $a = 0$, when designing the PI controller for $M_s = 1.4$.

values of k_d. However, optimization of k_i over the robustness region often gives controllers with undesirable properties. This can be understood from the plot of the robustness region in Figure 6.18, which shows that the largest value of k_i occurs at an edge. Such a solution is undesirable because small changes in controller parameters give drastic changes in k_i. This is also illustrated in Figure 6.24, which shows intersections of the robustness surface for fixed values of k_d. The figure shows that for PI control ($k_d = 0$) the envelope is smooth at the maximum $k_i = 0.2$ which occurs for $k = 0.4$. Integral gain k_i can be increased substantially by introducing derivative action. With higher values of k_d the maximum of k_i does, however, occur at an edge. Integral gain has its maximum $k_i = 0.9$ for $k = 0.925$ and $k_d = 2.86$. The performance is very sensitive to variations in the controller parameters at the maximum. Figure 6.24 shows that a marginal increase of proportional gain makes the system unstable. The controller that maximizes k_i also has other drawbacks, which are illustrated by the following example.

EXAMPLE 6.25—THE DERIVATIVE CLIFF
Consider a process with the transfer function

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

Maximizing integral gain k_i subject to the robustness constraint $M_s \leq 1.4$, gives the controller parameters $k = 0.925$, $k_i = 0.9$, and $k_d = 2.86$. The Nyquist plot of the loop transfer function is shown in Figure 6.25. Notice that the Nyquist curve has a loop. This will always occur when the maximum occurs where the envelope has an edge. The controller obtained has excessive phase lead, which is obtained by having a PID controller with complex zeros, $T_i < 4T_d$. In the particular case we have $T_i = 0.33T_d$. Time plots showing the response of

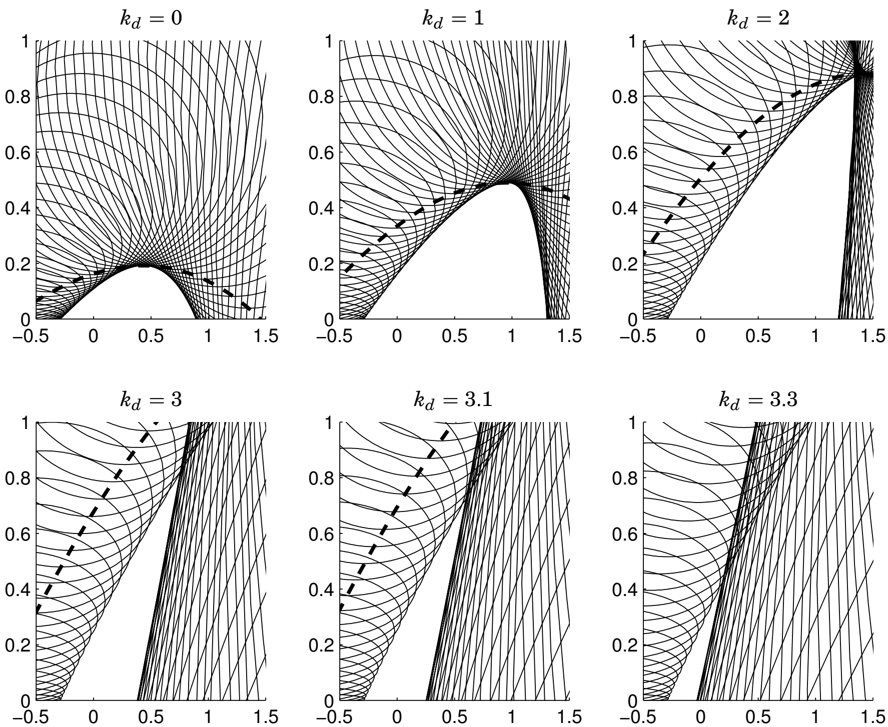

Figure 6.24 Cuts of the robustness region for constant derivative gain k_d. The curves are computed for PID control of the process $P(s) = 1/(s+1)^4$. Notice the sharp corners of the region for large k_d (the derivative cliff).

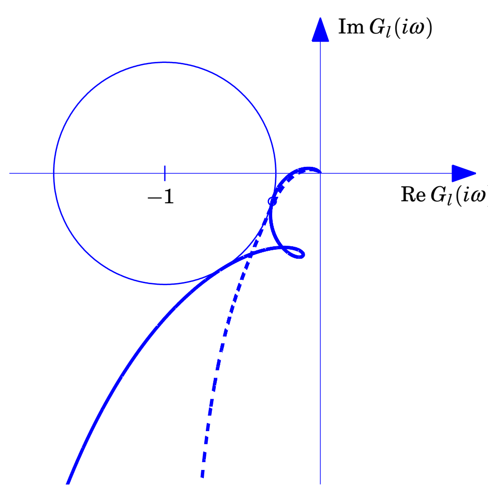

Figure 6.25 Nyquist curve of the loop transfer function for PID control of the process $P(s) = 1/(s+1)^4$, with a controller having parameters $k = 0.925$, $k_i = 0.9$, and $k_d = 2.86$.

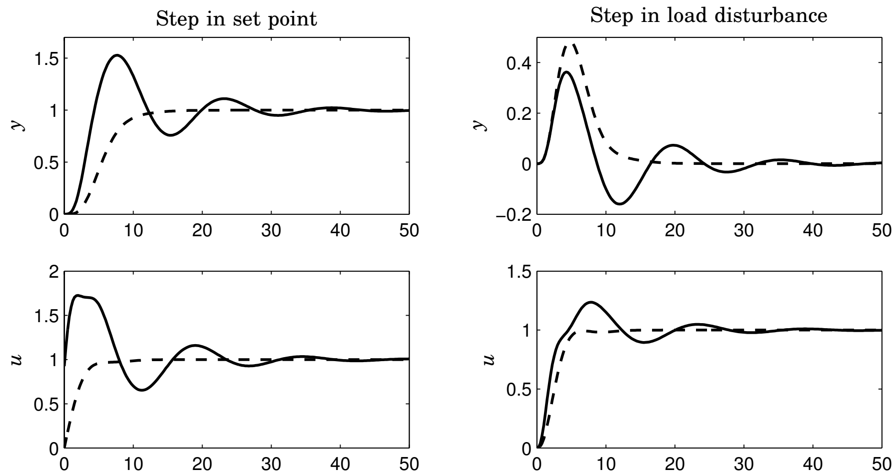

Figure 6.26 Time responses for PID control of the process $P(s) = 1/(s+1)^4$, with controller having parameters $k = 0.925$, $k_i = 0.9$, and $k_d = 2.86$ (solid lines) and $k = 1.1$, $k_i = 0.36$, and $k_d = 0.9$ (dashed lines).

the system to step changes in set point and load disturbances are shown in Figure 6.26. The responses are oscillatory.

For comparison we have shown Nyquist plots and time plots for a PID controller where $T_i = 4T_d$. The controller parameters are $k = 1.1$, $k_i = 0.36$, and $k_d = 0.9$. The responses of this controller are better, even if the peak in the response to load disturbances is larger. □

Avoiding the Derivative Cliff

There are several ways to modify the design problem to avoid the difficulties associated with the derivative cliff. One way is to introduce conditions that do not allow the Nyquist curve to have loops. Another alternative is to require that $T_i > \alpha T_d$. It has also been attempted to fix derivative gain to the value obtained by a PD controller. This does not eliminate the loops on the Nyquist curve in all cases. Maximization of k_i can also be replaced by maximizing the absolute integral error due to load disturbances.

The MIGO Method

After many attempts it has been found that a simple solution is to restrict derivative gain so that the maximum occurs at a point where $\partial k_i/\partial k = 0$. This avoids having a maximum at a ridge. The algorithm is straightforward.

ALGORITHM 6.2—MIGO DESIGN OF PID CONTROLLER

1. Fix derivative gain k_d. Find controller parameters by solving (6.44); then compute controller gains from (6.43).

2. Compute the value of M for a range of frequencies around ω^*, and test the robustness constraint $M \geq M_{crit}$.

3. Increase k_d until the largest value that satisfies the robustness constraint is obtained.

$\square$

A good initial value of integral gain is the value obtained for a PD controller. This particular design method is called MIGO (M constrained Integral Gain Optimization).

An Algorithm for a Controller in Series Form

It frequently happens that the MIGO design method gives controller parameters such that $T_i < 4T_d$. In Section 3.2 it was shown that such controllers cannot be implemented in series form. It is therefore of interest to have controllers where the parameters are constrained to $T_i \geq 4T_d$. When the ratio $n = T_i/T_d \geq 4$, the controller can be written as

$$C(s) = k\left(1 + \frac{1}{sT_i} + sT_d\right) = k'\frac{(T_i's+1)(T_d's+1)}{T_i's}, \tag{6.53}$$

where

$$\begin{aligned} k &= k'\frac{T_i' + T_d'}{T_i'} \\ T_i &= T_i' + T_d' \\ T_d &= \frac{T_i'T_d'}{T_i' + T_d'}. \end{aligned} \tag{6.54}$$

Introducing $n' = T_i'/T_d'$, it also follows that

$$n = \frac{(1+n')^2}{n'}. \tag{6.55}$$

Notice that $n' = 1$ corresponds to $n = 4$.

It follows from Equation 6.54 that $T_i = nT_d$ gives the following relation between the controller parameters

$$k_i = \frac{k^2}{nk_d}.$$

A simple algorithm for maximizing the integral gain of a PID controller with $T_i = nT_d$ subject to a robustness constraint will now be developed. We first make the observation that PID control of the process $P(s)$ gives the loop transfer function

$$G_l(s) = P(s)C(s) = k'\frac{(1+sT_i')(1+sT_d')}{sT_i'}P(s) = k'\frac{(1+sT_i')}{sT_i'}\Big((1+sT_d')P(s)\Big).$$

This is identical to the loop transfer function for PI control of the process

$$P'(s) = (1+sT_d')P(s).$$

Since there are efficient algorithms for PI control we obtain the following iterative algorithm.

Table 6.11 Controller parameters obtained by loop-shaping design with $M_s = 1.4$ for a process with the transfer function $P(s) = (s+1)^{-4}$.

Controller	K	k_i	k_d	b	T_i	T_d	IAE
PD	1.333	0	1.333	1	0	1	∞
PI	0.433	0.192	0	0.14	2.25	0	5.20
PID MIGO	1.305	0.758	1.705	0*	1.72	1.31	2.25
PID $T_i = 4T_d$	1.132	0.356	0.900	0.9	3.18	0.80	2.51

ALGORITHM 6.3—DESIGN OF PID CONTROLLER WITH $T_i = 4T_d$

1. Start by designing a PI controller for the process $P(s)$. This gives a controller with the integral time $T_i = k/k_i$. Set $T_1' = T_i/2$ and $j = 1$.
2. Design a PI controller for the process $P'(s) = (1 + sT_1')P(s)$. Let the integral time of the controller be T_i'. Set $T_{j+1}' = (T_j' + T_i')/2$ and repeat until T_j' converges to T'. Let the controller gain be k'.
3. The controller parameters are $k = 2k'$, $T_i = 2T'$ and $T_d = T'/2$.

□

Examples

The design method will be illustrated by two examples.

EXAMPLE 6.26—FOUR EQUAL LAGS
Consider a system with the transfer function

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

Table 6.11 summarizes properties of PD, PI, and PID controllers designed for $M_s = 1.4$. The PD controller was designed by maximizing proportional gain; the PI and PID controllers by maximizing integral gain. A PID controller with the additional constraint $T_i = 4T_d$ was also designed. Responses to set-point changes and load disturbances for are shown in Figure 6.27.

The PID controllers have better performance than the PI controller. Integral gain is 2 to 3 times larger and IAE a factor of 2 smaller. The controller PID MIGO has $T_i = 1.3T_d$. The table shows that performance is decreased when controller parameters are constrained to $T_i = 4T_d$. Notice that many commercial PID controllers have the constraint $T_i \geq 4T_d$ built in, because they are based on the series form; see (3.10).

The parameter b is calculated as described in Section 5.3. The calculation for the PID MIGO controller shows that the overshoot cannot be reduced sufficiently by using zero set-point weight, which is indicated by the entry 0* in the table. In this case it is recommended to use a proper feedforward design for a system with two degrees of freedom. Such a design can improve set-point response significantly. □

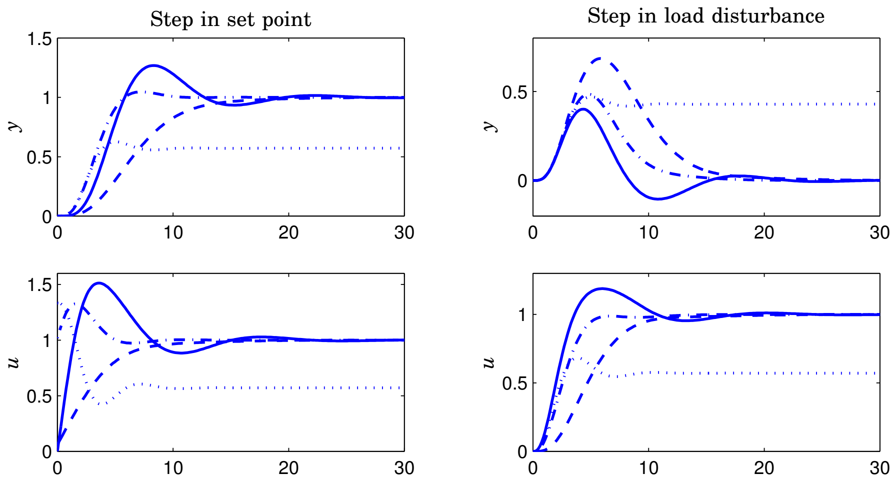

Figure 6.27 Responses for the system $P(s) = (s+1)^{-4}$ with the controllers in Table 6.11 to unit step changes in set point (left) and load disturbances (right). The dotted lines show responses with the PD controller, dashed with the PI controller, dash-dotted with the PID controller with parameters constrained to $T_i = 4T_d$, and solid lines with the PID controller designed using the MIGO method.

EXAMPLE 6.27—FOUR WIDELY DISTRIBUTED LAGS

Consider a system with the transfer function

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

Table 6.12 summarizes properties of PD, PI and PID controllers. All controllers were designed with the constraint that the maximum sensitivity is not larger than $M_s = 1.4$. The PD controller was designed by maximizing proportional gain, and the PI and PID controllers by maximizing integral gain. A PID with the additional constraint $T_i = 4T_d$ was also designed. Responses to set-point changes and load disturbances for the different controllers are shown in Figure 6.28.

Table 6.12 and Figure 6.28 show that derivative action improves performance drastically. The proportional gains of the controllers with derivative

Table 6.12 Controller parameters obtained by loop-shaping design with $M_s = 1.4$ for a process with the transfer function $P(s) = 1/(s+1)(0.1s+1)(0.01s+1)(0.001s+1)$.

Controller	K	k_i	k_d	b	T_i	T_d	IAE
PD	91.7	0	4.4	1	0	0.048	∞
PI	4.21	8.53	0	1	0.494	0	0.1044
PID MIGO	85.5	1488	3.87	0	0.057	0.045	0.00143
PID $T_i = 4T_d$	86.7	518	3.63	0.6	0.168	0.042	0.00143

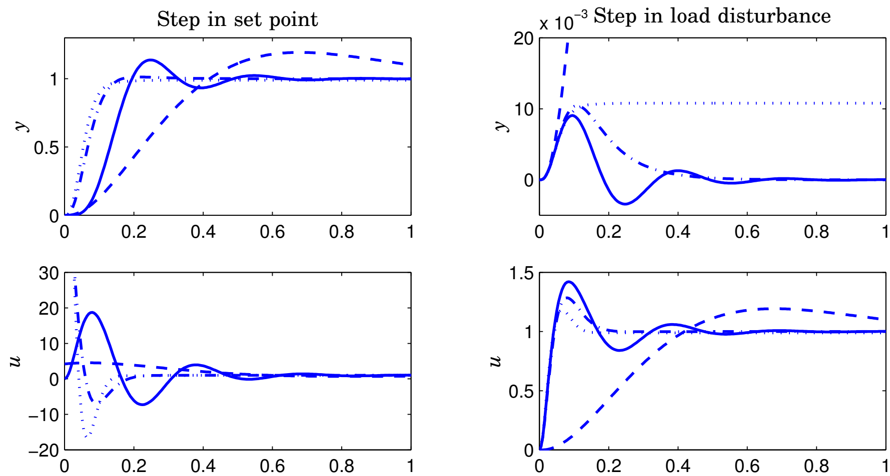

Figure 6.28 Responses for the system $P(s) = 1/(s+1))(0.1s+1)(0.01s+1)(0.001s+1)$ with the controllers in Table 6.12. The dotted lines show responses with the PD controller, dashed with the PI controller, dash-dotted with the PID controller with parameters constrained to $T_i = 4T_d$, and solid lines with the PID controller designed using the MIGO method. The load-disturbance response for the PI controller is out of scale.

action are around 90, while the PI controller has the gain 4.2. It follows from (4.40) that the largest peak of the load-disturbance response is around 0.01 for controllers with derivative action and about 20 times larger for PI control. The peak is so large that the load disturbance response for PI control is way outside the graph. The response time is also drastically increased when derivative action is used. The integral gains of controllers with derivative action are also much larger than for PI control. □

It is interesting to compare Examples 6.26 and 6.27. For the system with four equal lags in Example 6.26 the integral gain can be increased by a factor of 3 by introducing derivative action, while it can be increased by a factor of 200 for the system with distributed lags in Example 6.27. The main differences between the system is that the system in Example 6.27 is lag dominated; it has a normalized time delay $\tau = 0.07$. The system in Example 6.26 has $\tau = 0.33$. The normalized time delay is a good indicator for the benefits of derivative action. In Chapter 7 it will be shown that the large performance improvement with derivative action is possible for processes with small normalized time delays (lag dominated processes).

6.9 Summary

A number of techniques for designing PID controllers have been presented in this chapter, starting with methods of the Ziegler-Nichols type, where process dynamics were characterized by a few features that could be obtained from simple experiments. These methods have been very influential and have been used extensively by vendors. In spite of their popularity there are two drawbacks with the Ziegler-Nichols method, the fundamental assumption of quarter amplitude damping, which results in systems with very bad robustness, and the limited process knowledge used. Methods which avoids both difficulties will be developed in Chapter 7.

Standard methods for control system design can also be adapted to design of PID controllers. When using analytical techniques there is a correspondence between model and controller complexity, and it is necessary to approximate process dynamics by first- and second-order systems. Model reduction techniques are therefore necessary to apply pole placement to PID control. In particularly it is necessary to approximate the time delay. The unmodeled dynamics limit the performance that can be achieved, and the closed-loop poles that can be chosen.

Another way to use pole placement is to fix a pole configuration and to determine both the controller parameters and the magnitude of the poles. In this way it is possible to use second-order models to design PI controllers and third-order models to design PID controllers. Another way to apply pole placement to PID control is to place the dominant poles only. The advantage of this approach is that it can be applied to models of arbitrary order. A particular pole placement technique called lambda tuning, which has been used extensively in the process industry, is given particular attention.

A number of so-called algebraic design methods have also been discussed. In these methods the closed-loop transfer function is given and the controller parameters are obtained by algebraic calculations. The controller parameters can also be determined by optimization methods, where it is attempted to optimize criteria that specify performance subject to various constraints. There are many methods reflecting the richness of the control problem. Two methods, BO and SO, which are commonly used in motion control, have been given particular attention.

A novel design method developed by the authors and their coworkers is also presented. In this method it is attempted to optimize disturbance attenuation subject to constraints on robustness. The method gives a simple way to balance attenuation of load disturbances with the injection of measurement noise that is inevitable when feedback is used. Combining this method with set-point weighting, or more elaborate feedforward, gives a nice way to also achieve good response to set-point changes. The method can be viewed as an adaptation of robust design to PID control.

6.10 Notes and References

There is a very large literature on tuning of PID controllers. Good general sources are the books [Smith, 1972; Deshpande and Ash, 1981; Shinskey, 1988; McMillan, 1983; Corripio, 1990; Suda *et al.*, 1992; Oquinnaike and Ray, 1994; Marlin, 2000; Wang and Cluett, 2000; Wang *et al.*, 2000; Quevedo and Escobet, 2000; Cominos and Munro, 2002; Seborg *et al.*, 2004; O'Dwyer, 2003; Choi and Chung, 2004; Michael and Moradi, 2005]. The books clearly show the need for a variety of techniques, simple tuning rules, as well as more elaborate procedures that are based on process modeling, formulation of specifications, and control design. Even if simple heuristic rules are used, it is important to realize that they are not a substitute for insight and understanding. Successful controller tuning cannot be done without knowledge about process modeling and control theory. It is also necessary to be aware that there are many different types of control problems and consequently many different design methods. To only use one method is as dangerous as to only believe in empirical tuning rules. Control problems can be specified in many different ways. A good review of different ways to specify requirements on a control system is given in [Truxal, 1955; Maciejowski, 1989; Boyd and Barratt, 1991]. To formulate specifications it is necessary to be aware of the factors that fundamentally limit the performance of a control system.

The seminal papers [Ziegler and Nichols, 1942; Ziegler and Nichols, 1943] are the first attempts to develop systematic methods for tuning PID controllers. An interesting perspective on these paper is given in an interview with Ziegler; see [Blickley, 1990]. The CHR-method, described in [Chien *et al.*, 1952], is a modification of the Ziegler-Nichols method. This is one of the first papers where it is mentioned that different tuning methods are required for set-point response and for load disturbance response. Good response to load disturbances is often the relevant criterion in process control applications. Notice that the responses can be tuned independently by having a controller that admits a two-degree-of-freedom structure. The usefulness of a design parameter is also mentioned in the CHR-paper. In spite of its shortcomings, the Ziegler-Nichols method has been the foundation for many tuning methods; see [Tan and Weber, 1985; Mantz and Tacconi, 1989; Hang *et al.*, 1991]. Tuning charts were presented in [Wills, 1962b; Wills, 1962a; Fox, 1979].

The loop-shaping methods were inspired by classical control design methods based on frequency response; see [Truxal, 1955]. Applications to PID control are found in [Pessen, 1954; Habel, 1980; Chen, 1989; Yuwana and Seborg, 1982].

The idea of algebraic design was presented in [Truxal, 1955] and [Newton *et al.*, 1957] as a systematic method of design to given specifications; a more recent presentation is found in [Boyd and Barratt, 1991]. Algebraic design was applied to process control in [Smith, 1957; Atherton, 1999; Hansen, 2000]. The original papers on the λ-tuning method are [Dahlin, 1968] and [Higham, 1968]. The method is sometimes called the Dahlin method; see [Deshpande and Ash, 1981]. The method is very popular in the pulp and paper industry where it has been used to develop standardized tuning procedures; see [Sell, 1995] and [Anonymous, 1997]. Lambda tuning is closely related to the Smith predictor and the internal model controller; see [Smith, 1957; Chien, 1988; Chien and Fruehauf, 1990; Rivera *et al.*, 1986]. The tuning techniques developed in [Smith and Murrill, 1966; Pemberton, 1972a; Pemberton, 1972b; Smith *et al.*, 1975; Hwang and Chang, 1987] are other examples of the analytical approach to design. In [Rivera *et al.*, 1986] it was shown that internal model control reduces to PI and PID control when proper approximations of the time delay are done. A novel algebraic design method, described in [Hansen, 2000; Hansen, 2003] is used in a PID controller developed by Foxboro. An interesting feature is that the desired response is given as a high-order system.

The analytical tuning method gives controllers that cancel poles and zeros in the transfer function of the process. This leads to lack of observability or controllability. There are severe drawbacks in this as has been pointed out many times, e.g., in [Chien and Fruehauf, 1990; Shinskey, 1991b; Morari and Lee, 1991]. The response to load disturbances will be very sluggish for processes with lag dominated dynamics. A modification that does not cancel the process pole is given in [Chien and Fruehauf, 1990]. Skogestad's internal model controller is presented in [Skogestad, 2003]. This controller avoids the cancellation by an ad hoc modification of integral time for lag dominated dynamics.

Many methods for control design are based on optimization techniques. This approach has the advantage that it captures many different aspects of the design problem. There is also powerful software that can be used. A general discussion of the use of optimization for control design is found in [Boyd and Barratt, 1991]. The papers [Rovira *et al.*, 1969; Lopez *et al.*, 1969] give controllers that are optimized with respect to the criteria ISE, IAE, and ITAE. Other applications to PID control are given in [Hazebroek and van der Waerden, 1950; Wolfe, 1951; Oldenburg and Sartorius, 1954; van der Grinten, 1963; Lopez *et al.*, 1967; Marsili-Libelli, 1981; Yuwana and Seborg, 1982; Patwardhan *et al.*, 1987; Wong and Seborg, 1988; Polonoyi, 1989; Zhuang and Atherton, 1991]. The methods BO and SO were introduced in [Kessler, 1958a; Kessler, 1958b]. A discussion of these methods with many examples is found in [Fröhr, 1967; Fröhr and Orttenburger, 1982].

Pole placement is a straightforward algebraic design method much used in control engineering; see [Truxal, 1955]. It has the advantage that the closed-loop poles are specified directly. Many other design methods can also be interpreted as pole placement. The papers [Elgerd and Stephens, 1959; Graham and Lathrop, 1953] show how many properties of the closed-loop system can be deduced from the closed-loop poles. This gives good guidance for choosing the suitable closed-loop poles. An early example of pole placement is [Cohen and Coon, 1953; Coon, 1956a; Coon, 1956b]. It may be difficult to choose desired closed-loop poles for high-order systems. This is avoided by specifying only a few poles, as in the dominant pole design method described in [Persson, 1992; Persson and Åström, 1992; Persson and Åström, 1993].

The development of robust control was a major advance of control theory which made it possible to explicitly account for robustness in control design; see [Doyle *et al.*, 1992; Horowitz, 1993; Green and Limebeer, 1995], [Skogestad and Postlethwaite, 1996; Zhou *et al.*, 1996; Vinnicombe, 2000]. These ideas were applied to PID control in the papers [Panagopoulos *et al.*, 1997; Åström *et al.*, 1998; Panagopoulos and Åström, 2000; Panagopoulos, 2000; Kristensson, 2003]. The method discussed in Section 6.8 are based on these papers.

There are comparatively few papers on PID controllers that consider the random nature of disturbances. The papers [van der Grinten, 1963; Goff, 1966a; Fertik, 1975] are exceptions.

There are many papers on comparisons of control algorithms and tuning methods. The paper [McMillan, 1986] gives much sound advice; other useful papers are [Miller *et al.*, 1967; Gerry, 1987; Gerry, 1999].