1-1 What Is a Feedback Controller?

Consider the simple process shown in Fig. 1-1. The level in the tank is to be maintained "near" a target value by manipulating the valve on the inlet stream. Now, place the "as-yet-undefined" controller in Fig. 1-2. The controller must sense the level and decide how to adjust the valve. Notice that for the controller to work properly

1. There must be a way of measuring the tank level (the "level sensor") and a way of transmitting the measured signal to the controller.
2. Equally important there must be a way of transmitting the controller decision or controller output to the valve.
3. At the valve there must be a way of converting the controller output signal into a mechanical movement to either close or open the valve (the "actuator").

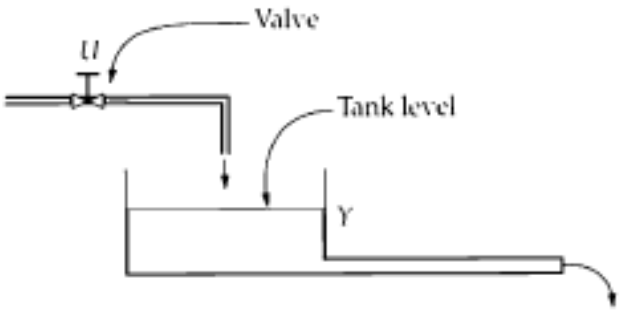


Figure 1-1 A tank of liquid (a process).

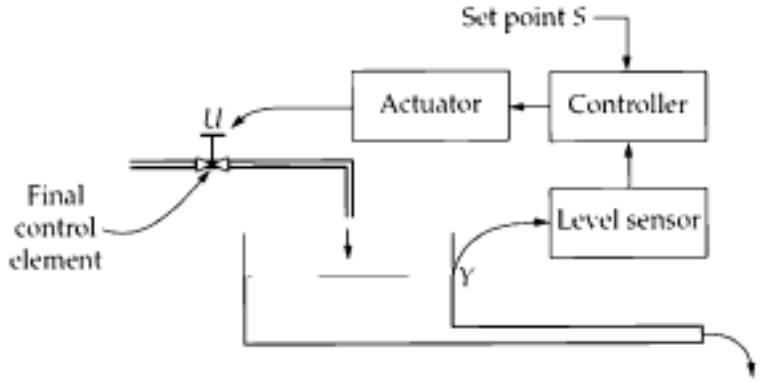


Figure 1-2 A tank of liquid with a controller added.

An abstract generalization of the above example is shown in Fig. 1-3, which is a schematic block diagram. The lower box represents the process (the tank of liquid) The input to the process is *U* (the valve position on the inlet pipe). The output is *Y* (the tank level). The process

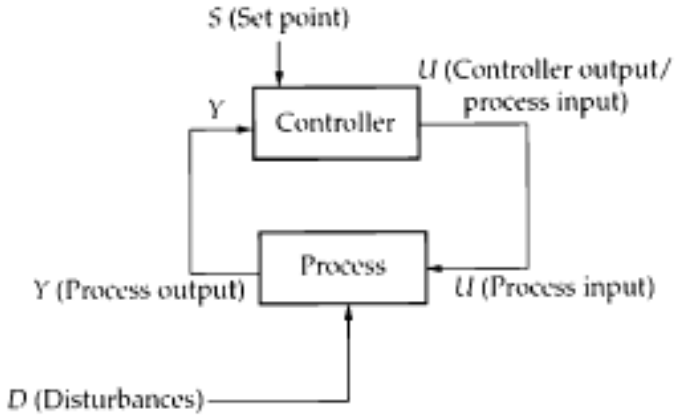


Figure 1-3 Block diagram of a control system.

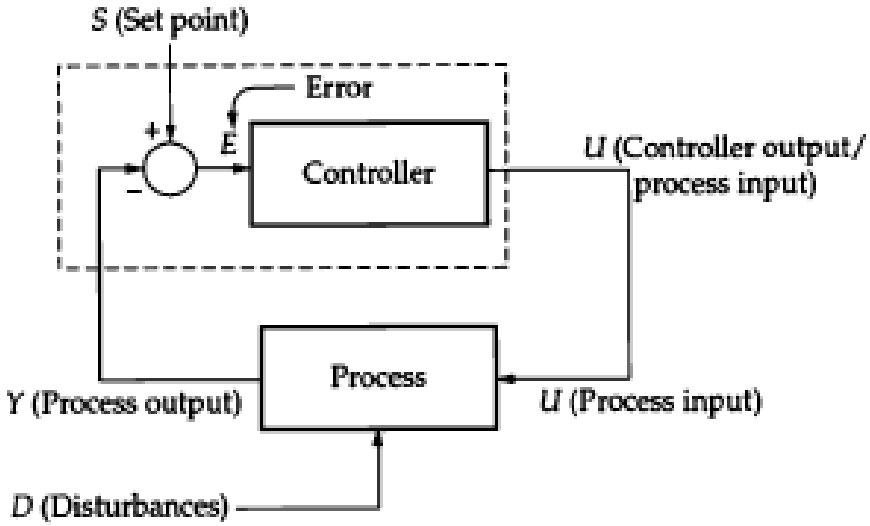


Figure 1-4 Block diagram of a control system showing the error.

1-2 What Is a Feedforward Controller?

Before getting into a deep discussion of a feedforward controller, let's develop a slightly modified version of our tank of liquid. Consider Fig. 1-5, which shows a large tank, full of water, sitting on top of a large hotel (use your imagination here, please). This tank is filled in the same manner as the one in the previous figures. However, this tank supplies water to the sinks, toilets, and showers in

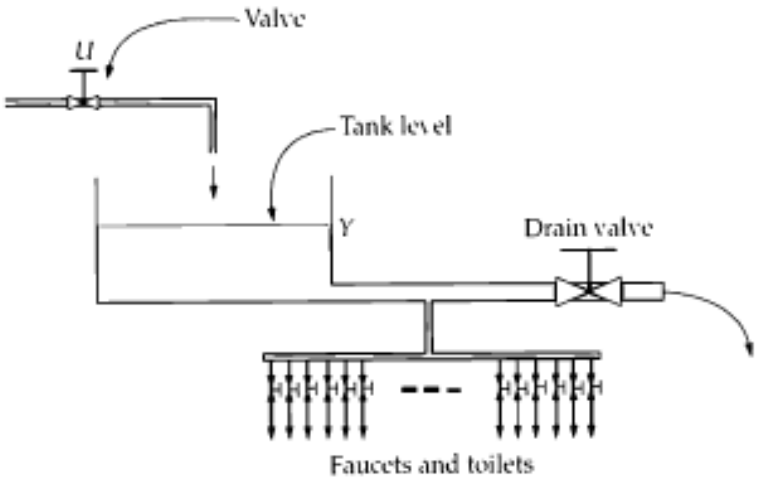


Figure 1-5 Large hotel water tank.

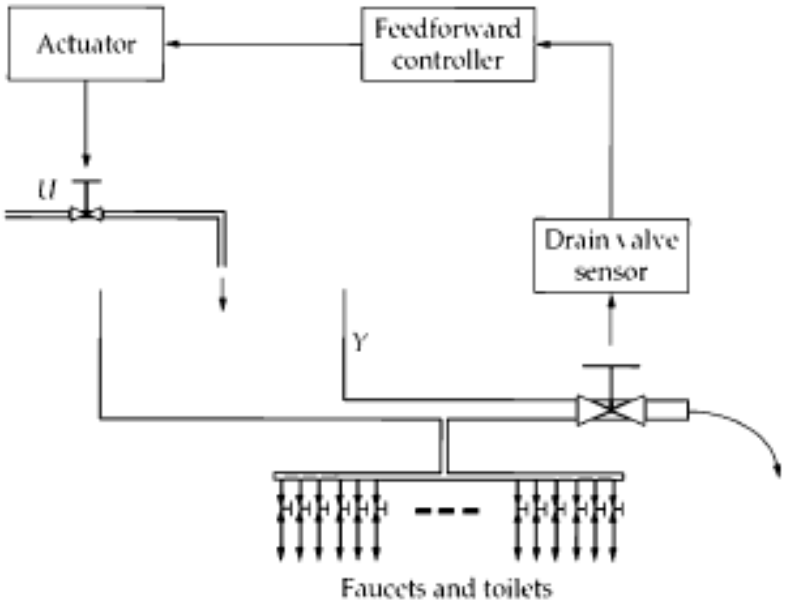


Figure 1-6 A feedforward controller.

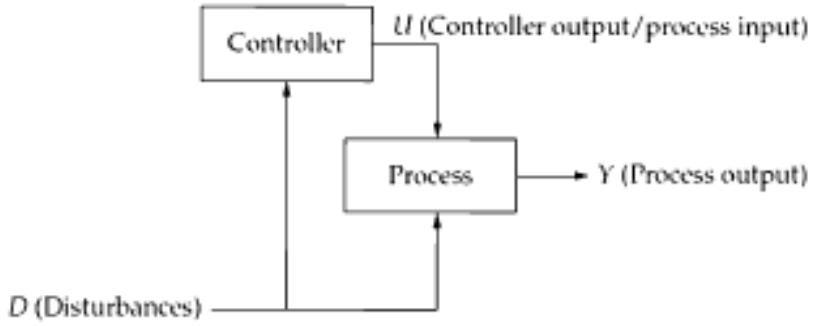


Figure 1-7 Feedforward controller block diagram.

1-3 Process Disturbances

Referring back to Fig 1-5, the tank on the hotel roof, let's spend some time discussing the impact of the faucets, the toilet flushings, and the drain valve on the tank level. First, consider the response of the tank level to a step change in the drain valve position. That is, we suddenly crank the drain valve from its initial constant position to a new, say more open, position and hold it there indefinitely. Figure 1-8 shows the response This kind of a disturbance is considered *deterministic* because one would usually know the exact time and amount of the valve adjustment.

1-5 Combining Feedforward and Feedback Controllers

Figure 1-11 shows how feedforward and feedback controllers can be combined for our hotel example and Fig. 1-12 shows an abstraction of

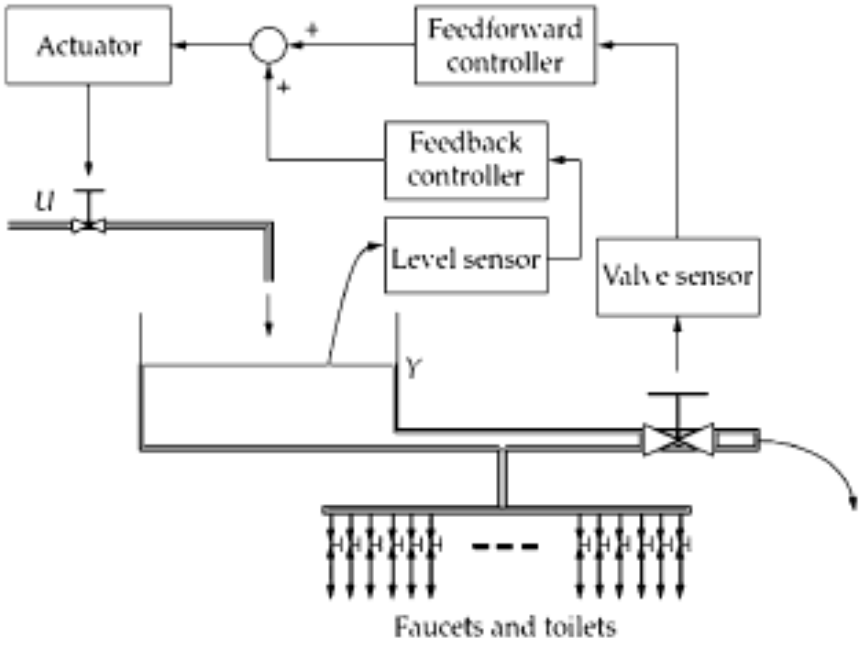


Figure 1-11 A feedforward/feedback controller.

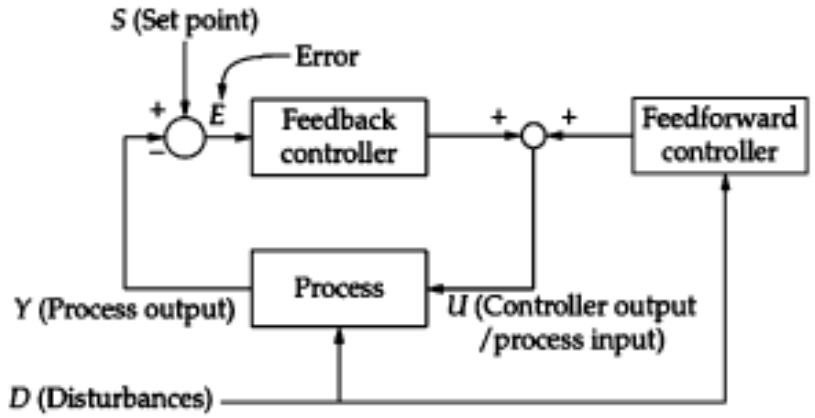


Figure 1-12 A feedforward/feedback controller block diagram.

1-7 An Example of Controlling a Noisy Industrial Process

To illustrate the impact of feedback control on noisy processes, consider a molten glass delivery forehearth shown in Fig. 1-14. Since the reader may not have a glass-manufacturing background, a little

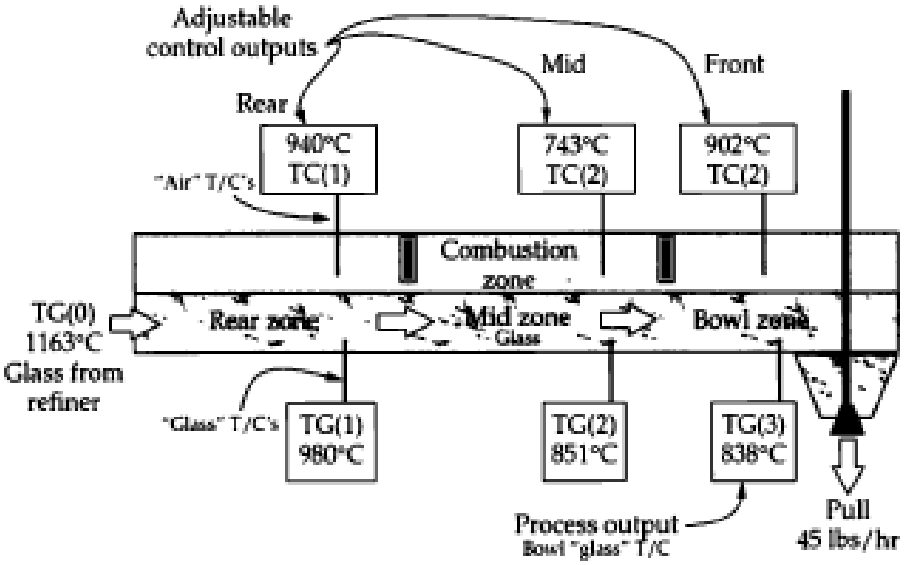


Figure 1-14 A molten glass forehearth.

explanation of the process depicted in Fig. 1-14 is necessary. The forehearth is a rectangular duct made of refractory material about 1 ft wide, about 16 ft long, and about 6 in deep. Molten glass at a relatively high temperature, here 1163°C, enters the forehearth from a so-called refiner. The forehearth is designed to cool the glass down to a suitable forming temperature, in this case 838°C. There is a gas combustion zone above the glass where the energy loss from the glass is controlled by maintaining the gas (*not the glass*) temperatures at desired values via controllers, the details of which we will gloss over for the time being.

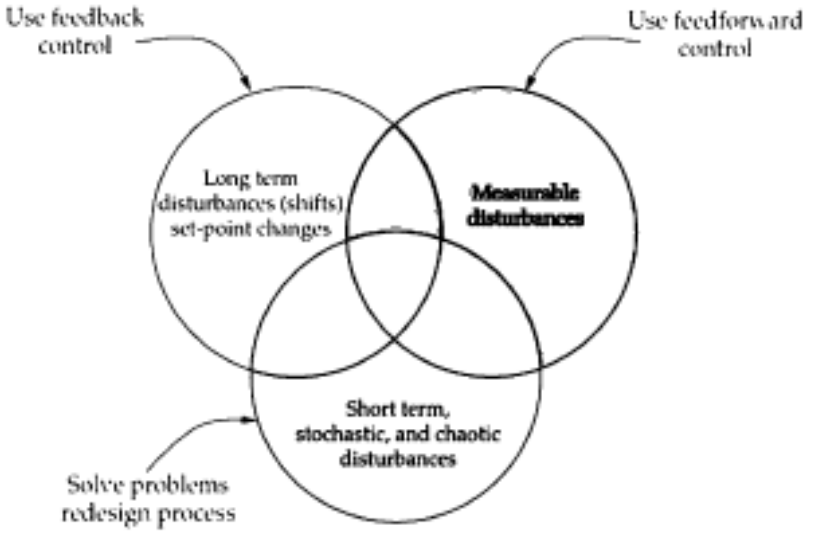


Figure 1-18 Different approaches for different problems.

Finally, if the main challenge is trying to maintain a process output satisfactorily near a set point in the face of persistent stochastic disturbances then the best approach probably should be the formation of a problem-solving team to deal with both the process and the environment.

1-8 What Is a Control Engineer?

So far we have implied that a control engineer designs control algorithms. In fact, the title of control engineer can mean many things. The following list, in no particular order, covers many of these "things":

1. Installer of control/instrumentation equipment (sometimes called an "instrumentation engineer"): In my experience this is the most prevalent description of a control engineer's activities. In this case, the actual design of the control algorithm is usually quite straightforward. The engineer usually purchases an off-the-shelf controller, installs it in an instrumentation panel, probably of her design, and then proceeds to make the controller work and get the process under control. This often is not trivial. There may be control input sensor problems. For example, the input signal may come from a thermocouple in an electrically heated bath of some kind and there may be serious common and normal mode voltages riding on the millivolt signal representing the thermocouple value. There may be control output actuator problems. There may be challenging process dynamics problems, which require careful controller tuning. In many ways, instrumentation engineering can be the most challenging aspect of control engineering.

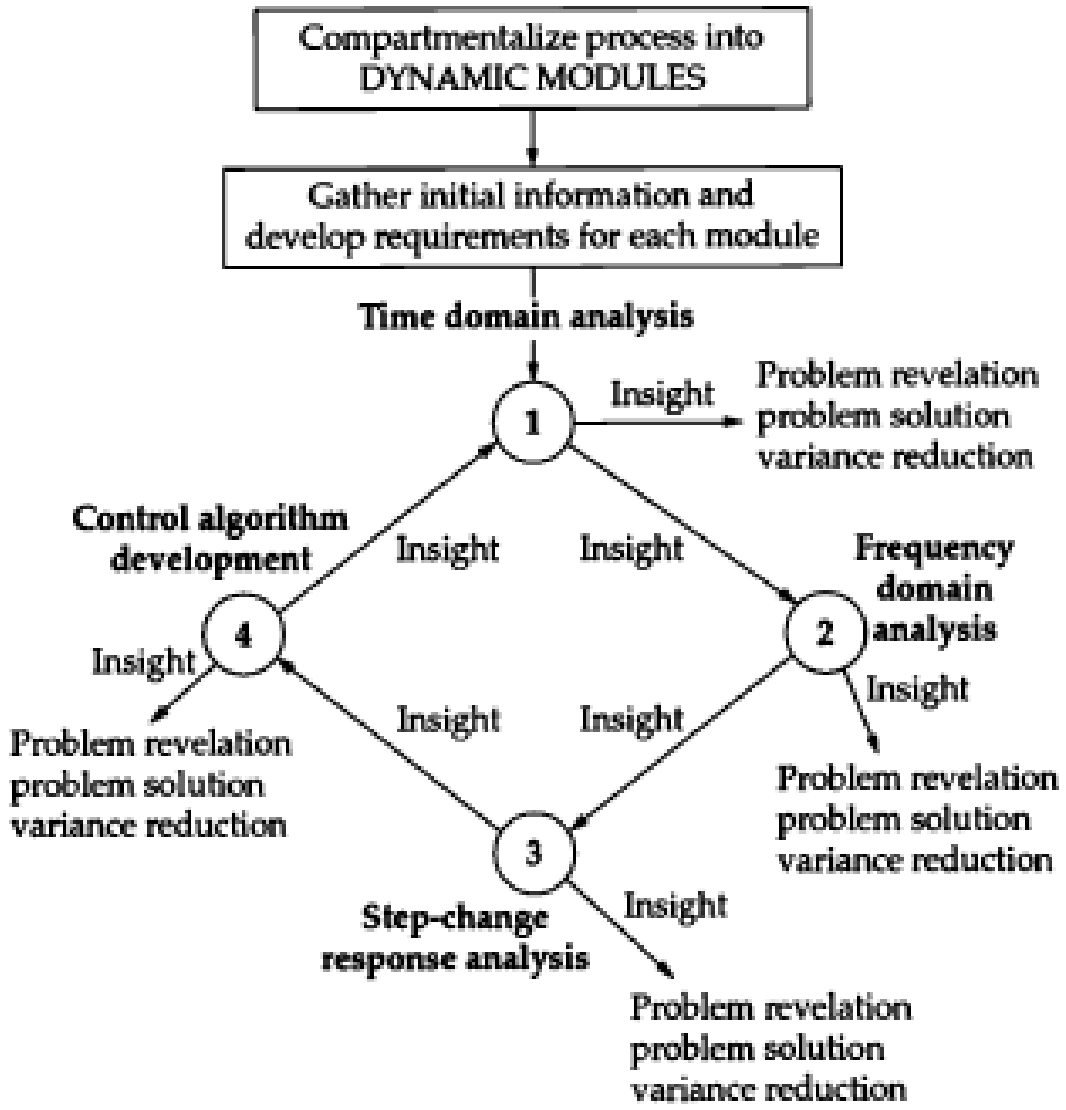


FIGURE 2-1 The diamond road map.

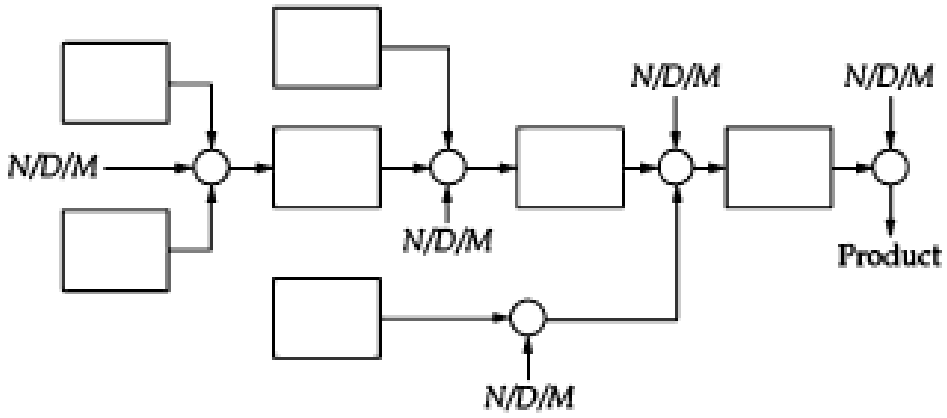


Figure 2-2 A complex process with many sources of noise (*N*), disturbances (*D*), and malfunctions (*M*) .

located at various points in the block diagram that have similar short-term trends due to these disturbances and malfunctions. A good project manager can have both approaches active and complementary.

Time Domain Analysis

Now that a module has been identified and the specifications gathered, it is time to "look" at the process in the simplest most logical way—in the time domain. This means collecting data on selected process variables local to the module and studying how they behave alone and when compared to each other. Before starting to collect the data the team should agree on the key process variables to collect and on what frequency to sample them. This may require installing some new sensors and even installing some data-acquisition equipment.

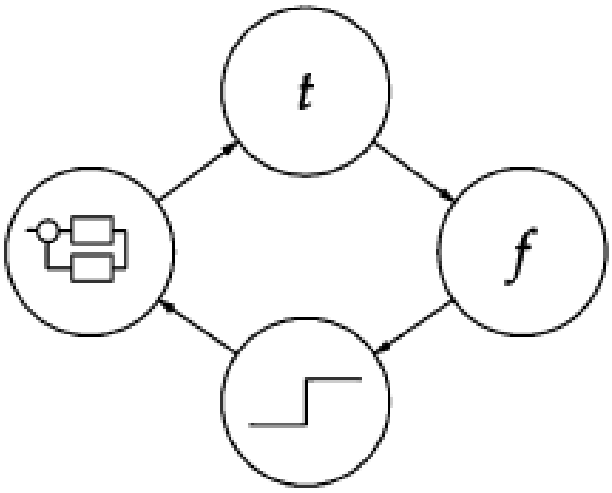


Figure 2-6 The diamond road map icon.

team can basically obviate the need for the activities in the "control development corner."

However, should an algorithm be developed and installed, one would move on to the top of the diamond and study the controlled process in the time domain. Having done this, one could continue around the diamond gaining insight and solving problems. For mnemonic purposes the diamond is symbolized in Fig. 2-6.

2-3 Dealing with Control Algorithms Bundled with the Process

What if you are selling a product such as an optical amplifier and you want to augment your product with a controller that will, say, maintain a desired optical output power? Now, you are bundling the process to be controlled with the controller and forming a product that contains two components. This is quite a different situation compared to that covered in Sec. 2-2.

What Is the Problem?

The product now has two components that can have problems: the process and the controller. If the product is constructed so that only the final output, say the optical power in the case of an optical amplifier, can be monitored then how do you diagnose problems? Is it the process or is it the controller?

Separation and Success

The key to success lies in designing the product with ports that will allow the problem solver to tap into internal signals, namely the controller input and the controller output. With these signals available, the problem solver can isolate the controller from the process.

Problem Solving with Bundling

The key to problem solving is having access to both the so-called process *and* the controller. If ports are in place then problem solving can be divided immediately into verifying that the process and the controller are performing properly. Without these ports problem solving becomes a guessing game (see Fig. 2-8).

There are ancillary benefits to having the extra ports. During product development, the ability to monitor the process and controller separately can allow for parallel beneficial development (see Fig. 2-9). If the learning about the process and controller are concurrent and interactive, their development can be also be interactive—leading to a synergism and a better final product.

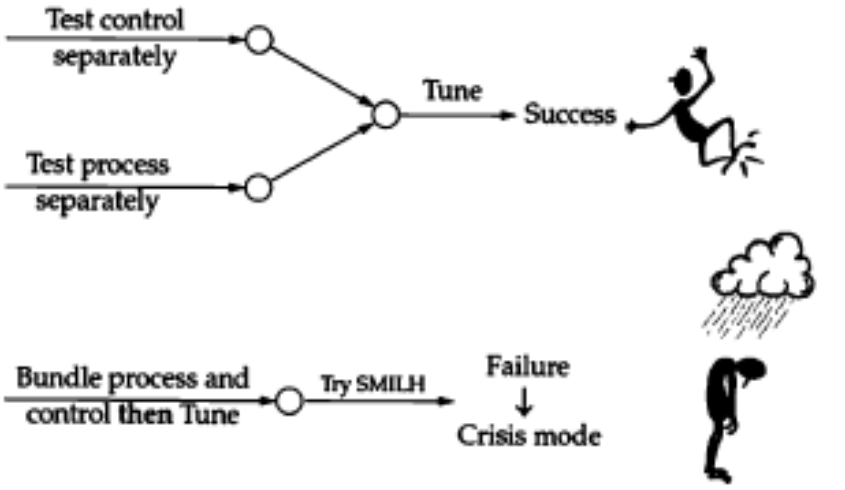


Figure 2-8 Bundling process and control as part of a saleable product—testing the process and the algorithm separately.

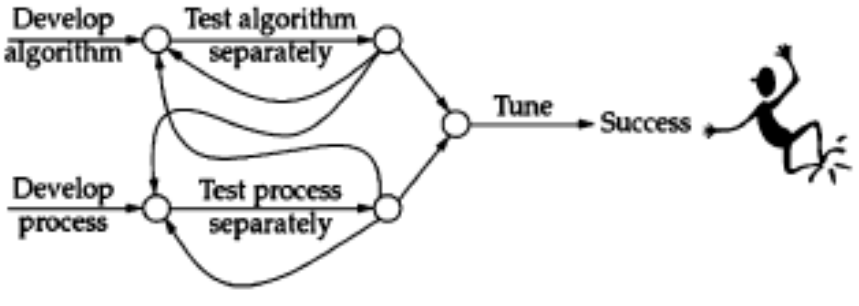


Figure 2-9 The benefits of separation: interaction, evolution, synergism, and problem isolation.

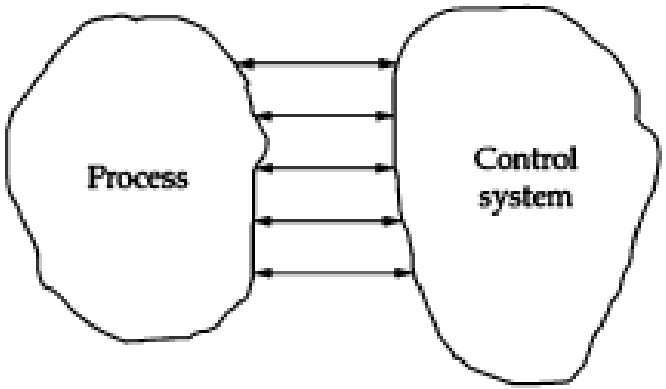


Figure 2-10 Interactions between process and control system.

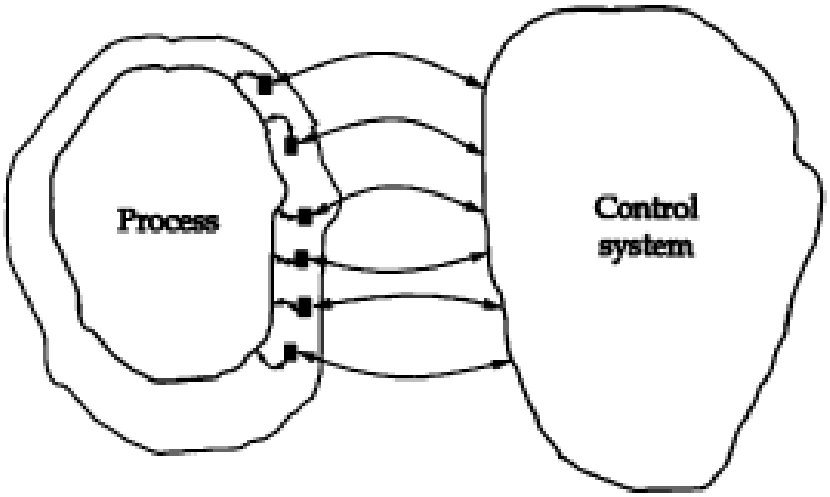


Figure 2-11 Interactions between process and control system with interfaces.

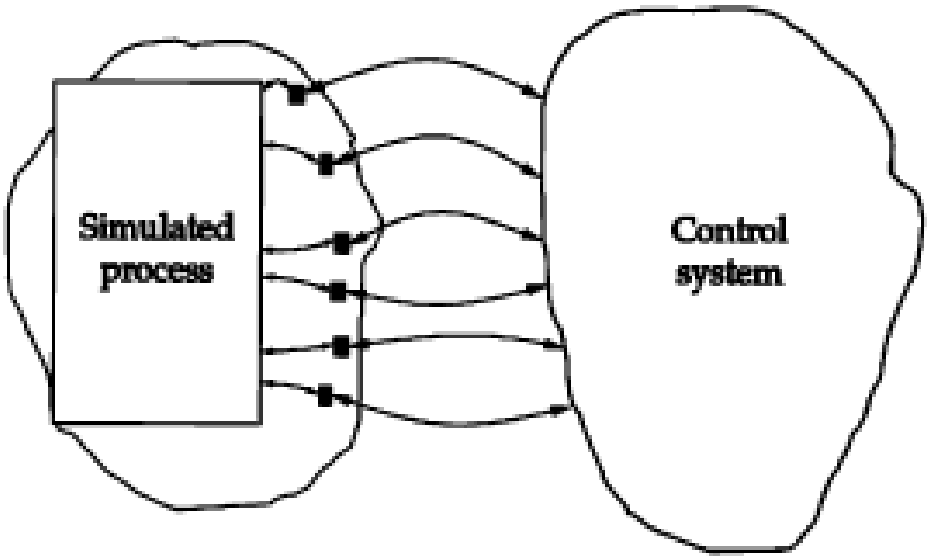


Figure 2-12 Interaction between the simulated process and the control system.

3-1 The First-Order Process—an Introduction

Let's go back to the tank of water introduced in Chap. 1 (Fig. 3-1). It will be our prototypical first-order process. The dynamic analysis of this tank often consists of studying the step response, which is shown in Fig. 3-2. Here the process input U, the valve, is given a step at time $t = 9$, from an initial value of zero to unity. The process output Y, the tank level, begins to rise and appears to *line out* at a value of 2.0. For convenience, we have chosen the initial value of the valve and the tank level to be zero. In general, these quantities could have almost any initial value but this graph would still apply if the reader is willing to allow us to subtract these nonzero initial values, that is, normalizing the initial values of these quantities to zero.

To proceed we need some nomenclature. First, let the *change* in the process input be signified as ΔU. The symbol Δ usually signifies a change in the quantity following it (or upon which it *operates*). Second, let the resulting change in the tank level be signified as ΔY.

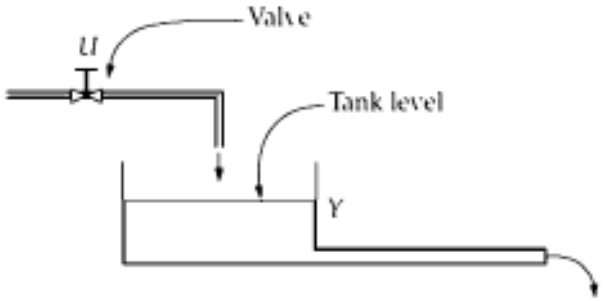
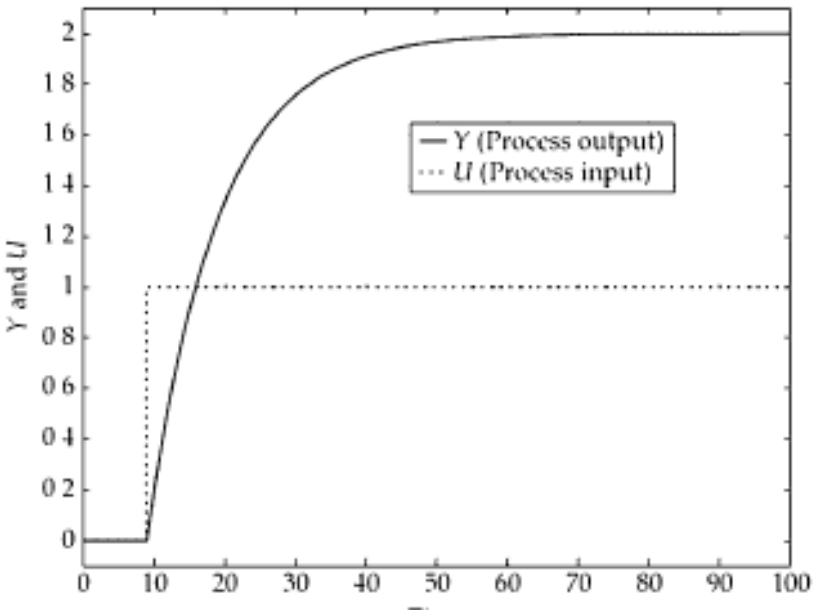


Figure 3-1 A tank of liquid (a process).



3-2-2 Solution of the Continuous Time Domain Model

We will solve the model for the first-order process developed in Sec. 3-2-1 for a variety of conditions. First, consider the case where the process input, initially zero, is assigned the nonzero value of U_c at time zero and held there for all time. The reader hopefully will recognize that this situation is the same as giving the process a step change in the process input at time zero. Hence, if the equation can be solved, the step-change response will be obtained and we can compare the results with the graph in Fig. 3-2.

For this case of a constant process input, Eq. (3-8) becomes

$$\tau \frac{dY}{dt} + Y = gU_c \tag{3-10}$$

There are many ways of solving this equation and a couple of them are reviewed in App. E. The solution for the case of a constant input is

$$Y = Y_0 e^{-\frac{t}{\tau}} + gU_c\left(1 - e^{-\frac{t}{\tau}}\right) \tag{3-11}$$

We will refer to this equation repeatedly in this book so the reader should be comfortable with its development before proceeding.

3-2-3 The First-Order Model and Proportional Control

Although optional, it would be quite helpful if the reader is able to follow the math in App. E used to arrive at the solution of the differential equation for the first-order model. We will now take a little side trip and see what can be learned from this model from the control point of view.

Let's tack a simple "proportional" controller onto our model and see if we can control the process output to a desired set point. Our starting point is the first-order model for the process to be controlled

$$\tau \frac{dY}{dt} + Y = gU \tag{3-15}$$

Our goal is to try to keep the process output Y "acceptably near" the set point S by adjusting U in some fashion. The simplest "fashion" is to form an error

$$e = S - Y \tag{3-16}$$

and to manipulate U in proportion to the error, that is,

$$U = ke = k(S - Y) \tag{3-17}$$

where k is the proportional control gain.

Before proceeding, let's think about Eq. (3-17) with reference to the water tank. Assume that initially Y is equal to the set point S so that e is initially zero. Also, assume that the nominal initial values have been subtracted from all of the quantities, so Y, S, e, and U are initially zero. If S is stepped up, then e would become nonzero and positive. This would mean that U would increase, assuming that k is positive. An increase in U means more flow into the tank and the level Y should rise. Okay, at least the control algorithm has the correct signs and moves the controller output in the right direction.

Schematically, this feedback control system can be presented as a block diagram (Fig. 3-3). This is a classic schematic that will reappear many times in many forms in the balance of this book. Note how the process output Y is fed back and subtracted from the set point S producing the error E which is fed to the controller which produces the process input U.

Combine Eqs. (3-15) and (3-17)

$$\tau \frac{dY}{dt} + Y = gU$$

$$U = k(S - Y)$$

and get

$$\tau \frac{dY}{dt} + Y = gk(S - Y)$$

$$\tau\frac{dY}{dt} + Y = gU$$

$$U = k(S - Y)$$

and get

$$\tau\frac{dY}{dt} + Y = gk(S - Y)$$

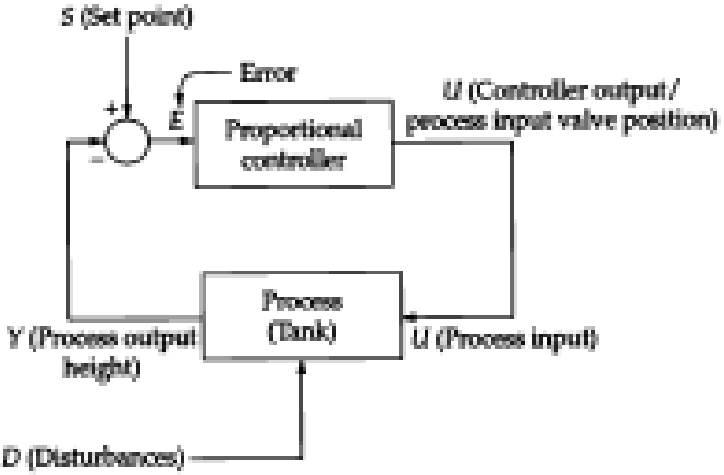


Figure 3-3 A feedback controller.

A little rearrangement, which the reader should verify, will yield

$$\left(\frac{\tau}{1+gk}\right)\frac{dY}{dt} + Y = \frac{gk}{1+gk}S \qquad (3\text{-}18)$$

This is the differential equation that describes the closed-loop system containing the process under simple proportional control. It has the same form as Eq. (3-10) except for the following replacements

$$\tau \Rightarrow \frac{\tau}{1+gk} \qquad g \Rightarrow \frac{gk}{1+gk} \qquad U \Rightarrow S$$

Therefore, by observation, we can obtain the solution to a process under simple proportional feedback control subject to a step in the set point (from 0 to S_r) at time zero. That is, Eq. (3-11), with the above substitutions, becomes

$$Y = Y_0 e^{-t\left(\frac{1}{\tau/gk}\right)} + \frac{gk}{1+gk}S_r\left[1 - e^{-t\left(\frac{1}{\tau/gk}\right)}\right] \qquad (3\text{-}19)$$

Underdamping

Finally, consider the case when

$$(1+gk)^2 < 4\tau g I$$

so,

$$I > \frac{(1+gk)^2}{4\tau g} \tag{3-29}$$

(Note how the integral gain is greater than that for critical damping.) The argument inside the square root is now negative. But we know that

$$\sqrt{-1} = j \qquad \text{and} \qquad \sqrt{-4} = 2j$$

and because of the inequality in Eq. (3-29), the roots are

$$a_1, a_2 = -\frac{(1+gk)}{2\tau} \pm \frac{j\sqrt{4\tau g I - (1+gk)^2}}{2\tau} = \alpha \pm j\beta \tag{3-30}$$

where $\alpha < 0$ and $\beta > 0$ are real numbers. This means that the solution will have exponential terms with imaginary arguments (see App. B) as in

$$e^{(\alpha+j\beta)t} \qquad \text{or} \qquad e^{\alpha t}e^{j\beta t}$$

The $e^{\alpha t}$ term (with $\alpha < 0$) means that the transient response will die away, but what about the other factor? Euler's equation (see App. B) can be useful here.

$$e^{j\beta t} = \cos(\beta t) + j\sin(\beta t)$$

3-3 The Laplace Transform

In the last section we had a little trouble with the second-order differential equation. In this section we introduce a tool, the Laplace transform, which will remove some of the problems associated with differential equations but with the cost of having to learn a new concept. The theory of the Laplace transform is dealt with in App. F so we will start with a simple *recipe* for applying the tool to the first-order differential equation.

The first-order model in the time domain is

$$\tau\frac{dY}{dt}+Y=gU \tag{3-31}$$

To move to the Laplace transform domain, the derivative operator is simply replaced by s, the so-called Laplace transform operator, and wiggles are placed over the symbols Y and U since they are in a new domain

$$\tau s\tilde{Y}+\tilde{Y}=g\tilde{U} \tag{3-32}$$

Before dealing with Eq. (3-32) consider some Laplace transform transition rules in the box:

$$\frac{d}{dt}\Rightarrow s$$

$$Y(t)\Rightarrow \tilde{Y}(s)$$

$$U(t)\Rightarrow \tilde{U}(s)$$

$$C\Rightarrow \frac{C}{s}$$

$$\int_0^t Y(u)du\Rightarrow \frac{\tilde{Y}(s)}{s}$$

$$\lim_{s\to 0} s\tilde{Y}(s)=Y(\infty)$$

Most control books have extensive tables giving the transforms for a wide variety of time functions. Note the following comments about the contents of the box given in Sec. 3-3.

1. All initial values must be zero. (Later on, nonzero initial conditions will be covered.
2. The differential operator d/dt is replaced with s.
3. The integral operator $\int_0^t \ldots du$ is replaced with $1/s$.
4. The quantity C is a constant.
5. The last equation in the box is really not a transform rule. Rather it is the *final value theorem* and it shows how one can find the final value in the time domain if one has the Laplace transform. The basis for these rules and the final value theorem are given in App. F.

3-3-1 The Transfer Function and Block Diagram Algebra

The introduction of the transfer function $G_p(s)$ in Eq. (3-33) is useful because of the block diagram interpretation (Fig. 3-9). The expression in the box multiplies the input to the box to give the box's output.

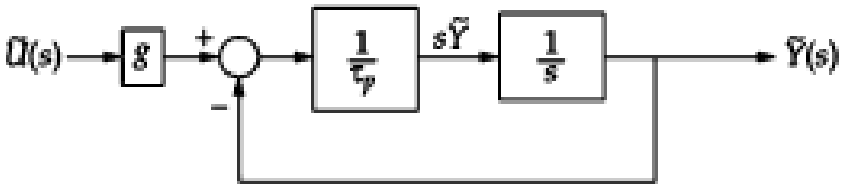
Alternatively, one can play some games with Eq.(3-33) and get

$$\begin{aligned} \tilde{Y} &= \frac{g}{\tau s + 1}\tilde{U} \\ \tilde{Y} + \tau s\tilde{Y} &= g\tilde{U} \\ \tau s\tilde{Y} &= g\tilde{U} - \tilde{Y} \\ \tilde{Y} &= \frac{1}{s}\left(\frac{1}{\tau}\right)\left(g\tilde{U} - \tilde{Y}\right) \end{aligned} \tag{3-34}$$

The last line of Eq. (3-34) suggests that (1) there is some integration going on via the $1/s$ operator and (2) there is some negative feedback since $\tilde{Y}$ is on the right-hand side of the equation with a minus sign. That last line of Eq. (3-34) can be interpreted using block algebra as shown in Fig. 3-10.

The reader should wade through Fig. 3-10 and deduce what each box does. The process output Y is fed back to a summing junction

Figure 3-9 The transfer function in block form.



$$\tilde{Y}(s) = \frac{g}{\tau s + 1}\tilde{U}(s)$$

$$\tau\frac{dY}{dt} + Y = gU$$

Figure 3-10 Block diagram showing integration and negative feedback as part of the process model.

3-3-2 Applying the New Tool to the First-Order Model

Returning to Eq. (3-33), assume that the time domain function $U(t)$ is a step function having a constant value of U_c. Therefore, it will be treated as a nonzero constant for $t \geq 0$. As with all of our variables, $U(t)$ is assumed to be zero for $t < 0$. The Laplace transform for U_c (see App. F and/or the box given in Sec. 3-3) is

$$\bar{U} = \frac{U_c}{s}$$

and Eq. (3-33) becomes

$$\bar{Y} = \frac{g}{\tau s + 1}\frac{U_c}{s} \tag{3-35}$$

To invert this transform to get $Y(t)$, Eq. (3-35) needs to be simplified to a point where we can recognize a familiar form and match it up with a time domain function. *Partial fractions* can be used to split Eq. (3-35) into two simpler terms. Referring to App. F the reader can verify that the new expression for $\bar{Y}$ is

$$\begin{aligned}\bar{Y} &= \frac{g}{\tau s + 1}\frac{U_c}{s} \\ &= \frac{gU_c}{s} - \frac{gU_c}{s + \frac{1}{\tau}} \\ &= gU_c\left(\frac{1}{s} - \frac{1}{s + \frac{1}{\tau}}\right)\end{aligned}$$

We already know the time domain functions for the Laplace transforms, namely,

3-3-3 The Laplace Transform of Derivatives

According to the recipe, the derivative in Eq. (3-31) was replaced by the operator s. App. F shows that the basis for this comes directly from the definition of the Laplace transform, which, for a quantity $Y(t)$, is

$$L\{Y(t)\} = \int_0^\infty dt\, e^{-st} Y(t) \tag{3-36}$$

Note that $e^{-st}Y(t)$ is integrated from $t = 0$ to $t = \infty$. It may seem like a technicality but the integration starts at zero so the value of the quantity $Y(t)$ for $t < 0$ is of no interest and is assumed to be zero. If the quantity has a nonzero initial value, say Y_0, then strictly speaking we have to look at it as

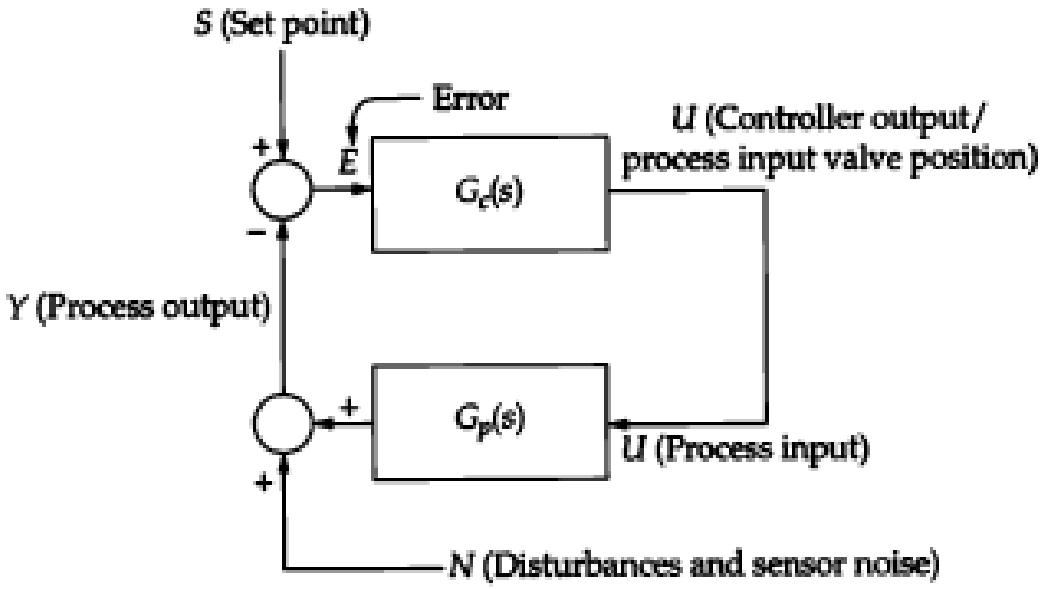
$$Y_0 = \lim_{t \to 0^+} Y(t) = Y(0^+)$$

That is, Y_0 is the initial value of $Y(t)$ when $t = 0$ is approached from the right or from positive values of t. So, effectively, a nonzero initial value corresponds to a step change at $t = 0$ from the Laplace transform point of view. This subtlety comes into play when one evaluates the Laplace transform of the derivative, as in

$$L\left\{\frac{dY}{dt}\right\} = \int_0^\infty dt\, e^{-st} \frac{dY}{dt}$$

The evaluation of this equation presents a bit of a challenge so I put the gory details in App. F for the reader to check if she wishes. However, after all the dust settles the result is

$$\boxed{L\left\{\frac{dY}{dt}\right\} = s\bar{Y} - Y(0^+)} \tag{3-37}$$


S (Set point)
Error
E
+
−
$G_c(s)$
U (Controller output/
process input valve position)
Y (Process output)
+
$G_p(s)$
U (Process input)
+
N (Disturbances and sensor noise)

3-3-6 Zeros and Poles

This section will repeatedly refer to Eq. (3-40) which is

$$\tilde{Y} = \frac{gks + gI}{\tau s^2 + (1 + gk)s + gI} \frac{S_c}{s}$$

The numerator in Eq. (3-40), namely, $gks + gI$, has one zero. That is, the value $s = -I/k$ causes this term to be zero, so the zero of this factor is $-I/k$.

The denominator in Eq. (3-40), namely,

$$(\tau s^2 + (1 + gk)s + gI)\, s$$

has the same form as the quadratic in Eq. (3-25) with one extra factor. Therefore, the denominator in Eq. (3-40) has three *zeros* (values at which a quantity equals zero). Conventionally, we say that Eq. (3-40) has three *poles* (values at which the quantity becomes infinite) and one zero (the value at which the quantity becomes zero).

Partial Fractions and Poles

Applying the quadratic equation solver, the poles of Eq. (3-40) are found to be

$$-\frac{1 + gk}{2\tau} \pm \frac{\sqrt{(1 + gk)^2 - 4\tau gI}}{2\tau} \qquad \text{and} \qquad 0.0 \qquad (3\text{-}44)$$

Two of the roots in Eq. (3-44) are the same as those obtained in Eq. (3-30). Assume for the time being, that the argument of the radical in Eq. (3-44) is positive so that the poles will all be zero or negative real numbers.

To make the following partial fraction algebra a little easier I will factor out τ so that the coefficient of s^2 is unity and Eq. (3-40) becomes

$$\tilde{Y} = \frac{(gks + gI)S_c}{\tau\left(s^2 + \frac{(1 + gk)}{\tau}s + \frac{gI}{\tau}\right)s} = \frac{\frac{(gks + gI)S_c}{\tau}}{(s - s_1)(s - s_2)s} \qquad (3\text{-}45)$$

4-1 Onward to the Frequency Domain

4-1-1 Sinusoidally Disturbing the First-Order Process

Instead of disturbing our tank of liquid with a step change in the input flow rate, consider an input flow rate that varies as a sinusoid about some nominal value as shown in Fig. 4-1. The figure suggests

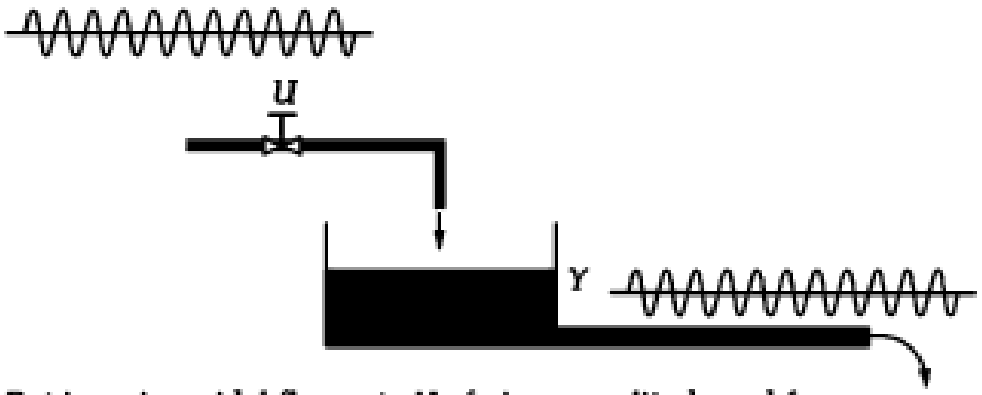


FIGURE 4-1 Frequency response of tank of liquid.

that if the input varies sinusoidally so will the level (and the output flow rate, too). Assume that the input flow rate is described by

$$U(t) = U_c + A_U \sin(2\pi ft)$$

4-1-3 A Little Mathematical Support in the Laplace Transform Domain

From Chap. 3 and App. F, the transfer function for the process described by Eq. (4-1) can be obtained directly from Eq. (3-33) by setting $g = 1.0$, resulting in

$$\tilde{Y} = \frac{1}{\tau s + 1}\tilde{U} = G_p\tilde{U}$$
$$G_p = \frac{1}{\tau s + 1} \tag{4-6}$$

Now, another trick! Let $s = j2\pi f$ and find the magnitude and the phase of the result

$$G_p(j2\pi f) = \frac{1}{\tau j2\pi f + 1}$$
$$= \frac{1}{\tau j2\pi f + 1} \cdot \frac{-\tau j2\pi f + 1}{-\tau j2\pi f + 1}$$
$$= \frac{-\tau j2\pi f + 1}{(\tau 2\pi f)^2 + 1}$$
$$= \frac{1}{(\tau 2\pi f)^2 + 1} - j\frac{\tau 2\pi f}{(\tau 2\pi f)^2 + 1}$$

As shown in App. B, the magnitude of a complex quantity is the square root of the real and imaginary parts

$$\left|G_p(j2\pi f)\right| = \frac{1}{\sqrt{(\tau 2\pi f)^2 + 1}} \tag{4-7}$$

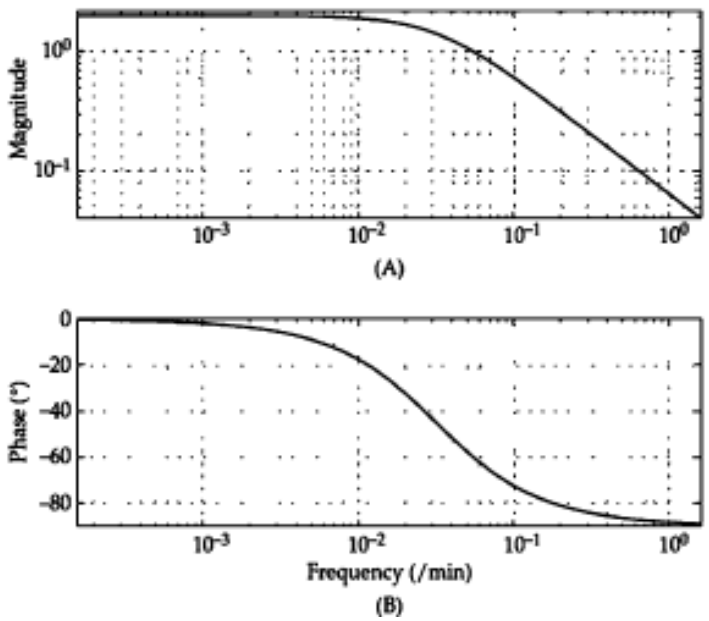
and the phase is the angle whose tangent is the ratio of the imaginary to the real part

$$\theta = \tan^{-1}\left(\frac{-\tau 2\pi f}{1}\right) = -\tan^{-1}(\tau 2\pi f) \tag{4-8}$$

The last two equations turn out to be the same as Eq. (4-5). So, we see one more reason why the Laplace transform can be so useful: there is an easy, straightforward path from the Laplace domain to the frequency domain. All you have to do is accept the serendipitous effect of replacing the Laplace operator by $j\omega$. There is one caveat. The result of making the substitution gives the *steady-state* sinusoidal solution after the transients have died out—remember, when you feed a sinusoid to a process, the process output requires some time to evolve toward a sinusoidal function. Refer to App. B where the full solution, including the transient part, is given.

4-1-4 A Little Graphical Support

How can this information be presented more compactly? Try plotting the amplitude ratio and the phase lag versus the frequency. For this example the result, called a Bode plot, would be as shown in Fig. 4-5,



Question 4-3 What would the slope at large frequencies be if dB units were used?

Answer For large frequencies $dB \cong 20\log_{10}(\tau\omega)$ so the slope would be –20 dB per decade change in frequency.

Third, when $\tau\omega = 1$, the magnitude in dB is given by

$$20\log_{10}\left|G_p(j\omega)\right| = 20\log_{10}\frac{1}{\sqrt{(1)^2+1}} = 20\log_{10}\frac{1}{\sqrt{2}}$$

$$= -20\log_{10}\sqrt{2} = -3.0103 \text{ dB}$$

$$\left|G_p(j\omega)\right| = \frac{1}{\sqrt{(1)^2+1}} = 0.7071$$

Thus, the graph of $|G|$ is approximately 0.7071 or 3 dB down when $\tau\omega = 1$. This frequency, $\omega_{cor} = 1/\tau$ or $f_{cor} = 1/(2\pi\tau)$ is called the *corner* frequency.

The shape of the $|G|$ curve can be approximated with two straight lines—one is horizontal from small $\tau\omega$ to the point where $\tau\omega = 1$. The second has a slope of –1 and starts at $\tau\omega = 1$.

The phase in Eq. (4-8) shows a left-hand asymptote at 0° and a right-hand asymptote at 90°. When $\tau\omega = 1$, the phase is 45°.

Figure 4-9 shows a Bode plot for the case of unity gain and a time constant of 10.0. The magnitude is plotted against $\tau\omega$ on a log-log

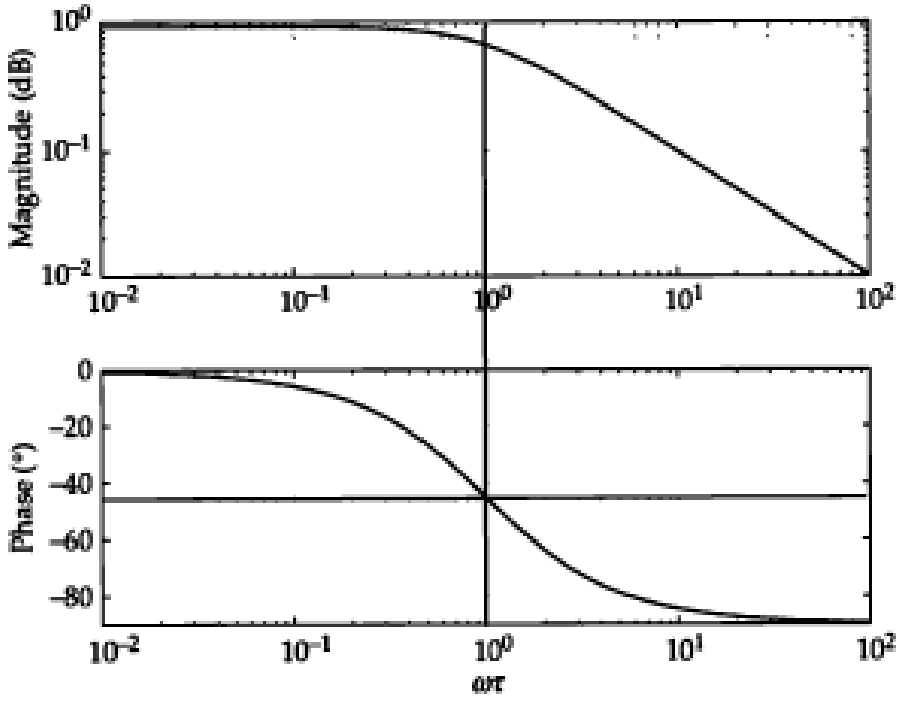


Figure 4-9 Asymptotes for first-order Bode plot.

4-2 How Can Sinusoids Help Us with Understanding Feedback Control?

In the Sec. 4-1, the input flow rate was varied sinusoidally and the output flow rate was observed. This was an open-loop disturbance with no control involved. Now, let's dreg up the closed-loop schematic that we talked about in Chap. 3. There is one change, however. For the time being, the process output will be the process outlet flow rate, so Eq. (4-1) with its unity gain describes the behavior of our process. The process input will still be the process input flow rate.

In Fig. 4-10 note that the set point is varied sinusoidally and the feedback loop is cut just before the process output is fed back and subtracted from the set point. We will focus on the output of the cut line as a response to the sinusoidally varying set-point input. The gain and phase of the output at the cut point will be called the open-loop gain and phase.

However, before looking at that input/output relationship, consider what happens at the point where the process output is subtracted from the set point. An equivalent diagram appears in the upper right-hand corner of Fig. 4-10. Here the subtraction is broken up into a negation followed by an addition. What happens to a sinusoid (the process output) that is negated before it is added to the set point? When the sign is changed from positive to negative the process variable immediately experiences a phase lag of 180° or, in other words, negating a quantity causes it to have a phase of −180°. To see this, look at Fig. 4-11.

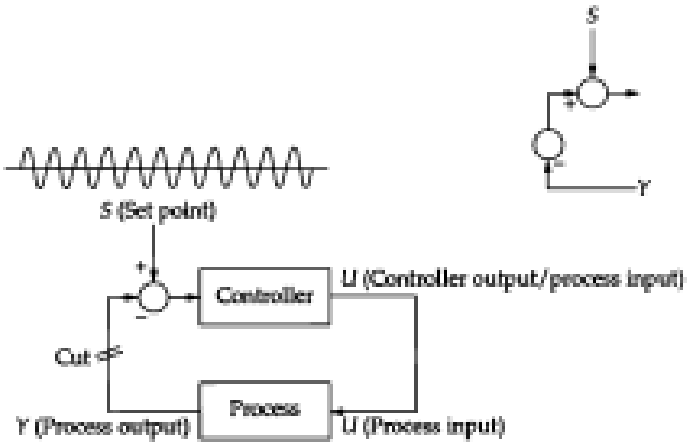
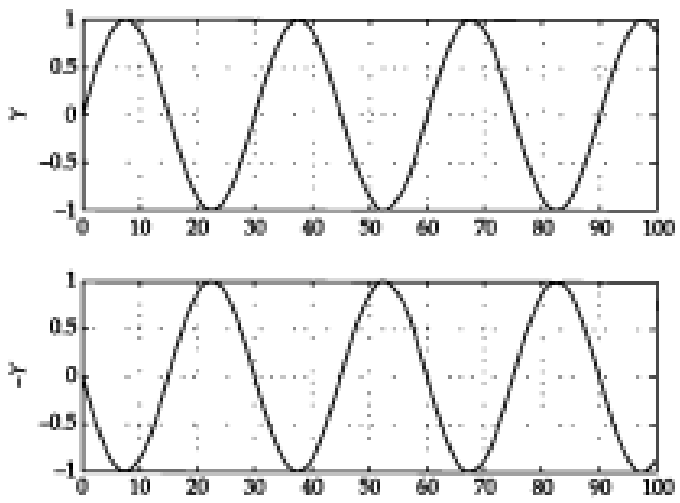


Figure 4-10 Varying the set point and cutting the loop.



4-3 The First-Order Process with Feedback Control in the Frequency Domain

Back in Sec. 3-2-3 the first-order process under feedback control was studied. With proportional-only control, the differential equation describing the controlled system was first order and suggested that the response was bounded under all conditions. When integral control was added the order of the describing differential equation jumped to two but the response still appeared to be bounded although at high integral control gains there could be underdamped behavior.

Before we can use the results of this section for actual design we need to discuss some process models that have more phase lag than the simple first-order system. These models are not only more complicated but they also are able to describe important characteristics of real processes. But for the time being, we will stick with the first-order process.

When the Laplace transform was applied to this first-order process under PI (proportional-integral) control in Chap. 3, the behavior

of the controlled system was shown to be described by the two transfer functions, one for the process G_p and one for the controller G_c. The algebraic development of the transfer function for response of the process output just before the summing point in response to the set point is as follows:

$$\begin{aligned}
\tilde{Y} &= G_p \tilde{U} \\
\tilde{U} &= G_c \tilde{E} \\
\tilde{Y} &= G_p G_c \tilde{E} \\
\tilde{Y} &= G_p G_c \tilde{S} \\
\frac{\tilde{Y}}{\tilde{S}}\Big|_{\text{openloop}} &= G_p G_c = \frac{g}{\tau s + 1}\frac{ks + I}{s} = G
\end{aligned} \tag{4-9}$$

Note that $\tilde{E} = \tilde{S}$ because there is no feedback connection—yet.

Had there been feedback and had the loop actually been *closed*, the algebra would have been carried out as follows:

$$\begin{aligned}
\tilde{Y} &= G_p \tilde{U} \\
\tilde{U} &= G_c \tilde{E} \\
\tilde{Y} &= G_p G_c \tilde{E} \\
\tilde{Y} &= G_p G_c (\tilde{S} - \tilde{Y}) \\
\tilde{Y} + G_p G_c \tilde{Y} &= G_p G_c \tilde{S}
\end{aligned}$$

and a phase. The simplest way to get the overall magnitude and phase is to look at each factor separately and convert it into a magnitude and a phase. After this conversion, the magnitudes can be multiplied and divided as necessary and the phases can be added or subtracted as necessary:

$$G(j\omega) = G_p(j\omega)G_c(j\omega) = \frac{g}{\tau j\omega + 1}\frac{kj\omega + I}{j\omega}$$

$$= \frac{ge^{j0}}{\sqrt{(\tau\omega)^2 + 1}\,e^{j\theta_1}}\frac{\sqrt{(k\omega)^2 + I^2}\;e^{j\theta_2}}{1e^{j\pi/2}}$$

$$= \frac{g\sqrt{(k\omega)^2 + I^2}}{\sqrt{(\tau\omega)^2 + 1}}e^{j(\theta_2 - \theta_1 - \pi/2)}$$

$$= |G|e^{j\theta}$$

$$\theta = \theta_2 - \theta_1 - \pi/2$$

$$\theta_1 = \tan^{-1}(\tau\omega) \qquad (4\text{-}11)$$

$$\theta_2 = \tan^{-1}\left(\frac{k\omega}{I}\right)$$

$$|G| = \frac{g\sqrt{(k\omega)^2 + I^2}}{\sqrt{(\tau\omega)^2 + 1}} \qquad \theta = \tan^{-1}\left(\frac{k\omega}{I}\right) - \tan^{-1}(\tau\omega) - \frac{\pi}{2}$$

The development of these equations used the simple algebra of complex exponentials where the magnitudes multiply and the angles add.

Question 4-4 Could you derive the appropriate equations for the magnitude and angles for the case where $I = 0$?

Answer With $I = 0$ and $k = 1$, Eq. (4-9) simplifies to the equation in Sec. 4-2 for the first-order process without control. See Eqs. (4-7) and (4-8).

Question 4-5 Could you derive the appropriate equations for the magnitude and angles for the case where $k = 0$?

Answer Take the limit as $k \to 0$ in Eq. (4-11) and remember that the angle whose tangent is zero is zero. The result will be

4-4 A Pure Dead-Time Process

Consider the process depicted in Fig. 4-22. Imagine many small buckets nearly contiguous such that when the inlet flow rate is continuous so is the outlet flow rate. With this in mind, Fig. 4-22 suggests that the process output Y will be identical to the process input U except with a shift in time, namely,

$$Y(t) = U(t - D) \tag{4-12}$$

where D is the dead time. If the conveyor belt speed is v and the distance between the filling and dumping points is L then the dead time would be $D = L/v$.

Figure 4-23 shows the step response of a process having a dead time of 8 time units. The process gain is unity—what goes in comes out unattenuated and unamplified. The time constant is zero but there is a dead time between the step in the input and the response of the output.

So much for the time domain. What does the Bode plot for the pure dead-time process look like? Figure 4-24 shows magnitude and phase plotted for linear frequency and Fig. 4-25 shows the same thing plotted with logarithmic frequency. Both figures support our contention that the amplitude ratio of the output to the input is unaffected by frequency. However, the phase lag of the output

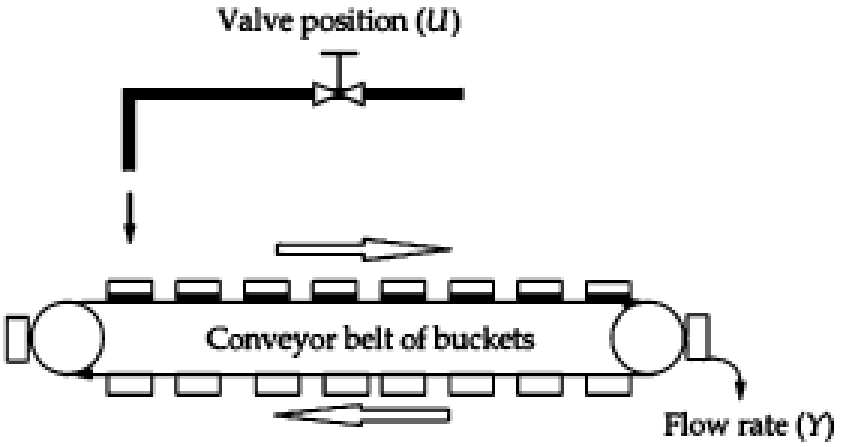


Figure 4-22 A dead-time process. Imagine many small buckets together so that the flow is effectively continuous.

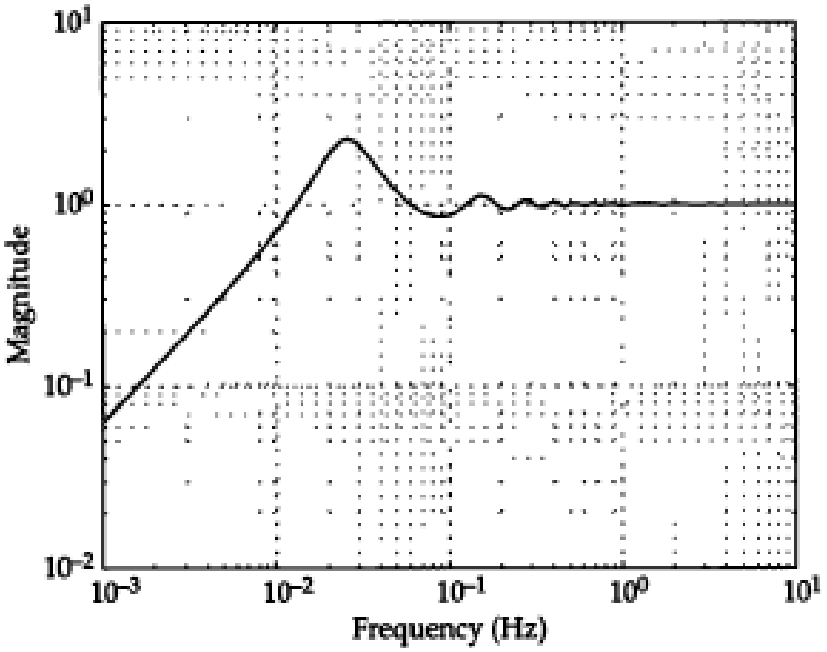


Figure 4-31 Error transmission curve for integral-only control of pure dead-time process.

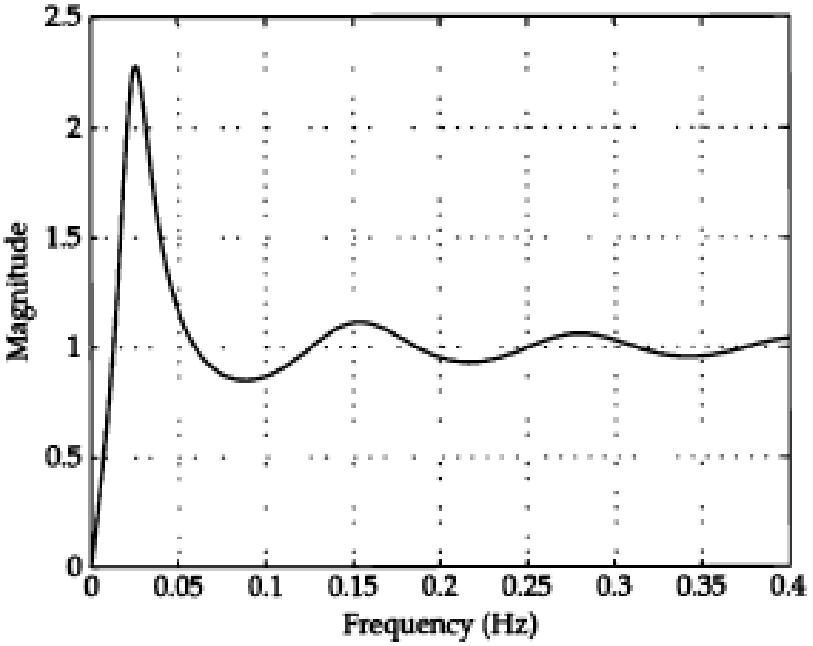


Figure 4-32 Error transmission curve for integral-only control of pure dead-time process (linear axes).

4-5 A First-Order with Dead-Time (FOWDT) Process

Consider Fig. 4-33 where the tank of liquid has been placed upstream of the pure dead-time process. The placement of the buckets-on-the-belt ahead of the tank suggests a dead time in series with a first-order process. Please do not be confused by the length of the pipe at the outlet of the tank. Let's assume that it is actually relatively short and that the pipe diameter is small so that the transit time of the liquid spent in the pipe is negligible compared to the time spent in the buckets on the belt.

Figure 4-34 shows the open-loop step-change response of the process for the case of $g = 2.5$, $\tau = 10$, $D = 8$. This is the first example process in this chapter that has had a nonunity gain.

In the continuous time domain, this model would be described by an extension of the first-order model:

$$\tau \frac{dy}{dt} + y = gU(t - D) \tag{4-16}$$

In the Laplace domain, the open-loop transfer function is

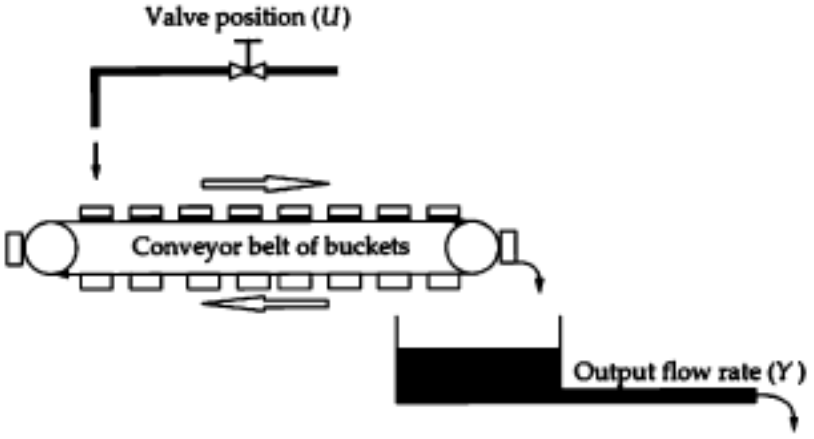
$$G_p(s) = e^{-sD} \frac{g}{\tau s + 1} \tag{4-17}$$

After applying $s = j\omega$, the magnitude and phase can be found as follows:

$$G_p(j\omega) = e^{-j\omega D} \frac{g}{j\tau\omega + 1} = |G_p| e^{j\theta}$$

$$|G_p| = \frac{g}{\sqrt{(\tau\omega)^2 + 1}} \tag{4-18}$$

$$\theta = -\tan^{-1}(\tau\omega) - \omega D$$



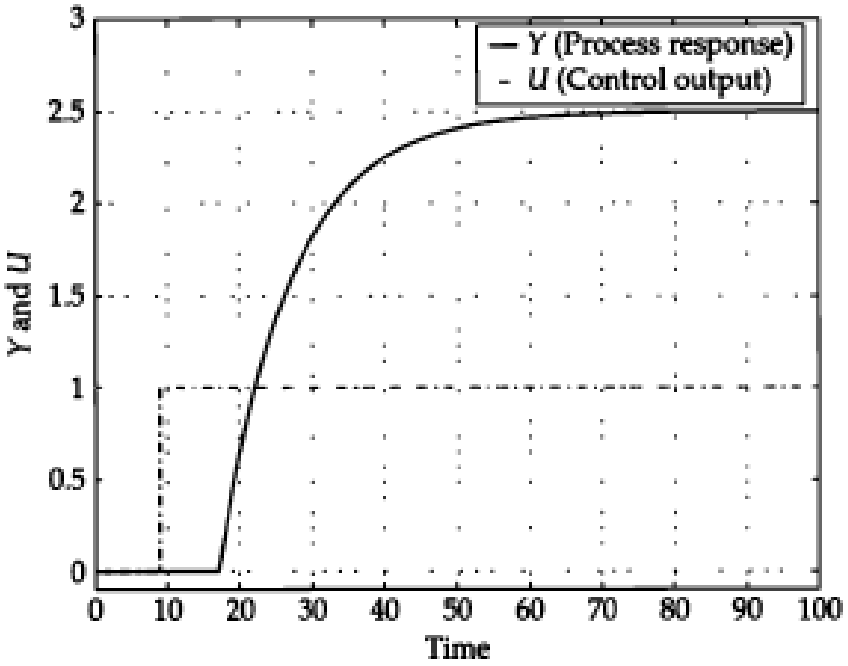


Figure 4-34 Open-loop step response of FOWDT process ($\tau = 10$, $D = 8$, $g = 2.5$).

The magnitude is identical to that of the dead time–less first-order model in Eq. (4-7) but the phase lag is increased by the contribution of the dead time.

The Bode plot for the open-loop transfer function is given in Fig. 4-35. Note the circles that indicate unity magnitude and −180° phase.

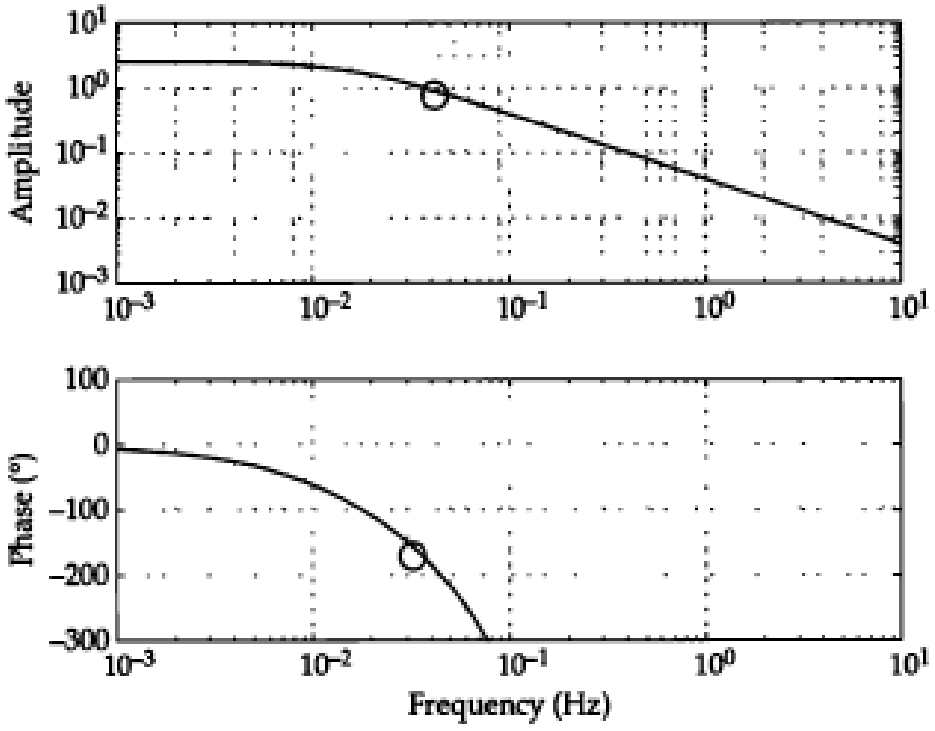


Figure 4-35 Bode plot for FOWDT process ($\tau = 10$, $D = 8$, $g = 2.5$).

4-5-1 The Concept of Minimum Phase

The FOWDT process is an example of a nonminimum phase model (NMP), which means there are other processes that have the same magnitude but have less phase lag. The first-order process (without dead time) is such a model. Figure 4-36 shows the open-loop Bode plot of a FOWDT process along with that of the minimum phase (MP) model. We will not use this concept but the manager may come across it and needs to be aware of it.

4-5-2 Proportional-Only Control

If the proportional-only control gain k is unity then the Bode plot for the open-loop transfer function would be identical to that in Fig. 4-35 and it would appear that this system would be unstable because at a frequency of about 0.05 Hz the phase is –180° and at that same frequency the amplitude is a little over 2.0. Likewise at a frequency of about 0.2 Hz the amplitude is about 1.0 while the phase is off-scale so it is at most –300°. This suggests that stability might be obtained by applying a control gain less than unity such that the overall gain is reduced below unity. To make it *just stable* the overall gain should equal unity or

$$gk = 1 = 2.5 \cdot k$$

$$k = 0.4$$

To obtain a little gain margin let us reduce the proportional control gain to 0.3 so that the overall gain is 0.75 instead of 2.5. Then the Bode

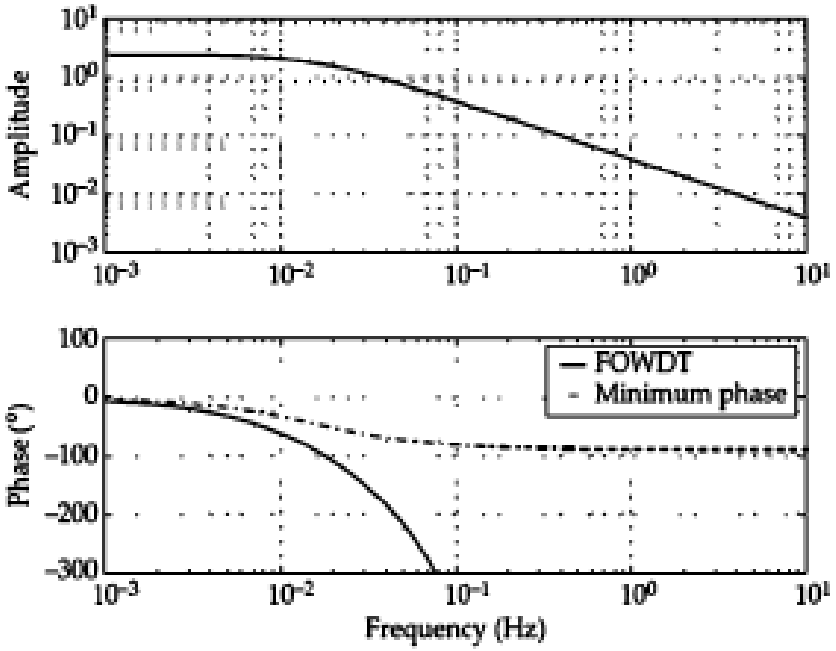


Figure 4-36 Bode plot for FOWDT model and the MP model ($g = 0.75$).

4-5-3 Proportional-Integral Control of the FOWDT Process

Adding integral control causes the open-loop transfer function to become

$$G_p G_c(s) = e^{-sD}\frac{g}{\tau s+1}\frac{ks+I}{s}$$

Applying $s = j\omega$ gives a relatively messy expression—but the reader might try wading through the following algebra—it's worth it:

$$G_p G_c(j\omega) = e^{-j\omega D}\frac{g}{\tau j\omega+1}\frac{kj\omega+I}{j\omega}$$

$$= \left(e^{-j\omega D}\right)\left(\frac{g}{\sqrt{(\tau\omega)^2+1}\,e^{j\tan^{-1}(\tau\omega)}}\right)\frac{\left(\sqrt{(k\omega)^2+1}\,e^{j\tan^{-1}\left(\frac{k\omega}{I}\right)}\right)}{\left(\omega e^{j\frac{\pi}{2}}\right)}$$

$$= \frac{g}{\sqrt{(\tau\omega)^2+1}}\frac{\sqrt{(k\omega)^2+I}}{\omega}\frac{e^{-j\omega D}e^{j\tan^{-1}\left(\frac{k\omega}{I}\right)}}{e^{j\tan^{-1}(\tau\omega)}e^{j\pi}}$$

$$= \frac{g}{\sqrt{(\tau\omega)^2+1}}\frac{\sqrt{(k\omega)^2+I}}{\omega}e^{-j\omega D+j\tan^{-1}\left(\frac{k\omega}{I}\right)-j\tan^{-1}(\tau\omega)-j\frac{\pi}{2}}$$

$$= |G|e^{j\theta}$$

$$|G| = \frac{g}{\sqrt{(\tau\omega)^2+1}}\frac{\sqrt{(k\omega)^2+I}}{\omega} \tag{4-19}$$

$$\theta = -\omega D - \tan^{-1}(\tau\omega) + \tan^{-1}\left(\frac{k\omega}{I}\right) - \frac{\pi}{2}$$

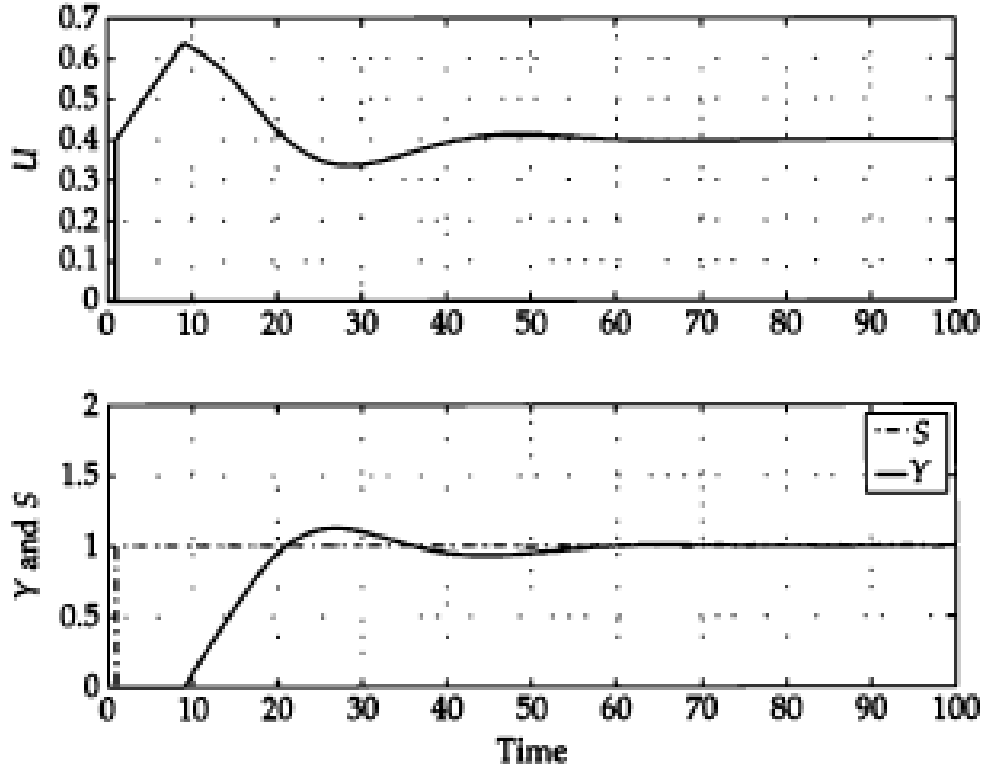


Figure 4-39 Response of PI controlled FOWDT process to unit set-point step.

4-7 Partial Summary and a Slight Modification of the Rule of Thumb

Our approach has been to find a proportional control gain, called the critical gain, that makes the open-loop amplitude ratio unity or the open-loop phase lag 180°. In the former case we reduce the critical proportional control gain to make the phase lag less than 180° by about 45°. In the latter case we reduce the critical proportional control gain by a factor of 0.5.

Therefore, as a starting point we are trying the find the critical values of ω and k_c, namely ω_c and k_c, such that the open-loop gain has a magnitude of unity and a phase of −180°. The App. B shows that the complex number −1.0 has a magnitude of unity and a phase of −180°, so we are really trying to find values of ω_c and k_c that satisfy the following equation:

$$G_p(j\omega_c)G_c(j\omega_c) = G(j\omega_c) = -1$$

or

$$\left|G(j\omega_c)\right|e^{i\theta} = -1 = 1e^{-i\pi} \tag{4-25}$$

or

$$\left|G(j\omega_c)\right| = 1$$

$$\theta(j\omega_c) = -\pi$$

Since several of the closed-loop transfer functions have a denominator of $1+G_cG_p$, it follows that finding the poles of these transfer functions is equivalent to solving Eq. (4-25).

If the proportional gain is set equal to k_c the performance should be on a cusp between instability and stability. That is, the process with the controller should experience sustained oscillations.

The critical values for proportional-only control of the FOWDT process would be the solution of the following two equations that come directly from Eq. (4-25):

5-1 Third-Order Process without Backflow

Figure 5-1 shows three independent tanks—independent in the sense that each downstream tank does not influence its upstream neighbor. Each tank in the series of three can be treated like the single tank we treated earlier except that the outlet flow rate of the upstream tank feeds into the next tank down the line. The single tank is described by

$$\rho A \frac{dL}{dt} + \frac{L}{R} = F \tag{5-1}$$

or

$$\tau \frac{dL}{dt} + L = RF \qquad \tau = \rho AR$$

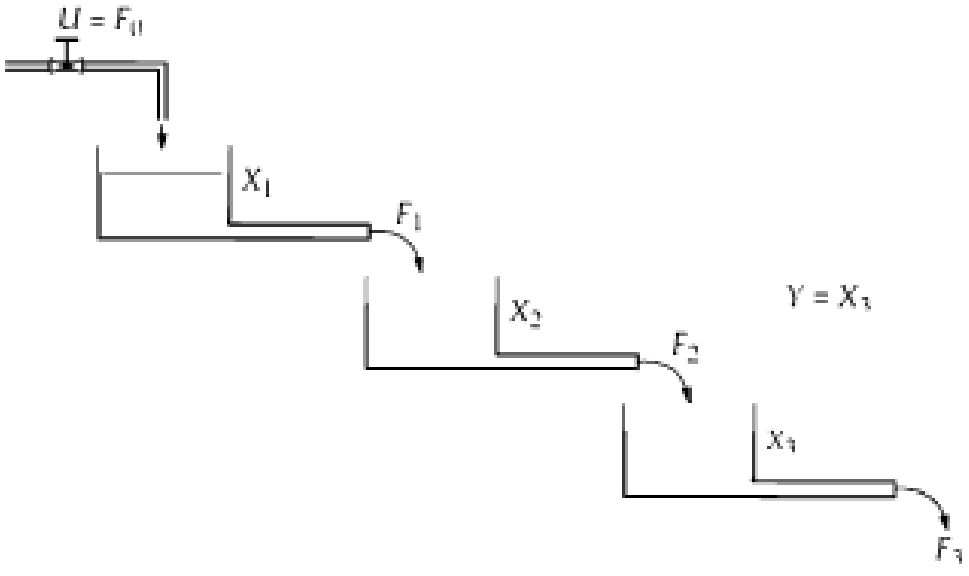


FIGURE 5-1 Three independent tanks.

The equations describing their behavior can be derived from Eq. (5-1) by inspection:

$$\begin{aligned}
\rho A_1 \frac{dX_1}{dt} + \frac{X_1}{R_1} &= F_0 \\
\rho A_2 \frac{dX_2}{dt} + \frac{X_2}{R_2} &= \frac{X_1}{R_1} \\
\rho A_3 \frac{dX_3}{dt} + \frac{X_3}{R_3} &= \frac{X_2}{R_2} \\
Y = X_3 \qquad U &= F_0
\end{aligned} \tag{5-2}$$

As we will see later on, X_1, X_2, and X_3 are the *states* of the system and the last line in Eq (5-2) says that the process output Y is the third element of the state.

The Laplace Transform Version

By inspection, the reader should be able to rewrite Eq. (5-2) in the Laplace domain.

$$\begin{aligned}
\left(\rho A_1 s + \frac{1}{R_1}\right)\bar{X}_1(s) &= \bar{U}(s) \\
\left(\rho A_2 s + \frac{1}{R_2}\right)\bar{X}_2(s) &= \frac{1}{R_1}\bar{X}_1(s) \\
\left(\rho A_3 s + \frac{1}{R_3}\right)\bar{X}_3(s) &= \frac{1}{R_2}\bar{X}_2(s) \\
\bar{Y}(s) &= \bar{X}_3(s)
\end{aligned} \tag{5-3}$$

These equations can be combined in the Laplace domain, eliminating the $\bar{X}_i$. First, rewrite Eq. (5-3) slightly, introducing three time constants.

$$\begin{aligned}
(R_1\rho A_1 s + 1)\bar{X}_1(s) &= \bar{U}(s) \qquad \tau_1 = R_1\rho A_1 \\
(R_2\rho A_2 s + 1)\bar{X}_2(s) &= \frac{R_2}{R_1}\bar{X}_1(s) \qquad \tau_2 = R_2\rho A_2 \\
(R_3\rho A_3 s + 1)\bar{X}_3(s) &= \frac{R_3}{R_2}\bar{X}_2(s) \qquad \tau_3 = R_3\rho A_3
\end{aligned}$$

Starting with the last equation, eliminate the three $\bar{X}_i$ and develop an expression for the process transfer function G.

Second, eliminate $\bar{X}_1$ and $\bar{X}_2$ to get $\bar{X}_3$:

The Matrix (State-Space) Version

Return to the time domain and rearrange Eq. (5-2) slightly

$$\frac{dX_1}{dt} = -\frac{X_1}{\tau_1} + \frac{R_1}{\tau_1} F_0$$

$$\frac{dX_2}{dt} = \frac{R_2}{R_1}\frac{X_1}{\tau_2} - \frac{X_2}{\tau_2} \qquad (5\text{-}6)$$

$$\frac{dX_3}{dt} = \frac{R_3}{R_2}\frac{X_2}{\tau_3} - \frac{X_3}{\tau_3}$$

$$Y = X_3$$

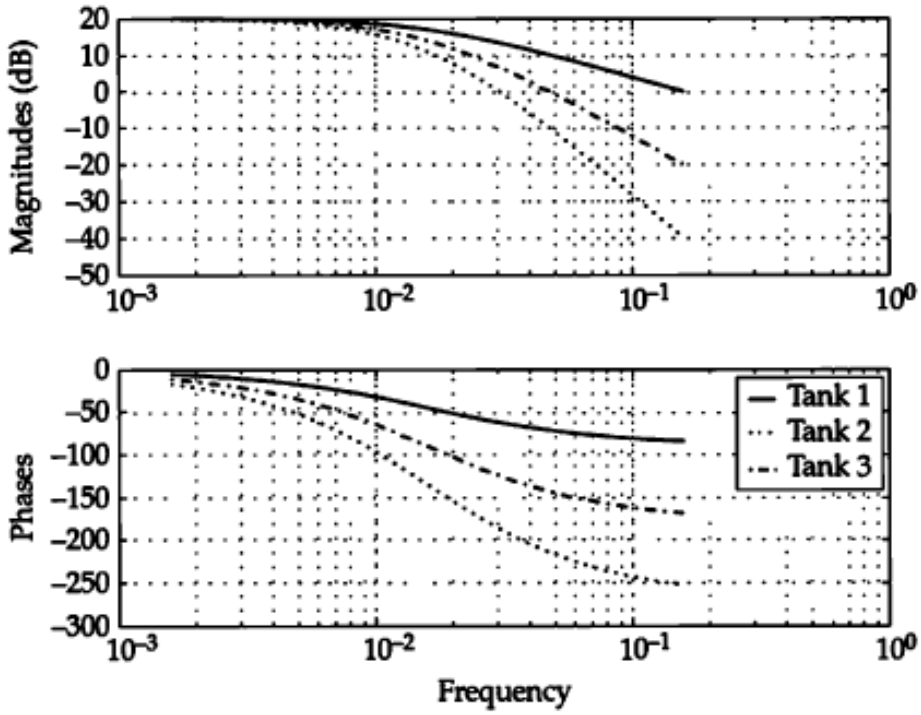


Figure 5-3 Bode plot of three-tank process with no backflow.

$$\frac{d}{dt}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\tau_1} & 0 & 0 \\ \frac{R_2}{R_1}\frac{1}{\tau_2} & -\frac{1}{\tau_2} & 0 \\ 0 & \frac{R_3}{R_2}\frac{1}{\tau_3} & -\frac{1}{\tau_3} \end{pmatrix}\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} + \begin{pmatrix} \frac{R_1}{\tau_1} \\ 0 \\ 0 \end{pmatrix} F_0 \qquad (5\text{-}7)$$

$$Y = (0 \quad 0 \quad 1)\begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix}$$

In general, many linear models of processes can be written to fit the general format of

$$\frac{d}{dt}X = AX + BU$$
$$Y = CX \qquad (5\text{-}8)$$

For this particular process the vectors and matrices occurring in Eq. (5-8) are

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \qquad A = \begin{pmatrix} -\frac{1}{\tau_1} & 0 & 0 \\ \frac{R_2}{R_1}\frac{1}{\tau_2} & -\frac{1}{\tau_2} & 0 \\ 0 & \frac{R_3}{R_2}\frac{1}{\tau_3} & -\frac{1}{\tau_3} \end{pmatrix} \qquad B = \begin{pmatrix} \frac{R_1}{\tau_1} \\ 0 \\ 0 \end{pmatrix} \qquad U = F_0$$

$$C = (0 \quad 0 \quad 1) \qquad Y = X_3$$

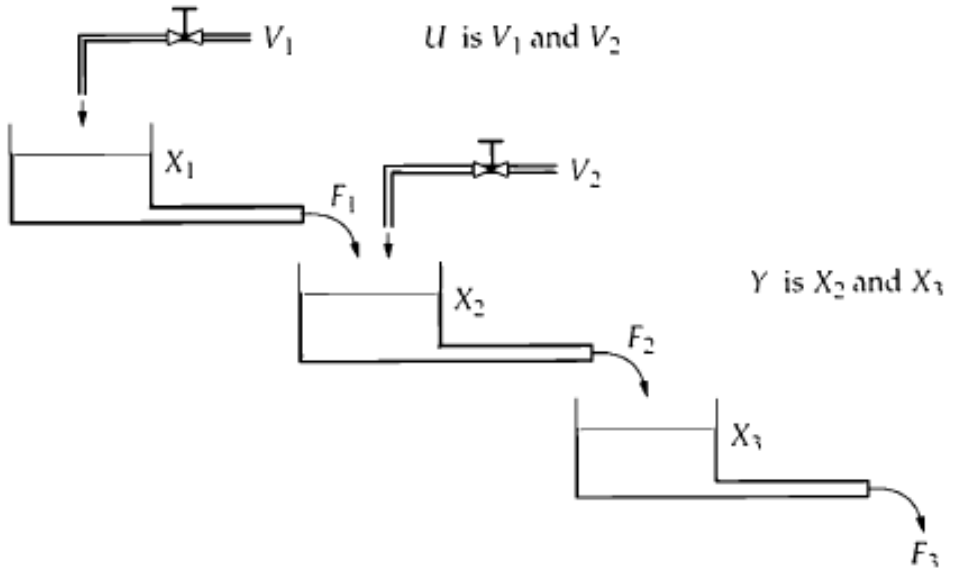


FIGURE 5-4 Three independent tanks with different inputs and outputs.

The matrices B and C change to accommodate the two-dimensional process, input and output, which have become two-dimensional column vectors. The matrix A and the state X are the same.

$$X = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \quad A = \begin{pmatrix} -\frac{1}{\tau_1} & 0 & 0 \\ \frac{R_2}{R_1}\frac{1}{\tau_2} & -\frac{1}{\tau_2} & 0 \\ 0 & \frac{R_3}{R_2}\frac{1}{\tau_3} & -\frac{1}{\tau_3} \end{pmatrix} \quad B = \begin{pmatrix} \frac{R_1}{\tau_1} & 0 \\ 0 & \frac{R_2}{\tau_2} \end{pmatrix} \quad U = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$$

$$Y = CX$$

$$C = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad Y = \begin{pmatrix} X_2 \\ X_3 \end{pmatrix}$$

5-2 Third-Order Process with Backflow

Figure 5-5 shows an interconnected three-tank system with forward *and* backflow. If we treat each tank separately, the equations of Chap. 3 derived from mass balances can be applied immediately.

$$\begin{aligned}
\rho A_1 \frac{dX_1}{dt} &= U - \frac{X_1 - X_2}{R_{12}} \\
\rho A_2 \frac{dX_2}{dt} &= \frac{X_1 - X_2}{R_{12}} - \frac{X_2 - X_3}{R_{23}} \\
\rho A_3 \frac{dX_3}{dt} &= \frac{X_2 - X_3}{R_{23}} - \frac{X_3}{R_3} \\
Y &= X_3
\end{aligned} \tag{5-11}$$

The variables X_1, X_2, and X_3 represent the levels in each of the tanks. The net flow leaving the first tank is

$$\frac{X_1 - X_2}{R_{12}}$$

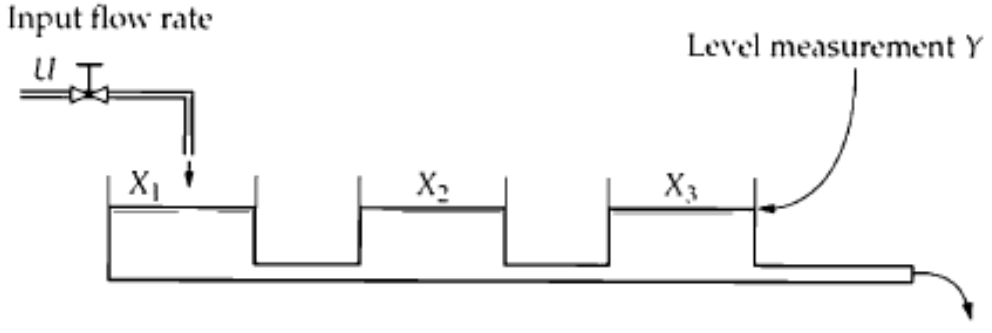


FIGURE 5-5 A three-tank system with backflow.

5-3 Control of Three-Tank System with No Backflow

These two example processes have a potential for control problems because at high frequencies the phase lag approaches 270°. To make it even more interesting, let's try integral-only control which we know adds an immediate 90° of phase lag to whatever is being controlled. With integral-only control the open-loop transfer function for the three-tank process with no backflow becomes

$$G(s) = G_p(s)G_c(s) = \frac{R_3}{\tau_3 s+1}\frac{1}{\tau_2 s+1}\frac{1}{\tau_1 s+1}\frac{I}{s} \qquad G_c = \frac{I}{s}$$

$$G(j\omega) = \frac{R_3}{\tau_3 j\omega+1}\frac{1}{\tau_2 j\omega+1}\frac{1}{\tau_1 j\omega+1}\frac{I}{j\omega}$$

$$= \frac{R_3}{\sqrt{(\tau_3\omega)^2+1}\,e^{j\theta_3}}\frac{1}{\sqrt{(\tau_2\omega)^2+1}\,e^{j\theta_2}}\frac{1}{\sqrt{(\tau_1\omega)^2+1}\,e^{j\theta_1}}\frac{I}{e^{j\frac{\pi}{2}}}$$

$$= \frac{R_3 e^{-j\theta_3}e^{-j\theta_2}e^{-j\theta_1}e^{-j\frac{\pi}{2}}}{\sqrt{(\tau_3\omega)^2+1}}\frac{1}{\sqrt{(\tau_2\omega)^2+1}}\frac{1}{\sqrt{(\tau_1\omega)^2+1}}\frac{I}{\omega}$$

$$= \frac{R_3 e^{-j(\theta_3+\theta_2+\theta_1+\frac{\pi}{2})}}{\sqrt{(\tau_3\omega)^2+1}\sqrt{(\tau_2\omega)^2+1}\sqrt{(\tau_1\omega)^2+1}}\frac{I}{\omega}$$

$$\theta_i = \tan^{-1}(\tau_i\omega) \quad i = 1,2,3 \tag{5-14}$$

$$\theta_i = \tan^{-1}(\tau_i \omega) \quad i = 1,2,3 \tag{5-14}$$

The Bode plot of the process with (I = 1) and without integral control is shown in Fig. 5-8. Note how the addition of integral-only control raises the amplitude curve and lowers the phase curve.

The critical points where the phase equals –180° or where the amplitude or overall gain equals unity (or 0 dB) can be found graphically from Fig. 5-8. This suggests that if $I = 1$ were to be used, instability would result. (Can you convince yourself of this?) To get the magnitude plot down below unity, when the phase equals –180°, requires that we lower I significantly. Figure 5-8 suggests that we might want to lower the gain by at least as much as 50 dB or a factor of 0.003. It is a bit difficult to estimate this using the graph.

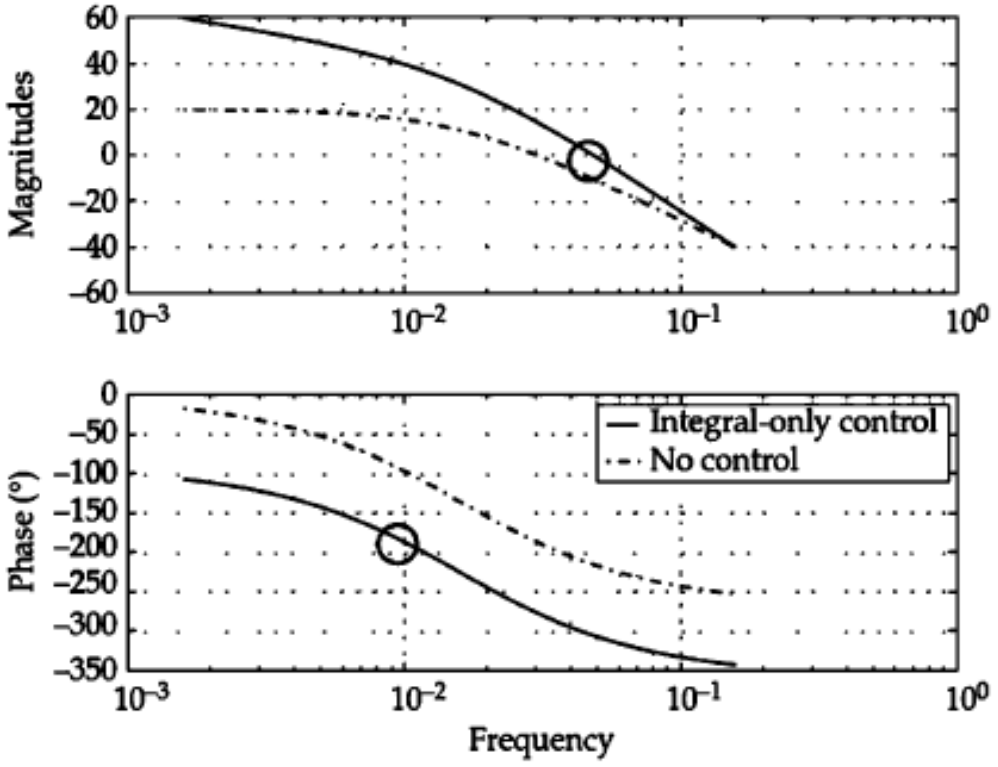


Figure 5-8 Bode plot for no backflow three-tank process with and without integral control.

Alternatively, a two-dimensional valley-seeking algorithm can be easily constructed to find the values of I and ω that minimize the following quantity

$$\left(\frac{R_3}{\sqrt{(\tau_3\omega)^2+1}\sqrt{(\tau_2\omega)^2+1}\sqrt{(\tau_1\omega)^2+1}}\frac{I}{\omega}-1\right)^2 + \left(\tan^{-1}(\tau_1\omega)+\tan^{-1}(\tau_2\omega)+\tan^{-1}(\tau_3\omega)+\frac{\pi}{2}+\pi\right)^2 \tag{5-15}$$

which is equivalent to solving Eq. (4-25) in Chap. 4 which is repeated here:

$$G_p(j\omega_c)G_c(j\omega_c)=G(j\omega_c)=-1$$

or

$$\left|G(j\omega_c)\right|e^{j\theta}=-1=1e^{-j\pi}$$

or

$$\left|G(j\omega_c)\right|=1 \qquad \left|G(j\omega_c)\right|-1=0$$

or

$$\theta(j\omega_c)=-\pi \qquad \theta(j\omega_c)+\pi=0$$

For this case, the quantity in Eq. (5-15) is minimized by $I=0.0089$ and $f=\omega/2\pi=0.00929$ Hz.

A Matlab script to carry out this minimization is

Closed-Loop Performance in the Frequency Domain

Figure 5-12 shows the Bode plot for the *closed*-loop system under integral-only control (shown in Fig. 5-10). Here the magnitude and phase of

$$\frac{\tilde{Y}}{\tilde{S}} = \frac{G_c G_p}{1 + G_c G_p}$$

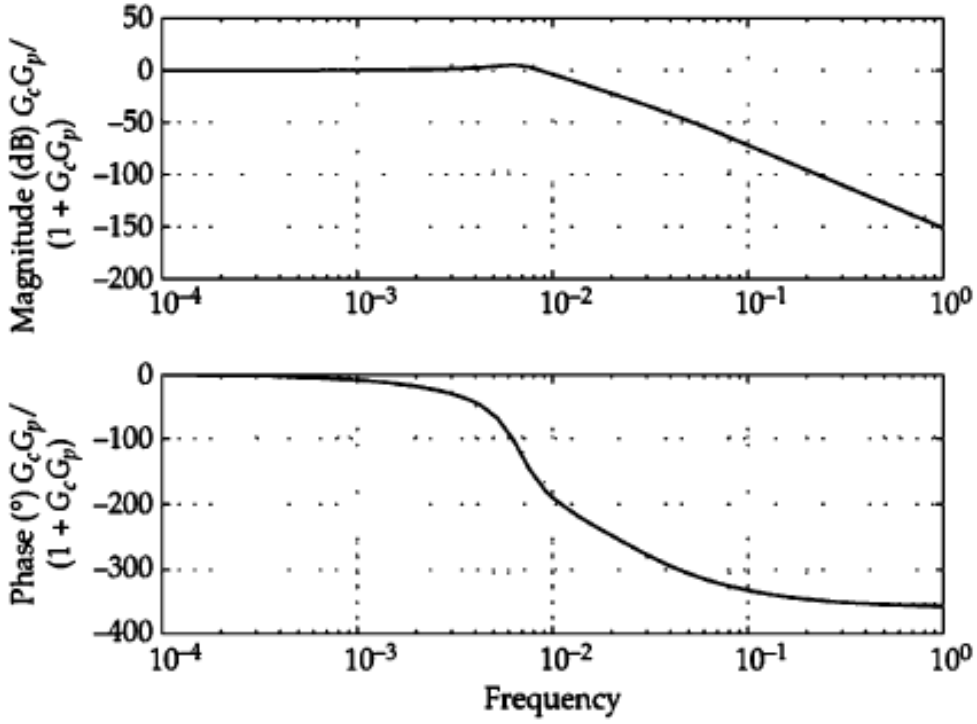


Figure 5-12 Bode plot of no backflow, integral-only control, three-tank, closed-loop.

5-4 Critical Values and Finding the Poles

In Chap. 3, the poles of the appropriate transfer function for the controlled system could be found by solving Eq. (3-50) or

$$G_c G_p = -1 \tag{5-16}$$

In these last couple of examples, we have shown that, after replacing s with $j\omega$, Eq. (5-16) essentially becomes a complex equation

$$G_c(j\omega_c)G_p(j\omega_c) = e^{j\pi} \tag{5-17}$$

where the dependence on the critical radian frequency ω_c is shown. (Note that Eq. (5-17) is a consequence of the expression $e^{j\pi} = -1$ that we mentioned in App. B). Actually, Eq. (5-17) depends on both ω_c and the critical control gain k_c (if the control is proportional-only). If the control is integral-only then the critical control gain would be I_c. Since Eq. (5-17) is now a complex equation, there are real and imaginary parts. Therefore, there are two equations in the two unknowns, ω_c and k_c. This argument suggests that the pole-finding approach and the Bode plot approach are basically the same.

5-5 Multitank Processes

Expand the concept presented in Fig. 5-1 to N tanks, each with no backflow, and specify that all N tanks have the same volume and that the interconnecting piping is the same. Therefore, all tanks will have the same time constant, say 1.0 after scaling, and the same resistance to flow. The ith tank will be described by

$$\tau \frac{dx_i}{dt} + x_i = x_{i-1} \qquad i = 1,\ldots,N$$

CHAPTER 6

An Underdamped Process

6-1 The Dynamics of the Mass/Spring/Dashpot Process

All of the example processes mentioned so far have been "overdamped" in that the open-loop step response does not generate overshoot or oscillations of any kind. The first-order process really has no choice—its behavior is dictated by its gain and time constant. The three tank third-order process has an inflection point in the step response but it will never oscillate or "ring" when subjected to a step change in the process input with no feedback control. These overdamped processes are typical of most of the real-live industrial processes that I faced for most of my career. However, near the end I got involved in some new photonics processes that were underdamped and posed many new challenges.

When we close the control loop on the overdamped processes we could get underdamped and even unstable behavior when the feedback was aggressive but the processes by themselves could not exhibit this kind of performance.

Not so with the so-called mass/spring/dashpot process shown in Fig. 6-1. To derive an equation that describes its behavior one needs to apply Newton's second law of motion:

$$\sum F = m\frac{d^2y}{dt^2} \tag{6-1}$$

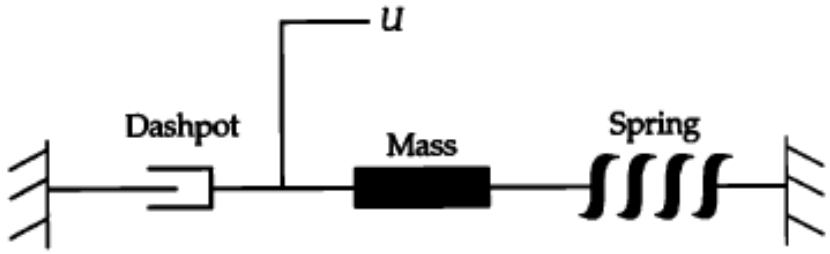


Figure 6-1 Mass/spring/dashpot process.

in $-B(dy/dx)$. The coefficient of friction is B. Finally, the third component is the applied external force U, which is also the process input.

With this in mind, Eq. (6-1) becomes

$$m\frac{d^2y}{dt^2} = -B\frac{dy}{dt} - ky + gU \tag{6-2}$$

By convention, Eq. (6-2) is rewritten as

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n\frac{dy}{dt} + \omega_n^2 y = g\omega_n^2 U$$
$$\omega_n = \sqrt{\frac{k}{m}} \qquad \zeta = \frac{B}{2\sqrt{km}} \tag{6-3}$$

where the damping factor ζ and the natural frequency ω_n appear as functions of the mass, spring constant, and coefficient of friction. When the damping factor ζ varies between 0 and 1 the behavior is underdamped. When $\zeta = 1$ the behavior is critically damped and when $\zeta > 1$ the behavior is overdamped. The natural frequency is effectively the frequency of the "ringing" that the mass experiences after a disturbance. A higher natural frequency means a faster response and higher frequency ringing. The natural frequency has units of radians/sec and is related to f_n, the frequency in cycles/sec, as follows:

$$\omega_n = 2\pi f_n$$

Alternatively, Eq. (6-3) can be written as

$$\tau^2\zeta^2\frac{d^2y}{dt^2} + 2\tau\zeta^2\frac{dy}{dt} + y = gU$$
$$\tau = \frac{2m}{B} \qquad \zeta = \frac{B}{2\sqrt{km}} \tag{6-4}$$

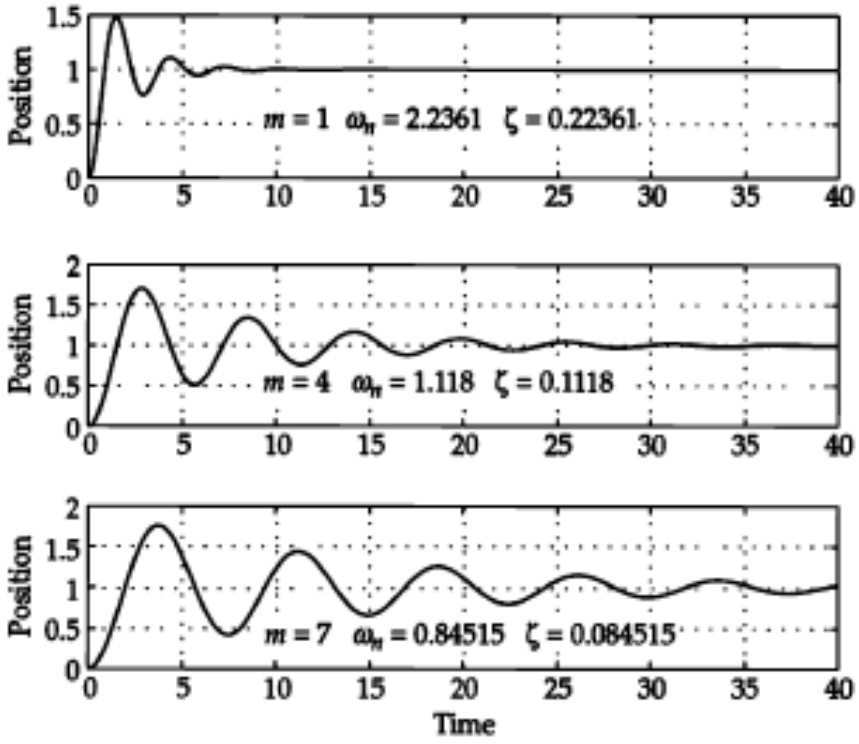


Figure 6-3 Effect of mass on dashpot dynamics.

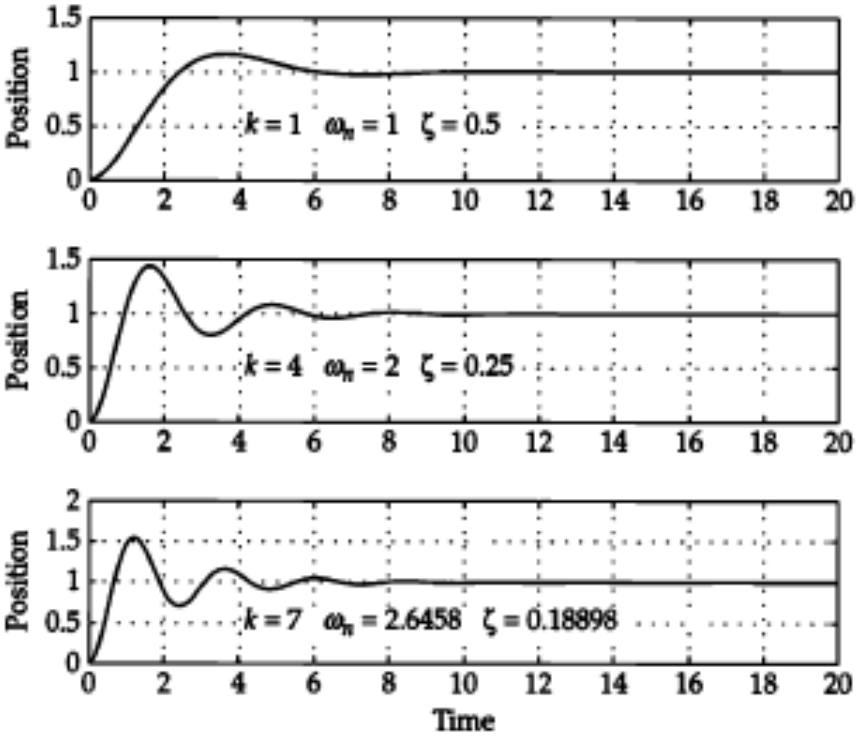


Figure 6-4 Effect of spring constant on dashpot dynamics.

6-2-1 Time Domain

As is our usual approach, we could attempt to solve Eq. (6-3) by trying a solution for the homogeneous part of the form:

$$y_h = Ce^{at}$$

This would generate a quadratic equation for a, which would lead to a homogeneous solution that had two exponential terms. For the time being this approach will be sidestepped.

6-2-2 Laplace Domain Solution

Alternatively, let's go directly to the Laplace domain and take the Laplace transform of Eq. (6-3), as in

$$s^2\tilde{y}(s) + 2\zeta\omega_n s\,\tilde{y}(s) + \omega_n^2\,\tilde{y}(s) = g\,\omega_n^2\,\tilde{U}(s)$$

$$\frac{\tilde{y}(s)}{\tilde{U}(s)} = G_p(s) = \frac{g\,\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{g\,\omega_n^2}{(s-p_1)(s-p_2)} \tag{6-5}$$

The poles of the transfer function are located at the roots of the quadratic in the denominator:

$$p_1,\ p_2 = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1} = a \pm jb$$

If the damping factor ζ is less than unity, these poles become complex conjugates and the solution will contain sinusoidal components suggesting underdamped behavior, as in

$$y_h(t) \sim C_1e^{p_1t} + C_2e^{p_2t} = C_1e^{(a+jb)t} + C_2e^{(a-jb)t}$$

where Euler's formula $e^{a+jb} = e^a[\cos(b) + j\sin(b)]$ can be used to bring in the sinusoids.

Figure 6-6 shows how the roots (or poles) move in the s-plane as the damping factor changes from 0.1 to 1.1. For this case, the natural frequency was kept constant at 100 Hz. When $\zeta = 1.1$, the poles are both real but when $\zeta = 0.1$ both poles nearly lie on the imaginary axis. When $\zeta = 1$ the poles are the same and real.

6-2-3 Frequency Domain

Letting $s = j\omega$ in Eq. (6-5), which gives

$$\frac{\tilde{y}(j\omega)}{\tilde{U}(j\omega)} = \frac{g\omega_n^2}{(j\omega)^2 + 2\zeta\omega_n(j\omega) + \omega_n^2} \tag{6-6}$$

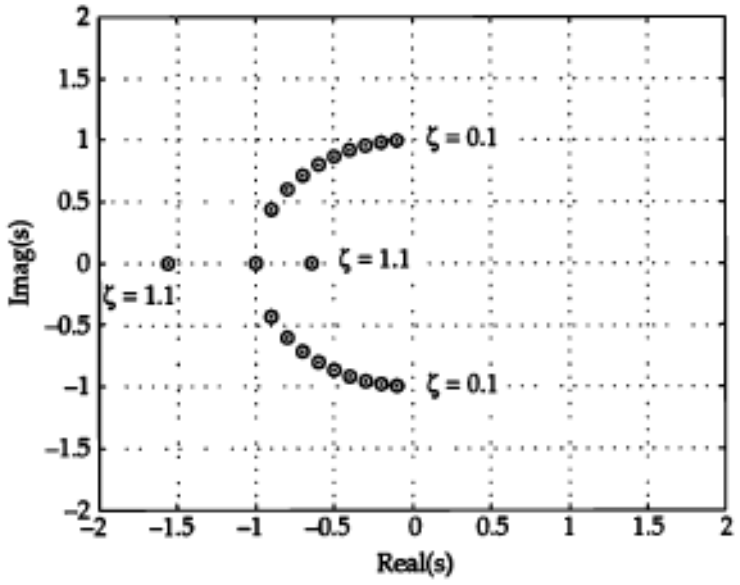


Figure 6-6 Poles of second-order model; $\zeta = 0.1$ to 1.1.

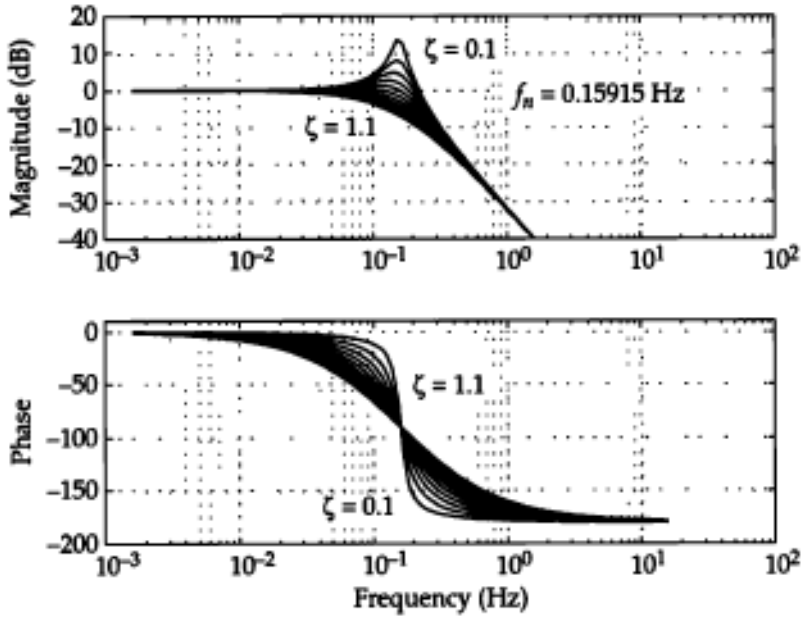


Figure 6-7 Typical second-order Bode diagram showing effect of damping.

6-2-4 State-Space Representation

Let's start with the time domain representation:

$$\frac{d^2y}{dt^2} + 2\zeta\omega_n \frac{dy}{dt} + \omega_n^2 y = g\omega_n^2 U \tag{6-7}$$

we construct two elements of the state as

$$x_1 = y$$

$$x_2 = \frac{dy}{dt}$$

Substituting these in Eq. (6-7) generates two first-order equations

$$\frac{dx_1}{dt} = x_2$$

$$\frac{dx_2}{dt} + 2\zeta\omega_n x_2 + \omega_n^2 x_1 = g\omega_n^2 U$$

These two equations can be written in matrix form as follows:

$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ g\omega_n^2 \end{pmatrix} U$$

$$\frac{dx}{dt} = Ax + BU \tag{6-8}$$

$$A = \begin{pmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{pmatrix}$$

$$B = \begin{pmatrix} 0 \\ g\omega_n^2 \end{pmatrix}$$

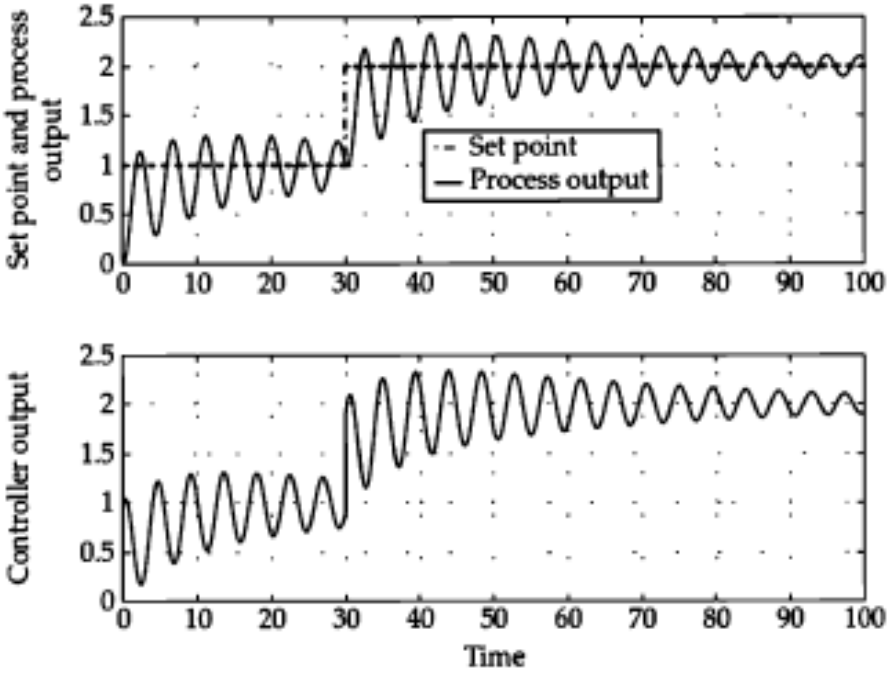
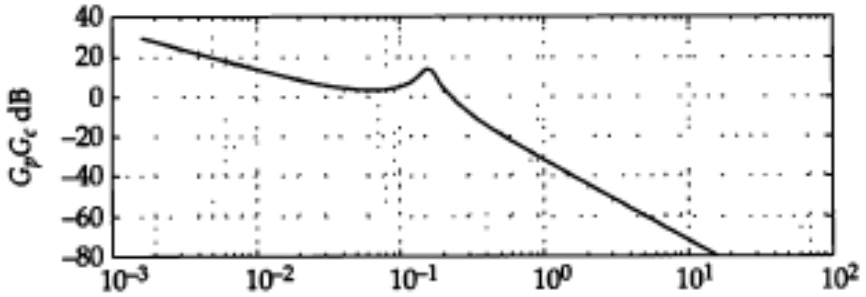


Figure 6-8 Set-point step-change response under PI control.



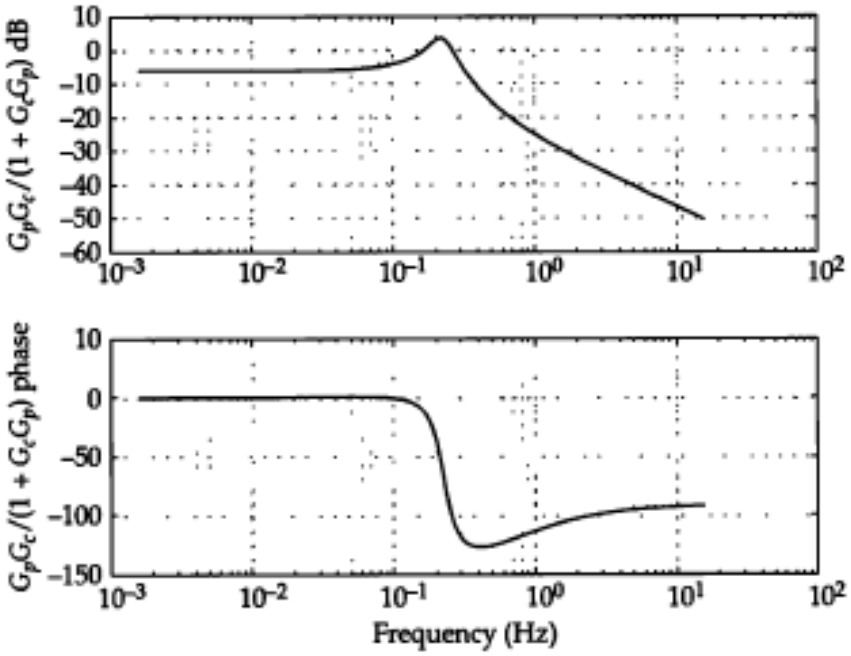


Figure 6-10 Closed-loop Bode plot with PI control—mass/spring/dashpot process.

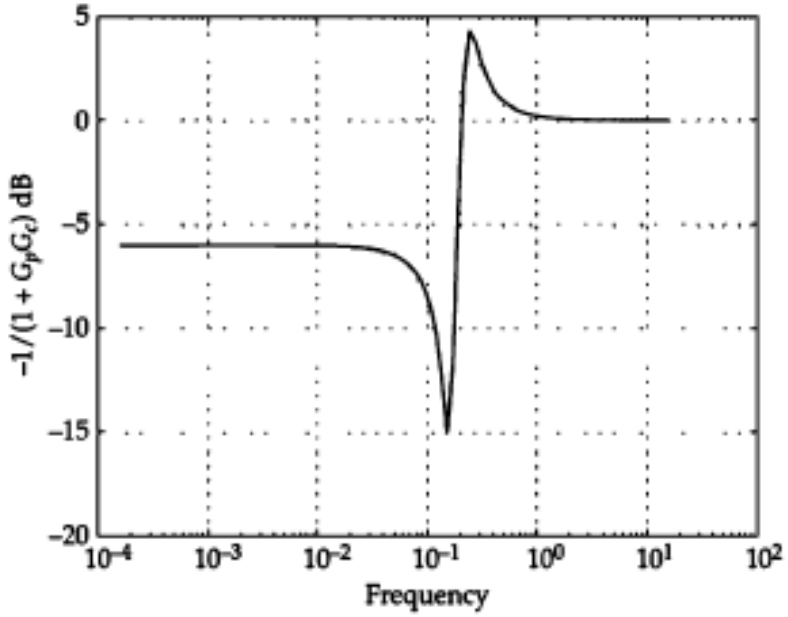


Figure 6-11 Error transmission curve for mass/spring/dashpot with PI control.

6-4 Derivative Control (PID)

The proportional-integral-derivative (PID) control algorithm has an additional term proportional to the derivative of the error:

$$U(t) = K_c\left(e(t) + I\int_0^t due(u) + D\frac{de(t)}{dt}\right)$$

$$\tilde{U}(s) = K_c\left(\tilde{e}(s) + \frac{I}{s}\tilde{e}(s) + Ds\tilde{e}(s)\right)$$

$$= K_c\left(1 + \frac{I}{s} + Ds\right)\tilde{e}(s)$$

$$\frac{\tilde{U}(s)}{\tilde{e}(s)} = G_c = K_c\left(1 + \frac{I}{s} + Ds\right) = K_c\frac{s + I + Ds^2}{s}$$

We have ignored the initial value of $U(t)$ and the presence of the error term in the derivative—both problems will be dealt with in

Chaps. 9 and 11. Furthermore, an overall control gain K_c has been introduced to be consistent with wide usage among control engineers.

Unlike the PI control algorithm, PID has two zeros in the numerator of G_c

$$\frac{\tilde{U}(s)}{\tilde{e}(s)} = G_c = K_c\frac{s + I + Ds^2}{s} = K_c\frac{(s - s_1)(s - s_2)}{s}$$

$$s_1,\ s_2 = \frac{-1 \pm \sqrt{1 - 4DI}}{2D}$$

which can be complex conjugates if $4DI > 1$. Therefore, these potentially complex zeroes in G_c *might* ameliorate the presence of the complex poles in G_p:

$$G_pG_c = \frac{1}{s^2 + 2\zeta s + 1}K_c\frac{s + I + Ds^2}{s}$$

Tuning the PID algorithm for the dashpot process was done by trial and error. We kept the proportional and integral gains of the previous simulation for PI and started with a conservative value for D and increased it until satisfactory control was obtained with $D = 4.0$. Figure 6-12 shows the poles of G_p and the zeroes of G_c for the PID controller and for the PI controller used in Sec. 6-3. Figure 6-13 shows the poles of closed-loop transfer function $(G_pG_c)/[1 + G_pG_c]$.

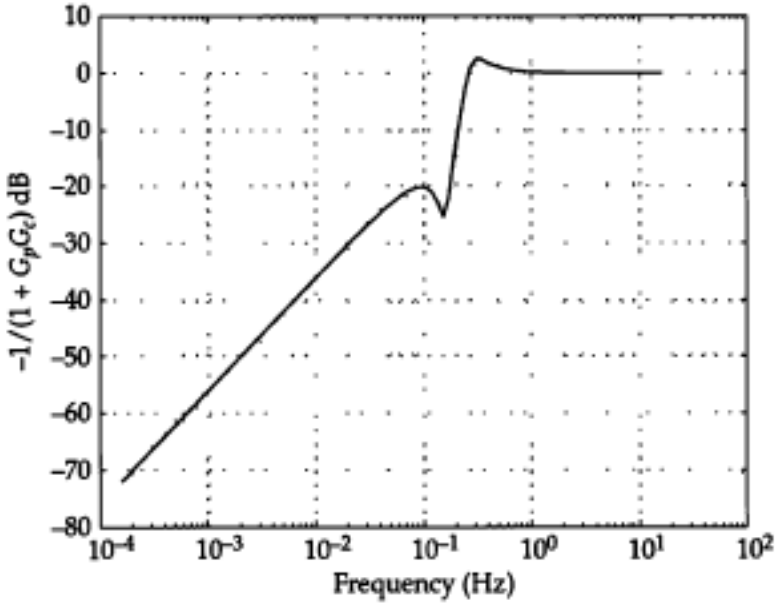


Figure 6-17 Error transmission curve for dashpot with PID control—mass/spring/dashpot process with derivatives.

6-4-1 Complete Cancellation

Perhaps the reader is wondering: what would happen if the zeros of the PID controller were chosen to exactly match those of the process? That is, what if:

$$\frac{-1 \pm \sqrt{1-4DI}}{2D} = -\zeta \pm \sqrt{\zeta^2 - 1}$$

$$D = \frac{1}{2\zeta} \qquad I = \frac{1}{2\zeta}$$

This would cause the open-loop transfer function to become

$$G_p G_c = \frac{1}{s^2 + 2\zeta s + 1} K_c \frac{s + I + Ds^2}{s} = \frac{K_c}{s}$$

and the closed-loop transfer function would be

$$\frac{Y}{S} = \frac{G_c G_p}{1 + G_c G_p} = \frac{\frac{K_c}{s}}{1 + \frac{K_c}{s}} = \frac{1}{\frac{s}{K_c} + 1}$$

6-4-2 Adding Sensor Noise

At this point, as a manager, you might be impressed to the point where you would conclude that the addition of derivative was the best thing since sliced bread (aside from the preceding comments about the extreme response to set-point steps). However, when the process output is noisy, troubles arise. For the purposes of this simulation exercise, we will add just a little white sensor noise (to be defined later) to the PI and the PID simulations. Figures 6-18 and 6-19 show the impact of adding a small amount of sensor noise on the process output signal for PI and PID. The added noise is barely discernible when PI control is used but when the same amount of

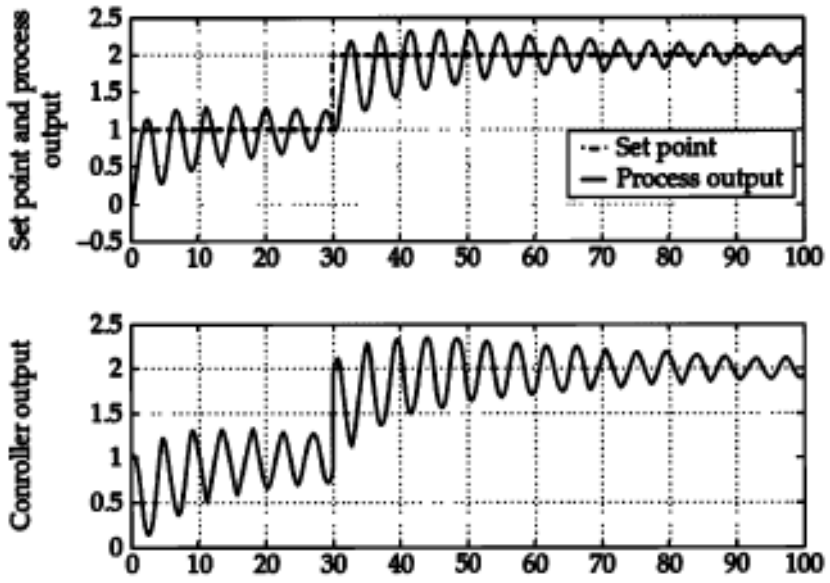
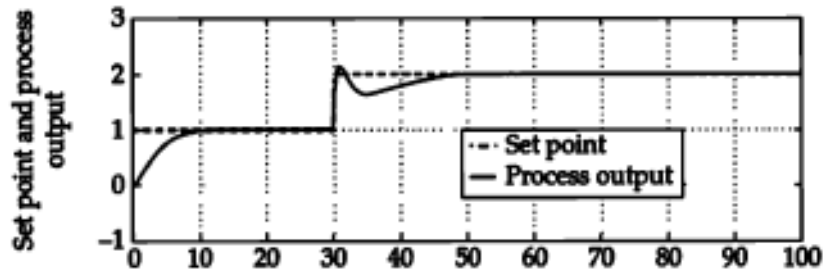


Figure 6-18 Response of PI control with added noise.

process noise is added to PID control there is quite a change in the control output. The addition of the derivative component to the control algorithm still drives the process output to set point without oscillation, but there is a tremendous price to pay in the activity of the control output. Also, note the spike in the output at $t = 30$ when the



6-4-3 Filtering the Derivative

The moral of this short story is to be careful about adding derivative because it greatly amplifies noise and sudden steps. Adding a first-order filter (with a time constant of 1.0) to the derivative partially addresses the problem as shown in Fig. 6-21. The outrageous control output activity has been ameliorated but there is still ringing.

Using the Laplace transform is the easiest way to present the filtered derivative:

$$\frac{\tilde{U}(s)}{\tilde{e}(s)} = G_c = K_c\left(1 + \frac{I}{s} + D\frac{s}{\tau_D s + 1}\right)$$

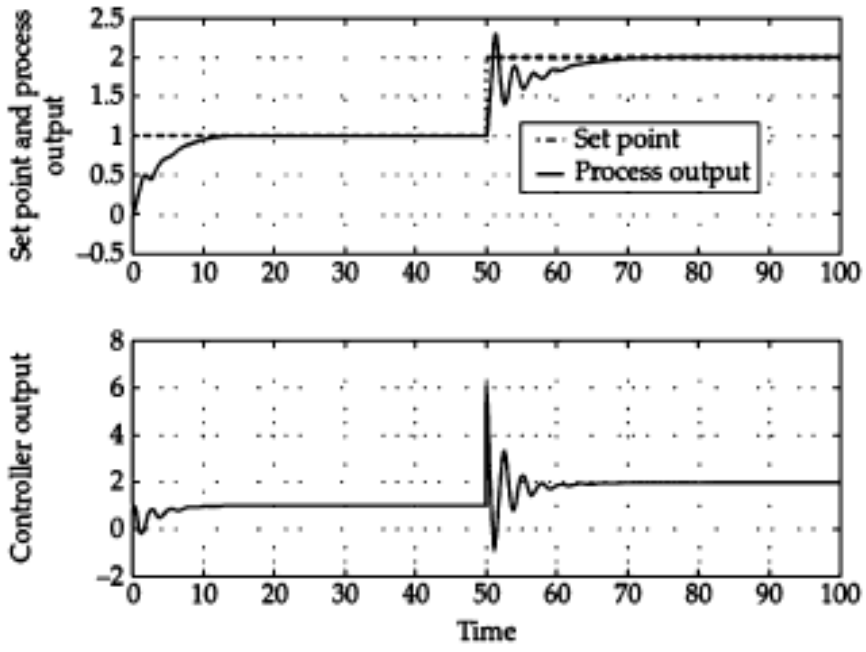


Figure 6-21 PIfD set-point step-change response with added noise.

Note the presence of the $1/(\tau_D s + 1)$ factor in the derivative term. For our simulation, the filter time constant τ_D was chosen to be equal to $1/\omega_u = 1$. Since this algorithm will most likely be implemented as a digital filter, its detailed discussion will be deferred until the discrete time domain is introduced in Chap. 9. However, why do you suppose that modifying the derivative term by the factor:

$$\frac{1}{\tau_D s + 1}$$

6-5 Compensation before Control—The Transfer Function Approach

Since the dashpot process has given us so much trouble, another approach will be taken in this section. We are going to modify the process by feeding the process variable and its first derivative back with appropriate gains. The gains will be chosen to make the modified process behave in a way more conducive to control.

Without compensation, the dashpot Laplace transform from Eq. (6-5) is

$$s^2\tilde{y}(s) + 2\zeta\omega_n s\,\tilde{y}(s) + \omega_n^2\,\tilde{y}(s) = g\,\omega_n^2\,\tilde{U}(s) \tag{6-9}$$

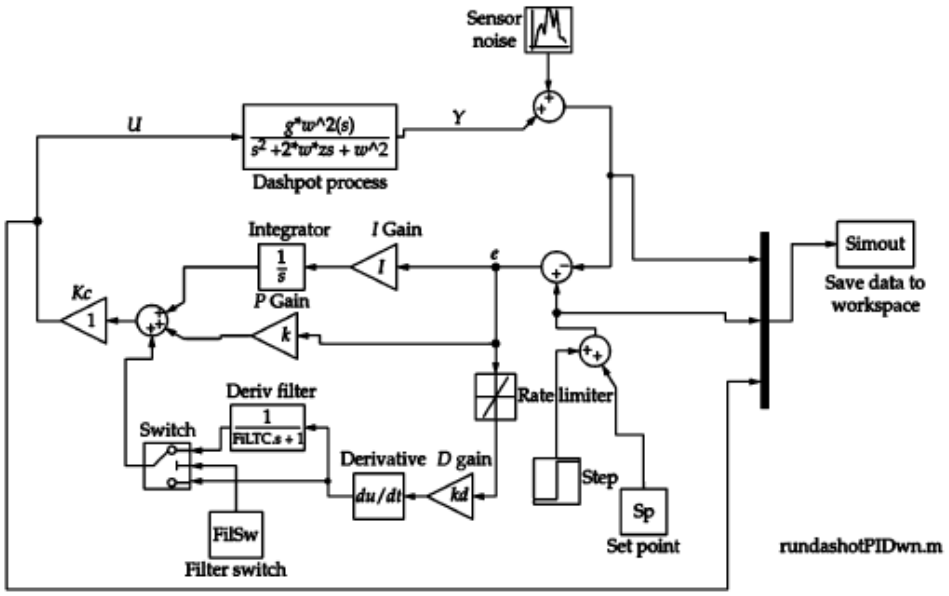


Figure 6-23 `Simulink` block diagram for dashpot control.

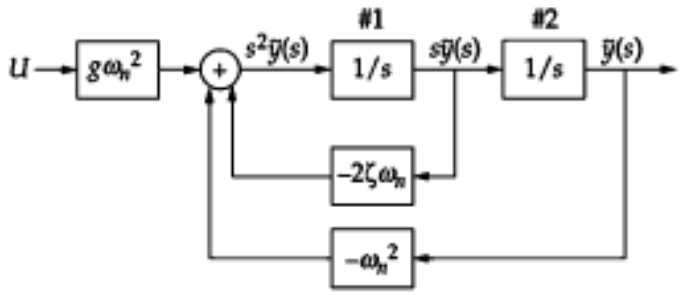


Figure 6-24 The dashpot model before compensation.

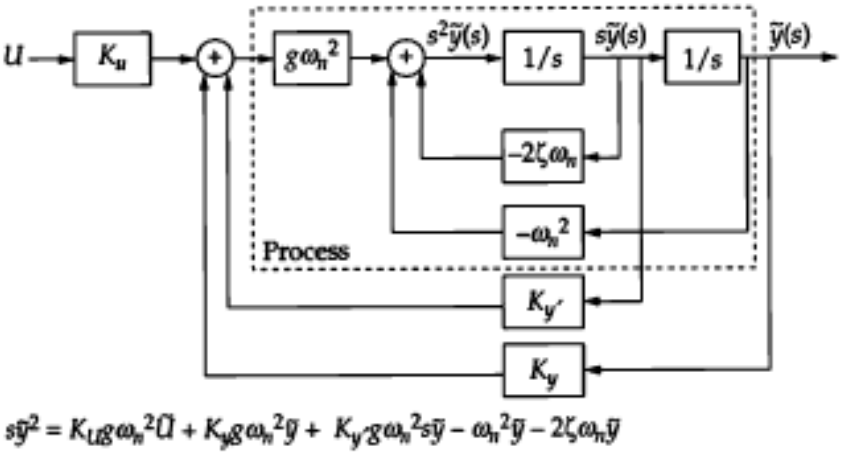


Figure 6-25 The dashpot model with states fed back.

Figure 6-25 shows a modified block diagram where y and dy/dx are fed back *again*, this time with gains K_y and $K_{y'}$. Note that no control is being attempted yet. We are feeding these signals back to create a new *modified* process that will have more desirable properties. Everything inside the dotted line box represents the structure of the original process. All the lines and blocks outside the box represent the added compensation. The Laplace transform of the modified system is

$$s^2\tilde{y} = K_U g\omega_n^2\tilde{U} + K_y g\omega_n^2\tilde{y} + K_{y'} g\omega_n^2 s\tilde{y} - \omega_n^2\tilde{y} - 2\zeta\omega_n\tilde{y} \qquad (6\text{-}11)$$

The logic behind the structure of this block diagram is the same as that for the unmodified process shown in Fig. 6-24. Three gains, K_y, $K_{y'}$, and K_U have been introduced.

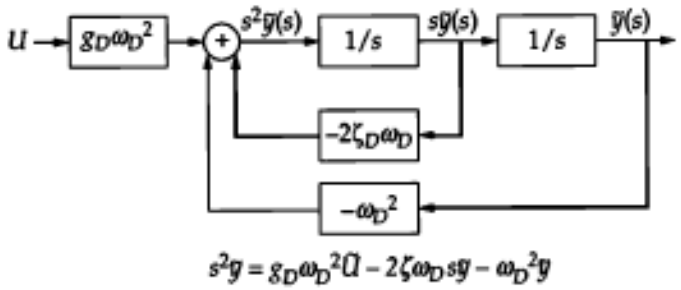
The values for these gains will be chosen so that the modified process looks like a *desired* process shown in Fig. 6-26. The Laplace transform for the desired system is

$$s^2\tilde{y}(s) = -2\zeta_D\omega_D s\tilde{y}(s) - \omega_D^2\tilde{y}(s) + g_D\omega_D^2\tilde{U}(s) \qquad (6\text{-}12)$$

K_y, $K_{y'}$, and K_U have been introduced.

The values for these gains will be chosen so that the modified process looks like a *desired* process shown in Fig. 6-26. The Laplace transform for the desired system is

$$s^2\tilde{y}(s) = -2\zeta_D\omega_D s\,\tilde{y}(s) - \omega_D^2\,\tilde{y}(s) + g_D\omega_D^2\,\tilde{U}(s) \qquad (6\text{-}12)$$



$$s^2y = g_D\omega_D^2\tilde{U} - 2\zeta\omega_D sy - \omega_D^2 y$$

Figure 6-26 The desired dashpot model.

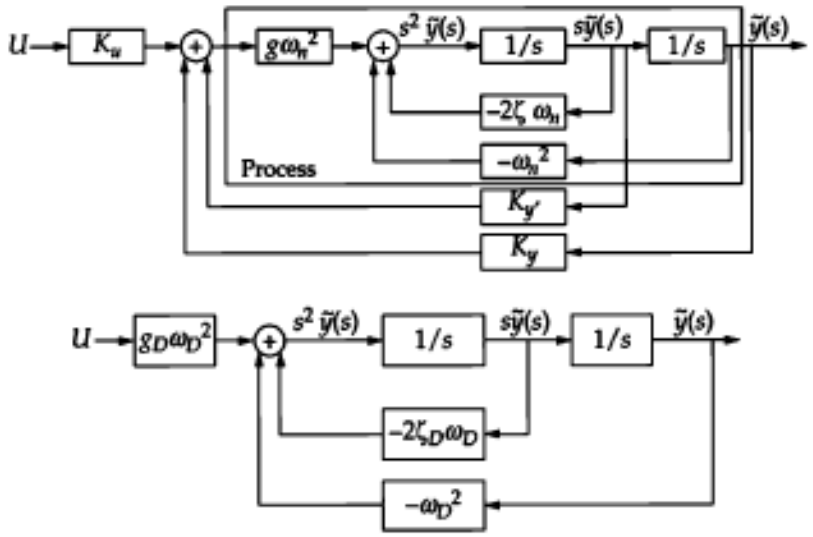


Figure 6-27 Choose K_u, $K_{y'}$, K_y to make the compensated and desired models identical.

Note that this desired process has the same structure as the original process but the parameters, g_D, ζ_D, and ω_D are yet to be specified. We will specify the values and then find the values of K_y, $K_{y'}$, and K_U that will make them happen (Fig. 6-27). As you might expect, we would want the damping parameter ζ_D to be greater than that of the original process so that there is less ringing. Likewise, we might want to make the natural frequency ω_D greater than ω_n so that the response would be quicker. To make life simple, g_D is chosen to be unity.

$$s^2\tilde{y} = K_U g\omega_n^2\tilde{U} + \left(K_y g\omega_n^2 - \omega_n^2\right)\tilde{y} + \left(K_{y'} g\omega_n^2 - 2\zeta\omega_n\right)s\tilde{y}$$
$$s^2\tilde{y}(s) = g_D\omega_D^2\tilde{U} - \omega_D^2\tilde{y}(s) - 2\zeta_D\omega_D s\,\tilde{y}(s) \qquad (6\text{-}13)$$

Comparing the coefficients of $\tilde{U}$, $\tilde{y}$, and $s\tilde{y}$ gives the following expressions.

$$K_U g\omega_n^2 = g_D\omega_D^2$$
$$K_y g\omega_n^2 - \omega_n^2 = -\omega_D^2$$
$$K_{y'} g\omega_n^2 - 2\zeta\omega_n = -2\zeta_D\omega_D$$

which can be solved for K_y, $K_{y'}$, and K_u as in

$$K_U = \frac{g_D\omega_D^2}{g\omega_n^2}$$
$$K_y = \frac{\omega_n^2 - \omega_D^2}{g\omega_n^2} \qquad (6\text{-}14)$$
$$K_{y'} = \frac{2\zeta\omega_n - 2\zeta_D\omega_D}{g\omega_n^2}$$

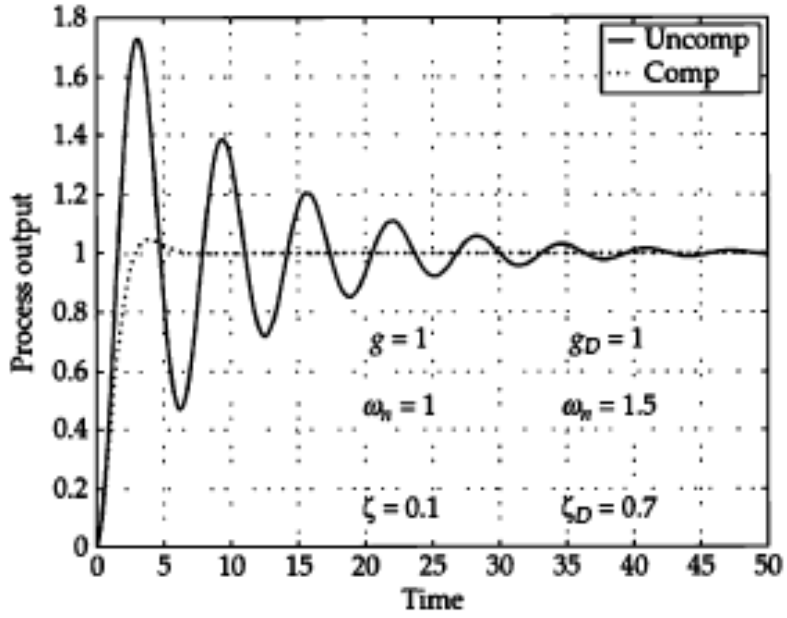


Figure 6-28 Effect of compensation.

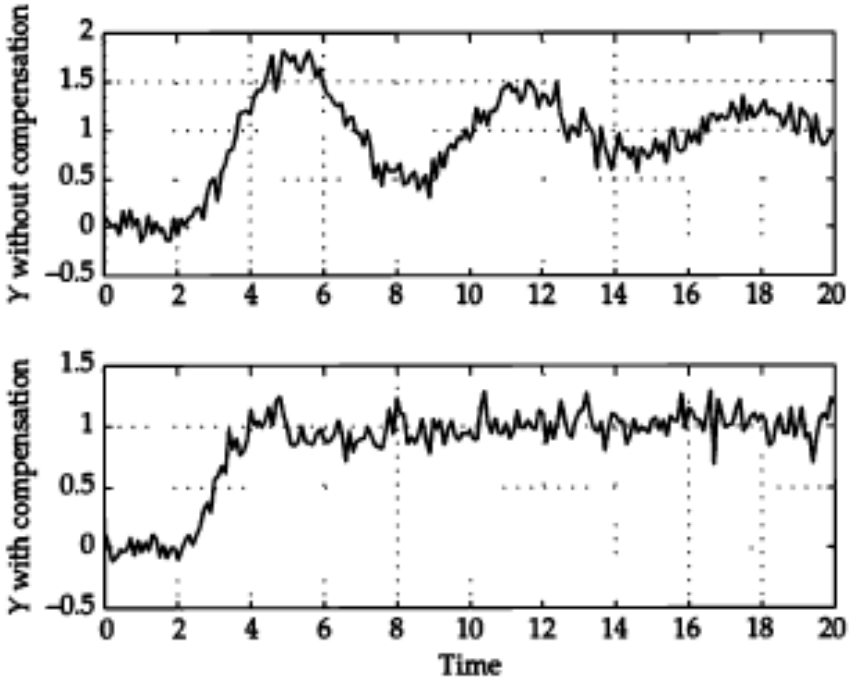


Figure 6-29 Effect of compensation in the face of noise using a filtered derivative for *y'*.

6-6 Compensation before Control—The State-Space Approach

The state-space model for the dashpot process is

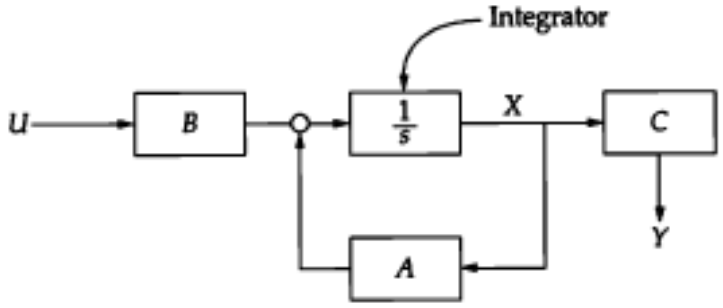
$$\frac{d}{dt}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} 0 \\ g\omega_n^2 \end{pmatrix} U$$

$$\frac{dx}{dt} = Ax + BU$$

$$s\tilde{x} = A\tilde{x} + B\tilde{U} \tag{6-15}$$

$$A = \begin{pmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{pmatrix}$$

$$B = \begin{pmatrix} 0 \\ g\omega_n^2 \end{pmatrix}$$



$$s\tilde{X} = A\tilde{X} + B\tilde{U}$$

$$\tilde{X} = \frac{1}{s}(A\tilde{X} + B\tilde{U}) \qquad X = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \text{Position} \\ \text{speed} \end{pmatrix}$$

Figure 6-30 A state-space block diagram.

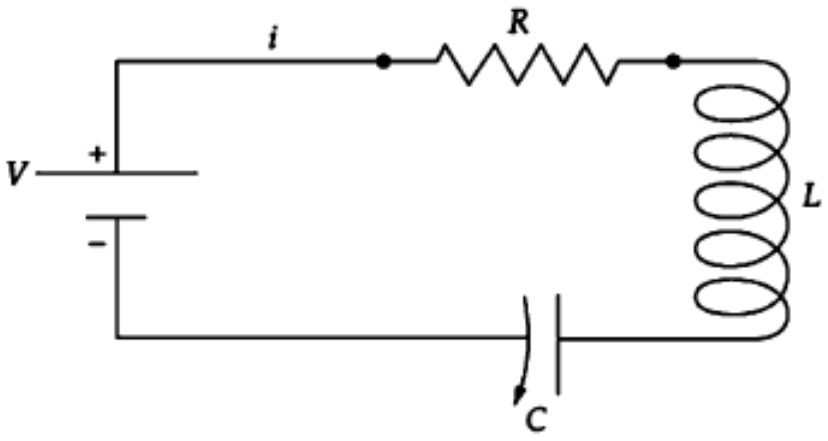
6-7 An Electrical Analog to the Mass/Dashpot/Spring Process

Consider the *RLC* circuit in Fig. 6-32 where R refers to resistance of the resistor, L the inductance of the coil, and C the capacitance of the capacitor. The applied voltage is V and it will also be the process input U. The voltage over the resistor is iR where i is the process output Y. The voltage over the capacitor is

$$\frac{1}{C}\int_0^t i(u)du$$

and the voltage over the inductor is

$$L\frac{di}{dt}$$



These three voltages have to add up to match the applied voltage.

$$V = iR + \frac{1}{C}\int_0^t i(u)du + L\frac{di}{dt} \tag{6-20}$$

Eq. (6-20) could be differentiated to get rid of the integral. Alternatively, the equation could be transformed to the Laplace domain yielding

$$\tilde{V} = \tilde{i}R + \frac{\tilde{i}}{Cs} + Ls\tilde{i}$$

The output/input transfer function is

$$\frac{\tilde{i}}{\tilde{V}} = \frac{\tilde{Y}}{\tilde{U}} = G = \frac{1}{R + \frac{1}{Cs} + Ls} = \frac{Cs}{LCs^2 + RCs + 1} = \frac{\frac{1}{L}s}{s^2 + \frac{R}{L}s + \frac{1}{LC}}$$

This expression looks similar to Eq. (6-5), which is repeated here as

$$\frac{\tilde{y}(s)}{\tilde{U}(s)} = G_p(s) = \frac{g\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

This suggests that

$$\omega_n^2 \Leftrightarrow \frac{1}{LC} \qquad 2\zeta\omega_n \Leftrightarrow \frac{R}{L}$$

which further suggests

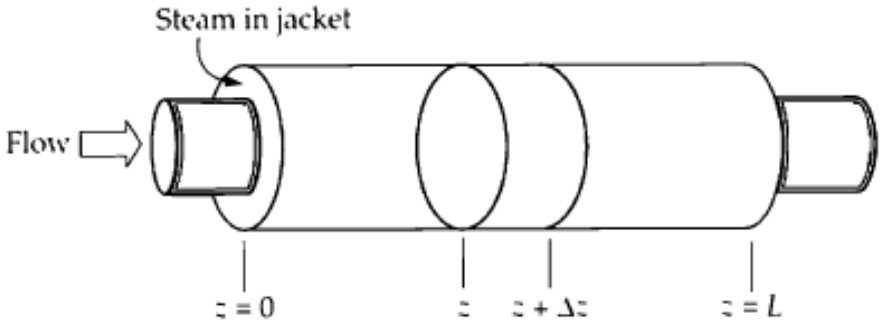
$$\omega_n = \sqrt{\frac{1}{LC}} \qquad \zeta_n = \frac{R}{2}\sqrt{\frac{C}{L}}$$

Therefore, the *RLC* process has the potential of behaving in an underdamped manner similar to that of the mass/dashpot/spring process. For example, with R, C, and L chosen such that $\varsigma < 1$, the step response will exhibit damped oscillations with a frequency of ω_n.

7-1 The Tubular Energy Exchanger—Steady State

Consider Fig. 7-1 which shows a jacketed tube of length L. A liquid flows through the inside tube. The jacket contains a fluid, say steam, from which energy can be transferred to the liquid in the tube. To describe how this process behaves in steady state, a simple energy balance can be made, not over the whole tube but over a small but finite section of the tube. Several assumptions (and idealizations) must be made about this new process.

1. The steam temperature T_s in the jacket is constant along the whole length of the tube. The tube length is L. The steam temperature can vary with time but not space.
2. The tube is cylindrical and has a cross-sectional area of $A = \pi D^2 / 4$ where D is the diameter of the inner tube.
3. The liquid flows in the tube as a plug at a speed v. That is, there is no radial variation in the liquid temperature. There is axial temperature variation of the liquid due to the heating effect of the steam in the jacket but there is no axial transfer of energy by conduction within the fluid. This is equivalent to saying the radial diffusion of energy is infinite compared to axial diffusion. The temperature of the flowing liquid therefore is a function of the axial displacement z, as in $T(z)$.



4. There is a small disc placed at some arbitrary location z along the tube that has cross-sectional area A and thickness Δz. This disc will be used to derive the model describing equation.
5 The liquid properties of density ρ, heat capacity C_p, thermal conductivity k are constant (independent of position and of temperature).
6. The flux of energy between the steam in the jacket and the flowing liquid is characterized by an overall heat transfer coefficient U.

A thermal energy balance over the disc of thickness Δz at location z will describe the steady-state behavior of the tube exchanger. The result is given in Eq (7-1) which is boxed below. You might want to skip to that location if derivations are not your bag. Otherwise, the derivation proceeds as follows

Energy rate in at z due to convection: $vA\rho C_p T(z)$

Energy rate out at $z+\Delta z$ due to convection: $vA\rho C_p T(z+\Delta z)$

Energy rate in from jacket: $U(\pi D\Delta z)\left[T_c - T\left(\frac{z+z+\Delta z}{2}\right)\right]$

In this last term the energy rate is proportional to the difference between the jacket temperature T_c and the liquid temperature in the middle of the disc, at the point

$$(z+z+\Delta z)/2$$

The energy balance then becomes

$$vA\rho C_p T(z) + U(\pi D\Delta z)\left[T_c - T\left(\frac{z+z+\Delta z}{2}\right)\right] = vA\rho C_p T(z+\Delta z)$$

After a slight rearrangement and after dividing all terms by Δz one gets

$$\frac{vA\rho C_p T(z+\Delta z) - vA\rho C_p T(z)}{\Delta z} = U\pi D\left[T_c - T\left(\frac{z+z+\Delta z}{2}\right)\right]$$

The thickness of the disc is decreased to differential or infinitesimal size as in

$$\lim_{\Delta z \to 0} \frac{vA\rho C_p T(z+\Delta z) - vA\rho C_p T(z)}{\Delta z} = \lim_{\Delta z \to 0} U\pi D \times \left[T_s - T\left(\frac{z+z+\Delta z}{2}\right)\right]$$

From App. A one sees that the above equation contains the definition of the derivative of T with respect to z, as in

$$vA\rho C_p \frac{dT}{dz} = U\pi D[T_s - T(z)] \tag{7-1}$$

This ordinary differential equation describes the steady-state behavior of the idealized jacketed tube energy exchanger. From Chap. 3 we already know how to solve this equation if we know an inlet temperature, as in $T(0) = T_0$.

If Eq. (7-1) is rearranged slightly, the reader can see the similarity to the equation for the liquid tank presented in Chap. 3.

$$\frac{vA\rho C_p}{U(\pi D)} \frac{dT}{dz} + T = T_s$$

$$\psi \frac{dT}{dz} + T = T_s \tag{7-2}$$

$$\psi = \frac{vA\rho C_p}{U\pi D} = \frac{vD\rho C_p}{4U}$$

The reader has seen Eq. (7-2) before, at least structurally. By inspection, the reader can arrive at a solution to Eq. (7-2) as

$$T(z) = T_0 e^{-\frac{z}{\psi}} + \left(1 - e^{-\frac{z}{\psi}}\right) T_s \tag{7-3}$$

The parameter ψ can be considered as a kind of "space constant," somewhat analogous to the time constant used in transient analysis. In fact, ψ is the tube length needed for $T(z)$ to reach 63% of the jacket temperature T_s.

7-2 The Tubular Energy Exchanger—Transient Behavior

The dynamic behavior can be described by a partial differential equation that also evolves from a thermal energy balance over a small disc of length Δz located somewhere in the interior of the tube and over a moment in time of length Δt. The balance proceeds as in Sec. 7-1 but with one more term—the temporal accumulation of thermal energy in the disc. The temperature now depends on both the axial distance z and the time t, as in $T(z,t)$. Furthermore, the jacket temperature T_s may now depend on time but, as specified above, it is not a function of axial position. A second balance could be written for the steam in the jacket; however, for the time being, the dynamics of the steam are

assumed to be much quicker than those of the liquid flowing through the tube.

As with the steady-state derivation in Sec. 7-1, the reader can skip to the result in Eq. (7-6) which is boxed. For the adventurous, the balance proceeds as follows.

Energy rate at z due to convection at time t during the interval Δt:

$$vA\rho C_p T(z,t)\Delta t$$

Energy rate out at $z+\Delta z$ due to convection at time t during the interval Δt:

$$vA\rho C_p T(z+\Delta z,t)\Delta t$$

Energy rate in from jacket at time t during the interval Δt:

$$U(\pi D\Delta z)\left[T_s - T\left(\frac{z+z+\Delta z}{2},t\right)\right]\Delta t$$

Energy rate in from jacket at time t during the interval Δt:

$$U(\pi D\Delta z)\left[T_s - T\left(\frac{z+z+\Delta z}{2}, t\right)\right]\Delta t$$

Accumulation of energy in the disc between time t and time $t+\Delta t$ in the volume $A\Delta z$:

$$A\Delta z\rho C_p T\left(\frac{z+z+\Delta z}{2}, t+\Delta t\right) - A\Delta z\rho C_p T\left(\frac{z+z+\Delta z}{2}, t\right)$$

Entering the various elements into the balance equation

$$\text{In} - \text{out} = \text{accumulation}$$

gives

$$vA\rho C_p T(z,t)\Delta t + U(\pi D\Delta z)\left[T_s - T\left(\frac{z+z+\Delta z}{2}, t\right)\right]\Delta t - vA\rho C_p T(z+\Delta z, t)\Delta t$$

$$= A\Delta z\rho C_p T\left(\frac{z+z+\Delta z}{2}, t+\Delta t\right) - A\Delta z\rho C_p T\left(\frac{z+z+\Delta z}{2}, t\right)$$

Dividing by $\Delta t\Delta z$ and doing a little rearranging gives

$$vA\rho C_p \frac{T(z,t) - T(z+\Delta z, t)}{\Delta z} + U(\pi D)\left[T_s - T\left(\frac{z+z+\Delta z}{2}, t\right)\right]$$

$$= A\rho C_p \frac{T\left(\frac{z+z+\Delta z}{2}, t+\Delta t\right) - T\left(\frac{z+z+\Delta z}{2}, t\right)}{\Delta t}$$

If the space element Δz and the time element Δt are both decreased to an infinitesimally small size, then the following partial differential equation results.

$$-vA\rho C_p \frac{\partial T}{\partial z} + U(\pi D)[T_s - T(z,t)] = A\rho C_p \frac{\partial T}{\partial t} \tag{7-4}$$

Dividing all terms by $A\rho C_p$ and remembering that $A = \pi D^2/4$, gives

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial z} = \frac{4U}{D\rho C_p}[T_s - T(z,t)] \tag{7-5}$$

The quantity $D\rho C_p / 4U$ has units of sec, so Eq. (7-5) could be written as

$$\begin{aligned} &\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial z} = \frac{1}{\tau_T}[T_s - T(z,t)] \\ &\tau_T = \frac{D\rho C_p}{4U} = \frac{\psi}{v} \end{aligned} \tag{7-6}$$

where τ_T has units of time and is a time constant. Equation (7-6) is a *partial* differential equation describing the time–space behavior of the temperature in the tube. It is subject to initial conditions, such as $T(z,0) = T_0$, $0 \le z \le L$, and a boundary condition on the inlet, such as $T(0,t) = T_i$, $t > 0$. Since we have added the dependence on time, this process model can be used in simulations to test control algorithms.

As an aside, the quantity

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial z}$$

is often called the *total derivative* or the *convective derivative* of temperature and is sometimes given the symbol DT / Dt.

7-2-1 Transfer by Diffusion

The model in Eq. (7-6) describes the transfer of energy along the tube by convection. Energy can also be transported axially by molecular diffusion where the rate is proportional to the axial gradient of temperature, as in $-k(\partial T / \partial z)$ where k is the thermal conductivity. If one modifies the above energy balance on an element of length Δz by adding the contribution of diffusion, the result is

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial z} = \frac{4U}{D\rho C_p}[T_s - T(z,t)] + \frac{k}{\rho C_p}\frac{\partial^2 T}{\partial z^2} \tag{7-7}$$

where the added mechanism of transport is described by the term $(k / \rho C_p) / (\partial^2 T / \partial z^2)$. The presence of this term makes the solution procedure significantly more difficult and we will not refer to Eq. (7-7) until later in this chapter when lumping is discussed.

7-3 Solution of the Tubular Heat Exchanger Equation

There are a variety of approaches to solve Eq. (7-6) but we will pick the one using the tools already developed in this book and the one that will lend itself to using the frequency domain to gain insight. This means transforming the time dependence out of Eq. (7-6) using the Laplace transform. This will leave us with a first-order ordinary differential equation in the spatial dimension z which we can solve using standard techniques. The details are given in App. F.

The result of applying the Laplace transform to Eq. (7-6) is

$$s\tilde{T} + v\frac{d\tilde{T}}{dz} = \frac{1}{\tau_T}(\tilde{T}_s - \tilde{T}) \tag{7-8}$$

7-3 Solution of the Tubular Heat Exchanger Equation

There are a variety of approaches to solve Eq. (7-6) but we will pick the one using the tools already developed in this book and the one that will lend itself to using the frequency domain to gain insight. This means transforming the time dependence out of Eq. (7-6) using the Laplace transform. This will leave us with a first-order ordinary differential equation in the spatial dimension z which we can solve using standard techniques. The details are given in App. F.

The result of applying the Laplace transform to Eq. (7-6) is

$$s\tilde{T} + v\frac{d\tilde{T}}{dz} = \frac{1}{\tau_T}(\tilde{T}_s - \tilde{T}) \tag{7-8}$$

You should convince yourself that Eq. (7-8) is indeed the result of multiplying Eq. (7-6) by $\exp(-st)$ and integrating over $[0,\infty]$ with respect to time. In any case, after the dust has settled, Eq. (7-8) is a first-order ordinary differential equation of the form

$$v\frac{d\tilde{T}}{dz} + \left(s + \frac{1}{\tau_T}\right)\tilde{T} = \frac{\tilde{T}_s}{\tau_T} \tag{7-9}$$

where the Laplace variable s is just a parameter. Remember that $\tilde{T}_s$ is the Laplace transform of the jacket temperature which we specified could be a function of time but not of axial position, that is, T_s or $\tilde{T}_s$ is not a function of z.

Now, how do we solve Eq. (7-9)? We could apply the Laplace transform again with a different variable, say p, instead of s and remove the spatial dimension or we could solve the ordinary differential equation by trying a solution of the form $Ce^{\alpha z}$. Both of these approaches have been used elsewhere in this book. The details of the solution are presented in App. F and the result is

$$\tilde{T}(z,s) = \tilde{T}_0 e^{az} + \tilde{T}_s \frac{1 - e^{az}}{\tau_T s + 1} \tag{7-10}$$

where

$$a = -\frac{1}{v}\left(s + \frac{1}{\tau_T}\right)$$

7-3-1 Inlet Temperature Transfer Function

Equation (7-10) contains two transfer functions of interest. The first transfer function shows how the inlet temperature affects the outlet temperature (at $z = L$):

$$\frac{\tilde{T}(L,s)}{\tilde{T}_0(s)} = e^{-\frac{L}{v}\left(s+\frac{1}{\tau_T}\right)} = e^{-s\frac{L}{v}} e^{-\frac{L}{v\tau_T}} = e^{-st_D} e^{\frac{t_D}{\tau_T}} \tag{7-11}$$

where $t_D = L/v$ is the average residence time or delay time for the tube. Equation (7-11) ignores the impact of T_s and shows that $T(L,t)$ lags $T(0,t)$ by t_D and is attenuated by a constant factor of e^{-t_D/τ_T}. This makes physical sense based on the assumptions of plug flow for the liquid. Thus, when T_0 is the input, Eq. (7-11) suggests that the response of $T(L,t)$ behaves as dead-time process with an attenuation factor.

Question 7-1 What does a time plot of this response look like and is it physically realistic?

Answer A sharp step in the inlet propagates through the reactor as a sharp step in the liquid temperature. Thus plug flow is idealistic because there is bound to be some axial mixing either from turbulence or diffusion. If Eq. (7-7) were solved, the propagation would be more realistic with less sharpness. Later when lumping is discussed this issue of idealistic sharpness will be revisited.

7-3-2 Steam Jacket Temperature Transfer Function

Equation (7-10) yields a second transfer function relating the steam jacket temperature to the outlet temperature.

$$\frac{\tilde{T}(L,s)}{\tilde{T}_s(s)} = \frac{1 - e^{-st_D} e^{-\frac{t_D}{\tau_T}}}{\tau_T s + 1} \tag{7-12}$$

We will use this transfer function later on when assessing the feasibility of controlling the outlet temperature by manipulating the steam temperature. The denominator of Eq. (7-12) has appeared before so we can expect τ_T to act as a time constant in a way similar to previous transient analyses.

7-4 Response of Tubular Heat Exchanger to Step in Jacket Temperature

Let the jacket temperature be a step of size U_c at time zero. Equation (7-12) becomes

$$\tilde{T}(L,s) = \frac{1 - e^{-st_D} e^{\frac{t_D}{\tau_T}}}{\tau_T s + 1} \frac{U_c}{s} \tag{7-13}$$

Appendix F shows that the inversion of Eq. (7-13) gives

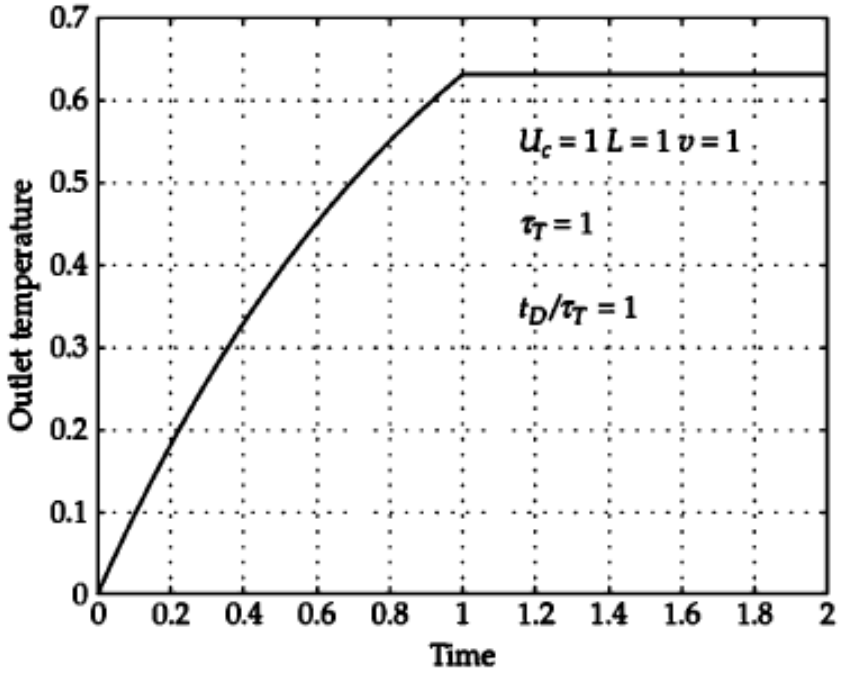
$$T(L,t) = U_c \left(1 - e^{-\frac{t}{\tau_T}}\right) - U_c e^{\frac{t_D}{\tau_T}} \left(1 - e^{-\frac{t - t_D}{\tau_T}}\right) \hat{U}(t - t_D) \tag{7-14}$$

Appendix F also explains the nature of the unit step function $\hat{U}(t)$.

7-4-1 The Large-Diameter Case

Figure 7-3 shows the behavior of $T(L,t)$ for the case where $U_c = 1$, $L = 1$, $\tau_T = 1$, $t_D = 1$, and $v = 1$. Note for $t > t_D$, the outlet temperature is constant at

$$T(L,\infty) = U_c \left(1 - e^{-\frac{t_D}{\tau_T}}\right) = 0.6321$$



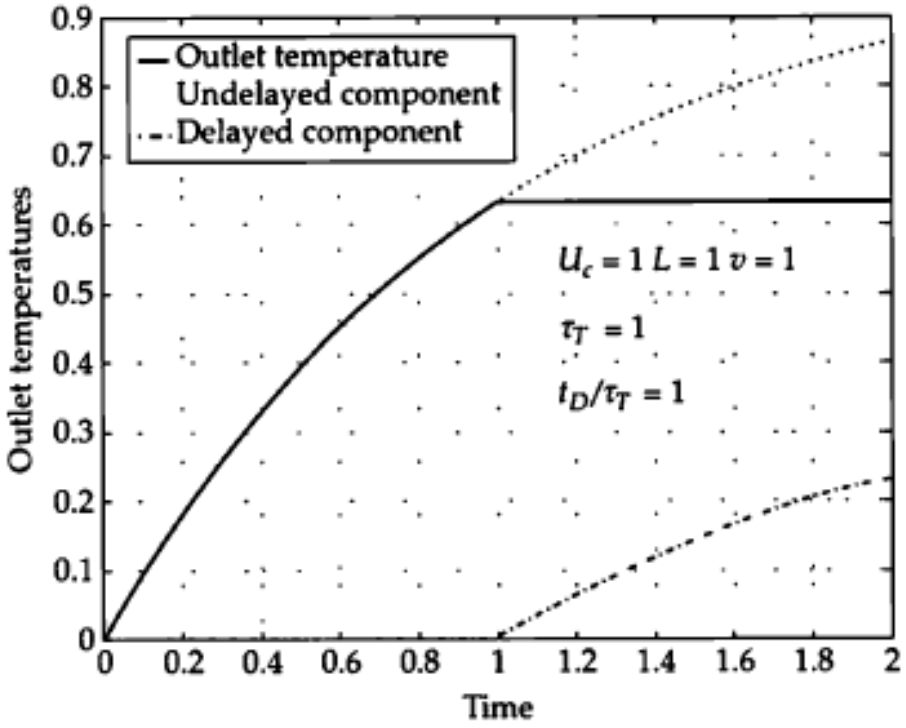


Figure 7-4 Components of the outlet temperature for large-diameter tube.

7-5 Studying the Tubular Energy Exchanger in the Frequency Domain

We wish to analyze the effect of a sinusoidal variation in the jacket temperature on the liquid outlet temperature. Start with the transfer function between the process output, which is the liquid temperature as it emerges from the tube at $z = L$, and the process input/output control which in this case is the jacket temperature, given in Eq. (7-12). Make the usual substitution of $s \rightarrow j\omega$:

$$\frac{\tilde{T}(L, j\omega)}{\tilde{T}_s(j\omega)} = \frac{1 - e^{-\frac{t_D}{\tau_T}} e^{-j\omega t_D}}{\tau_T j\omega + 1}$$

Appendix F shows that this transfer function can be reformed in terms of magnitude and phase as

$$\left|\frac{\tilde{T}(L, j\omega)}{\tilde{T}_s(j\omega)}\right| = \sqrt{\frac{1 - 2e^{-\frac{t_D}{\tau_T}} \cos^2(\omega t_D) + e^{-2\frac{t_D}{\tau_T}}}{(\tau\omega)^2 + 1}}$$

$$\theta = \tan^{-1}\left(\frac{e^{-\frac{t_D}{\tau_T}} \sin(\omega t_D)}{1 - e^{-\frac{t_D}{\tau_T}} \cos(\omega t_D)}\right) - \tan^{-1}(\tau_T j\omega) \qquad (7\text{-}15)$$

Figure 7-7 shows a Bode plot for this process model for the large-diameter tube exchanger where $L = 1$, $\tau_T = 1$, and $v = 1$.

First, note that the magnitude and phase curves start to decrease near the corner frequency which is

$$\omega_{cor} = \frac{1}{\tau_T} = 1$$

$$f_{cor} = \frac{\omega_{cor}}{2\pi} = 0.159 \text{ Hz}$$

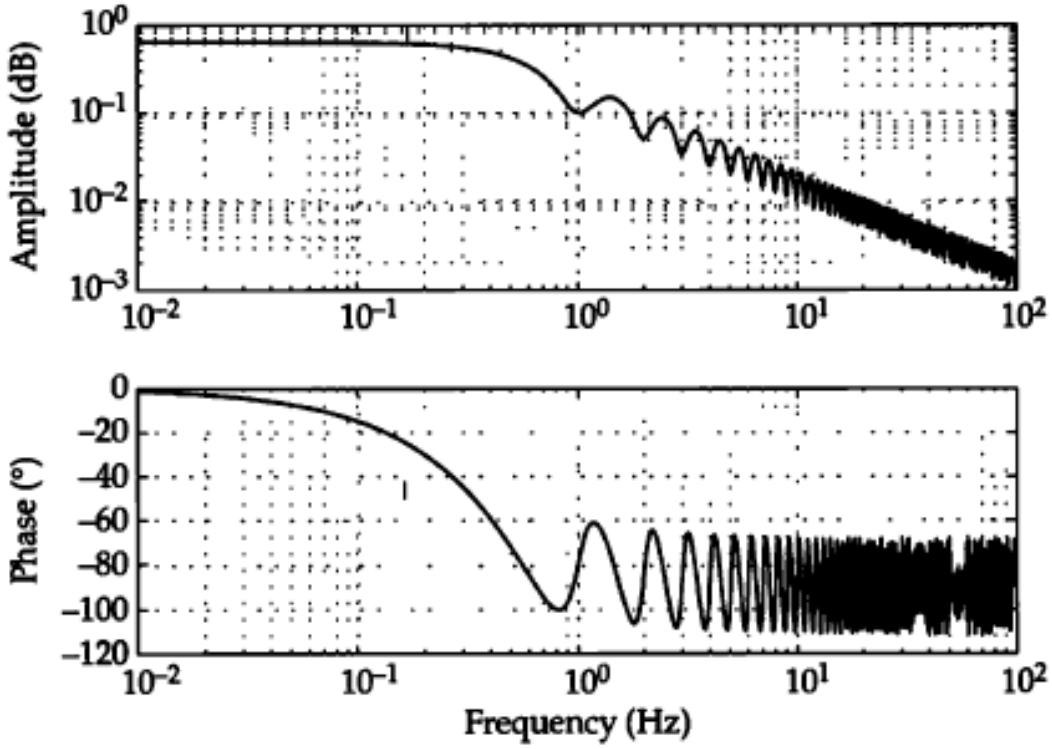
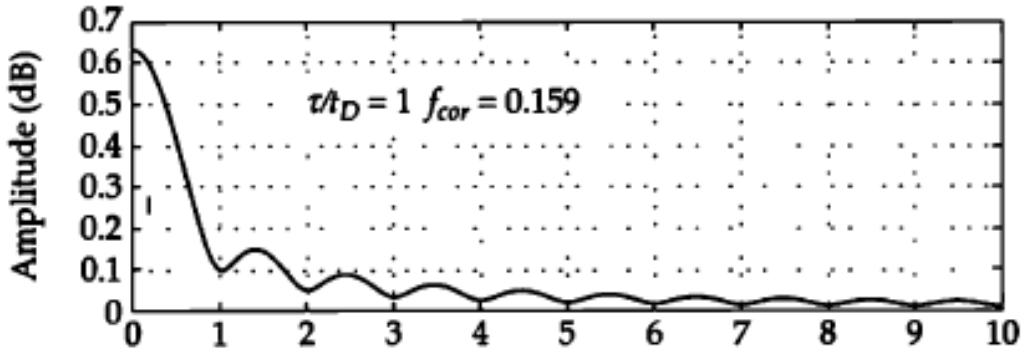


Figure 7-7 Bode plot for large-diameter tube exchanger.

Second, note the "resonances" or "ripples." The large-diameter tube exchanger Bode plot is replotted in Fig. 7-8. There appears to be a ripple that has peaks at multiples of 1 Hz. This makes sense because the residence time t_D is 1.0 sec.



7-6 Control of the Tubular Energy Exchanger

The open loop step-change response and the Bode plot suggest that there should not be too much trouble if feedback control is attempted. The large-diameter tubular energy exchanger has some idiosyncrasies but the total phase lag varies about 90° (why?—because, despite the distributed nature of the process, it is still basically first order).

This section starts by applying PI control to the large-diameter tubular exchanger where $L = 1$, $v = 1$, $\tau_T = 1$. Since the process gain is nominally unity, an initial proportional gain of unity was tried. This was increased to 3.0 by trial and error simulation using a Simulink model. Then the integral gain was increased slowly until a value of 2.0 was found to be satisfactory. Figure 7-11 shows the response to a unit step in the set point at time zero.

Implicit in this control scheme is the presence of a slave loop that will manipulate a valve so as to affect the steam jacket temperature set point which is the control output of the master loop that we are

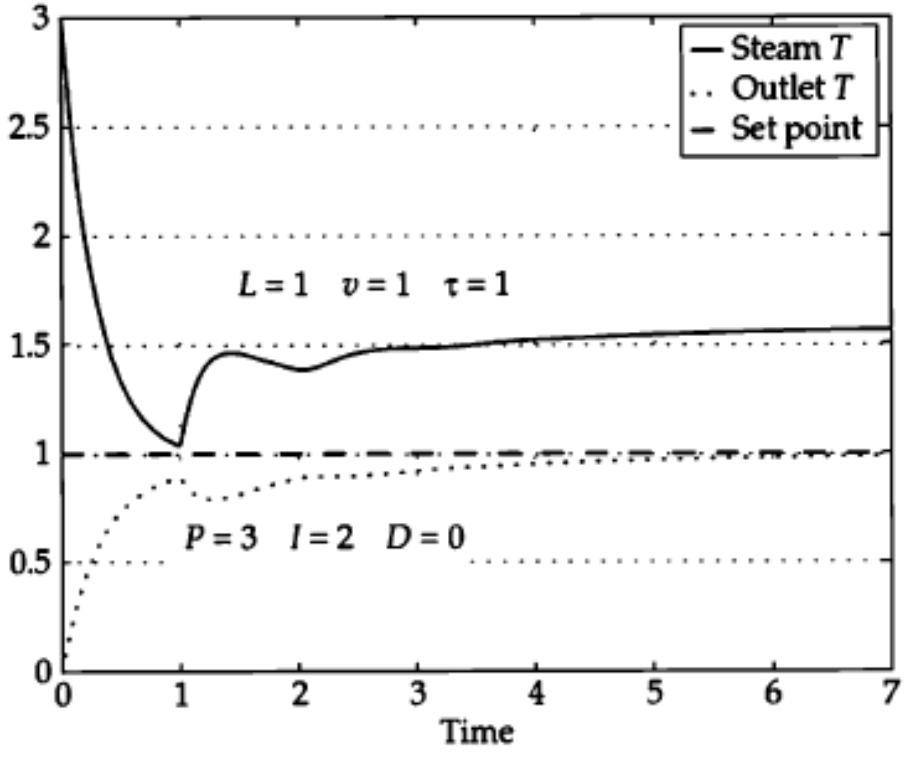


Figure 7-11 PI control of the large-diameter tubular heat exchanger.

7-7 Lumping the Tubular Energy Exchanger

7-7-1 Modeling an Individual Lump

Often, process analysts like to approximate distributed models, described by partial differential equations, with lumped models, described by ordinary differential equations. The tubular exchanger could be approximated in this way. For example, consider Fig. 7-14 where the tubular exchanger is to be modeled by N tanks. The N-tank model has the following characteristics:

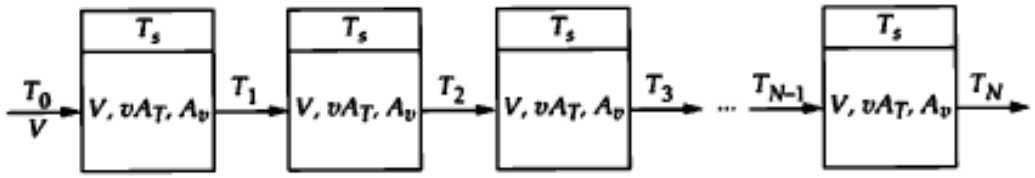


Figure 7-14 N-lump approximation to tubular exchanger.

1. Each tank is completely mixed in the sense that the exit temperature is the same as the temperature throughout the lump or tank. These lumps are often called *continuous stirred tanks* (CSTs).
2. Each tank is jacketed and is exposed to the jacket temperature T_s. Although this need not be case in general, each tank sees the same jacket temperature. That is, the jacket temperature does not vary from tank to tank.
3. The temperature leaving the kth tank is the inlet temperature for the k+1th tank.
4. Parameterwise, each tank is identical. This is not necessary but it does make the mathematics more manageable.

7-7-2 Steady-State Solution

First, in steady state we would like the N-tank model to look something like the steady-state solution obtained in Sec. 7-1 which was

$$T(z) = T_0 e^{-\frac{z}{\Psi}} + \left(1 - e^{-\frac{z}{\Psi}}\right) T_s$$

The N-tank steady-state solution is quite simple since the derivative in Eq. (7-17) is zero, so

$$T_k = g_s T_s + g_i T_{k-1}$$

$$T_1 = g_s T_s + g_i T_0$$

$$T_2 = g_s T_s + g_i T_1 = g_s T_s + g_i (g_s T_s + g_i T_0) = g_s T_s (1 + g_i) + g_i^2 T_0$$

$$T_3 = g_s T_s + g_i T_2 = g_s T_s + g_i (g_s T_s + g_i T_1) = g_s T_s (1 + g_i + g_i^2) + g_i^3 T_0$$

...

$$T_N = g_s T_s (1 + g_i + g_i^2 + \cdots + g_i^{N-1}) + g_i^N T_0 \tag{7-19}$$

When a quantity like g_i is less than unity, the sum of the geometric series, contained in the parentheses in Eq. (7-19), can be written compactly as

$$T_N = g_s T_s \frac{1 - g_i^N}{1 - g_i} + g_i^N T_0 \tag{7-20}$$

Remember that

1. g_s and g_i depend on the area for energy transfer A_T of each tank.
2. $g_s + g_i = 1$.

3. N increases A_T, the energy transfer area for the kth tank decreases so as to maintain the total energy transfer area constant.

Therefore, Eq. (7-20) becomes

$$T_N = T_0 g_i^N + \left(1 - g_i^N\right) T_s \tag{7-21}$$

This is to be compared with Eq. (7-3)

$$T(L) = T_0 e^{-\frac{L}{\Psi}} + \left(1 - e^{-\frac{L}{\Psi}}\right) T_s$$

This suggests that if the tube length L is divided up so that each tank has length $L/N = \Delta z$, then $g_i^N = e^{-L/\Psi}$ or $g_i = e^{-L/N\Psi} = e^{-\Delta z/\Psi}$. Thus, $N \to \infty$, $g_s \to 0$, and $g_i \to 1$. This exercise suggests that the lumping approach is approximately similar to the continuous approach, at least in steady state.

7-7-3 Discretizing the Partial Differential Equation

An alternative approach to lumping returns to the partial differential equation in Eq. (7-6) and replaces the partial derivative with respect to axial distance z with a finite difference, as in

$$\frac{\partial T}{\partial t} + v\frac{\partial T}{\partial z} = \frac{1}{\tau_T}(T_s - T)$$

$$\tau_T = \frac{D\rho C_p}{4U} = \frac{\Psi}{v}$$

$$T \Rightarrow T_k \qquad \frac{\partial T}{\partial z} \Rightarrow \frac{T_k - T_{k-1}}{\Delta z} \tag{7-22}$$

$$\frac{dT_k}{dt} + v\frac{T_k - T_{k-1}}{\Delta z} = \frac{1}{\tau_T}(T_s - T_k)$$

$$\frac{dT_k}{dt} = -\left(\frac{v}{\Delta z} + \frac{1}{\tau_T}\right)T_k + \frac{v}{\Delta z}T_{k-1} + \frac{1}{\tau_T}T_s$$

If the reader makes the following substitutions

$$A_p = \frac{\pi D^2}{4} \qquad A_T = \pi D \Delta z \qquad V = \frac{\pi D^2}{4}\Delta z \qquad \tau_T = \frac{D\rho C_p}{4U}$$

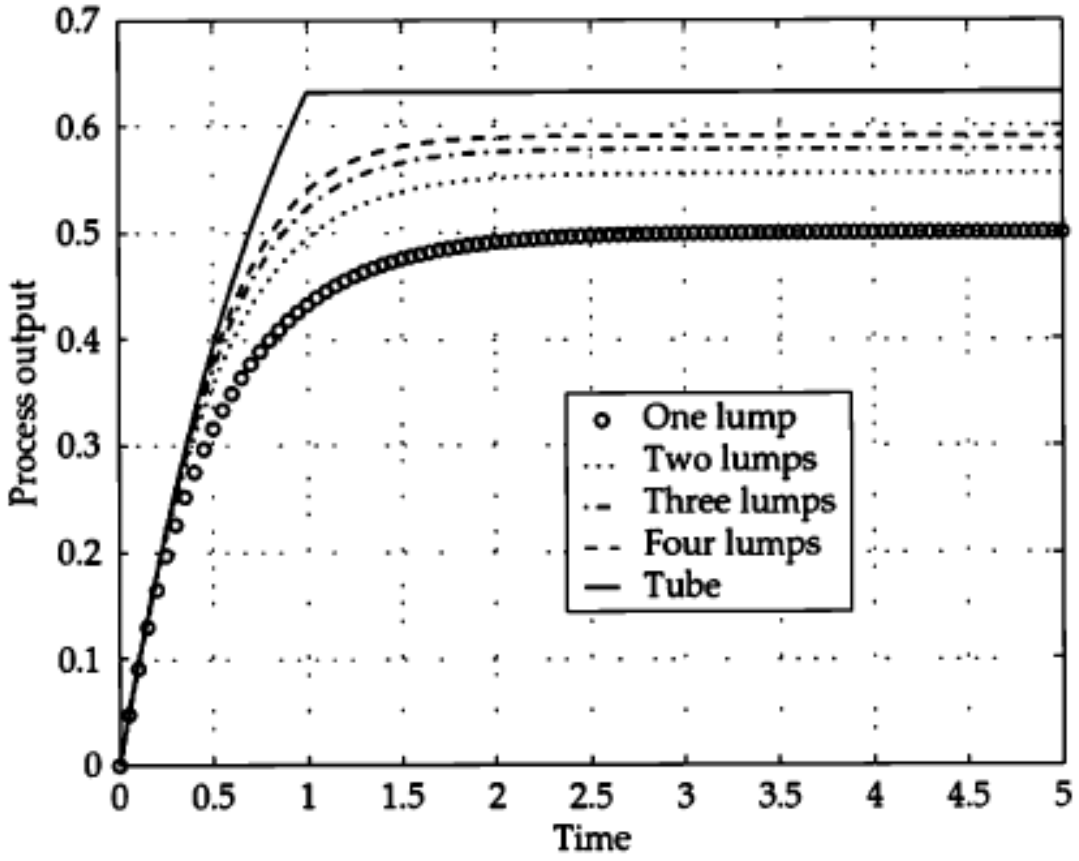


Figure 7-15 Response to steam temperature jacket T_s—matching large tubular reactor with lumped models, $\tau_t = 1.0$.

CHAPTER 8

Stochastic Process Disturbances and the Discrete Time Domain

Developing a successful control algorithm often requires proper identification of the disturbances. In Chap. 1, unautocorrelated process disturbances (white noise) and autocorrelated process disturbances were presented using the "large hotel water tank" example. In this chapter these terms and concepts will be revisited with a little more rigor using the autocorrelation, the line spectrum, the cumulative line spectrum, and the expectation operator. The ability of a PI controller to deal with different kinds of disturbances will be discussed.

Chapter 9 will revisit the discrete time domain and introduce the Z-transform.

8-1 The Discrete Time Domain

In the previous chapters, for the most part, the time domain was considered as continuous. Differential equations were derived based on this concept. Laplace transforms were used to solve these differential equations and also to provide a path to the frequency domain which was also considered continuous. In this chapter the time domain will be discrete in the sense that a data stream will now consist of a sequence of numbers usually sampled at a constant interval of time. For example, a data stream might consist of samples of a temperature $T(t)$, as is

$$T(t_1), T(t_2), \ldots, T(t_N)$$

or

$$T_1, T_2, \ldots, T_N$$

with the sample-instants in time being equally spaced, in the sense that

$$t_i = t_{i-1} + h \qquad i = 1,2, \ldots$$

where h is the sampling interval, which will be assumed to be constant unless otherwise stated. The sampling frequency is $1/h$.

Instead of differential equations where the independent variable is continuous time, there will be algebraic equations with the independent variable being an index, such as i, to an instant of time. A simple example of an indexed equation would be a running sum of a data stream consisting of sampled values of the variable x, as in

$$S_i = S_{i-1} + x_i \qquad i = 2, 3, \ldots, N$$

The average of the $x_1, x_2, \ldots$ data after N samples would be

$$\bar{x}_N = \frac{1}{N} S_N$$

The sample average $\bar{x}_N$ is an estimator of the population mean μ, which we will discuss in more detail later in this chapter.

8-2 White Noise and Sample Estimates of Population Measures

Consider a data stream of infinite extent

$$w_1, w_2, \ldots$$

from which N contiguous samples have been taken. Figure 8-1 shows two views of the data stream. The infinite data stream represents a population having certain *population characteristics* and the subset of size N mentioned above is a sample of that population. The subset has certain *sample characteristics*, which can be used as estimates of the population characteristics. The data shown in Fig. 8-1 will soon be shown to be samples of "white noise." For now, we simply refer to it as a stochastic sequence. The word "stochastic," means "nondeterministic" in that the value at time t_i does not completely determine the value at time t_{i+1}. In the white noise stochastic sequence shown in Fig. 8-1, the value at t_i has no influence whatsoever on the value at t_{i+1}. In other non–white stochastic sequences to be covered later in the chapter, the value at t_{i+1} still is nondeterministic but the value at t_i does have an influence. Note that the two streams shown in Fig. 8-1 have different sample standard deviations.

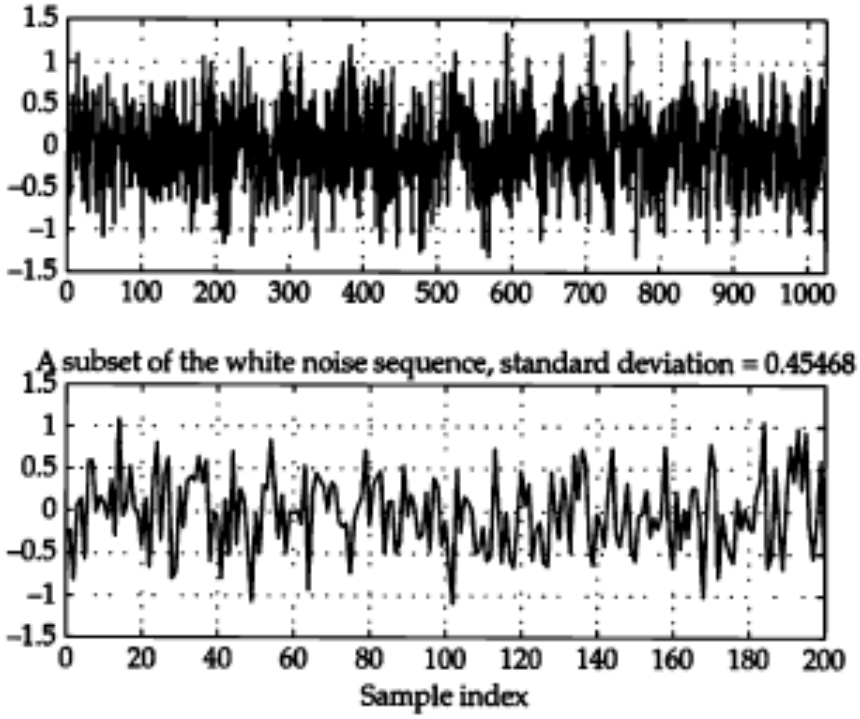


Figure 8-1 Two views of a white noise sequence with $\sigma = 0.515$.

8-2-1 The Sample Average

The sample average of the finite subset of the infinite stream of data is

$$\bar{w} = \frac{\sum_{k=1}^{N} w_k}{N} = \frac{w_1 + w_2 + \cdots + w_N}{N} \tag{8-1}$$

The average of many stochastic sequences used in this book will be removed from the data stream. If that were the case here, the subset would have a zero average. Equation (8-1) applies to any data stream, white noise, or otherwise. As mentioned in Sec. 8-1, the sample average is an estimate of the population mean μ which we will discuss later on in this chapter.

8-2-2 The Sample Variance

A measure of the strength of the variation of $w_1, w_2, \ldots$ about its average (which may be zero if the average has been removed) is the sample variance V_w, defined as

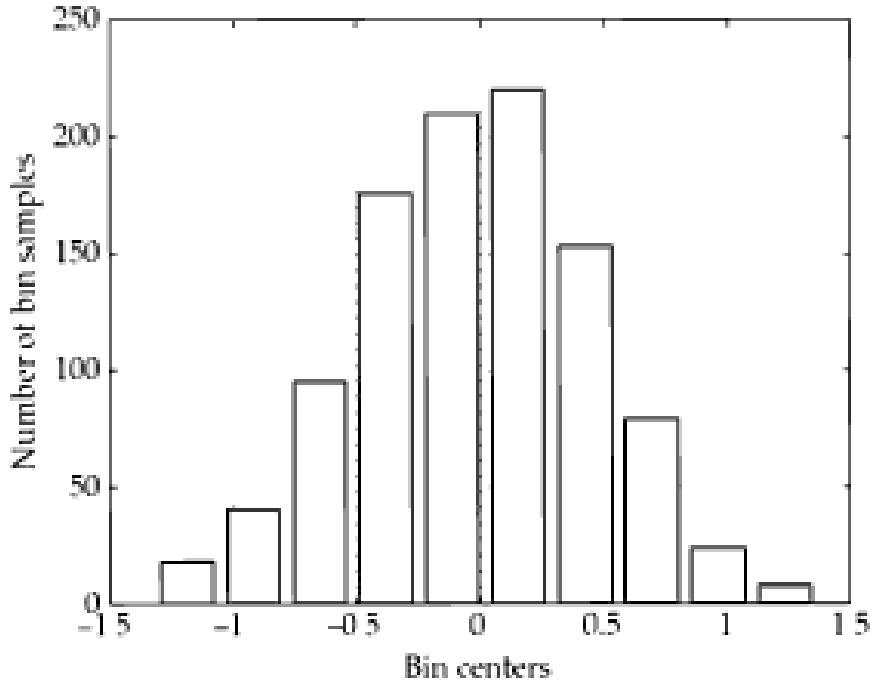
$$V_w = \frac{1}{N}\sum_{k=1}^{N}(w_k - \bar{w})^2 \tag{8-2}$$

8-2-3 The Histogram

The values in the data stream shown in Fig. 8-1 seem to cluster about the average of approximately zero. A picture of how these values are distributed is given by the histogram in Fig. 8-2 where the range over which the data stream varies is divided into 10 "bins" or cells and the number occurring in each bin is plotted versus the center of each bin. The histogram augments and extends the sample variance to give the analyst a feel for how the elements of the data stream vary about the average. In effect, the histogram is a sample estimate of the population's probability distribution In this case, the population probability distribution is normal or Gaussian and is given by

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

We will return to the probability distribution later on in this chapter. Figure 8-3 shows the shape of two normal probability distributions. Each has zero mean but one has a standard deviation of 1.0 while that for other is 1.5. Note how these curves qualitatively match that of the



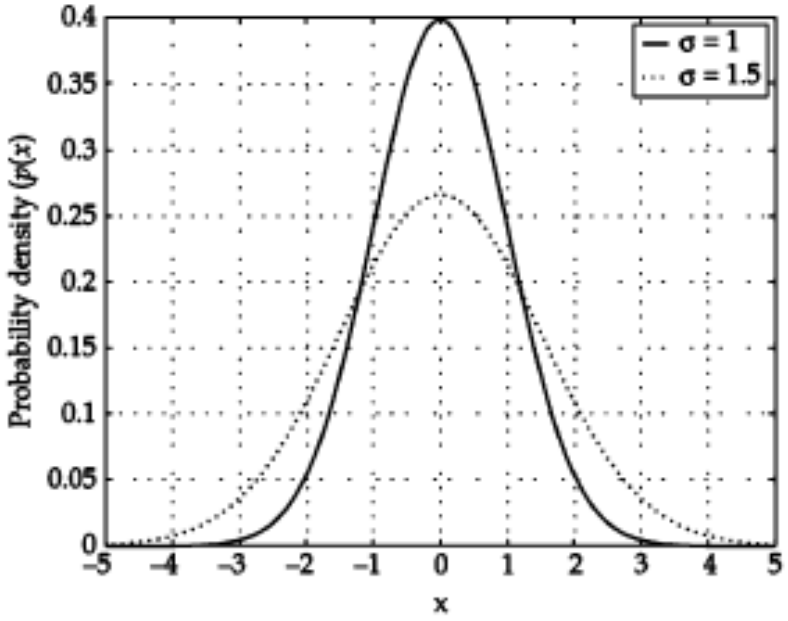


Figure 8-3 Two normal or Gaussian probability distributions.

histogram in Fig. 8-2. The histogram gives no insight into how the elements are interrelated in time. To gain some insight in that area we need other tools such as the autocorrelation and the line spectrum.

8-2-4 The Sample Autocorrelation

The white noise data stream in Fig. 8-1 is *unautocorrelated* because each sample, w_i, is independent of each and every other sample. That is, w_i is independent of w_k for every $k \neq i$. This condition could also be described using a lag index n, where samples w_i and w_{i+n} are considered uncorrelated in the sense that an average of the products of w_i and w_{i+n} taken over a set of N samples would be so close to zero as to be insignificant.

The sample estimate of the autocorrelation, $r_w(n)$, which uses the lag index n as a parameter, is one way to characterize this condition and is defined as follows:

$$r_w(n) = \frac{1}{(N-n)V_w} \sum_{k=1}^{N-n} (w_k - \bar{w})(w_{k+n} - \bar{w}) \tag{8-3}$$

The sample estimate of the autocorrelation is basically an average of the product of the lagged products over the available data set for all possible lags. As the lag size increases, the size of the data set available for the calculation in Eq. (8-3) becomes smaller and the estimate becomes less reliable.

CHAPTER 9

The Discrete Time Domain and the Z-Transform

The discrete time domain is important because (1) most data collected during a process analysis consists of samples at points in time separated by a constant interval, (2) most custom control algorithms are implemented digitally, and (3) concepts like white noise and the delta function for pulses are physically realizable in this domain. (Remember how the Dirac delta function in the continuous time domain had no specific shape and had to be defined in terms of an integral.) In the previous chapter stochastic processes defined in the discrete time domain were introduced. Here, several familiar continuous time model equations will be discretized and the Z-transform will be introduced. Just as the Laplace transform aided and abetted our attempts to solve problems and gain insight in the continuous time domain, the Z-transform will be used in the discrete time domain.

There are a couple of ways to introduce the Z-transform: (1) using the backshift operator in a manner similar to using the Laplace operator s to replace derivatives, or (2) deriving the Z-transform from the Laplace transform of a sampled time function. The latter approach is quite elegant and more general but I think it is best placed in App. I. Therefore, in Secs. 9.1 and 9.2 the backshift operator approach will naturally fall out of the discretization of the first-order model. With the new tool in hand, several other models, algorithms, and filters will be recast and studied in the Z-transform domain. As with the Laplace transform, there will be a transition to the frequency domain where more insight will be gained. The chapter closes with a discussion of fitting discrete time domain data to models.

This will perhaps be the longest chapter in the book, so you might want to break your reading plan into four parts. In the first part, you will learn about the Z-transform. In the second part, you will see how several unconventional control algorithms can be designed using

9-1 Discretizing the First-Order Model

This and the following three sections are busy. The first-order process model is studied for a special case where the process input is a series of steps. The describing equation will be modified slightly when time is discretized. The result will be rewritten using the backshift operator which will lead to the Z-transform. This necessitates a discussion of sampling and holding. The discretized unity-gain first-order model is then reinterpreted as a discrete time filter.

Back in Chap. 3 we presented the first-order model in the continuous time domain via the differential equation

$$\tau \frac{dy}{dt} + y = gU(t) \tag{9-1}$$

In the Laplace domain we wrote the transfer function between process input and output as

$$\frac{\tilde{Y}(s)}{\tilde{U}(s)} = G_p(s) = \frac{g}{\tau s + 1} \tag{9-2}$$

When Eq. (9-1) or (9-2) was solved for the case where the process input U is a step change we obtained

$$y(t) = y_0 e^{-\frac{t}{\tau}} + gU_c\left(1 - e^{-\frac{t}{\tau}}\right) \tag{9-3}$$

Before moving to the discrete time domain, let's apply Eq. (9-3) for one time increment of size h over which $U(t)$ is held constant at U_0.

$$y(h) = y_0 e^{-\frac{h}{\tau}} + gU_0\left(1 - e^{-\frac{h}{\tau}}\right) \tag{9-4}$$

Equation (9-4) moves information at $t = 0$, namely y_0 and U_0, to $t = h$ to produce $y(h)$. The information at $t = h$ can be moved to $t = 2h$ by reapplying Eq. (9-4) suitably modified, as in

$$y(2h) = y(h) e^{-\frac{h}{\tau}} + gU_1\left(1 - e^{-\frac{h}{\tau}}\right) \tag{9-5}$$

Note that the value of $U(t)$ over the interval $h \le t < 2h$ is held constant at U_1 which may be different from U_0. We have effectively moved to the discrete time domain by breaking the time variable up into samples spaced apart by h sec. This means that time is described by

$$t \Rightarrow t_i = hi \qquad i = 0, 1, 2, \ldots$$

The process output becomes

$$y(t) \Rightarrow y_i \qquad i = 0, 1, 2, \ldots$$

The process input becomes

$$U(t) \Rightarrow U_i \qquad i = 0, 1, 2, \ldots$$

With this in mind, Eqs. (9-4) or (9-5) can be written as

$$y_i = y_{i-1}e^{-\frac{h}{\tau}} + gU_{i-1}\left(1 - e^{-\frac{h}{\tau}}\right) \qquad i = 0, 1, 2, \ldots \tag{9-6}$$

Spend a few moments thinking about Eq. (9-6). It is the same as Eq. (9-4) except that it is applied over the time interval from t_{i-1} to t_i during which U_{i-1} is held constant. Figure 9-1 shows how a first-order process with a time constant of 5.0 sec and a gain of 1.1 responds to a

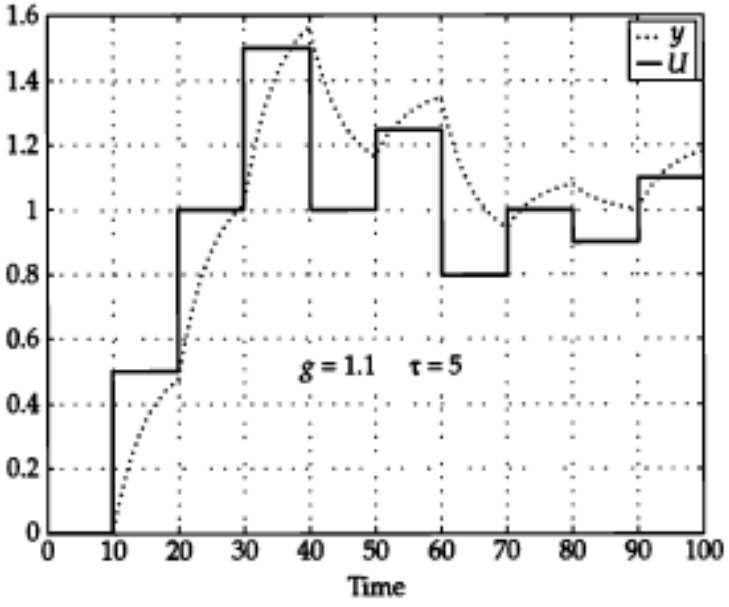


Figure 9-1 Response of first-order model to a series of steps in the process input.

9-2 Moving to the Z-Domain via the Backshift Operator

The quantity y in Eq. (9-6) can be converted to its Z-transform counterpart $\tilde{y}(z)$ by introducing the backshift operator z^{-1} defined as follows

$$y(t-h) = z^{-1}\, y(t) \tag{9-7}$$

where h is the sampling interval. I have not added the "hat" to the variable $y(t)$ because Eq. (9-7) is in a kind of limbo between the time and the z domains. It is perhaps better to proceed with the actual definition of the Z-transform which is

$$\hat{y}(z) = Z[y(t)] = \sum_{k=0}^{\infty} z^{-k} y(k) \tag{9-8}$$

Note that this is a weighted sum over all of the sampled values of $y(t)$ as compared to the Laplace transform which is a weighted integral over all the continuously variable values of $y(t)$. In both cases, values of $y(t)$ are considered zero for $t < 0$.

If the variable $y(t)$ is shifted in time one sample, Eq. (9-8) would be

$$\begin{aligned} Z[y(t-1)] &= \sum_{k=0}^{\infty} z^{-k} y(k-1) \\ &= \sum_{p=-1}^{\infty} z^{-(p+1)} y(p) \\ &= z^{-1} \sum_{p=-1}^{\infty} z^{-p} y(p) \\ &= z^{-1} \sum_{p=0}^{\infty} z^{-p} y(p) \\ &= z^{-1} Z[y(t)] \\ &= z^{-1} \hat{y}(z) \end{aligned} \tag{9-9}$$

The summation index in the third line of Eq. (9-9) is changed to zero because, like the Laplace transform, $y(t)$ is assumed to be zero for $t < 0$. The manipulations in Eq. (9-9) should convince the reader that z^{-1} is a backshift operator.
With this in mind, Eq. (9-6) becomes

$$y_i = y_{i-1}e^{-\frac{h}{\tau}} + gU_{i-1}\left(1 - e^{-\frac{h}{\tau}}\right) \qquad i = 0, 1, 2, \ldots$$

$$\bar{y}(z) = e^{-\frac{h}{\tau}}z^{-1}\bar{y}(z) + g\left(1 - e^{-\frac{h}{\tau}}\right)z^{-1}\bar{U}(z) \tag{9-10}$$

As with the Laplace transform, Eq. (9-10) is algebraic and can be solved for $\bar{y}$ as in

$$\bar{y}(z) = \frac{g\left(1 - e^{-\frac{h}{\tau}}\right)z^{-1}}{1 - e^{-\frac{h}{\tau}}z^{-1}}\bar{U}(z) \tag{9-11}$$

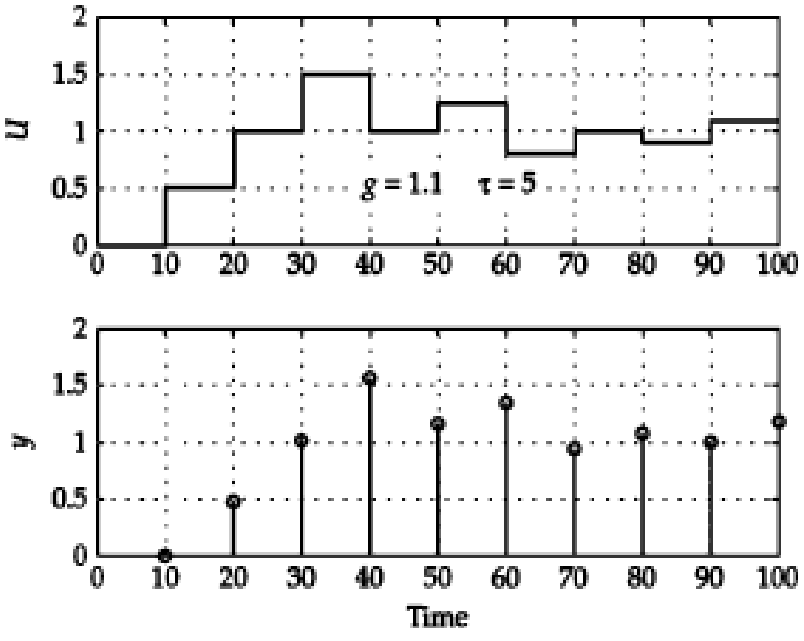
or

$$\bar{y}(z) = G(z)\bar{U}(z) \tag{9-12}$$

where $G(z)$ is the transfer function in the Z-domain for the first-order process model. For this multiplication in Eq. (9-12) to be valid, the time domain variable $U(t)$ has to behave as in Fig. 9-1.

9-3 Sampling and Zero-Holding

In Eq. (9-10), $U(t)$ is a series of steps, as if it were the output of a digital/analog (D/A) on a microprocessor. Alternatively, and more elegantly, one can say that $U(t)$ has been put through a sampler and a *zero-order hold device* which samples the value of $U(t)$ at time t_i and holds it for a period of h sec. The device releases it at time t_{i+1} at which time the device samples the new value of $U(t)$ and holds it. There is no zero-order hold device associated with $y(t)$ so it is considered as a sequence of isolated sampled values that exist only at time t_i, $i = 0, 1, 2, \ldots$. The sampled variable $y(t)$ could also be considered as a train of spikes with the height of each spike equal to the value of $y(t_i)$ as depicted in Fig. 9-2. Figure 9-3 shows how the zero-order hold device is introduced. Note the samplers that act on $U(t)$ and on $Y(t)$.



Time

Figure 9-2 Response to a series of steps in the process input, alternative view.

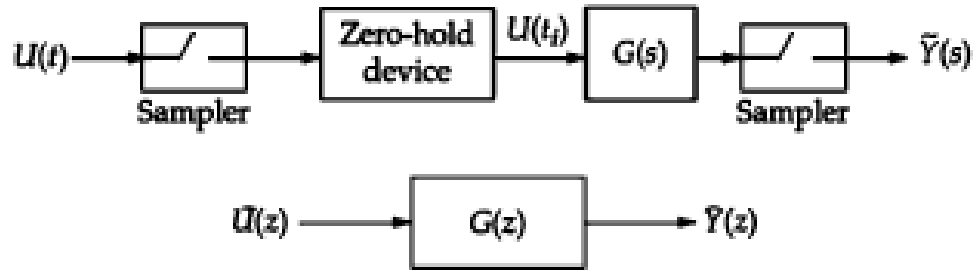


Figure 9-3 The sampler zero-hold device as part of the Z-transform transfer function.

In App. I, the zero-order hold is studied in detail and it is shown that the transfer function in Eq. (9-11) can be written as

$$\hat{Y}(z) = \frac{g\left(1 - e^{-\frac{h}{\tau}}\right)z^{-1}}{1 - e^{-\frac{h}{\tau}}z^{-1}}\hat{U}(z) = Z[\Pi_h(s)G(s)]\hat{U}(z) \tag{9-13}$$

where

$$\Pi_h(s) = \frac{1 - e^{-sh}}{s} \qquad G(s) = \frac{g}{\tau s + 1}$$

$$G(z) = \frac{g\left(1 - e^{-\frac{h}{\tau}}\right)z^{-1}}{1 - e^{-\frac{h}{\tau}}z^{-1}}$$

This transfer function can come from the discrete time domain with a stepped process input augmented by the backshift operator or from the Laplace transform domain where transfer functions in s are converted to Z-transforms in z.

If $y(t)$ turns out to be the input to another stage as in Fig. 9-4, then the Z-transform cannot be used without some consideration. Note that no sampler is applied to $y(t)$ before it becomes an input to the box represented by $H(s)$.

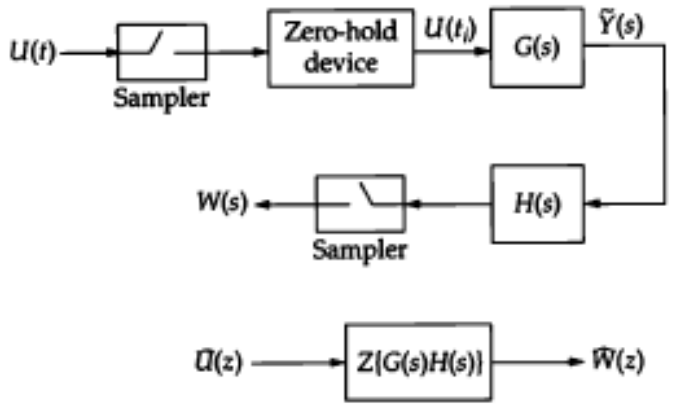


Figure 9-4 The zero-hold device as part of the Z-transform.

The correct expression to yield $\hat{W}(z)$ would be

$$\hat{W}(z) = Z\{\Pi_h(s)G(s)H(s)\}\hat{U}(z) \tag{9-14}$$

However, if a sampling device were to be placed between stages $G(s)$ and $H(s)$ then the expression for $\hat{W}(z)$ would be

$$\begin{aligned}\hat{W}(z) &= Z\{\Pi_h(s)G(s)\}\,Z\{H(s)\}Z\{\tilde{U}(s)\} \\ &= G(z)H(z)\hat{U}(z)\end{aligned} \tag{9-15}$$

To complete the concept of sampling and holding, consider the case where there is a sampler applied to the input but there is no hold as is shown in Fig. 9-5. Now the process is responding to a train of spikes and the response in Fig. 9-6 is quite different from that in Fig. 9-1. Note the first-order response to each of the U spikes (or un-zero-held samples).

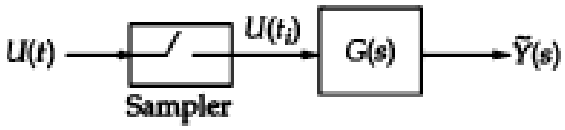


Figure 9-5 Removing the zero-hold device.

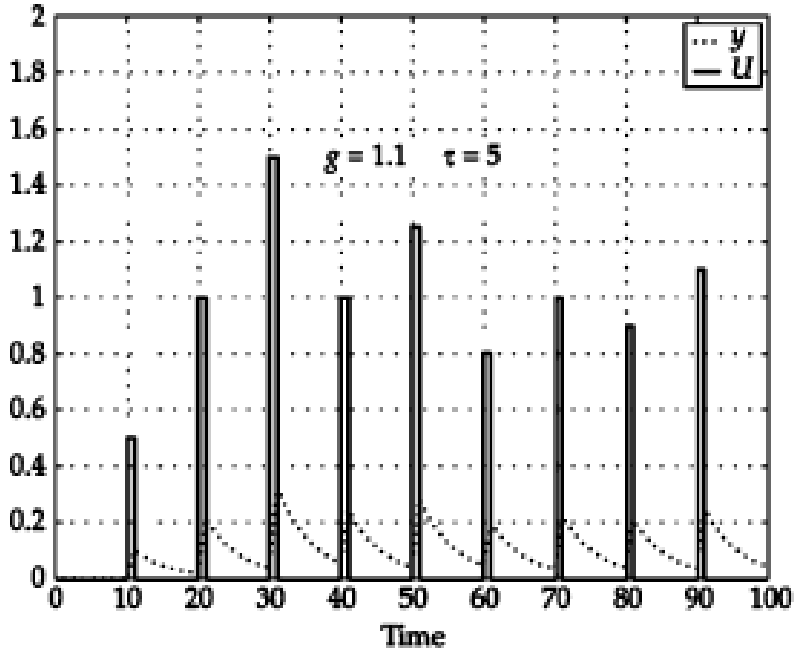


Figure 9-6 Response to a series of spikes in the process input.

9-4 Recognizing the First-Order Model as a Discrete Time Filter

If in Eq. (9-6) we set the gain to unity, then

$$y_i = y_{i-1}e^{-\frac{h}{\tau}} + U_{i-1}\left(1 - e^{-\frac{h}{\tau}}\right) \qquad i = 0, 1, 2, \ldots$$

can be written as

$$\begin{aligned} &y_i = ay_{i-1} + bU_{i-1} \qquad i = 0, 1, 2, \ldots \\ &a = e^{-\frac{h}{\tau}} \qquad \tau = -\frac{h}{\ln a} \qquad b = 1 - a \end{aligned} \tag{9-16}$$

Equation (9-16) is the widely used unity-gain discrete first-order filter with a time constant τ (also sometimes called the "exponential filter").

For the special case where the input U is white noise, one can apply the expected value operator to the square of Eq. (9-16) and obtain the variance reducing property of the discrete first-order filter, as in

$$\begin{aligned} E\{y^2\} = V_y &= E\{(ay_i + bU_{i-1})^2\} \\ &= E\{a^2y_i^2 + 2abyU_{i-1} + b^2U_{i-1}^2\} \\ &= a^2V_y + b^2V_U \end{aligned}$$

$$V_y = \frac{b^2}{1-a^2}V_U = \frac{(1-a)^2}{(1-a)(1+a)}V_U = \frac{1-a}{1+a}V_U \tag{9-17}$$

9-5 Descretizing the FOWDT Model

In Chap. 4, the FOWDT model was presented as

$$\tau\frac{dy}{dt}+y=gU(t-D) \tag{9-18}$$

In the Laplace domain, the FOWDT model transfer function was written as

$$\frac{\bar{y}(s)}{\bar{U}(s)}=G_p(s)=e^{-sD}\frac{g}{\tau s+1} \tag{9-19}$$

In Chap. 4, we rushed to the frequency domain without solving Eq. (9-18) for a constant process input. To obtain the solution in the discrete time domain, adjust the indices of U, as in

$$\begin{aligned} y_i &= y_{i-1}e^{-\frac{h}{\tau}}+g\left(1-e^{-\frac{h}{\tau}}\right)U_{i-1-n} \qquad i=0,1,2,\ldots \\ \bar{Y}(z) &= e^{-\frac{h}{\tau}}z^{-1}\bar{Y}(z)+g\left(1-e^{-\frac{h}{\tau}}\right)z^{-1-n}\bar{U}(z) \end{aligned} \tag{9-20}$$

where dead time is $D=nh$. For the time being, assume that the dead time D can be exactly divided up into n increments, each of size h.

The transfer function between y and U follows from Eq. (9-20):

$$\frac{\bar{Y}(z)}{\bar{U}(z)}=G(z)=\frac{g\left(1-e^{-\frac{h}{\tau}}\right)z^{-1-n}}{1-e^{-\frac{h}{\tau}}z^{-1}} \tag{9-21}$$

where, again, the process input is passed through a sampler and a zero-order hold.

9-10 Using the Z-Transform to Design Control Algorithms

To design a discrete time domain controller we start again with a feedback control loop but we insert a sampler and a zero-order hold as in Fig. 9-7. The block diagram algebra is similar to that for the Laplace transform except that one has to ensure that the location of the samplers and zero-order holds make sense. Here, the controller error is formed, sampled, and fed to the controller (which is probably implemented digitally) as a train of pulses. The

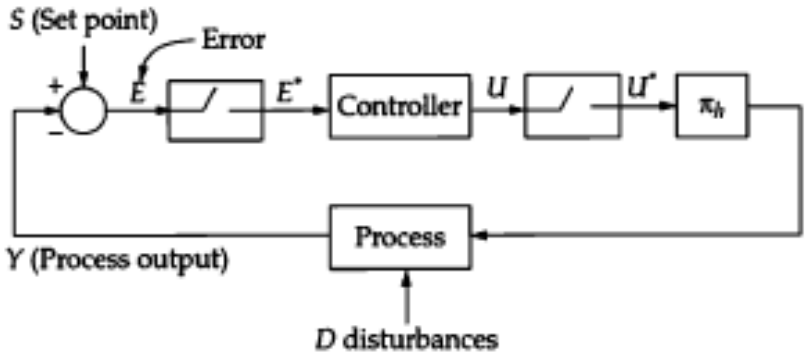


Figure 9-7 A feedback controller with a sampler and a zero-order hold.

output of the controller, U, is sampled, held and fed to the process as a sequence of steps.

$$\hat{Y} = Z\{\Pi_h(s)G_p(s)\}\hat{U}(z)$$

$$\hat{U} = G_c\hat{E}$$

$$\hat{Y} = Z\{\Pi_h(s)G_p(s)\}G_c(z)\hat{E}$$

$$\hat{E} = \hat{S} - \hat{Y} \qquad (9\text{-}41)$$

$$\hat{Y} = Z\{\Pi_h(s)G_p(s)\}G_c(z)(\hat{S} - \hat{Y})$$

$$\hat{Y} = \frac{Z\{\Pi_h(s)G_p(s)\}G_c(z)}{1 + Z\{\Pi_h(s)G_p(s)\}G_c(z)}\hat{S} = H\hat{S}$$

This look messy, but the only difference between Eq. (9-41) and Eq. (3-42) is the additional factor of the zero-order hold $\Pi_h(z)$ which is dealt with in detail in App. I.

As in Sec. 9.9, the desired relationship between the process output and the set point is specified as

This look messy, but the only difference between Eq. (9-41) and Eq. (3-42) is the additional factor of the zero-order hold $\Pi_h(z)$ which is dealt with in detail in App. I.

As in Sec. 9.9, the desired relationship between the process output and the set point is specified as

$$\frac{\tilde{Y}}{\tilde{S}} = Q(z) \tag{9-42}$$

Equations (9-41) and (9-42) are combined as before giving the expression for the controller

$$\frac{\Pi_h(z)G_p(z)G_c(z)}{1+\Pi_h(z)G_p(z)G_c(z)} = Q(z)$$

or

$$\boxed{\frac{\tilde{U}}{\tilde{E}} = G_c(z) = \frac{Q(z)}{1-Q(z)}\frac{1}{\Pi_h(z)G_p(z)}} \tag{9-43}$$

If we are dealing with first-order models, the desired response would be characterized by the discrete first-order unity gain transfer function

$$\tilde{Y}(z) = \frac{\left(1-e^{-\frac{h}{\tau_d}}\right)z^{-1}}{1-e^{-\frac{h}{\tau_d}}z^{-1}}\tilde{S}(z) = Q(z)\tilde{S}(z) \tag{9-44}$$

Continuing the first-order approach, the expression for $\Pi_h(z)G_p(z)$ would be obtained from Eq. (9-13):

$$\tilde{Y}(z) = \frac{g\left(1-e^{-\frac{h}{\tau_p}}\right)z^{-1}}{1-e^{-\frac{h}{\tau_p}}z^{-1}}\tilde{U}(z) = Z\{\Pi_h(s)G(s)\}\tilde{U}(z) \tag{9-45}$$

The only difference between Eqs. (9-44) and (9-45) is the gain and the subscript on the time constant.

The algebraic crank-turning yields

$$G_c(z) = \frac{\left(1-e^{-\frac{h}{\tau_d}}\right)\left(1-z^{-1}e^{-\frac{h}{\tau_p}}\right)}{g\left(1-e^{-\frac{h}{\tau_p}}\right)(1-z^{-1})} \tag{9-46}$$

Continuing the first-order approach, the expression for $\Pi_h(z)G_p(z)$ would be obtained from Eq. (9-13):

$$\hat{Y}(z) = \frac{g\left(1-e^{-\frac{h}{\tau_p}}\right)z^{-1}}{1-e^{-\frac{h}{\tau_p}}z^{-1}}\hat{U}(z) = Z\{\Pi_h(s)G(s)\}\hat{U}(z) \tag{9-45}$$

The only difference between Eqs. (9-44) and (9-45) is the gain and the subscript on the time constant.

The algebraic crank-turning yields

$$G_c(z) = \frac{\left(1-e^{-\frac{h}{\tau_d}}\right)\left(1-z^{-1}e^{-\frac{h}{\tau_p}}\right)}{g\left(1-e^{-\frac{h}{\tau_p}}\right)(1-z^{-1})} \tag{9-46}$$

The control algorithm is therefore

$$\hat{U}(z) = \frac{\left(1-e^{-\frac{h}{\tau_d}}\right)}{g\left(1-e^{-\frac{h}{\tau_p}}\right)}\frac{\left(1-z^{-1}e^{-\frac{h}{\tau_p}}\right)}{(1-z^{-1})}\hat{E}(z) = \frac{B_d}{gB}\frac{(1-Az^{-1})}{(1-z^{-1})}$$

where (9-47)

$$A = e^{-\frac{h}{\tau_p}} \qquad A_d = e^{-\frac{h}{\tau_d}} \qquad B = 1-A \qquad B_d = 1-A_d$$

A little (actually, a lot of) algebraic manipulation of Eq. (9-47) gives

$$\hat{U}(z) = \frac{B_d}{gB}\hat{E}(z) + \frac{B_d}{g}\frac{z^{-1}}{1-z^{-1}}\hat{E}(z) \tag{9-48}$$

The first term on the right-hand side of Eq. (9-48) is the proportional component and the second term is the integral component with an extra delay of one sample in the numerator. In practice, one would remove the extra delay because, from a common sense point of view, it adds nothing to the performance.

Therefore, converting to the discrete time domain we have a PI control algorithm:

$$\begin{aligned}\Delta U_i &= \frac{B_d}{gB}\Delta E_i + \frac{B_d}{g}E_i \\ &= P\Delta E_i + IhE_i \\ U_i &= U_{i-1} + \Delta U_i\end{aligned} \tag{9-49}$$

with the following tuning rules.

$$\boxed{P = \frac{B_d}{B} = \frac{1 - e^{-\frac{h}{\tau_d}}}{g\left(1 - e^{-\frac{h}{\tau}}\right)} \qquad I = \frac{1}{h}\frac{B_d}{g} = \frac{1 - e^{-\frac{h}{\tau_d}}}{gh}} \tag{9-50}$$

Question 9-1 If the control interval h is decreased to an infinitesimal value, will the tuning rules in Eq. (9-50) evolve into those of Eq. (9-36)?

Answer Yes, they would and I will leave it to the reader as an exercise. Sorry, unless you trust me you will have to work it out on your own.

As with the tuning rules given in Eq. (9-36), these in Eq. (9-50) are practical. When the digitally implemented PI controller has a control or sampling interval that is quite small relative to the dominant process time constant, these two sets of tuning rules are virtually identical and I would recommend using the former.

9-11 Designing a Control Algorithm for a Dead-Time Process

Before we get started, we have to understand that there really is no panacea for controlling processes with a dead time. The controllability of a process with a dead time can never be as good as that for a process without dead time, no matter how fancy the control algorithm is. Consider the following sequence of events. A disturbance causes the process output to deviate from the set point. You, acting as the controller, immediately initiate a control move to address the deviation. There will be no response to that control move until the dead time has elapsed. During that dead-time period, more nasty things can happen to the process output but you still haven't seen the effect of your initial action. This situation often leads to impatient and aggressive moves that cause more trouble than the original disturbance.

9-13-3 A Double-Pass Filter

As the reader can tell from the Bode plots, the first-order filter carries out attenuation, which is usually a desirable thing. However, it also adds phase lag to the output and this can sometimes be a problem when the data is analyzed graphically. For example, you might want to plot the raw data over the filtered data and you would probably want the filtered and unfiltered streams to be in phase. To address this problem, the filter is sometimes applied twice; once in the forward direction and once in the backward direction, as in

$$\begin{aligned} y_i &= ay_{i-1} + bU_i \qquad i = 1, 2, \ldots, N \qquad y_0 = U_0 \\ w_{i-1} &= aw_i + by_{i-1} \qquad i = N, N-1, \ldots, 1 \qquad w_N = y_N \end{aligned} \tag{9-64}$$

In the forward (left to right) direction there is a phase lag introduced but in the reverse direction (right to left) the additional phase cancels the phase lag of the first pass.

In the Z-transform domain, Eq. (9-64) becomes

$$\bar{y} = \frac{b}{1 - az^{-1}} \bar{U}$$

$$\bar{w} = \frac{z^{-1}b}{z^{-1} - a} \bar{y}$$

or, after combining

$$\begin{aligned}\tilde{w} &= \frac{bz^{-1}}{(z^{-1}-a)}\frac{b}{(1-az^{-1})}\tilde{U} \\ \frac{\tilde{w}}{\tilde{U}} &= \frac{b^2}{1-a(z^{-1}+z)+a^2}\end{aligned} \tag{9-65}$$

This filter is now a second-order filter and in the denominator, the term $z^{-1}+z$, after the substitution $z=e^{j\Omega}=\cos(\Omega)+j\sin(\Omega)$, is simply $2\cos(\Omega)$. Therefore, in the frequency domain, the filter's transfer function is real, as in

$$\frac{\tilde{w}(j\Omega)}{\tilde{U}(j\Omega)} = \frac{b^2}{1-2a\cos(\Omega)+a^2} \tag{9-66}$$

and the phase is zero—meaning no lag. So, the filtered signal is completely in phase with the input signal.

Figures 9-16 and 9-17 compare the first-order filter (which would have phase lag) and the double-pass filter. The first plot compares the frequency domain magnitudes and the second compares how they filter noisy data.

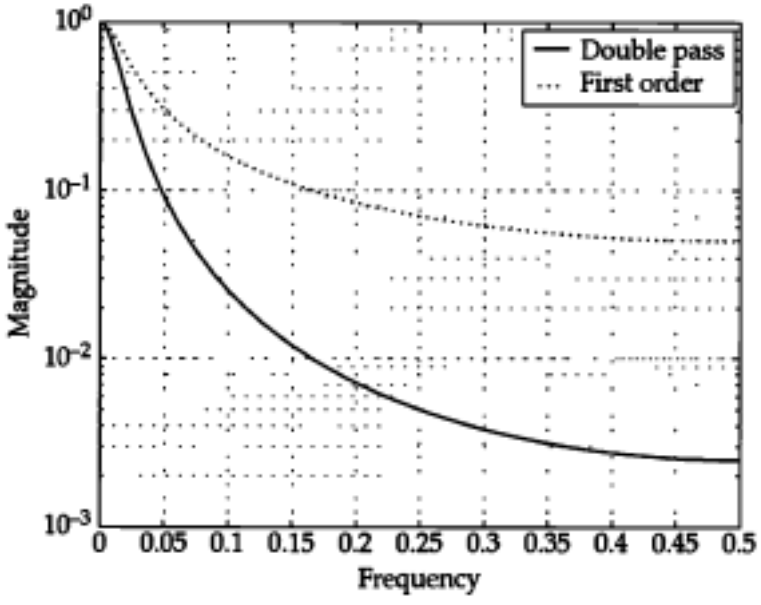


Figure 9-16 First-order versus double-pass filter, Bode plot.

9-13-4 High-Pass Filters

The differentiator filter in the Laplace domain would be simply s. Its implementation is not obvious but you can probably conceive of a way to hook up a capacitor and a resistor to do it. If not, don't worry; the digital approach is probably simpler and more direct.

In the discrete time domain, differentiation is approximated by the difference filter, the simplest of which is the backward differencer:

$$\begin{aligned} y_i &= \frac{U_i - U_{i-1}}{h} \\ \frac{\hat{y}}{\hat{U}} &= \frac{1 - z^{-1}}{h} \\ \frac{\tilde{y}(\Omega)}{\tilde{U}(\Omega)} &= \frac{1}{h}\sqrt{2 - 2\cos(\Omega)}\, e^{i\theta} \\ \theta &= \tan^{-1}\left(\frac{\sin\Omega}{1 - \cos\Omega}\right) \end{aligned} \tag{9-67}$$

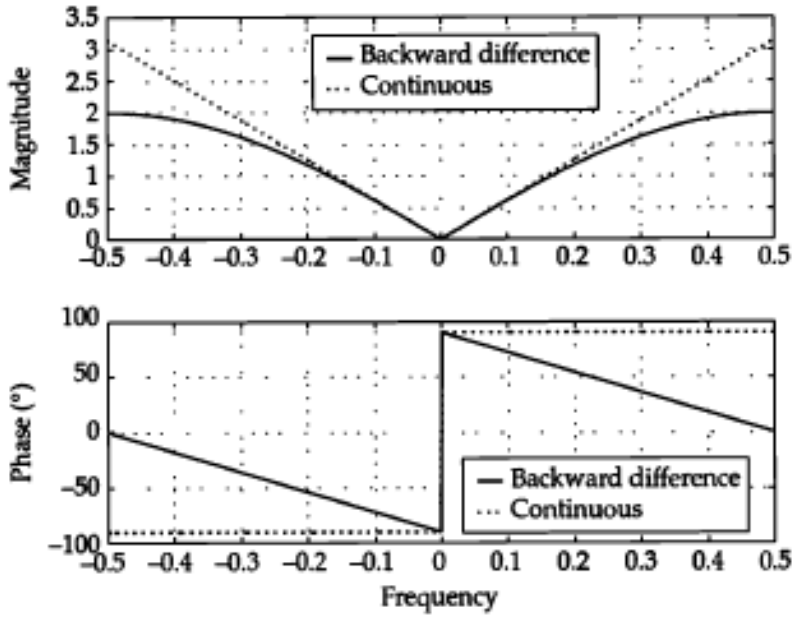


Figure 9-18 Backward difference Bode plot.

higher the frequency, a central difference is sometimes used.

$$y_i = \frac{U_i - U_{i-2}}{2h} \tag{9-68}$$

which can be considered as the average of two backward differences, as in,

$$\frac{1}{2}\left(\frac{1-z^{-1}}{h} + \frac{z^{-1}-z^{-2}}{h}\right) = \frac{1-z^{-2}}{2h}$$

The Bode plot for this difference operator is shown in Fig. 9-19. Note that both the backward and central differences have zero gain at zero frequency but the central difference has zero gain at the folding frequency, also.

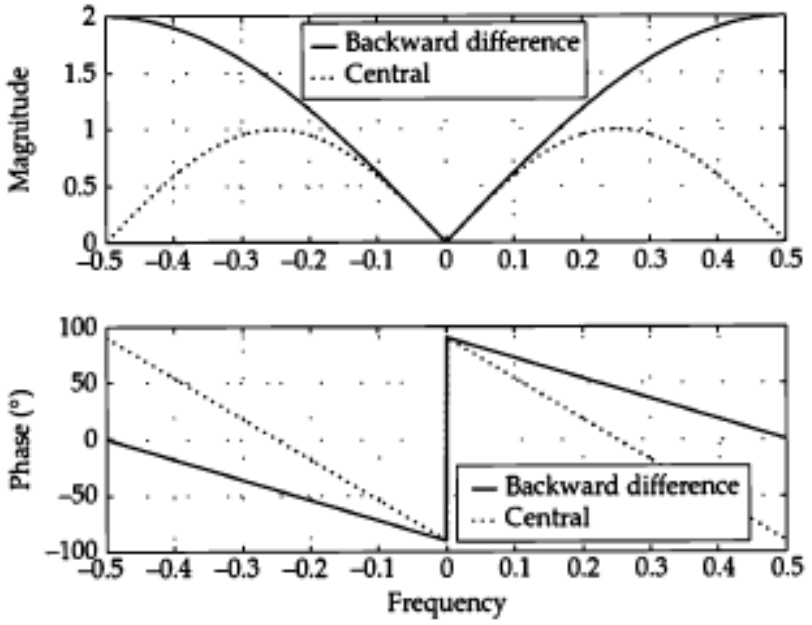


Figure 9-19 Backward and central differences, Bode plot.

9-14 Frequency Domain Filtering

There are a plethora of sophisticated computer-aided design techniques for moving average and autoregressive filters. A simple alternative uses frequency domain filtering where the Fourier transform of the data is multiplied by a magnitude factor which removes part of the spectrum. The modified transform is inverted back to the time domain yielding the filtered data. For example, should the analyst wish to remove a band of frequencies from the data, he might apply the factor shown in Fig. 9-20 to the transformed data in the frequency domain. This factor suggests that components in a signal having frequencies greater than about 0.21 Hz would be removed while those with frequencies less than 0.21 Hz would be passed unattenuated. For the readers who have read App. C, it might be worth noting that when multiplying in the frequency domain, one convolves in the time domain. However, that is a detail that is a little bit beyond the scope of this section.

Finally, note that the factor is symmetrical about the folding frequency of 0.5 Hz. Figure 9-21 shows the factor (after scaling to make it more presentable) and the spectrum of the signal to be filtered (two sinusoids, one in the pass band and one not). Figure 9-22 shows the result of applying the filter. One needs only the fast Fourier transform (and some code) to use this filtering method. The following is a crude Matlab script that carries out frequency-domain filtering.

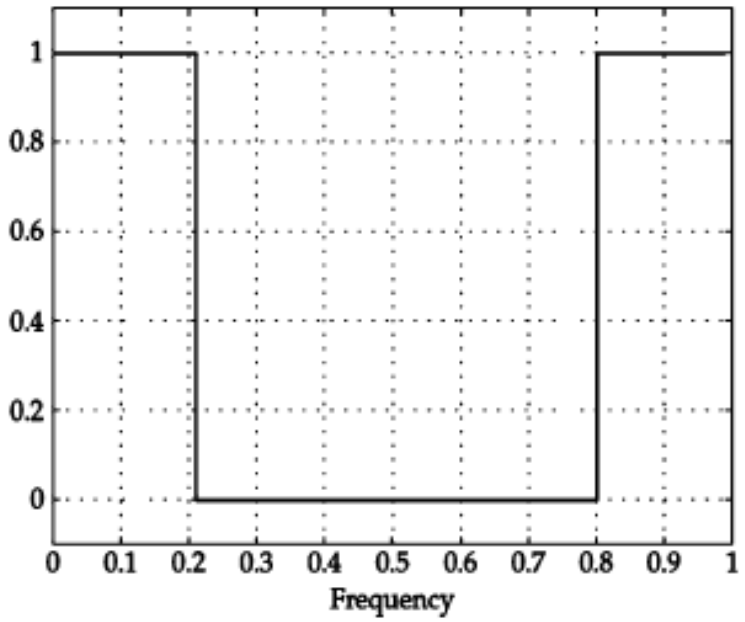


Figure 9-20 Factor to be applied to signal spectrum.

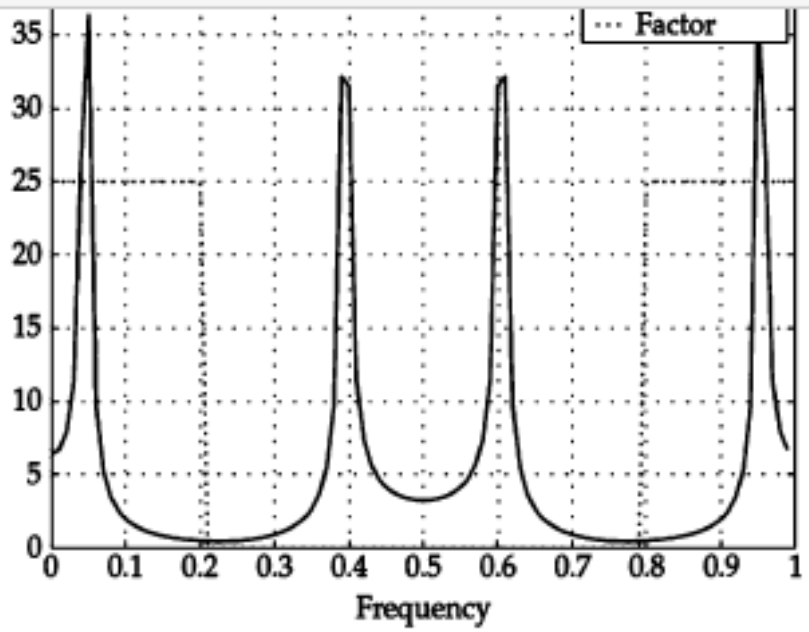


Figure 9-21 Signal spectrum (magnitude) and filtering factor (scaled by 25).

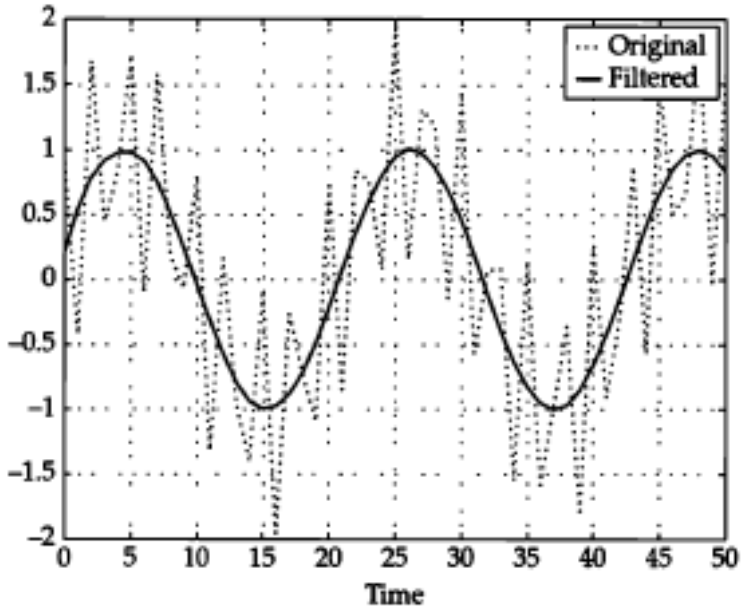


Figure 9-22 Original and filtered signal.

Estimating the State and Using It for Control

In Chap. 5, matrices and the concept of the state were introduced. In Chap. 6 an underdamped process was studied where the state consisted of the position and speed (or derivative of the position) of the mass in the mass/spring/dashpot process. We showed that regular proportional-integral (PI) control, which uses only the position, did not do a good job for this process. However, the proportional-integral-derivative (PID) algorithm which uses the position and the speed, that is, the state, performed significantly better. Another method fed back the state to create a new process that had better dynamic characteristics. It appears that an estimate of the state can play a crucial part in the successful control of a process.

This chapter will present a method that combines a model of the actual process with process measurement(s) to produce an estimate of the state. It will be applied to the control problem posed in Chap. 6. The method, called the Kalman filter, was developed in the late 1950s. To use the Kalman filter, one must find values of a vector called the Kalman gain. Two ways to find this gain will be presented. The first is based on choosing variances associated with the process model and with the process noise. The second is based on placing the eigenvalues of the system.

The Kalman filter will also be applied to the three-tank problem presented in Chap. 5. However, the variables to be controlled will be extended to include all the three tank levels and the variables to be manipulated will be extended to include all the input flow rates. The resulting multidimensional control algorithm will contain integral control and will be tuned by placing the eigenvalues of the controlled system. For the sake of comparison, the same three-tank process will be controlled by three separate PI controllers. Finally, the state-space control approach will be applied to a lumped approximation of the tubular energy exchanger process presented in Chap. 7.

10-1 An Elementary Presentation of the Kalman Filter

The Kalman filter combines the predicted value of the state from a model with suitably adjusted process measurements to provide an estimate of the state. The first component of the Kalman filter is the process model.

10-1-1 The Process Model

Consider the continuous time case where the process is described by

$$\begin{aligned} \frac{d}{dt}x &= Ax + BU \\ Z &= Hx \end{aligned} \tag{10-1}$$

where x is a $(n, 1)$ vector, A is a (n, n) matrix, B is a (n, m) matrix, U is a $(m, 1)$ vector, and H is the (p, n) "measurement" matrix. The quantity Z, a $(p, 1)$ vector, is the measured quantity. If all the elements of the state are measurable, then $p = n$ and the H matrix is square. If some of the states are not measurable, then $p < n$. In the case of the underdamped process, it might be the position that is the only part of the state available for measurement and therefore, $n = 2$, $p = 1$.

The discrete time version of Eq. (10-1) is developed in App. H as

$$\begin{aligned} x_{i+1} &= \Phi x_i + \Gamma U_i \\ Z_i &= H x_i \end{aligned} \tag{10-2}$$

This discrete time model is augmented by two sources of noise, as follows:

$$\begin{aligned} x_{i+1} &= \Phi x_i + \Gamma U_i + w_i \\ Z_i &= H x_i + v_i \end{aligned} \tag{10-3}$$

where w is sometimes called *process noise* and can represent the error between the model and the actual process. The symbol v is sometimes called *measurement noise*. Both of these stochastic processes are considered to be white, have zero mean, have a normal distribution, and have covariances (with zero lag), symbolized by matrices Q and R, respectively. The covariance matrix was introduced in Chap. 8. In the scalar case we will use σ_w^2 and σ_v^2. The covariance matrix Q is a measure of the model uncertainty and the covariance matrix R is a measure of the measurement noise.

10-1-2 The Premeasurement and Postmeasurement Equations

In many texts the derivation of the Kalman filter appears, in my humble opinion, to be one of the most convoluted exercises in control engineering theory. I will not attempt to derive it here. If the reader thinks, after the presentation in this section, that she needs to delve into the derivation for a better understanding, I recommend *Applied Optimal Estimation*, edited by Arthur Gelb. This book was first published in 1974 and is still probably one of the most readable books around. Do not attempt to read Rudolf Kalman's original paper! As an interesting alternative, one might visit the Internet and see what the Wikipedia has to say about the Kalman filter.

There are two stages in the estimation: Before the measurement and after the measurement. A quantity estimated before the measurement is taken (using the model) will have (–) appended to its symbol. Quantities estimated after the measurement is taken will have the (+) appendage.

Before a measurement is taken at the kth sample time, the model can be used to generate an estimate at time t_k, as in

$$\hat{X}_k(-) = \Phi \hat{X}_{k-1}(+) + \Gamma U_{k-1} \tag{10-4}$$

Equation (10-4) gives the premeasurement state estimate $\hat{X}_k(-)$ at time t_k based on knowledge of the process input U_{k-1} and the postmeasurement estimate of the state from time t_{k-1} which is $\hat{X}_{k-1}(+)$. Using the model to predict a value at time t_k based on information at time t_{k-1} is sometimes referred to *one-step extrapolation*. Note that the tilde symbolizes that the quantity is an estimate of the true value X_k.

The postmeasurement estimate at time t_k is calculated from

$$\hat{X}_k(+) = K_k Z_k + (I - K_k H)\hat{X}_k(-) \tag{10-5}$$

where Z_k is the measurement at time t_k and K_k is the (n, p) Kalman gain vector at time t_k. Equation (10-5) suggests that the postmeasurement is a weighted sum of the measurement Z_k and the premeasurement model-based estimate $\hat{X}_k(-)$. Equation (10-5) can also be written as

$$\hat{X}_k(+) = \hat{X}_k(-) + K_k[Z_k - H\hat{X}_k(-)] \tag{10-6}$$

which shows that the postmeasurement estimate $\hat{X}_k(+)$ is equal to the premeasurement model-based estimate $\hat{X}_k(-)$ plus a correction

10-1-3 The Scalar Case

Temporarily consider the scalar first-order case where $n = 1$, $p = 1$, and $m = 1$. In this case the state is one-dimensional and the measurement of the state is available but may be noisy. The premeasurement state estimate, via extrapolation, is

$$\hat{X}_k(-) = e^{-\frac{h}{\tau}}\hat{X}_{k-1}(+) + g\left(1 - e^{-\frac{h}{\tau}}\right)U_{k-1} \tag{10-7}$$

and the postmeasurement correction is

$$\hat{X}_k(+) = \hat{X}_k(-) + K_k[Z_k - \hat{X}_k(-)] \tag{10-8}$$

where all of the quantities are scalars. If the model is quite accurate, and the measurement is noisy, then $\sigma_w = 0$ and σ_v would be relatively large and you might expect that K_k would be small. Conversely, if the model is only approximate but the measurement is quite good, then $\sigma_v = 0$ and σ_w would be relatively large and you would expect K_k to be significant.

10-4 Using the Kalman Filter for Control

In Sec. 6-5 the state of the underdamped process was fed back to make the compensated system behave differently, namely, without the ripples. We then applied integral-only control to the compensated system with reasonable success. The state was constructed from the measured position and the estimated filtered derivative.

In this section, the Kalman filter will be used to estimate the two components of the state, which will be fed back just as in Sec. 6-5. In addition, the estimated position will be used in an integral-only control loop. Figure 10-6 shows a condensed version of a Matlab Simulink model of the controlled system. If you are not familiar with Simulink, treat the figure as a block diagram. Box 1 contains the compensation gain K_u, which is applied to the controller output. Box 2 contains the two compensation gains, which are applied to the state

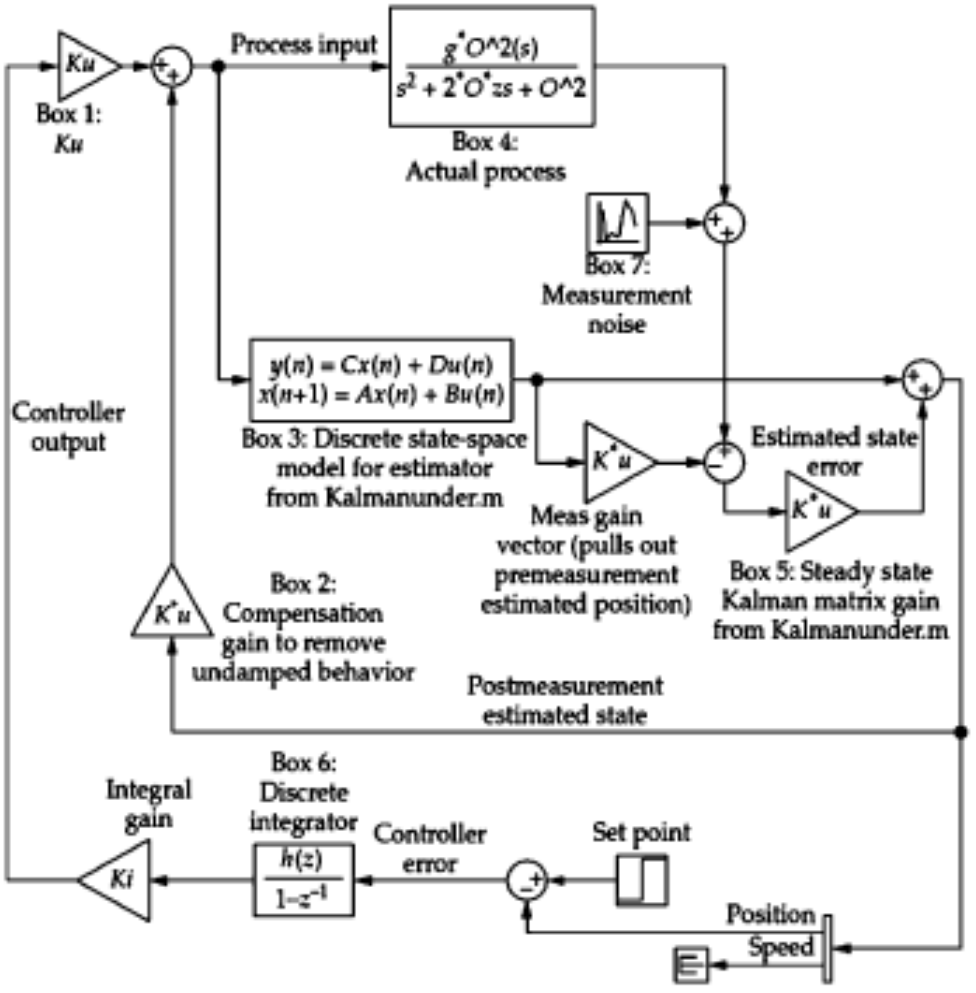


Figure 10-6 Simulink model: control using the Kalman filter.

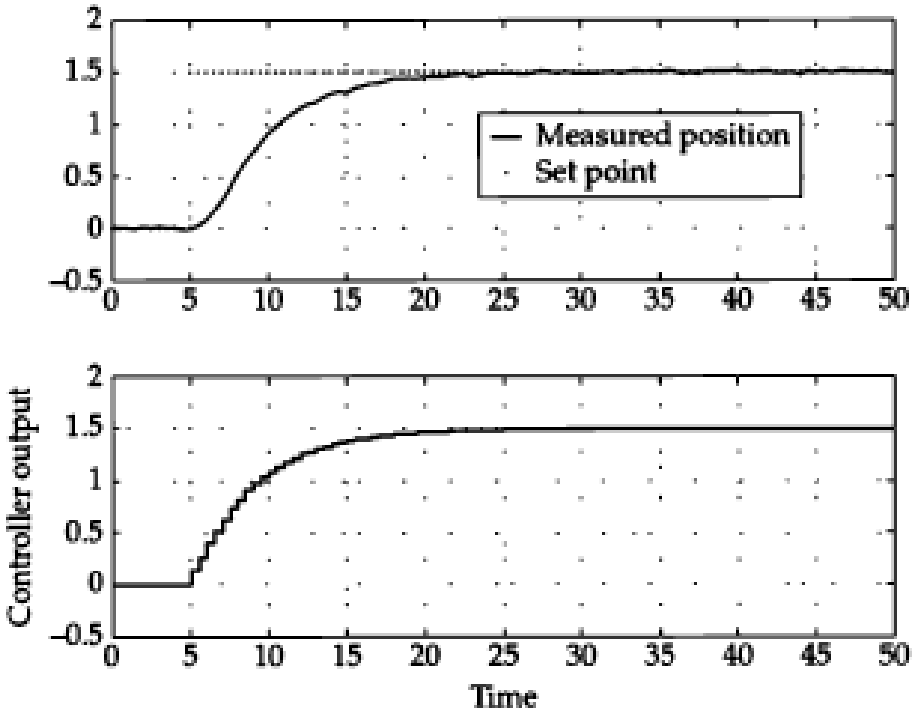


Figure 10-7 Closed-loop control of compensated underdamped process.

10-5 Feeding Back the State for Control

In previous chapters one-dimensional proportional-only control was accomplished by feeding back the process variable and subtracting it from the set point. In state-space one could apply this approach as follows.

$$\begin{aligned} X_k &= \Phi X_{k-1} + \Gamma U_{k-1} \\ U_k &= K_c(S_k - X_k) \end{aligned} \tag{10-25}$$

where K_c is a feedback gain. The state X_k can be available through measurements or through estimation via the Kalman filter.

When these two equations are combined we get

$$\begin{aligned} X_k &= \Phi X_{k-1} + \Gamma K_c(S_{k-1} - X_{k-1}) \\ &= (\Phi - \Gamma K_c)X_{k-1} + BK_cS_{k-1} \end{aligned} \tag{10-26}$$

which describes the dynamics of a closed-loop system. As with the Kalman filter equations in Sec. 10-1-2, this is an indexed equation that has a homogeneous and nonhomogeneous part. The homogeneous part, namely,

$$X_k^h = (\Phi - \Gamma K_c)X_{k-1}^h \tag{10-27}$$

has a solution of the form

$$X_k^h = C\lambda^k$$

which, when applied to Eq. (10-27), gives

$$C\lambda^k = (\Phi - \Gamma K_c)C\lambda^{k-1} \tag{10-28}$$

Equation (10-28) can be rearranged to give

$$(\Phi - \Gamma K_c - \lambda I)C = 0$$

where, for a solution to exist, λ must satisfy the eigenvalue-yielding equation of

$$\left|\Phi - \Gamma K_c - \lambda I\right| = 0 \tag{10-29}$$

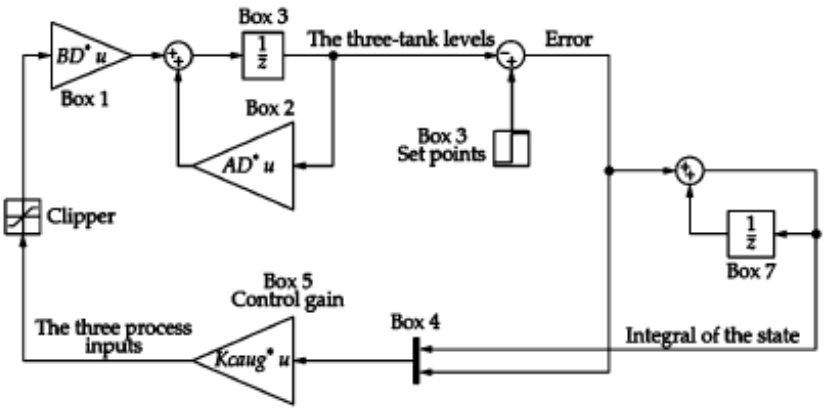


Figure 10-9 Simulink model: augmented state control of three-tank system with back flow.

11-1 The Strange Motel Shower Stall Control Problem

You are on a business trip for your company, visiting a far-flung plant in the hinterlands. You check into the local motel/hotel and decide to take a shower (Fig. 11-1).

Not being familiar with the plumbing in this motel you have to develop a strategy for adjusting the shower water temperature before getting into the shower. I suggest that it would be something like that shown in Fig. 11-2.

Let's try to quantify the algorithm outlined in Fig. 11-2. The stick figure (you) is standing outside the stall and sampling the shower head spray. Once you have turned the valve you might carry out the following steps:

1. Sample the water temperature at time t_i, $i = 1$, with your finger (the start of "digital" control). You will not have a numerical value but we will still denote the temperature by $T(t_i)$.
2. Adjust the valve to an amount that is proportional to the perceived error $E(t_i)$

$$E(t_i) = S - T(t_i)$$

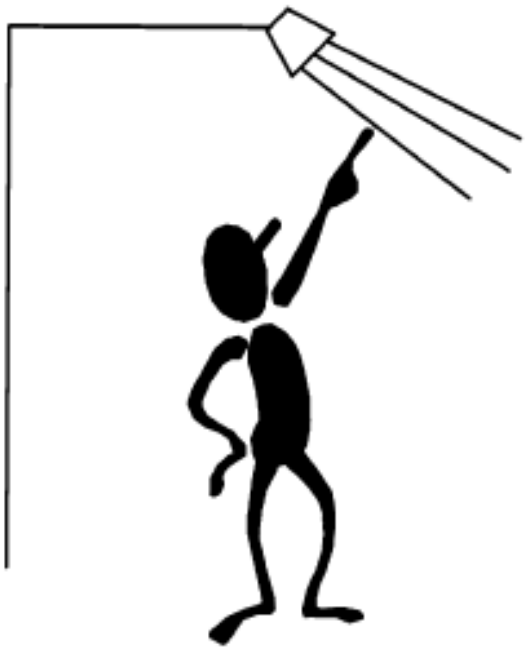


Figure 11-1 The Strange Motel shower stall control algorithm.

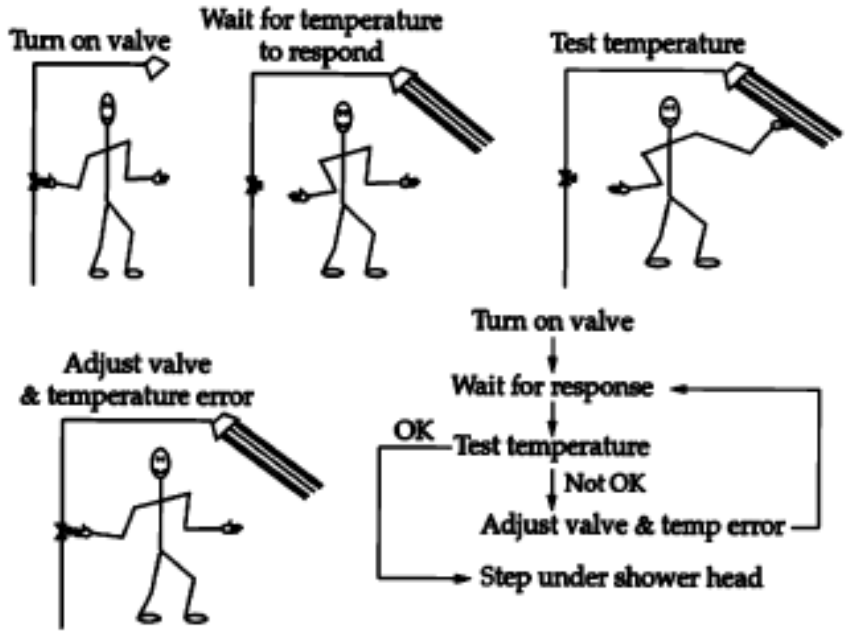


Figure 11-2 The Strange Motel shower stall control strategy.

As with the temperature, you will not have a numerical value for the error but you probably will have a feeling for the deviation as to sign and approximate amount. The adjustment will be

$$\Delta U(t_i) = K E(t_i) = K[S - T(t_i)]$$

$$U(t_i) = U(t_{i-1}) + \Delta U(t_i)$$

where K is a proportionality constant that is a measure of your patience and aggressiveness and ΔU is the change in the valve position. Note that the second equation simply says that you added the increment (positive or negative) to the previous valve position.

3. Wait a period of time h for the water temperature to respond to your adjustment. This wait time will probably include any dead time and at least one time constant. The time is now $t_i = t_{i-1} + h$, $i = 2$. Note that you have implicitly incremented the time index i.
4. Sample the water at time t_i with your fingers and go to step 2.

You would continue this loop until the error is perceived to be

with the transport of the water through the piping, followed by a first-order-like response. The wait time h is long enough for the expiration of the dead time and 99% of the time constant response.

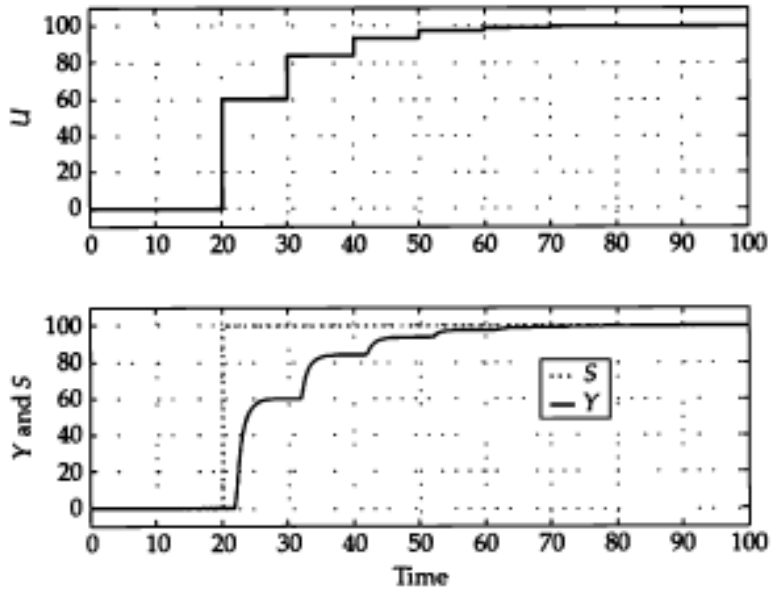


Figure 11-3 Conservative Strange Motel shower stall control.

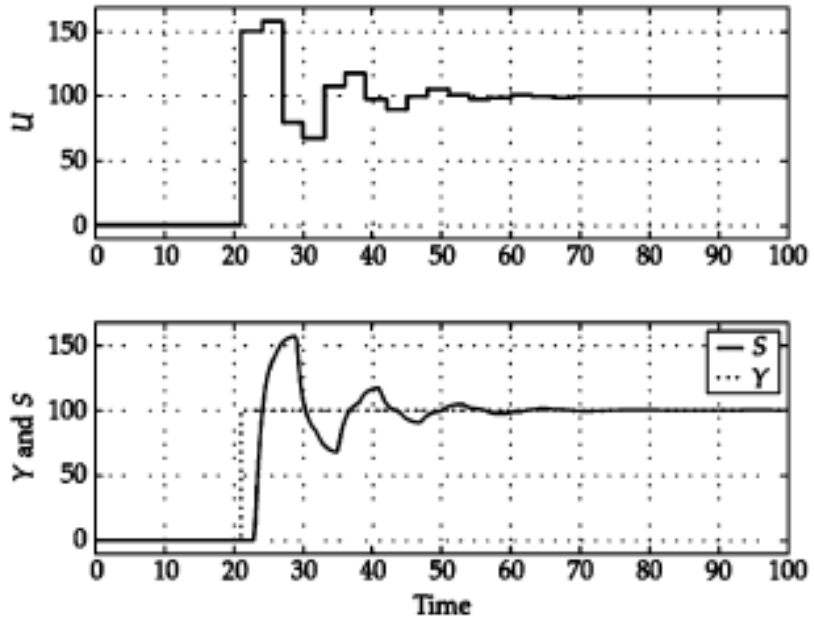


Figure 11-4 Aggressive Strange Motel shower stall control.

If you used the same strategy but had less patience and felt more

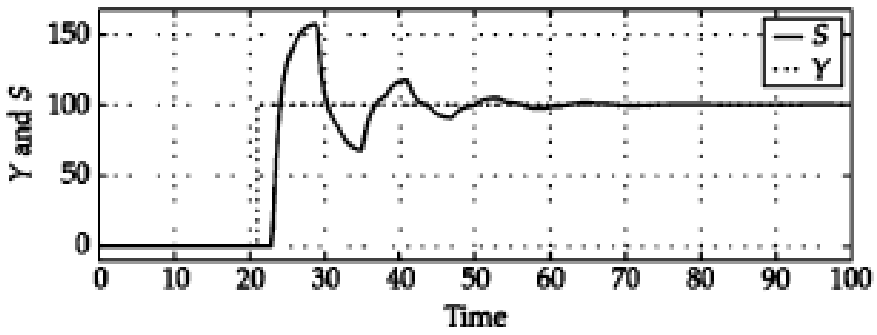


Figure 11-4 Aggressive Strange Motel shower stall control.

If you used the same strategy but had less patience and felt more aggressive, the results might be like those shown in Fig. 11-4.

In this case, you did not wait for the full response of the temperature and your adjustment sizes were greater for the same perceived size of the error. As a result, there was overshoot, although the desired temperature may have been arrived at earlier than with the more conservative strategy of Fig. 11-3.

The control strategy fits the closed-loop structure that we have been using in the rest of the book as shown in Fig. 11-5.

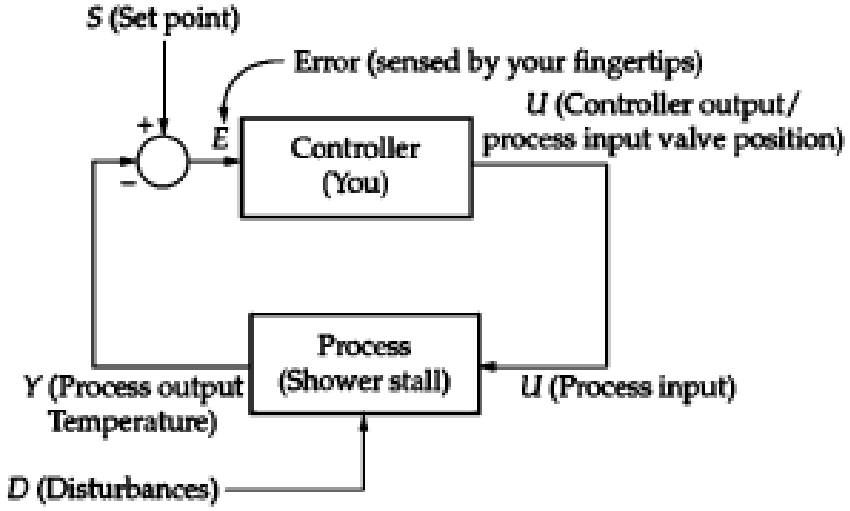


Figure 11-5 The Strange Motel shower stall control strategy, block diagram.

11-3-2 Proportional-Only Control

Figure 11-8 shows the effect of removing the integral control for the same conditions as those in Fig. 11-7. Here the control output jumps to 50 at $t = 3$ and stays there until the process output starts to respond at $t = 5$. During this period there is no control output movement because the error does not change. When the process responds, ΔE is negative and the control output backs off and moves around by a small amount until the error stops changing. Unfortunately, when the process output and the error stop changing, the latter is not zero. Since there is no integral component to continue to work on the constant but nonzero error, there will be an offset between the process output and the set point.

11-3-3 Proportional-Integral-Derivative Control

Adding derivative to Eq. 11-4 gives

$$\begin{aligned} \Delta U(t_i) &= I h E(t_i) + P \Delta E(t_i) + D_g \, \Delta[\Delta E(t_i)] \\ U(t_i) &= U(t_{i-1}) + \Delta U(t_i) \end{aligned} \tag{11-5}$$

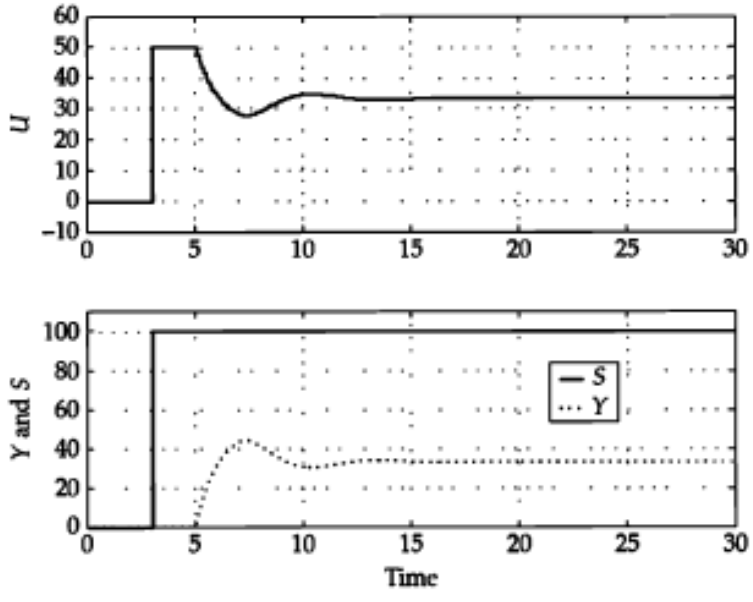


Figure 11-8 Proportional-Only control of shower stall temperature.

11-4 Cascade Control

Figure 11-13 shows the familiar water tank in a slightly different configuration. The source of the process input is a secondary tank that has an input flow rate of unknown origin. The valve is adjusted to maintain the level in the primary tank. Now, what would happen if there were a significant disturbance in the secondary tank? This disturbance would first cause the flow rate to the primary tank to vary. This flow rate variation would cause the primary tank level to deviate from set point. The control loop would then adjust the valve in an attempt to bring the level back to the set point.

The process output, namely the primary tank level, experiences a significant deviation in response to the upstream disturbance. For the controlled system to react to the disturbance, an error has (and will) show up in the primary tank process output. Figure 11-14 shows the set point being stepped at time $t = 1$. Later on, at time $t = 30$ there is a disturbance in the secondary tank and Fig. 11-14 shows the resulting disturbance in the primary tank level.

This problem can be addressed if a second flow-control loop is added, as shown in Fig. 11-15. In this case, the flow rate coming into the primary tank is controlled to a flow-rate set point generated by the level control loop Should there be a disturbance in the secondary tank, it will be sensed by the flow-rate controller and quickly corrected such that there may be little or no variation in the primary tank level.

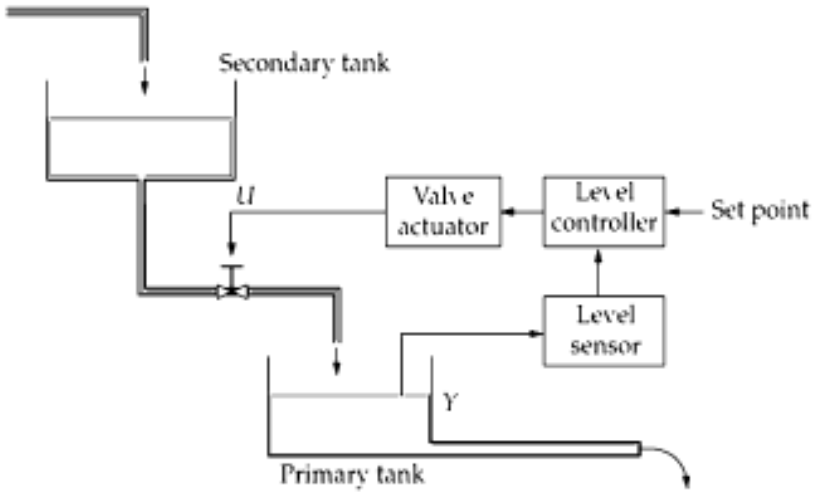


Figure 11-13 A single control loop.

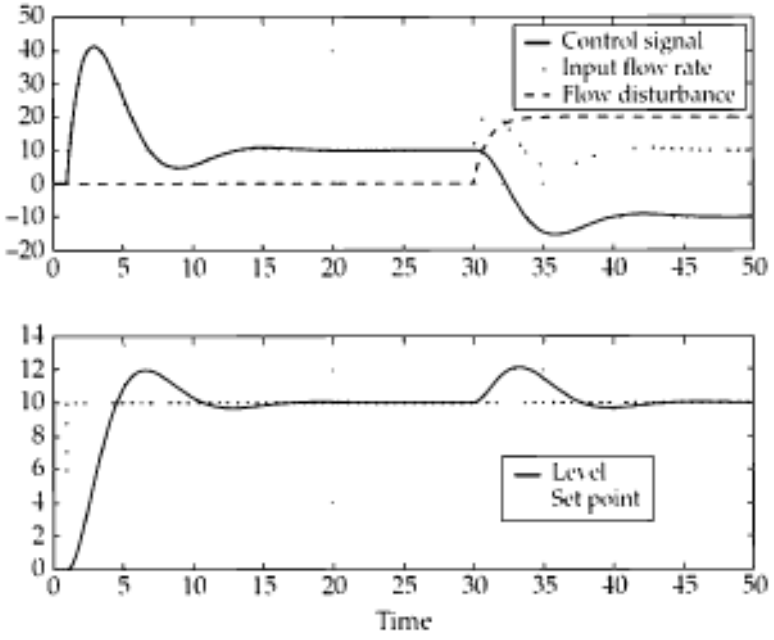


Figure 11-14 Single-loop control performance.

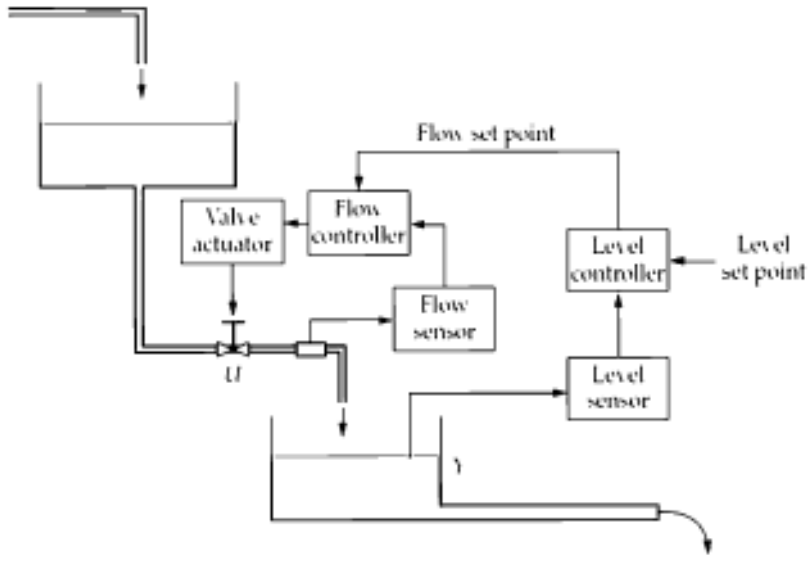


Figure 11-15 Cascade control.

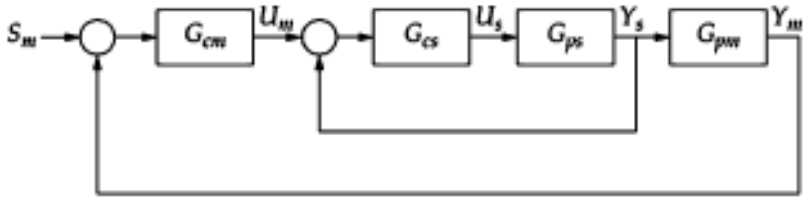


Figure 11-16 Cascade control schematic.

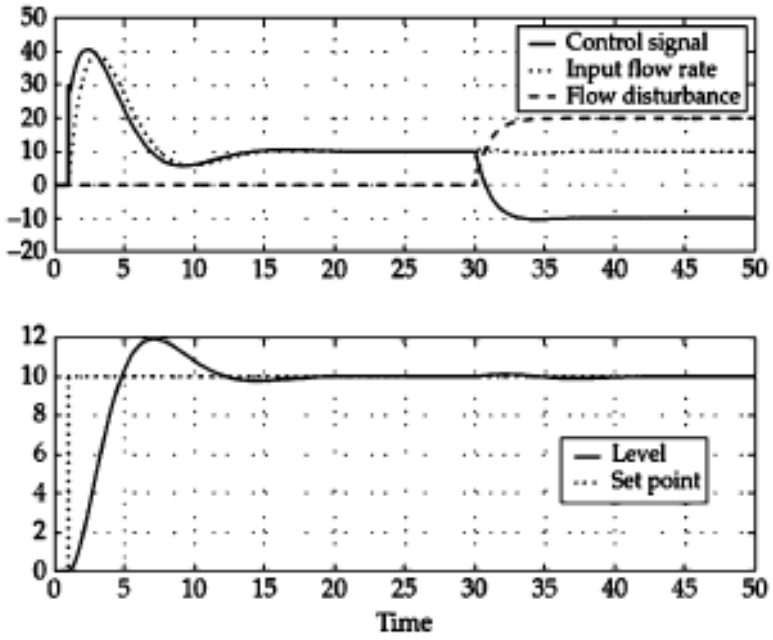


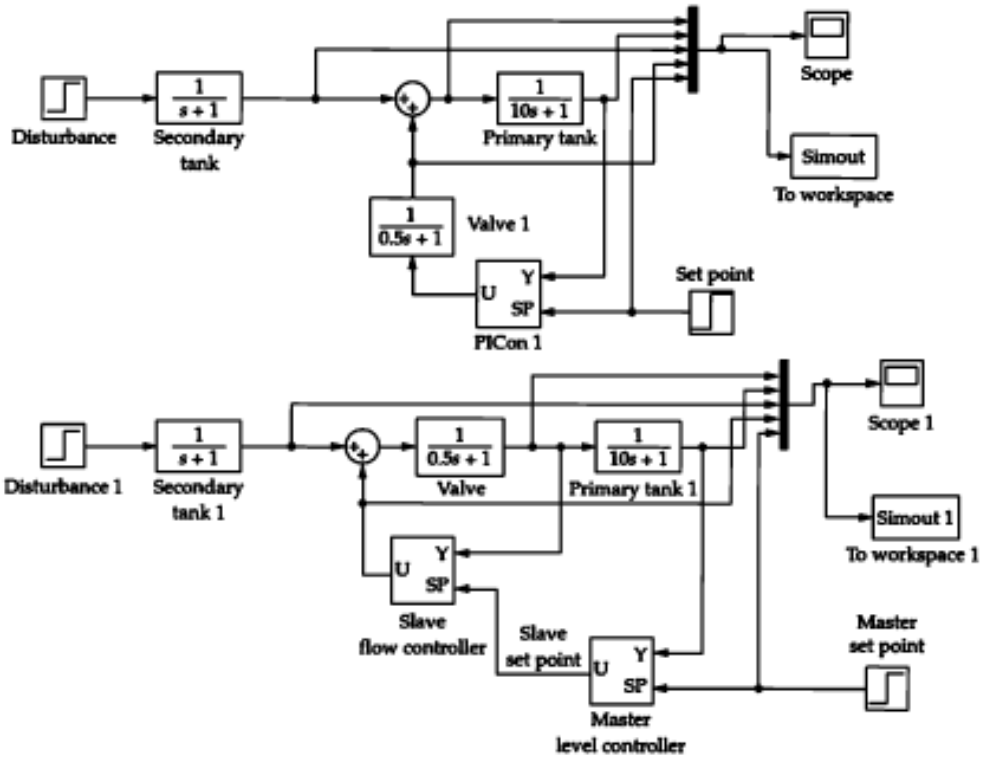
Figure 11-17 Cascade control performance.

tank is smaller than the primary with the same process gain but with a time constant of 1.0 time units. The flow-controller dynamics are even quicker with a gain of unity and a time constant of 0.5 time units. As in Fig. 11-14, there is a disturbance in the secondary tank level at time $t = 30$. Figure 11-17, when compared to Fig. 11-14, shows the improvement in performance by using cascade control.

The Matlab Simulink model used to generate the simulations in Figs. 11-14 and 11-17 is given in Fig. 11-18.

Cascade control, sometimes with several levels of embedded master/slave structure, is widely used in industry. It is especially effective where a secondary loop is much faster than a primary loop.

In Chap. 1, Sec. 1-7, cascade control appeared in an example process that tended to behave like a molten glass forehearth. The master control loop reads the glass temperatures via a thermocouple and sends a temperature set point to the combustion zone slave controller.



11-5 Control of White Noise—Conventional Feedback Control versus SPC

In the 1980s there was a great rush to a relatively old concept that was relabeled statistical process control (SPC). Although statisticians will go into cardiac arrest at this description, SPC is basically an alarm system that detects non–white noise riding on the signal of a process variable. Most SPC systems are based on the so-called WECO rules that were published by Western Electric in 1956. These rules claim that a process is "out of control" when one or more of the following conditions are satisfied:

1. One sample of the process output has deviated from the nominal value (probably a set point) by three standard deviations.
2. Two out of three samples have deviated from the nominal by two standard deviations.
3. Three out of four samples have deviated from the nominal by one standard deviation.
4. Eight samples in succession have occurred above or below the median line.

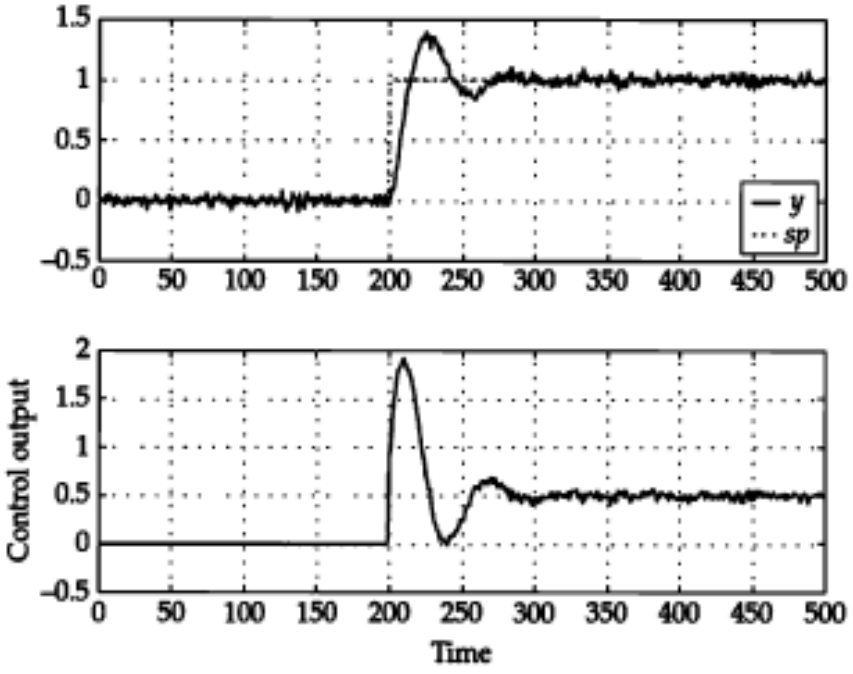


Figure 11-19 Control of a long-time constant process subject to white noise.

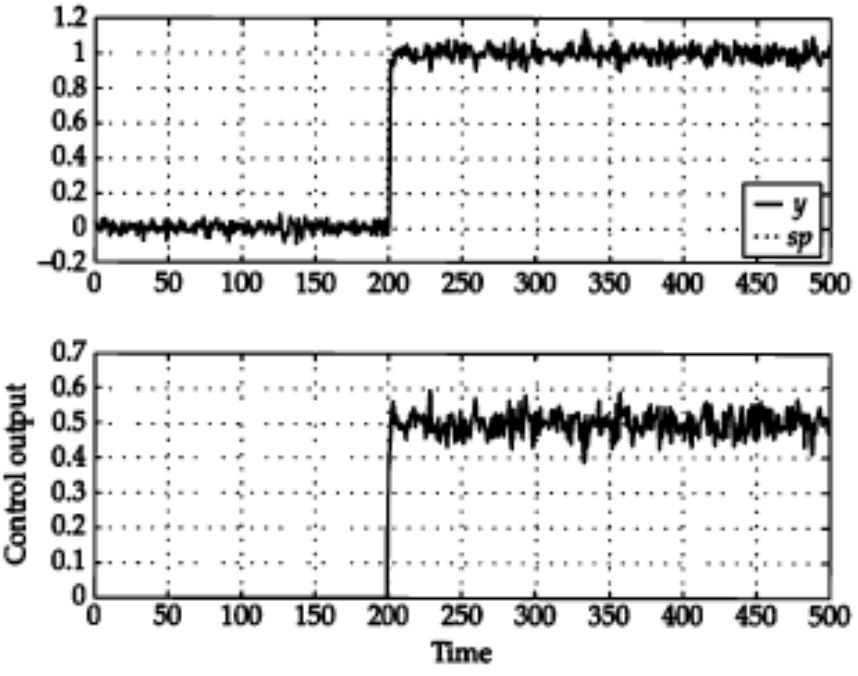


Figure 11-20 Control of a short-time constant process subject to white noise.

11-6 Control Choices

We have stopped the deluge of different control algorithms that you or your control engineer can choose from. This does not imply that there are not more—there definitely are—however, I think we have covered the "big picture" of control algorithms.

The proportional-only control algorithm was presented first in Chap. 3 and then again in this chapter. For industrial situations it would probably not be your first choice. However, it occurs in many places. For example, your automobile engine coolant flow is regulated by a thermostatic valve. When the engine is cold, the thermostat closes the valve to restrict coolant flow and allow the engine to quickly reach a satisfactory operating temperature. As the engine heats up, the thermostat opens the valve and allows more coolant to circulate. The movement of the valve is proportional to the temperature of the coolant and there really is no set point as such. There also is no history of engine temperatures available to the thermostat so there is no integral effect that might be able to slowly work the temperature back to the desired value.

The proportional-integral control algorithm is the workhorse of the process control industry. In my opinion, it should be the first choice. Before some more sophisticated approach is taken it should be conclusively shown why PI is not acceptable.

The PI tuning rules were presented in Chap. 9

$$P = \frac{\tau_p}{g\tau_d} \qquad I = \frac{1}{g\tau_d}$$

or

$$K_c = \frac{\tau_p}{g\tau_d} \qquad \tau_I = \tau_d$$

11-7 Analysis and Design Tool Choices

We started with the simple first-order process model and used an ordinary differential equation in the continuous time domain to describe its behavior. As the models became more involved, the Laplace transform was used to move from the continuous time domain to the *s*-domain where differential equations became algebraic equations and life was often simpler. Laplace transforms were used to generate transfer functions which in turn could be used in a block diagram algebra that opened up many new methods of design and analysis. The dynamics of process models were shown to be characterized by the location of poles in the *s*-plane.

A simple substitution allowed us to move from the Laplace *s*-domain to the frequency domain where we could use concepts like phase lag, phase margin, and gain margin to develop insight into dealing with dynamics, both open loop and closed loop, often without having to solve differential or algebraic equations.

Matrices were shown to be a compact method of dealing with higher dimensional problems. The state-space approach brought us back to the time domain but presented us with an enlarged kit of tools. Eigenvalues of certain matrices were shown to be equivalent to the poles of transfer functions.

The movement from the continuous time domain to the discrete time domain was facilitated by the Z-transform where another simple substitution allowed us to move to the frequency domain to develop more insight. The state-space approach was represented in this new domain.

Finally, the Kalman filter was introduced and shown to provide a means of estimating the state from a noisy measurement if a process model was available. Several control approaches using the Kalman filter and the state-space concept were presented.