2

Process Models

2.1 Introduction

Mathematical models are commonly used to describe the behavior of processes. Models give a unified way to treat systems of widely different types, and they make it possible to introduce a number of useful concepts. The models are also essential for simulation and control design. In this chapter we will review some of the models that are commonly used for PID control. The models try to capture some aspects of the process that are relevant for control. Many different types of models are used.

The steady-state behavior of a process can be captured by a function that tells the steady-state value of the process variable for given values of the manipulated variable. Such models are discussed in Section 2.2.

To control a system it is necessary to describe process dynamics. For the purpose of control it is often sufficient to describe small deviations from an equilibrium. In this case the behavior can be modeled as a linear dynamical system. This is a very rich field with many useful concepts and tools, which form the core of control theory. Different ways to describe process dynamics are discussed in Section 2.3. The ideas of transient response and frequency response are introduced as well as the important concepts of step response, impulse responses, and transfer functions.

Special techniques for modeling process dynamics have traditionally been used in PID control. The idea is to characterize process dynamics by a few features. This is discussed in Section 2.4, where features such as average residence time, apparent time delay, apparent time constant, normalized delay, ultimate gain, ultimate frequency, and gain ratio are introduced.

In Section 2.5 we introduce some particular models that are widely used for PID control. These models are introduced in terms of their transfer functions. The important concepts of normalization are also introduced in that section, as well as nonlinearities. The examples introduced in Section 2.5 will be used extensively in the book.

Disturbances are an important aspect of a control problem. In Section 2.6 we describe some models that are used to describe disturbances. Section 2.7 describes simple methods for obtaining the models, and Section 2.8 describes

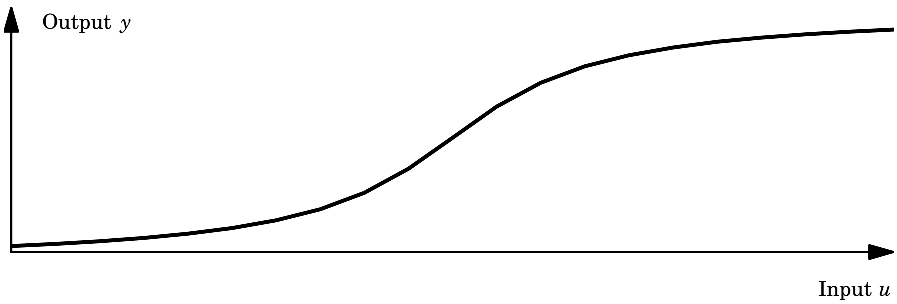


Figure 2.1 Static process characteristic, which shows process output y as a function of process input u under steady-state conditions.

some techniques used to simplify a complicated model. The chapter is summarized in Section 2.9, and references are given in Section 2.10.

2.2 Static Models

It is natural to start by describing the stationary behavior of the process. This can be done by a curve that shows the steady-state value of the process variable y (the output) for different values of the manipulated variable u (the input); see Figure 2.1. This curve is called a static model or a static process characteristic. All process investigations should start with a determination of the static process model. It can be used to determine the range of control signals required to change the process output over the desired range, to size actuators, and to select sensor resolution. The slope of the curve in Figure 2.1 tells how much the process variable changes for small changes in the manipulated variable. This slope is called the static gain of the process. Large variations in the gain indicate that the control problem may be difficult.

The static model can be obtained experimentally in several ways. A natural way is to keep the input at a constant value and measure the steady-state output. This gives one point on the process characteristics. The experiment is repeated to cover the full range of inputs.

An alternative procedure is to make a closed-loop experiment where the output of the system is kept constant by feedback and the steady-state value of the input is measured.

The experiments required to determine the static process model often give a good intuitive feel for how easy it is to control the process, and if there are many disturbances. Data for steady-state models can also be obtained from on-line measurements.

Sometimes process operations do not permit the experiments to be done as described above. Small perturbations are normally permitted, but it may not be possible to move the process over the full operating range. In such a case the experiment must be done over a long period of time. It is possible to provide a control system with facilities to automatically determine the static process model during operation; see Chapter 10.

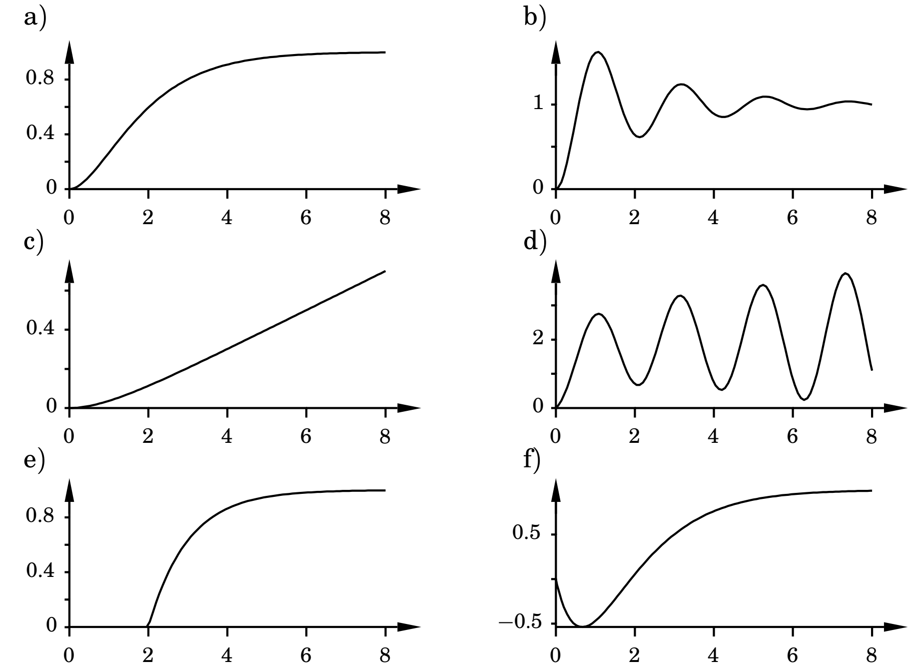


Figure 2.2 Open-loop step responses.

2.3 Dynamic Models

A static process model like the one discussed in the previous section tells the steady-state relation between the input and the output signal. A dynamic model should give the relation between the input and the output signal during transients. It is naturally much more difficult to capture dynamic behavior. This is, however, essential when dealing with control problems.

Qualitative Characterization of Process Dynamics

Before attempting to model a system it is often useful to give a crude characterization of its dynamical behavior. To describe the dynamical behavior we will simply show the response of the system to a step change in the manipulated variable. This is called the step response of the system or the process reaction curve.

One distinction is between stable and unstable systems. The step response of a stable system goes to a constant value. An unstable system will not reach a steady state after a step change. Systems with integrating action are a typical example of an unstable system. In early process control literature, stable systems were called self-regulating systems.

Many properties of a system can be obtained directly from the step response. Figure 2.2 shows step responses that are typically encountered in process control.

In Figure 2.2a, the process output is monotonically changed to a new stationary value. This is the most common type of step response encountered in process control. In Figure 2.2b, the process output oscillates around its final stationary value. This type of process is uncommon in process control. One case where it occurs is in concentration control of recirculation fluids. In mechanical designs, however, oscillating processes are common where elastic materials are used, e.g., weak axles in servos, spring constructions, etc. The systems in Figures 2.2a and 2.2b are stable, whereas the system shown in Figures 2.2c and 2.2d are unstable. The system in Figure 2.2c is an integrating process. Examples of integrating processes are level control, pressure control in a closed vessel, concentration control in batches, and temperature control in well isolated chambers. The common factor in all these processes is that some kind of storage occurs in them. In level, pressure, and concentration control, storage of mass occurs, while in the case of temperature control there is a storage of energy. The system in Figure 2.2e has a long time delay. The time delay occurs when there are transportation delays in the process. The system in Figure 2.2f is a non-minimum phase system. Notice that the output initially moves in the wrong direction. The water level in boilers often reacts like this after a step change in feed water flow.

Linear Time-Invariant Systems

There is a restricted class of models, called linear time-invariant systems, that can often be used. Such models describe the behavior of systems for small deviations from an equilibrium. Time-invariant means that the behavior of the system does not change with time. Linearity means that the superposition principle holds. This means that if the input u_1 gives the output y_1 and the input u_2 gives the output y_2 it then follows that the input $au_1 + bu_2$ gives the output $ay_1 + by_2$.

A nice property of linear time-invariant systems is that their response to an arbitrary input can be completely characterized in terms of the response to a simple signal. Many different signals can be used to characterize a system. Broadly speaking, we can differentiate between transient and frequency responses.

In a control system we typically focus on only two signals, the control signal and the measured variable. Process dynamics deals with the relation between those signals. This means that it includes dynamics in actuators, process, and sensors. The dynamics are often dominated by process dynamics. In some cases it is, however, the sensors and actuators that give the major contribution to the dynamics. For example, it is very common that there are long filter-time constants in temperature sensors. There may also be measurement noise and other imperfections. There may also be significant dynamics in the actuators. To do a good job of control, it is necessary to be aware of the physical origin of the process dynamics to judge if a good response in the measured variable actually corresponds to a good response in the physical process variable. Even if the attention is focused on the measured variable it is useful to always keep in mind that the process variable is the signal that really matters.

Physical Modeling—Differential Equations

A traditional way to obtain a process model is to use basic physical laws such as mass, momentum and energy balances. Such descriptions typically lead to

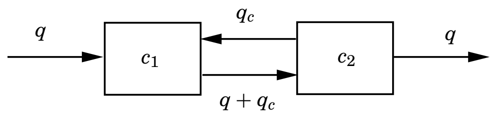


Figure 2.3 Schematic diagram of a system consisting of two tanks.

a mathematical model in terms of a differential equation. We illustrate this with two examples.

EXAMPLE 2.1—STIRRED TANK
Consider an ideal stirred tank reactor. Let the reactor volume be V and the volume flow rate through the reactor be q. The manipulated variable is the concentration u of the inflow, and the process variable y is the concentration in the reactor. A mass balance for the reactor gives

$$V\frac{dy}{dt} = q(u - y).$$

The parameter $T = V/q$, which has dimension time, is the average residence time of particles that enter the reactor. It is also called the time constant of the system. □

The system in Example 2.1 is of first order because only one variable is required to account for the storage in the tank. This is possible because the tank is well stirred so the concentration is constant throughout the tank. In more complicated cases many variables are required to account for the storage of mass, energy, and momentum. This is illustrated in the next example.

EXAMPLE 2.2—COUPLED TANKS
Consider the system shown in Figure 2.3, which is composed of two well-stirred tanks. Assume that each tank has volume V, that the inflow and the outflow are q, and that the reflux flow is q_c. Furthermore, let the input be the concentration of the inflow $u = c_{in}$, and let the output be the concentration of the outflow, $y = c_{out}$. When the tanks are well stirred the mass balance can be characterized by the concentrations in the tanks. The mass balances for the tanks become

$$\begin{aligned} V\frac{dc_1}{dt} &= -(q + q_c)c_1 + q_c c_2 + qu \\ V\frac{dc_2}{dt} &= (q + q_c)c_1 - (q + q_c)c_2 \\ y &= c_2. \end{aligned}$$

□

The model in the example consists of two differential equations of first order. There are two differential equations because the system is completely described by mass balances and the storage of mass can be captured by two variables. Similar descriptions are obtained for more complicated systems, but the number of equations increases with the complexity of the system. The differential equation may also be nonlinear if there are nonlinear transport phenomena.

The model in Example 2.2 consists of a system of first-order differential equations. If we are only interested in the relations between the input u and the output y a linear model can also be described by a differential equation of higher order, i.e.,

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \ldots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + \ldots + b_n u. \tag{2.1}$$

The number n is equal to the number of variables required to account for the storage. This is one of the standard models used in automatic control.

The differential equation (2.1) is characterized by two polynomials

$$\begin{aligned} a(s) &= s^n + a_1 s^{n-1} + \cdots + a_n \\ b(s) &= b_1 s^{n-1} + \cdots + b_n, \end{aligned} \tag{2.2}$$

where the polynomial $a(s)$ is called the characteristic polynomial. The zeros of the polynomial $a(s)$ are called the *poles* of the system, and the zeros of the polynomial $b(s)$ are called the *zeros* of the system.

The differential equation (2.1) has a solution of the form

$$y(t) = \sum kC_k(t) e^{\alpha_k t} + \int_0^t g(t-\tau) u(\tau) d\tau, \tag{2.3}$$

where α_k are the poles of the system and $C_k(t)$ are polynomials (constants if the poles are distinct). The first term of the above equation depends on the initial conditions and the second on the input. The function g has the same form as the first term of the right-hand side of (2.3). The poles thus give useful qualitative insight into the properties of the system.

In more complicated situations it may be more difficult to account for the storage of mass momentum and energy. We illustrate with a simple example.

EXAMPLE 2.3—TIME DELAY

Consider a system where mass is transported on a conveyor belt. Let the input $u(t)$ be the mass flow rate onto the belt, and let the output $y(t)$ be the mass flow out of the belt. The input-output relation for the system is then

$$y(t) = u(t - L), \tag{2.4}$$

where L is the time it takes for a particle to pass the belt. To account for the storage of mass on the belt it is necessary to specify the mass distribution on the belt. The output is thus simply the delayed input. This system is therefore called a time delay or a transport delay. A time delay is also called a dead time. The model (2.4) also describes the concentration in a pipe with no mixing. □

Other physical systems such as heat conduction and diffusion give rise to models in terms of partial differential equations; examples of such models are given in Section 2.5.

An attractive feature of physical models is that the parameters of the equation can be related to physical quantities such as volumes, flows, and material constants. Complicated models can also be constructed by dividing a system into subsystems, deriving simple models for each subsystem and combining the simple models.

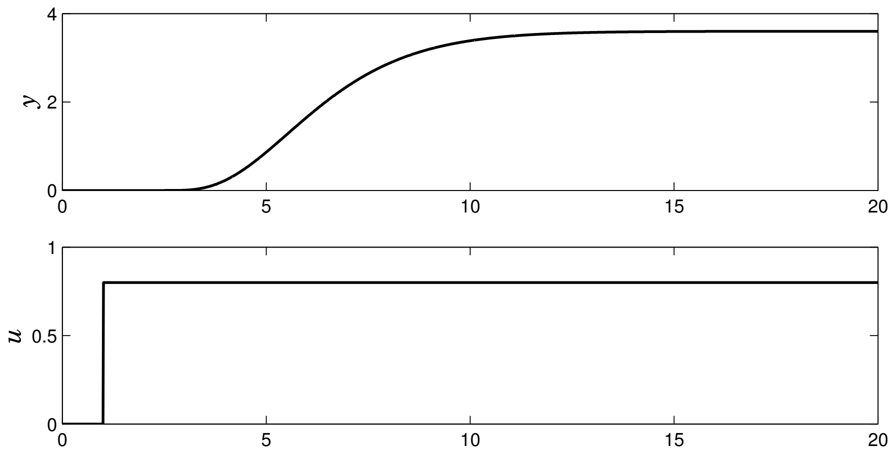


Figure 2.4 The lower curve shows an input signal in the form of a step, and the upper curve shows the response of the system to the step.

State Models

The notion of *state* is an important concept in system dynamics. The state is a collection of variables that summarize the past behavior of the system and admits a prediction of the future under the assumption that future inputs are known. For the system in Example 2.2, which consists of two tanks, the state is simply the concentrations c_1 and c_2 in the tanks. In general, the state is the variables required to describe storage of mass, momentum, and energy of a system. Sometimes it is necessary to use infinitely many variables to describe storage in a system. For the system in Example 2.3 the state at time t is the past inputs over an interval of length L, i.e., $\{u(\tau), t - L \leq \tau < t\}$.

Transient Responses

An alternative to describing models by differential equations is to focus directly on the input-output behavior. Dynamics can in principle be described by a large table of input signals and corresponding output signals. This approach, which is called transient response, is perhaps the most intuitive way to characterize process dynamics. A very nice property of linear time-invariant systems is that the table can be described by one pair of signals. The particular input signal is often chosen so that it is easy to generate experimentally. Typical examples are steps, pulses, and impulses. Recall that typical step responses were shown in Figure 2.2.

Because of the superposition principle the amplitude of the signals can be normalized. For simplicity it is common practice to normalize by dividing the output with the magnitude of the input step. It is also common practice to translate the curve so that the step starts at time $t = 0$. It is then sufficient to show the output only. This practice will be followed in this book. For example, in Figure 2.4 the output should be divided by 0.8 and translated one unit to the left. In early process control literature the step response was also called the reaction curve.

The output generated by an arbitrary input can be computed from the step response. Let $h(t)$ be the response to a unit step. The output $y(t)$ to an arbitrary input signal $u(t)$ is then given by

$$y(t) = \int_0^t u(\tau)\,\frac{dh(t-\tau)}{dt}\,d\tau = \int_0^t u(\tau)g(t-\tau)d\tau, \tag{2.5}$$

where we have introduced $g(t)$ as the derivative of the step response $h(t)$. The function $g(t)$ is called the impulse response of the system because it can be interpreted as the response of the system to a very short impulse with unit area.

The Transfer Function

The formula (2.5) can be simplified significantly by introducing Laplace transforms. The Laplace transform $F(s)$ of a time function $f(t)$ is defined as

$$F(s) = \int_0^\infty e^{-st} f(t)dt. \tag{2.6}$$

Assuming that the system is initially at rest, i.e., $y(t) = 0$ and $u(t) = 0$ for $t \le 0$, and using Laplace transforms, Equation 2.5 can be written as

$$Y(s) = G(s)U(s), \tag{2.7}$$

where $U(s)$, $Y(s)$, and $G(s)$ are the Laplace transforms of $u(t)$, $y(t)$, and $g(t)$, respectively. The function $G(s)$ is called the transfer function of the system. The transfer function $G(s)$ is also the Laplace transform of the impulse response $g(t)$.

The formula given by (2.7) has a strong intuitive interpretation. The Laplace transform of the output is simply the Laplace transform of the input multiplied by the transfer function of the system. This is one of the main reasons for using Laplace transforms when analyzing linear systems. Analysis of linear systems is reduced to pure algebra. A nice feature is that processes, controllers, and signals are described in the same way.

Equation 2.7 can also be used to define the transfer function as the ratio of the Laplace transforms of the input and the output of a system. As illustrations we will give the transfer function for some systems.

EXAMPLE 2.4—STIRRED TANK

The stirred tank in Example 2.1 has the transfer function

$$G(s) = \frac{1}{sV/q + 1} = \frac{1}{sT+1}, \tag{2.8}$$

where the quantity $T = V/q$, which has dimension time, is called the time constant of the system. □

EXAMPLE 2.5—TIME DELAY
Consider the system describing a transport delay in Example 2.3. Assuming that $u(t) = 0$ for $-L \leq t \leq 0$ we find

$$Y(s) = \int_0^\infty e^{-st} y(t) dt = \int_0^\infty e^{-st} u(t-L) dt = e^{-sL} U(s).$$

The transfer function of a transport delay is thus

$$G(s) = e^{-sL}. \tag{2.9}$$

□

Equation 2.7 implies that it is easy to obtain the transfer function of interconnected system. This is illustrated by the following example.

EXAMPLE 2.6—FIRST-ORDER SYSTEM WITH TIME DELAY (FOTD)
Consider a system that is a stirred tank that is fed by a pipe with no mixing. Multiplying the transfer function of the tank in Example 2.4 with the transfer function of a time delay in Example 2.5 we find that the system has the transfer function

$$G(s) = \frac{1}{1+sT} e^{-sL}. \tag{2.10}$$

This model is very common in process control. It is called a first-order system with a time delay or a FOTD system for short. □

Another nice property of Laplace transforms is that the transform of a derivative is given by the formula

$$\int_0^\infty e^{-st} f'(t) dt = s \int_0^\infty e^{-st} f(t) dt - f(0) = sF(s) - f(0).$$

If the initial value of the time function is zero it follows that differentiation of a time function corresponds to multiplication of the Laplace transform with s. Similarly, it can be shown that integration of a signal corresponds to dividing the Laplace transform with s. This gives a very simple rule for manipulating differential equations where initial values are zero. Simply replace functions with their corresponding Laplace transforms and derivatives by s. The relation between signals is then obtained by simple algebra.

EXAMPLE 2.7—GENERAL DIFFERENTIAL EQUATION
Consider the system described by the differential equation (2.1). Assuming that the system is initially at rest and taking Laplace transforms of (2.1) we get

$$(s^n + a_1 s^{n-1} + \ldots + a_n) Y(s) = (b_1 s^{n-1} + b_2 s^{n-2} + \ldots + b_n) U(s),$$

where $Y(s)$ is the Laplace transform of the output, and $U(s)$ the Laplace transform of the input. The transfer function of the system is the ratio of the Laplace transforms of output and input, i.e.,

$$G(s) = \frac{Y(s)}{U(s)} = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \ldots + b_n}{s^n + a_1 s^{n-1} + \ldots + a_n}. \tag{2.11}$$

□

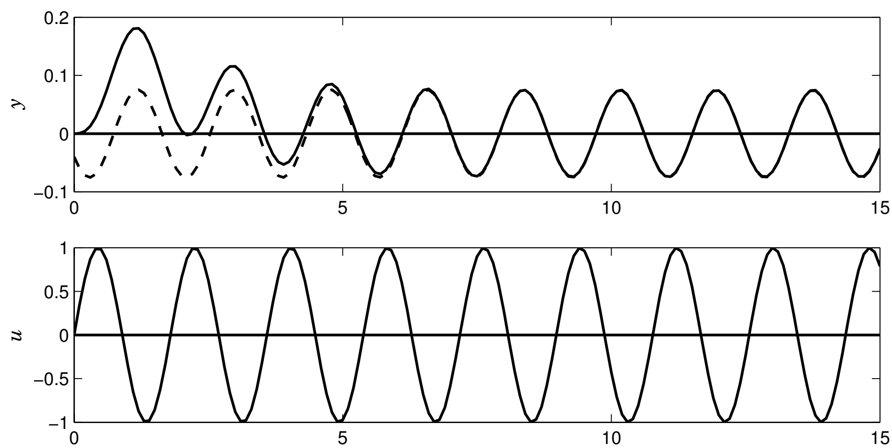


Figure 2.5 Illustration of frequency response. The input signal u is a sinusoid, and the output signal y becomes sinusoidal after a transient. The dashed line shows the steady-state response to the sinusoidal input.

EXAMPLE 2.8—PID CONTROLLER
The PID controller given by Equation 1.5 is a dynamical system with the transfer function

$$C(s) = \frac{U(s)}{E(s)} = K\left(1 + \frac{1}{sT_i} + sT_d\right). \tag{2.12}$$

□

The last two examples illustrate that transfer functions can be obtained from differential equations by inspection. The rule is simply to replace derivatives by s, integrals by $1/s$, and time functions by their transforms. The transfer functions are then obtained as the ratio between signals.

Frequency Response

Another way to characterize the dynamics of a linear time-invariant system is to investigate the response to sinusoidal input signals, an idea that goes back to the French mathematician Fourier. Frequency response is less intuitive than transient response, but it gives other insights.

Consider a stable linear system. If the input signal to the system is a sinusoid, then the output signal will also be a sinusoid after a transient (see Figure 2.5). The output will have the same frequency as the input signal. Only the phase and the amplitude are different. If the input signal is $u(t) = u_0 \sin \omega t$ the steady-state output is

$$y(t) = a(\omega)u_0 \sin\big(\omega t + \varphi(\omega)\big).$$

The steady-state relations between the output and a sinusoidal input with frequency ω can be described by two numbers: the amplitude ratio and the phase. The amplitude ratio is the output amplitude divided by the input amplitude,

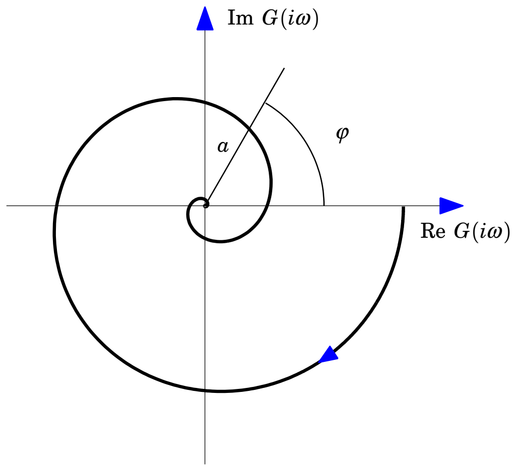


Figure 2.6 The Nyquist curve of a system is the locus of the complex number $G(i\omega)$ as ω goes from 0 to ∞.

and the phase is the phase shift of the output in relation to the input. The functions $a(\omega)$ and $\varphi(\omega)$ give amplitude ratio and phase for all frequencies. The functions $a(\omega)$ and $\varphi(\omega)$ are related to the transfer function in the following way.

$$G(i\omega) = a(\omega)e^{i\varphi(\omega)}. \tag{2.13}$$

The values of the transfer function for imaginary arguments thus describe the steady-state transmission of sinusoidal signals, and $G(i\omega)$ is called the frequency response function of the system.

The Nyquist Plot

There are very useful graphical illustrations of the frequency response. The complex number $G(i\omega)$ can be represented by a vector with length $a(\omega)$ that forms angle $\varphi(\omega)$ with the real axis (see Figure 2.6). When the frequency goes from 0 to ∞, the vector describes a curve in the plane, which is called the frequency curve or the Nyquist curve.

The Nyquist curve gives a complete description of the system. It can be determined experimentally by sending sinusoids of different frequencies through the system. This may, however, be time consuming. It can also be determined from other signals.

The Bode Plot

The Bode plot is another graphical representation of the transfer function. The Bode plot of a transfer function consists of two curves, the gain curve and the phase curve; see Figure 2.7. The amplitude or gain curve shows the amplitude ratio $a(\omega) = |G(i\omega)|$ as a function of the frequency ω. The phase curve shows the phase $\varphi(\omega) = \arg G(i\omega)$ as a function of the frequency ω. The frequency is given in logarithmic scales on both curves, either in rad/s or Hz. The gain is also given in logarithmic scales. The angle is given in linear scales. The Bode

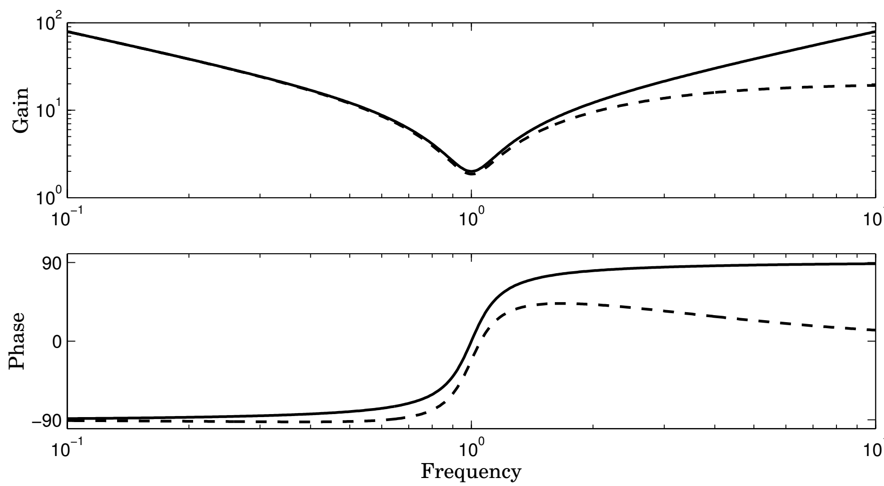


Figure 2.7 Bode plot of an ideal PID controller (solid lines) and a controller with a filter (dashed). The upper curve shows the gain curve $|G(i\omega)|$, and the lower diagram shows the phase curve $\arg G(i\omega)$. The controller has high gain for low frequencies, and the phase is $-90°$. The ideal controller also has high gain at high frequencies and the phase is $90°$. The controller with a filter has constant gain for high frequencies.

plot gives a good overview of the properties of a system over a wide frequency range. Because of the scales the gain curve also has linear asymptotes.

2.4 Feature-Based Models

Sometimes it is desirable to have a crude characterization of a process based on only a few features. The features should be chosen so that they are meaningful with good physical interpretation. They should also be easy to determine experimentally. This way of describing dynamics has a long tradition in process control. It is useful to start with a crude classification of step responses as illustrated in Figure 2.2.

Process Gain

For stable processes the steady-state behavior can be described by one parameter, the process gain K_p. For processes with integration a constant input gives in steady state an output that changes with a constant rate. This behavior can be captured by the rate constant K_v.

Average Residence Time

It is also useful to find a few parameters to characterize process dynamics. The time behavior of stable system with positive impulse response can be characterized with the parameter

$$T_{ar} = \frac{\int_0^\infty tg(t)dt}{\int_0^\infty g(t)dt}, \tag{2.14}$$

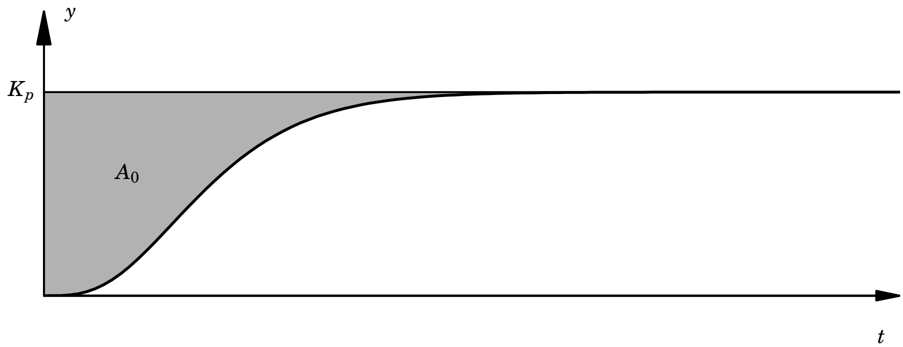


Figure 2.8 Illustrates the area method for determining the average residence time.

which is called the average residence time. The average residence time is a rough measure of how long it takes for the input to have a significant influence on the output. Notice that the function $g(t)/\int g(t)dt$ can be interpreted as a probability density if $g(t) \geq 0$.

The average residence time can be calculated from the step response in the following way

$$T_{ar} = \frac{\int\limits_0^\infty (h(\infty) - h(t))dt}{K_p} = \frac{A_0}{K_p}, \tag{2.15}$$

where $h(t)$ is the step response and $K_p = G(0)$ is the static process gain. Notice that $K_p = h(\infty)$ and that A_0 is the shaded area in Figure 2.8.

Average Residence Time and Transfer Functions

The average residence time can be computed very conveniently from the transfer function. Since the transfer function is the Laplace transform of the impulse response we have

$$G(s) = \int_0^\infty e^{-st} g(t)dt.$$

Differentiation of this expression with respect to s gives

$$G'(s) = -\int_0^\infty e^{-st} tg(t)dt.$$

Setting $s = 0$ in these expressions it then follows from the definition of the average residence time (2.14) that

$$T_{ar} = -\frac{G'(0)}{G(0)}. \tag{2.16}$$

This formula will now be illustrated by a few examples.

EXAMPLE 2.9—AVERAGE RESIDENCE TIME FOR STIRRED TANK
The transfer function for the stirred tank in Example 2.4 is

$$G(s) = \frac{1}{1 + sT}.$$

We have

$$G'(s) = -\frac{T}{(1 + sT)^2},$$

and it follows from (2.16) that the average residence time is

$$T_{ar} = T = \frac{V}{q}.$$

The average residence time is thus the ratio of the volume and the flow through the tank. □

EXAMPLE 2.10—AVERAGE RESIDENCE TIME FOR TIME DELAY
The transfer function for the time delay in Example 2.5 is

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

We have

$$G'(s) = -Le^{-sL},$$

and it follows from (2.16) that the average residence time is

$$T_{ar} = L.$$

The average residence time is thus equal to the time delay. □

EXAMPLE 2.11—AVERAGE RESIDENCE TIME FOR CASCADED SYSTEMS
A system that is the cascade combination of two stable linear systems with transfer functions $G_1(s)$ and $G_2(s)$ has the transfer function

$$G(s) = G_1(s)G_2(s).$$

Differentiation gives

$$G'(s) = G_1'(s)G_2(s) + G_1(s)G_2'(s).$$

It follows from (2.16) that the average residence time is

$$T_{ar} = -\frac{G_1'(0)G_2(0) + G_1(0)G_2'(0)}{G_1(0)G_2(0)} = -\frac{G_1'(0)}{G_1(0)} - \frac{G_2'(0)}{G_2(0)}.$$

The average residence time is the sum of the residence times of each system. □

It follows from this example that the average residence time for the FOTD model in Example 2.6 is $T_{ar} = L + T$.

A system with the transfer function

$$G(s) = \frac{K_p(1 + sT_1)(1 + sT_2)}{(1 + sT_3)(1 + sT_4)(1 + sT_5)} e^{-sL}.$$

has the average residence time $T_{ar} = T_3 + T_4 + T_5 + L - T_1 - T_2$.

Models with Two Parameters

A very simple way to characterize the dynamics of a stable process is to use the gain K_p and the average residence time T_{ar}. This gives the following models

$$\begin{aligned} G(s) &= \frac{K_p}{1+sT_{ar}} \\ G(s) &= K_p\, e^{-sT_{ar}}, \end{aligned} \tag{2.17}$$

where dynamics is either represented by a lag or a time delay.

Apparent Time Delay and Apparent Time Constant

Systems with essentially monotone step responses are very common in process control. Such systems can be modeled as first-order systems with time delay with the transfer function

$$G(s) = \frac{K_p}{1+sT}\, e^{-sL}. \tag{2.18}$$

To emphasize that the parameters L and T are approximate they are referred to as the *apparent time delay* and the *apparent time constant*, or the *apparent lag*, respectively. The average residence time is $T_{ar} = L + T$. The parameter

$$\tau = \frac{L}{T_{ar}} = \frac{L}{L+T}, \tag{2.19}$$

which has the property $0 \le \tau \le 1$, is called the *normalized time delay* or the *normalized dead time*. This parameter can be used to characterize the difficulty of controlling a process. It is sometimes also called the *controllability ratio*. Roughly speaking, processes with small τ are easy to control, and the difficulty in controlling the system increases as τ increases. Systems with $\tau = 1$ correspond to processes with pure time delay, which are difficult to control well.

Ultimate Gain and Ultimate Period

So far we have used features that are based on the transient response. It is also possible to use features of the frequency response. Models can be characterized in terms of their phase lags and the frequency, where the systems have a given phase lag. For this purpose, we introduce ω_φ to denote the frequency where the phase lag is φ degrees, and we introduce $K_\varphi = |G(i\omega_\varphi)|$ to denote the process gain at ω_φ. The frequencies ω_{90} and ω_{180} and the corresponding process gains K_{90} and K_{180} are of particular interest for PID control. These frequencies correspond to the intersections of the Nyquist curve with the negative imaginary and real axes; see Figure 2.9. They also have nice physical interpretations. Consider a process with pure proportional control. If the controller gain is increased the process will start to oscillate, and it will reach the stability limit when the controller gain is $K_u = 1/K_{180}$. The oscillation will have the frequency ω_{180}. This frequency is called the ultimate frequency. The parameter K_u is called the ultimate gain or the critical gain. The parameters K_{90} and ω_{90} have similar interpretations for a process with pure integral control.

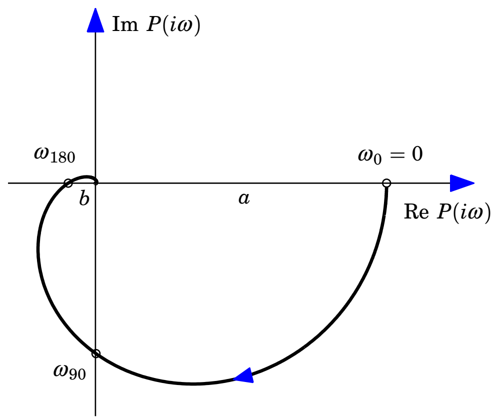


Figure 2.9 Nyquist curve with the points ω_0, ω_{90} and ω_{180}. The gain ratio κ is the ratio of the distances a and b.

The Gain Ratio

The *gain ratio* is an additional parameter that gives useful information about the system. This parameter is defined as

$$\kappa = \frac{K_{180}}{K_p} = \frac{|G(i\omega_{180})|}{G(0)}. \tag{2.20}$$

It is an indicator of how difficult it is to control the process. Processes with a small κ are easy to control. The difficulty increases with increasing κ. The parameter is also the ratio between the distances a and b in the Nyquist plot; see Figure 2.9.

Parameter κ is also related to the normalized time delay τ. For the FOTD model given by Equation 2.18 the parameters τ and κ are related in the following way:

$$\tau = \frac{\pi - \arctan\sqrt{1/\kappa^2 - 1}}{\pi - \arctan\sqrt{1/\kappa^2 - 1} + \sqrt{1/\kappa^2 - 1}}. \tag{2.21}$$

This relation is close to linear as is shown in Figure 2.10. This relation holds approximately for many other systems. As a crude approximation we can thus equate κ and τ. For small values a better approximation is given by $\kappa = 1.6\tau$. For the FOTD model it is also possible to find the parameters L and T from κ and ω_{180} using the following equations

$$\begin{aligned} T &= \frac{1}{\omega_{180}}\sqrt{\kappa^{-2} - 1} \\ L &= \frac{1}{\omega_{180}}(\pi - \arctan\kappa^{-2} - 1) \\ K &= \frac{|G(i\omega_{180})|}{\kappa}. \end{aligned} \tag{2.22}$$

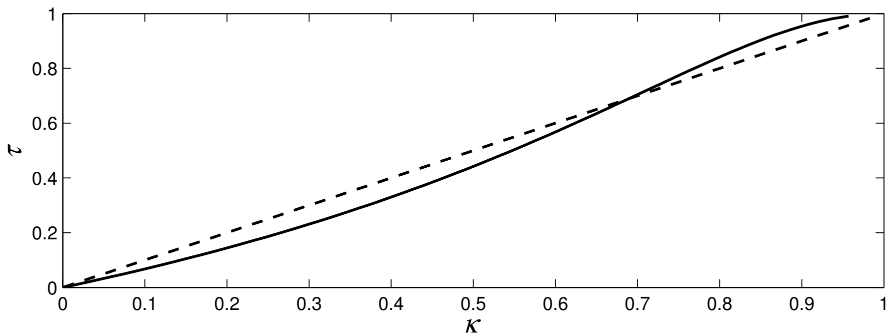


Figure 2.10 The normalized time delay τ as a function of gain ratio κ for the system (2.18). The dashed line shows the straight line approximation $\kappa = \tau$.

2.5 Typical Process Models

Much of the dynamic behavior encountered in control is relatively simple. Processes are designed in such a way that they should be easy to control. If PID control is used it is thus natural that simple process models are used. In this section we will discuss some of the models that are commonly used in connection with PID control. Most of these models are characterized by a few parameters only.

The FOTD Model

A process model that is commonly used in process control has the transfer function (2.18). It is simple and it describes the dynamics of many industrial processes approximately. A comparison with Examples 2.3 and 2.4 shows that it can represent the dynamics of a stirred tank with a pipe without mixing. The model is characterized by three parameters: the (static) gain K_p, the time constant T, and the time delay L. The time constant T is also called the lag. The step response of the model (2.18) is

$$h(t) = K_p \left(1 - e^{-(t-L)/T}\right).$$

Since the average residence time is $T_{ar} = L + T$, the value of the step response at this time becomes

$$h(T_{ar}) = K_p \left(1 - e^{-1}\right) \approx 0.63 K_p.$$

The average residence time can thus be determined as the time when the step response has reached 63 percent of its steady-state value.

Two parameters of the model (2.18) correspond to scaling of the axes and can be reduced by normalization. They can be chosen as the gain and the average residence time. This means that if the output is scaled by the gain $K_p = G(0)$ and time by the average residence time T_{ar} the response is completely characterized by one parameter, the normalized dead time τ. The system is a pure time delay for $\tau = 1$ and a first-order system or a pure lag for $\tau = 0$.

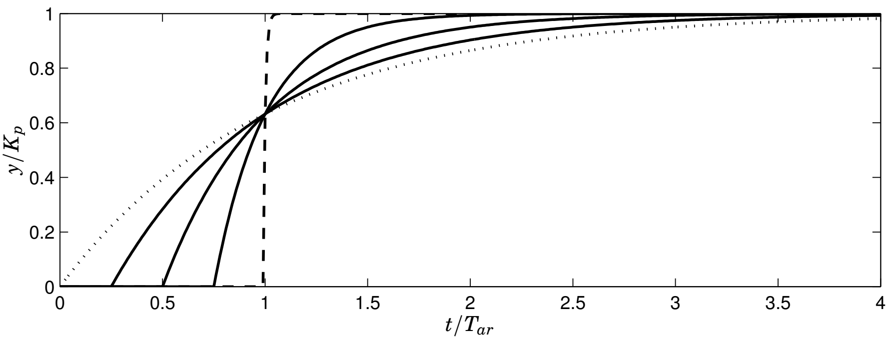


Figure 2.11 Normalized step responses of the FOTD model (2.18) for different values of the normalized time delay. The normalized time delay is $\tau = 0$ (dotted), 0.25, 0.5, 0.75 and 0.99 (dashed).

Figure 2.11 shows the normalized step responses for different values of τ. Notice that all curves intersect at one point $t = T_{ar}$ because of the normalization.

Noninteracting Tanks or Multiple Lags

The transfer function (2.8) represents the dynamics of a simple tank. The upper part of Figure 2.12 shows a system that is a cascade combination of n tanks. This system has the transfer function

$$G_n(s) = \frac{K_p}{(1+sT)^n}. \tag{2.23}$$

where n is the number of tanks. Since a first-order system is also called a lag the system is also called a multiple lag system. Notice that this formula holds only if the outflow of each tank only depends on its level. This means that there is no interaction between the tanks.

The average residence time is

$$T_{ar} = -\frac{G'(0)}{G(0)} = nT.$$

The model (2.23) has the impulse response

$$g(t) = \frac{K_p}{(n-1)!}\frac{t^{n-1}}{T^n}e^{-t/T}, \tag{2.24}$$

which has its maximum

$$\max g(t) = \frac{K_p(n-1)^{n-2}}{T(n-2)!}e^{-n+1},$$

for $t = (n-1)T$. The unit step response is

$$h(t) = K_p\left(1 - \left(1 + \frac{t}{T} + \frac{t^2}{2T^2} + \ldots + \frac{t^{n-1}}{(n-1)!T^{n-1}}\right)\right)e^{-t/T}. \tag{2.25}$$

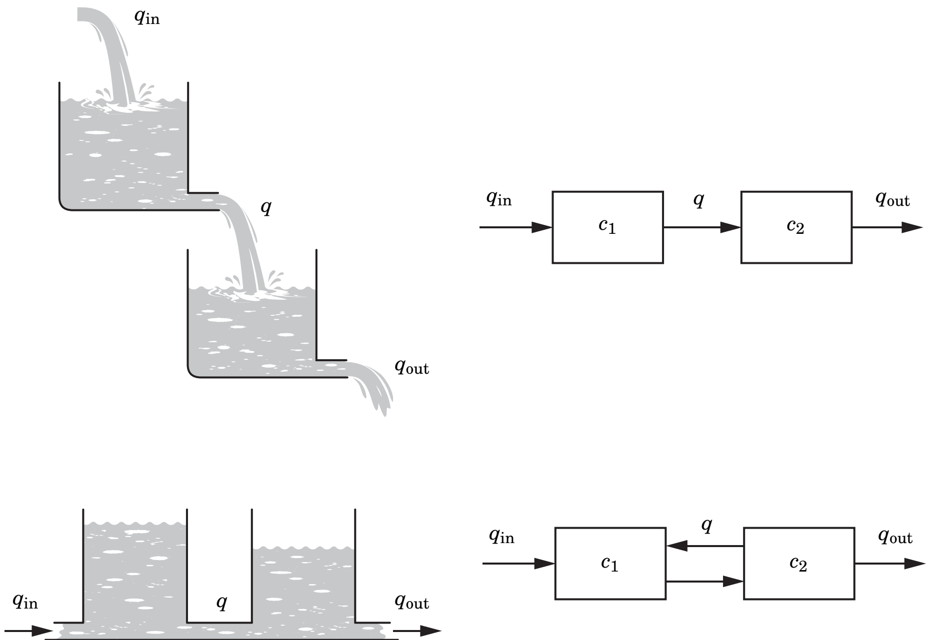


Figure 2.12 Cascaded tanks and corresponding block diagram representations. The upper tanks are noninteracting, and the lower are interacting.

The step response is characterized by three parameters, K_p, n, and T. The number of parameters can be reduced by normalization. Parameters K_p and T only influence the scaling of the axes. The shape of the step response is thus uniquely given by the parameter n. Normalized step responses for different values of n are shown in Figure 2.13. The step responses are close but not equal at $t = T_{ar}$. As n goes to infinity we have

$$\lim_{n\to\infty} G_n(s) = K_p e^{-t/T_{ar}}.$$

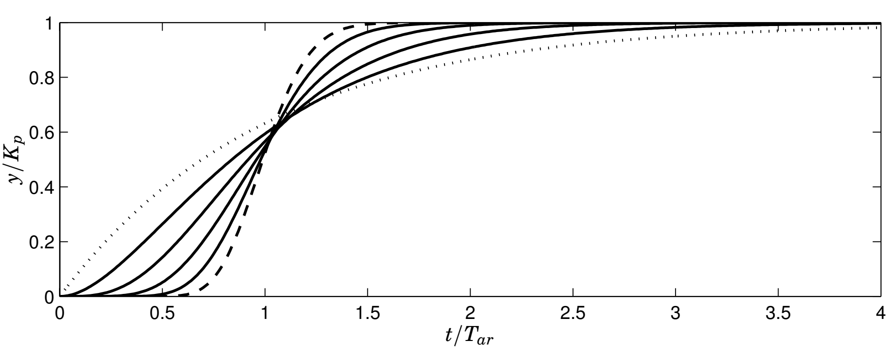


Figure 2.13 Normalized step responses for the processes $G_n(s) = 1/(1+sT)^n$ for $n = 1$ (dotted), 2, 4, 8, 16, and 32 (dashed).

Table 2.1 Apparent time constant T_e, apparent time delay L_e, average residence time T_{ar}, and normalized time delay τ for the process (2.23).

n	2	3	4	5	6	7	8	16	32
T_e	1.86	2.44	2.91	3.32	3.68	4.01	4.31	6.23	8.90
L_e	0.28	0.81	1.43	2.10	2.81	3.55	4.31	10.78	24.67
T_{ar}	2.14	3.25	4.34	5.42	6.49	7.56	8.62	17.02	33.57
τ	0.13	0.25	0.33	0.39	0.43	0.47	0.50	0.63	0.73

For large n the system thus approaches a pure time delay. Figure 2.13 shows, however, that very large values of n are required to get a good approximation of the step response of a time delay.

The transfer function $G_n(s)$ can be approximated by an FOTD system. The apparent time constants and time delays for the approximation are given in Table 2.1.

Multiple Interacting Tanks—Distributed Lags

The dynamics of cascaded tanks are very different if the tanks are interacting. In the system shown in the lower part of Figure 2.12 the outflow of a tank depends on the levels of the neighboring tanks. Let x_k be the level of tank k. The control variable u is the inflow to the first tank, and let the output be the outflow of tank n. Assume that the tanks have unit cross-section, and assume that the flow from tank k to tank $k+1$ is $x_k - x_{k-1}$. The mass balances for the tanks are

$$\begin{aligned} \frac{dx_1}{dt} &= -x_1 + x_2 + u \\ &\vdots \\ \frac{dx_k}{dt} &= x_{k-1} - 2x_k + x_{k+1} \\ &\vdots \\ \frac{dx_n}{dt} &= x_{n-1} - 2x_n. \end{aligned} \tag{2.26}$$

This is a state model with n states. The state variables represent the levels in the different tanks. The system is also called a distributed lag. With a unit step input the equilibrium values of the states are $x_k = n-k+1$. The characteristic polynomials of systems having different order are

$$\begin{aligned} d_1 &= s+1 \\ d_2 &= s^2 + 3s + 1 \\ d_n &= (s+2)d_{n-1} - d_{n-2}. \end{aligned}$$

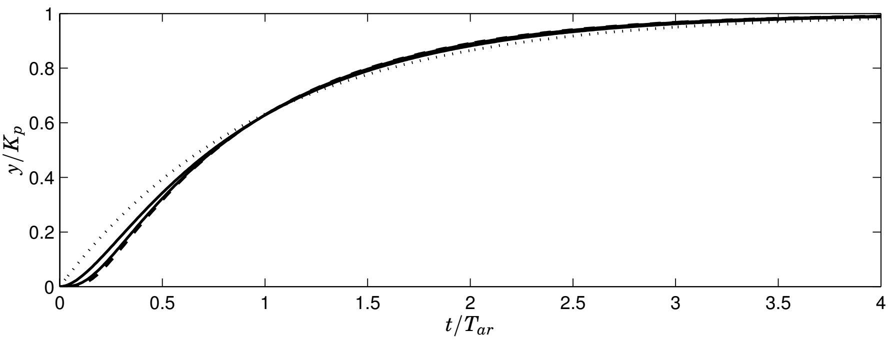


Figure 2.14 Normalized step responses for interacting tanks, (2.26), for $n = 1$ (dotted), 2, 4, and 8 (dashed).

The transfer functions for a few values of n are given by

$$
\begin{aligned}
G_1(s) &= \frac{1}{s+1} \\
G_2(s) &= \frac{1}{s^2+3s+1} \\
G_4(s) &= \frac{1}{s^4+7s^3+15s^2+10s+1} \\
G_8(s) &= \frac{1}{s^8+15s^7+91s^6+286s^5+495s^4+462s^3+210s^2+36s+1}.
\end{aligned}
$$

The average residence time is the ratio of the total steady-state volume to the flow, hence

$$T_{ar} = \frac{n(n+1)}{2}.$$

This is also the coefficient of the s-term in the denominator of the transfer function. As the number of tanks increases we have asymptotically for large n

$$G_n(s) \approx \frac{1}{\cosh \sqrt{2T_{ar}s}}.$$

These transfer functions are very different from the transfer function (2.23) of noninteracting tanks.

Figure 2.14 shows normalized step responses for interacting tanks. Notice that the responses are very similar for larger values of n. A comparison with Figure 2.13 shows that there is a significant difference between interacting and noninteracting tanks.

Another Version of Interacting Tanks

The model (2.26) is not the only way to interconnect tanks. Another configuration is shown in Figure 2.15. For simplicity we have shown a system with

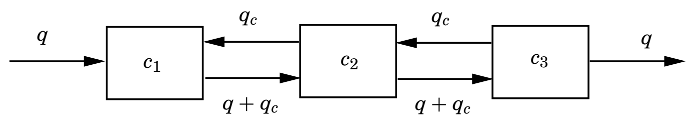


Figure 2.15 Schematic diagram of three cascaded tanks with recirculation.

three tanks. The system consists of identical stirred tanks with a forward flow q and a recirculation flow q_c. Let V be the tank volume, u the concentration of the inflow, c_k the concentration in the kth tank, and $y = c_n$ the concentration in the outflow. The mass balances for a system with n tanks are

$$
\begin{aligned}
V\frac{dc_1}{dt} &= -(q+q_c)c_1 + q_c c_2 + qu \\
&\vdots \\
V\frac{dc_k}{dt} &= (q+q_c)c_{k-1} - (q+2q_c)c_k + q_c c_{k+1} \\
&\vdots \\
V\frac{dc_n}{dt} &= (q+q_c)c_{n-1} - (q+q_c)c_n \\
y &= c_n.
\end{aligned}
$$

This is also a state model where the states are the concentrations in the different tanks. The transfer functions for a few values of n are

$$
\begin{aligned}
G_1(s) &= \frac{q}{Vs+q} \\
G_2(s) &= \frac{q(q+q_c)}{Vs^2 + 2V(q+q_c)s + q(q+q_c)} \\
G_3(s) &= \frac{2q_c(q+q_c)^2}{(Vs+q+q_c)(V^2s^2 + (2q+3q_c)s + 2q_c(q+q_c))} \\
&= \frac{2q_c(q+q_c)^2}{V^3s^3 + (3q+4q_c)V^2s^2 + (q+q_c)(q+3q_c)Vs + 2q_c(q+q_c)^2}.
\end{aligned}
$$

The static gain is $K_p = 1$ and the average residence time is

$$
T_{ar} = \frac{nV}{q}.
$$

The recirculation flow has a major impact on the dynamics. For $q_c = 0$ there is no interaction, and the system is equivalent to the model given by (2.23). As $q_c/q \to \infty$ the model is equivalent to the model given by (2.26). The model with recirculation thus makes it possible to interpolate between the models with noninteracting and distributed lags.

Figure 2.16 shows the step response of a system of n:th order for different values of the recirculation flow.

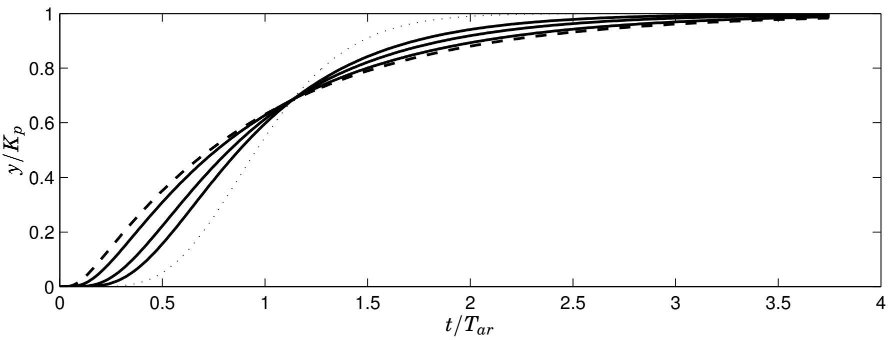


Figure 2.16 Normalized step responses for eight tanks with recirculation. The recirculation ratio is $q_c/q = 0$ (dotted), 1, 2, 5 and 10 (dashed).

Oscillatory Systems

The model (2.18) cannot describe systems with oscillatory responses. A simple model for such systems is given by the transfer function

$$G(s) = \frac{K_p}{1 + 2\zeta sT + (sT)^2}. \tag{2.27}$$

This model has three parameters: static gain K_p, time constant T, and relative damping ζ. The parameter $1/T$ is also called undamped natural frequency. The step responses can be normalized by the gain and the time constant. Its shape is then determined by one parameter only. The step responses are shown in Figure 2.17. For $\zeta < 1$ the step response has its maximum

$$M = K_p e^{-\frac{\pi\zeta}{\sqrt{1-\zeta^2}}},$$

which occurs at

$$t_{max} = \frac{2\pi T}{\sqrt{1-\zeta^2}}.$$

The position of the maximum increases with increasing ζ, and it becomes infinite for $\zeta = 1$ when the overshoot disappears. The transfer function is then

$$G(s) = \frac{K_p}{(1+sT)^2},$$

and the step response is

$$h(t) = K_p\Big(1 - e^{-t/T} - \frac{t}{T}e^{-t/T}\Big).$$

The Bode plots of the systems are shown in Figure 2.18.

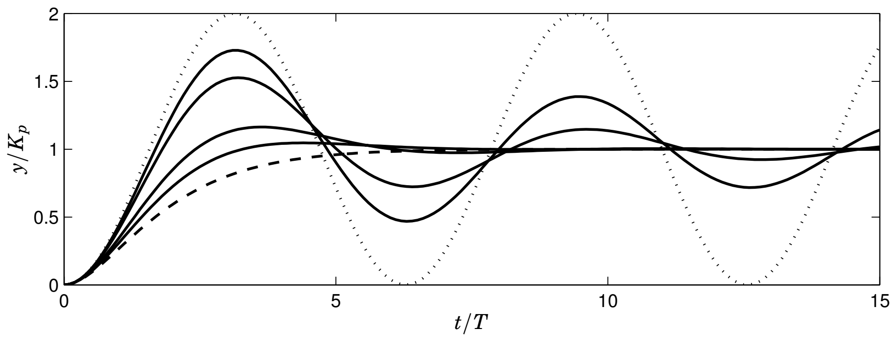


Figure 2.17 Normalized step responses of oscillatory systems (2.27) with $\zeta = 0$ (dotted), 0.1, 0.2, 0.5, 0.7, and 1.0 (dashed).

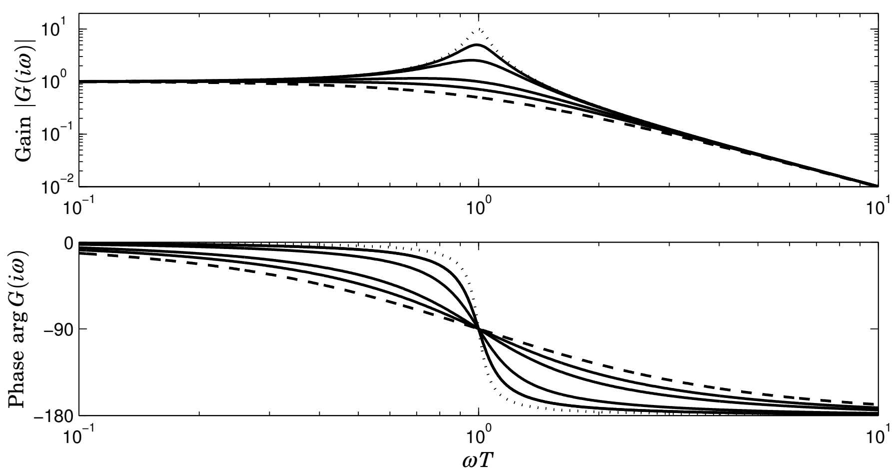


Figure 2.18 Bode plots of oscillatory systems (2.27) with $\zeta = 0.05$ (dotted), 0.1, 0.2, 0.5, 0.7, and 1.0 (dashed).

Processes with Integration

Integrating processes will not reach steady state during open-loop conditions. In practice, the same is true for processes with very long time constants. Asymptotically, the output will change at constant rate after a step change in the control signal. In the early process control literature these systems were said to be without self-regulation because the process variable did not reach a steady state after a disturbance. Many methods for PID tuning also treat such systems separately. Models for such systems are obtained simply by dividing the transfer function of a process with self-regulation by s.

A combination of an integrator and a time delay is a common model. The transfer function is

$$G(s) = \frac{K_v}{s}\, e^{-sL}. \tag{2.28}$$

This model is characterized by two parameters, a gain and a time delay. The

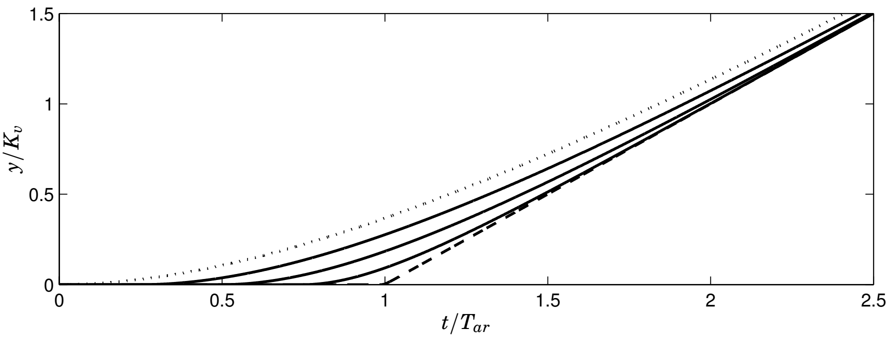


Figure 2.19 Normalized step responses for the FOTDI model (2.30) for $\tau = 0$ (dotted), 0.25, 0.5, 0.75, and 0.99 (dashed).

integrating gain is denoted by a special symbol K_v, which tells how fast the output increases in steady state after a unit input step change. The parameter K_v has dimension frequency.

A combination of a lag and an integrator is a model that is commonly used to describe simple drive systems. This model has the transfer function

$$G(s) = \frac{K_v}{s(1+sT)}. \tag{2.29}$$

The transfer function $K_v/(1+sT)$ represents the transfer function from the voltage of the drive system to the rate of rotation, and the integrator represents the relation between angular rate and angle.

A slightly more complicated model is obtained by adding integration to the standard model (2.18).

$$G(s) = \frac{K_v}{s(1+sT)}\, e^{-sL}. \tag{2.30}$$

We call this the FOTD model with integration or FOTDI for short. This process can be normalized in the same way as the model (2.18) by introducing the normalized time delay given by (2.19). The normalized step responses of the FOTDI model are shown in Figure 2.19.

Systems with Inverse Responses

The systems discussed so far do not have any zeros. Systems that are represented as a parallel connection of several systems can have transfer functions of the type

$$G(s) = \frac{1+sT}{s^2+1.4s+1}. \tag{2.31}$$

This system has a zero at $s = -1/T$, which may have a significant influence on the response of the system. Figure 2.20 shows the step response of this system for $T = -2$, -1, 0, 1, and 2. Notice that the overshoot of the step response increases with increasing positive values of T. Also notice that the output signal initially moves in the wrong direction when T is negative. Such

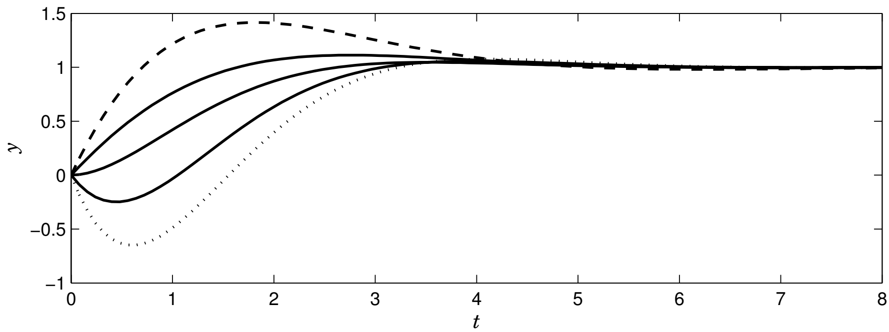


Figure 2.20 Step responses for the model (2.31) for $T = -2$ (dotted), -1, 0, 1 and 2 (dashed).

systems are said to have inverse responses. Systems with inverse responses are difficult to control. Examples of such systems are level dynamics in steam generators, dynamics of hydroelectric power stations, dynamics of backing cars, etc.

Heat Conduction

Temperature control is a very common application of PID control. Some models that are directly based on physics will now be discussed. Consider an infinitely long rod with thermal diffusivity λ. Assume that there is no radial heat transfer and that the input is the temperature at the left end of the rod. The transfer function to a point at the distance a from the left end point is

$$G(s) = e^{-\sqrt{sT}}, \tag{2.32}$$

where $T = a^2/\lambda$. The impulse response of the system is given by

$$h(t) = \frac{\sqrt{T}}{2\sqrt{\pi}t^{3/2}}\, e^{-\frac{T}{4t}}. \tag{2.33}$$

This impulse response has the property that all its derivatives are zero for $t = 0$, which means that the initial response of the system is very slow. The impulse response has a maximum at $t = T/6$. For large values of t the impulse response decays very slowly as $t^{-1.5}$. The step response of the system is

$$y(t) = 1 - \operatorname{erf}\sqrt{\frac{T}{4t}} = 1 - \frac{2}{\sqrt{\pi}} \int_{-\infty}^{\sqrt{T/4t}} e^{-x^2} dx. \tag{2.34}$$

The step and impulse responses are shown in Figure 2.21. Notice that the temperature starts to rise very slowly initially. After a rapid rise it also approaches the steady state very slowly.

We will now instead consider the situation when the right-hand side is isolated. The transfer function then becomes

$$G(s) = \frac{1}{\cosh\sqrt{sT}}. \tag{2.35}$$

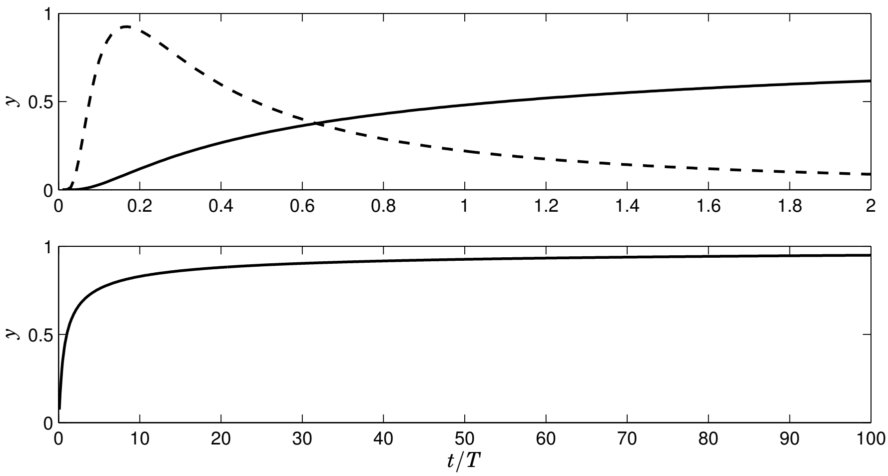


Figure 2.21 Step response (solid) and impulse response (dashed) for a system with the transfer function $e^{-\sqrt{sT}}$. The upper curves show the step responses and the impulse response. The lower curve shows the step response in a different time scale.

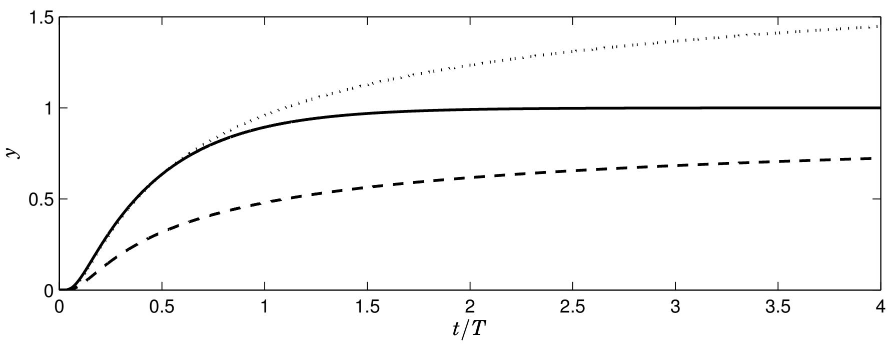


Figure 2.22 Step responses for the transfer function $1/\cosh\sqrt{sT}$ (solid), $e^{-\sqrt{sT}}$ (dashed) and $2e^{-\sqrt{sT}}$ (dotted).

This is a system with infinitely many lags with time constants $4T/\pi^2$, $T/9\pi^2$, $4T/25\pi^2$, $4T/49\pi^2$,.... This transfer function is also called a distributed lag.

The step response of this transfer function is shown in Figure 2.22. Notice that the response approaches the steady-state value faster than the system (2.32). The step response of the systems are thus quite different. The isolation of the right end of the rod makes it much easier to transfer heat into the system. A simple calculation shows that the average residence time for the system is

$$T_{ar} = -G'(0) = \frac{T}{2}. \tag{2.36}$$

The system (2.32) with the transfer function $e^{-\sqrt{sT}}$ has infinite residence time which reflects the fact that the impulse response decays very slowly; compare with Figure 2.21.

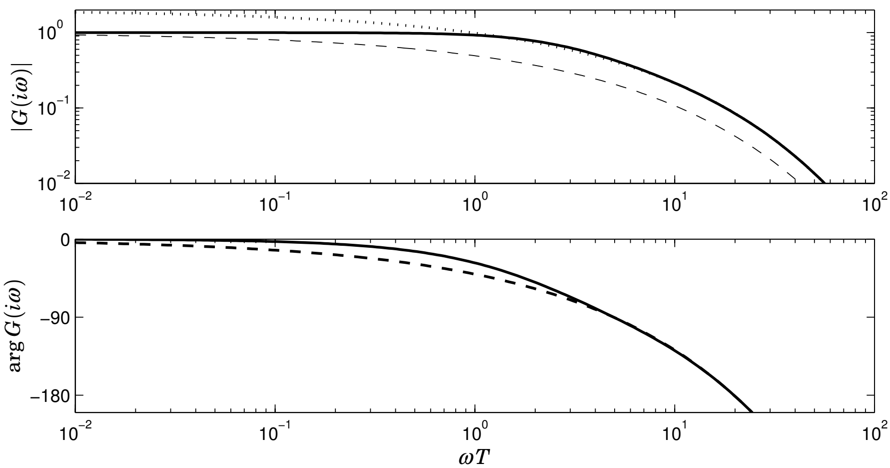


Figure 2.23 Bode plots for the transfer functions $1/\cosh\sqrt{sT}$ (solid lines), $e^{-\sqrt{sT}}$ (dashed). For comparison we also show the gain curve for the transfer function $2e^{-\sqrt{s}}$ (dotted).

We will now investigate the frequency responses of the systems (2.32) and (2.35). It can be shown that both transfer functions have a phase lag of 180° at the same frequency.

$$\omega_{180} = \frac{2\pi^2}{T}. \tag{2.37}$$

The magnitudes of the transfer functions at ω_{180} are given by

$$|e^{-\sqrt{i\omega_{180}T}}| = e^{-\pi} \approx 0.04321$$
$$\frac{1}{|\cosh\sqrt{i\omega_{180}T}|} = \frac{2e^{-\pi}}{1+e^{-2\pi}} \approx 0.08627.$$

At the frequency where the phase lag is 180° the gain of the system (2.35) is thus very close to twice as high as the gain for the system (2.32). The Bode plots of the system are shown in Figure 2.23. Notice that for frequencies above 2 rad/s there are very small differences between the transfer functions $2e^{-\sqrt{s}}$ and $1/\cosh\sqrt{sT}$ even if the step responses differ significantly. This observation is very important for the design of control systems. Figure 2.22 also shows that the step responses for the transfer functions $2e^{-\sqrt{s}}$ and $1/\cosh\sqrt{sT}$ are very close.

A Heat Exchanger

The transfer function from input temperature to output temperature of an ideal heat exchanger is

$$G(s) = \frac{1}{sT}(1 - e^{-sT}). \tag{2.38}$$

The step and impulse responses of this system are shown in Figure 2.24. Notice that the step response settles to the final value at time $t = T$ and that the

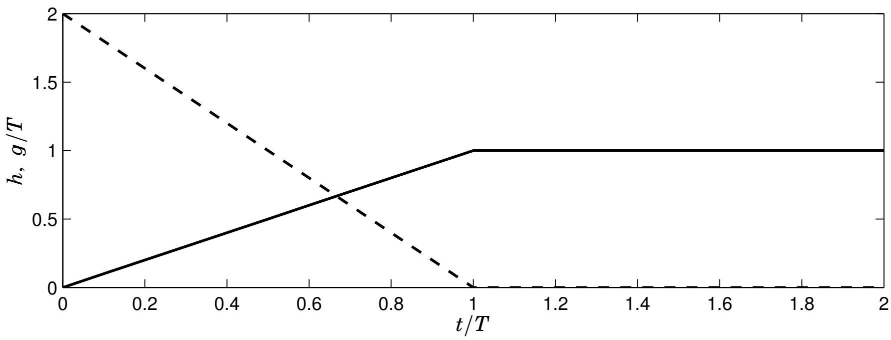


Figure 2.24 Normalized step (solid) and impulse (dashed) responses for the transfer function (2.38) of an ideal heat exchanger.

impulse response is zero after that time. This reflects the fact that once liquid has passed through the heat exchanger its temperature is no longer influenced. The average residence time of the system is

$$T_{ar} = \frac{T}{2}.$$

The frequency response of the system is

$$G(i\omega) = \frac{1}{i\omega T}\left(1 - e^{-i\omega T}\right).$$

The transfer function is zero for $\omega T = 2n\pi$. This is clearly seen in the Nyquist curve of the transfer function in Figure 2.25. An interesting property of this transfer function is that

$$\arg G(i\omega) = -\frac{\omega T}{2}, \quad \text{for } \omega T < 2\pi.$$

A Continuous Stirred Tank Reactor

Consider a continuous-time stirred tank reactor where the reaction $\mathcal{A} \to \mathcal{B}$ takes place. The reaction is exothermic, and reaction heat is removed by a coolant. The system is modeled by mass and energy balances. The mass balance is

$$\frac{dc}{dt} = \frac{q}{V}(c_f - c) - k(T)c, \tag{2.39}$$

where c [kmol$/m^3$] is the concentration of species $\mathcal{A}$, c_f the concentration of $\mathcal{A}$ in the feed, q [m^3/s] the volume flow rate, V [m^3] the reactor volume, and $k(T)$ [s^{-1}] the reaction rate which is a function of temperature

$$k(T) = k_0 e^{-E/RT}. \tag{2.40}$$

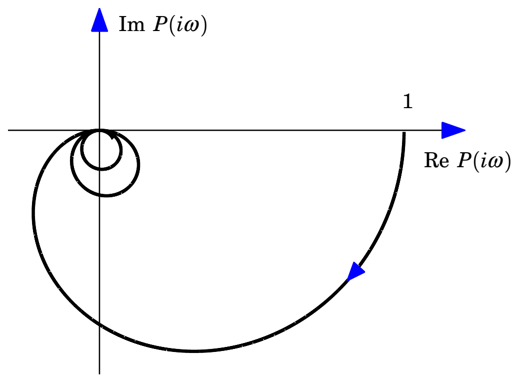


Figure 2.25 Nyquist plot of the transfer function $G(s)$ (2.38) of an ideal heat exchanger.

Table 2.2 Parameters of the exothermic continuous-time stirred tank reactor.

q	$0.002\ m^3/s$	T_f, T_c	300 K
V	$0.1\ m^3$	E/R	8750 K
ρ	$1000\ kg/m^3$	c_f	2 kmol/m^3
C_p	4 kJ/kgK	UA	50 W/K
k_0	$3\times10^8\ s^{-1}$	ΔH	-5×10^5 kJ/kmol

The first term on the right-hand side represents the mass flow rate and the second term represents the rate of removal of $\mathcal{A}$ through the reaction.

The energy balance can be written as

$$\frac{dT}{dt} = \frac{q}{V}(T_f - T) + k(T)\frac{-\Delta H}{\rho C_p}c + \frac{UA}{\rho V C_p}(T_c - T), \tag{2.41}$$

where ΔH [kJ/kmol] is the reaction heat, ρ [kg/m^3] the density of the species $\mathcal{A}$, C_p [kJ/kgK] specific heat, U [$J/min/K/m^2$] the heat transfer coefficient, A [m^2] the area, T_c [K] the cooleant temperature, and T_f [K] the feed temperature. The first term on the right-hand side represents the energy flow rate of the system, the second term represents the power generated by the reaction, and the last term represents the energy removal rate through cooling. Typical parameters are given in Table 2.2.

We will first analyse the steady-state solutions. In steady state it follows from (2.40) that

$$c = \frac{1}{1 + Vk(T)/q}c_f.$$

The power generated by the reaction is

$$P_g = \frac{k(T)}{1 + Vk(T)/q}(-\Delta H)c, \tag{2.42}$$

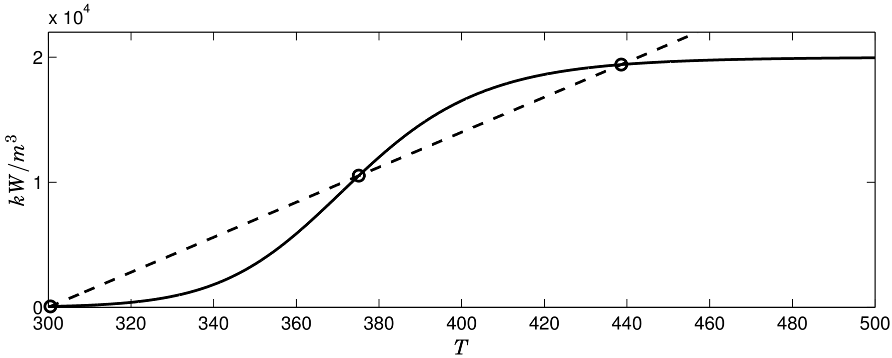


Figure 2.26 Steady-state heat generation rate (solid) and heat removal rate (dashed) as function of temperature. The equilibria are marked with ∘.

and the rate of removal of energy is

$$P_r = \frac{q\rho C_p}{V}(T - T_f) + UA(T - T_c). \tag{2.43}$$

Equating P_g and P_r gives an equation in one variable to determine the reaction temperature T. A graphical solution gives insight as illustrated in Figure 2.26, which shows P_g and P_r as functions of temperature. There are three equilibria where the curves intersect at $T = 300.5$, 375.1, and 438.6. The equilibrium at $T = 375.1$ is unstable because the rate of heat generation is larger than the rate heat removal if temperature is increased. The other equilibria are stable. Approximating the dynamics in the neighborhood of the unstable equilibrium gives the following linear model of the system

$$\begin{aligned}\frac{dx_1}{dt} &= -0.0422x_1 + 0.0013x_2\\ \frac{dx_2}{dt} &= 2.7746x_1 - 0.0064x_2 + 0.15u,\end{aligned} \tag{2.44}$$

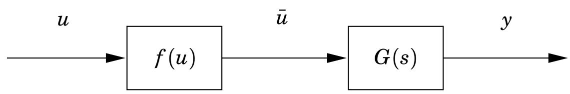
where $x_1 = c - c_0$, $x_2 = T - T_0$, and $u = T_c - T_{c_0}$ and c_0, T_0, and T_{c_0} are the equilibrium values. The transfer function is

$$P(s) = -\frac{0.15s + 0.0063}{s^2 + 0.048631s - 0.003359} = -0.15\frac{s + 0.04220}{(s + 0.08717)(s - 0.03854)}$$

The system has the pole $s = 0.03854$ in the right half plane.

Nonlinear Black Models

The static model discussed in Section 2.2 could be nonlinear. The dynamic models discussed so far have, however, been linear. Since nonlinearities are common in practice it is highly desirable to have nonlinear models. Valves, actuators, and sensors may be nonlinear; the process dynamics itself can also be a)



b)

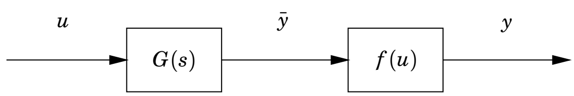


Figure 2.27 A Hammerstein model a), and a Wiener model b).

nonlinear. General models for nonlinear dynamics are complicated, and there are no good methods for designing PID controllers for such systems.

Fortunately, there are special classes of models that are well suited for PID control. A system may be represented as a combination of a static nonlinearity and a linear dynamical system. Such models are quite simple, and they are nicely adapted to PID control, but there are nonlinear systems that cannot be modeled well using this approach.

The nonlinearity can be before the linear part as shown in Figure 2.27a. This model is called a Hammerstein model. It is a good model for a system with a nonlinear actuator, for example, a nonlinear valve.

The nonlinearity can also be placed after the linear dynamical system. This gives a Wiener model, which is illustrated in the block diagram in Figure 2.27b. The Wiener model is a good representation for a system with a nonlinear sensor, for example, a pH electrode.

If the process is nonlinear the dynamics are varying with the operating conditions. Ideally, the controller should be tuned with respect to these variations. A conservative approach is to tune the controller for the worst case and accept degraded performance at other operating conditions. Another approach is to find a measurable variable that is well correlated with the process nonlinearity. Such a variable is called a scheduling variable. The controller is then tuned for a few values of the scheduling variable. Controller parameters for intermediate values may be obtained by interpolation. This approach to generating a nonlinear controller is called gain scheduling. It will be discussed in more detail in Section 9.3.

It is easy to compensate for the nonlinearity for a system that is described by a Wiener or a Hammerstein model by using a nonlinear controller composed of a PID controller and a static nonlinearity. The linear PID controller is designed as if the system was linear. When the process has a nonlinearity at the input we simply pass the control signal through the inverse of the nonlinearity. If the nonlinearity is at the output, as for the Wiener model, we simply pass the sensor signal through an inverse of the nonlinearity before feeding the measured signal to the controller. Many PID controllers have a facility to introduce a nonlinearity characterized as a piecewise linear function.

2.6 Models for Disturbances

So far, we have only discussed models of process dynamics. Disturbances are another important aspect of the control problem. In fact, without disturbances and process uncertainty there would be no need for feedback. There is a special branch of control, stochastic control theory, that deals explicitly with disturbances. This has had little impact on the tuning and design of PID controllers. For PID control, disturbances have mostly been considered indirectly, e.g., by introducing integral action. As our ambitions increase and we strive for control systems with improved performances it will be useful to consider disturbances explicitly. In this section, therefore, we will present some models that can be used for this purpose. Models for disturbances are useful for simulation, diagnostics, and performance evaluation.

The Nature of Disturbances

We distinguish between three types of disturbances, namely, set-point changes, load disturbances, and measurement noise. In process control, most control loops have set points that are constant over long periods of time with occasional changes. An appropriate model is therefore a piecewise constant signal. Set-point changes are typically known beforehand. Good response to set-point changes is the major issue in drive systems.

Load disturbances are disturbances that enter the control loop somewhere in the process and drive the system away from its desired operating point. Load disturbances typically have low frequency. Efficient reduction of load disturbances is a key issue in process control systems.

Measurement noise represents disturbances that distort the information about the process variables obtained from the sensors. Measurement noise is often a high-frequency disturbance. It is often attempted to filter the measured signals to reduce the measurement noise. Filtering does, however, add dynamics to the system.

The Character of Disturbances

One way to get a first estimate of the disturbances is to log the measured variable. The measured signal has contributions both from load disturbances and measurement noise. If there are large variations it is often useful to investigate the sensor to reduce some of the measurement noise. Filtering may also be useful. Filtering should be done in such a way that it does not impair control.

The process variations may have very different character. Some examples are given in Figure 2.28. The disturbances can be classified as pulses (a), steps (b), ramps (c), and periodic (d). It is useful to compute statistics such as mean values, variances, and maximum deviation. It is also useful to plot a histogram of the amplitude distribution of the disturbances.

Simple Models

It is useful to have simple models for disturbances for simulation and evaluation of control strategies. Models that are typically used are shown in Figure 2.28. The impulse is a mathematical idealization of a pulse whose duration

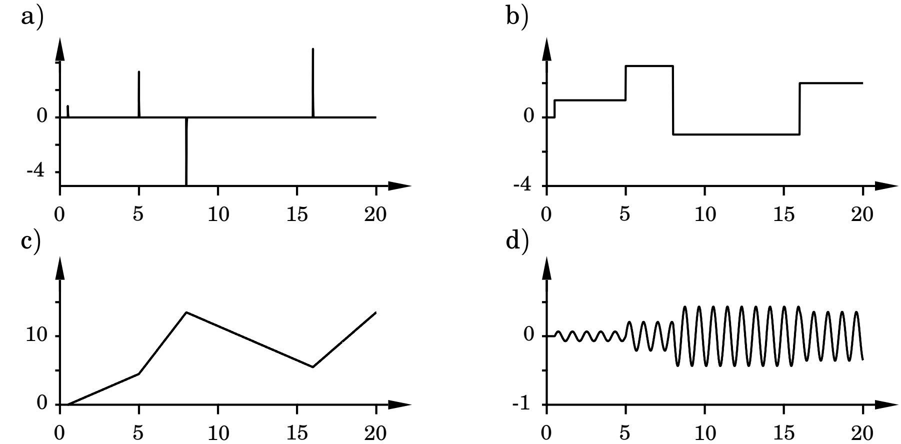


Figure 2.28 Different types of disturbances: a) impulses, b) steps, c) ramps, and d) sinusoids.

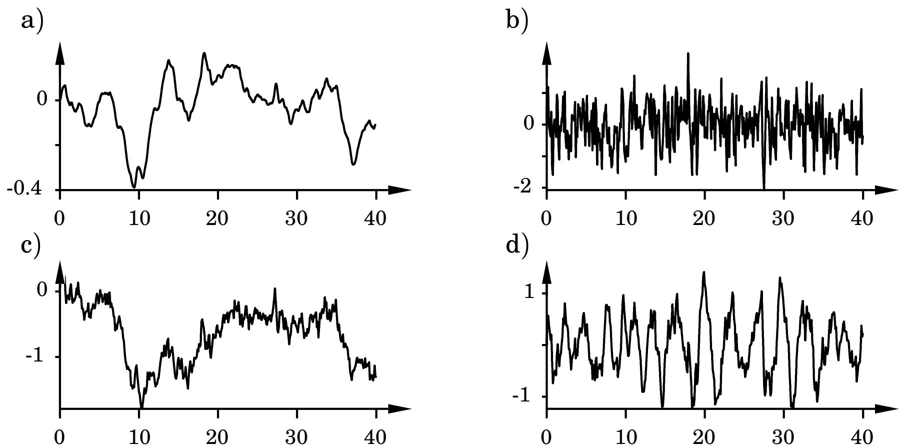


Figure 2.29 Examples of stochastic disturbances.

is short in comparison with the time scale. The signals are essentially deterministic. The only uncertain elements in the impulse, step, and ramp are the times when they start and the signal amplitude. The uncertain elements of the sinusoid are frequency, amplitude, and phase.

Random Fluctuations

Disturbances may also be more irregular as is shown in Figure 2.29. There are well developed concepts and techniques for dealing with random fluctuations that are described as stochastic processes. There are both time domain and frequency domain characterizations. In the frequency domain the random disturbances are characterized by the spectral density function $\phi(\omega)$. The variance

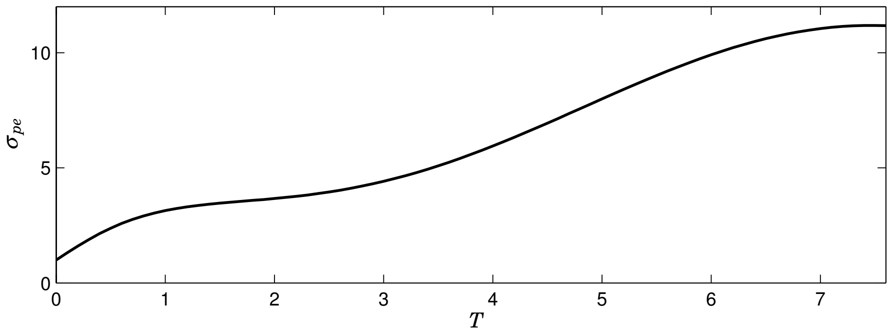


Figure 2.30 Prediction error σ_{pe} as a function of prediction time T_p.

of the signal is given by

$$\sigma^2 = \int_{-\infty}^{\infty} \phi(\omega) d\omega.$$

The spectral density tells how the variation of the signal is distributed on different frequencies. The value

$$2\phi(\omega)\Delta\omega$$

is the average energy in a narrow band of width $\Delta\omega$ centered around ω. A signal where $\phi(\omega)$ is constant is called white noise. Such a signal has its energy equally distributed on all frequencies.

There are efficient techniques to compute the spectral density of a given function. If the spectral density is known it is possible to evaluate how the variations in the process variable are influenced by different control strategies.

Prediction of Disturbances

When controlling important quality variables in a process it is often of interest to assess the improvements that can be achieved and to determine if a particular control strategy gives a performance that is close to the achievable limits. This can be done as follows. The process variable $y(t)$ is logged during normal operation with or without control. By analyzing the fluctuations it is possible to determine how accurately the process variable can be predicted T_p time units into the future based on present and past values of y. Let $\hat{y}(t + T_p|t)$ be the best prediction of $y(t+T_p)$ based on $y(\tau)$ for all $\tau < t$. By plotting the variance of the prediction error $y(t + T_p) - \hat{y}(t + T_p|t)$ as a function of the prediction time we obtain the curve shown in Figure 2.30. For large prediction times the prediction error is equal to the variance of the process variable, approximately $\sigma_{pe} = 12$ in the figure. The best control error that can be achieved is the prediction error at a prediction time T_p corresponding to the time delay of the process and the sampling time of the controller. This can be achieved with a so called minimum variance controller. See Section 8.6. The figure indicates that variances less that 5 can be obtained if T_p is less than 3.4. Further reductions are possible for smaller T_p, but variances less than 1 cannot be achieved even if T_p is very short. By comparing this with the actual variance we get an assessment of the achievable performance. This is discussed in more detail in Chapter 10. There is efficient software for computing the prediction error and its variance from process data.

2.7 How to Obtain the Models

In previous sections we have briefly mentioned how the models can be obtained. In this section we will give a more detailed discussion of methods for determining the models. There are two broad types of methods that can be used. One is physical modeling, and the other is modeling from data.

Physical modeling uses first principles to derive the equations that describe the system. The physical laws express conservation of mass, momentum, and energy. They are combined with constitutive equations that describe material properties. When deriving physical models a system is typically split into subsystems. Equations are derived for each subsystem, and the results are combined to obtain a model for the complete system. Simple examples were given in Section 2.3. Physical modeling is often very time consuming. There are often difficult decisions on suitable approximations. The models obtained can, however, be very useful since they have a sound physical basis. They also give considerable insight into the dependence of the model on the physical parameters. A simple way to start is to model dynamics as first-order systems where the time constants are the ratio of storage and flow.

Modeling from data is an experimental procedure. Data is generated by perturbing the input signal (the manipulated variable) and recording the system output. The experiment can also be performed under closed-loop conditions, for example, by perturbing the set point of a controller or the controller output. It is then attempted to find a model that fits the data well. There are several important issues to consider; selection of input signals, selection of a suitable model structure, parameter adjustments, and model validation. Ideally, the experimental conditions should be chosen to be as similar as possible to the intended use of the model. The parameter adjustment can be made manually for crude models or by using optimization techniques.

Static Models

Static models are easy to obtain by observing the relation between the input and the output in steady state. For stable, well-damped processes the relation can be obtained by setting the input to a constant value and observing the steady-state output. The procedure is then repeated for different values of the input until the full range is covered. For systems with integration it is convenient to use a controller to keep the output at a constant value. The set point of the controller is then changed so that the full signal range is covered. Effects of disturbances can be reduced by taking averages.

The Bump Test

The bump test is a simple procedure that is commonly used in process control. It is based on an experimental determination of the step response. To perform

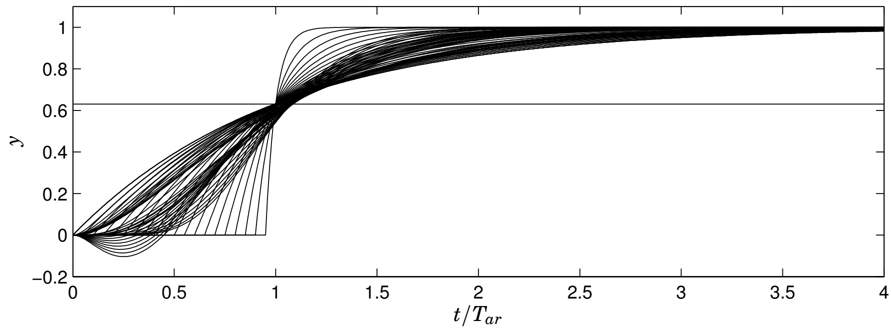


Figure 2.31 Step responses for a large batch of stable systems. The responses have been normalized to give the same average residence time.

the experiment the system is first brought to steady state. The manipulated variable is changed rapidly to a new constant value and the output is recorded. The measured data is scaled to correspond to a unit step. The change in the manipulated variable should be large in order to get a good signal-to-noise ratio but it should not be so large that the process behavior is not linear. The allowable magnitude is also limited by process operation. It is also useful to record the fluctuations in the measurement signal when the control signal is constant. This gives data about the process noise. It is good practice to repeat the experiment for different amplitudes of the input signal and at different operating conditions. This gives an indication of the signal ranges when the model is linear. It also indicates if the process changes with the operating conditions.

By inspection of the step response it is possible to make a crude classification of the dynamics of the system into the categories shown in Figure 2.2. A model with a few parameters is then fitted to the data.

The Average Residence Time

The average residence time is a simple way to characterize the response time of systems with essentially monotone step responses. Figure 2.31 step responses for a large batch of systems that are normalized to give the same average residence time. (The transfer functions for the systems are given in Section 7.1.) The figure shows that all step responses are close for $t = T_{ar}$. For all processes in the test batch we have $0.99 < T_{63}/T_{ar} < 1.08$. The average residence time can thus be estimated as the time T_{63} where the step response has reached 63 percent of its final value.

The FOTD model

The parameters of the FOTD model given by Equation 2.18 can be determined from a bump test as illustrated in Figure 2.32. The static gain K_p is simply determined from the steady-state values of the signals before and after the step change. The apparent time delay L is given by the point where the steepest tangent intersects the steady-state level before the step change. The average residence time $T_{ar} = T + L$ is determined as the time T_{63} where the step

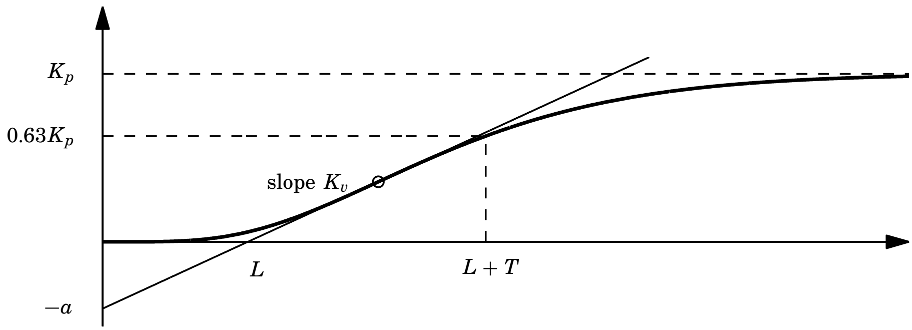


Figure 2.32 Unit step response of a process and a procedure used to determine the process parameters K_p, L, T, and K_v of an FOTD model. The point of largest slope is denoted by $\circ$.

response has reached 63 percent of its final steady-state value. This gives the correct results for the FOTD-model (see Figure 2.11) and approximate results for many other models (see Figures 2.13, 2.14, 2.16, and 2.31). The velocity gain K_v is the slope of the steepest tangent.

Similar methods can be used when the input signal is a pulse instead of a step. Pulses may be used when it is not permitted to use a step. This is common in medical and biological applications and is less common in process control. Ramp response analysis is common when analyzing servo drives and hydraulic systems.

The Integral and Time Delay Approximation

The model parameters of the model (2.28), which has the transfer function

$$G(s) = \frac{K_v}{s}\, e^{-sL} = \frac{a}{sL}\, e^{-sL}, \tag{2.45}$$

can also be determined from a bump test as indicated in Figure 2.32. The velocity constant K_v is the steepest slope of the step response, and the intersections of this tangent with the vertical and horizontal axes give a and L, respectively. The model given by Equation 2.45 is the basis for the Ziegler-Nichols tuning procedure discussed in Chapter 6.

The Doublet-Pulse Method

A variation of the bump test is to excite the process by a doubled pulse as is illustrated in Figure 2.33. The pulse amplitude a is chosen so that the response is well above the noise level, and the pulse width T_p is chosen a little longer than the time delay of the process. The maximum $y_{\max}$ and the minimum $y_{\min}$ and the times $t_{\max}$ and $t_{\min}$ when they occur are determined. Simple calculations show that for an FOTD system with the transfer function (2.10) we

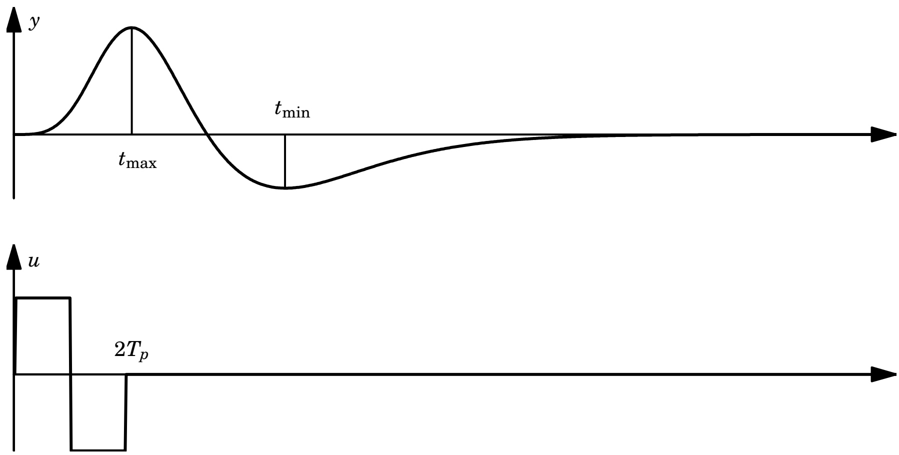


Figure 2.33 Determination of the parameters of an FOTD model by exciting the process by a doublet pulse.

have

$$\begin{aligned} y_{\max} &= aK_p\big(1 - e^{-T_p/T}\big) \\ y_{\min} &= -aK_p\big(1 - e^{-T_p/T}\big)^2 \\ t_{\max} &= L + T_p \\ t_{\min} &= L + 2T_p. \end{aligned}$$

It follows from these equations that

$$\begin{aligned} \frac{y_{\min}}{y_{\max}} &= -1 + e^{-T_p/T} \\ \frac{y_{\max}^2}{y_{\min}} &= -aK_p, \end{aligned}$$

and we get the following simple equations for the parameters of the model

$$\begin{aligned} K &= -\frac{y_{\max}^2}{a y_{\min}} \\ T &= \frac{T_p}{\log(1 + y_{\max}/y_{\min})} \\ L &= t_{\max} - T_p \\ L &= t_{\min} - 2T_p. \end{aligned} \tag{2.46}$$

The fact that the time delay L can be estimated in two ways can be used to asses if a process can be modeled by an FOTD model.

The selection of the pulse time T_p can be determined automatically, for example, as the time when the output has changed a specified amount. The method can be applied to SOTD models, but the formulas are more complicated.

The main advantages of using a doublet pulse is that the process output returns to its original value after the perturbation, and the time required to

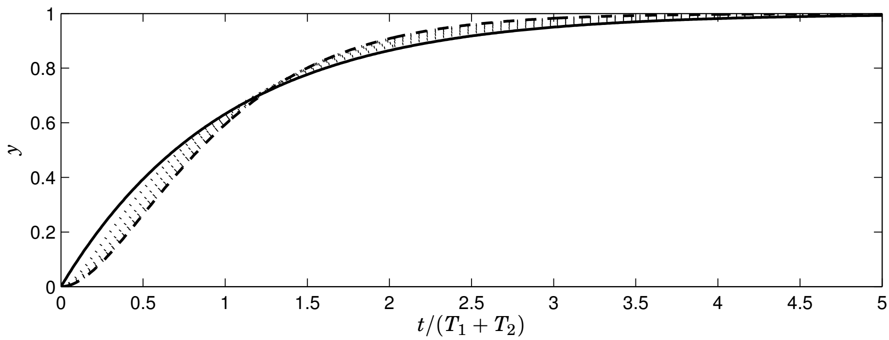


Figure 2.34 Normalized step responses for the system (2.48) for $T_2/T_1 = 0.1, \ldots, 1$.

determine the dynamics is short because it is not necessary to wait for steady state as for the bump test. The disadvantages of the method are that it is difficult to determine times when the extrema occurs accurately, and the estimate of the gain is poor because the excitation of the pulse is mainly in the high-frequency regime. Another disadvantage is that the method cannot be applied to oscillatory systems.

The SOTD Model

The model

$$G(s) = \frac{K_p}{(1+sT_1)(1+sT_2)} e^{-sL_1}, \tag{2.47}$$

which is a natural generalization of the FOTD model (2.18), is called the second-order model with time delay or SOTD model. Without loss of generality it can be assumed that $T_2 \leq T_1$. The step response of the system (2.47) is

$$y(t) = \begin{cases} K_p\left(1 - \dfrac{T_1}{T_1 - T_2} e^{-(t-L_1)/T_1} - \dfrac{T_2}{T_2 - T_1} e^{-(t-L_1)/T_2}\right) & \text{if } T_1 \neq T_2 \\ K_p\left(1 - e^{-(t-L_1)/T_1} - \dfrac{t}{T_1} e^{-(t-L_1)/T_1}\right) & \text{if } T_1 = T_2. \end{cases} \tag{2.48}$$

The normalized step responses for different ratios T_2/T_1 are shown in Figure 2.34. The responses have been normalized so that all systems have the same average residence time. All step responses are quite close, and they are almost identical for $t/(T_1 + T_2) \approx 1.3$. Since the separation of the curves is so small it is difficult to determine the parameters T_1 and T_2 robustly from the step response, particularly if there is a small amount of noise. Other inputs that excite the system better are necessary to determine the parameters reliably. The figure shows that it would be easier to determine the parameters based on an impulse response, which could be obtained by differentiating the step response.

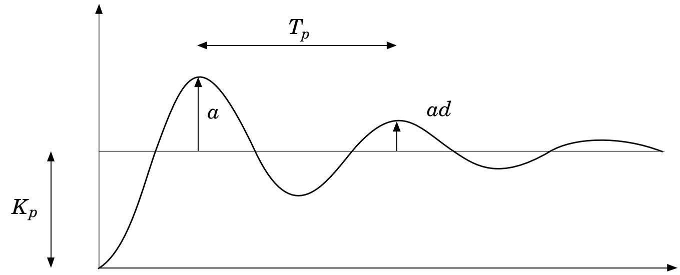


Figure 2.35 Graphical determination of mathematical models for systems with an oscillatory step response.

An Oscillatory System

The model (2.18) cannot model systems with oscillatory responses. A simple model for such systems is given by the transfer function (2.27), which has three parameters: the static gain K_p, the undamped natural frequency $1/T$, and the relative damping ζ. These parameters can be determined approximately from the step response as indicated in Figure 2.35. Parameters T and ζ are related to time period T_p and decay ratio d as follows.

$$d = e^{-2\zeta\pi/\sqrt{1-\zeta^2}} \qquad T_p = \frac{2\pi T}{\sqrt{1-\zeta^2}} \tag{2.49}$$

or

$$\zeta = \frac{1}{\sqrt{1+(2\pi/\log d)^2}} \qquad T = \frac{\sqrt{1-\zeta^2}}{2\pi}T_p. \tag{2.50}$$

The accuracy of the model is limited by the limited excitation obtained with a step or a pulse. Measurement errors and difficulty in obtaining steady state are other factors that limit the accuracy. Some improvements can be made by using optimization for fitting the parameters. Typically, it is difficult to determine more than three parameters from a step response unless the experimental conditions are exceptional.

Frequency Response

In frequency response analysis a sinusoidal signal is instead introduced, and the steady-state response is analyzed. An advantage with frequency response analysis is that very accurate measurements can be made by using correlation techniques. The long experimental times is a drawback.

It is also possible to introduce an arbitrary signal as a perturbation. The frequency response can be obtained as the ratio of the Fourier transforms of the output and the input signals. It is also possible to fit the parameters of a model with given structure to the data.

A nice feature of using signals other than steps is that it is possible to make a trade-off between signal amplitude and duration.

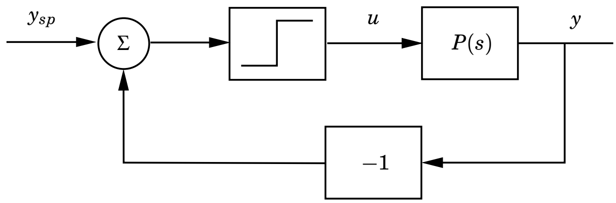


Figure 2.36 Block diagram of a process with relay feedback.

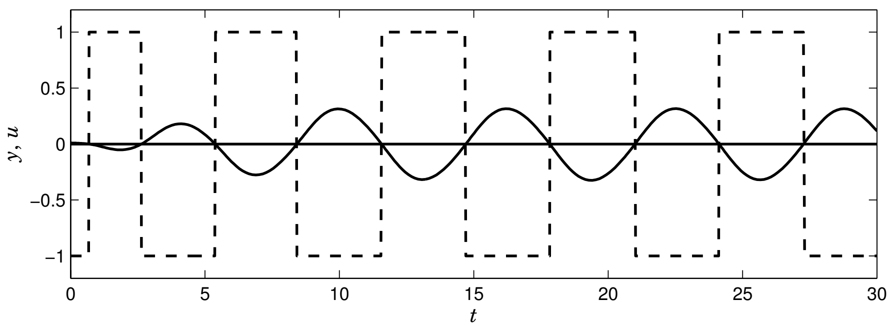


Figure 2.37 Relay output u (dashed) and process output y (solid) for a system under relay feedback.

Relay Feedback

There is a very special technique that is particularly suited to determine ω_{180} and K_{180}. This has been used very effectively for tuning PID controllers. The idea is the observation that it is possible to create an oscillation with the ultimate frequency automatically by using relay feedback.

To make the experiment the system is connected in a feedback loop with a relay function as shown in Figure 2.36. For many systems there will then be an oscillation (as shown in Figure 2.37) where the control signal is a square wave and the process output is close to a sinusoid. Notice that the process input and output have opposite phase.

To explain how the system works, assume that the relay output is expanded in a Fourier series and that the process attenuates higher harmonics effectively. It is then sufficient to consider the first harmonic component of the input only. The input and the output then have opposite phase, which means that the frequency of the oscillation is equal to ω_{180}. If d is the relay amplitude, the first harmonic of the square wave input has amplitude $4d/\pi$. Let a be the amplitude of the process output. The process gain at ω_{180} is then given by

$$K_{180} = \frac{\pi a}{4d}. \tag{2.51}$$

Notice that the relay experiment is easily automated. Since the amplitude of the oscillation is proportional to the relay output, it is easy to control it by adjusting the relay output. Also notice in Figure 2.37 that a stable oscillation is established very quickly. The amplitude and the period can be determined after about 20 s only, in spite of the fact that the system is started so far from the equilibrium that it takes about 8 s to reach the correct level. The average residence time of the system is 12 s, which means that it would take about 40 s for a step response to reach steady state.

The SOTD Model—Combined Step and Frequency Response

It was mentioned previously that the parameters of the SOTD model cannot be determined reliably from step response data. Good estimates can, however, be obtained by combining step and frequency response data. The idea is that the step response gives K_p and T_{63} and the frequency response method gives the ultimate frequency $\omega_u = \omega_{180}$ and the ultimate gain $K_u = 1/K_{180}$. This gives the equations

$$\begin{aligned} K_p^2 K_u^2 &= (1+\omega_u^2 T_1^2)(1+\omega_u^2 T_2^2) \\ \pi &= \arctan \omega_u T_1 + \arctan \omega_u T_2 + \omega_u L_1. \end{aligned} \tag{2.52}$$

Combined with the data K_p and T_{63} the parameters are then given by Equations 2.48 and 2.52 which gives four equations for the four unknown.

$$\begin{aligned} 0 &= \begin{cases} 0.37 - \dfrac{T_1}{T_1 - T_2} e^{-(T_{63}-L_1)/T_1} - \dfrac{T_2}{T_2 - T_1} e^{-(T_{63}-L_1)/T_2} & \text{if } T_1 \neq T_2, \\ 1 - e^{-(T_{63}-L_1)/T_1} - \dfrac{T_{63}}{T_1} e^{-(T_{63}-L_1)/T_1} - 0.63 & \text{if } T_1 = T_2 \end{cases} \\ 0 &= (1+\omega_u^2 T_1^2)(1+\omega_u^2 T_2^2) - K_p^2 K_u^2 \\ 0 &= \arctan \omega_u T_1 + \arctan \omega_u T_2 + \omega_u L_1 - \pi. \end{aligned} \tag{2.53}$$

These equations can be solved iteratively, but this is complicated since we have to take care of the special cases when the parameters T_1 and T_2 are equal or zero.

An alternative method is to iterate the ratio $a = T_2/T_1$ until the equations match. Parameter K_p is determined as the static gain of the step response. The equation (2.52) for the ultimate gain then becomes

$$(1+\omega_u^2 T_1^2)(1+a^2\omega_u^2 T_1^2) = K_p^2 K_u^2.$$

This equation has the solution

$$T_1 = \frac{1}{a\omega_u\sqrt{2}} \sqrt{\sqrt{4a^2 K_p^2 K_u^2 + (1-a^2)^2} - 1 - a^2}.$$

The parameters T_2 and L_1 are then given by

$$\begin{aligned} T_2 &= aT_1 \\ L_1 &= \frac{\pi - \arctan \omega_u T_1 - \arctan \omega_u T_2}{\omega_u}. \end{aligned}$$

The step response given by (2.48) can then be computed as a function of a, and the parameter a can be iterated to match the value of T_{63}.

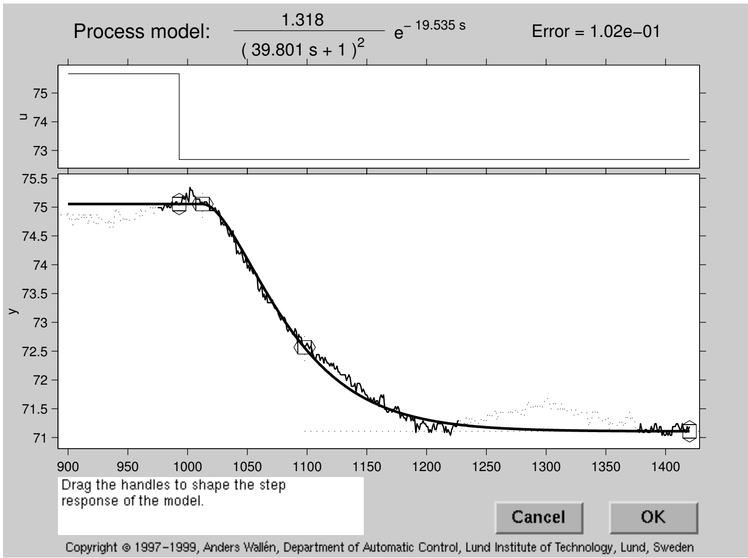


Figure 2.38 Computer screen from a tool for process modeling. From [Wallén, 2000]

Modeling Tools

There are several modeling tools that are very useful. They make it possible to enter process data in the form of sequences of input-output data from bump tests or other process experiments. Models of different structure can be selected, and their parameters can be fitted to the data using some optimization procedure. The tools also permit selection of parts of data sequences used in the analysis.

Figure 2.38 shows the computer screen for a particular system. A model structure can be chosen from a menu. When data has been entered a preliminary model can be fitted by manually dragging the handles shown in the figure. The handles represent the start of the step, the initial level, the final level, and the time when the response has reached 63 percent of its final value. The model parameters are displayed. Optimization can then be used to improve the fit.

The particular tool illustrated in the figure also allows use of a nonlinear model as illustrated in Figure 2.39. In this case a static model is first fitted to input-output data obtained from a static experiment. A dynamic model is then fitted as indicated in Figure 2.39. Both Wiener and Hammerstein models are tried to see which gives the best fit. The particular example is from a tank system where the outflow is a nonlinear function of level. In this case the Wiener model gives the best fit because the nonlinearity appears at the system output. Notice in Figure 2.39 that the input steps are of equal size, but the magnitude of the output response changes significantly. This data cannot be well matched by a linear model.

The interactive tools give a very good feel for the relations between the parameters and the response and the sensitivity of the parameters. It is also very effective to combine simple manual fits with numeric optimization. Most tools

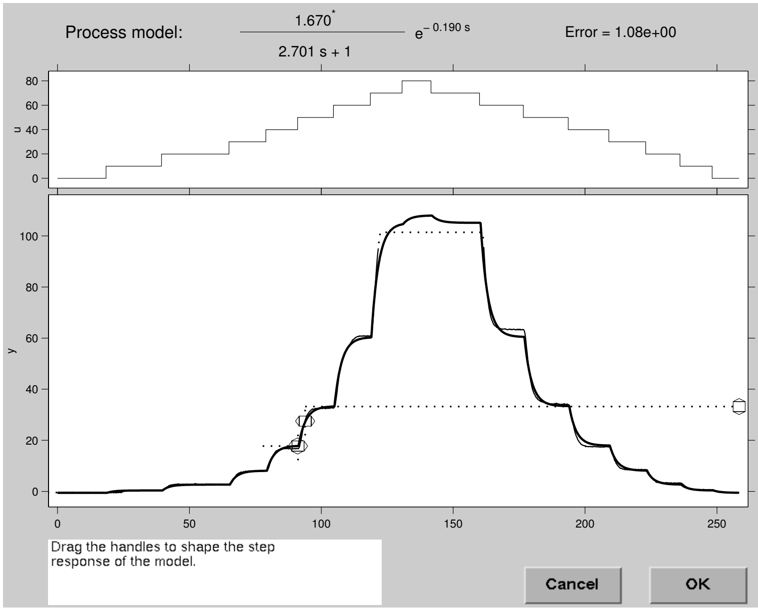


Figure 2.39 Illustrates computer-based nonlinear modeling. From [Wallén, 2000]

permit fitting a simple bump test, but there are also tools that permit general input signals. They also make it possible to determine noise characteristics and the prediction curve shown in Figure 2.30.

2.8 Model Reduction

Many methods for tuning PID controllers are based on simple models of process dynamics. To use such methods it is necessary to have methods for simplifying a complicated model. A typical case is when a model is obtained by combining models for subsystems. To find suitable approximations it is necessary to specify the purpose of the model. For tuning PID controllers this can be done by specifying the frequency range of interest. This can be done simply by specifying the highest frequency ω^* where the model is valid. For PI control the frequency ω^* is about ω_{145}, the frequency where the phase lag of the process is 145°. The reason for this is that a PI controller always has a phase lag. For a PID controller, which can provide phase lead, the frequency ω^* can be chosen as ω_{180}.

Model reduction starts with a model represented by the transfer function $G(s)$. The transfer function is first factored as

$$G(s) = G_l(s)\frac{1}{1+sT_s}G_h(s). \tag{2.54}$$

The low frequency factor $G_l(s)$ has all its poles and zeros and time delays at frequencies around ω^* or at lower frequencies. The high-frequency factor $G_h(s)$

has dynamics at frequencies higher than ω^*. The time constant T_s represents an intermediate pole. The factorization can always be done in such a way that the high frequency factor $G_h(s)$ has the property $G_h(0) = 1$.

For design of PID controllers the model (2.54) will be simplified to

$$
\begin{aligned}
G(s) &= \frac{K_p}{1+sT} e^{-sL} \\
G(s) &= \frac{K_p}{(1+sT_1)(1+sT_2)} e^{-sL}.
\end{aligned}
\tag{2.55}
$$

The reason for these choices is that there are methods for designing PID controllers for models of this type. These models are particularly suitable for typical process control problems where the dynamics have essentially monotone step responses.

The Low-Frequency Factor

The low-frequency factor will normally only contain one or two modes. If the system has multiple poles they can be approximated by the transfer function

$$G_l(s) = \frac{K_p}{1+sT_e} e^{-sL_e}.$$

where T_e and L_e are obtained from Table 2.1. In this way we obtain a low-frequency factor of first or second order, which is required for PID control. If the model is more complex it is necessary to reduce ω^* or to use a more complex controller.

Approximation of Fast Modes

There are several ways to approximate the fast modes. A simple way is to characterize the high-frequency part by its average residence time T_{arh}. This is illustrated by the following example.

EXAMPLE 2.12—APPROXIMATION OF FAST MODES
Consider a system where the high-frequency factor is

$$G_h(s) = \frac{(1+sT_1)(1+sT_2)}{(1+sT_3)(1+sT_4)(1+sT_5)} e^{-sL}.$$

This system has the average residence time

$$T_{arh} = T_3 + T_4 + T_5 + L - T_1 - T_2.$$

□

Compare this with Section 2.4, which shows how average residence times are computed. When using digital control half the sampling period should also be added to T_{arh}.

Skogestad's Half Rule

Having simplified the low- and high-frequency factors we have obtained a low-frequency factor of the form given by (2.55) or (2.58) and a characterization of the high-frequency factor by its average residence time T_{arh}. Skogestad has suggested that the intermediate time constant T_s in (2.54) is approximated by adding $T_s/2$ to the time delay of the model and $T_s/2$ to its time constant. The reduced model then becomes

$$
\begin{aligned}
G(s) &= \frac{K_p}{1+s(T+T_s/2)} e^{-s(L+T_{arh}+T_s/2)} \\
G(s) &= \frac{K_p}{(1+sT_1)(1+s(T_2+T_{arh}/2))} e^{-s(L+T_{arh}/2)}.
\end{aligned} \tag{2.56}
$$

The model error is characterized by $T_{arh} + T_s/2$, which means that it must be required that $\omega^*(T_{arh} + T_s/2)$ is sufficiently small. A reasonable value is that it is less than 0.1 or 0.2, which means that the neglected dynamics has a phase lag of 6 to 12 degrees.

Approximating Slow Modes by Integrators

Modes that are much slower than ω^* can be approximated by integrators. For example, if $\omega^* T$ or $\omega^* T_1$ are larger than 5 to 10, the model (2.56) can be approximated by

$$
\begin{aligned}
G(s) &= \frac{K_p}{1+s(T+T_{arh}/2)} e^{-s(L+T_{arh}/2)} \approx \frac{K_p}{s(T+T_{arh}/2)} e^{-s(L+T_{arh}/2)} \\
G(s) &= \frac{K_p}{(1+sT_1)(1+s(T_2+T_{arh}/2))} e^{-s(L+T_{arh}/2)} \\
&\approx \frac{K_p}{sT_1(1+s(T_2+T_{arh}/2))} e^{-s(L+T_{arh}/2}.
\end{aligned} \tag{2.57}
$$

Another Model Representation

For some design techniques it is desirable to have models of the form

$$
\begin{aligned}
G(s) &= \frac{b}{s+a} \\
G(s) &= \frac{b_1 s + b_2}{s^2 + a_1 s + a_2},
\end{aligned} \tag{2.58}
$$

which do not have any time delays. These forms can also be used for oscillatory systems. The models given by (2.55) can be converted to the form (2.58) by using the approximation

$$
e^{-sT} \approx \frac{1 - sT/2}{1 + sT/2}. \tag{2.59}
$$

Time delays and zeros in the right half plane are the features of a system that ultimately limits the achievable performance. These properties are preserved by the above approximation.

Examples

Model reduction will now be illustrated with a few examples.

EXAMPLE 2.13—MODEL REDUCTION
Consider a system described by the transfer function

$$G(s) = \frac{K_p}{(1+s)(1+0.1s)(1+0.01s)(1+0.001s)}. \tag{2.60}$$

We have $\omega_{90} = 3$ and $\omega_{180} = 31.6$, which gives the ranges of ω^*. Let us first consider model reduction for a design with $\omega^* = 3$. The low-frequency factor is

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

the mid-frequency factor is $T_s = 0.1$, and the average residence time of the high-frequency part is $T_{arh} = 0.011$. Skogestad's half rule gives the model

$$\tilde{G}(s) = \frac{K_p}{1+1.05s}e^{-0.061s}.$$

Requiring that $\omega^*(T_{arh} + T_s/2) < 0.2$ we find that the model can be used for designs with $\omega^* < 3.3$.

For $\omega^* = 31.6$ the low-frequency factor becomes

$$G_l(s) = \frac{K_p}{(1+s)(1+0.1s)},$$

the mid-frequency time constant is $T_s = 0.01$, and the average residence time of the high-frequency part is then $T_{arh} = 0.001$. The half rule gives the model

$$\tilde{G}(s) = \frac{K_p}{(1+s)(1+0.105s)}e^{-0.006s}.$$

Requiring that $\omega^*(T_{arh} + T_s/2) < 0.2$ we find that the model can be used for designs with $\omega^* < 33$.

The approximations are illustrated in Figure 2.40. □

A Warning

The fact that the step responses of two systems are similar does not imply that the systems are similar under feedback control. This is illustrated by the following example.

EXAMPLE 2.14—SIMILAR OPEN LOOP – DIFFERENT CLOSED LOOP
Systems with the transfer functions

$$G_1(s) = \frac{100}{s+1}, \qquad G_2(s) = \frac{100}{(s+1)(1+0.025s)^2}$$

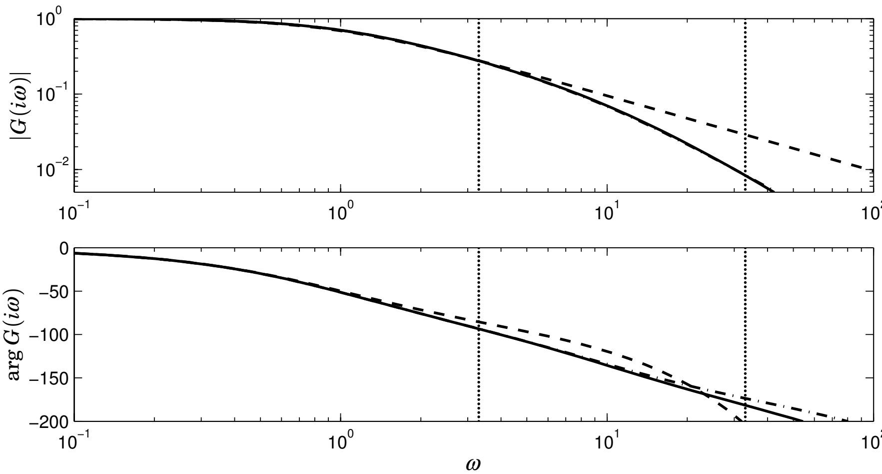


Figure 2.40 Bode plots of the original system (solid line) and the approximations for frequencies $\omega^* < 3.3$ (dashed) and $\omega^* < 33$ (dash-dotted).

have very similar open-loop responses as illustrated in Figure 2.41. The differences between the step responses are barely noticeable in the figure. The closed-loop systems obtained with unit feedback have the transfer functions

$$G_{1cl} = \frac{100}{s+101}, \qquad G_{2cl} = \frac{100}{(1+0.01192s)(1-0.001519s+0.0005193s^2)}.$$

The closed-loop systems are very different since the system P_{2cl} is unstable.

□

It is also possible to have the opposite situation, namely, systems whose closed-loop behavior are very similar even if their open-loop behavior are very different.

EXAMPLE 2.15—DIFFERENT OPEN LOOP – SIMILAR CLOSED LOOP
The systems with the transfer functions

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

have very different open-loop properties because one system is unstable and the other is stable. The closed-loop systems obtained with unit feedback are, however,

$$P_{1cl}(s) = \frac{100}{s+101} \qquad P_{2cl}(s) = \frac{100}{s+99},$$

which are very close. □

The paradoxes in the examples can be resolved by considering the frequency ranges that are relevant for closed-loop control. In Example 2.14 the closed-loop system bandwidth of relevance is about 100 rad/s. This corresponds to time

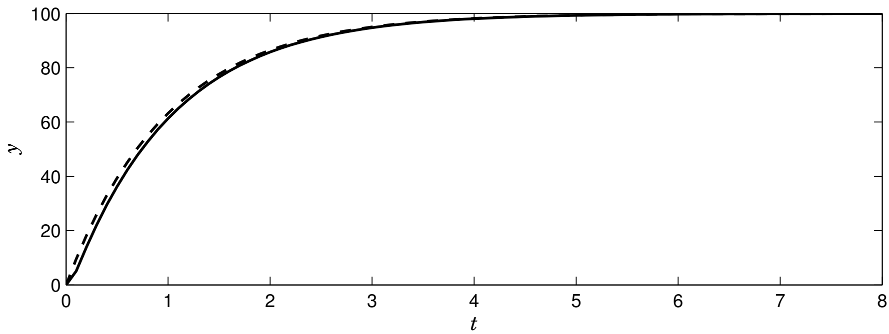


Figure 2.41 Step responses for systems with the transfer functions $G_1(s) = 100/(s+1)$ (dashed) and $G_2(s) = 100/((s+1)(1+0.025s)^2)$ (solid).

constants of about 0.01 s. A closer examination shows that the step responses in Figure 2.41 are indeed quite different at that time scale even if the general appearance of the step responses are very similar. In Example 2.15 the closed-loop bandwidth is also about 100 rad/s, which corresponds to a time scale of 0.01 s. At that time scale the open-loop systems are very similar even if one model is stable and the other unstable. It is a good rule to be aware of the relevant frequency ranges and to analyze the Bode plots. This is one of the main reasons for using frequency response.

2.9 Summary

Modeling is an important aspect of controller tuning. The models we need should describe how the process reacts to control signals. They should also describe the properties of the disturbances that enter the system. Most work on tuning of PID controllers has focused on the process dynamics, which is also reflected in the presentation in this chapter.

A number of methods for determining the dynamics of a process have been presented in this chapter. Some are very simple: they are based on a direct measurement of the step response and simple graphical constructions. Others are based on the frequency response. It has been shown that very useful information can be generated from relay feedback experiments. Such experiments are particularly useful because the process is brought into self-oscillation at the ultimate frequency, which is of considerable interest for design of controllers.

The simple methods are useful in field work when a controller has to be tuned and few tools are available. The methods are also useful to provide understanding as well as being references when more complicated methods are assessed. We have also presented more complicated methods that require significant computations.

Models of different complexity have been presented. Many models were characterized by a few parameters. Such models are useful for many purposes and are discussed in Chapter 6. When using such models it should be kept in mind that they are approximations.

When deriving the models we also introduced two dimension-free quantities, the normalized time delay τ and the gain ratio κ. These parameters make it possible to make a crude assessment of the difficulty of controlling the process. Processes with small values of κ and τ are easy to control. The difficulty increases as the values approach 1. Tuning rules based on τ and κ are provided in Chapter 7.

To summarize: When deriving a simple model to be used for PID controller tuning, it is important to ensure that the model describes the process well for the typical input signals obtained during the process operations. The amplitude and frequency distribution of the signal is of importance. Model accuracy may be poor if the process is nonlinear or time varying. Control quality can be improved by gain scheduling or adaptive control. It is also important to know what kind of disturbances are acting on the system and which limitation they impose.

2.10 Notes and References

Early efforts in modeling using differential equations were made independently by [Maxwell, 1868] and [Vyshnegradskii, 1876] in connection with analysis of engines with centrifugal governors. The idea of modeling a process by its reaction curve (step response) emerged in the 1930s. The reaction curve was approximated by an FOTD model (2.18) in [Callender *et al.*, 1936]. The reaction curve was also used in [Ziegler and Nichols, 1942]. Frequency response arguments were used in [Ivanoff, 1934] who investigated a temperature-control loop using the model given by (2.32). Frequency response was also used by [Ziegler and Nichols, 1942]. An early reference to the notion of block diagram is found in [Mason and Philbrick, 1940].

Process modeling is a key element in understanding and solving a control problem. Good presentations of modeling are found in standard textbooks on control, such as [Eckman, 1945; Buckley, 1964; Cannon, 1967; Smith, 1972; Luyben, 1990; Shearer and Kulakowski, 1990]. The books [Oquinnaike and Ray, 1994; Marlin, 2000; Bequette, 2003; Rawlings and Ekerdt, 2002; Seborg *et al.*, 2004] are of particular interest for process control. These books have much material on many different modeling techniques. Similar presentations are given in [Gille *et al.*, 1959; Harriott, 1964; Oppelt, 1964; Takahashi *et al.*, 1972; Deshpande and Ash, 1981; Shinskey, 1996; Stephanopoulos, 1984; Hägglund, 1991]. The books [Tucker and Wills, 1960] and [Lloyd and Anderson, 1971] are written by practitioners in control companies. There are also books that specialize in modeling for control system design; see [Wellstead, 1979; Nicholson, 1980; Nicholson, 1981; Close and Frederick, 1993].

By the mid-1950s frequency response was very well established as manifested by a symposium organized as part of the annual meeting of the American Society of Mechanical Engineering in 1953. The proceedings of the symposium were published in the book [Oldenburg, 1956]. A nice overview of step and frequency response methods is given in the paper [Rake, 1980]. Additional details are given in [Strejc, 1959; Anderssen and White, 1971; Anderssen and White, 1970]. The doublet method is discussed in [Shinskey, 1994], and the method of moments is described in [Gibilaro and Lees, 1969].

The relay method is introduced in [Åström and Hägglund, 1984b], and it is elaborated in [Hägglund and Åström, 1991; Schei, 1992; Hang and Åström, 2002]. The describing function method is well documented in [Atherton, 1975] and [Gelb and Velde, 1968]. A method to estimate what is today called an ARX model was developed in [Åström and Bohlin, 1965] and applied to modeling and control of paper machines in [Åström, 1967]. There are many books on parameter estimation: [Ljung and Söderström, 1983; Ljung, 1998; Söderström and Stoica, 1989; Bohlin, 1991; Johansson, 1993]. Many useful practical aspects on system identification are given in [Isermann, 1980].

Modeling has been greatly enhanced by simulation. The first simulation of a control system with PID control was made at the University of Manchester using a copy of the differential analyzer developed by Vannevar Bush; see [Callender *et al.*, 1936]. The differential analyser was also used in [Ziegler and Nichols, 1942] to develop tuning rules. Pneumatic simulators built from components of pneumatic controllers were used early by equipment manufacturers. The first electronic analog computer developed by Philbrick had a major impact, and the use of simulation increased drastically. The rapid development of digital computing has made it possible for every engineer to have simulation tools on his lap; see [Åström *et al.*, 1998]. Many of the simulation programs used today mimic the diagrams used to program early analog computers in the 1950s. There are major efforts underway to combine experiences of process modeling with advances in computing science to develop a new generation of languages and tools for process modeling; see [Elmqvist *et al.*, 1998] and [Tiller, 2001].

There are many methods for model reduction. Early work was reported in [Ziegler and Nichols, 1943]. A nice survey is found in [Glover, 1990]. One method that is geared to PID control is presented in [Fröhr and Orttenburger, 1982]. The half-rule was developed in [Skogestad, 2003] as a simple method that works well for the purpose of tuning PID controllers.