Analysis of Linear Continuous Systems

Transient-Response Analysis and Steady-State Error Analysis

The first step in analyzing a control system was to derive a mathematical model of the system. Once such model is obtained, various methods are available for the analysis of system performance.

In analyzing and designing control systems we must have a basis of comparison of performance of various control systems. This basis may be set up by specifying particular test input signals and by comparing the responses of various systems to these signals. The use of test signals can be justified because of a correlation existing between the response characteristics of a system to a typical test input signal and the capability of the system to cope with the actual input signal.

The aim of analysis is the study of system's behavior in transient and steady-state when the model of the system and the input signals are known, [1], [2].

Typical test signals

The commonly used test input signals are those of step functions, ramp functions, impulse functions, sinusoidal functions and the like. With these test signals, mathematical and experimental analysis of control systems can be carried out easily since the signals are very simple functions of time.

If the input of a control system are gradually changing functions of time, then a ramp function of time may be a good test signal. Similarly, if a system is subjected to sudden disturbances a step function may be a good test signal, and for a system subjected to shock inputs, an impulse function may be a test. Once a control system is designed on the basis of test signals, the performance of the system in response to actual inputs is generally satisfactory.

The use of such test signals enables one to compare the performance of all systems on the same basis.

Transient response and steady-state response

The time response of a system consists of two parts: the transient and the steady-state response, as shown in Figure 1. By transient response we mean that which goes from

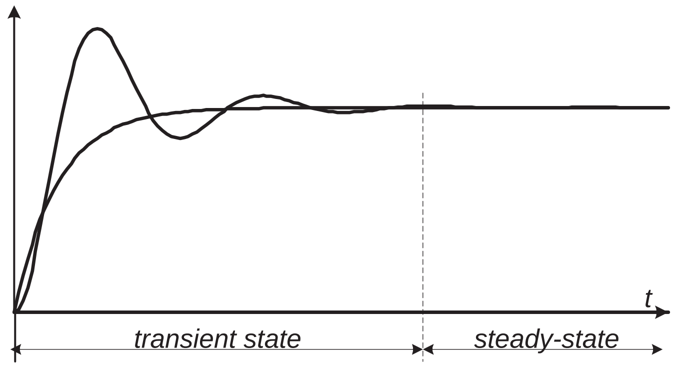

Figure 1: Transient and steady-state response

the initial to the final state. By steady-state response we mean the manner in which the system output behaves as t approaches infinity.

Absolute stability, relative stability and steady-state error

In designing a control system, we must be able to predict the dynamic behavior of the system from a knowledge of the components. The most important characteristics of the dynamic behavior of a control system is absolute stability, that is, whether the system is stable or unstable. A control system is in equilibrium if, in the absence of any disturbance or input, the output stays in the same state. A linear time-invariant control system is stable if the output eventually comes back to its equilibrium state when the system is subjected to a disturbance. A linear time-invariant system is unstable if either oscillation of the output continues forever or the output diverges from the bound from its equilibrium state when the system is subjected to a disturbance. Actually, the output of a physical system may increase to a certain extent but may be limited by mechanical "stops" or the system may breakdown or become nonlinear after the output exceeds a certain magnitude so that the linear differential equation no longer apply. We shall discuss the stability problem in another chapter.

Important system behavior to which we must give careful consideration includes relative stability and steady-state error. Since a physical control system involves energy storage, the output of the system, when subjected to an input, cannot follow the input immediately but exhibits a transient response before a steady state can be reached. The transient response of a control system often exhibits damped oscillations before reaching a steady state. If the output of a system at steady state does not exactly agree with the input, the system is said to have *steady-state error*. This error is indicative of the accuracy of the system. In analyzing a control system, we must examine transient-response behavior, such as the time required to reach a new steady state and the value of the error while following an input signal, as well as the steady-state behavior.

Errors in control systems can be attributed to many factors. Changes in the reference input will cause unavoidable errors during transient periods and may also cause steady-state errors. Imperfections in the system components, such as static friction, aging or deterioration will cause errors at steady state. In this chapter we shall not discuss errors due to imperfections of the system components. Rather, we shall investigate a type of steady-state error that is caused by the incapability of the system to follow particular types of inputs.

Usually the system is decomposed in simple elements of at most second order, that simplifies the method of analysis and also we can know the contribution of each element to the system behavior.

If the simple elements do not accumulate energy (they don't have inertia or time constants) they are called *ideal elements*, otherwise they are *real elements.*

The number of time constants at the denominator of the transfer function define the *order* of the element.

The behavior of simple elements can be studied using some characteristic parameters:

- Time constant, T
- Time delay constant, T_m
- Damping factor ζ
- Natural frequency ω_n
- Gain constant, K

The main steps in system analysis are:

- Write the equation or system of equations that define the relationship between the input and output signals. In the case of linear elements these are differential equations.
- Determine the transfer function or transfer matrix and identification of main parameters: gain factors, damping factors, natural frequencies, time constants etc.
- Determine the output signal for a test input signal (step, ramp, impulse, sine etc.)
- Graphical or analytical study of each element behavior emphasizing the influence of system parameters.

First-order systems

Consider a first-order system shown in Figure 2. Physically, this system may represent an RC circuit, a thermal system, or the like. The input-output relationship is given by:

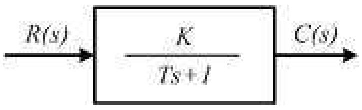

Figure 2: First-order system

$$\frac{C(s)}{R(s)} = \frac{K}{Ts+1} \tag{1}$$

Example. RC circuit Consider the RC circuit shown in Figure 3. The differential

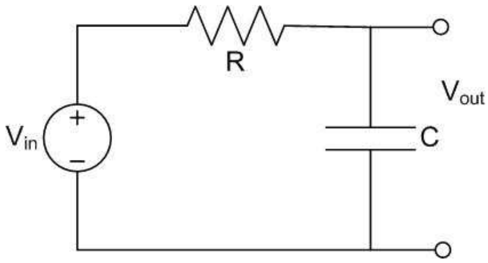

Figure 3: RC circuit

equation that describe the relationship between the input and output voltage (V_{in} and V_{out}) is obtained directly from the Kirchoff's laws:

$$V_{in} = V_R + V_{out}$$

where

$$V_R = Ri_R = Ri_C, \quad i_C = C\frac{dV_{out}}{dt}$$

and we obtain:

$$V_{in} = RC\frac{dV_{out}}{dt} + V_{out}$$

If we apply the Laplace transform to this equation, when all the initial conditions are zero, the transfer function is calculated as:

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{RCs+1}$$

The gain $K = 1$ and the time constant is $T = RC$.

In the following we shall analyze the system responses to such inputs as the unit step, unit ramp and unit impulse functions. The initial conditions are assumed to be zero. Note that all systems having the same transfer function will exhibit the same output in response to the same input. For any given physical system, the mathematical response can be given a physical interpretation.

Unit-step response of first-order systems

Since the Laplace transform of the unit step function is $1/s$, substituting $R(s) = 1/s$ into equation (1), we obtain:

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

Expanding $C(s)$ into partial fraction gives:

$$C(s) = K\left(\frac{1}{s} - \frac{T}{Ts+1}\right)$$

Taking the inverse Laplace transform, we obtain:

$$c(t) = K\left(1 - e^{-t/T}\right), \quad (t \geq 0) \tag{2}$$

Equation (2) states that initially the output $c(t)$ is zero and finally approaches K. One important characteristic of such an exponential response curve $c(t)$ is that at $t = T$ the value of $c(t)$ is $0.632K$, or the response has reached 63.2% of its total change. This may be easily seen by substituting $t = T$ in $c(t)$. That is:

$$c(T) = K(1 - e^{-1}) = 0.632K$$

Note that the smaller the time constant T, the faster the system response. Another important characteristic of the exponential response curve is that the slope of the tangent at $t = 0$ is K/T, since:

$$\frac{dc(t)}{dt} = \frac{K}{T}e^{-t/T}|_{t=0} = \frac{K}{T} \tag{3}$$

The output would reach the final value at $t = T$ if it maintained its initial speed of response. From equation (3) we see that the slope of the response curve $c(t)$ decreases monotonically from K/T at $t = 0$ to zero at $t = \infty$.

The exponential response curve $c(t)$ given by equation (2) is shown in Figure 4. In one time constant, the exponential response curve has gone from 0 to 63.2% of the final value. For $t \geq 4T$ the response remains within 2% of the final value. As seen from equation (2), the steady-state is reached mathematically only after an infinite time. In practice, however, a reasonable estimate of the response time is the length of time the response curve needs to reach the 2% line of the final value, or four time constants.

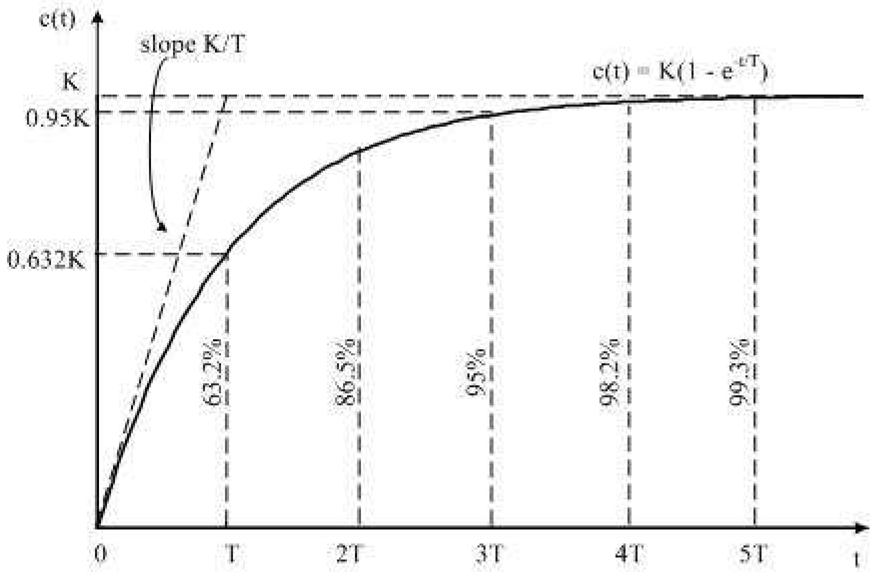

Figure 4: First-order system step response

Unit-ramp response of first-order systems

Since the Laplace transform of the unit ramp function is $1/s^2$, we obtain the output of the system of Figure 2 as:

$$C(s) = \frac{K}{Ts+1}\frac{1}{s^2}$$

Expanding $C(s)$ into partial fraction gives

$$C(s) = K\left(\frac{1}{s^2} - \frac{T}{s} + \frac{T^2}{Ts+1}\right)$$

Taking the inverse Laplace transform of this equation we obtain:

$$c(t) = K\left(t - T + Te^{-t/T}\right), \quad (t \geq 0) \tag{4}$$

A plot of the unit ramp response of a general first-order system is shown in Figure 5.

In case the gain factor $K = 1$, the error signal $e(t)$, calculated as the difference between the input and output signal, is:

$$e(t) = r(t) - c(t) = T(1 - e^{-t/T})$$

As t approaches infinity, $e^{-t/T}$ approaches zero, and thus the error signal $e(t)$ approaches T or

$$e(\infty) = T$$

The unit ramp input and the system output are shown in Figure 6. The error in tracking the unit-ramp input is equal to T for sufficiently large T. The smaller the time constant T, the smaller the steady-state error in following the ramp input.

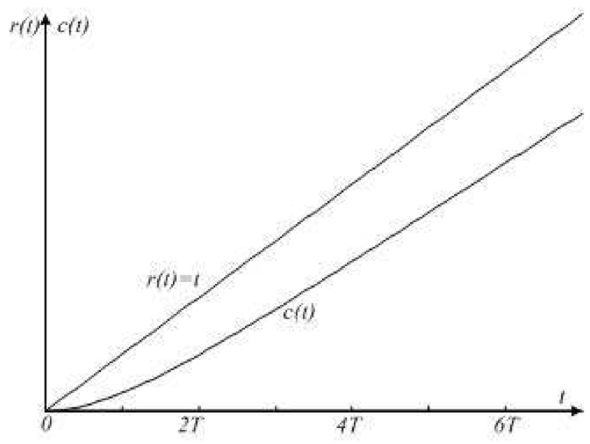

Figure 5: First-order system ramp response

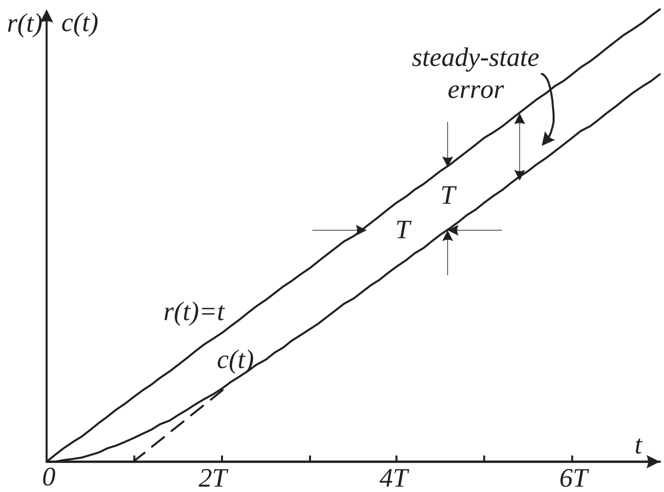

Figure 6: First-order system ramp response ($K = 1$)

Unit-impulse response of first-order systems

For the unit-impulse input, $R(s) = 1$ and the output of the system of Figure 2 can be obtained as:

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

or,

$$c(t) = \frac{K}{T}e^{-t/T}, \quad (t \geq 0)$$

The response curve is shown in Figure 7

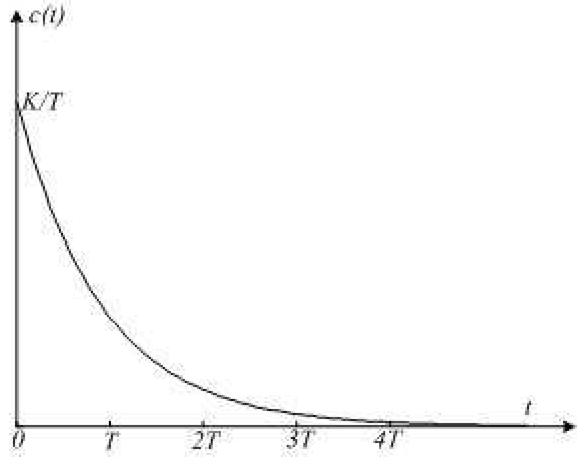

Figure 7: First-order system impulse response

An important property of linear time-invariant system

In the analysis above, it has been shown that for a unit-ramp input the output $c(t)$ is

$$c(t) = K(t - T + Te^{-t/T}), \quad (t \geq 0),$$

for the unit-step input, which is the derivative of the unit-ramp input $c(t)$ is

$$c(t) = K(1 - e^{-t}), \quad (t \geq 0)$$

and for the unit-impulse, which is the derivative of the step input, the output is:

$$c(t) = \frac{K}{T}e^{-t/T}, \quad (t \geq 0)$$

Comparison of system response to these three inputs clearly indicates that the response to the derivative of an input signal can be obtained by differentiating the response of the system to the original signal. It can also be seen that the response to the integral of the original signal can be obtained by integrating the response of the system to the original signal and by determining the integration constants from the zero output initial condition. This is a property of linear time-invariant (LTI) systems. Linear time-varying and nonlinear systems do not posses this property.

Second-order systems

Consider the second-order system shown in Figure 8. where the system transfer function

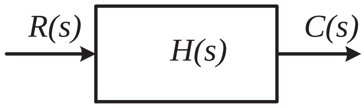

Figure 8: Second-order system

is written in the generalized form:

$$H(s) = \frac{C(s)}{R(s)} = \frac{1}{\frac{1}{\omega_n^2}s^2 + \frac{2\zeta}{\omega_n}s + 1} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2}$$

The dynamic behavior of the second-order system can be described in terms of two parameters: *the natural frequency* ω_n, and *the damping factor* ζ.

Example. Consider the system with the transfer function:

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

The system parameters are then calculated from:

$$\frac{1}{\omega_n^2} = 1; \quad \frac{2\zeta}{\omega_n} = 1; \quad \Rightarrow \omega_n = 1; \quad \zeta = \frac{1}{2}$$

If $0 < \zeta < 1$, the poles of the are complex conjugates and lie in the left-half s-plane. The system is then called **underdamped** and the transient response is oscillatory.

The roots of the characteristic equation

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0$$

are:

$$s_{1,2} = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$

They are complex for $0 < \zeta < 1$ and real for $\zeta \geq 1$.

If $\zeta = 1$, the system is called **critically damped**. **Overdamped systems** correspond to $\zeta > 1$. The transient response of critically damped and overdamped systems do not oscillate. If $\zeta = 0$, the transient response does not die out.

Step-response of second-order systems

The transfer function of the second-order system is:

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n^2 s + \omega_n^2}$$

and consider the input $r(t) = 1, \quad (t \geq 0)$ a unit step. Then $R(s) = 1/s$ and the system response is:

$$C(s) = \frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n^2 s + \omega_n^2)}$$

1. **Underdamped case** $(0 < \zeta < 1)$

   The poles of the transfer function are complex conjugates and $C(s)$ can be expanded in partial fractions:

$$C(s) = \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{1}{s} - \frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} \tag{5}$$

   where $\omega_d = \omega_n\sqrt{1 = \zeta^2}$ is called the **damped natural frequency**.

   Using a Laplace transform table we obtain:

$$\pounds^{-1}\left[\frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t} cos\omega_d t$$

$$\pounds^{-1}\left[\frac{\omega_d}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] = e^{-\zeta\omega_n t} sin\omega_d t$$

   Hence, the inverse Laplace transform of equation (5) is obtained as:

$$\pounds^{-1}[C(s)] = c(t) = 1 - e^{-\zeta\omega_n t}\left[cos\omega_d t + \frac{\zeta}{\sqrt{1 - \zeta^2}} sin\omega_d t\right] =$$

$$= 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}} \cdot sin\left(\omega_d t + arctan\frac{\sqrt{1 - \zeta^2}}{\zeta}\right), \quad (t \geq 0) \tag{6}$$

   From equation (6) it can be seen that the frequency of transient oscillation is the damped natural frequency ω_d and thus varies with the damping ration ζ.

   If the damping ratio ζ is equal to zero, the response becomes undamped and oscillations continue indefinitely. The response $c(t)$ for the zero damping case may be obtained by substituting $\zeta = 0$ in equation (6), yielding:

$$c(t) = 1 - \cos\omega_n t, \quad (t \geq 0) \tag{7}$$

   Thus, from equation (7) we see that ω_n represents the undamped natural frequency of the system. That is, ω_n is that frequency at which the system would oscillate if the damping were decreased to zero.

2. **Critically damped case**, $(\zeta = 1)$

   If the two poles of $C(s)/R(s)$ are equal the system is a critically damped one. For a unit step input $R(s) = 1/s$, $C(s)$ can be written as:

$$C(s) = \frac{\omega_n^2}{(s + \omega_n)^2 s} \tag{8}$$

   The inverse Laplace transform of equation (8) may be found as:

$$c(t) = 1 - e^{-\omega_n t}(1 - \omega_n t), \quad , (t \geq 0)$$

3. **Overdamped case**, ($\zeta > 1$)

   In this case, the two poles of $C(s)/R(s)$ are negative real and undamped. For a unit step, $C(s)$ can be written:

$$C(s) = \frac{\omega_n^2}{s(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1})} \tag{9}$$

   By partial fraction expansion and taking the inverse Laplace transform of equation (9) the system output $c(t)$ is:

$$\begin{aligned} c(t) = 1 + & \frac{1}{2\sqrt{\zeta^2 - 1}(\zeta + \sqrt{\zeta^2 - 1})} e^{-(\zeta + \sqrt{\zeta^2 - 1})\omega_n t} - \\ & - \frac{1}{2\sqrt{\zeta^2 - 1}(\zeta + \sqrt{\zeta^2 - 1})} e^{-(\zeta - \sqrt{\zeta^2 - 1})\omega_n t} \\ & = 1 + \frac{\omega_n}{2\sqrt{\zeta^2 - 1}} \left(\frac{e^{-s_1 t}}{s_1} - \frac{e^{-s_2 t}}{s_2} \right) \end{aligned} \tag{10}$$

   where $s_1 = (\zeta + \sqrt{\zeta^2 - 1})\omega_n$ and $s_2 = (\zeta - \sqrt{\zeta^2 - 1})\omega_n$ are the system poles. Thus, the response $c(t)$ includes two decaying exponential terms.

The transient response of a second-order system for various values of the damping ratio ζ is shown in Figure 9. As ζ decreases, the closed-loop roots approach the immaginary

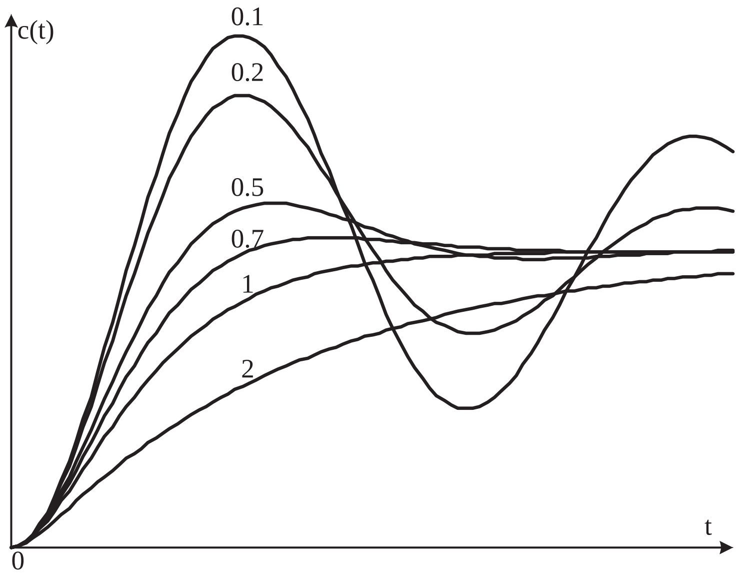

Figure 9: Step response of a second-order system for various values of damping ratio

axis and the response becomes increasingly oscillatory.

Definitions of transient-response specifications

In many practical cases, the desired performance characteristics of control systems are specified in terms of time-domain quantities. Systems with energy storage cannot respond instantaneously and exhibit transient responses whenever they are subjected to inputs or disturbances.

The transient response of a practical control system often exhibits damped oscillations before reaching steady-state. It is common to specify the following:

1. Rise time

2. Peak time

3. Maximum overshoot

4. Settling time

These specifications are defined in what follows and are shown graphically in Figure 10.

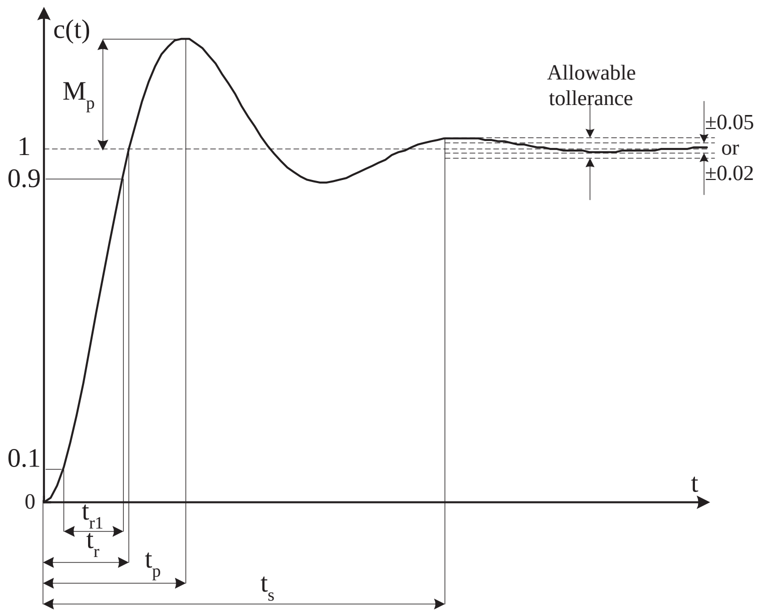

Figure 10: Second-order system underdamped response

1. Rise time, t_r : the time required for the response to rise from 10% to 90%, 5% to 95% or 0% to 100% of its final value. For underdamped second-order systems the 0% to 100% rise time is normally used. For overdamped systems, the 10% to 90% rise time is commonly used.

2. Peak time, t_p : the time required for the response to reach the first peak of the overshoot.

3. **Maximum (percent) overshoot** M_p : the maximum peak value of the response curve measured from the steady-state value of the response:

$$M_p = \frac{c(t_p) - c(\infty)}{c(\infty)} \cdot 100\%$$

The amount of the maximum (percent) overshoot directly indicates the relative stability of the system.

4. **Settling time,** t_s : the time required for the response curve to reach and stay within a range about the final value, of size specified by absolute percentage of the final value (usually 2% or 5%). The settling time is related to the largest time constant of the control system.

The time-domain specifications just given are quite important since most control systems are time-domain systems; that is they must exhibit acceptable time responses. This means, that the control system must be modified until the transient response is satisfactory. Note that if we specify the values of t_r, t_p, t_s and M_p, then the shape of the response curve is virtually determined.

Except for certain applications where oscillations cannot be tolerated, it is desirable that the transient response be sufficiently damped. Thus, for a desirable transient response of a second-order system, the damping ratio must be between 0.4 and 0.8. Smaller values of ζ yield excessive overshoot in the transient response and a system with a large value of ζ responds sluggishly.

Second-order systems and transient response specifications

In the following we shall obtain the rise time, peak time, maximum overshoot and settling time of the second-order system response. The values will be obtained in terms of ζ and ω_n. The system is assumed to be underdamped.

1. **Rise time,** t_r . Referring to equation (6) we obtain the rise time t_r by letting $c(t_r) = 1$ or

$$c(t_r) = 1 = 1 - e^{-\zeta\omega_n t_r}(cos\omega_d t_t + \frac{\zeta}{\sqrt{1-\zeta^2} sin\omega_d t_r})$$

Since $e^{-\zeta\omega_n t_r} \neq 0$, we obtain:

$$cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2} sin\omega_d t_r} = 0$$

or

$$tan\omega_d t_r = -\frac{\zeta}{\sqrt{1-\zeta^2}} = -\frac{\omega_d}{\sigma}$$

Thus, the rise time t_r is:

$$t_r = \frac{1}{\omega_d} \cdot arctan\left(\frac{\omega_d}{-\sigma}\right) = \frac{\pi - \beta}{\omega_d}$$

where β and σ are defined in Figure 11.

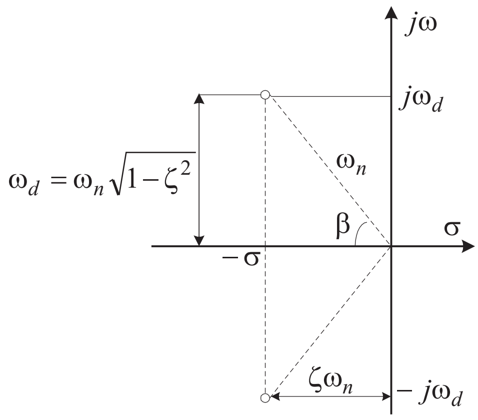

Figure 11: Second-order system poles

2. **Peak time,** t_p . Referring to equation (6) we may obtain the peak time by differentiating $c(t)$ with respect to time and letting the derivative equal zero, or:

$$\frac{dc(t)}{dt}|_{t=t_p} = sin(\omega_d t_p) \cdot \frac{\omega_n}{\sqrt{1-\zeta^2}} \cdot e^{-\zeta\omega_n t_p} = 0$$

This yields the following equation:

$$sin(\omega_d t_p) = 0$$

or

$$\omega_d t_p = 0, \pi, 2\pi, 3\pi, ...$$

Since the peak time corresponds to the first peak overshoot, $\omega_d t_p = \pi$. Hence

$$t_p = \frac{\pi}{\omega_d}$$

The peak time corresponds to one half cycle of the frequency of damped oscillation.

3. **Maximum (percent) overshoot,** M_p occurs at the peak time or at $t = t_p = \frac{\pi}{\omega_d}$. Thus, from equation (6) M_p is obtained as:

$$M_p = c(t_p) - 1 = -e^{-\zeta\omega_n\pi/\omega_d}\left(cos\pi + \frac{\zeta}{\sqrt{1-\zeta^2}} sin\pi\right)$$

or

$$M_p = e^{-\pi\zeta/\sqrt{1-\zeta^2}}$$

4. **Settling time,** t_s . For an underdamped second-order system, the transient response is obtained from equation (6) as:

$$c(t) = 1 - e^{-\zeta\omega_n t} \cdot sin\left(\omega_d t + arctan\frac{\sqrt{1-\zeta^2}}{\zeta}\right)$$

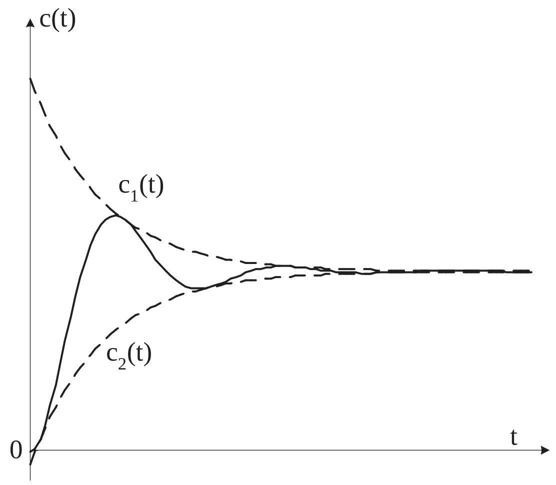

Figure 12: Step response of a second-order system and the envelope curves

The curves $c_{1,2}(t) = \pm e^{-\zeta\omega_n t}/\sqrt{1-\zeta^2}$ are the envelope curves of the transient response for a unit-step input. The response curve $c(t)$ always remains within a pair of the envelope curves as shown in Figure 12. The two envelope curves, as well as $c(t)$ will reach 2% from the final value approximately when

$$e^{-\zeta\omega_n t_s} < 0.02$$

or

$$\zeta\omega_n t_s \cong 4$$

Therefore we have:

$$t_s = \frac{4}{\zeta\omega_n}$$

Example. Consider the closed-loop system with the transfer function:

$$H(s) = \frac{25}{s^2 + 6s + 25}$$

From the general form of a second-order system transfer function:

$$H(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n + \omega_n^2}$$

the system parameters are:

$$\omega_n = 5, \quad \zeta = 0.6$$

Then we obtain:

1. The damped natural frequency is:

$$\omega_d = \omega_n\sqrt{1-\zeta^2} = 5\sqrt{1-0.6^2} = 4$$

and the poles negative real part:

$$\sigma = -\zeta\omega_n = -3.$$

According to Figure 11, the angle β is:

$$\beta = arctan\frac{\omega_d}{\sigma} = 0.93$$

2. *The rise time* is:

$$t_r = \frac{\pi - \beta}{\omega_d} = \frac{3.14 - 0.93}{4} = 0.55sec$$

3. *The peak time* is:

$$t_p = \frac{\pi}{\omega_d} = \frac{3.14}{4} = 0.78sec$$

4. *Maximum overshoot* will be:

$$M_p = e^{-\pi\zeta/\sqrt{1-\zeta^2}} = 0.095$$

The maximum percent overshoot is then: $M_p = 9.5\%$.

5. For the 2% criterion *the settling time* is:

$$t_s = \frac{4}{\sigma} = \frac{4}{\zeta\omega_n} = \frac{4}{3} = 1.33sec$$

In Figure 13 is shown the step response of the system. The values of the system parameters can be seen also from the plot.

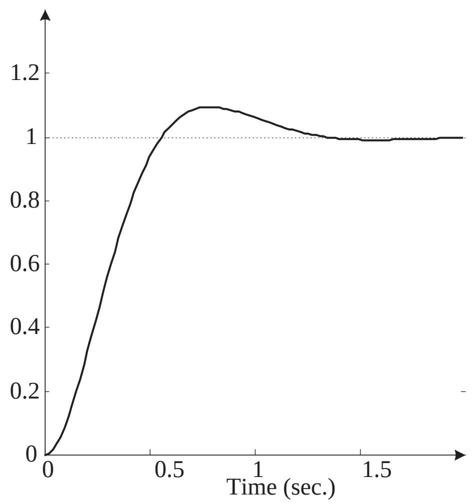

Figure 13: Step response

Steady-State Error

As a further requirement in analyzing a control system one must examine and compare the final steady-state error for an open-loop and a closed-loop system. The **steady-state error** is the error after the transient response has decayed, leaving only the steady-state response.

For an open-loop system having the input $r(t)$ and the output $c(t)$, the error signal is:

$$e(t) = r(t) - c(t)$$

and the **steady-state error,** e_{ss} is defined as:

$$e_{ss} = \lim_{t\to\infty} e(t) = \lim_{t\to\infty}(r(t) - c(t))$$

The Laplace transform of the error signal for the open-loop system shown in Figure 14 is:

$$E(s) = R(s) - C(s) = R(s) - G(s)R(s) = (1 - G(s))R(s)$$

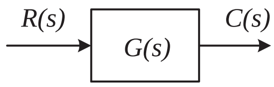

Figure 14: Open-loop system

The Laplace transform of the error for unity feedback closed-loop system, shown in Figure 15, is obtained from:

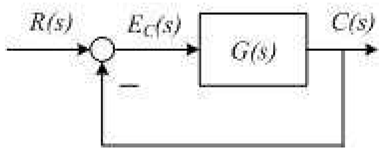

Figure 15: Closed-loop system

$$E_c(s) = R(s) - C(s) = R(s) - G(s)C(s)$$

or

$$E_c(s) = \frac{1}{1 + G(s)}R(s)$$

To calculate the steady-state error, we utilize the final value theorem, which is:

$$\lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s)$$

For the open-loop system we obtain:

$$e_{ss} = \lim_{s\to 0} s(1 - G(s))R(s)$$

and for the closed-loop system:

$$e_{ss} = \lim_{s\to 0} s\left(\frac{1}{1+G(s)}R(s)\right)$$

Therefore, using a unit step input as a comparable input, we obtain for the open-loop system

$$e_{ss} = \lim_{s\to 0} s(1 - G(s))(\frac{1}{s}) = \lim_{s\to 0}(1 - G(s)) = 1 - G(0)$$

For the closed-loop system, we have:

$$e_{ss} = \lim_{s\to 0} s\left(\frac{1}{1+G(s)}\right)\left(\frac{1}{s}\right) = \frac{1}{1+G(0)}$$

Example. The advantage of the closed-loop system in reducing the steady-state error of the system resulting from parameter changes and calibration errors may be illustrated by an example. Let us consider a system with a process transfer function

$$G(s) = \frac{k}{Ts+1}$$

which would represent a thermal control process, a voltage regulator, or a water-level control process. For a specific setting of the desired input variable, which may be represented by the normalized unit step input function, we have $R(s) = 1/s$. Then, the steady-state error of the open-loop system is:

$$e_{ss} = 1 - G(0) = 1 - k$$

The error for the closed-loop system of Figure 15 is:

$$e_{ss} = \frac{1}{1+G(0)} = \frac{1}{1+k}$$

For the open-loop system, one would calibrate the system so that $k = 1$ and the steady-state error is zero. For the closed-loop system one would set a large gain k, for example $k = 100$. Then the closed-loop system steady-state error is $e_{ss} = 1/101$.

Effect of an additional zero

Consider a system with the transfer function $H(s)$. The system step response can be expressed as:

$$y(t) = \pounds^{-1}\left[Y(s)\right] = \pounds^{-1}\left[\frac{H(s)}{s}\right]$$

Now suppose we add a zero at $-a$ and divide the transfer function with a so the gain of the new system is unchanged. The new transfer function is:

$$H_z(s) = \frac{s+a}{a}H(s) = \frac{s}{a}H(s) + H(s)$$

The step response of the system $H_z(s)$ result:

$$y_z(t) = \pounds^{-1}\left[Y_z(s)\right] = \pounds^{-1}\left[\frac{1}{s}H_z(s)\right] = \pounds^{-1}\left[\frac{1}{s}\left(\frac{s}{a}H(s) + H(s)\right)\right] = \frac{1}{a}\dot{y}(t) + y(t)$$

(since $\pounds^{-1}[sY(s)] = \dot{y}(t)$)

If a is small, or the zero is close to the imaginary axis, $1/a$ is large and the step response of $H_z(s)$ will increase with the quantity $1/a \cdot \dot{y}(t)$. The effect of addition of a zero is the increase of the overshoot.

Example. Consider the system with the transfer function:

$$\text{System 1:} \quad H_1(s) = \frac{1}{s^2+s+1}$$

We add a zero at -1 and obtain:

$$\text{System 2:} \quad H_2(s) = \frac{s+1}{s^2+s+1}$$

The step responses for these two systems are presented in Figure 16 Now consider the

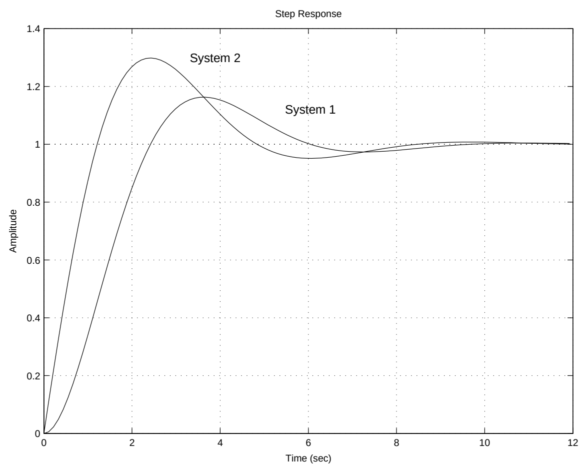

Figure 16: Effect of addition of a zero at -1

zero was added at -10 and the gain of the new system was divided by the value 10 so the total gain is equal to one (the gain of system 1).

$$\text{System 3:} \quad H_3(s) = \frac{0.1(s+10)}{s^2+s+1}$$

The step responses were compared in Figure 17. The general effect of additional zero to the second order system is the increase of overshoot.

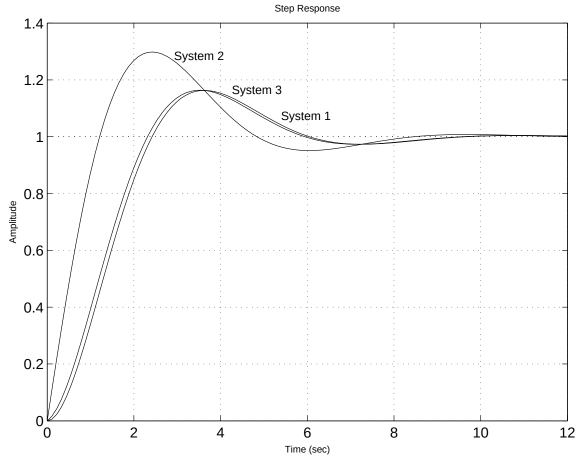

Figure 17: Effect of addition of a zero at -1 and -10

Transient response of higher-order systems

Consider a system with unit step input $R(s) = 1/s$ an output $C(s)$ and the transfer function H(s). Then C(s) can be written:

$$C(s) = H(s) \cdot R(s) = \frac{a^m s^m + ... + a_1 s + a_0}{s(b^n s^n + ... + b_1 s + b_0)}, \quad (m \leq n) \tag{11}$$

The transient response of this system to any given input can be obtained by computer simulation. If an analytical expression for the transient response is desired then it is necessary to factor the denominator polynomial. The poles of $C(s)$ consist of real poles and complex conjugates poles. A pair of complex-conjugates poles yields a second order term in s. Since the factored form of the higher-order characteristic equation consists of first- and second-order terms, equation (11) can be rewritten:

$$C(s) = \frac{K \prod_{i=1}^{m}(s + z_i)}{s \prod_{j=1}^{q}(s + p_j) \prod_{k=1}^{r}(s^2 + 2\zeta_k \omega_k s + \omega_k^2)}$$

where $q + 2r = n$. If the closed-loop poles are distinct $C(s)$ can be expanded into partial fractions as follows:

$$C(s) = \frac{a}{s} + \sum_{j=1}^{q} \frac{a_j}{s + p_j} + \sum_{k=1}^{r} \frac{b_k(s + \zeta_k \omega_k) + c_k \omega_k \sqrt{1 - \zeta_k^2}}{s^2 + 2\zeta_k \omega_k s + \omega_k^2}$$

From this last equation we see that the response of a higher-order system is composed of a number of terms involving the simple functions found in the responses of first- and second-order systems. The unit-step response $c(t)$, the inverse Laplace transform of $C(s)$ is then:

$$c(t) = a + \sum_{j=1}^{q} a_j e^{-p_j t} + \sum_{k=1}^{r} b_k e^{-\zeta_k \omega_k t} cos \omega_k \sqrt{1 - \zeta_k^2} t + \sum_{k=1}^{r} b_k e^{-\zeta_k \omega_k t} sin \omega_k \sqrt{1 - \zeta_k^2} t$$

If all the closed-loop poles lie in the left-half s-plane, then the exponential terms and the damped exponential terms will approach zero as time increases. The steady-state value of the output is $c(\infty) = a$.

Let us assume that the system considered is a stable one. Then the closed-loop poles that are located far from the $j\omega$ (imaginary) axis have large negative real parts. The exponential terms that correspond to these poles decay very rapidly to zero.

The poles located nearest the $j\omega$ axis correspond to transient response terms that decay slowly. Those poles that have dominant effects on the transient-response behavior are called **dominant poles**. They are most important among all closed-loop poles.

Note. Consider a first-order system with the transfer function

$$H_k(s) = \frac{k}{Ts+1} = k \cdot H(s)$$

or a second-order system with the transfer function:

$$H_k(s) = k \cdot \frac{\omega_n^2}{s^2 + 2\zeta\omega_n + \omega_n^2} = k \cdot H(s)$$

The system response when the input is a unit step $R(s) = 1/s$ is:

$$c(t) = \pounds^{-1}[H_k(s) \cdot R(s)] = \pounds^{-1}[\frac{k \cdot H(s)}{s}] = k \cdot \pounds^{-1}[\frac{H(s)}{s}]$$

That is, the time response is $k \cdot c(t)$ where $c(t)$ is the response of the system with a unity gain k. The final value of the response, in case of a stable system will be $c(\infty) = k$.

System approximation using the concept of dominant poles

Consider a system with a transfer function:

$$H(s) = \frac{k(s+a)}{(\frac{1}{\omega_n^2}s^2 + \frac{2\zeta}{\omega_n}s + 1)(Ts+1)}$$

It has a zero at $z_1 = -a$ and three poles: two complex poles $p_{1,2} = -\zeta\omega_n \pm j\sqrt{1-\zeta^2}$ and one real pole at $p_3 = -1/T$. If, for example, the real pole is far from the $j\omega$ axis and the complex poles are dominant, the system order can be reduced by neglecting the real pole. Because the steady-state value of the system response to test inputs must remain the same, the gain factor must be multiplied by the absolute value of the time constant or $1/pole$ (or divided by the absolute value of the pole).

Example 1. Consider a third-order system with the transfer function:

$$H_1(s) = \frac{s+2}{(s^2+2s+2)(s+10)}$$

The system poles are: $p_1, 2 = -1 \pm j$ and $p_3 = -10$. Because the real pole is located 10 times far from the imaginary axis than the complex poles, p_3 can be neglected. The system gain will be divided by $|p_3| = 10$ and obtain:

$$H_2(s) = \frac{0.1(s+2)}{s^2+2s+2}$$

The step responses obtained for the two systems are almost the same, as shown in Figure 18.

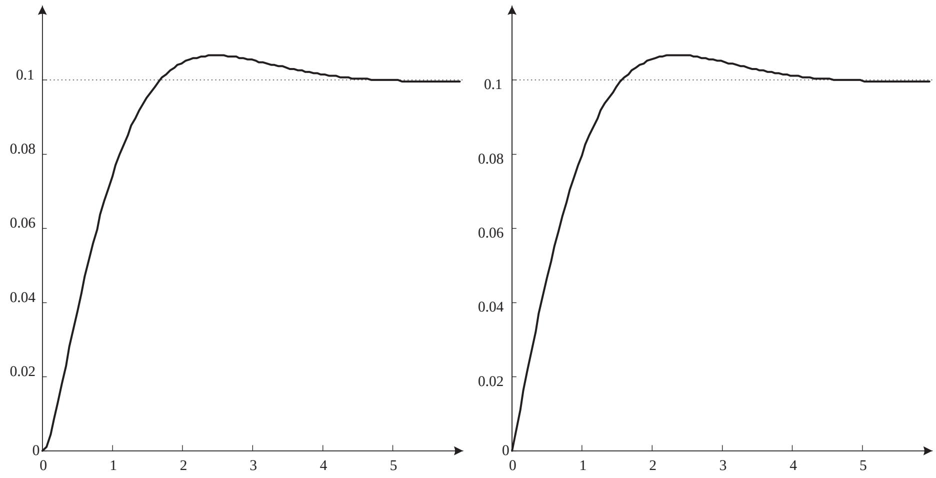

Figure 18: Comparation of two step responses

Example 2. Consider a system with the transfer function:

$$H_1(s) = \frac{62.5(s+2.5)}{(s^2+6s+25)(s+6.25)}$$

The system poles are: $p_1, 2 = -3 \pm 4 \cdot j$ and $p_3 = -6.25$. The gain is equal to 1 and the steady state value for a step input must be maintained. As an approximation we neglect the real pole and obtain:

$$H_2(s) = \frac{10(s+2.5)}{s^2+6s+25}$$

Using a computer simulation the step responses of the two systems are shown in Figure 19. The effect of neglecting the real pole was to increase the overshoot and reduce the settling time. This happened because the real pole was not very far from the complex poles and cannot be neglected for a good approximation.

Systems with transport lag

Figure 20 shows a thermal system in which hot water is circulated to keep the temperature of a chamber constant. In this system, the measuring element is placed downstream a distance L from the furnace, the air velocity is v and $T = L/v$ sec would elapse before any change in the furnace temperature was sensed by the thermometer. Such a delay in measuring, delay in controller action, or delay in actuator operation, and the like is called **transport lag** or **dead time**. Dead time is present in most process control systems.

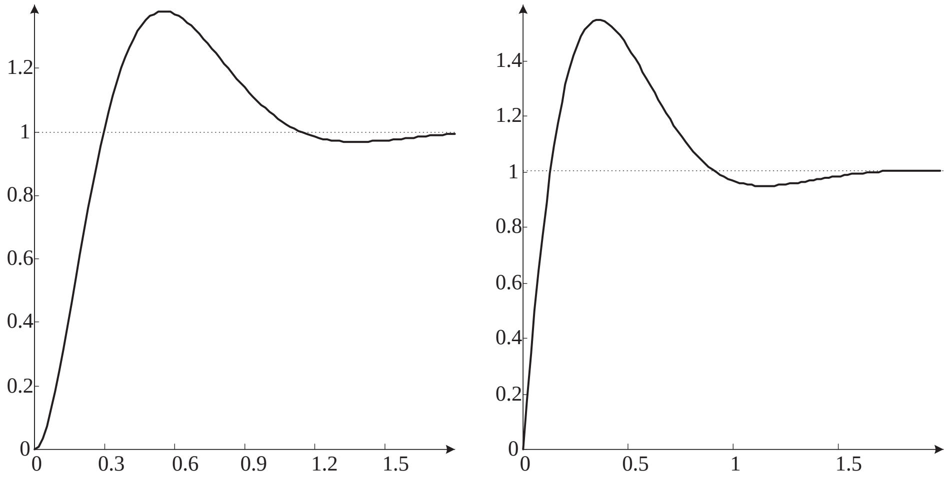

Figure 19: Comparation of two step responses

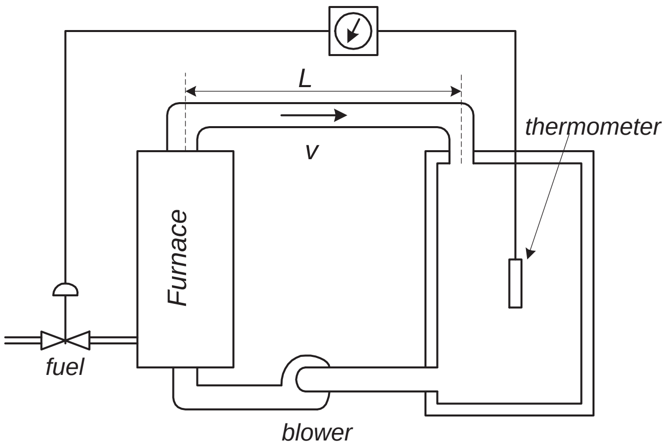

Figure 20: Thermal process

A **dead time** is the time interval between the start of an event at one point in a system and its resulting action at another point in the system.

The input $x(t)$ and the output $y(t)$ of a transport lag or dead time element are related by

$$y(t) = x(t - T)$$

where T is the dead time. The transfer function of transport lag or dead time is given by:

$$H(s) = \frac{\pounds[x(t-T)]}{\pounds[x(t)]} = \frac{X(s)e^{-sT}}{X(s)} = e^{-sT}$$

A linear system which exhibits dead time, is defined by the differential equation:

$$\sum_{j=0}^{m} a_j \frac{d^j x(t-T)}{dt^j} = \sum_{j=0}^{n} b_j \frac{d^j y(t)}{dt^j}$$

where $x(t)$ is the input signal, $y(t)$ is the output, and T the dead time. The Laplace transform of the differential equation which describes the system will give:

$$e^{-sT} \cdot \sum_{j=0}^{m} a_j s^j \pounds[x(t)] = \sum_{j=0}^{n} b_j s^j \pounds[y(t)]$$

and the transfer function is:

$$H(s) = e^{-sT} \frac{\sum_{j=0}^{m} a_j s^j}{\sum_{j=0}^{n} b_j s^j} = \frac{Y(s)}{X(s)}$$

If we use a Taylor series expansion for e^{-sT}:

$$e^{-sT} = 1 - Ts + \frac{1}{2!} T^2 s^2 - \frac{1}{3!} T^3 s^3 + ...$$

and the truncated Taylor series expansion of a ratio of two polynomials:

$$\frac{1 + \alpha Ts}{1 + \beta Ts} = 1 + (\alpha + \beta) Ts - \beta(\alpha - \beta) T^2 s^2 + \beta^2 (\alpha - \beta) T^3 s^3,$$

we can approximate the dead time transfer function as a ratio of two polynomials, usually called a *Pade approximation*:

$$a) \quad e^{-sT} = \frac{1 - \frac{1}{2} Ts}{1 + \frac{1}{2} Ts}$$

$$b) \quad e^{-sT} = \frac{1 - \frac{1}{2} Ts + \frac{1}{12} T^2 s^2}{1 + \frac{1}{2} Ts + \frac{1}{12} T^2 s^2}$$

PID - The Basic Technique for Feedback Control

Introduction

A feedback controller is designed to generate an *output* that causes some corrective effort to be applied to a *process* so as to drive a measurable *process variable* towards a desired value known as the *setpoint.* The *controller* uses an actuator to affect the process and a *sensor* to measure the results. Figure 1 shows a typical feedback control system with blocks representing the dynamic elements of the system and arrows representing the flow of information, generally in the form of electrical signals. Virtually all feedback controllers determine their output by observing the *error* between the setpoint and the actual process variable measurement.

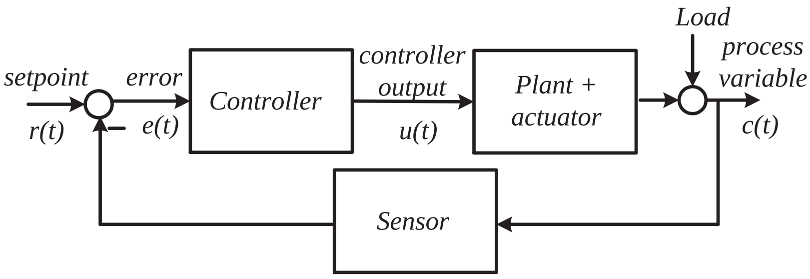

Figure 1: Control system

Errors occur when an operator changes the setpoint intentionally or when a process load changes the process variable accidentally.

Example. In warm weather, a home thermostat is a familiar controller that attempts to correct temperature of the air inside a house. It measures the room temperature with a thermocouple and activates the air conditioner whenever an occupant lowers the desired room temperature or a random heat source raises the actual room temperature. In this example, the house is the process, the actual room temperature inside the house is the process variable, the desired room temperature is the setpoint, the thermocouple is the sensor, the activation signal to the air conditioner is the controller output, the air conditioner itself is the actuator, and the random heat sources (such as sunshine and warm bodies) constitute the loads on the process.

PID Control

PID (proportional-integral-derivative) is the control algorithm most often used in industrial control. Despite the abundance of sophisticated tools, including advanced controllers, the PID controller is still the most widely used in modern industry,controlling more than 95 % of closed-loop industrial processes. It is implemented in industrial single loop controllers, distributed control systems and programmable logic controllers (PLC).

A PID controller performs much the same function as a thermostat but with a more elaborate algorithm for determining its output. It looks at the current value of the error, the integral of the error over a recent time interval, and the current derivative of the error signal to determine not only how much of a correction to apply, but for how long. Those three quantities are each multiplied by a *tuning constant* and added together to produce the current controller output $u(t)$ as in equation 1:

$$u(t) = K_P e(t) + K_I \int_0^t e(\tau) d\tau + K_D \frac{de(t)}{dt} \tag{1}$$

In this equation, K_P is the *proportional tuning constant*, K_I is the *integral tuning constant*, K_D is the *derivative tuning constant*, and the error $e(t)$ is the difference between the setpoint $r(t)$ and the process variable $c(t)$ at time t. If the current error is large or the error has been sustained for some time or the error is changing rapidly, the controller will attempt to make a large correction by generating a large output. Conversely, if the process variable has matched the setpoint for some time, the controller will leave well enough alone.

Tuning a PID controller is setting the K_P, K_I, and K_D tuning constants so that the weighted sum of the proportional, integral, and derivative terms produces a controller output that steadily drives the process variable in the direction required to eliminate the error.

How to best tune a PID controller depends upon how the process responds to the controller's corrective efforts. Consider a sluggish process that tends to respond slowly. If an error is introduced abruptly (as when the setpoint is changed), the controller's initial reaction will be determined primarily by the derivative term in equation 1. This will cause the controller to initiate a burst of corrective efforts the instant the error changes from zero. The proportional term will then come in to play to keep the controller's output going until the error is eliminated.

After a while, the integral term will also begin to contribute to the controller's output as the error accumulates over time. In fact, the integral term will eventually come to dominate the output signal, since the error decreases so slowly in a sluggish process. Even after the error has been eliminated, the controller will continue to generate an output based on the history of errors that have been accumulating in the controller's integrator. The process variable may then overshoot the setpoint, causing an error in the opposite direction.

If the integral tuning constant is not too large, this subsequent error will be smaller than the original, and the integral term will begin to diminish as negative errors are added to the history of positive ones. This whole operation may then repeat several times until both the error and the accumulated error are eliminated. Meanwhile, the derivative term will continue to add its share to the controller output based on the derivative of the oscillating error signal. The proportional term too will come and go as the error increases or decreases.

Now suppose the process responds quickly to the controller's efforts. The integral term in equation 1 will not play as dominant a role in the controller's output since the errors will be so short lived. On the other hand, the derivative term will tend to be larger since the error will change rapidly.

Clearly, the relative importance of each term in the controller's output depends on the behavior of the controlled process. Determining the best mix suitable for a particular application is the essence of controller tuning. For the sluggish process, a large value for the derivative tuning constant K_D might be advisable to accelerate the controller's reaction to an error that appears suddenly. For the fast-acting process, however, an equally large value for K_D might cause the controller's output to fluctuate as every change in the error (including extraneous changes caused by measurement noise) is amplified by the controller's derivative action.

Hundreds of mathematical and heuristic techniques for selecting appropriate values for the tuning constants have been developed over the last 50 years.

There are basically three schools of thought on how to select K_P, K_I, and K_D values to achieve an acceptable level of controller performance.

1. The first method is the simple trial-and-error approach. Experienced control engineers

seem to know just how much proportional, integral, and derivative action to add or subtract to correct the performance of a poorly tuned controller. Unfortunately, intuitive tuning procedures can be difficult to develop since a change in one tuning constant tends to affect the performance of all three terms in the controller's output.

For example, turning down the integral action reduces overshoot. This in turn slows the rate of change of the error and thus reduces the derivative action as well.

2. The analytical approach to the tuning problem is more rigorous. It involves a mathematical model of the process that relates the current value of the process variable to its current rate of change plus a history of the controller's output.

   There are hundreds of analytical techniques for translating model parameters into tuning constants. Each approach uses a different model, different controller objectives, and different mathematical tools.

3. The third approach to the tuning problem is a compromise between purely heuristic trial-and-error techniques and the more rigorous analytical techniques. It was originally proposed in 1942 by John G. Ziegler and Nathaniel B. Nichols of Taylor Instruments and remains popular today because of its simplicity and its applicability to many real-life processes.

Tuning PID controllers

The block diagram of a simplified control system is shown in Figure 2. In practice the output

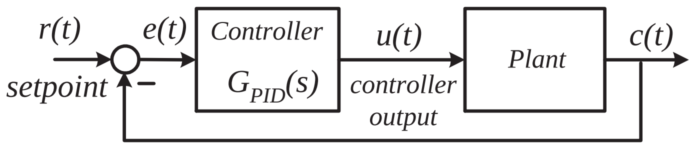

Figure 2: Control system

of a PID controller is given by:

$$u(t) = K_p \left[e(t) + \frac{1}{T_i} \int_0^t e(\tau) d\tau + T_d \frac{de(t)}{dt} \right] \tag{2}$$

The transfer function of a PID controller is:

$$G_{PID}(s) = \frac{U(s)}{E(s)} = K_p \left(1 + \frac{1}{T_i s} + T_d s \right)$$

where K_p = proportional gain
T_i = integral time
T_d = derivative time

If a mathematical model of the plant can be derived, then it is possible to apply various design techniques for determining parameters of the controller that will meet the transient and steady-state specifications of the closed-loop system.

If the plant is so complicated and its mathematical model cannot be easily obtained, then analytical approach to the design of a PID controller is not possible. Then we must resort to experimental approaches to the design of PID controllers.

The process of selecting the controller parameters to meet given performance specifications is known as controller tuning.

Ziegler and Nichols (1942) proposed rules for tuning PID controllers (for determining values of K_p, T_i and T_d) based on the transient response characteristics of a given plant. Such determination of the parameters of PID controller can be made by engineers on site by experiments on the plant.

There are two methods called Ziegler-Nichols tuning rules. In both they aimed at obtaining 25% maximum overshoot in step response.

The Ultimate Sensitivity Method

The goal is to achieve a marginally stable controller response. A P-controller is used first to control the system as shown in Figure 3. Set $T_i = \infty$, and $T_d = 0$.

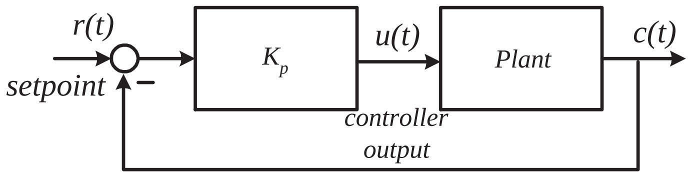

Figure 3: Control system with proportional control

Using the proportional control action only, increase K_p from 0 to a critical value K_0 where the output, $c(t)$, first exhibits sustained oscillations. (If the output does not exhibits sustained oscillations for whatever value K_p may take, then this method does not apply.)

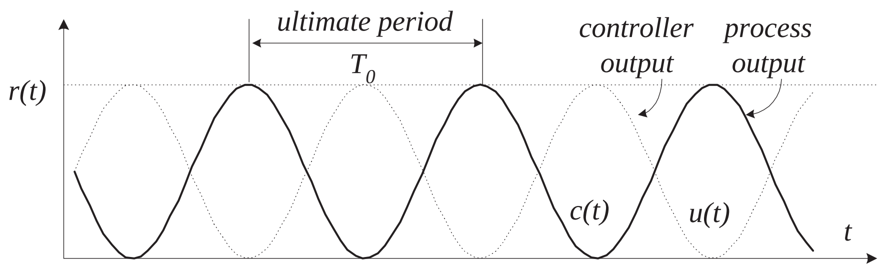

Figure 4: Ultimate system response

When the response shown in Figure 4 is obtained, the result is termed the *ultimate gain* setting that causes a continuous sinusoidal response in the process output.

Thus, the critical gain and the corresponding period T_0 are experimentally determined.

Ziegler and Nichols suggested to set the values of the parameters K_p, T_i and T_d according to the formula shown in Table 1.

Notice that the PID controller tuned by the ultimate sensitivity method of Ziegler-Nichols

Type of controller	K_p	T_i	T_d
P	$0.5K_0$	∞	0
PI	$0.45K_0$	$1/1.2T_0$	0
PID	$0.6K_0$	$0.5T_0$	$0.125T_0$

Table 1: PID Parameters

gives:

$$G_{PID}(s) = K_p\left(1+\frac{1}{T_is}+T_ds\right) = 0.6K_0\left(1+\frac{1}{0.5T_0s}+0.125T_0s\right)$$
$$= 0.075K_0T_0\frac{(s+4/T_o)^2}{s}$$

Thus, the PID controller has a pole at the origin and double zeros at $s=-4/T_0$.

Derivative cautions. Derivative action is applied only one time when the process output moves away from the setpoint. Derivative works on rate of change. If the process output rate of change is caused by noise, the derivative may cause over-corrections. Control loops likely to have significant noise are pressure and flow. Level can also be noisy when stirring/aggitating or splashing occurs.

Ziegler-Nichols tuning rules have been widely used to tune PID controllers in process control where the plant dynamics are not precisely known. Over many years such tuning rules proved to be very useful.

Ziegler-Nichols tuning rules can, of course, be applied to plants where dynamics are known.

Example Consider the control system shown in Figure 5 in which a PID controller is used to control a system with the transfer function $G(s)$. The PID controller has the transfer function

$$G_{PID}(s) = K_p\left(1+\frac{1}{T_is}+T_ds\right)$$

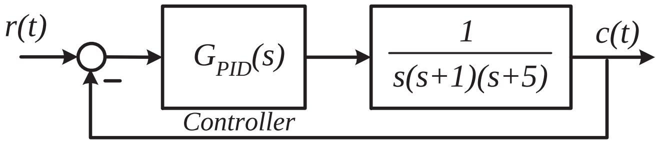

Figure 5: PID-controlled system

Although many analytical methods are available for the design of a PID controller for the present system let us apply a Ziegler Nichols tuning rule for determination of the values of parameters K_p, T_i, T_d. Then obtain a unit-step response curve and check to see if the designed system exhibits approximately 25% maximum overshoot. If the maximum overshoot is excessive, make a fine tuning and reduce the amount of the maximum overshoot to approximately 25%.

By setting $T_i = \infty$ and $T_d = 0$, we obtain the closed-loop transfer function as follows:

$$\frac{C(s)}{R(s)} = \frac{K_p}{s(s+1)(s+5) + K_p}$$

The value of K_p that makes the system marginally stable so that sustained oscillation occurs can be obtained by use of Routh's stability criterion. Since the characteristic equation for the closed-loop system is:

$$s^3 + 6s^2 + 5s + K_p = 0$$

the Routh array becomes as follows:

$$\begin{array}{ccll} s^3 & : & 1 & 5 \\ s^2 & : & 6 & K_p \\ s^1 & : & (30 - K_p)/6 & \\ s^0 & : & K_p & \end{array}$$

Examining the coefficients of the first column of the Routh table, we find that sustained oscillation will occur if $K_p = 30$. Thus, the critical gain is $K_0 = 30$.

With the gain K_p set equal to $K_0 (= 30)$, the characteristic equation becomes:

$$s^3 + 6s^2 + 5s + 30 = 0$$

or

$$(s+1)(s^2+5) = (s+1)(s+\omega_n^2) = 0$$

from which we find the frequency of the sustained oscillation to be $\omega_n = \sqrt{5}$. Hence, the period of the sustained oscillation is:

$$T_0 = \frac{2\pi}{\omega_n} = \frac{2\pi}{\sqrt{5}} = 2.81$$

Referring to Table 1, we determine K_p, T_i, T_d as follows:

$$\begin{aligned} K_p &= 0.6K_0 = 18 \\ T_i &= 0.5T_0 = 1.405 \\ T_d &= 0.125T_0 = 0.35 \end{aligned}$$

The transfer function of the PID controller is thus:

$$G_{PID}(s) = K_p \left(1 + \frac{1}{T_i s} + T_d s\right) = 18 \left(1 + \frac{1}{1.405s} + 0.35s\right) = \frac{6.32(s+1.42)^2}{s}$$

The PID controller has a pole at the origin and a zero at $s = -1.42$. The closed-loop transfer function is:

$$\frac{C(s)}{R(s)} = \frac{G_{PID}(s)G(s)}{1 + G_{PID}(s)G(s)} = \frac{6.32s^2 + 18s + 12.81}{s^4 + 6s^3 + 11.32s^2 + 18s + 12.81}$$

The unit step response of this system can be obtained easily by using a computer simulation as shown in Figure 6. The maximum overshoot in the unit-step response is approximately 60%. The amount of maximum overshoot is excessive. It can be reduced by fine tuning the control parameters. Such fine tuning can be made on the computer. We find that keeping $K_p = 18$ and by moving the double zero of the PID controller to $s = -0.65$, that is using the PID-controller:

$$G_{PID}(s) = 18 \left(1 + \frac{1}{3.07s} + 0.76s\right) = \frac{13.84(s+0.65)^2}{s}$$

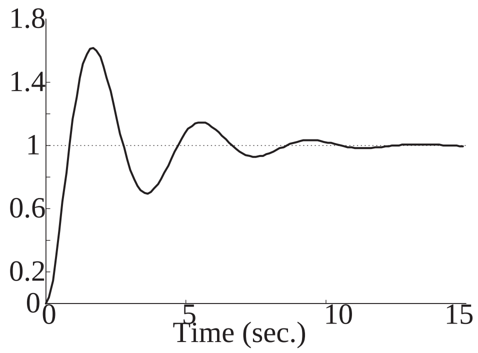

Figure 6: Unit-step response of PID-controlled system

the maximum overshoot in the unit step response ca be reduced to approximately 18% (see Figure 7 - left). If the proportional gain K_p is increased to 39.42 without changing the new location of the double zero, that is using the PID controller:

$$G_{PID}(s) = 39.42\left(1 + \frac{1}{3.07s} + 0.76s\right) = \frac{30.322(s+0.65)^2}{s}$$

then the speed of response in increased, but the maximum overshoot is also increased to approximately 28%, as shown in Figure 7 - right. Since the maximum overshoot in this case is

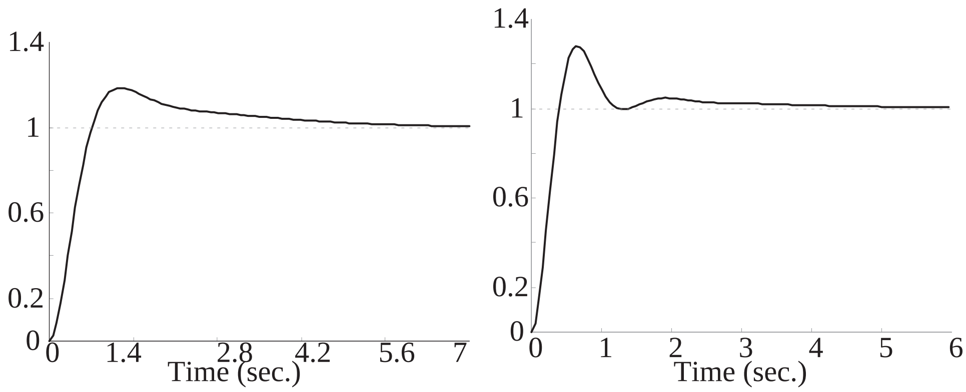

Figure 7: Unit-step response of PID-controlled system

fairly close to 25% and the response is faster we may consider that the last controller designed is acceptable. Then the tuned values of K_p, T_i, T_d are:

$$K_p = 39.42, \;\; T_i = 3.07, \;\; T_d = 0.76$$

The important thing to note here is that the values suggested by the Ziegler-Nichols tuning rule has provided a starting point for fine tuning.

Ziegler-Nichols transient response method

The method is also known as *reaction curve (open-loop) method.* The philosophy of open loop testing is to begin with a steady-state process, make a step change to the final control element and record the results of the process output, as shown in Figure 8.

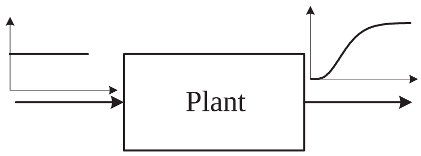

Figure 8: Open-loop step response)

Zeigler-Nichols' transient response method will work on any system that has an open-loop step response that is an essentially critically damped or overdamped character like that shown in Figure 9. Information produced by the open-loop test is the open-loop gain K, the loop apparent deadtime L, and the loop time constant, T. The transfer function of the plant may

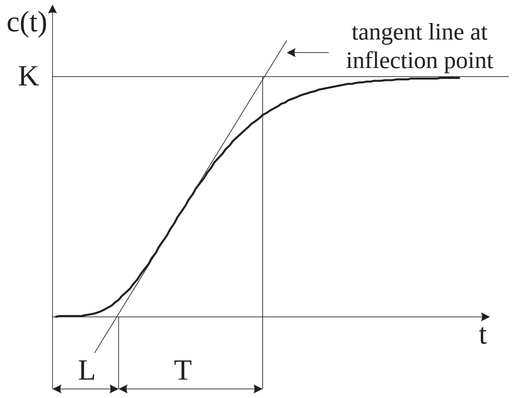

Figure 9: Open-loop step response (S-shaped response curve)

then be approximated by a first-order system with a transport lag:

$$\frac{C(s)}{U(s)} = \frac{Ke^{-Ls}}{Ts+1}$$

Ziegler and Nichols suggested to set the values of K_p, T_i and T_d according to the formula shown in Table 2.

Type of controller	K_p	T_i	T_d
P	T/L	∞	0
PI	$0.9T/L$	$L/0.3$	0
PID	$1.2T/L$	$2L$	$0.5L$

Table 2: PID Parameters

Notice that the PID controller tuned by the transient response method gives:

$$G_{PID}(s) = K_p \left(1 + \frac{1}{T_i s} + T_d s\right) = 1.2\frac{T}{L}\left(1 + \frac{1}{2Ls} + 0.5Ls\right) = 0.6T\frac{(s + 1/L)^2}{s}$$

Thus the PID controller has a pole at the origin and double zeros at $s = -1/L$.

Zeigler-Nichols tuning methods, however, tend to produce systems whose transient response is rather oscillatory and so will need to be tuned further prior to putting the system into closed-loop operation.

Bibliography

[1] Richard C. Dorf and Robert H. Bishop, *Modern Control Systems*, Addison-Wesley, 1995.

[2] Katsuhiko Ogata, *Modern Control Engineering*, Prentice Hall International Editions, 1990.