CHAPTER 3

One-Degree of Freedom Internal Model Control

Objectives of the Chapter

- Present the one-degree of freedom (1DF) Internal Model Control (IMC) structure and describe its properties.
- Provide design methods for 1DF IMC systems with perfect models that give the best possible performance consistent with noisy measurements for inherently stable processes.

Prerequisite Reading

Chapter 2, "Continuous-Time Models"
Appendix A, "Review of Basic Concepts"
Appendix B, "Frequency Response Analysis"

3.1 INTRODUCTION

This chapter introduces methods for designing feedback controllers to force the output of an inherently stable process to (1) respond in a desired manner to a setpoint change, and (2) counter the effects of disturbances that enter directly into the process output. To enable us to carry out a quantitative controller design, we assume that we have a mathematical model of the process that allows us to predict how the process output (sometimes also called the process variable) responds to the control effort (e.g., how a process flow responds to the opening or closing of a valve) and to disturbances. Further, to keep these initial discussions as simple as possible, we assume that (1) the mathematical model is a perfect representation of the process, (2) the process is linear, and (3) there are no constraints on the control effort so it can take on any value between plus and minus infinity. Chapter 7 extends the results of this chapter to the case of imperfect models and uncertain processes. Chapter 5 shows how the controller designs obtained in this chapter and Chapter 4 can be implemented so as to accommodate control effort saturation. Chapter 6 shows how the IMC designs of this and the next chapter can be converted into PID controllers. While the treatment of nonlinear process models is beyond the scope of this text, many of the controller design concepts for linear models carry over to fairly broad classes of nonlinear process models (see, for example, Kravaris, 1987).

The IMC structure (Garcia & Morari, 1982) given in Figure 3.1 is central to our discussions on the design of controllers. Its conceptual usefulness lies in the fact that it allows us to concentrate on the controller design without having to be concerned with control system stability *provided that the process model* $\tilde{p}(s)$ *is a perfect representation of a stable process* $p(s)$.

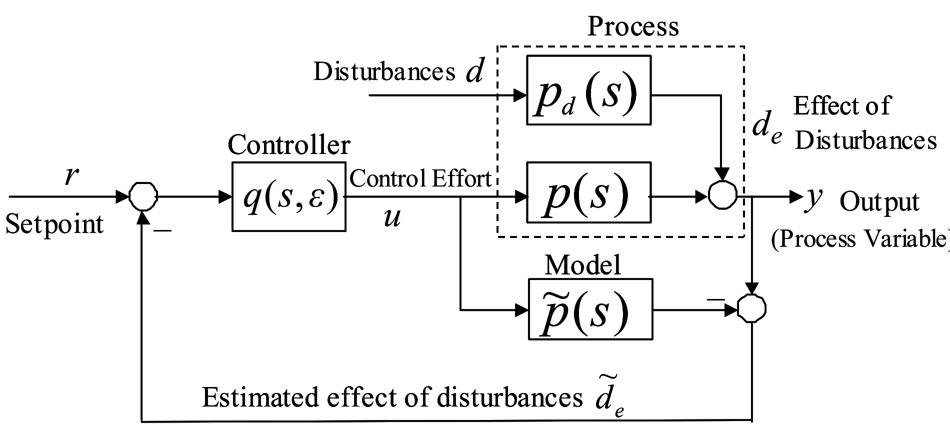

Figure 3.1 The IMC system.

As we shall see in Chapter 7, even if the model is imperfect, it is still possible to design the controller $q(s,\varepsilon)$ without concern for system stability, and then select the tuning parameter ε in $q(s,\varepsilon)$ to assure stability, provided only that the process $p(s)$ is inherently stable. In addition, if the controller gain is the inverse of the *model gain*, then the process output $y(s)$ will eventually reach and maintain the setpoint r (in the absence of new disturbances), provided only that the process and model gains have the same sign and that the controller is tuned so as to assure stability. For these reasons, the structure of Figure 3.1 is well suited for exploring ideal control system performance. Further, since the structure of Figure 3.1 can be rearranged into other structures, it can be used to obtain the controller for these other structures. We shall show how this is done for classical PID feedback control in Chapter 6.

While investigators have made use of concepts similar to those of IMC to design optimal feedback controllers since the late 1950s (Newton, Gould, & Kaiser, 1957), it was not until 1974 that the German investigator Frank first proposed utilizing the structure shown in Figure 3.1 to control processes. In 1979, Brosilow recognized that the IMC structure was at the core of both his inferential control system (Brosilow, 1979; Joseph & Brosilow, 1978a, 1978b; Tong & Brosilow, 1978) and the Smith Predictor (Smith, 1957), and proposed methods for designing the controller $q(s,\varepsilon)$. Morari and his coworkers, in a series of papers (Garcia & Morari, 1982, 1985a, 1985b; Morari, 1983, 1985; Morari, Skogestad, & Rivera 1984; Morari & Zafiriou, 1989), greatly expanded on the design methods for $q(s,\varepsilon)$ and placed the methodology in a sound theoretical framework. During this period, it also became clear that the IMC structure underlies the industrially important model predictive controllers known as IDCOM (Richalet, 1978), DMC (Cutler & Ramaker, 1979), and QDMC (Garcia & Morshedi, 1986; Prett & Garcia, 1988).

3.2 PROPERTIES OF IMC

3.2.1 Transfer Functions

An easy way to develop the transfer functions between the inputs d and r and the process output y is to first redraw Figure 3.1 as a simple feedback system, as shown in Figure 3.2, and then apply the following rule:

The transfer function between any input and the output of a single-loop feedback system is the forward path transmission from the input to the output divided by one plus the loop transmission for negative feedback.

For the feedback controller $c(s)$ of Figure 3.2, the rule gives

$$c(s) \equiv \frac{u(s)}{e(s)} = \frac{q(s)}{1 - q(s)\,\tilde{p}(s)}. \tag{3.1}$$

The negative term in the denominator of Eq. (3.1) arises from the positive feedback around $q(s)$.

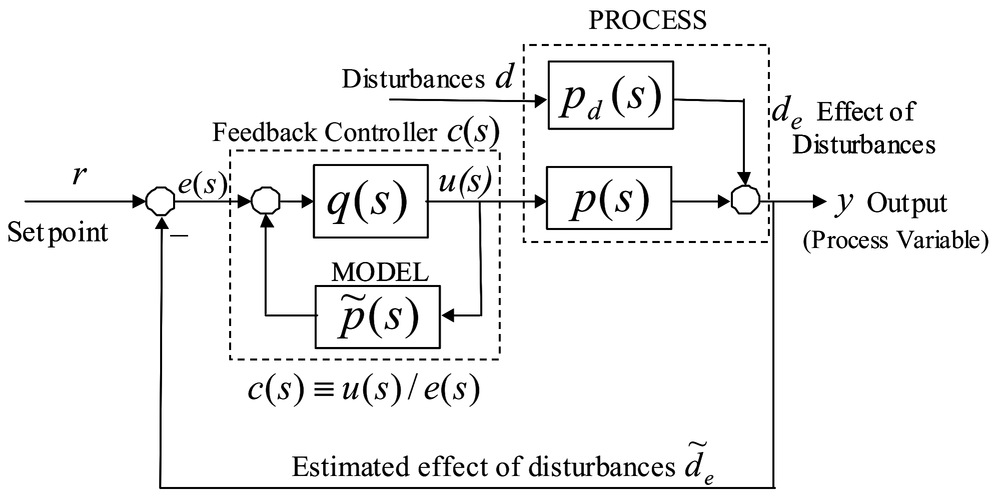

Figure 3.2 Alternate IMC Configuration.

The input-output relationships for Figure 3.2 are given by

$$\frac{y(s)}{r(s)} = \frac{p(s)c(s)}{1 + p(s)c(s)} \tag{3.2}$$

$$\frac{y(s)}{d(s)} = \frac{p_d(s)}{1 + p(s)c(s)} \tag{3.3}$$

$$\frac{u(s)}{r(s)} = \frac{c(s)}{1 + p(s)c(s)} = \left(\frac{y(s)}{r(s)}\right)p^{-1}(s) \tag{3.4}$$

$$\frac{u(s)}{d(s)} = \frac{-\,p_d(s)c(s)}{1 + p(s)c(s)} = \left(-\,\frac{y(s)}{d(s)}\right)c(s) \tag{3.5}$$

Substituting Eq. (3.1) into Equations (3.2) and (3.3) and clearing fractions gives

$$y(s) = \frac{p(s)q(s)r(s)}{1 + \left(p(s) - \widetilde{p}(s)\right)q(s)} \tag{3.6a}$$

$$y(s) = \frac{\left(1 - \widetilde{p}(s)q(s)\right)p_d(s)d(s)}{1 + \left(p(s) - \widetilde{p}(s)\right)q(s)} \tag{3.6b}$$

3.2.2 No Offset Property of IMC

The steady-state gain of any stable transfer function can be obtained by replacing the Laplace variable *s* with zero (see Chapter 2). If Equations (3.6a) and (3.6b) are stable, and if we choose the steady-state gain of the controller $q(0)$ to be the inverse of the model gain ($\widetilde{p}(0)q(0) = 1$), then the gain of the denominator of Equations (3.6a) and (3.6b) is $p(0)q(0)$. Thus the gain between the setpoint r(s) and $y(s)$ is one; the gain between the disturbance $d(s)$ and $y(s)$ is zero, and there is no steady-state deviation of the process output from the setpoint.

An ideal control system would force the process output to track its setpoint instantaneously and perfectly suppress all disturbances so that they do not affect the output. That is, the ideal controller would accomplish

$$y(s) = r(s) \tag{3.7a}$$

and

$$y(s)/d(s) = 0. \tag{3.7b}$$

From Equations (3.6a) and (3.6b), the above requires that

$$p(s)q(s) = 1 \quad \text{and} \quad \widetilde{p}(s) = p(s). \tag{3.8}$$

Thus, for perfect control, we need a perfect model, and from Eq. (3.8), the controller must perfectly invert that perfect model. Unfortunately, one never has a perfect model, and if the model has any dynamics at all (i.e., it is not just a gain), no controller can perfectly invert the process model. A controller can, however, come very close to inverting the process model. Just how close it can come is the subject of our next section, where we discuss controller design assuming perfect models. The design methodology always incorporates a tuning parameter that can slow down the control system response sufficiently to accommodate most modeling errors for inherently stable processes. Calculation of the tuning parameter for various descriptions of anticipated modeling errors is treated in Chapter 7.

The next section is limited to controller design of stable processes. IMC controller design for unstable processes is discussed in Chapter 4.

3.3 IMC DESIGNS FOR NO DISTURBANCE LAG

This section discusses the case where the disturbance lag $p_d(s)$ (c.f. Figure 3.2) is unity. For design methods when the disturbance lag is not one, see Chapter 4. The discussion begins by considering a process modeled by a first-order lag plus deadtime. Such models approximate the behavior of many chemical and petroleum processes. Since in this section we always assume that the model is perfect, we shall suppress the tilde notation.

Example 3.1 IMC Controller for an FOPDT Process

$$p(s) = \frac{Ke^{-Ts}}{\tau s + 1}; \; p_d(s) = 1 \tag{3.9}$$

The inverse of the process $p(s)$ given in Eq. (3.9) is

$$p^{-1}(s) = \frac{\tau s + 1}{K} e^{Ts}. \tag{3.10}$$

The only term in Eq. (3.10) that can be realized (i.e., be physically constructed) directly in a controller is the inverse of the process gain K. The term e^{Ts} represents an unrealizable prediction of future outputs,[1] while the term $(\tau s + 1)$ requires an unrealizable pure (i.e., unfiltered) differentiation of the process output. A real-time unfiltered differentiation of a continuous time signal is not realizable,[2] and even if it were, it would not be implemented because of unacceptable amplification of the noise on the measured process output y. Thus the best[3] that can be done is to implement the controller $q(s)$ for the process given by Eq. (3.9) as

$$q(s) = \frac{\tau s + 1}{K(\varepsilon s + 1)}, \tag{3.11}$$

where ε = a filter time constant or tuning parameter chosen to avoid excessive noise amplification and to accommodate modeling errors.

For modest or small modeling errors, the filter time constant ε will be less than the process time constant τ and the controller $q(s)$, given by Eq. (3.11), will be a lead network. That is, the frequency response of the controller will show that its magnitude increases from

---

[1] The inverse transform of $f(s)e^{Ts}$ is $f(t+T)$. The IMC controller of Figure 3.1 operates on the setpoint minus the disturbance estimate. Therefore if the controller has a term of the form e^{Ts}, implementation of the controller would require a prediction of the disturbance T units of time in the future, and an exact prediction is impossible unless we have a priori knowledge of the disturbance.

[2] Cannot be implemented by any physical device, such as a digital or analog computer.

[3] Best in the sense that as ε approaches zero, the output response to a step setpoint change approaches a step output change after one dead time. It is not possible to improve on such a response without a priori knowledge.

$1/|K|$ at low frequencies to $\tau/\varepsilon|K|$ at high frequencies, while the phase angle goes from zero at low frequencies to $\tan^{-1}\tau/\varepsilon$ at high frequencies ($\tan^{-1}\tau/\varepsilon$ approaches 90° as ε approaches zero).

The perfect model loop response transfer function is given by $pq(s)$ (see Eq. (3.8) with $p(s) = \tilde{p}(s)$). Using Equations (3.9) and (3.11) for $p(s)$ and $q(s)$ gives

$$y(s) = \frac{e^{-Ts}}{(\varepsilon s + 1)} r(s) + \left(1 - \frac{e^{-Ts}}{\varepsilon s + 1}\right) d(s) \cdot \tag{3.12}$$

The time response of y, given by Eq. (3.12), to a unit step change in setpoint r is shown in Figure 3.3 for $T = 1$ and $\varepsilon = .05$ and 1.0.

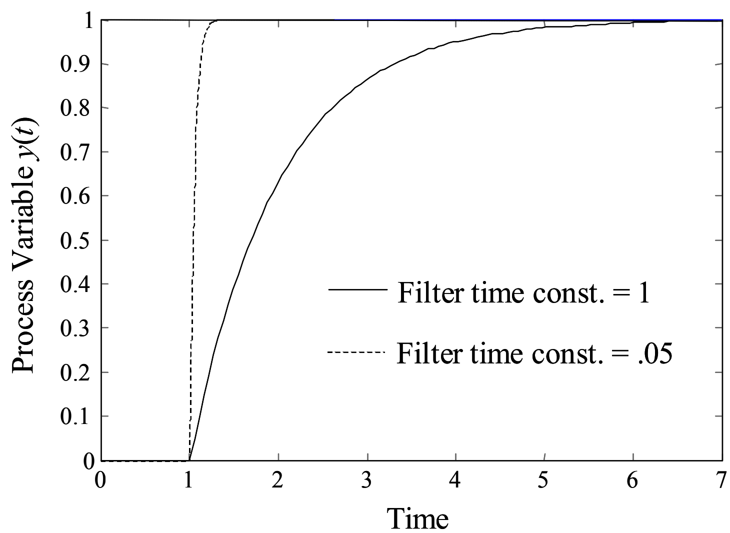

Figure 3.3 Perfect model IMC step response for first-order lag plus dead time.

The choice of the filter parameter ε in Eq. (3.12) depends on the allowable noise amplification by the controller and on modeling errors. Methods for choosing the filter time constant to accommodate modeling errors are discussed in Chapter 7. To avoid excessive noise amplification, we recommend that the filter parameter ε be chosen so that the high frequency gain of the controller is not more than 20 times its low frequency gain. For controllers that are ratios of polynomials, this criterion can be expressed as

$$|q(\infty)/q(0)| \leq 20. \tag{3.13}$$

The criterion given by Eq. (3.13) arises from the standard industrial practice of limiting the high frequency gain of a PID controller to no more than 20 times the low frequency controller gain, which is usually referred to simply as the controller gain (see Appendix A).

Factors of 5 and 10 are also frequently encountered in practice. While Eq. (3.13) limits only the infinite frequency gain of the controller, the complete design methodology presented later in this section limits the gain to less than 20 over all frequencies.

♦

The controller design method for the first order lag and dead time process generalizes easily to processes of the form

$$p(s) = \frac{N(s)}{D(s)} e^{-Ts}, \tag{3.14}$$

where $N(s)$ and $D(s)$ are polynomials in s.

The design of IMC controllers for Eq. (3.14) depends mainly on the characteristics of $N(s)$, as discussed in the following sections.

3.4 DESIGN FOR PROCESSES WITH NO ZEROS NEAR THE IMAGINARY AXIS OR IN THE RIGHT HALF OF THE *S*-PLANE

When $N(s)$ has no zeros in the right half of the s-plane or near the imaginary axis, the inverse of the model is stable and not overly oscillatory. In this case, the IMC controller for Eq. (3.14) can be chosen as

$$q(s) = \frac{D(s)}{N(s)\,(\varepsilon s + 1)^r}, \tag{3.15}$$

where r = the relative order[4] of $N(s)/D(s)$.

From Eq. (3.13), the filter time constant ε in Eq. (3.15) must satisfy

$$\varepsilon \geq \left(\lim_{s \to \infty} \frac{D(s)N(0)}{20 s^r N(s) D(0)} \right)^{1/r}. \tag{3.16}$$

As before, the limit given by Eq. (3.16) ensures that the high frequency gain of the controller is not more than 20 times its low frequency gain. The actual values of the filter time constant will more often be dictated by modeling errors and will be computed as shown in Chapter 7.

The form of the filter (i.e., $1/(\varepsilon s + 1)^r$) is somewhat arbitrary. It was chosen because it is the simplest form with a single adjustable parameter, ε, that provides an

[4] For transfer functions that are ratios of polynomials, relative order is defined as the order of the denominator minus the order of the numerator.

overdamped response[5] and makes q(s) realizable. Such a filter has the great merit of simplicity at the possible price of being suboptimal. There is also no incentive to use a filter order, r, greater than the minimum required to make the IMC controller realizable, because when there are modeling errors, higher order filters lead to slower responses, as shown in Chapter 7. Choosing a filter whose order is the same as the relative order of the model leads to a controller, $q(s)$, whose relative order is zero.

3.5 DESIGN FOR PROCESSES WITH ZEROS NEAR THE IMAGINARY AXIS

If the term $N(s)$ in Eq. (3.14) contains complex roots with low damping ratios,[6] then such terms can cause an IMC controller formed like that given in Equations (3.15) and (3.16) to amplify noise excessively at intermediate frequencies. There are two relatively simple options to reduce such excessive noise amplification. The first option is to increase the filter time constant sufficiently to reduce the peak to an acceptable level. This option generally requires large filter time constants, excessively increasing the settling time of the control system, and is therefore *not recommended***.** The second, and recommended, option is to not invert low damping ratio zeros. Rather, form a controller similar to that given by Eq. (3.15), but with the damping ratio in the original polynomials in $N(s)$ modified so as to be sufficiently large to avoid excessive noise amplification at intermediate frequencies. An example should help clarify the suggested design procedure.

Example 3.2 A Process with Low Damping Ratio Zeros

Consider the process given by

$$p(s) = \frac{s^2 + 0.001\,s + 1}{(s+1)^4}. \tag{3.17a}$$

By Equations (3.15) and (3.16), the controller and filter time constant would be

$$q(s) = \frac{(s+1)^4}{(s^2 + .001\,s + 1)\,(\varepsilon s + 1)^2} \tag{3.17b}$$

---

[5]Any transfer function of the form $1/\prod_i(\tau_i s+1)$ is said to yield an overdamped response because the frequency response of such transfer functions is everywhere less than one. Therefore, a plot of the time response of the output of the transfer function is smoother than its input.

[6]The damping ratio ζ is defined for second-order polynomials of the form $D(s) = \tau^2 s^2 + 2\zeta\tau s + 1$. If $0 \le \zeta < 1$, then the roots of this polynomial are complex conjugates. Damping ratios less than .4 (but greater than zero) are generally considered to be low because transfer functions of form 1/D(s) yield oscillatory responses to step inputs. A damping ratio of zero is characteristic of a system whose response to a step input oscillates continuously (i.e., without damping).

$$\varepsilon \geq 1/\sqrt{20} \cong .22. \tag{3.17c}$$

The frequency response of $q(s)$ given by Equations (3.17b) and (3.17c) is in shown Figure 3.4. The magnitude of the peak in Figure 3.4 is actually 3810. A filter time constant of 20 would be required to reduce the noise amplification at frequencies around 1.0 to a factor of 20. The settling time of the control system with such a controller exceeds 100 units. As we shall see, a controller with the same form as that given by Eq. (3.17b), but with a different damping ratio in the denominator, gives a much faster response.

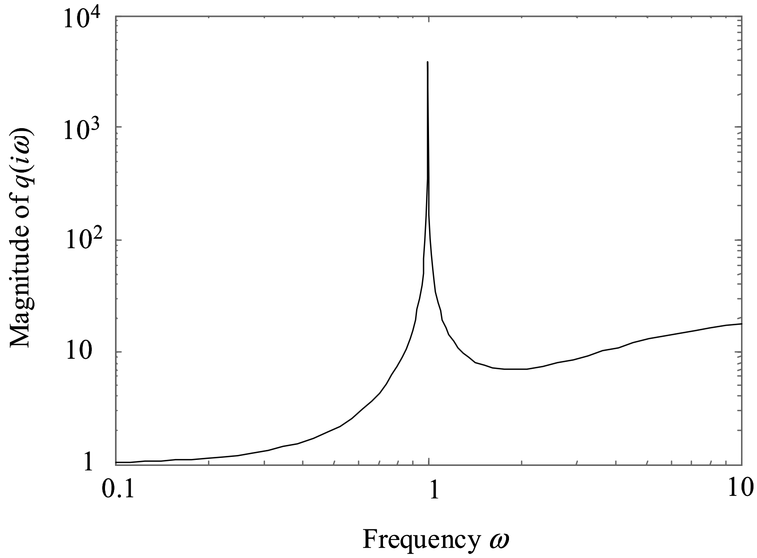

Figure 3.4 Frequency response of $q(s) = (s + 1)^4/((s^2 + 2\zeta s + 1)(.22s + 1)^2)$.

A better controller than that of Eq. (3.17b) is

$$q(s) = \frac{(s+1)^4}{(s^2+2\zeta s + 1)\,(.22s + 1)^2}. \tag{3.18a}$$

To reduce the controller frequency response peak so that its magnitude is only 20 (with $\varepsilon = .22$, as before) requires a damping ratio, ζ, of 0.1. The resulting loop response $pq(s)$ is

$$p(s)\,q(s) = \frac{s^2 + .001\,s + 1}{(s^2 + 2\zeta s + 1)\,(.22s + 1)^2}. \tag{3.18b}$$

Figure 3.5 shows the loop response given by Eq. (3.18b) for a damping ratio of 0.1 as well as a damping ratio of 0.5. As is apparent from the figure, a controller damping ratio of .5 gives a less oscillatory response than that given by a controller damping ratio of .1. On the other hand, the response for ζ = .5 is more sluggish than that for ζ = .1. In such cases the engineer has to use process knowledge to select the most appropriate controller

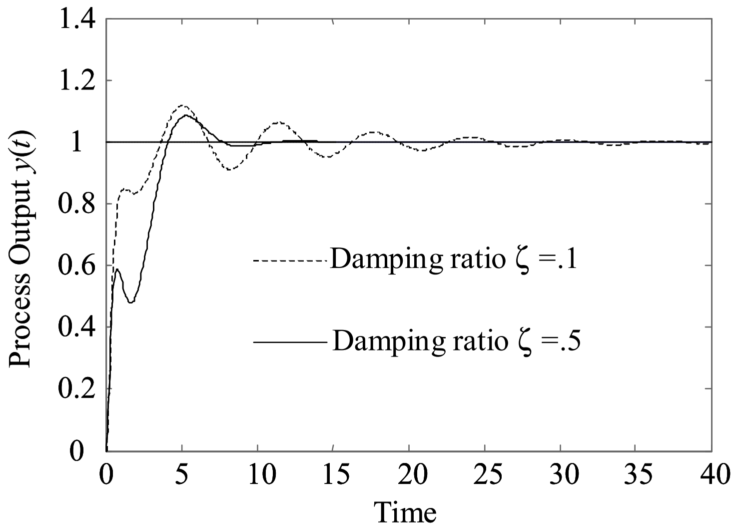

Figure 3.5 Perfect model loop response for $p(s) = (s^2 + .001s + 1)/(s + 1)^4$ with $q(s) = (s + 1)^4/((s^2 + 2\zeta s + 1)(.22s + 1)^2)$.

♦

3.6 DESIGN FOR PROCESSES WITH RIGHT HALF PLANE ZEROS

When $N(s)$ in Eq. (3.14) has factors of the form $(-\tau s+1)$ or $(\tau^2 s^2 - 2\tau\zeta s + 1)$, with τ and ζ greater than zero, its inverse is unstable. In this case the IMC controller cannot be formed as given by Eq. (3.15). The integral square error (ISE)[7] optimal choice of controller for such cases is to invert that portion of the model which has zeros in the left half plane and add poles at the mirror image of the right half plane zeros (Morari & Zafiriou, 1989). That is, we assume that the model given by Eq. (3.14) can be rewritten as

$$p(s) = \frac{N_-(s)\, N_+(s)}{D(s)}\, e^{-Ts}, \tag{3.19a}$$

[7] $ISE \equiv \int_0^\infty (y(t) - r(t))^2\, dt.$

where $N_{-}(s)$ contains only left half plane zeros, none of which have small damping ratios.

$N_{+}(s)$ contains only right half plane zeros, and can be written as

$$N_{+}(s) = \prod_{i,j} (-\tau_i s + 1)\ (\tau_j^2 s^2 - 2\tau_j \zeta_j s + 1) \qquad \tau_i, \tau_j > 0; \quad 0 < \zeta < 1. \tag{3.19b}$$

Notice that the gain of $N_{+}(s)$ is one.

Before designing the IMC controller, we strongly recommend that the model be put in time constant form (i.e., the numerator and denominator are factored into products of the form $(\pm \tau s + 1)$, $(\tau^2 s^2 \pm 2\tau\zeta s + 1)$ so that it is easy to form $N_{+}(s)$ and $N_{-}(s)$. The MATLAB functions *tcf* and *tfn* provided with IMCTUNE were developed specifically to put transfer functions into time constant form, and to facilitate their manipulation in this form. There are also other software programs that can be used to accomplish the desired factorization as described in Section 3.9.

The ISE optimal IMC controller for Eq. (3.19a) is

$$q(s) = \frac{D(s)}{N_{-}(s)\ N_{+}(-s)\ (\varepsilon s + 1)^r}, \tag{3.20}$$

where the zeros of $N_{+}(-s)$ are all in the left half plane and are the mirror images of the zeros of $N_{+}(s)$. r = relative order of $N(s)/D(s)$ as before.

The choice of controller given by Eq. (3.20) results in a loop response given by

$$y(s) = pq(s) = \prod_{i,j} \left(\frac{-\tau_i s + 1}{\tau_i s + 1} \right) \left(\frac{\tau_j^2 s^2 - \tau_j \zeta_j s + 1}{\tau_j^2 s^2 + \tau_j \zeta_j s + 1} \right) \frac{e^{-Ts}}{(\varepsilon s + 1)^r} \qquad \tau_i, \tau_j > 0; \quad 0 < \zeta_j < 1. \tag{3.21}$$

The loop response given in Eq. (3.21) is optimal in an ISE sense for a filter time constant ε of zero, and is suboptimal for finite ε. Also, when ε is zero, the loop transfer function given by Eq. (3.21) is called all-pass, since the magnitude of the frequency response is one over all frequencies.

Example 3.3 One Right Half Plane Zero

The process model is

$$p(s) = \frac{(s-1)}{27\ (s + 1/3)^3}. \tag{3.22a}$$

Putting Eq. (3.22a) in time constant form yields

$$p(s) = \frac{-1(-s+1)}{(3s+1)^3}. \tag{3.22b}$$

The IMC controller is

$$q(s) = \frac{-1(3s+1)^3}{(s+1)(\varepsilon s+1)^2}. \tag{3.22c}$$

The resulting loop response is

$$pq(s) = \frac{(-s+1)}{(s+1)(\varepsilon s+1)^2}. \tag{3.22d}$$

Figure 3.6 compares the step response of the ISE optimal loop transmission given by Eq. (3.21) with $\varepsilon = 0$ to step responses of suboptimal responses obtained by increasing and decreasing the controller time constant. Notice that the faster response obtained with $pq(s) = (-s+1)/(.5s+1)$ comes at the expense of a more negative initial response. Thus, for this simple example, the ISE optimal response is also qualitatively the best compromise between a more sluggish response and a faster response with a more negative initial response.

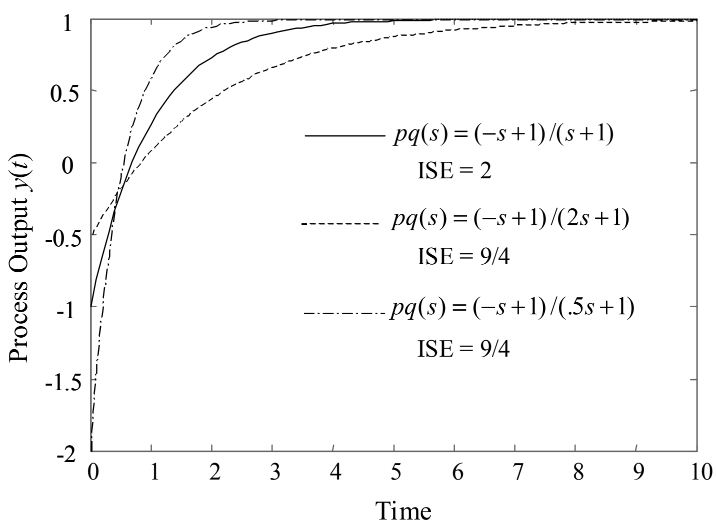

Figure 3.6 Response of processes with one right half plane zero.

The transfer function given by the controller of Eq. (3.22c) does not result in an optimal response to a step setpoint change unless ε is zero. We could get closer to an optimal transfer function by selecting the IMC controller as

$$q(s) = \frac{(3s+1)^3}{((1-2\varepsilon)s+1)(\varepsilon s+1)^2} \quad \text{for } \varepsilon \le 0.5. \qquad (3.22e)$$

The controller given by Eq. (3.22e) was obtained by forcing the coefficient of the linear term of its expanded denominator to be one, which is the same as the linear term in the denominator of Eq. (3.22c) when ε is zero. The loop response then becomes that given by Eq. (3.22f).

$$p(s)q(s) = \frac{(-s+1)}{((1-2\varepsilon)s+1)(\varepsilon s+1)^2} = \frac{(-s+1)}{((1-2\varepsilon)\varepsilon^2 s^3 + (2-3\varepsilon)\varepsilon s^2 + s + 1)}\,. \qquad (3.22f)$$

The cubic and quadratic terms in s in the denominator of Eq. (3.22f) are small relative to its linear term so that Eq. (3.22f) approaches the optimal transfer function given by Eq. (3.22d) with ε equal to zero. Notice, however, that Eq. (3.22e) is valid only for ε ≤ .5. For larger values of the filter time constant, Eq. (3.22e) is not stable.

The approach used to obtain the controller given by Eq. (3.22e) can be used to develop nearly optimal controllers for arbitrary nonminimum[8] phase processes. However, such controllers will generally be useful only for small filter time constants.

♦

Example 3.4 Two Right Half Plane Zeros

When the initial process response to a step is in a direction opposite to that of the final steady state, as in the previous example, the process is said to exhibit an inverse response. Processes with an odd number of right half plane zeros exhibit inverse responses. Processes with an even number of right half plane zeros do not have inverse responses, since the initial value at time zero plus is always in the direction of the steady-state[9] as shown in Figure 3.7 for the loop response given by

$$pq(s) = (-s+1)^2/(\tau s+1)^2 \qquad (3.23)$$

[8] A minimum phase transfer function has only zeros in the left half of the s-plane (i.e., the zeros are all negative). A non-minimum phase transfer function has one or more zeros in the right half of the s-plane. The terminology arises from the fact that changing the time constants of the numerator from positive to negative always results in increasing the phase lag of the transfer function at all frequencies. For example, the phase lag of $(s + 1)/(2s + 1)$ is always less than the phase lag of $(-s + 1)(2s + 1)$. A deadtime is often called a nonminimum phase element because it cannot be inverted and is in that way similar to a nonminimum phase transfer function, which also does not have a stable inverse.

[9] For a process with an odd number of right half plane zeros, the sign of the coefficient of the highest order s term is opposite to that of the zeroth order coefficient. For a process with an even number of right half plane zeros, the sign of the coefficient of the highest order s term is the same as that of the coefficient of the zeroth order coefficient. Therefore, by the initial value theorem, the sign of the first non-zero derivative is opposite to, or the same as, the sign of the steady-state gain in the case of an odd or even number of right half plane zeros, respectively.

with $\tau = .5$, 1, and 2.

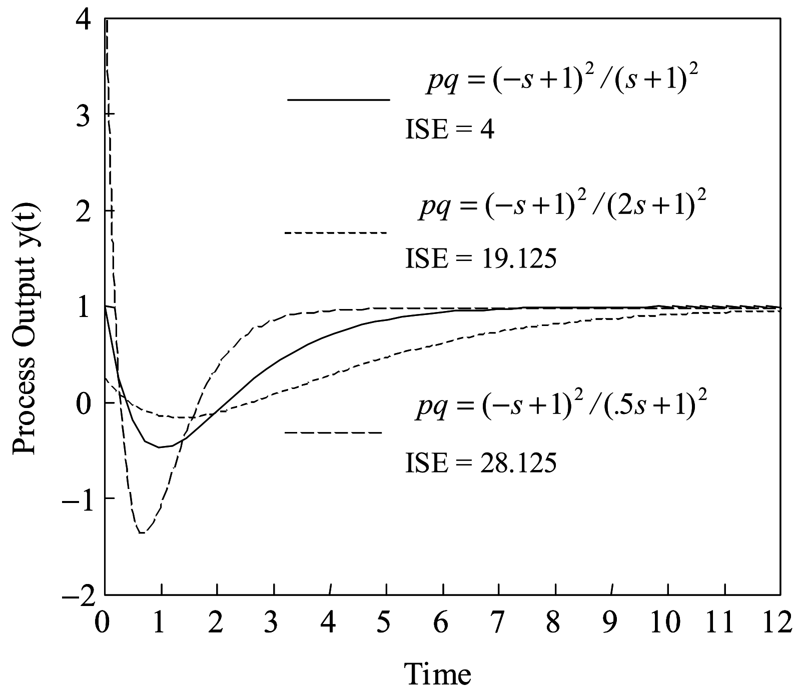

Figure 3.7 Response of processes with two right half plane zeros.

Notice that the ISE optimal response in Figure 3.7 is again that which also gives the qualitatively best compromise between a sluggish response and a response with too much initial overshoot.

♦

3.7 PROBLEMS WITH MATHEMATICALLY OPTIMAL CONTROLLERS

In both Figure 3.6 and Figure 3.7 the qualitatively and quantitatively optimal response is that given by the all-pass loop response. However, it can happen that the all-pass response, while ISE optimal, is not qualitatively the best response and may not even be an acceptable response, as shown by Example 3.5 following.

Example 3.5 A Process with an Infinite Number of RHP Zeros

The transfer function given by Eq. (3.24) below is one that might arise between a process variable and a control effort in a multivariable system, where one or more of the other process variables are under closed-loop control.

$$p(s) = \left(1 + 3e^{-s}/(s+1)\right)/(s+1) \tag{3.24}$$

$p(s)$ has two complex zeros in the right half plane at $0.214 \pm 2.10i$. These zeros were calculated by replacing the dead time in Eq. (3.24) by a five over five Padé[10] approximation and then factoring the resulting sixth order numerator of $p(s)$. The factored form of the approximated transfer function $\hat{p}(s)$ is

$$\hat{p}(s) = \frac{4(.225\,s^2 - .0964\,s + 1)\,(.0156\,s^2 + .0348\,s + 1)\,(.00235\,s^2 + .0616\,s + 1)}{(s+1)^2\,(.137\,s + 1)\,(.0175\,s^2 + 235\,s + 1)\,(.0138\,s^2 + .128\,s + 1)}. \tag{3.25}$$

The ISE optimal controller for (3.25) is

$$q(s) = \frac{.25(s+1)^2\,(.137s + 1)\,(.0175s^2 + .235s + 1)\,(.0138s^2 + .128s + 1)}{(.225s^2 + .0964s + 1)\,(.0156s^2 + .0348s + 1)\,(.00235s^2 + .0616s + 1)\,(\varepsilon s + 1)}. \tag{3.26}$$

To achieve $\left|q(\infty)/q(0)\right| \le 20$ requires $\varepsilon \ge 0.2$ The resulting loop response is

$$\hat{p}(s)q(s) = \frac{(.225s^2 - .0964s + 1)}{(.225s^2 + .0964s + 1)\,(\varepsilon s + 1)}. \tag{3.27}$$

Figure 3.8a shows the response of Eq. (3.27) to a step with $\varepsilon = 0$ and $\varepsilon = 0.2$. The oscillatory response in Figure 3.8a is due to the low damping ratio in the denominator of Eq. (3.27), which is only 0.101. Increasing the damping ratio to 0.5 in the first term of the denominator of Eq. (3.26) by increasing the coefficient of 0.0964 to 0.48 gives a qualitatively better response, that is, one with less overshoot and no oscillations. However, this improved response has a *higher* ISE. The associated controller and loop response are given by

$$q(s) = \frac{.25(s+1)^2\,(.137s + 1)\,(.0175s^2 + .235s + 1)\,(.0138s^2 + .128s + 1)}{(.225s^2 + .48s + 1)\,(.0156s^2 + .0348s + 1)\,(.00235s^2 + .0616s + 1)\,(.2s + 1)} \tag{3.28}$$

---

[10] A Padé approximation to the exponential e^{-Ts} is a ratio of polynomials of order m in the numerator, and n in the denominator, whose coefficients are chosen so that the ratio of polynomials approximates the exponential to within terms of order $n + m + 1$ in s. That is, the Maclaurin series expansion in s of the exponential and its Padé approximation agree through terms of order $m + n$. Padé approximations to the exponential e^{-Ts} of degree n over n (i.e., of order accuracy $2n$) for any $n > 0$ are available in the control toolbox of MATLAB and in Program CC.

$$\hat{p}(s)q(s) = \frac{(.225s^2 - .0964s + 1)}{(.225s^2 + .48s + 1)(.2s + 1)}. \tag{3.29}$$

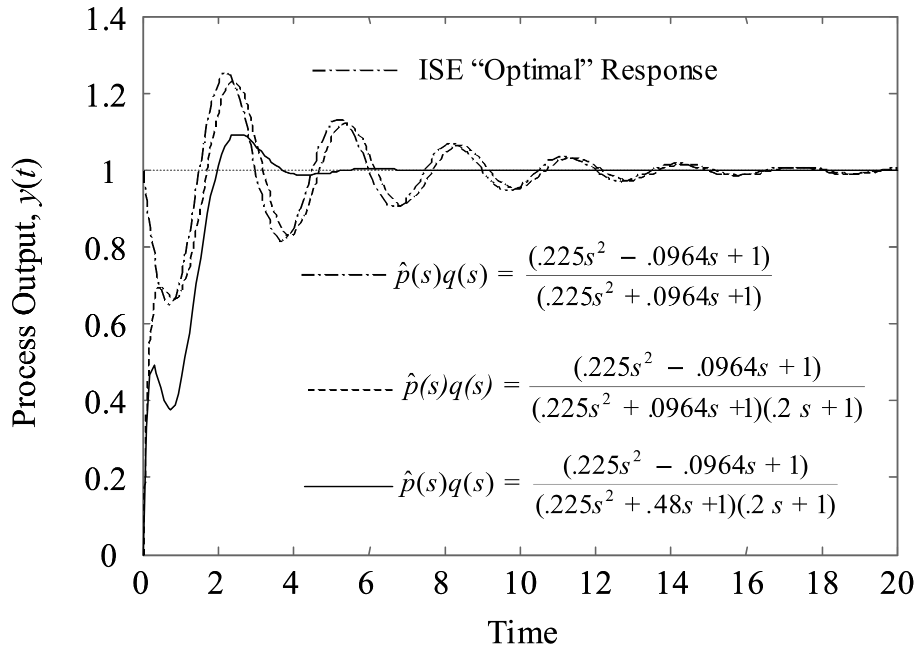

Figure 3.8a Optimal versus suboptimal responses to a step setpoint change for the process, $p(s) = (1 + 3e^{-s}/(s + 1))/(s + 1)$.

The responses in Figure 3.8a were obtained with a rather complex controller, as given by Eq. (3.26) and its modifications. It is reasonable to ask whether there is a much simpler suboptimal controller that might do nearly as well. Indeed, the control system that uses the simple controller given by

$$q(s) = \frac{.25(s+1)}{(.05s + 1)} \tag{3.30}$$

produces a very good step setpoint response, as shown in Figure 3.8b. The above controller arises from the recognition that the numerical value of the process steady-state gain is four, that only the term $1/(s + 1)$ in $p(s)$ is easily invertible, and that the step response of $p(s)$ is monotonic. Figure 3.8b compares the loop responses using the controllers given by Eq. (3.26) with $\varepsilon = 0$, with Equations (3.28) and (3.30).

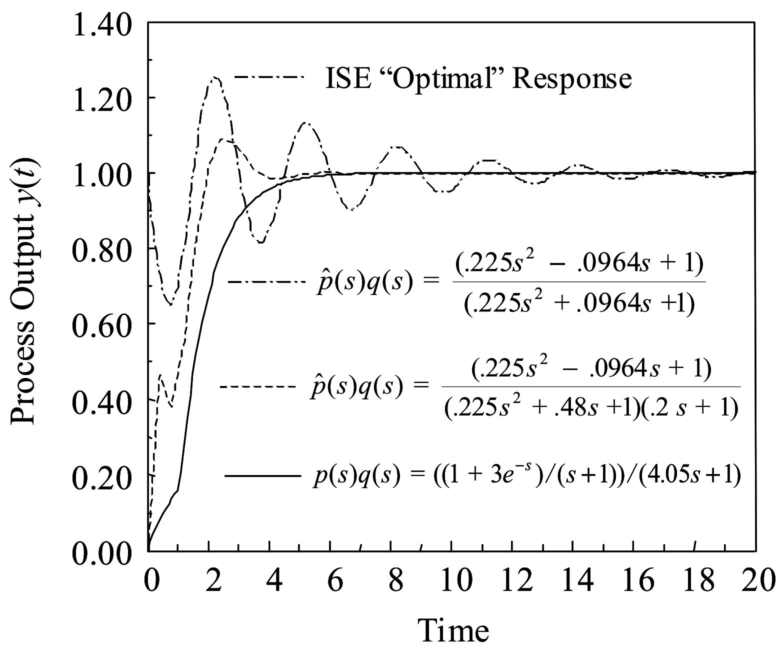

Figure 3.8b Optimal versus suboptimal responses to a step setpoint change for the process, $p(s) = (1 + 3e^{-s}/(s + 1))/(s + 1)$

♦

The controller given by Eq. (3.30) is not only simpler than that given by Eq. (3.26), but also yields a response with no overshoot and achieves the final steady state in about the same time. Many engineers, including the authors, will prefer the response obtained with the simple controller to that obtained with the more complex controller. Thus, this example demonstrates that optimality, like beauty, is in the eye of the beholder. As the foregoing examples attempt to demonstrate, controller designs that minimize the ISE can be quite useful, but they should not be applied by rote. The engineer should always look at the ideal loop response given by the loop transmission $pq(s)$ to judge whether the optimal controller gives desirable responses. If not, then a trial and error procedure is usually necessary to obtain a controller that does give an acceptable loop behavior. A systematic way to obtain such a design is to start with a controller $q(s)$ that is just the inverse of the model gain and then attempt to improve on that controller by including in it increasing portions of the model inverse.

Lest the reader get the misimpression that qualitative improvements in performance over that obtained with the ISE optimal controller are only possible when the optimal response is oscillatory, we offer the following example.

Example 3.6 A Process with an Infinite Number of RHP Zeros

$$p(s) = 1 - Ke^{-Ts} \; ; \; |K| > 1 \tag{3.31}$$

The above process has an infinite set of zeros in the right half plane at $(\ell nK - 2n\pi i)/T$ if $K > 1$, and at $(\ell n(-K) - (2n+1)\pi i)/T$ if $K < -1$. We can form a controller $q(s)$ with poles at the mirror image of the zeros of Eq. (3.31) as

$$q(s) = \frac{1}{(e^{-Ts} - K)}, \tag{3.32}$$

so that

$$p(s)q(s) = \frac{1 - Ke^{-Ts}}{e^{-Ts} - K}. \tag{3.33}$$

The response of Eq. (3.33) to a unit step is

$$y(t) = y(nT) \; ; \quad nT \le t < (n+1)\,T$$
$$y(nT) = 1 - (1 + 1/K)\,(1/K)^n. \tag{3.34}$$

For the case $K = 2$, the first several values of $y(nT)$ are $y(0) = -1/2$, $y(T) = 1/4$, $y(2T) = 5/8$, and $y(3T) = 13/16$, as shown in Figure 3.9 for $T = 1$. Clearly, the optimal response approaches unity in a staircase fashion.

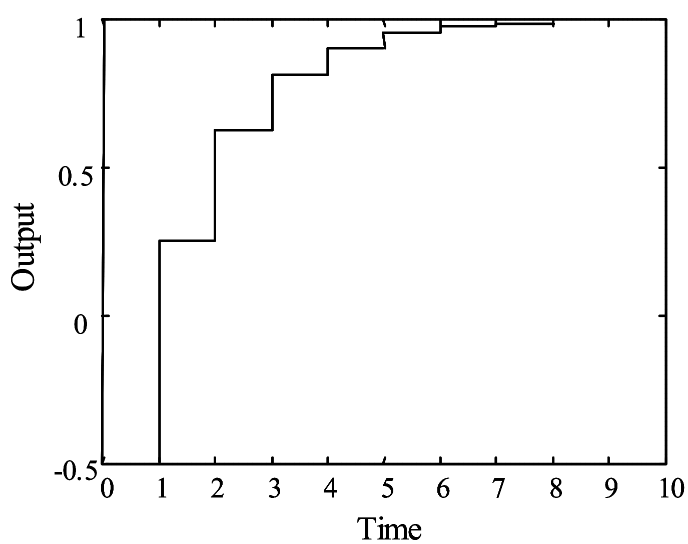

Figure 3.9 Loop response of Eq. (3.33) to a unit step input.

The ISE for the output $y(nT)$ of Eq. (3.33) is given by

$$ISE = T\sum_{n=0}^{\infty}(1 - y(nT))^2.$$

Substituting Eq. (3.34) into the above gives

$$ISE = T\sum_{n=0}^{\infty}(1 + 1/K)^2(1/K^2)^n.$$

Recognizing that $\sum_{n=0}^{\infty}(1/K^2)^n = \dfrac{1}{1 - 1/K^2}$ gives

$$ISE = T(1 + 1/K)^2/(1 - 1/K^2). \tag{3.35}$$

For $K = 2$, the ISE from Eq. (3.35) is 3T. Now consider the controller that is simply the reciprocal of the model gain; that is

$$q(s) = \frac{1}{1 - K}. \tag{3.36}$$

The loop response for this controller is

$$y(t) = y(nT)\ ;\ nT \le t \le (n+1)T \tag{3.37}$$

$$y(0) = -1$$

$$y(nT) = 1,\ n = 1, 2, \ldots\infty.$$

The ISE for the above is $4T$ (vs. $3T$ before), but the output reaches and maintains its desired value after one dead time.

♦

3.8 Modifying the Process to Improve Control System Performance

Another interesting feature of the process in Example 3.6 is that the response can be substantially improved by modifying the process. Let us rewrite Eq. (3.31) as

$$p(s) = e^{-T_0 s} - Ke^{-Ts};\ |K| > 1 \tag{3.38a}$$

If T_0 in Eq. (3.38a) is taken as zero, then Eq. (3.38a) is the same as Eq. (3.31). If, however, T_0 is taken as T, then Eq. (3.38a) becomes

$$p^*(s) = e^{-Ts} - Ke^{-Ts} = (1-K)\,e^{-Ts}, \tag{3.38b}$$

where p* = the modified process.

Now, our simple controller, given by Eq. (3.36), gives

$$\begin{aligned} y(t) &= 0; \quad 0 \le t < T \\ y(t) &= 1; \quad T \le t. \end{aligned} \tag{3.38c}$$

The above has an ISE of T, which is substantially better than that given by Eq. (3.35) using the optimal controller given by Eq. (3.32). Recall that when $K = 2$, the optimal ISE was $3T$. This shows that small design changes to the process can lead to significant improvements in control system performance. Notice that the process given by Eq. (3.31) responds instantaneously to a step change in control effort. However, the process given by Eq. (3.38a), with $T_0 = T$, responds to a step change in control effort only after a dead time of T units. Thus, we have improved control system performance by, in effect, increasing the deadtime of the process in going from Eq. (3.31) to Eq. (3.38a). For more information on how to improve performance by process modifications that *increase* certain dead times, see Psarris & Floudas (1990).

3.9 Software Tools for IMC Design

An important aid in the design of IMC controllers is the ability to display the time response of the loop transmission pq as in Figures 3.5 to 3.8. This can be done quite conveniently using Program CC, MATLAB with SIMULINK or Vissim, or similar types of software. We provide a MATLAB version 5.3.1 (or later) suite of m-files IMCTUNE.m on the web site for this text[11] that allows the user to enter the process model $\tilde{p}(s)$ and the portion of the model $p_m(s)$, inverted by the IMC controller in a convenient form. The program yields the output time response as one of its options under the Results and Simulations menu. IMCTUNE is also capable of producing PID controllers and of tuning the IMC controller to accommodate process uncertainty as described in Chapters 6 and 7. IMCTUNE requires the Optimization and Control System toolboxes in order to be used with MATLAB and SIMULINK.

To simplify the IMC design task, it is often convenient to convert a transfer function into time constant form wherein the numerator and denominator are factored into products of the form $(\pm\,\tau s + 1)$, $(\tau^2 s^2 \pm 2\tau\zeta s + 1)$. This can be done in Program CC simply by the command *tcf,g*, where *g* is the transfer function that is desired in time constant form. MATLAB m-files, *tcf.m* and *tfn.m*, provided with the IMCTUNE software, provide the same functionality. Information on how to use *tcf.m* and *tfn.m* can be obtained by typing *help*, followed by the function name (e.g., *help tcf*).

[11] http://www.phptr.com/brosilow/

3.10 SUMMARY

This chapter presents IMC design techniques for inherently stable linear processes where the disturbance enters directly into the process output (i.e., the disturbance lag is either unity or has very fast time constants relative to the process time constants). The controller design methods differ depending on the location of the zeros of the numerator of the process transfer function.

Controllers for processes with no right half plane zeros, or zeros near the imaginary axis, can be obtained simply by inverting the entire model except any multiplicative delay. That is, if $\tilde{p}(s) = g(s)e^{-Ts}$ then $q(s) = g^{-1}(s)f(s)$,

where $f(s) = 1/(\varepsilon s + 1)^r$,

r = the relative order of $g(s)$,

ε = an adjustable filter time constant.

We recommend choosing the filter time constant ε to avoid excessive high frequency noise amplification by using the criterion

$$|q(\infty)/q(0)| \le N$$

where N is between 10 and 20.

If the term $g(s)$ has complex zeros with small damping ratios, then the controller, q(s), might amplify noise more than is desirable at some midrange frequency (i.e., $|q(i\omega_c)/q(0)| > \mathrm{N}$ for some ω_c). When this occurs, we recommend either (1) not inverting the zeros that cause the noise amplification, or (2) increasing the damping ratio of the denominator terms in the controller $q(s)$ that arise from inverting zeros with small damping ratios.

If the term $g(s)$ has right half plane zeros, then the ISE optimal IMC controller inverts all of $g(s)$ except the right half plane zeros, and poles are added to the controller q(s) so that the loop transmission $\tilde{p}(s)q(s)$ is an all-pass system cascaded with the filter f(s). That is, the poles of $\tilde{p}(s)q(s)$ are at the mirror image of the right half plane zeros of $\tilde{p}(s)q(s)$. As usual, the filter order is the relative order of the model, $g(s)$.

Finally, if the ISE optimal IMC controller described in the previous paragraph is overly complex (as might occur when the right half plane zeros of the model are due to transcendental terms such as $(p_1(s) + p_2(s)e^{-Ts})$, where $p_1(s)$ and $p_2(s)$ are polynomials in s, or if the loop response is undesirable because of excessive oscillations or overshoot, then we recommend rejecting the ISE optimal IMC controller in favor of a simpler controller. A systematic method of obtaining such a controller is to start with a controller that is just the inverse of the model gain, then attempt to improve the loop response $\tilde{p}(s)q(s)$ by including in the controller increasing portions of the model inverse.

Problems

3.1 Derive all the transfer functions from Eq. (3.1) to Eq. (3.6).

3.2 Use SIMULINK or other simulation software to obtain the step responses for each of the examples in the chapter. Note that only the transfer functions of the loop response need be simulated, since all of the examples are for perfect model responses to a step setpoint change.

3.3 Simulate a step disturbance response for Example 3.1 with $p_d(s) = 1$. How does this response relate to the step setpoint response?

3.4 Find an IMC controller for each of the following process models. The controller must satisfy

$$\max_{\omega}\left|\frac{q(i\omega)}{q(0)}\right| \le 20$$

with equality holding at some frequency (possibly infinity). State the rationale for your choice of the part of the model that the controller inverts. Recall the statement in Section 3.6 that "*Before designing the IMC controller, we strongly recommend that the model be put in time constant form.*"

a. $$\frac{s^2+2s+.25}{s^4+6.5s^3+15s^2+14s+4}$$

b. $$\frac{32s^2+.8s+2}{s^4+6.5s^3+15s^2+14s+4}$$

c. $$\frac{(16s^2+.4s+1)}{16s^4+40.5s^3+18.25s^2+3s+1}$$

d. $$\frac{(s-1)e^{-s}}{s^2+s+1}$$

e. $$3\frac{d^2y}{dt}+4\frac{dy}{dt}+3y(t)=u(t)-\frac{d}{dt}u(t)$$

f. $$\frac{(s^2+2s+.25)e^{-s}}{s^4+6.5s^3+15s^2+14s+4}$$

g. $$\frac{(s-1)(s^2-s+1)}{s^4+6.5s^3+15s^2+14s+4}$$

h. $$\frac{s^3-1.9s^2+.8s-2}{(s^2+.2s+1)(3s+1)^2(s+1)}$$

3.5 Find an IMC controller for each of the following process models. The IMC controllers should satisfy the same noise amplification criterion as that for Problem 3.4. Note, however, that the inverse of the numerator in each of the following process models can be inverted using a simple negative feedback loop that has unity in the forward path and Ke^{-Ts} in the feedback path, with appropriate values of K and T. This is not to say, however, that such inverses will not be very oscillatory, or even unstable. Thus, in developing an IMC controller for each of the following, the student should decide whether or not to invert the numerator exactly; approximate it by replacing the exponential with a Padé approximation, and then invert all or part of the numerator; or perhaps simply invert the numerator gain. To complicate matters further, the filter time constant necessary to satisfy the same noise amplification criterion as that for Problem 3.4 depends on which option is selected, and if a Padé approximation is used, it even depends on whether highest order terms in the Padé polynomials are odd or even. We suggest that before designing a controller for the following problems, the student get a step response of the process model.

The student should note that, as in most engineering situations, there is no single correct answer. In such cases the simplest controller is often the best controller. To compare controllers, we recommend that the student simulate the output response to a step setpoint change.

a. $$\frac{1+0.5e^{-3s}}{(s+2)^2(s+.5)^2}$$

b. $$\frac{1+.95e^{-3s}}{(s+2)^2(s+.5)^2}$$

c. $$\frac{1-.95e^{-3s}}{(s+2)^2(s+.5)^2}$$

d. $$\frac{1+2e^{-s}}{s+1}$$

3.6 Why do the IMC controller filter time constants in the IMC controllers for all the processes of Problem 3.5 change depending on whether an even or odd order Padé approximation for the exponential term is used in forming the controller?

References

Brosilow, C. B. 1979. “The Structure and Design of Smith Predictors from the Viewpoint of Inferential Control.” *Joint Automatic Control Conference*, June 17, Denver.

Cutler, C. R., and B. L. Ramaker. 1980. “Dynamic Matrix Control—A Computer Control Algorithm.” AIChE National Meeting, Houston, TX; also *Proc. Joint Automatic Control Conf.*, San Francisco.

Garcia, C. E., and A. M. Morshedi. 1986. “Quadratic Programming Solution of Dynamic Matrix Control (QDMC).” *Chemical Engineering Communications* 46, 73–87.

Garcia, C. E., and M. Morari. 1982. “Internal Model Control –1. A Unifying Review and Some New Results.” *Ind. Eng. Chem. Process Des. & Dev*. 21, 308–323.

Joseph, B., and C. B. Brosilow. 1978a. “Inferential Control of Processes, Part I, Steady-state Analysis and Design.” *AIChE J*. 24(3), 485–492.

Joseph, B., and C. B. Brosilow. 1978b. “Inferential Control of Processes, Part III, Construction of Optimal and Suboptimal Dynamic Simulators,” *AIChE J.24* (3), 500–509.

Kravaris, C., and C. B. Chung. 1987. “Nonlinear State Feedback Synthesis by Global Input-Output Linearization.” *AIChE J*., 33, 592–603.

MATLAB with SIMULINK, © The MathWorks, *www.mathworks.com*.

Morari, M. 1983. “Design of Resilient Processing Plants-III: A General Framework for the Assessment of Dynamic Resilience.” *Chem. Eng. Sci*., 38, 1881–1891.

Morari, M., S. Skogestad, and D E. Rivera. 1984. “Implications of Internal Model Control for PID Controllers.” *Proc. of American Control Conference*, San Diego, CA, 661–666.

Morari, M. 1985. “Robust Stability of Systems with Integral Control.” *IEEE Trans. Autom. Control*, AC-30, 574–577.

Morari, M., and F. Zafiriou. 1989. *Robust Process Control.* Prentice-Hall, NJ.

Newton, G. C., L. A. Gould, and J. F. Kaiser. 1957. *Analytic Design of Feedback Controls*, John Wiley & Sons, NY.

Psarris, A., and C. Floudas. 1990. “Improving Dynamic Operability in MIMO Systems with Time Delays.” *Chem. Eng. Sci*. 45, 3505–3524.

Prett, D. M., and C. E. Garcia. 1988. *Fundamental Process Control.* Butterworths, MA.

Program CC, © Systems Technology, *www. systemstech.com/ progcc1.htm*.

Richalet, J. A. 1978. “Model Predictive Heuristic Control: Applications to an Industrial Process.” *Automatica* 14, 413–428.

Smith, O. J. M. 1975. “Closer Control of Loops with Dead Time.” *Chem. Eng. Progress* 53(5).

Tong, M., and C. B. Brosilow. 1978. "Inferential Control of Processes, Part II, The Structure and Dynamics of Inferential Control Systems." I. 24(3), 492–500.

VisSim, © Visual Solutions, *www.vissol.com*.