Contents

Chapter 3

Analysis of linear continuous systems

3.1 System response

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

In analyzing and designing control systems we must have a basis for performance comparison 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, (**?**), (**?**).

In the following we will consider systems described by transfer functions $H(s)$ with one known input signal $r(t)$. The output $c(t)$ will be computed from:

$$c(t) = L^{-1}[H(s)R(s)]$$

3.1.1 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.

3.1.2 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 3.1. By transient response we mean that which goes from the initial to the final state. By steady-state response we mean the manner in which the system output behaves as t approaches infinity.

In designing a control system, we must be able to predict the dynamic behavior of the system from a knowledge of the components.

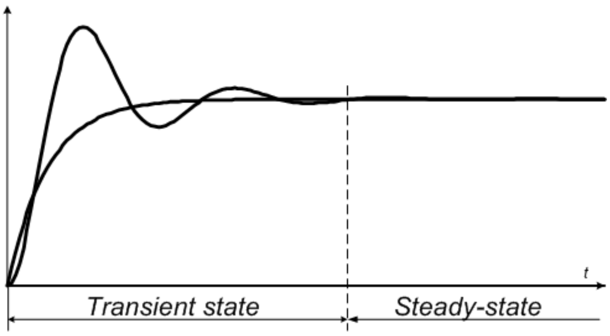


Figure 3.1: Transient and steady-state response

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.

The behavior of simple systems 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 systems, these are differential equations.
- Determine the transfer function or transfer matrix and identify the system 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.

3.2 First-order systems

Consider a first-order system shown in Figure 3.2.

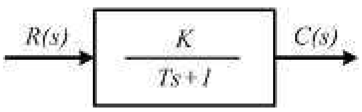


Figure 3.2: First-order system

Physically, this system may represent an RC circuit, a thermal system, or the like. The input-output model, or the transfer function, is given by:

$$H(s) = \frac{C(s)}{R(s)} = \frac{K}{Ts+1} \tag{3.1}$$

In equation (3.1), the constant K is the *gain*, and the coefficient of s in the first-order polynomial in the denominator, T, is the *time constant*. In this section, the time constant T is a positive real number.

Example 3.2.1 *Electrical system. Consider the RC circuit shown in Figure 3.3. The differential equation*

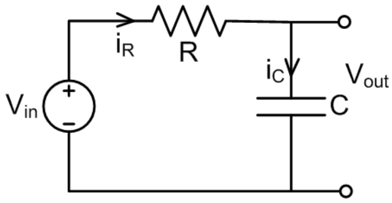


Figure 3.3: RC circuit

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

$$V_{in}(t) = V_R(t) + V_{out}(t)$$

where

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

and we obtain:

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

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

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

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

Example 3.2.2 *Mechanical system. Consider the mechanical system shown in Figure 3.4, where m is the mass of the car, $u(t)$ is an external force (the input signal), $y(t)$ is the velocity of the car (the output signal) and b is the friction coefficient.*

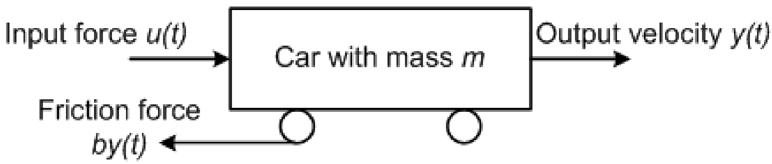


Figure 3.4: Mechanical system

If $y(t)$ is the velocity of the car, the first derivative $dy(t)/dt$ is the acceleration. Then, from the equilibrium of forces, we have the following differential equation:

$$m\frac{dy(t)}{dt} = u(t) - by(t) \text{ or } m\frac{dy(t)}{dt} + by(t) = u(t)$$

If we apply the Laplace transform, considering the initial conditions to be zero, the transfer function results as:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{1}{ms+b}$$

The gain and time constant of the transfer function obtained above can be determined by dividing the denominator by b. Thus:

$$H(s) = \frac{\frac{1}{b}}{\frac{m}{b}s + 1}$$

The gain is $K = 1/b$ and the time constant $T = m/b$.

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.

3.2.1 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 (3.1), 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 the system response as:

$$c(t) = K\left(1 - e^{-t/T}\right), \quad (t \geq 0) \tag{3.2}$$

The exponential response curve $c(t)$ given by equation (3.2) is shown in Figure 3.5.

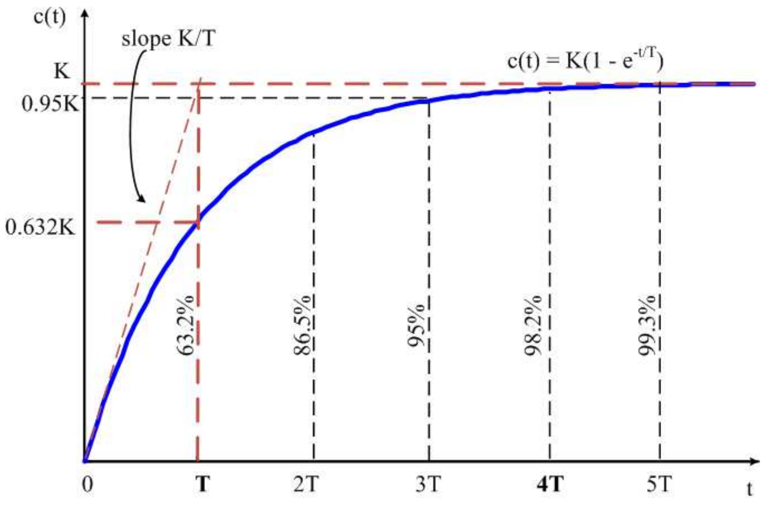


Figure 3.5: First-order system step response

Equation (3.2) 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:

$$c(T) = K(1 - e^{-1}) = K(1 - 0.368) = 0.632K$$

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 (3.2), 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. This can be seen by substituting $t = 4T$ in $c(t)$ given by relation (3.2):

$$c(4T) = K(1 - e^{-4T/T}) = K(1 - e^{-4}) = K(1 - 0.018) = 0.982K$$

The time required for the output to reach 98% from its final value (and to remain within 2% of its final value) is called **the settling time**. For a first-order system it is equal to 4 time constants:

$$t_s = 4T$$

Another important characteristic of the exponential response curve is that the slope of the tangent at $t = 0$ is K/T, since:

$$\frac{dc(t)}{dt} = \frac{K}{T} e^{-t/T}|_{t=0} = \frac{K}{T} \tag{3.3}$$

If the line crossing the origin and having this initial slope is extended until it intersects the steady-state asymptote (see Figure 3.5), the intersection will occur exactly at one time constant $t = T$.

From equation (3.3) we see that the slope of the response curve $c(t)$ decreases monotonically from K/T at $t = 0$ to zero at $t = \infty$.

Notice that the smaller the time constant T, the faster the system response (see Figure 3.6).

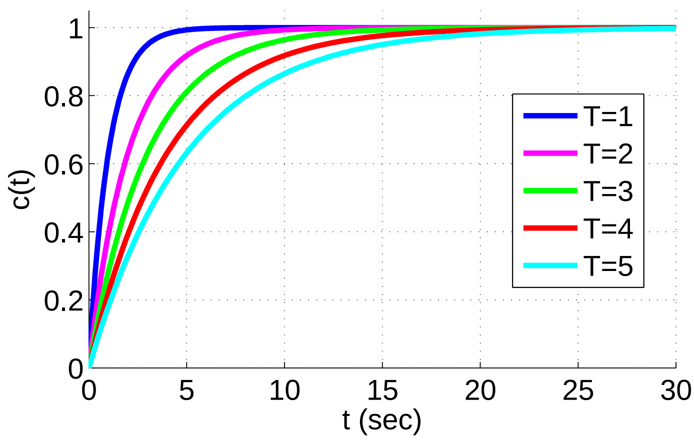


Figure 3.6: First-order system step response for various time constants

3.2.2 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 3.2 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:

$$c(t) = K\left(t - T + Te^{-t/T}\right), \quad (t \geq 0) \tag{3.4}$$

A plot of the unit ramp response of a general first-order system is shown in Figure 3.7.

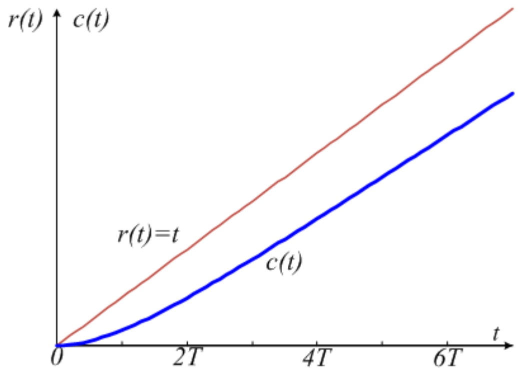


Figure 3.7: First-order system ramp response

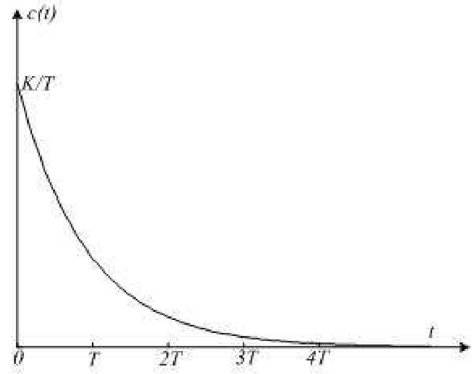


Figure 3.8: First-order system impulse response

When time approaches infinity $t \to \infty$, the exponential term in (3.4) will approach 0 and the system response will follow asymptotically a line with the equation:

$$c(t) = K\,(t - T)$$

In practice, the steady-state of the system response occurs also at 4 time constants, similar to the step response of the first-order system, because at this moment of time the exponential term is sufficiently close to zero:

$$e^{-4T/T} = e^{-4} = 0.0183$$

thus, the settling time is $t_s = 4T$.

3.2.3 Unit-impulse response of first-order systems

For the unit-impulse input, $R(s) = 1$ and the output of the system of Figure 3.2 can be obtained as:

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

or,

$$c(t) = \frac{K}{T}e^{-t/T}, \quad (t \geq 0) \tag{3.5}$$

The response curve is shown in Figure 3.8.

From (3.5), the response curve starts at K/T for the initial time $t = 0$ and approaches 0 as $t \to \infty$. The steady-state can be considered reached at the settling time $t_s = 4T$.

3.2.4 A note on the system gain

Consider a first-order system with the transfer function

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

where the gain is 1, and a first-order system with the gain K:

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

When the input is a unit step $R(s) = 1/s$, the step response for the first system is:

$$c(t) = L^{-1}[H(s)R(s)] = L^{-1}[\frac{H(s)}{s}] \tag{3.6}$$

and

$$c_k(t) = L^{-1}[H_k(s) \cdot R(s)] = L^{-1}[\frac{K \cdot H(s)}{s}] = K \cdot L^{-1}[\frac{H(s)}{s}] = K \cdot c(t) \tag{3.7}$$

for the second system.

Thus, the gain constant will influence the system response only in magnitude. Consider, for example a first-order system with the transfer function:

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

The system has a time constant $T = 1$ and the settling time: $t_s = 4T = 4sec$, for any value of K. The system responses for a unit step input and values of $K = 1, 2, 3, 4$ are illustrated in Figure 3.9. The steady-state value of the output, for a unit step input is K.

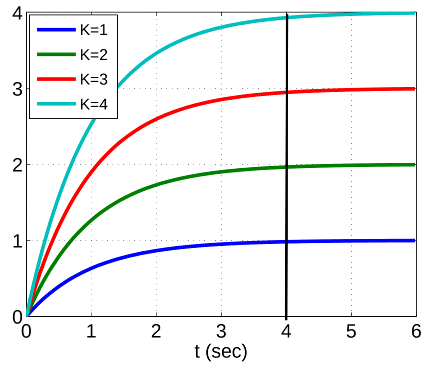


Figure 3.9: Step response for various values of K

Relations (3.6), (3.7) are actually valid for any system described by a transfer function, not only for first-order systems. Therefore, in the following section, the gain of the system is taken as equal to 1 to simplify the calculus. A different gain will not change the characteristics of the transient response, it will influence only the magnitude of the response (proportional to K) and the steady-state value.

3.3 Second-order systems

Consider the second-order system shown in Figure 3.10 where the system transfer function is written in the

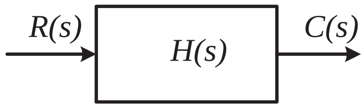


Figure 3.10: Second-order system

generalized form:

$$H(s) = \frac{C(s)}{R(s)} = \frac{1}{\frac{1}{\omega_n^2}s^2 + \frac{2\zeta}{\omega_n}s + 1} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{3.8}$$

The system gain is 1, but any other value of the gain will change only the magnitude of the system response, as discussed in section 3.2.4.

All coefficients of the polynomial in the denominator are positive. The special case when some of them have negative values will be discussed in the section presenting the problem of system stability.

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 3.3.1 *Consider the system with the transfer function:*

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

The natural frequency and the damping factor are calculated from:

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

The roots of the characteristic equation of the second order system:

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

(or the poles of the system) are:

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

The poles of a second-order system in the form (3.8) are:

- complex conjugates for $0 < \zeta < 1$ and lie in the left half s-plane. The system is called **underdamped** and the transient response is oscillatory
- complex conjugates on the imaginary axis for $\zeta = 0$. The system is called **undamped** and the transient response does not die out.
- real for $\zeta \geq 1$ and the system is called **overdamped**. If $\zeta = 1$ the system is **critically damped**. The transient response of critically damped and overdamped systems do not oscillate.

3.3.1 Step-response of second-order systems

The transfer function of the second-order system is:

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

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

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

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

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

   $$C(s) = \frac{1}{s} - \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2} = \frac{1}{s} - \frac{s + \zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} - \frac{\zeta\omega_n}{(s + \zeta\omega_n)^2 + \omega_d^2} \tag{3.9}$$

   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 (3.9) is obtained as:

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

   or, after some calculations:

   $$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1 - \zeta^2}} \cdot \sin\left(\omega_d t + arctan\frac{\sqrt{1 - \zeta^2}}{\zeta}\right), \quad (t \geq 0) \tag{3.10}$$

   From equation (3.10), the frequency of transient oscillation is the damped natural frequency $\omega_d = \omega_n\sqrt{1 - \zeta^2}$ and thus varies with the damping ratio ζ and the natural frequency ω_n. Since for the underdamped case ζ is positive and smaller than 1, the frequency of oscillations is mainly influenced by ω_n. See for example the step responses of second-order systems for various values of ω_n when ζ is kept constant, illustrated in Figure 3.11. Notice also that a variable ω_n does not change the maximum value of the system output.

   The step responses of second order systems when ω_n is constant and ζ takes various values between 0 and 1 is presented in Figure 3.12. The frequency of oscillations is slightly changed by the variable ζ, but the maximum value of the system response depends on the damping factor.

2. **Undamped case** $(\zeta = 0)$. If the damping ratio ζ is equal to zero, the response is undamped and oscillations continue indefinitely. The transfer function is:

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

   and it has two complex poles on the imaginary axis $s_{1,2} = \pm\omega_n j$. The response $c(t)$ for the zero damping case may be obtained by substituting $\zeta = 0$ in equation (3.9), yielding:

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

   or, taking the inverse Laplace transform:

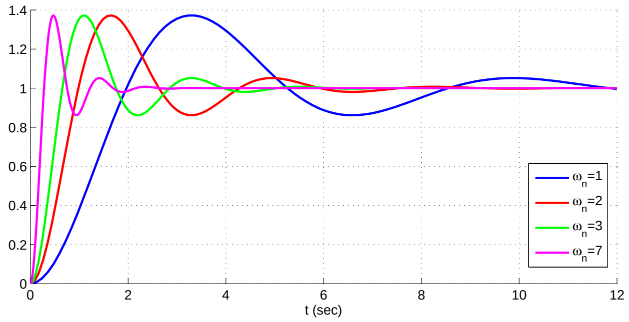


Figure 3.11: Step response of an underdamped second order system for various values of ω_n

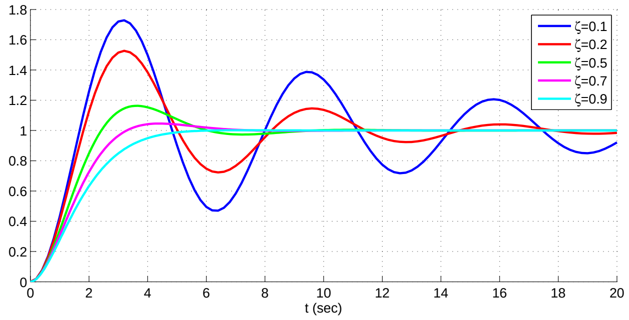


Figure 3.12: Step response of an underdamped second order system for various values of ζ

$$c(t) = 1 - \cos \omega_n t, \quad (t \geq 0) \tag{3.11}$$

Thus, from equation (3.11) we see that ω_n is 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.

The step responses of three undamped second-order systems ($\zeta = 0$) for various values of the natural frequency ω_n are illustrated in Figure 3.13.

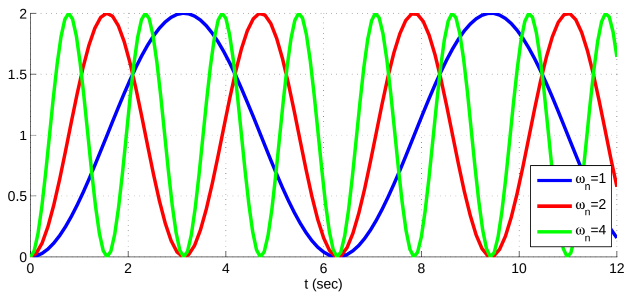


Figure 3.13: Step response of an undamped second order system for various values of ω_n

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

   If $\zeta = 1$ the transfer function of the second-order system is:

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

   The system poles are real and equal: $s_1 = s_2 = -\omega_n$ and the system is critically damped.

For a unit step input $R(s) = 1/s$, the output $C(s)$ can be written as:

$$C(s) = \frac{\omega_n^2}{(s + \omega_n)^2 s} \tag{3.12}$$

The inverse Laplace transform of equation (3.12) may be found as:

$$c(t) = 1 - e^{-\omega_n t}(1 - \omega_n t), \quad , (t \geq 0) \tag{3.13}$$

The step response of a critically damped system ($\zeta = 1$) is shown in Figure 3.14 for various values of ω_n.

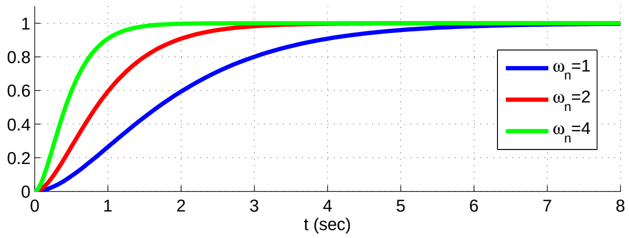


Figure 3.14: Step response of critically damped second order system for various values of ω_n

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

In this case, the two poles of $H(s) = C(s)/R(s)$ are negative real and unequal:

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

For a unit step input, the output $C(s)$ can be written:

$$C(s) = \frac{\omega_n^2}{s(s - s_1)(s - s_2)} \tag{3.14}$$

By partial fraction expansion the output (3.14) is:

$$C(s) = \frac{\omega_n^2}{s_1 s_2}\frac{1}{s} + \frac{\omega_n^2}{s_1(s_1 - s_2)}\frac{1}{s - s_1} - \frac{\omega_n^2}{s_2(s_1 - s_2)}\frac{1}{s - s_2}$$

To simplify the relation, compute:

$$s_1 s_2 = (-\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1})(-\zeta\omega_n - \omega_n\sqrt{\zeta^2 - 1}) = \omega_n^2$$

and:

$$s_1 - s_2 = -\zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1} + \zeta\omega_n + \omega_n\sqrt{\zeta^2 - 1} = 2\omega_n\sqrt{\zeta^2 - 1}$$

then, $C(s)$ becomes:

$$C(s) = \frac{1}{s} + \frac{\omega_n}{2\sqrt{\zeta^2 - 1}}\left(\frac{1}{s_1(s - s_1)} - \frac{1}{s_2(s - s_2)}\right)$$

Taking the inverse Laplace transform, the system output $c(t)$ is:

$$c(t) = 1 + \frac{\omega_n}{2\sqrt{\zeta^2 - 1}}\left(\frac{e^{s_1 t}}{s_1} - \frac{e^{s_2 t}}{s_2}\right) \tag{3.15}$$

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

The step response of an overdamped second order system is illustrated in Figure 3.15 for various valued of the damping factor ζ.

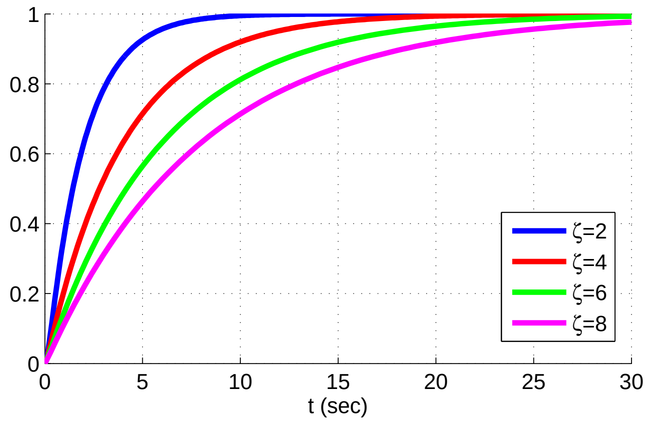


Figure 3.15: Step response of an overdamped second order system for various values of ζ

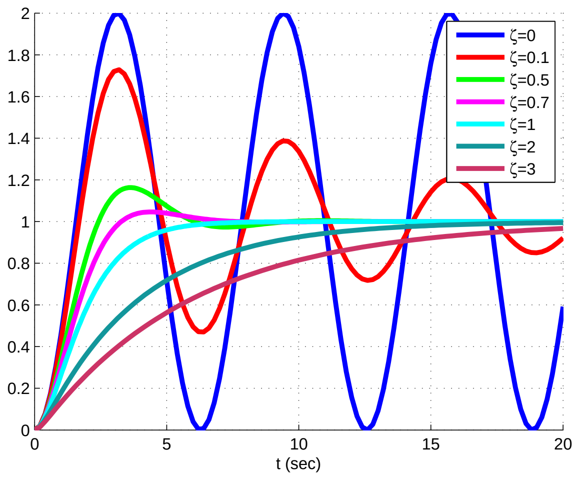


Figure 3.16: Step response of a second-order system for various values of damping ratio

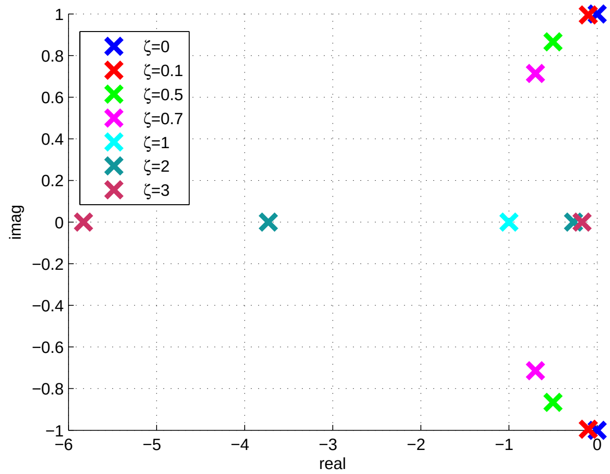


Figure 3.17: Poles of a second-order system for various values of damping ratio

All cases presented above for a second-order system are summarized in Figure 3.16 for various values of the damping ratio ζ, when the natural frequency is constant $\omega_n = 1$.

As ζ decreases, the system poles approach the imaginary axis (see Figure 3.17) and the response becomes increasingly oscillatory.

3.4 The unit step response and transient-response specifications

In many practical cases, the desired performance characteristics of control systems are specified in terms of time-domain quantities. The following transient response characteristics to a unit step input are commonly used:

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 3.18.

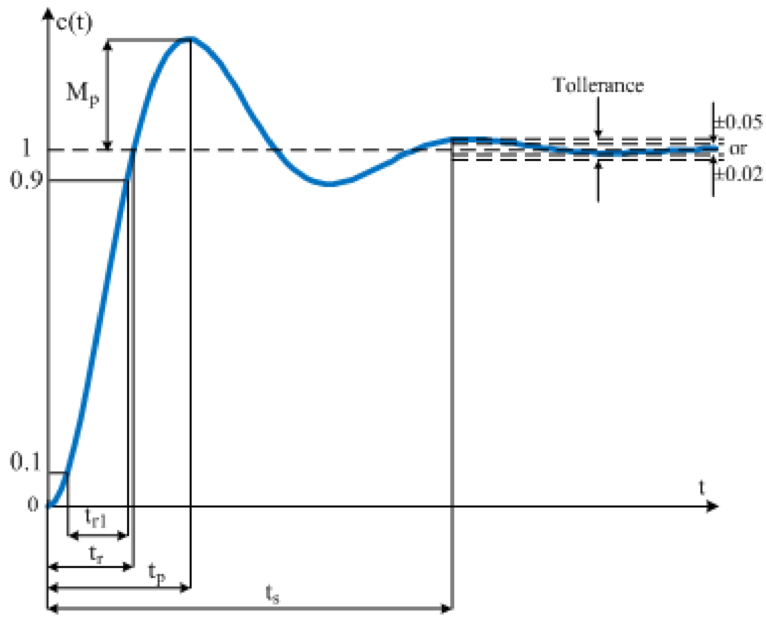


Figure 3.18: Second-order system underdamped response

1. Rise time, t_r : the time required for the response to rise from 10% to 90%, or 0% to 100% of its final value. For underdamped 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, or maximum, peak. The peak time is not defined for overdamped systems.

3. Maximum overshoot M_p : the maximum peak value of the response curve measured from the steady-state value of the response. Very often the overshoot is expressed as percentage of the steady-state value (**percent overshoot,** $M_{p\%}$) and computed from:

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

where $c(\infty)$ is the final (steady-state) value of the output.

4. Settling time, t_s : the time required for the response curve to reach and stay within a range about the final value, usually 2% or 5% of the final value.

3.4.1 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 step response of an underdamped second-order system. The values will be obtained in terms of ζ and ω_n.

1. Rise time, t_r . Referring to equation (3.10) we obtain the rise time t_r by letting $c(t_r) = 1$ or

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

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

$$\sin\left(\omega_d t_r + arctan\frac{\sqrt{1-\zeta^2}}{\zeta}\right) = 0$$

or

$$\omega_d t_r + arctan\frac{\sqrt{1-\zeta^2}}{\zeta} = k\pi, \quad k \in \mathbf{Z} \tag{3.16}$$

The first value of k from (3.16) for which $t_r > 0$ is $k = 1$.

$$\omega_d t_r + arctan\frac{\sqrt{1-\zeta^2}}{\zeta} = \pi \quad \Rightarrow \quad \omega_d t_r = \pi - arctan\frac{\sqrt{1-\zeta^2}}{\zeta}$$

Thus, the rise time t_r is:

$$t_r = \frac{1}{\omega_d} \cdot \left(\pi - arctan\frac{\sqrt{1-\zeta^2}}{\zeta}\right) = \frac{\pi - \beta}{\omega_d} \tag{3.17}$$

where $\beta = arctan\frac{\sqrt{1-\zeta^2}}{\zeta}$ is defined in Figure 3.19. The poles of an underdamped second-order system are $s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$. The real part of the poles is negative and equal to $-\zeta\omega_n$ and the imaginary part is $\omega_d = \omega_n\sqrt{1-\zeta^2}$, as presented in Figure 3.19. The angle β is the angle between the negative real axis and the line that connects the pole s_1 to the origin.

2. Peak time, t_p . From equation (3.10) we may obtain the peak time by differentiating $c(t)$ with respect to time and letting the derivative equal zero (the peak time is the time of the first maximum). The output signal $c(t)$ is:

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

where $\beta = arctan\frac{\sqrt{1-\zeta^2}}{\zeta}$, as shown in Figure 3.19. Notice also, from the same figure that: $\cos\zeta = \zeta$ and $\sin\beta = \sqrt{1-\zeta^2}$.

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

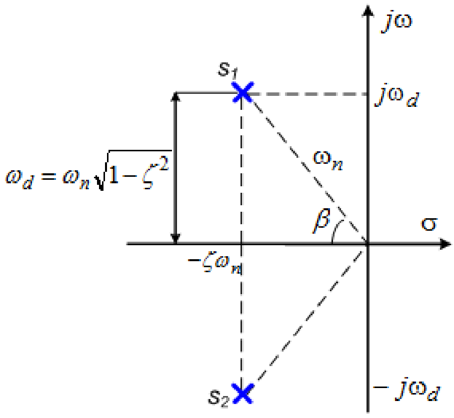


Figure 3.19: Second-order system poles

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

The derivative is zero at the peak time t_p and we obtain:

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

This yields the following equation:

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

or

$$\omega_d t_p = k\pi, \quad k = 0, 1, 2, ...$$

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

$$t_p = \frac{\pi}{\omega_d} \tag{3.18}$$

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

3. **Maximum overshoot,** M_p , occurs at the peak time or at $t = t_p = \frac{\pi}{\omega_d}$. The steady-state value of the output is $c(\infty) = 1$. Thus, from equation (3.10) the value of the maximum overshoot M_p is obtained as the difference between the maximum value and the steady-state value of the output:

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

$$M_p = -\frac{e^{-\pi\zeta/\sqrt{1-\zeta^2}}}{\sqrt{1-\zeta^2}}\sin(\pi+\beta) = -\frac{e^{-\pi\zeta/\sqrt{1-\zeta^2}}}{\sqrt{1-\zeta^2}}(-\sin\beta) = \frac{e^{-\pi\zeta/\sqrt{1-\zeta^2}}}{\sqrt{1-\zeta^2}}\sqrt{1-\zeta^2}$$

or

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

For a second-order system having the gain 1 and a unit step input, the percent overshoot is:

$$M_{p\%} = \frac{c(t_p) - c(\infty)}{c(\infty)}\cdot 100\% = \frac{c(t_p)-1}{1}\cdot 100\% = e^{-\pi\zeta/\sqrt{1-\zeta^2}}\cdot 100\% \tag{3.19}$$

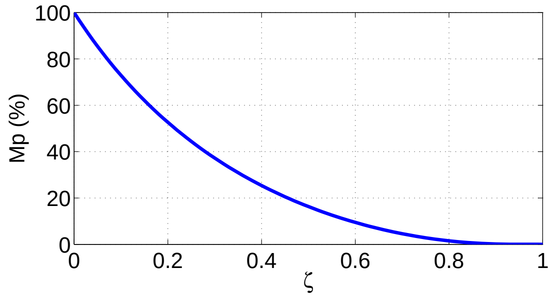


Figure 3.20: M_p versus ζ

The overshoot is a function only of the damping ratio ζ and the relationship between them is shown in Figure 3.20. As illustrated in the figure, the percent overshoot decreases as the damping ratio increases.

4. **Settling time,** t_s . For an underdamped second-order system, the transient response is obtained from equation (3.10):

$$c(t) = 1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \cdot \sin(\omega_d t + \beta) \tag{3.20}$$

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 3.21. The time it takes the decaying sinusoid in relation (3.20) to reach 2% of its steady-state value is

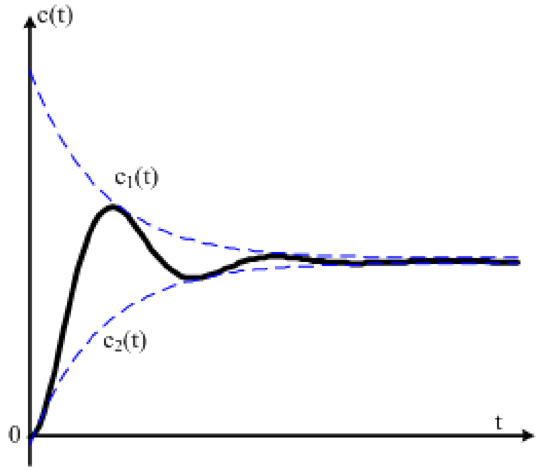


Figure 3.21: Step response of an underdamped second-order system and the envelope curves

the settling time. It is the time when the envelope curves reach 0.02, or:

$$\frac{e^{-\zeta\omega_n t_s}}{\sqrt{1-\zeta^2}} = 0.02$$

Then:

$$t_s = \frac{-\ln 0.02\sqrt{1-\zeta^2}}{\zeta\omega_n}$$

As ζ varies between 0 and 0.9, the expression $-\ln 0.02\sqrt{1-\zeta^2}$ varies between 3.9 and 4.7, (**?**). Then, a common approximation of the settling time is:

$$t_s = \frac{4}{\zeta\omega_n} \tag{3.21}$$

Example 3.4.1 *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 system poles are:*

$$s_{1,2} = -\zeta\omega_n \pm \sqrt{1-\zeta^2}j = -3 \pm 4j$$

2. *The damped natural frequency (the imaginary part of the poles) is:*

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

   and the poles negative real part:

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

   According to Figure 3.19, the angle β is:

$$\beta = arctan\frac{\omega_d}{\zeta\omega_n} = 0.93\ (rad)$$

3. *The rise time is:*

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

4. *The peak time is:*

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

5. *Maximum overshoot will be:*

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

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

6. *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 3.22 is shown the step response of the system. The values of the system parameters can be seen also from the plot.

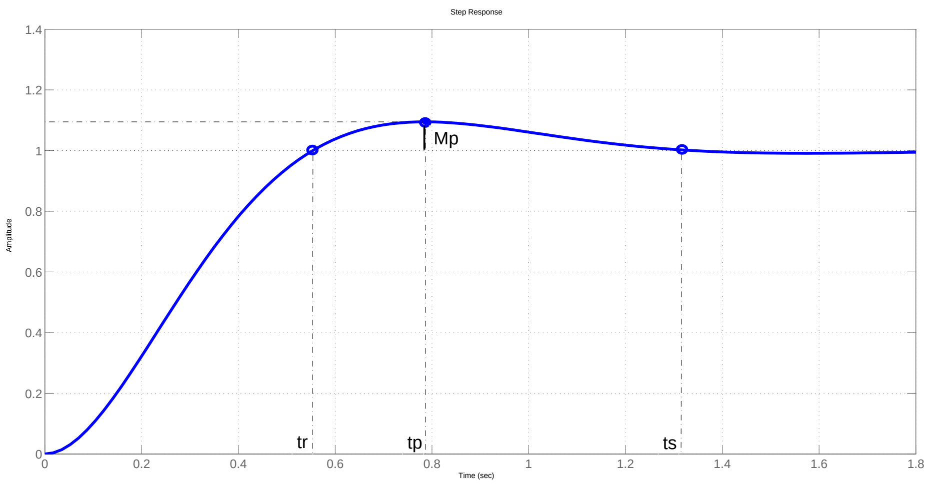


Figure 3.22: Step response

3.5 Steady-state error

3.5.1 Introduction. The final value theorem

One of the objectives of most control systems is that the system output response follows a specific reference signal accurately in the steady state. If the output of a system at steady state does not exactly agree with the input, the difference between the reference signal and the output in the steady state is defined as **the steady-state error**, (**?**). This error is indicative of the accuracy of the system. Errors in control systems can be attributed to many factors, for example because of friction and other imperfections, or the natural structure of the system, (**??**).

The steady-state error can be computed only for systems that have a steady state, i.e. are stable. Stability analysis of linear continuous systems will be presented in 3.10, but consider a stable system one that has all poles with negative real parts, or located in the left half s-plane.

Example 3.5.1 *The unit step response of three systems is presented in Figure 3.23, together with the step input (green). The difference between the prescribed input signal (unit step) and the the output at steady state is the steady-state error.*

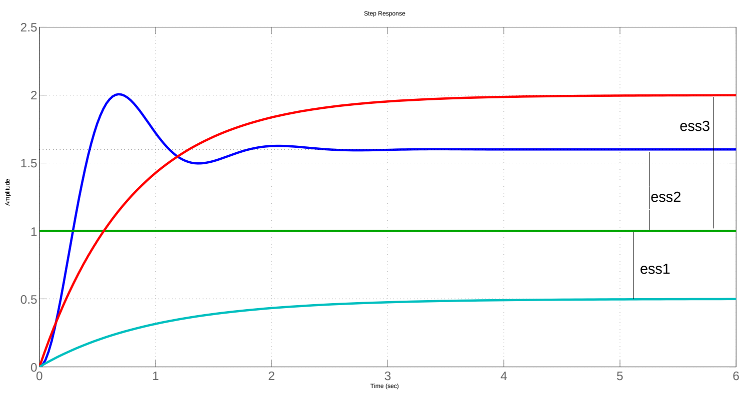


Figure 3.23: Steady-state error for a unit step input

For a 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 value of the error, e_{ss} is:

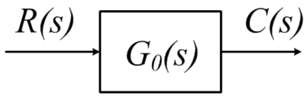


Figure 3.24: A system

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

The Laplace transform of the error signal for the system shown in Figure 3.24 is:

$$E(s) = R(s) - C(s) = R(s) - G_0(s)R(s) = (1 - G_0(s))R(s)$$

The *final value theorem* states that:

$$\text{if } \lim_{t\to\infty} e(t) \text{ exists, then: } \lim_{t\to\infty} e(t) = \lim_{s\to 0} sE(s)$$

For the system in Figure 3.24 we obtain:

$$e_{ss} = \lim_{s\to 0} s(1 - G_0(s))R(s)$$

3.5.2 Steady-state error for unity feedback systems

For unity feedback systems, the steady-state error can be calculated from the closed-loop transfer function $G_0(s)$ or the open-loop transfer function $G(s)$ (see Figure 3.25), where:

$$G_0(s) = \frac{G(s)}{1 + G(s)}$$

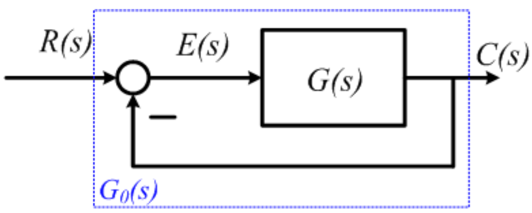


Figure 3.25: Closed-loop system

The error signal in Figure 3.25, as a function of time, is:

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

The Laplace transform of the error for unity feedback closed-loop system, is obtained from the transfer function of the closed-loop system $G_0(s)$ as:

$$E(s) = R(s) - C(s) = R(s) - G_0(s)R(s) = (1 - G_0(s))R(s)$$

or, from the open-loop transfer function:

$$E(s) = (1 - G_0(s))R(s) = \left(1 - \frac{G(s)}{1 + G(s)}\right) R(s) = \frac{1}{1 + G(s)} R(s)$$

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

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

Using the closed-loop transfer function, we obtain:

$$e_{ss} = \lim_{s \to 0} s(1 - G_0(s))R(s)$$

and using the open-loop transfer function:

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

Example 3.5.2 *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 $R(s) = 1/s$.

For the closed-loop system of Figure 3.25, the steady-state error is:

$$e_{ss} = \lim_{s \to 0} s \frac{1}{1 + G(s)} \frac{1}{s} = \frac{1}{1 + G(0)} = \frac{1}{1 + \frac{k}{T \cdot 0 + 1}} = \frac{1}{1 + k}$$

For the closed-loop system a large gain k, for example $k = 100$, will reduce the steady-state error to a small value. In this case the closed-loop system steady-state error is $e_{ss} = 1/101$.

3.5.3 System type and static error constants

Consider a unity feedback closed-loop control system (Figure 3.26) with the open-loop transfer function:

$$G(s) = \frac{k \prod_{i=1}^{M} (s + z_i)}{s^N \prod_{j=1}^{Q} (s + p_j)}$$

The transfer function $G(s)$ involves the term s^N in the denominator, that represents a pole of multiplicity N at the origin. The zeros of the transfer function: $-z_i$, $i = \overline{1, M}$ and the poles: $-p_j$, $j = \overline{1, Q}$ can be real or complex conjugate numbers.

We define the **system type** to be the value of N in the denominator, or, equivalently, the number of pure integrations in the forward path. A system with no poles at the origin, or $N = 0$ is a *Type 0* system,

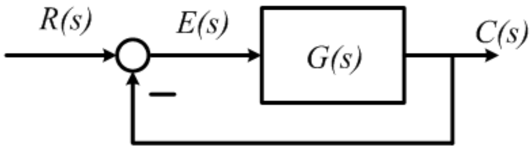


Figure 3.26: Closed-loop system

a system having one pole at the origin, or $N = 1$, is a *Type 1* system, a system with $N = 2$ is a *Type 2* system, etc.

The steady state error may be calculated according to:

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

where R(s) is the system input. For typical test inputs, the steady-state error can be written in terms of *static error constants*. In this chapter, only the unit step and unit ramp inputs will be considered.

Step input. When $R(s) = 1/s$, the steady-state error for a unit step input is:

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

where K_p is the **static position error constant**:

$$K_p = \lim_{s\to 0} G(s)$$

The steady-state error and the static error constants depend upon the type-number N.

Type 0 system: $N = 0$**.** The steady-state error for a unit step input is:

$$e_{ss} = \lim_{s\to 0}\frac{1}{1+G(s)} = \lim_{s\to 0}\frac{1}{1+\dfrac{k\prod_{i=1}^{M}(s+z_i)}{\prod_{j=1}^{Q}(s+p_j)}} = \frac{1}{1+\dfrac{k\prod_{i=1}^{M}z_i}{\prod_{j=1}^{Q}p_j}} = \text{constant}$$

and the static position error constant is:

$$K_p = \lim_{s\to 0}G(s) = \lim_{s\to 0}\frac{k\prod_{i=1}^{M}(s+z_i)}{\prod_{j=1}^{Q}(s+p_j)} = \frac{k\prod_{i=1}^{M}z_i}{\prod_{j=1}^{Q}p_j} = \text{constant}$$

Type 1 or higher: $N \geq 1$**.** The steady-state error for a unit step input is:

$$e_{ss} = \lim_{s\to 0}\frac{1}{1+G(s)} = \lim_{s\to 0}\frac{1}{1+\dfrac{k\prod_{i=1}^{M}(s+z_i)}{s^N\prod_{j=1}^{Q}(s+p_j)}} = \lim_{s\to 0}\frac{s^N\prod_{j=1}^{Q}(s+p_j)}{s^N\prod_{j=1}^{Q}(s+p_j)+k\prod_{i=1}^{M}(s+z_i)} = 0$$

and the static position error constant is:

$$K_p = \lim_{s\to 0}G(s) = \lim_{s\to 0}\frac{k\prod_{i=1}^{M}(s+z_i)}{s^N\prod_{j=1}^{Q}(s+p_j)} = \infty$$

Ramp input. When $R(s) = 1/s^2$, the steady-state error for a ramp input is:

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

where K_v is the **static velocity error constant**:

$$K_v = \lim_{s \to 0} sG(s)$$

Type 0 system: $N = 0$. The steady-state error for a ramp input is:

$$e_{ss} = \lim_{s \to 0} \frac{1}{sG(s)} = \lim_{s \to 0} \frac{1}{s\dfrac{k\prod_{i=1}^{M}(s+z_i)}{\prod_{j=1}^{Q}(s+p_j)}} = \infty$$

and the static velocity error constant is:

$$K_v = \lim_{s \to 0} sG(s) = \lim_{s \to 0} s\frac{k\prod_{i=1}^{M}(s+z_i)}{\prod_{j=1}^{Q}(s+p_j)} = 0$$

Type 1 system: $N = 1$. The steady-state error for a ramp input is:

$$e_{ss} = \lim_{s \to 0} \frac{1}{sG(s)} = \lim_{s \to 0} \frac{1}{s\dfrac{k\prod_{i=1}^{M}(s+z_i)}{s\prod_{j=1}^{Q}(s+p_j)}} = \frac{\prod_{j=1}^{Q} p_j}{k\prod_{i=1}^{M} z_i} = \text{constant}$$

and the static velocity error constant is:

$$K_v = \lim_{s \to 0} sG(s) = \lim_{s \to 0} s\frac{k\prod_{i=1}^{M}(s+z_i)}{s\prod_{j=1}^{Q}(s+p_j)} = \frac{k\prod_{i=1}^{M} z_i}{\prod_{j=1}^{Q} p_j} = \text{constant}$$

Type 2 or higher: $N \geq 2$. The steady-state error for a ramp input is:

$$e_{ss} = \frac{1}{\lim_{s \to 0} sG(s)} = \lim_{s \to 0} \frac{1}{s\dfrac{k\prod_{i=1}^{M}(s+z_i)}{s^N\prod_{j=1}^{Q}(s+p_j)}} = \lim_{s \to 0} \frac{s^{N-1}\prod_{j=1}^{Q}(s+p_j)}{k\prod_{i=1}^{M}(s+z_i)} = 0$$

and the static velocity error constant is:

$$K_v = \lim_{s \to 0} sG(s) = \lim_{s \to 0} s\frac{k\prod_{i=1}^{M}(s+z_i)}{s^N\prod_{j=1}^{Q}(s+p_j)} = \lim_{s \to 0} \frac{k\prod_{i=1}^{M}(s+z_i)}{s^{N-1}\prod_{j=1}^{Q}(s+p_j)} = \infty$$

All these results are summarized in Table 3.1.

System type	Step input	Ramp input
Type 0, $N = 0$	$K_p = const$, $e_{ss} = \dfrac{1}{1+K_p}$	$K_v = 0$, $e_{ss} = \infty$
Type 1, $N = 1$	$K_p = \infty$, $e_{ss} = 0$	$K_v = const$, $e_{ss} = \dfrac{1}{K_v}$
Type 2 or higher, $N \geq 2$	$K_p = \infty$, $e_{ss} = 0$	$K_v = \infty$, $e_{ss} = 0$

Table 3.1: Summary of static error constants and steady-state errors

Example 3.5.3 *Consider a unity negative feedback closed-loop system with the open-loop transfer function*

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

Evaluate system type, static error constants K_p, K_v and the steady-state errors for a unit step and a ramp input.

The open-loop system $G(s)$ has one pole at the origin, therefore it is a Type 1 system.

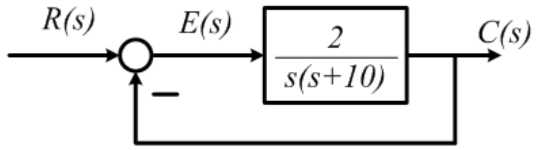


Figure 3.27: Closed-loop system

- *For a unit step input, compute the static position error constant:*

$$K_p = \lim_{s\to 0} G(s) = \lim_{s\to 0} \frac{2}{s(s+10)} = \infty$$

and the steady-state error:

$$e_{ss} = \frac{1}{1+K_p} = 0$$

- *For a unit ramp input compute the static velocity error constant*

$$K_v = \lim_{s\to 0} sG(s) = \lim_{s\to 0} s\frac{2}{s(s+10)} = \lim_{s\to 0} \frac{2}{s+10} = \frac{1}{5}$$

and the steady-state error:

$$e_{ss} = \frac{1}{K_v} = 5$$

3.6 The static gain (DC gain) of a stable system

Although the stability of linear continuous systems will be presented in a following section 3.10, consider a system with the transfer function $H(s)$ and all poles having negative real parts, or located in the left half s-plane.

The **static gain** or the **DC gain** of the system is the ratio of the steady-state output of the system to its steady-state input.

Example 3.6.1 *Consider two systems having the transfer functions $H_1(s)$ and $H_2(s)$ and the systems response for a unit step input shown in Figure 3.28.*

- *for system $H_1(s)$, the steady state value of the output is 0.5 and the input is 1. The static gain, or the DC gain of the system is $0.5/1 = 0.5$.*

- *for system $H_2(s)$, the steady state value of the output is 1 and the input is 1. The static gain, or the DC gain of the system is $1/1 = 1$.*

Example 3.6.2 *Consider two systems having the transfer functions $H_1(s)$ and $H_2(s)$ and the systems response for a step input $r(t) = 2$ shown in Figure 3.29.*

- *for system $H_1(s)$, the steady state value of the output is 1 and the input is 2. The static gain, or the DC gain of the system is $1/2 = 0.5$.*

- *for system $H_2(s)$, the steady state value of the output is 3 and the input is 2. The static gain, or the DC gain of the system is $3/2 = 1.5$.*

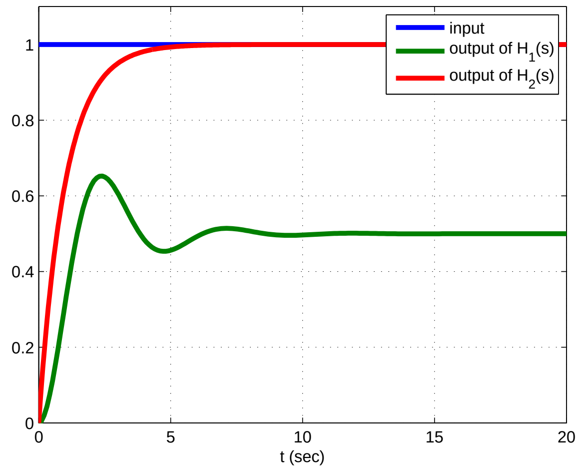


Figure 3.28: Step responses

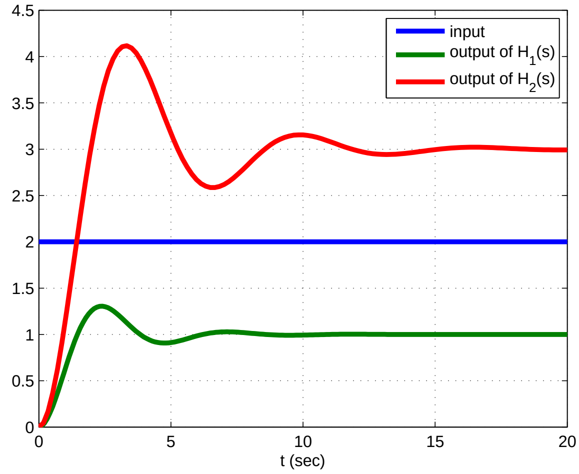


Figure 3.29: Step responses

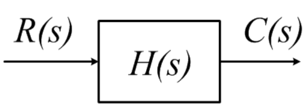


Figure 3.30: A system

In general, for a system having the transfer function $H(s)$ and a unit step input $r(t) = 1$ the output is given by:

$$C(s) = H(s)R(s) = H(s)\frac{1}{s}$$

The steady-state value of the output can be computed using the final value theorem:

$$c(\infty) = \lim_{t\to\infty} c(t) = \lim_{s\to 0} sC(s) = \lim_{s\to 0} sH(s)\frac{1}{s} = \lim_{s\to 0} H(s)$$

and the DC gain is:

$$DCgain = \frac{c(\infty)}{1} = \lim_{s\to 0} H(s)$$

For any constant (step) input $r(t) = A$, the output is given by:

$$C(s) = H(s)\frac{A}{s}$$

the steady-state value of the output is:

$$c(\infty) = \lim_{t\to\infty} c(t) = \lim_{s\to 0} sC(s) = \lim_{s\to 0} sH(s)\frac{A}{s} = A\lim_{s\to 0} H(s)$$

and the DC gain is:

$$DCgain = \frac{c(\infty)}{A} = \lim_{s\to 0} H(s)$$

Example 3.6.3 *Consider a system having the transfer function:*

$$H(s) = \frac{s+10}{(s+2)(s+4)(s^2+s+5)}$$

The DC gain is:

$$DCgain = \lim_{s\to 0} H(s) = \lim_{s\to 0} \frac{s+10}{(s+2)(s+4)(s^2+s+5)} = \frac{10}{2\cdot 4\cdot 5} = \frac{1}{4} = 0.25$$

If a unit step input is applied to this system, the steady-state value of the output will be:

$$c(\infty) = DCgain = 0.25$$

If the input is a constant signal $r(t) = 5$, the steady-state value of the output is:

$$c(\infty) = 0.25\cdot 5 = 1.25$$

3.7 Effect of an additional zero

Consider a system with the transfer function $H(s)$, a unit step input $R(s) = 1/s$ and the output $C(s)$.

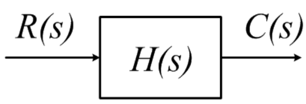


Figure 3.31: A system

The system step response can be expressed as:

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

Now suppose we add a zero at $-a$ and divide the transfer function with a so the DC gain (or the steady-state value) 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:

$$c_z(t) = \pounds^{-1}\left[C_z(s)\right] = \pounds^{-1}\left[\frac{1}{s}H_z(s)\right] = \pounds^{-1}\left[\frac{1}{s}\left(\frac{s}{a}H(s) + H(s)\right)\right] = \frac{1}{a}\dot{c}(t) + c(t)$$

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

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 $1/a \cdot \dot{c}(t)$. The effect of addition of a zero is the increase of the overshoot.

Example 3.7.1 . *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 3.32.

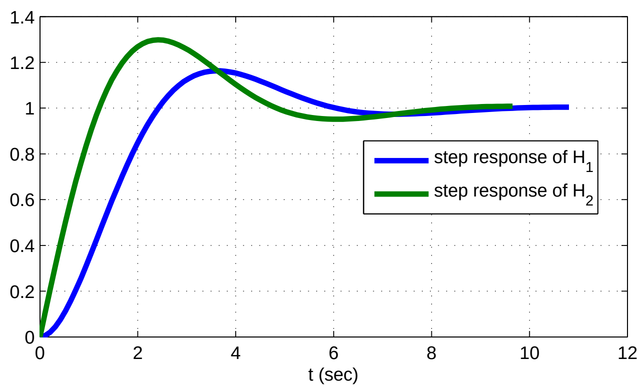


Figure 3.32: Effect of addition of a zero at -1

Now consider the 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 3.33.

The general effect of additional zero at -1 to the second order system is the increase of overshoot. If the additional zero is -10, then $a = 10$ and $1/a = 1/10 = 0.1$ is small. Thus, the contribution of the term $\dot{c}(t)$ in the total response is small and the step response of the system with a zero at -10 resembles the step response of the original system.

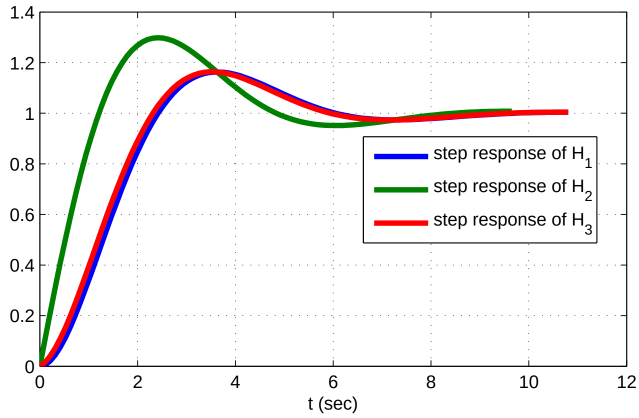


Figure 3.33: Effect of addition of a zero at -1 and -10

3.8 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:

$$C(s) = H(s) \cdot R(s) = \frac{a^m s^m + ... + a_1 s + a_0}{s(b^n s^n + ... + b_1 s + b_0)}, \quad (m \leq n) \tag{3.22}$$

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 (3.22) 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 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, (?):

$$c(t) = a + \sum_{j=1}^{q} a_j e^{-p_j t} + \sum_{k=1}^{r} b_k e^{-\zeta_k \omega_k t} cos \omega_k \sqrt{1 - \zeta_k^2} t + \sum_{k=1}^{r} c_k e^{-\zeta_k \omega_k t} sin \omega_k \sqrt{1 - \zeta_k^2} t \tag{3.23}$$

If all the 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 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**.

Example 3.8.1 *A real pole located at* -1 *will contribute to the total response of a system (3.23) with a (weighted) exponential* e^{-t}*, and a pole at* -10 *will contribute with an exponential* e^{-10t}*. A complex pole having the real part* -1 *and the imaginary part* 20 *gives a term of the type* $e^{-t}\sin 20t$ *and a complex pole with the real part* -10 *and the same imaginary part gives* $e^{-10t}\sin 20t$*. As presented in Figure 3.34, if the absolute value of the real part of the pole is large, the exponential approaches zero fast and its contribution in the system response vanishes quickly. These poles are located far from the imaginary axis. When the absolute value of the real part of the pole is smaller, or the pole is located near the imaginary axis, the corresponding exponential term decays slowly to zero and has a significant contribution to the system output.*

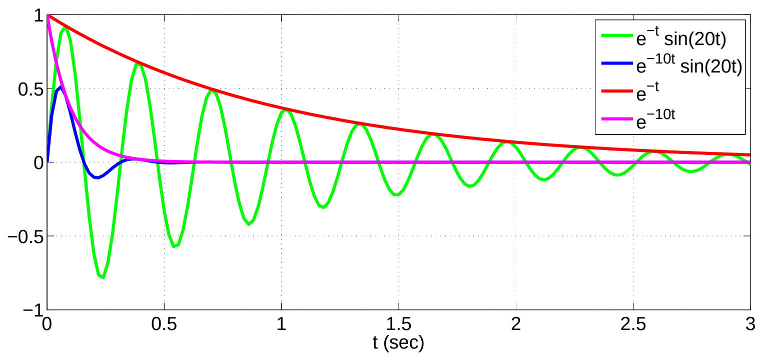


Figure 3.34: Exponential terms

3.8.1 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 $p_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2}$ 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 $1/pole$ (divided by the absolute value of the pole).

Example 3.8.2 *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 3.35.

Example 3.8.3 *Consider a system with the transfer function:*

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

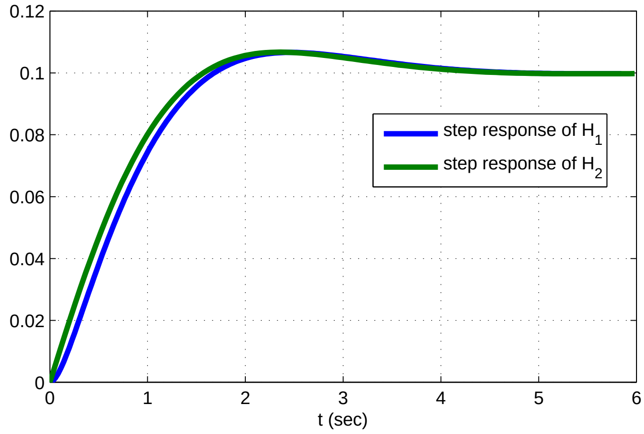


Figure 3.35: Comparation of two step responses

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 3.36. The effect of

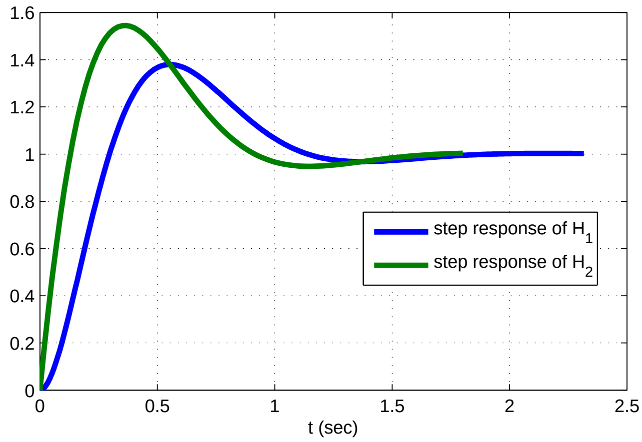


Figure 3.36: Comparation of two step responses

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.

3.9 Systems with time delay

Figure 3.37 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 $\tau = L/v$ 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 **time delay**, **transport lag** or **dead time**. The time delay is present in most process control systems.

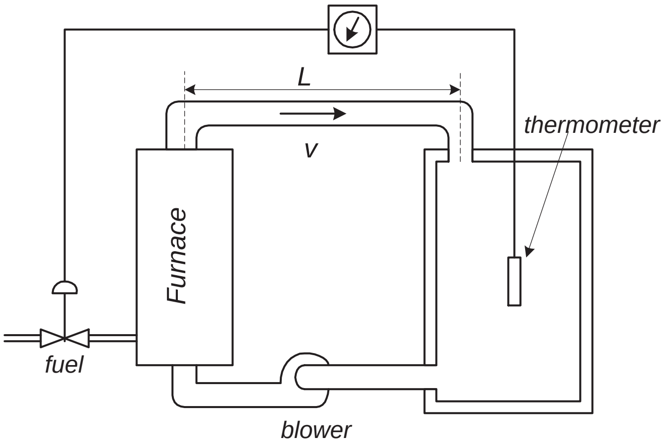


Figure 3.37: Thermal process

A **time delay** 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 $u(t)$ and the output $y(t)$ of a time delay element are related by

$$y(t) = u(t-\tau)$$

where τ is the time delay. The transfer function of pure time delay element is given by:

$$H(s) = \frac{\pounds[u(t-\tau)]}{\pounds[u(t)]} = \frac{U(s)e^{-s\tau}}{U(s)} = e^{-s\tau}$$

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

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

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

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

and the transfer function is:

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

If we use a Taylor series expansion for $e^{-s\tau}$:

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

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

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

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

$$a) \quad e^{-s\tau} = \frac{1-\frac{1}{2}\tau s}{1+\frac{1}{2}\tau s}$$

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

3.10 Stability of linear continuous systems

3.10.1 Definitions and stability in the s-plane

One of the most important characteristics of the dynamic behavior of a system is the stability. For linear systems the stability is related to the location of the roots of the characteristic equation.

Whether a linear system is stable or unstable is a property of the system itself and does not depend on the input or driving function of the system. The input contributes only to the steady-state response terms in the solution.

A system is *BIBO stable* if for a bounded input it has a bounded output (response).

A bounded input may be, for example, a step signal, or a sinusoidal signal. A ramp or an impulse are not bounded and cannot be used for stability analysis using this criterion.

Example 3.10.1 *, (?). The concept of stability can be illustrated by considering a right circular cone placed on a plane horizontal surface. If the cone is resting on its base and is tipped slightly, it returns to its original equilibrium position. This position and response is said to be stable. If the cone rests on its side and is displaced slightly, it rolls with no tendency to leave the position on its side. This position is designated as the neutral stability. On the other hand, if the cone is placed on its tip and released, it falls onto its side. This position is said to be unstable.*

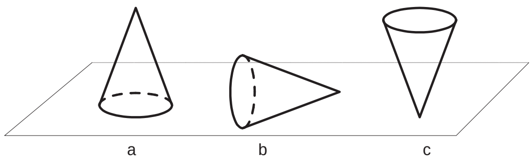


Figure 3.38: Stability

The **impulse response** can also be used as a basis for stability analysis. **A linear system is stable if and only if the absolute value of its impulse response, integrated over an infinite range, is finite**. A consequence of this definition is that the impulse response of a stable system will approach zero as time approaches infinity.

Consider a system with an input $r(t) = \delta(t)$ (ideal impulse) and an output $c(t)$. The responses of a stable, marginally stable or unstable system are presented in Figure 3.39.

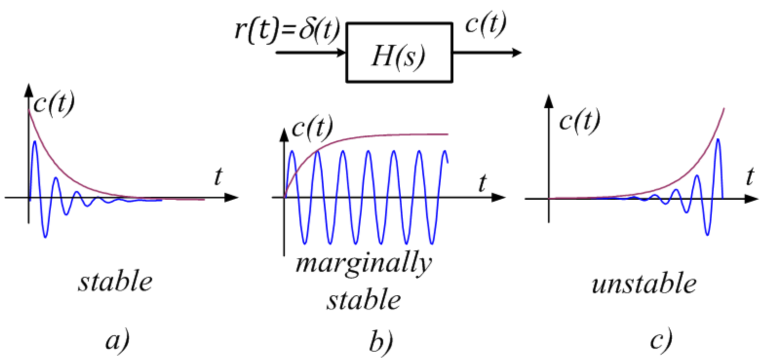


Figure 3.39: System response

- For a **stable system** the steady-state value of the impulse response is zero (Figure 3.39 a)):

$$\lim_{t\to\infty} c(t) = 0$$

- For a **marginally (critically) stable system** the steady-state value of the impulse response is different from zero, but it is limited, (Figure 3.39 b)):

$$0 < \lim_{t\to\infty} c(t) < \infty$$

- For an **unstable system** the steady-state value of the impulse response is not limited, (Figure 3.39 c)):

$$\lim_{t\to\infty} c(t) = \infty$$

The system transfer function is written as:

$$H(s) = \frac{C(s)}{R(s)} = \frac{k\prod_{i=1}^{m}(s+z_i)}{s^n\prod_{j=1}^{q}(s+\sigma_j)\prod_{k=1}^{r}(s^2+2\alpha_k s+(\alpha_k^2+\omega_k^2))}$$

The system poles are either real: $p_j = -\sigma_j$, complex conjugates $p_{k1,2} = \alpha_k \pm j\omega_k$, or poles at the origin of multiplicity N. The output response for an impulse input is then:

$$c(t) = \sum_{j=1}^{q} A_j e^{p_j t} + \sum_{k=1}^{r} B_k(\frac{1}{\omega_k})e^{\alpha_k t} sin\omega_k t \tag{3.24}$$

The location of the poles in the s-plane indicate the system response. The poles in the left-hand portion of the s-plane result in a decreasing response for disturbance inputs. Similarly, poles on the $j\omega$ axis and in the right-hand plane result in a neutral and an increasing response, respectively, for an impulse input.

This division of the s-plane is shown in Figure 3.40 where the location of the poles and the corresponding impulse response are represented with the same color.

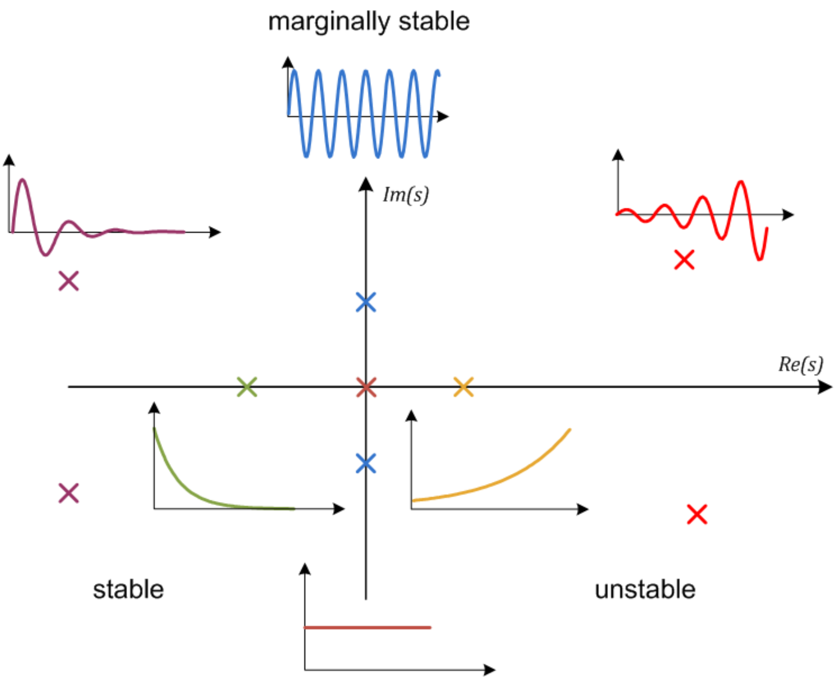


Figure 3.40: Stability in the s-plane

- Real positive poles (p_j) and complex poles with positive real parts (α_k) contribute in the system response with terms of the type $e^{p_j t}$ or $e^{\alpha_k t} sin\omega_k t$ that increase towards infinity.

- Real negative poles (p_j) and complex poles with negative real parts (α_k) contribute in the system response with terms of the type $e^{p_j t}$ or $e^{\alpha_k t} sin\omega_k t$ that approach zero as time approaches infinity.
- Complex poles located on the imaginary axis ($\pm j\omega_k$) give an undamped sinusoidal term in the system output
- One pole at the origin ($p_j = 0$) contributes with a constant term in the system response
- Poles at the origin of multiplicity greater than one $n > 1$ contribute in the system response with terms of the type At^{n-1} which increase towards infinity
- Complex poles located on the imaginary axis of multiplicity greater than one will give terms in the system response of the form $At^n cos(\omega t + \phi)$ which also approach infinity as time approaches infinity.

The last two cases in the list above are not included in Figure 3.40, but the following example shows the impulse response for systems having double poles on the imaginary axis.

Example 3.10.2 *Consider two system with double poles on the imaginary axis, with the transfer functions:*

$$H_1(s) = \frac{1}{s^2}, \qquad H_2(s) = \frac{1}{(s^2+1)^2}$$

H_1 has a double pole at the origin and H_2 has two pairs of complex poles at $\pm j$. The impulse responses, shown in Figure 3.41, increase towards infinity in both cases.

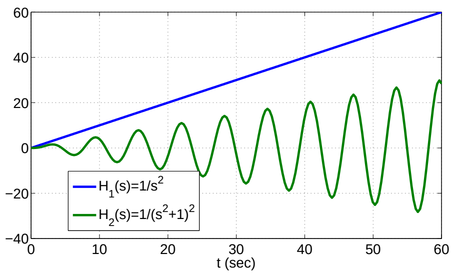


Figure 3.41: Double poles on the imaginary axis

Clearly, to obtain a zero steady-state of the impulse response, all exponentials in (3.24) must approach zero as time approaches infinity, or, all the poles of the system must be in the left-hand portion of the s-plane.

A necessary and sufficient condition for a system to be stable is that all the poles of the system transfer function have negative real parts.

We will call a system not stable if *not* all the poles are in the left-hand s-plane. If the characteristic equation has roots on the imaginary axis, with all the other roots in the left-hand plane, the steady-state output will be sustained oscillations for an impulse input. Such a system is called *marginally stable or critically stable.* For an unstable system, the characteristic equation has at least one root in the right-hand half plane or repeated $j\omega$ (imaginary) roots or multiple poles at the origin; for this case, the output is unbounded for any input.

To determine the stability of a dynamical system, one could determine the roots of the characteristic equation.

Example 3.10.3 • *A system with the transfer function:*

$$H_1(s) = \frac{1}{(s+1)(s+2)}$$

is stable because all the poles are negative: $p_1 = -1$ *and* $p_2 = -2$.

- *A system with the transfer function:*

$$H_2(s) = \frac{1}{(s+1)(s^2+2s+2)}$$

is stable because all the poles are located in the left half s-plane: one real negative poles $p_1 = -1$ *and a pair of complex poles* $p_{2,3} = -1 \pm j$ *with negative real parts.*

- *A system with the transfer function:*

$$H_3(s) = \frac{1}{s(s+1)}$$

is marginally stable because it has one pole at the origin $p_1 = 0$ *while the second pole is negative* $p_2 = -1$.

- *A system with the transfer function:*

$$H_3(s) = \frac{1}{s^2+4}$$

is marginally stable because it has a pair of complex poles on the imaginary axis $p_{1,2} = \pm 2j$.

- *A system with the transfer function:*

$$H_4(s) = \frac{1}{(s+1)(s+2)(s-1)}$$

is unstable because not all the poles are located on the left half s-plane: one pole is positive $p_1 = 1$ *and the other poles* $p_2 = -1$, $p_3 = -2$ *are negative. The impulse response will still approach infinity as time approaches infinity due to the positive pole.*

3.10.2 The Routh-Hurwitz stability criterion

The discussion and determination of stability has occupied the interest of many engineers. In the late 1800's, A. Hurwitz and E.J.Routh published independently a method of investigating the stability of linear systems. The Routh-Hurwitz stability method provides an answer to the question of stability by considering the characteristic equation of the system, (**?**).

The characteristic polynomial in the Laplace variable is written as:

$$q(s) = a_n s^n + a_{n-1}s^{n-1} + ... + a_1 s + a_0 \tag{3.25}$$

To analyze the stability of the system, it is necessary to determine whether any one of the roots of $q(s)$ lies in the right-hand half of the s-plane.

The procedure in the Routh's stability criterion is as follows:

1. Write the polynomial in s in the form (3.25) where the coefficients are real quantities. We assume that $a_0 \neq 0$ that is, any zero root has been removed.
2. *Necessary conditions.*
   (a) All the coefficients of the polynomial $q(s)$ must have the same sign if all the roots are in the left-hand half plane.

(b) All the coefficients of $q(s)$ for a stable system are nonzero.

However, although necessary, these requirements are not sufficient. That is, we immediately know the system is unstable if they are not satisfied, yet if they are satisfied, we must proceed to ascertain the stability of the system.

Example 3.10.4 *For example, in the characteristic polynomial $q(s) = s^3 + s^2 + 1$ the coefficient of s^1 is zero. In the polynomial $q(s) = s - 2$, the coefficients do not have the same sign. Therefore, the roots of these characteristic polynomials are not (all) located in the left-hand s-plane and the systems are not stable.*

3. *Sufficient condition.* If all the coefficients are positive, arrange the coefficients of the polynomial in the *Routh array* according to the following pattern:

$$
\begin{array}{rc|c|cccc}
s^n & : & a_n & a_{n-2} & a_{n-4} & a_{n-6} & \cdot\ \ \cdot \\
s^{n-1} & : & a_{n-1} & a_{n-3} & a_{n-5} & a_{n-7} & \cdot\ \ \cdot \\
s^{n-2} & : & b_1 & b_2 & b_3 & b_4 & \cdot\ \ \cdot \\
s^{n-3} & : & c_1 & c_2 & c_3 & c_4 & \cdot\ \ \cdot \\
\cdot & : & \cdot & \cdot & \cdot & \cdot & \cdot\ \ \cdot \\
\cdot & : & \cdot & \cdot & \cdot & \cdot & \cdot\ \ \cdot \\
s^2 & : & e_1 & e_2 & & & \\
s^1 & : & f_1 & & & & \\
s^0 & : & g_1 & & & &
\end{array}
$$

The coefficients b_1, b_2, and so on are evaluated as follows:

$$b_1 = \frac{-\begin{vmatrix} a_n & a_{n-2} \\ a_{n-1} & a_{n-3} \end{vmatrix}}{a_{n-1}} = \frac{a_{n-1}a_{n-2} - a_n a_{n-3}}{a_{n-1}}$$

$$b_2 = \frac{-\begin{vmatrix} a_n & a_{n-4} \\ a_{n-1} & a_{n-5} \end{vmatrix}}{a_{n-1}} = \frac{a_{n-1}a_{n-4} - a_n a_{n-5}}{a_{n-1}}$$

$$b_3 = \frac{-\begin{vmatrix} a_n & a_{n-6} \\ a_{n-1} & a_{n-7} \end{vmatrix}}{a_{n-1}} = \frac{a_{n-1}a_{n-6} - a_n a_{n-7}}{a_{n-1}}$$

The evaluation of b's is continued until the remaining ones are all zero. The same pattern is followed in evaluating the c's, ..e's, f, g, using the coefficients of the two previous rows. That is:

$$c_1 = \frac{-\begin{vmatrix} a_{n-1} & a_{n-3} \\ b_1 & b_2 \end{vmatrix}}{b_1} = \frac{b_1 a_{n-3} - a_{n-1} b_2}{b_1}$$

$$c_2 = \frac{-\begin{vmatrix} a_{n-1} & a_{n-5} \\ b_1 & b_3 \end{vmatrix}}{b_1} = \frac{b_1 a_{n-5} - a_{n-1} b_3}{b_1}$$

The process is continued until the n-th row is completed. The complete array of coefficients is triangular.

Routh's stability criterion states that *the number of roots of the characteristic polynomial $q(s)$ with positive real parts is equal to the number of changes in sign of the coefficients in the first column of the array.*

Example 3.10.5 *Consider the following polynomial:*

$$q(s) = s^4 + 2s^3 + 3s^2 + 4s + 5$$

Following the procedure presented above we'll construct the array of coefficients. The first two rows are obtained directly from the given polynomial. The remaining are obtained from the two rows above. If any coefficients are missing they may be replaced by zeros in the array.

$$\begin{array}{lcccc} s^4 & : & 1 & 3 & 5 \\ s^3 & : & 2 & 4 & 0 \\ s^2 & : & \frac{2\cdot3-1\cdot4}{2} = 1 & \frac{2\cdot5-1\cdot0}{2} = 5 & \\ s^1 & : & \frac{1\cdot4-2\cdot5}{1} = -6 & & \\ s^0 & : & \frac{-6\cdot5-1\cdot0}{-6} = 5 & & \end{array}$$

In this example, the number of changes in sign of the coefficients in the first column is two ($1 \to -6$ and $-6 \to 5$). This means that there are two roots with positive real parts. Indeed, if we calculate the roots of the polynomial we obtain: $p_1 = 0.2878 + 1.4161j$, $p_2 = 0.2878 - 1.4161j$, $p_3 = -1.2878 + 0.8579j$, $p_4 = -1.2878 - 0.8579j$.

Example 3.10.6 ***Special case***, *(?). If a first-column term in any row is zero, but the remaining terms are not zero or there is no remaining term, then the zero term is replaced by a very small positive number ϵ and the rest of the array is evaluated. For example consider the characteristic equation:*

$$s^3 + 2s^2 + s + 2 = 0$$

All coefficients have the same sign (are positive) and none of them is zero. The necessary conditions are fulfilled.

The Routh array of coefficients is:

$$\begin{array}{lccc} s^3 & : & 1 & 1 \\ s^2 & : & 2 & 2 \\ s^1 & : & 0 \approx \epsilon & \\ s^0 & : & 2 & \end{array}$$

If the sign of the coefficient above the zero (ϵ) is the same as the one below it, it indicates that there are a pair of imaginary roots. Actually, this equation has two roots at $\pm j$.

If, in other cases, the sign of the coefficient above zero is opposite that below it, it indicates that there is one sign change.

Example 3.10.7 *Application of Routh's stability criterion to control system analysis, (?) Routh stability criterion is useful to determine the effects of changing one or two parameters by examining the values that cause instability. In the following, we shall consider the problem of determining the stability range of a parameter value.*

Consider the system shown in Figure 3.42. Let us determine the range of k for stability. The closed-loop transfer function is:

$$\frac{C(s)}{R(s)} = \frac{k}{s(s+2)(s^2+s+1)+k}$$

The characteristic equation is:

$$s^4 + 3s^3 + 3s^2 + 2s + k = 0$$

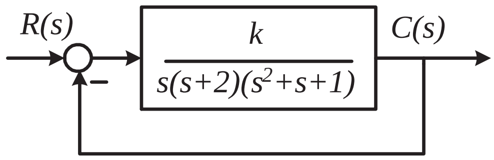


Figure 3.42: Closed-loop system

To fulfill the necessary conditions, k must be positive $k > 0$. All the other coefficients of the characteristic polynomial are positive and non-zero.

The array of coefficients is:

$$\begin{array}{ccccc} s^4 & : & 1 & 3 & k \\ s^3 & : & 3 & 2 & 0 \\ s^2 & : & 7/3 & k & \\ s^1 & : & 2-9k/7 & & \\ s^0 & : & k & & \end{array}$$

For stability, k must be positive (also imposed by the necessary condition), and all coefficients in the first column must be positive. Therefore:

$$2 - \frac{9 \cdot k}{7} > 0$$

or

$$0 < k < \frac{14}{9}$$

When $k = 14/9$, the system becomes marginally stable, the system response is oscillatory with constant amplitude.

3.11 Root locus analysis

3.11.1 Introduction and motivation

The basic characteristic of the transient response of a closed-loop system is closely related to the location of the closed-loop poles. If the system has a variable loop gain, then the location of the closed-loop poles depends on the value of the gain chosen. It is important therefore, that the designer knows how the closed-loop poles move in the s-plane as the loop gain is varied.

Example 3.11.1 *Consider a closed-loop negative feedback control system (Figure 3.43) with the open-loop transfer function $kG(s)$ where:*

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

and k is a positive variable gain.

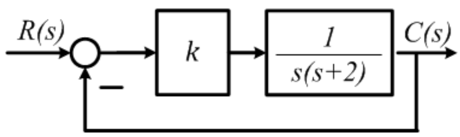


Figure 3.43: Closed-loop control system

For $k = 1$:

$$G_0(s) = \frac{1}{s^2 + 2s + 1}$$

The closed-loop poles are real and equal $s_{1,2} = -1$ and the step response of the closed-loop system is critically damped.

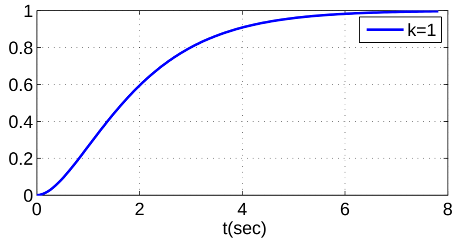


Figure 3.44: Step response for $k = 1$

For $k = 3$:

$$G_0(s) = \frac{3}{s^2 + 2s + 3}$$

The closed-loop poles are complex $s_{1,2} = -1 \pm \sqrt{2}j$ and the step response of the closed-loop system is underdamped with a small overshoot.

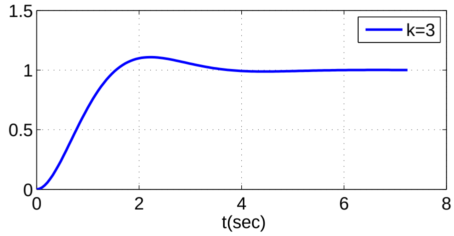


Figure 3.45: Step response for $k = 3$

The poles of the open-loop transfer function are 0 *and* -2 *and do not depend on the value of the variable system parameter* k*. The closed-loop transfer function is:*

$$G_0(s) = \frac{kG(s)}{1 + kG(s)} = \frac{k}{s^2 + 2s + k}$$

Clearly, the poles of the closed-loop system depend on k *and thus, the system response, for example to a unit step input, will change for various values of* k*.*

A method for finding the roots of the characteristic equation has been developed by W.R. Evans and used extensively in control engineering. The method, called *the root-locus method*, is one in which the roots of the characteristic equation are plotted for all values of a system parameter. The roots corresponding to a particular value of this parameter can then be located on the resulting graph. Note that the parameter is usually the gain, but any other variable of the open-loop transfer function may be used, (**?**).

The root locus is a plot of the roots of the characteristic equation of a closed-loop system for all values of a variable parameter in the system, $k \in [0, \infty)$.

3.11.2 The root-locus method

Consider the closed-loop system shown in Figure 3.47, where k is the open-loop gain, and the open-loop transfer function is:

$$H_d(s) = kG(s)H(s)$$

For $k = 50$:

$$G_0(s) = \frac{50}{s^2 + 2s + 50}$$

The closed-loop poles complex $s_{1,2} = -1 \pm 7j$ and the step response of the closed-loop system is underdamped with a large overshoot and more oscillations.

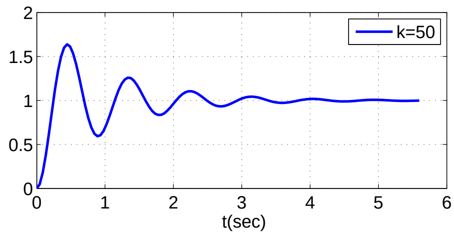


Figure 3.46: Step response for $k = 50$

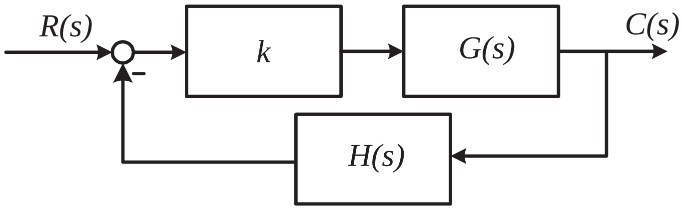


Figure 3.47: Closed-loop control system

The closed-loop transfer function is:

$$\frac{C(s)}{R(s)} = \frac{kG(s)}{1 + kG(s)H(s)} \tag{3.26}$$

The *characteristic equation* for this closed-loop system is obtained by setting the denominator of the right side of equation (3.26) equal to zero. That is,

$$1 + kG(s)H(s) = 0$$

or

$$kG(s)H(s) = -1 \tag{3.27}$$

The basic idea behind the root-locus method is that the values of s that make the transfer function around the loop $kG(s)H(s)$ equal -1 must satisfy the characteristic equation of the system.

$kG(s)H(s)$ is a ratio of polynomials in s. Since $kG(s)H(s)$ is a complex quantity, equation (3.27) can be split in two equations by equating the angles and magnitudes of both sides, to obtain:

$$\mid kG(s)H(s) \mid \angle kG(s)H(s) = -1 + j0$$

or:

Angle condition:

$$\angle kG(s)H(s) = \pm 180^o(2q+1), \quad q = 0, 1, 2, ... \tag{3.28}$$

Magnitude condition:

$$|kG(s)H(s)| = 1 \tag{3.29}$$

The values of s that fulfill the angle and magnitude conditions are the roots of the characteristic equation, or the closed-loop poles. A plot of points of the complex plane satisfying the angle condition alone is the root-locus. The roots of the characteristic equation (the closed-loop poles) corresponding to a given value of the gain can be determined from the magnitude condition, (**?**).

Remember 1 . ***Complex numbers.***

- *A complex number $s = a + jb$, with the real part a and the imaginary part b (Figure 3.48) has the absolute value (or* magnitude *or modulus):*

  $$|s| = |a + jb| = \sqrt{a^2 + b^2}$$

  and the phase *(or angle, or argument):*

  $$\angle s = \arg(s) = \arctan\frac{b}{a}, \quad \text{if} \;\; a > 0$$

  The phase is measured counterclockwise from the positive real axis.

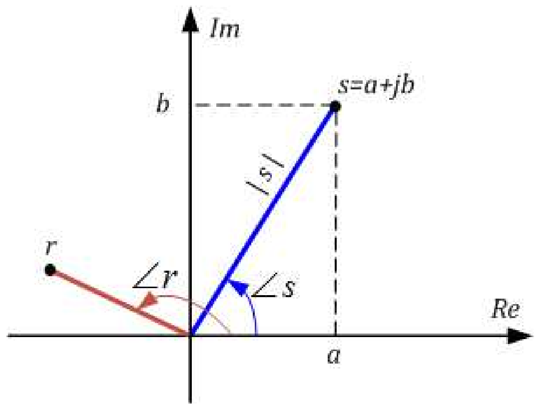


Figure 3.48: Magnitude and phase of a complex number

- *Consider two complex numbers:* $s_1 = a + jb$ *and* $s_2 = c + jd$. *The product of these complex numbers has the magnitude:*

$$|s_1 s_2| = |s_1| \cdot |s_2| = \sqrt{a^2 + b^2}\sqrt{c^2 + d^2}$$

and the phase:

$$\angle(s_1 s_2) = \angle s_1 + \angle s_2 = \angle(a + jb) + \angle(c + jd)$$

- *The ratio of two complex numbers has the magnitude:*

$$\left|\frac{s_1}{s_2}\right| = \frac{|s_1|}{|s_2|} = \frac{\sqrt{a^2 + b^2}}{\sqrt{c^2 + d^2}}$$

and the phase:

$$\angle\frac{s_1}{s_2} = \angle s_1 - \angle s_2 = \angle(a + jb) - \angle(c + jd)$$

- *Consider, for example, a complex number* r *and we want to compute the phase of* $r + 3$. *As shown in Figure 3.49, the points* r, $r + 3$, 0 *and* -3 *form a parallelogram. Therefore, the phase of* $r + 3$ *can be obtained as the angle (measured counterclockwise) from the positive real axis to the line that connects* r *and the root of* $r + 3$, *i.e.* -3.

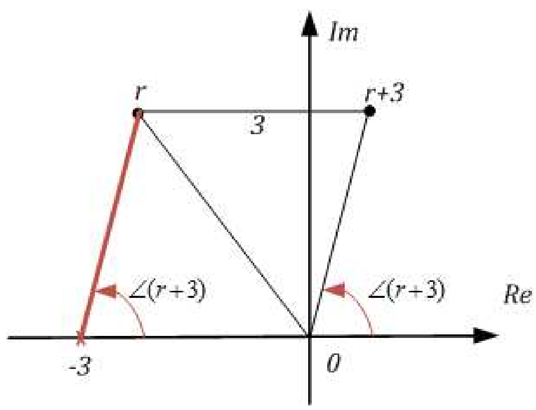


Figure 3.49: The phase of $r + 3$

Example 3.11.2 , *(?) Consider a closed-loop system with the open-loop transfer function* $kG(s)H(s)$ *given by:*

$$kG(s)H(s) = \frac{k(s + 3)}{(s + 5)(s^2 + 2s + 2)}$$

The open-loop poles are: $p_1 = -5$ *and* $p_{2,3} = -1 \pm j$. *The open-loop zero is* $z_1 = -3$.

For a testpoint $s = r$ *(in general a complex number), the angle of* $kG(s)H(s)$ *is:*

$$\angle kG(s)H(s)|_{s=r} = \angle k + \angle(r+3) - \angle(r+5) - \angle(r+1-j) - \angle(r+1+j) = 0 + \varphi_1 - \varphi_2 - \varphi_3 - \varphi_4$$

where φ_1, φ_2, φ_3, φ_4 *are measured counterclockwise as shown in Figure 3.50.*

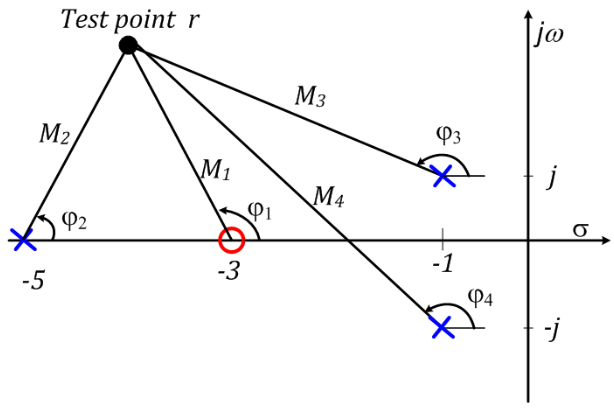


Figure 3.50: Angle measurement from a test point r to the open-loop poles and zeros

The magnitude of $G(s)H(s)$ *for this system is:*

$$|kG(s)H(s)|_{s=r} = \frac{kM_1}{M_2M_3M_4}$$

where M_1, M_2, M_3, M_4 *are the magnitudes of complex quantities in the numerator and denominator of the open-loop transfer function, as shown in Figure 3.50:*

$$M_1 = |r+3|, \quad M_2 = |r+5|, \quad M_3 = |r+1-j|, \quad M_4 = |r+1+j|$$

Example 3.11.3 ***Root locus of a second order system***

A root-locus plot of a simple second-order system is illustrated. Consider the closed-loop system shown in Figure 3.51. The open-loop transfer function $G(s)$ *is:*

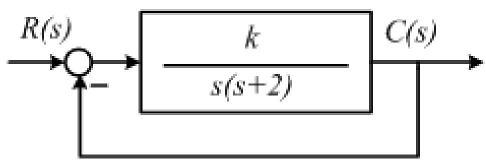


Figure 3.51: Closed-loop control system

$$G(s) = \frac{k}{s(s+2)}$$

The closed-loop transfer function is then:

$$G_0(s) = \frac{G(s)}{1+G(s)} = \frac{\frac{k}{s(s+2)}}{1+\frac{k}{s(s+2)}} = \frac{k}{s^2+2s+k}$$

The characteristic equation is:

$$1 + \frac{k}{s(s+2)} = 0 \quad \text{or } s^2 + 2s + k = 0$$

We wish to find the locus of the roots of this system as k is varied from zero to infinity.

We shall first obtain analytically the roots of the characteristic equation in terms of k and analyze them when k varies from 0 to ∞. In general, this is not a proper method to construct a root locus. An analytical solution of the characteristic equation is, in general, hard to find and complicated for systems of order higher than 2. The root locus method for systems of any order will be presented in the next section. This example introduces only the main ideas for a simple application.

The roots of the characteristic equation are:

$$s_1 = -1 + \sqrt{1-k}, \quad s_2 = -1 - \sqrt{1-k}$$

The roots are real for $k \leq 1$ and complex for $k > 1$. The locus of the roots corresponding to all positive values of k is plotted in Figure 3.52. The gain k is a parameter on the root locus. The motion of the poles

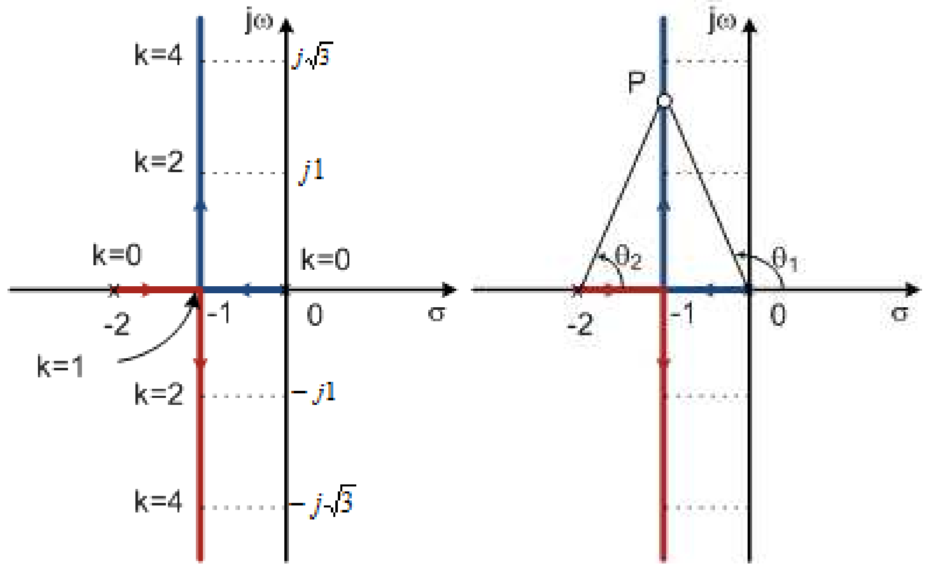


Figure 3.52: Root locus plot (left). Angles measured from a test point P (right)

with increasing k is shown by arrows.

- *For $k = 0$, the closed-loop poles are the same as the poles of the open-loop transfer function $G(s)$, i.e., $s_1 = 0$, $s_2 = -2$.*
- *As k increases in the interval $(0, 1)$, the closed-loop poles move towards the point $(-1, 0)$. Because the poles have real values, the system response in overdamped.*
- *For $k = 1$, the closed-loop poles are real and equal $s_1 = s_2 = -1$. The system is critically damped.*
- *As k increases from 1 to ∞, the closed-loop poles break away from the real axis and become complex. They are given by $s_{1,2} = -1 \pm \sqrt{k-1}j$. Since the real part of the poles is constant at -1, the closed-loop poles move along the vertical line $s = -1$. Because the poles are complex conjugates, they will move symmetrically with respect to the real axis. For $k > 1$ the system becomes underdamped and the system response is oscillatory.*

We shall show that any point on the root locus satisfies the angle condition. The angle condition given by equation (3.28) is:

$$\angle\frac{k}{s(s+2)} = -\angle s - \angle s+2 = \pm 180^o(2q+1), \quad q = 0,1,2,..$$

Consider point P on the root locus shown in Figure 3.52 (- right). The complex quantities s and $s+2$ have angles θ_1 and θ_2 respectively. Note that all angles are considered positive if they are measured in the counterclockwise direction. The sum of angles θ_1 and θ_2 is clearly 180^o, thus:

$$\angle\frac{k}{s(s+2)}|_{s=P} = -\theta_1 - \theta_2 = -180^o$$

If the point P is located on the real axis between -1 and 0, then $\theta_1 = 180^o$ and $\theta_2 = 0^o$. Thus, we see that any point on the root locus satisfies the angle condition. We also see that if the point P is not located on the root locus, then the sum of θ_1 and θ_2 is not equal to $\pm 180^o(2k+1)$. Thus, points that are not on the root locus do not satisfy the angle condition, and therefore, they cannot be closed-loop poles for any value of k.

Once we draw such a root locus plot we can determine the value of k that will yield a root, or a closed-loop pole, at a desired point. The corresponding value of k is determined from the magnitude condition, equation (3.29)

If, for example the closed-loop poles selected are $s = -1 \pm j$, then the corresponding value of k is found from:

$$|G(s)H(s)|_{s=-1+j} = \left|\frac{k}{s(s+2)}\right|_{s=-1+j} = 1$$

or

$$k = |s(s+2)|_{s=-1+j} = |-1+j|\cdot|-1+j+2| = |-1+j|\cdot|1+j| = \sqrt{1^2+1^2}\sqrt{1^2+1^2} = 2$$

From the root-locus diagram of Figure 3.52 we see the effects of changes in the value of k on the transient response behavior of the second-order system. An increase in the value of k will decrease the damping ratio ζ, resulting in an increase in the overshoot of the response. An increase in the value of k will also result in increases in the damped and undamped natural frequencies.

From the root-locus plot, it is clear that the closed-loop poles are always in the left-half of the s-plane; thus, no matter how much k is increased, the system remains always stable.

3.11.3 The root locus procedure

Consider a closed-loop system with a variable parameter $k \in [0, \infty)$ as shown in Figure 3.53.

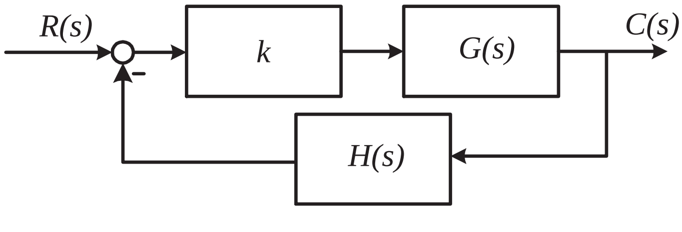


Figure 3.53: Closed-loop control system

The characteristic equation of the closed-loop system is given by:

$$1 + kG(s)H(s) = 0$$

where $kG(s)H(s)$ is the open-loop transfer function.

In the following, some rules for constructing a root locus are given.

1. Write the characteristic equation so that the variable parameter k appears as a multiplier

$$1 + kP(s) = 0.$$

2. Factor $P(s)$ in terms of n_p poles and n_z zeros

$$1 + k\frac{\prod_{i=1}^{n_z}(s + z_i)}{\prod_{i=1}^{n_p}(s + p_j)} = 0$$

3. Locate the open-loop poles and zeros of $P(s)$ in the s-plane with symbols: $\times$ - open-loop poles, o - open-loop zeros.

4. Determine the number of separate loci SL. $SL = n_p$, when $n_p \geq n_z$, where n_p is the number of open-loop poles and n_z the number of open-loop zeros.

5. Locate the segments of the real axis that are root loci:

   (a) On the real axis the locus lies to the left of an odd number of open-loop poles and zeros

   (b) Locus begins at an open-loop pole and ends at an open-loop zero (or infinity along to asymptotes- in case the number of zeros is smaller than the number of poles)

6. The root locus is symmetrical with respect to the horizontal real axis.

7. The loci proceed to infinity along asymptotes centered at σ_A and with angles Φ_A.

$$\sigma_A = \frac{\sum(p_j) - \sum(z_i)}{n_p - n_z}$$

$$\Phi_A = \frac{2q+1}{n_p - n_z} \cdot 180^o, \quad q = 0, 1, 2, ...(n_p - n_z - 1)$$

8. By utilizing the Routh-Hurwitz criterion, determine the point at which the locus crosses the imaginary axis (if it does so).

9. Determine the breakaway or break-in points on the real axis (if any)

   (a) Set $k = -\frac{1}{P(s)} = p(s)$, (from the characteristic equation $1 + kP(s) = 0$)

   (b) Obtain $dp(s)/ds = 0$

   (c) Determine roots of (b) or use graphical method to find maximum of $p(s)$.

10. Determine the angle of locus departure from complex poles and the angle of locus arrival at complex zeros, using the phase criterion

$$\angle P(s) = \pm 180^o(2q+1), \ at\ s = p_j\ or\ z_i.$$

11. If required, determine the root locations that satisfy the phase criterion

$$\angle P(s) = \pm 180^o(2q+1) \text{ at a root location } s_x$$

12. If required, determine the parameter value k_x at a specific root s_x using the magnitude condition:

$$k_x = \frac{\prod_{j=1}^{n_p} \mid s + p_j \mid}{\prod_{i=1}^{n_z} \mid s + z_i \mid} \mid_{s=s_x}$$

Example 3.11.4 *Consider a closed-loop system as the one shown in Figure 3.53, where $H(s) = 1$, and the open-loop transfer function:*

$$G(s) = \frac{k}{s(s+1)(s+2)}$$

Let us sketch the root-locus plot and determine the value of k so that the damping factor ζ of a pair of dominant complex-conjugate closed-loop poles is 0.5.

By following the procedure presented above section we obtain:

1. *The characteristic equation is written as:*

$$1 + kG(s) = 0$$

or

$$1 + \frac{k}{s(s+1)(s+2)} = 0$$

2. *The open-loop transfer function has no zeros, thus $n_z = 0$, and three poles, $p_1 = 0$, $p_2 = -1$, $p_3 = -2$, thus $n_p = 3$.*

3. *Locate the open-loop poles of the open-loop transfer function with symbol $\times$, as shown in Figure 3.54.*

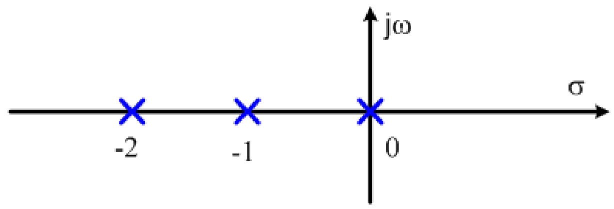


Figure 3.54: Location of the open-loop poles

4. *The locus has $n_p = 3$ branches.*

5. *Locate the segments of the real axis that are root loci:*

On the negative real axis, the locus exists between $p_1 = 0$ and $p_2 = -1$, and from $p_3 = -2$ to $-\infty$ (see Figure 3.55). All points on these segments are located on the left of an odd number of open-loop poles (or zeros). The points in $(0, \ -1)$ are on the left of one open-loop pole ($p_1 = 0$). The points in $(-2, -\infty)$ are on the left all of 3 open-loop poles. All the points between $p_2 = -1$ and $p_3 = -2$ are on the left of 2 (an even number) open loop poles therefore, this segment does not belong to the root locus.

The three branches of the root locus begin (for $k = 0$) at the open-loop poles and end (for $k \to \infty$) at infinity along to some asymptotes that will be computed. The arrows placed on the colored branches of the root locus in Figure 3.55 show the direction of k increasing from 0.

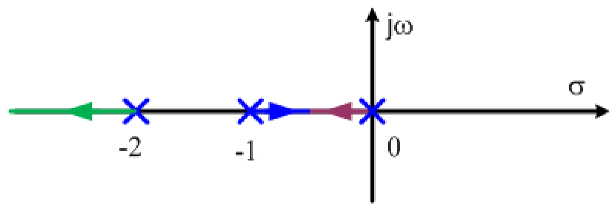


Figure 3.55: Location of the root locus on the real axis

6. The root locus is symmetrical with respect to the horizontal real axis. The points on the root locus are either real poles of the closed-loop system, located on the real axis, or complex conjugate poles located in the complex plane symmetrically with respect to the real axis.

7. The locus proceeds to infinity along asymptotes centered at σ_A and with angles Φ_A. Compute the center of the asymptotes and the angles:

$$\sigma_A = \frac{\sum(p_j) - \sum(z_i)}{n_p - n_z} = \frac{0-1-2}{3} = -1$$

$$\Phi_A = \frac{2q+1}{n_p - n_z} \cdot 180^o, \quad q = 0, 1, 2$$

or

$$\Phi_A = 60^o, 180^o, 300^o$$

Thus, there are three asymptotes. The one having the angle 180^o is the negative real axis. The asymptotes are shown in Figure 3.56.

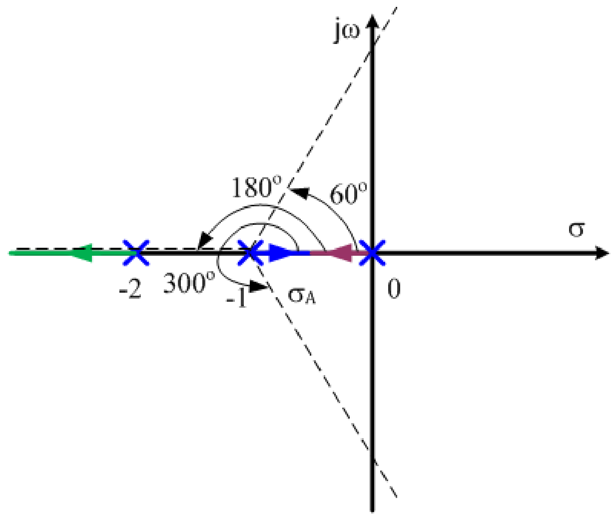


Figure 3.56: Asymptotes

8. The points at which the locus crosses the imaginary axis can be determined using the Routh-Hurwitz criterion. We shall compute the value of k for which the closed-loop system is marginally stable and the complex solutions of the characteristic equation located on the imaginary axis for this value of k. The characteristic equation is:

$$1 + \frac{k}{s(s+1)(s+2)} = 0 \quad \text{or} \quad s^3 + 3s^2 + 2s + k = 0$$

and the Routh array:

$$\begin{array}{lccc} s^3 & : & 1 & 2 \\ s^2 & : & 3 & k \\ s^1 & : & (6-k)/3 & \\ s^0 & : & k & \end{array}$$

The crossing points on the imaginary axis that makes the s^1 term in the first column equal zero is obtained from $(6-k)/3 = 0$, or $k = 6$. Replacing this value in the characteristic equation we obtain:

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

or

$$(s+3)(s^2+2)=0$$

The crossing points on the imaginary axis are the imaginary roots of the characteristic equation when $k=6$:

$$s_{1,2}=\pm j\sqrt{2}$$

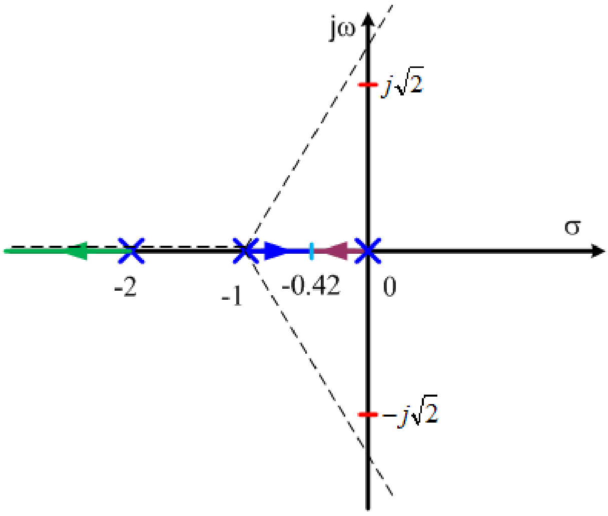


Figure 3.57: Intersection with the imaginary axis and the breakaway point

9. *Compute the breakaway point. From the the characteristic equation compute k:*

$$1+\frac{k}{s(s+1)(s+2)}=0 \quad \Rightarrow \quad k=p(s)=-s(s+1)(s+2)$$

The breakaway point on the real axis will be determined from the solution of

$$p'(s)=0$$

$$\frac{dp(s)}{ds}=-(3s^2+6s+2)=0$$

yields

$$s_1=-0.4226, \quad s_2=-1.5774$$

Since the breakaway point must lie between 0 and -1, $s_1=-0.4226$ corresponds to the actual breakaway point (Figure 3.57).

Based on the information obtained in the foregoing steps, the root locus obtained is shown in Figure 3.58.

In order to determine a pair of complex conjugate closed-loop poles such that the damping ratio $\zeta=0.5$ we have to remember some properties of the complex poles of the second order system:

- *The complex poles of a general second-order transfer function with the characteristic equation:*

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

are

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

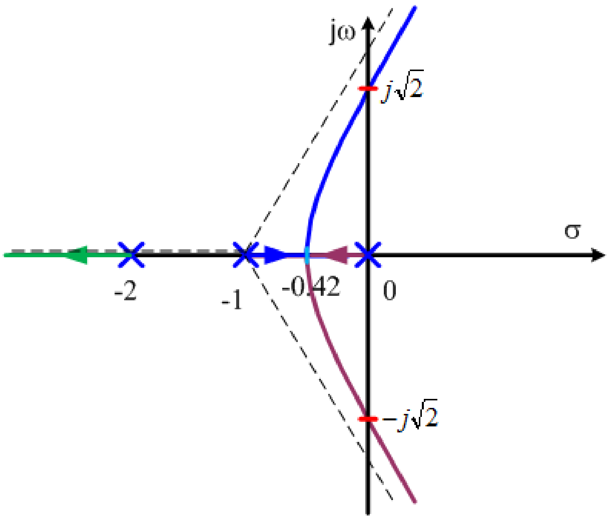


Figure 3.58: Root-locus plot

- *The complex poles are represented in Figure 3.59.*

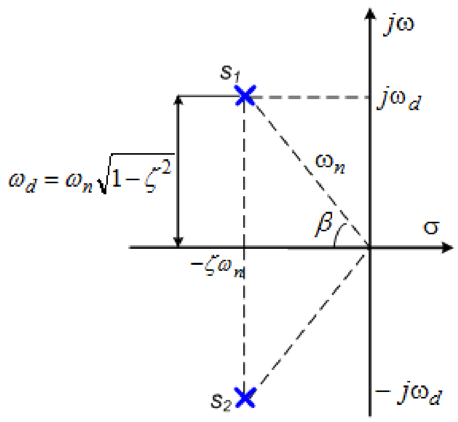
  From the figure, the distance between the origin and the pole s_1 is: $\sqrt{(\zeta\omega_n)^2 + \omega_n^2(1-\zeta^2)} = \omega_n$*, and the damping factor ζ results as follows:*

$$cos\beta = \frac{\zeta\omega_n}{\omega_n} = \zeta$$

or

$$\beta = arccos\zeta$$

Closed-loop poles with $\zeta = 0.5$ lie on the lines passing through the origin and making the angles $\pm arccos\zeta = \pm arccos 0.5 = \pm 60^o$ with the negative real axis, as shown in Figure 3.60.



[b]

Figure 3.59: Location of complex poles

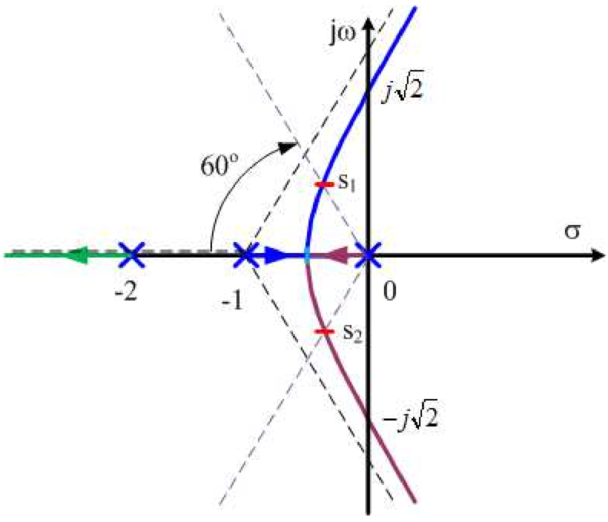


Figure 3.60: Root-locus plot and the poles with $\zeta = 0.5$

From Figure 3.60, such closed-loop poles having $\zeta = 0.5$ are obtained as follows:

$$s_{1,2} = -0.3337 \pm j0.5780$$

Note that these accurate values of the poles with $\zeta = 0.5$ can be obtained only from a very accurate root locus plot, constructed using a computer. From a root locus plot sketched following the procedure presented in this section we can only approximate the values of the required poles.

The value of k that yields such poles is found from the magnitude condition:

$$k = |s(s+1)(s+2)|_{s_{1,2}} = |-0.3337 + j0.5780| \cdot |-0.3337 + j0.5780 + 1| \cdot |-0.3337 + j0.5780 + 2| = 1.0383$$

Using this value of k, the third pole is found at $s_3 = -2.3326$ by replacing $k = 1.0383$ in the characteristic equation and computing the solutions.

3.11.4 Constructing the root locus for any variable parameter

In some cases the variable parameter k may not appear as a multiplying factor of the open-loop transfer function $G(s)H(s)$. In such cases it may be possible to rewrite the characteristic equation such that the procedure described in the previous section can be applied. The next example illustrates how to proceed in this case.

Example 3.11.5 *Consider a closed-loop system shown in Figure 3.61.*

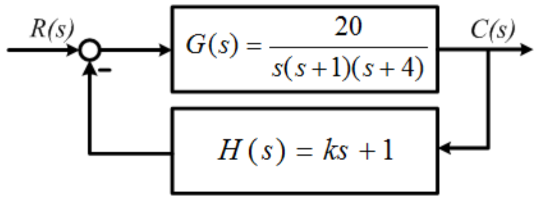


Figure 3.61: Closed-loop system

The open-loop transfer function is:

$$G(s)H(s) = \frac{20(1+ks)}{s(s+1)(s+4)}$$

Notice that the adjustable variable k does not appear as a multiplying factor of the open-loop transfer function. The characteristic equation of this system is:

$$1 + G(s)H(s) = 1 + \frac{20(ks+1)}{s(s+1)(s+4)} = \frac{s(s+1)(s+4) + 20(ks+1)}{s(s+1)(s+4)} = 0$$

or

$$s^3 + 5s^2 + 4s + 20 + 20ks = 0$$

By defining $20k = K$ and dividing by the sum of terms that do not contain k, we get:

$$1 + \frac{Ks}{s^3 + 5s^2 + 4s + 20} = 0 \text{ which is in the general form: } \quad 1 + KP(s) = 0$$

We shall now sketch the root locus of the system, from the new characteristic equation.

- *Factor $P(s)$ and obtain:*

$$P(s) = \frac{s}{s^3 + 5s^2 + 4s + 20} = \frac{s}{(s+5)(s^2+4)}$$

- *The poles and zeros of $P(s)$ are: one zero at $z_1 = 0$ and three poles: $p_{1,2} = \pm j2$, $p_3 = -5$. Thus the number of open-loop zeros and poles are $n_z = 1$, $n_p = 3$, respectively.*
- *The root locus will have $n_p = 3$ branches.*
- *On the real axis the region which starts at pole -5 to the origin (open-loop zero) belongs to the final root locus. The other two branches start on the imaginary axis, at the poles $\pm j2$ and approach the asymptotes for increasing K.*
- *The intersection of the two ($n_p - n_z = 3 - 1 = 2$) asymptotes with the real axis (the center of the asymptotes) can be found from:*

$$\sigma_A = \frac{-5 - j2 + j2 - 0}{2} = -\frac{5}{2} = -2.5$$

The angles of the asymptotes are:

$$\Phi_A = \frac{\pm 180^o(2q+1)}{2}, \quad q = 0, 1$$

or

$$\Phi_A = 90^o, 270^o$$

- *With this information available the root locus may be sketched and is presented in Figure 3.62.*

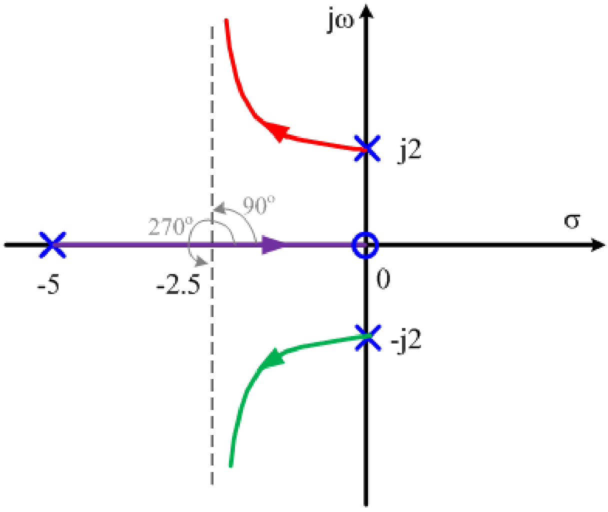


Figure 3.62: Root locus

Bibliography

Burns, R. S. (2001). *Advanced Control Engineering.* Butterworth-Heinemann.

Dorf, R. C. and Bishop, R. H. (2008). *Modern Control Systems.* Pearson Prentice Hall, 11th edition.

Nise, N. S. (2004). *Control Systems Engineering.* Wiley, 4th edition.

Ogata, K. (2002). *Modern Control Engineering.* Prentice Hall - Pearson Education International, 4th edition.