Transient Response

1 Introduction

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

In practice, the input signal to a control system is not known ahead of time but is random in nature, and the instantaneous input cannot be expressed analytically. Only in some special cases is the input signal known in advance and expressible analytically.

In analyzing and designing control systems, we must have a basis of comparison of performance. This may be set up by specifying particular test input signals and by comparing the responses of various systems.

Many design criteria are based on the response to such test signals or on the response of systems to changes in initial conditions (without any test signals). The use of test signals can be justified because of a correlation existing between the response characteristics to a typical test input signal and the capability of the system to cope with actual input signals.

1.1 Typical Test Signals

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

Which of these typical input signals to use for analyzing system characteristics may be determined by the form of the input that the system will be subjected to most frequently under normal operation. If the inputs to 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 of time may be a good test signal; and for a system subjected to shock inputs, an impulse function may be best. 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 many systems on the same basis.

1.2 Transient Response and Steady-State Response

The time response of a control system consists of two parts: the transient response and the steady-state response. By transient response, we mean that which goes from the initial state to the final state. By steady-state response, we mean the manner in which the system output behaves as t approaches infinity. Thus the system response $c(t)$ may be written as

$$c(t) = c_{tr}(t) + c_{ss}(t) \tag{1}$$

where the first term on the right-hand side of the equation is the transient response and the second term is the steady-state response.

1.3 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 characteristic 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 an initial condition. A linear time-invariant control system is critically stable if oscillations of the output continue forever. It is unstable if the output diverges without bound from its equilibrium state when the system is subjected to an initial condition.

Important system behavior (other than absolute stability) 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 practical 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 and steady-state behavior.

2 First-Order Systems

Consider the first-order system shown in Figure 1(a). Physically, this system may represent an RC circuit, thermal system, or the like. A simplified block diagram is shown in Figure 1(b). The input-output relationship is given by

$$\frac{C(s)}{R(s)} = \frac{1}{Ts+1} \tag{2}$$

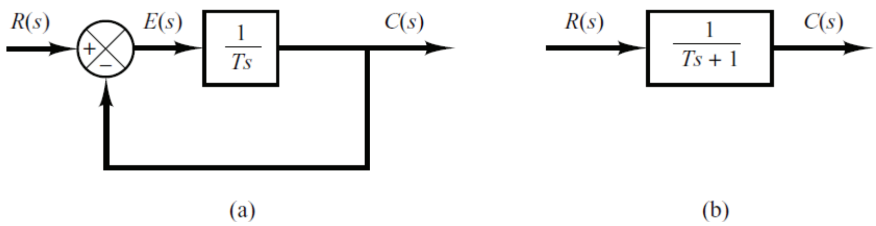

Figure 1: (a) Block diagram of a first-order system; (b) simplified block diagram.

2.1 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 (2), we obtain

$$C(s) = \frac{1}{Ts+1}\,\frac{1}{s} \tag{3}$$

Expanding $C(s)$ into partial fractions gives

$$C(s) = \frac{1}{s} - \frac{T}{Ts+1} = \frac{1}{s} - \frac{1}{s+1/T} \tag{4}$$

Taking the inverse Laplace transform of Equation (4), we obtain

$$c(t) = 1 - e^{-t/T},\ t \geq 0 \tag{5}$$

Equation (5) states that initially the output $c(t)$ is zero and finally it becomes unity. One important characteristic of such an exponential response curve $c(t)$ is that at $t = T$ the value of $c(t)$ is 0.632. This may be easily seen by substituting $t = T$ in $c(t)$. That is,

$$c(T) = 1 - e^{-1} = 0.632 \tag{6}$$

The exponential response curve $c(t)$ given by Equation (5) is shown in Figure 2. 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 line at $t = 0$ is $1/T$, since

$$\left.\frac{dc}{dt}\right|_{t=0} = \left.\frac{1}{T}e^{-t/T}\right|_{t=0} = \frac{1}{T} \tag{7}$$

The output would reach the final value at $t = T$ if it maintained its initial speed of response. From Equation (7) we see that the slope of the response curve c(t) decreases monotonically from $1/T$ at $t = 0$ to zero at $t = \infty$.

The exponential response curve $c(t)$ given by Equation (5) is shown in Figure 2. In one time constant, the exponential response curve has gone from 0 to 63.2%

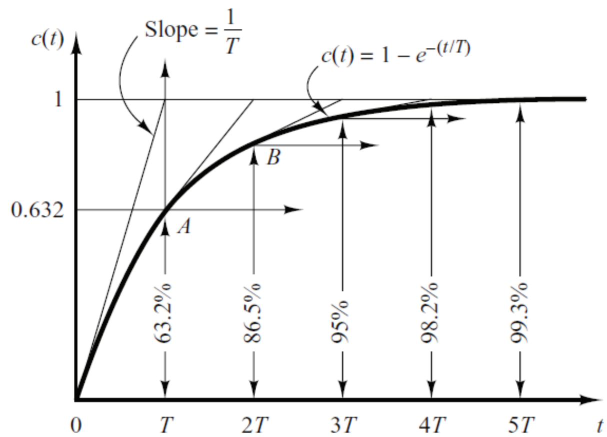

Figure 2: (a) Block diagram of a first-order system; (b) simplified block diagram.

of the final value. In two time constants, the response reaches 86.5% of the final value. At $t = 3T$, $4T$, and $5T$, the response reaches 95%, 98.2%, and 99.3%, respectively, of the final value. Thus, for $t \geq 4T$, the response remains within 2% of the final value. As seen from Equation (7), 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 and stay within the 2% line of the final value, or four time constants.

2.2 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 1(a) as

$$C(s) = \frac{1}{Ts+1}\,\frac{1}{s^2} \tag{8}$$

Expanding $C(s)$ into partial fractions gives

$$C(s) = \frac{1}{s^2} - \frac{T}{s} + \frac{T^2}{Ts+1} \tag{9}$$

Taking the inverse Laplace transform of Equation (9), we obtain

$$c(t) = t - T + Te^{-t/T},\ t \geq 0 \tag{10}$$

The error signal $e(t)$ is then

$$e(t) = r(t) - c(t) = T(1 - e^{-t/T}) \tag{11}$$

The unit-ramp input and the system output are shown in Figure 3. As t approaches infinity, $e^{-t/T}$ approaches zero, and thus the error signal $e(t)$ approaches T. The error in following 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.

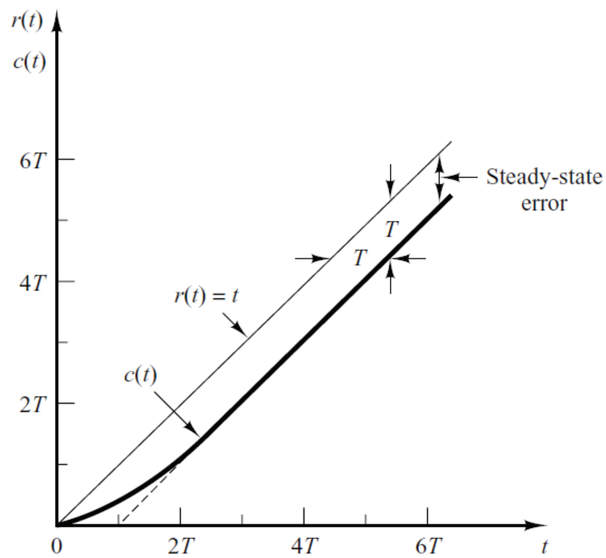

Figure 3: Unit-ramp response of the system shown in Figure 1(a).

2.3 Unit-Impulse Response of First-Order Systems

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

$$C(s) = \frac{1}{Ts + 1} \tag{12}$$

The inverse Laplace transform of Equation (12) gives

$$c(t) = \frac{1}{T}e^{-t/T}, \ t \geq 0 \tag{13}$$

The response curve given by Equation (13) is shown in Figure 4.

2.4 Important Properties of Linear Time-Invariant Systems

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

$$c(t) = t - T + Te^{-t/T}, \ t \geq 0 \tag{14}$$

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

$$c(t) = 1 - e^{-t/T}, \ t \geq 0 \tag{15}$$

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

$$c(t) = \frac{1}{T}e^{-t/T}, \ t \geq 0 \tag{16}$$

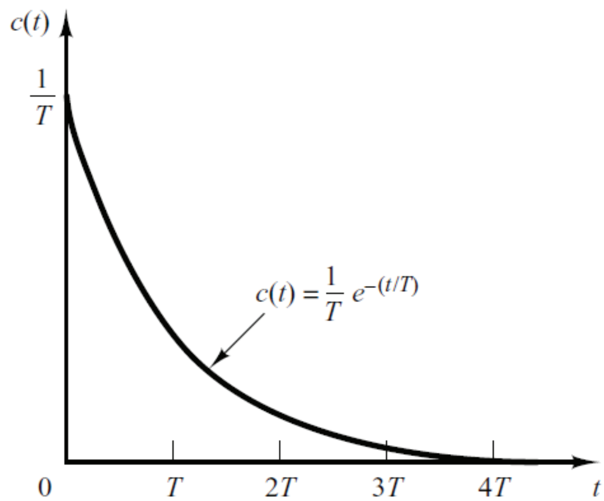

Figure 4: Unit-impulse response of the system shown in Figure 1(a).

Comparing the system responses 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 constant from the zero-output initial condition. This is a property of linear time-invariant systems. Linear time-varying systems and nonlinear systems do not possess this property.

3 Second-Order Systems

In this section, we shall obtain the response of a typical second-order control system to a step input, ramp input, and impulse input. Here we consider a servo system as an example of a second-order system.

3.1 Servo System

The servo system shown in Figure 5(a) consists of a proportional controller and load elements (inertia and viscous-friction elements). Suppose that we wish to control the output position c in accordance with the input position r. The equation for the load elements is

$$J\ddot{c} + b\dot{c} = T \tag{17}$$

where T is the torque produced by the proportional controller whose gain is K. By taking Laplace transforms of both sides of this last equation, assuming the

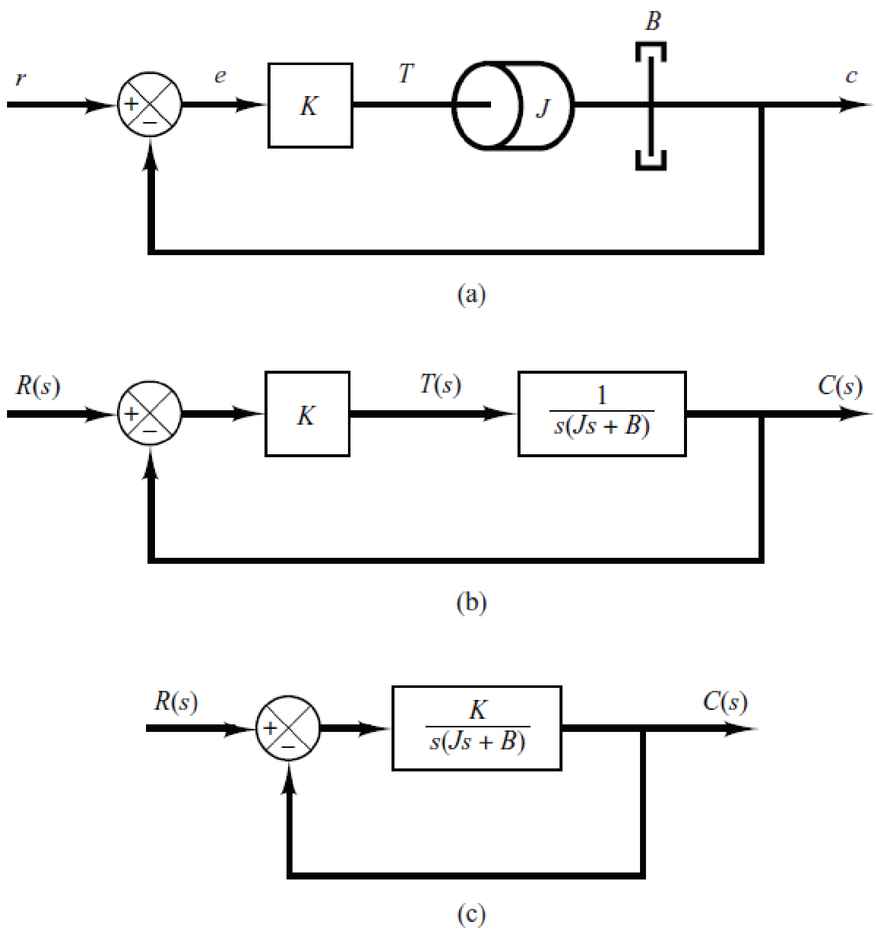

Figure 5: (a) Servo system; (b) block diagram; (c) simplified block diagram.

zero initial conditions, we obtain

$$Js^2C(s) + BsC(s) = T(s) \tag{18}$$

So the transfer function between $C(s)$ and $T(s)$ is

$$\frac{C(s)}{T(s)} = \frac{1}{s(Js + B)} \tag{19}$$

By using this transfer function, Figure 5(a) can be redrawn as in Figure 5(b), which can be modified to that shown in Figure 5(c). The closed-loop transfer function is then obtained as

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + Bs + K} = \frac{K/J}{s^2 + (B/J)s + (K/J)} \tag{20}$$

Such a system where the closed-loop transfer function possesses two poles is called a second-order system. Some second-order systems may involve one or two zeros.

3.2 Step Response of Second-Order System

The closed-loop transfer function of the system shown in Figure 5(c) is

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + Bs + K} \tag{21}$$

which can be rewritten as

$$\frac{C(s)}{R(s)} = \frac{K/J}{\left[s + B/(2J) + \sqrt{(B/(2J))^2 - K/J}\right]\left[s + B/(2J) - \sqrt{(B/(2J))^2 - K/J}\right]} \tag{22}$$

The closed-loop poles are complex conjugates if $B^2 - 4JK < 0$ and they are real if $B^2 - 4JK \geq 0$. In the transient-response analysis, it is convenient to write

$$\frac{K}{J} = \omega_n^2, \;\; \frac{B}{J} = 2\zeta\omega_n^2 = 2\sigma \tag{23}$$

where σ is called the attenuation; ω_n, the undamped natural frequency; and ζ, the damping ratio of the system. It is the ratio of the actual damping B to the critical damping

$$\zeta = \frac{B}{B_c} = \frac{B}{2\sqrt{JK}} \tag{24}$$

In terms of ζ and ω_n, the system shown in Figure 5(c) can be modified to that shown in Figure 6, and the closed-loop transfer function $C(s)/R(s)$ given by Equation (21) can be written

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{25}$$

This form is called the *standard form* of the second-order system.

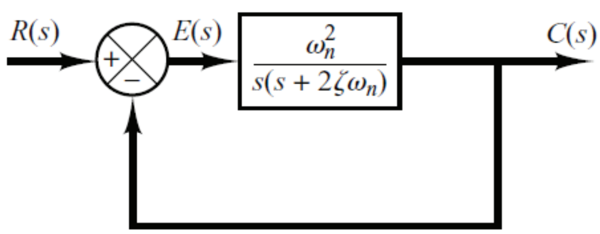

Figure 6: Second-order system.

The dynamic behavior of the second-order system can then be described in terms of two parameters ζ and ω_n. If $0 < \zeta < 1$, the closed-loop poles are complex conjugates and lie in the left-half s plane. The system is then called *underdamped*, and the transient response is oscillatory. If $\zeta = 0$, the transient response does not die out. If $\zeta = 1$, the system is called *critically damped*. *Overdamped* systems correspond to $\zeta > 1$.

We shall now solve for the response of the system shown in Figure 6 to a unit-step input. We shall consider three different cases: the underdamped ($0 < \zeta < 1$), critically damped ($\zeta = 1$), and overdamped ($\zeta > 1$) cases.

3.2.1 Underdamped case ($0 < \zeta < 1$):

In this case, $C(s)/R(s)$ can be written

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{(s + \zeta\omega_n + j\omega_d)(s + \zeta\omega_n - j\omega_d)} \tag{26}$$

where $\omega_d = \omega_n\sqrt{1 - \zeta^2}$. The frequency ω_d is called the *damped natural frequency*. For a unit-step input, $C(s)$ can be written

$$C(s) = \frac{\omega_n^2}{s(s^2 + 2\zeta\omega_n s + \omega_n^2)} \tag{27}$$

The inverse Laplace transform of Equation (27) can be obtained easily if $C(s)$ is written in the following form:

$$\begin{aligned} C(s) &= \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\ &= \frac{1}{s} - \frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} \end{aligned} \tag{28}$$

Referring to the Laplace transform

$$\begin{aligned} \mathfrak{L}^{-1}\left[\frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] &= e^{-\zeta\omega_n t}\cos\omega_d t \\ \mathfrak{L}^{-1}\left[\frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2}\right] &= e^{-\zeta\omega_n t}\sin\omega_d t \end{aligned} \tag{29}$$

Hence the inverse Laplace transform of Equation (27) is obtained as

$$\begin{aligned} \mathfrak{L}^{-1}\left[C(s)\right] &= c(t) \\ &= 1 - e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1 - \zeta^2}}\sin\omega_d t\right) \\ &= 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}}\sin\left(\omega_d t + \tan^{-1}\frac{\sqrt{1 - \zeta^2}}{\zeta}\right), \ t \geq 0 \end{aligned} \tag{30}$$

From Equation (30), it can be seen that the frequency of transient oscillation is the damped natural frequency ω_d and thus varies with the damping ratio ζ.

The error signal for this system is the difference between the input and output and is

$$e(t) = r(t) - c(t) = e^{-\zeta\omega_d t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right),\ t \geq 0 \tag{31}$$

This error signal exhibits a damped sinusoidal oscillation. At steady state, or at $t = \infty$, no error exists between the input and output.

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 (30), yielding

$$c(t) = 1 - \cos\omega_n t \tag{32}$$

Thus, from Equation (32), we see that ω_n represents the undamped natural frequency of the system. That is, the frequency at which the system output would oscillate if the damping were decreased to zero. If the linear system has any amount of damping, the undamped natural frequency cannot be observed experimentally. The frequency that may be observed is the damped natural frequency ω_d which is equal to $\omega_n\sqrt{1-\zeta^2}$. An increase in ζ would reduce the damped natural frequency ω_d. If ζ is increased beyond unity, the response becomes overdamped and will not oscillate.

3.2.2 Critically damped case ($\zeta = 1$):

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

$$C(s) = \frac{\omega_n^2}{s(s+\omega_n)^2} \tag{33}$$

The inverse Laplace transform of Equation (33) may be found as

$$c(t) = 1 - e^{-\omega_n t}(1 + \omega_n t),\ t \geq 0 \tag{34}$$

3.2.3 Overdamped case ($\zeta > 1$):

In this case, the two poles of $C(s)/R(s)$ are negative real and unequal. For a unit-step input, $R(s) = 1/s$ and $C(s)$ can be written

$$C(s) = \frac{\omega_n^2}{\left(s + \zeta\omega_n + \omega_n\sqrt{\zeta^2-1}\right)\left(s + \zeta\omega_n - \omega_n\sqrt{\zeta^2-1}\right)} \tag{35}$$

The inverse Laplace transform of Equation (35) is

$$\begin{aligned}c(t) &= 1 + \frac{1}{2\sqrt{\zeta^2-1}\left(\zeta+\sqrt{\zeta^2-1}\right)} e^{-\left(\zeta+\sqrt{\zeta^2-1}\omega_n t\right)} \\ &\quad - \frac{1}{2\sqrt{\zeta^2-1}\left(\zeta+\sqrt{\zeta^2-1}\right)} e^{-\left(\zeta-\sqrt{\zeta^2-1}\omega_n t\right)} \\ &= 1 + \frac{\omega_n}{2\sqrt{\zeta^2-1}}\left(\frac{e^{-s_1 t}}{s_1} - \frac{e^{-s_2 t}}{s_2}\right)\end{aligned} \tag{36}$$

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

When ζ is appreciably greater than unity, one of the two decaying exponentials decreases much faster than the other, so the faster-decaying exponential term (which corresponds to a smaller time constant) may be neglected.

A family of unit-step response curves $c(t)$ with various values of ζ is shown in Figure 7, where the abscissa is the dimensionless variable $\omega_n t$. The curves are functions only of ζ. These curves are obtained from Equations (30), (34), and (36). The system described by these equations was initially at rest.

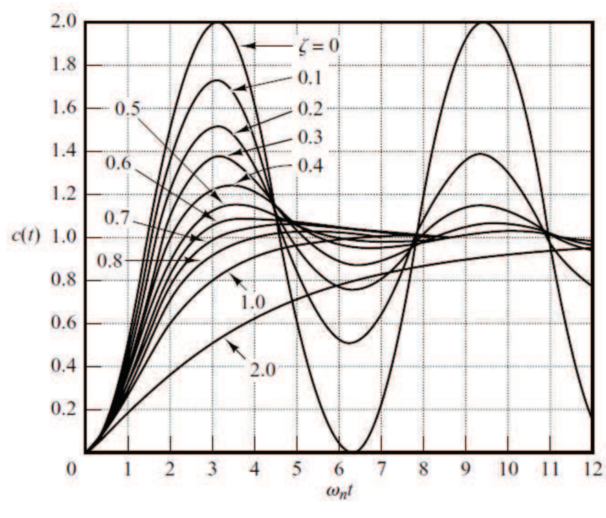

Figure 7: Unit-step response curves of the system shown in Figure 6.

Note that two second-order systems having the same ζ but different ω_n will exhibit the same overshoot and the same oscillatory pattern. Such systems are said to have the same relative stability.

From Figure 7, we see that an underdamped system with ζ between 0.5 and 0.8 gets close to the final value more rapidly than a critically damped or overdamped system. Among the systems responding without oscillation, a critically damped system exhibits the fastest response. An overdamped system is always sluggish in responding to any inputs. It is important to note that, for second-order systems whose closed-loop transfer functions are different from that given by Equation (25), the step-response curves may look quite different from those shown in Figure 7.

3.3 Definitions of Transient-Response Specifications

Frequently, the performance characteristics of a control system are specified in terms of the transient response to a unit-step input, since it is easy to generate and is sufficiently drastic.

The transient response of a system to a unit-step input depends on the initial conditions. For convenience in comparing transient responses of various systems, it is a common practice to use the standard initial condition that the system is at rest initially with the output and all time derivatives thereof zero. Then the response characteristics of many systems can be easily compared.

The transient response of a practical control system often exhibits damped oscillations before reaching steady state, see Figure 8. In specifying the transient-response characteristics of a control system to a unit-step input, it is common to specify the following:

1. Delay time, t_d. The delay time is the time required for the response to reach half the final value the very first time.

2. Rise time t_r. The rise time is 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 secondorder systems, the 0% to 100% rise time is normally used. For overdamped systems, the 10% to 90% rise time is commonly used.

3. Peak time, t_p. The peak time is the time required for the response to reach the first peak of the overshoot.

4. Settling time, t_s. The settling time is 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.

5. Maximum overshoot, M_p. The maximum overshoot is the maximum peak value of the response curve measured from unity. If the final steady-state value of the response differs from unity, then it is common to use the maximum percent overshoot. It is defined by

$$\text{Maximum percent overshoot } = \frac{c(t_p) - c(\infty)}{c(\infty)} \times 100\% \tag{37}$$

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

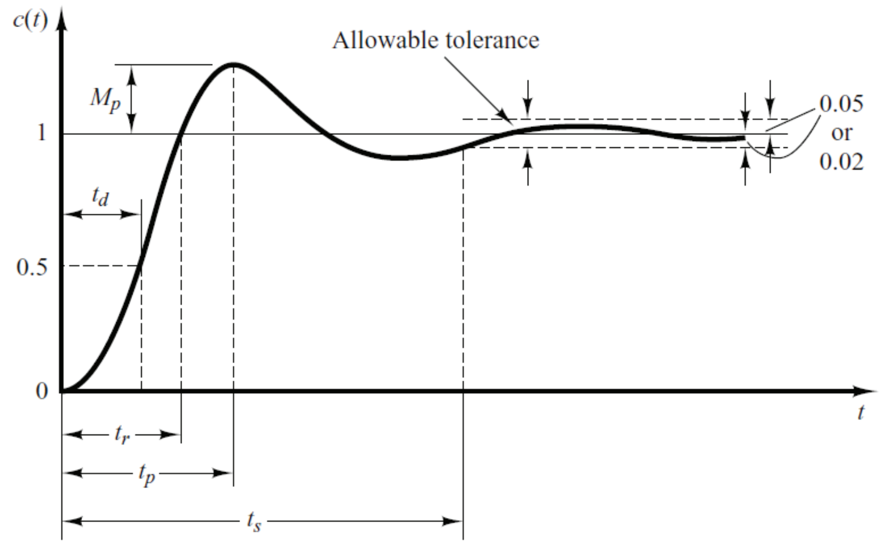

Figure 8: Unit-step response curve showing t_d, t_r, t_p, t_s, and M_p.

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 not all these specifications necessarily apply to any given case. For example, for an overdamped system, the terms peak time and maximum overshoot do not apply.

Except for certain applications where oscillations cannot be tolerated, it is desirable that the transient response be sufficiently fast and 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. Small values of ζ (that is, $\zeta < 0.4$) yield excessive overshoot in the transient response, and a system with a large value of ζ (that is, $\zeta > 0.8$) responds sluggishly.

3.4 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 given by Equation (25). These values will be obtained in terms of ζ and ω_n. The system is assumed to be underdamped.

Rise time t_r: Referring to Equation (30), we obtain the rise time t_r by letting $c(t_r) = 1$.

$$c(t_r) = 1 = 1 - e^{\zeta\omega_n t_r}\left(\cos\omega_n t_r + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t_r\right) \tag{38}$$

Since $e^{-\zeta\omega_n t} \neq 0$, we obtain from Equation (38) the following equation:

$$\cos\omega_n t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t_r = 0 \tag{39}$$

Since $\omega_n\sqrt{1-\zeta^2} = \omega_d$ and $\zeta\omega_n = \sigma$, we have

$$\tan\omega_d t_r = -\frac{\sqrt{1-\zeta^2}}{\zeta} = -\frac{\omega_d}{\sigma} \tag{40}$$

Thus, the rise time t_r is

$$t_r = \frac{1}{\omega_d}\tan^{-1}\left(-\frac{\omega_d}{\sigma}\right) = \frac{\pi-\beta}{\omega_d} \tag{41}$$

where angle β is defined in Figure 9. Clearly, for a small value of t_r , ω_d must be large.

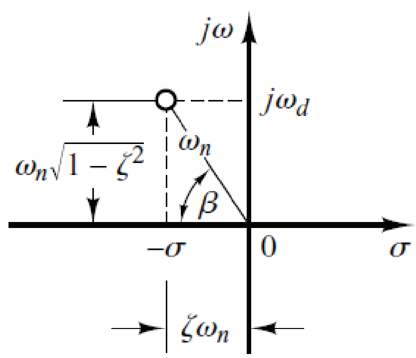

Figure 9: Definition of the angle β.

Peak time t_p: Referring to Equation (30), we may obtain the peak time by differentiating $c(t)$ with respect to time and letting this derivative equal zero. Since

$$\frac{dc}{dt} = \zeta\omega_n e^{-\zeta\omega_n t}\left(\cos\omega_d t + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\omega_d t\right) + e^{-\zeta\omega_n t}\left(\sin\omega_d t - \frac{\zeta\omega_d}{\sqrt{1-\zeta^2}}\cos\omega_d t\right) \tag{42}$$

and the cosine terms in this last equation cancel each other, dc/dt, evaluated at $t = t_p$, can be simplified to

$$\left.\frac{dc}{dt}\right|_{t=t_p} = \sin\omega_d t_p \frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t_p} = 0 \tag{43}$$

This last equation yields

$$\sin\omega_d t_p = 0 \Rightarrow \omega_d t_p = 0, \pi, 2\pi, \cdots \tag{44}$$

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

$$t_p = \frac{\pi}{\omega_d} \tag{45}$$

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

Settling time t_s: For an underdamped second-order system, the transient response is obtained from Equation (30) as

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin\left(\omega_d t + \tan^{-1}\frac{1-\zeta^2}{\zeta}\right), \; t \geq 0 \tag{46}$$

The curves $1 \pm (e^{-\zeta\omega_n t/\sqrt{1-\zeta^2}})$ are the envelope curves of the transient response to a unit-step input. The response curve $c(t)$ always remains within a pair of the envelope curves, as shown in Figure 10. The time constant of these envelope curves is $1/\zeta\omega_n$.

The speed of decay of the transient response depends on the value of the time constant $1/\zeta\omega_n$. For a given ω_n, the settling time t_s is a function of the damping ratio ζ. From Figure 7, we see that for the same ω_n and for a range of ζ between 0 and 1 the settling time t_s for a very lightly damped system is larger than that for a properly damped system. For an overdamped system, the settling time t_s becomes large because of the sluggish response.

For convenience in comparing the responses of systems, we commonly define the settling time t_s to be

$$t_s = \begin{cases} 4T = \frac{4}{\zeta\omega_n}, & 2\% \text{ criterion} \\ 3T = \frac{3}{\zeta\omega_n}, & 5\% \text{ criterion} \end{cases} \tag{47}$$

Note that the settling time is inversely proportional to the product of the damping ratio and the undamped natural frequency of the system. Since the value of ζ is usually determined from the requirement of permissible maximum overshoot, the settling time is determined primarily by the undamped natural frequency ω_n. This means that the duration of the transient period may be varied, without changing the maximum overshoot, by adjusting the undamped natural frequency ω_n.

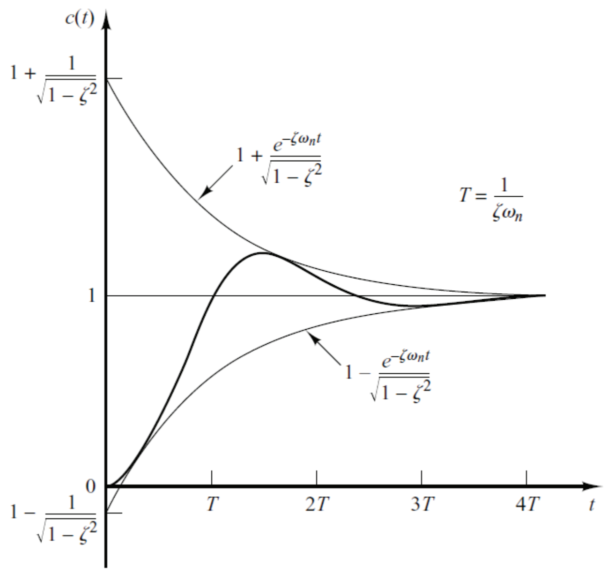

Figure 10: Pair of envelope curves for the unit-step response curve of the system shown in Figure 6.

Maximum overshoot M_p: The maximum overshoot occurs at the peak time or at $t = t_p = \pi/\omega_d$. Assuming that the final value of the output is unity, M_p is obtained from Equation (30) as

$$\begin{aligned} M_p = c(t_p) - 1 &= -e^{-\zeta\omega_n(\pi/\omega_d)}\left(\cos\pi + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin\pi\right) \\ &= e^{-(\sigma\omega_d)\pi} = e^{-(\zeta/\sqrt{1-\zeta^2})\pi} \end{aligned} \tag{48}$$

For rapid response ω_n must be large. To limit the maximum overshoot M_p and to make the settling time small, the damping ratio ζ should not be too small. The relationship between the maximum percent overshoot M_p and the damping ratio ζ is presented in Figure 11. Note that if the damping ratio is between 0.4 and 0.7, then the maximum percent overshoot for step response is between 25% and 4%.

3.5 Impulse Response of Second-Order Systems

For a unit-impulse input $r(t)$, the corresponding Laplace transform is unity, or $R(s) = 1$. The unit-impulse response $C(s)$ of the second-order system shown in Figure 6 is

$$C(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{49}$$

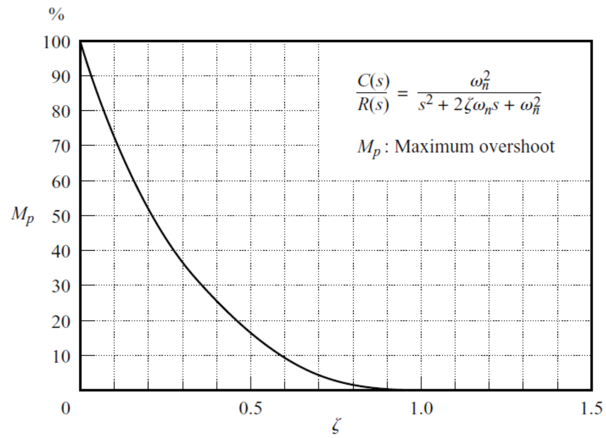

Figure 11: M_p versus ζ curve.

The inverse Laplace transform of this equation yields the time solution for the response $c(t)$ as follows:

For $0 \geq \zeta < 1$,

$$c(t) = \frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t} \sin \omega_n \sqrt{1-\zeta^2} t \tag{50}$$

For $\zeta = 1$

$$c(t) = \omega_n^2 t e^{-\omega_n t} \tag{51}$$

For $\zeta > 1$

$$c(t) = \frac{\omega_n}{2\sqrt{\zeta^2-1}} e^{-(\zeta-\sqrt{\zeta^2-1})\omega_n t} - \frac{\omega_n}{2\sqrt{\zeta^2-1}} e^{-(\zeta+\sqrt{\zeta^2-1})\omega_n t} \tag{52}$$

Note that without taking the inverse Laplace transform of $C(s)$ we can also obtain the time response $c(t)$ by differentiating the corresponding unit-step response, since the unit-impulse function is the time derivative of the unit-step function. A family of unit-impulse response curves given by Equations (50) and (51) with various values of ζ is shown in Figure 12. The curves $c(t)/\omega_n$ are plotted against the dimensionless variable $\omega_n t$, and thus they are functions only of ζ. For the critically damped and overdamped cases, the unit-impulse response is always positive or zero; that is, $c(t) \geq 0$. This can be seen from Equations (51) and (52). For the underdamped case, the unit-impulse response $c(t)$ oscillates about zero and takes both positive and negative values. The maximum overshoot for the unit-impulse response of the underdamped system occurs at,

$$t = \frac{\tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta}}{\omega_n \sqrt{1-\zeta^2}}, \ 0 < \zeta < 1 \tag{53}$$

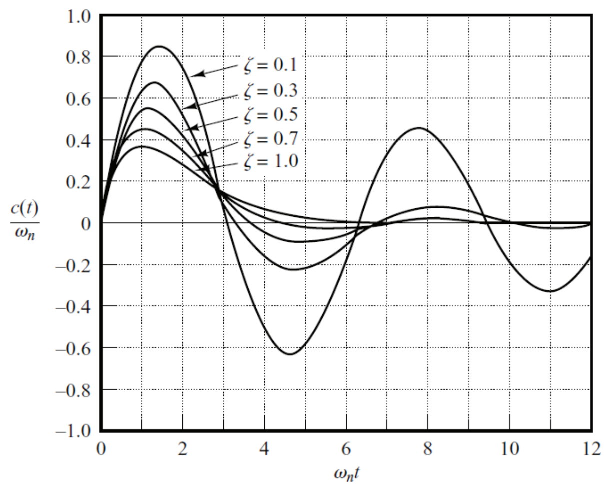

Figure 12: Unit-impulse response curves of the system shown in Figure 6.

by equating dc/dt to zero and solving for t. The maximum overshoot is, by substituting Equation (53) into Equation (50)

$$c(t)_{max} = \omega_n e^{-\left(\frac{\zeta}{\sqrt{1-\zeta^2}} \tan^{-1} \frac{\sqrt{1-\zeta^2}}{\zeta}\right)}, \quad 0 < \zeta < 1 \tag{54}$$

Since the unit-impulse response function is the time derivative of the unit-step response function, the maximum overshoot M_p for the unit-step response can be found from the corresponding unit-impulse response. That is, the area under the unitimpulse response curve from $t = 0$ to the time of the first zero, as shown in Figure 13, is $1 + M_p$, where M_p is the maximum overshoot (for the unit-step response) given by Equation (48). The peak time t_p (for the unit-step response) given by Equation (48) corresponds to the time that the unit-impulse response first crosses the time axis.

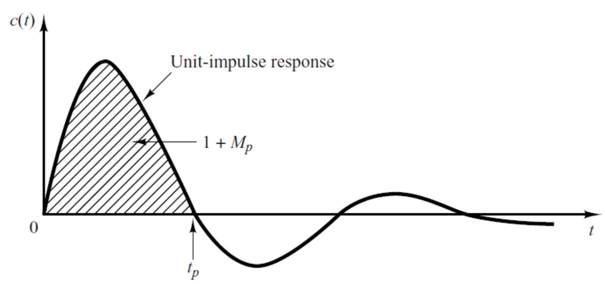

Figure 13: Unit-impulse response curve of the system shown in Figure 6.