EEET 3046 Control Systems (2020)
Lecture 4: Time response

Dr. Yee Wei Law ⟨yeewei.law@unisa.edu.au⟩

Contents

1 Introduction **1**

2 Transient response characteristics **2**

3 First-order systems **3**
3.1 Impulse response 3
3.2 Step response 4

4 Second-order systems **6**
4.1 Linking pole locations to performance 12

5 System responses with zeros **14**

6 Summary **17**

1 Introduction

We design controllers that are *stable* and that exhibit desirable *transient response characteristics*. These transient response characteristics include

- whether the system response will converge to a steady-state value,
- how long the system takes to get there, and
- how much overshoot the system typically undergoes to get there.

For this, we need to evaluate the transient response ("pre-steady-state response") of our systems, under various inputs. The most common inputs are

- Impulse input: $\delta(t)$, which is an infinitely large spike at $t = 0$.
- Step input: $u(t)$, which is 1 for $t \geq 0$, but 0 elsewhere.

This sounds incredibly difficult because systems can be arbitrarily complex. Fortunately, the nature of linear systems and the beauty of transfer functions make this task relatively easy, as explained below.

Detail: Impulse

In continuous time, an impulse is represented by the Dirac delta function, which has the property

$$\int_{a-\epsilon}^{a+\epsilon} f(t)\delta(t-a)\,\mathrm{d}t = f(a) \text{ for } \epsilon > 0.$$

In discrete time, a unit impulse is represented by the Kronecker delta function:

$$\delta[k] = \begin{cases} 1 & \text{for } k = 0; \\ 0 & \text{for } k \neq 0. \end{cases}$$

In the previous lectures, we learnt how to model systems using transfer functions, and then determine the stability of these transfer functions. In the ensuing discussion, think of a system as a transfer function, which can have a order of one, two, or more.

- If it is a first-order or second-order system, there are readily available formulas for calculating its transient response characteristics.
- If it is a higher-order system, we can often express it as a sum of first- and second-order subsystems; the transient response characteristics of the higher-order system can then be approximated by those of the *dominant* second-order subsystem.

This lecture follows the directions of [Nis15, Ch. 4].

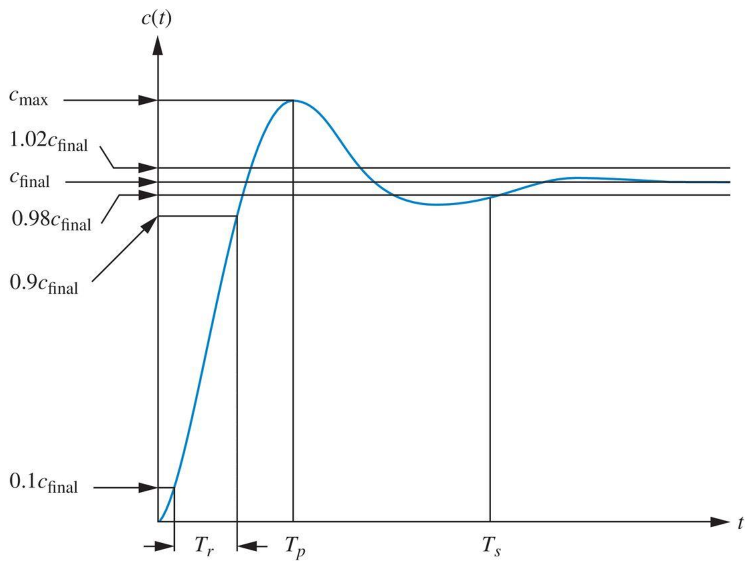


Figure 1: Characteristics of a transient response [Nis15, FIGURE 4.14]. T_r = rise time. T_p = peak time. T_s = settling time.

2 Transient response characteristics

The metrics we use to characterize the transient response of a system are defined as follows (see Figure 1):

Rise time is the time for a system response to rise from 10% to 90% of its steady-state value.

- This is the same definition used by MATLAB function `stepinfo`.
- This is sometimes defined as the time from 0 to steady-state value [Bol13], but some systems approach the steady-state value asymptotically, and never reach it.

Setting time is the time for a system response to reach and remain within $\pm 2\%$ of its steady-state value.

- Strictly speaking, this is the "2% settling time".
- This is the same definition used by MATLAB function `stepinfo`.
- The "5% settling time" is sometimes used.

Peak time is the time for a system response to reach the first, or maximum, peak.

Percent overshoot is defined as

$$\frac{\max\{c_{\max} - c_{\text{final}}, 0\}}{c_{\text{final}} - c_{\text{init}}} \times 100\%, \tag{1}$$

where $c_{\max}$ is the peak value, c_{final} is the steady-state value, c_{init} is the initial value (see Figure 1).

In the next two sections, we will learn how to calculate the transient response characteristics of first-order and second-order systems.

3 First-order systems

First-order systems are systems with one pole. In the previous lecture, we briefly learnt about poles and zeros. While poles determine the stability of a system, both poles and zeros shape the transient response of the system.

Consider the strictly proper first-order system

$$G(s) = \frac{1}{s+a}, \quad a \in \mathbb{R} \setminus \{0\}. \tag{2}$$

In this case, the system has a pole at $-a$, and no zero.

3.1 Impulse response

The *impulse response* of system (2) is the system's response to impulse input $r(t) = \delta(t)$:

$$Y(s) = G(s)R(s) = \frac{1}{s+a} \cdot 1 \implies y(t) = e^{-at}. \tag{3}$$

Note: The fact that the transfer function and impulse response of a system are equivalent has been mentioned in the previous lecture. It is clear that

- When the pole is negative ($a > 0$), the impulse response is a decaying exponential that converges to 0.
- When the pole is positive ($a < 0$), the impulse response is a growing exponential that diverges to ∞.

This confirms the fact mentioned in the previous lecture that a stable system must have poles with negative real parts.

Quiz 1

In Eq. (2), why did we not consider the case where a is complex?

If we add a zero at $-b$ to the system, the impulse response becomes

$$Y(s) = \frac{s+b}{s+a} = 1 + \frac{b-a}{s+a} \implies y(t) = \delta(t) + (b-a)e^{-at}. \tag{4}$$

Therefore we observe that while the zero does not affect stability, it does together with the pole shape the response.

3.2 Step response

The *step response* of system (2) is the system's response to step input $r(t) = u(t)$:

$$Y(s) = G(s)R(s) = \frac{1}{s+a} \cdot \frac{1}{s} = \frac{1}{a}\left(\frac{1}{s} - \frac{1}{s+a}\right) = \implies y(t) = \frac{1}{a}\left(1 - e^{-at}\right). \tag{5}$$

The constant term $1/a$ is the forced response (as well as the steady-state response), whereas the exponential term $-(1/a)e^{-at}$ is the natural response. Again, we can see the pole must be negative ($a > 0$) for the natural response to die away, i.e., for the system to be stable.

Definition: Time constant

In Eq. (5), a is called the *exponential frequency*, whereas its inverse, denoted $\tau \stackrel{\text{def}}{=} 1/a$, is called the *time constant* of the response. In the duration of a time constant, the step response increases from 0 to $1 - e^{-1} \approx 63\%$ of its steady-state value.

If we add a zero at $-b$ to the system, the step response becomes

$$Y(s) = \frac{s+b}{s+a} \cdot \frac{1}{s} = \frac{1}{a}\left(\frac{b}{s} + \frac{a-b}{s+a}\right) \implies y(t) = \frac{b}{a} + \frac{a-b}{a}e^{-at}. \tag{6}$$

Again, we observe that while the zero does not affect stability, it plays a role in shaping the response.

The step response of a first-order system is representative of many step responses we observe in the real world. Fortunately, the rise time and settling time of the step response of a first-order system are readily quantifiable:

Rise time, t_r Denote by t_{10} (t_{90}) the time when 10% (90%) of the steady-state value is reached. Then,

$$\begin{aligned} 1 - e^{-t_{10}/\tau} = 0.1 &\implies t_{10} = -\tau \ln 0.9, \\ 1 - e^{-t_{90}/\tau} = 0.9 &\implies t_{90} = -\tau \ln 0.1, \\ \therefore t_r = t_{90} - t_{10} &= \tau(-\ln 0.1 + \ln 0.9) = \tau \ln 9, \end{aligned}$$

$$\therefore t_r \approx 2.2\tau. \tag{7}$$

Settling time, t_s By definition,

$$1 - e^{-t_s/\tau} = 0.98,$$

$$\therefore t_s = -\tau \ln 0.02 = \tau \ln 50 \approx 4\tau. \tag{8}$$

Peak time is not applicable.

Percent overshoot is 0%.

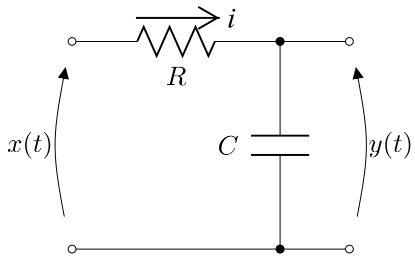


Figure 2: RC circuit for Example 1.

Example 1

In the previous lecture, we derived the transfer function for the RC circuit in Figure 2, which is a first-order system, as

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

where $\tau \stackrel{\text{def}}{=} RC$. Since $t_r \approx 2.2\tau = 2.2RC$ and $t_s \approx 4\tau = 4RC$, by picking suitable values of R and C, we can change the rise time and settling time of the step response of the RC circuit.

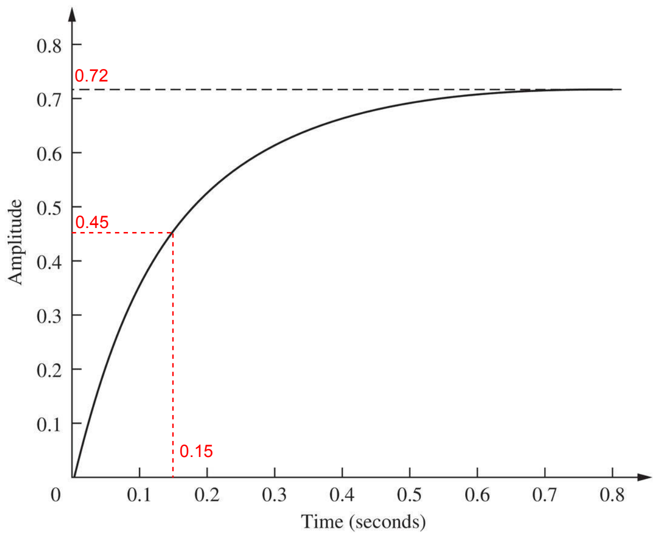


Figure 3: The observed step response of a system to be identified [Nis15, FIGURE 4.6].

Example 2

In this example, we will perform *system identification* on a supposedly first-order system, whose step response is shown in Figure 3. Suppose the system transfer function is

$$G(s) = \frac{K}{\tau s + 1},$$

in the so-called *time constant form*. Further suppose the step input is A, then the output response is

$$Y(s) = \frac{K}{\tau s+1} \cdot \frac{A}{s} = KA\left(\frac{1}{s} - \frac{\tau}{\tau s+1}\right) \implies y(t) = KA(1 - e^{-t/\tau}).$$

Figure 3 shows the steady-state value is 0.72, so if $A = 1$, then

$$K = 0.72.$$

Since τ is the time for the response to reach 63% of its steady-state value, τ is the time corresponding to $0.72 \times 0.63 \approx 0.45$, so by inspecting Figure 3,

$$\tau = 0.15.$$

Therefore, we estimate

$$G(s) = \frac{0.72}{(0.15)s+1} = \frac{4.8}{s+6.67}.$$

In fact, the curve in Figure 3 was generated for $\frac{5}{s+7}$, so our estimation is close.

4 Second-order systems

Second-order systems are systems with two poles. They are important because

- Many important systems are approximately second-order.
- They are of the lowest order that exhibit oscillations and overshoot, like what most physical processes do.
- A higher-order system can often be expressed as a sum of first- and second-order subsystems; and the transient response characteristics of the higher-order system can be approximated by those of the *dominant* second-order subsystem, i.e. the second-order subsystem with poles closest to the imaginary axis. Suppose

  $$G(s) = \frac{s^m + b_1 s^{m-1} + \cdots + b_m}{s^n + a_1 s^{n-1} + \cdots + a_n} = \frac{c_1}{s+p_1} + \cdots + \underbrace{\frac{c_2 s + d_2}{s^2 + 2\zeta_2\omega_2 s + \omega_2^2}}_{\text{Dominant subsystem}} + \cdots,$$

  and further suppose $s^2 + 2\zeta_2\omega_2 s + \omega_2^2$ has complex conjugate roots that are closest to the imaginary axis, then the corresponding second-order subsystem is dominant; and the time response of $G(s)$ can be approximated by the time response of that subsystem.

Quiz 2

Can you think of a sample higher-order system that cannot be decomposed into a sum of first- and second-order subsystems?

The canonical form of a second-order system is

$$G(s) = \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2}, \tag{9}$$

where ζ (read "zeta") is the **damping ratio**, and ω_n (read "omega n") is the **natural frequency**. The natural frequency is the system's frequency of oscillation in the absence of damping ($\zeta = 0$).

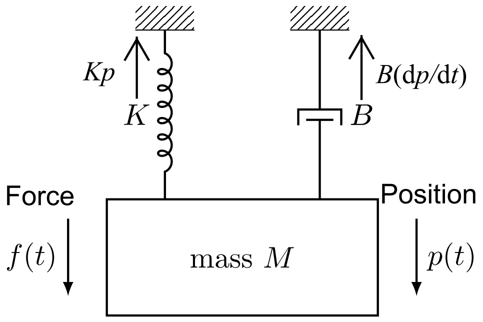


Figure 4: Mass-spring-damper system for Example 3.

Example 3

In the previous lecture, we derived the differential equation for the mass-spring-damper in Figure 4, which is a second-order system, as

$$M\ddot{p} + B\dot{p} + Kp = f.$$

Applying Laplace transformation, we get the transfer function

$$G(s) = \frac{P(s)}{F(s)} = \frac{1}{M}\frac{1}{s^2 + \frac{B}{M}s + \frac{K}{M}}, \tag{10}$$

where $P(s)$ and $F(s)$ are the Laplace transforms for $p(t)$ and $f(t)$ respectively. Upon inspecting $G(s)$, intuition suggests

- If B is small relative to M and the mass oscillates, then the oscillation frequency is related to K (a stiffer spring oscillates more) and M (a smaller inertia changes direction more easily). In fact, based on Eqs. (9)–(10), the natural frequency is

  $$\omega_n = \sqrt{\frac{K}{M}},$$

  consistent with our physical understanding.

- Provided the viscous damper has high enough damping B relative to the mass M, the mass does not oscillate. In fact, based on Eqs. (9)–(10), the damping ratio is

  $$\zeta = \frac{B}{2\omega_n M} = \frac{B}{2\sqrt{KM}},$$

  which is, again, consistent with our physical understanding. If B is really small, the mass-spring-damper will oscillate at a frequency close to $\omega_n = \sqrt{K/M}$.

It is extremely important to understand the effect of ζ on

- the location of the poles of the second-order system (see Figure 5); and
- the time response of the second-order system (see Figure 6).

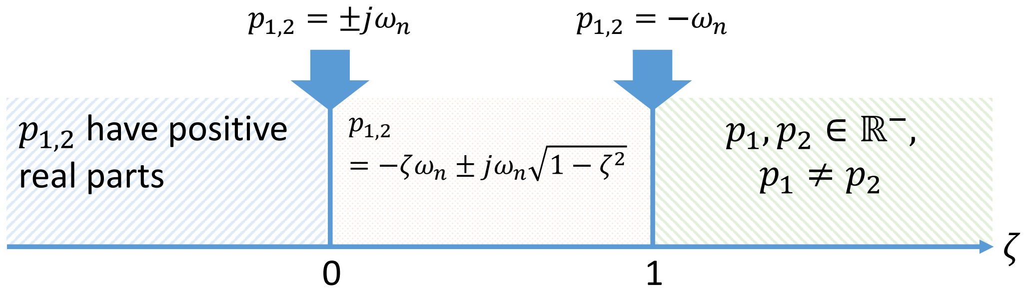


Figure 5: The location of the poles, denoted p_1 and p_2, depends the value of ζ.

The poles of the second-order system are readily calculable using the formula for solving quadratic equations:

$$p_{1,2} = \frac{1}{2}\left(-2\zeta\omega_n \pm \sqrt{4\zeta^2\omega_n^2 - 4\omega_n^2}\right) = -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2 - 1}$$
$$= \begin{cases} \text{has a positive real part} & \text{if } \zeta < 0 \text{ (unstable);} \\ \pm j\omega_n & \text{if } \zeta = 0 \text{ (undamped);} \\ -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} & \text{if } 0 < \zeta < 1 \text{ (underdamped);} \\ -\omega_n & \text{if } \zeta = 1 \text{ (critically damped);} \\ -\zeta\omega_n \pm \omega_n\sqrt{\zeta^2-1} & \text{if } 1 < \zeta \text{ (overdamped).} \end{cases} \tag{11}$$

Quiz 3

Which case in Eq. (11) does not apply to the mass-spring-damper?

For each of the cases in Eq. (11), a sample time response is sketched in Figure 6. Of particular interest is the *underdamped* case, where the control designer often specifies a nonzero percent overshoot, i.e., a value between 0 and 1 for ζ, in exchange for a shorter settling time than what critical damping or overdamping provides. The rest of this section is the most important part of this lecture, where we calculate the transient response characteristics of the second-order system.

Our starting point is an algebraic expression for the system's step response:

$$Y(s) = \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2} \cdot \frac{A}{s} = \frac{c_1}{s} - \frac{c_2 s + c_3}{s^2 + 2\zeta\omega_n s + \omega_n^2}. \tag{12}$$

Performing partial fraction expansion, we have

$$\begin{aligned} & c_1 s^2 + 2\zeta\omega_n c_1 s + c_1\omega_n^2 - c_2 s^2 - c_3 s = KA \\ \Longrightarrow\ & c_1 = c_2, \qquad 2\zeta\omega_n c_1 = c_3, \qquad c_1\omega_n^2 = KA \\ \Longrightarrow\ & c_1 = \frac{KA}{\omega_n^2}, \qquad c_2 = \frac{KA}{\omega_n^2}, \qquad c_3 = \frac{2\zeta\omega_n KA}{\omega_n^2}. \end{aligned}$$

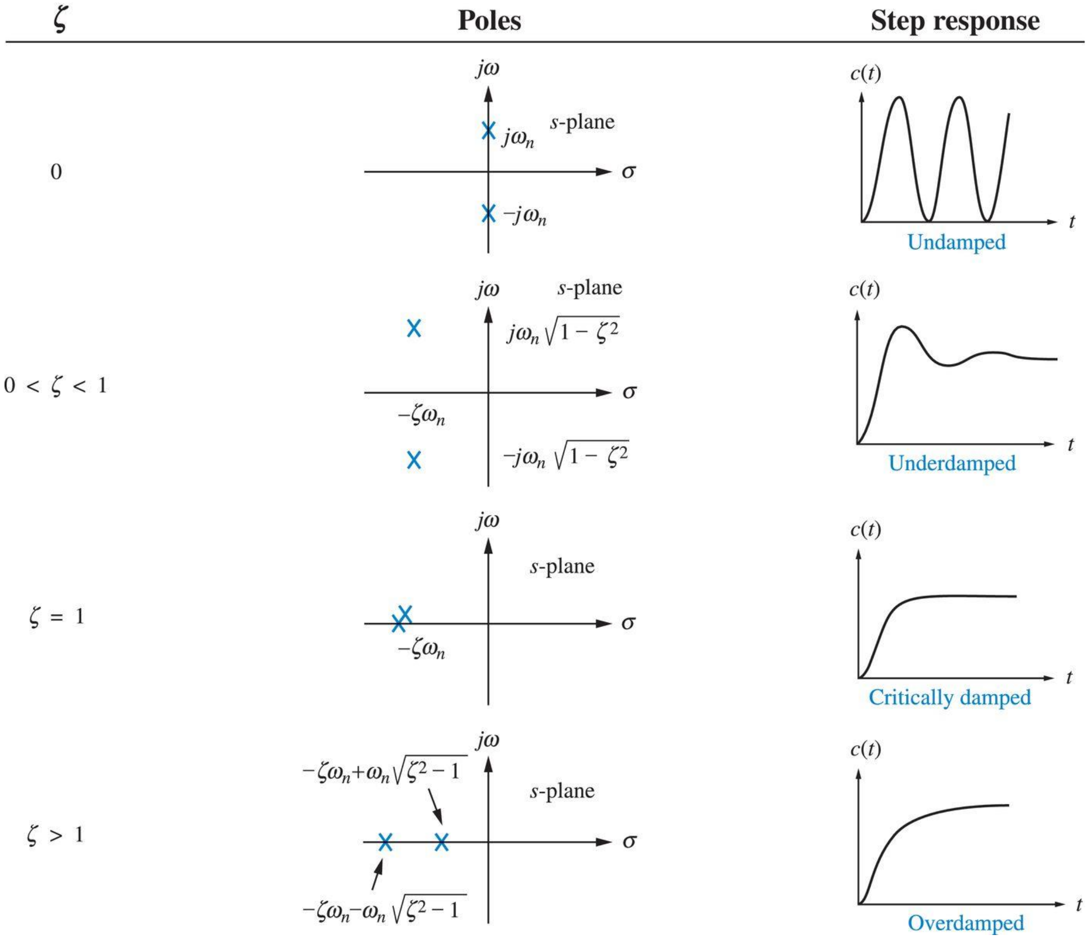
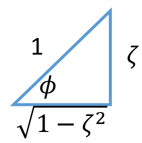


Figure 6: Step responses of second-order systems [Nis15, FIGURE 4.11].

Therefore,

$$
\begin{aligned}
Y(s) &= \frac{KA}{\omega_n^2}\left\{\frac{1}{s} - \frac{s+2\zeta\omega_n}{s^2+2\zeta\omega_n s+\omega_n^2}\right\} = \frac{KA}{\omega_n^2}\left\{\frac{1}{s} - \frac{(s+\zeta\omega_n) + \frac{\zeta}{\sqrt{1-\zeta^2}}\omega_n\sqrt{1-\zeta^2}}{(s+\zeta\omega_n)^2+\omega_n^2(1-\zeta^2)}\right\} \\
&= \frac{KA}{\omega_n^2}\left\{\frac{1}{s} - \frac{s+\zeta\omega_n}{(s+\zeta\omega_n)^2+\omega_n^2(1-\zeta^2)} - \left(\frac{\zeta}{\sqrt{1-\zeta^2}}\right)\frac{\omega_n\sqrt{1-\zeta^2}}{(s+\zeta\omega_n)^2+\omega_n^2(1-\zeta^2)}\right\} \\
\implies y(t) &= \frac{KA}{\omega_n^2}\left\{1 - e^{-\zeta\omega_n t}\cos(\omega_n\sqrt{1-\zeta^2}t) - \frac{\zeta}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_n t}\sin(\omega_n\sqrt{1-\zeta^2}t)\right\} \\
&= \frac{KA}{\omega_n^2}\left\{1 - e^{-\zeta\omega_n t}\left[\cos(\omega_n\sqrt{1-\zeta^2}t) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin(\omega_n\sqrt{1-\zeta^2}t)\right]\right\}.
\end{aligned}
\tag{13}
$$

We shall simplify the equation above by combining the cosine and sine terms into a single cosine term. Since $0 < \zeta < 1$, we can draw a right-angle triangle with two orthogonal sides of lengths ζ and $\sqrt{1-\zeta^2}$, and a hypotenuse of length $\sqrt{\zeta^2 + (\sqrt{1-\zeta^2})^2} = 1$. Then,



$$\tan\phi = \frac{\zeta}{\sqrt{1-\zeta^2}} \qquad \text{and} \qquad \cos\phi = \sqrt{1-\zeta^2}. \tag{14}$$

$$\begin{aligned}
\cos(\omega_n\sqrt{1-\zeta^2}t) + \frac{\zeta}{\sqrt{1-\zeta^2}}\sin(\omega_n\sqrt{1-\zeta^2}t) &= \frac{\cos\phi\cos(\omega_n\sqrt{1-\zeta^2}t) + \sin\phi\frac{\zeta}{\sqrt{1-\zeta^2}}\sin(\omega_n\sqrt{1-\zeta^2}t)}{\cos\phi} \\
&= \frac{\cos(\omega_n\sqrt{1-\zeta^2}t - \phi)}{\sqrt{1-\zeta^2}}.
\end{aligned} \tag{15}$$

Therefore, finally, the step response of the second-order system of Eq. (9) is

$$y(t) = \frac{KA}{\omega_n^2}\left\{1 - \frac{e^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}}\cos\left(\omega_n\sqrt{1-\zeta^2}t - \phi\right)\right\}, \text{ where } \phi = \arctan\frac{\zeta}{\sqrt{1-\zeta^2}}. \tag{16}$$

In Eq. (16), we can see the natural response is an exponentially damped sinusoid.

- The term $\zeta\omega_n = |\Re(p_{1,2})|$ determines how fast the exponential factor decays, hence it is called the **exponential damping frequency**.
- The term $\omega_n\sqrt{1-\zeta^2} = \Im(p_{1,2})$ is the frequency of the exponentially damped sinusoid, hence it is called the **damped frequency of oscillation**.

Note: $\Re$ represents the real part, whereas $\Im$ represents the imaginary part of a complex number.

With Eq. (16), we now derive formulas for the rise time, settling time, peak time, and overshoot.

Rise time, t_r There is no exact formula, but

$$t_r \approx 1.8/\omega_n, \tag{17}$$

or based on [DB11, Eq. (5.17)],

$$t_r \approx (2.16\zeta + 0.60)/\omega_n. \tag{18}$$

These formulas are not as useful as the formula for settling time.

Settling time, t_s After the settling time, $y(t)$ is constrained within a sinusoid of amplitude 0.02, so

$$\frac{e^{-\zeta\omega_n t_s}}{\sqrt{1-\zeta^2}} = 0.02 \implies t_s \approx \frac{4}{\zeta\omega_n}. \tag{19}$$

You will derive Eq. (19) in the tutorial under guidance.

Peak time, t_p To find the peak, we determine when $\dot{y}(t) = 0$. Instead of differentiating in the time domain, it is easier to perform the equivalent in the Laplace domain.

$$\begin{aligned}
\mathcal{L}\{\dot{y}(t)\} = sY(s) = s \cdot \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2} \cdot \frac{1}{s} &= \frac{K}{s^2 + 2\zeta\omega_n s + \omega_n^2} \\
&= \frac{K}{\omega_n\sqrt{1-\zeta^2}} \frac{\omega_n\sqrt{1-\zeta^2}}{(s+\zeta\omega_n)^2 + \omega_n^2(1-\zeta^2)} \\
\therefore \dot{y}(t_p) = 0 \implies & \frac{K}{\omega_n\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t_p} \sin(\omega_n\sqrt{1-\zeta^2}t_p) = 0 \\
\implies & \sin(\omega_n\sqrt{1-\zeta^2}t_p) = 0 \\
\implies & \omega_n\sqrt{1-\zeta^2}t_p = n\pi, \text{ where } n = 0, \pm 1, \pm 2, \ldots
\end{aligned} \tag{20}$$

Since $t_p > 0$, we can rule out negative n and zero n, hence,

$$t_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}}. \tag{21}$$

Percent overshoot, $\%OS$ At peak time, the step response reaches the peak overshoot:

$$\begin{aligned}
y(t_p) &= \frac{KA}{\omega_n^2}\left\{1 - \frac{e^{-\zeta\omega_n t_p}}{\sqrt{1-\zeta^2}}\cos(\omega_n\sqrt{1-\zeta^2}t_p - \phi)\right\} = \frac{KA}{\omega_n^2}\left\{1 - \frac{e^{-\zeta\pi/\sqrt{1-\zeta^2}}}{\sqrt{1-\zeta^2}}\cos(\pi - \phi)\right\} \\
&= \frac{KA}{\omega_n^2}\left\{1 + e^{-\zeta\pi/\sqrt{1-\zeta^2}}\right\}.
\end{aligned} \tag{22}$$

Since the steady-state value is 1, the percent overshoot is simply

$$\%OS = \exp\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right) \times 100\%. \tag{23}$$

Notice how the overshoot depends only on the damping ratio. When designing a controller, it is common practice to determine the damping ratio from a specified percent overshoot. With just some simple algebra, we can get

$$\zeta = -\frac{\ln(OS)}{\sqrt{\pi^2 + \ln^2(OS)}}, \tag{24}$$

where OS is in fraction, not percentage.

Now, we have all the formulas we need to

- calculate ζ, ω_n from the desired transient response characteristics (usually overshoot and settling time); and
- calculate the pole locations from ζ, ω_n using Eq. (11).

Note: When designing a closed-loop system, the pole locations are the locations of the **closed-loop poles**.

4.1 Linking pole locations to performance

When designing a controller, there is usually flexibility in placing the poles. It is essential to develop an understanding of how moving the poles in the s plane affects the performance. Recall the poles of an underdamped second-order system with damping ratio ζ and natural frequency ω_n are

$$p_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} = \omega_n e^{j(\pi\pm\theta)}, \text{ where } \theta = \arctan\frac{\sqrt{1-\zeta^2}}{\zeta}. \tag{25}$$

Thus,

- The settling time is related to the real part of the poles through Eqs. (19) and (25):

$$t_s \approx \frac{4}{\zeta\omega_n} = \frac{4}{|\Re(p_{1,2})|}. \tag{26}$$

  This means poles at the same distance away from the imaginary axis of the s plane provide the same amount of settling time (see Figure 7(a)).

> Quiz 4
>
> To *decrease* the settling time, should the poles be moved horizontally to the *left or right*?

- The peak time is related to the imaginary part of the poles through Eqs. (21) and (25):

$$t_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} = \frac{\pi}{\Im(p_{1,2})}. \tag{27}$$

  This means poles at the same distance away from the real axis of the s plane provide the same amount of peak time (see Figure 7(b)).

> Quiz 5
>
> To *decrease* the peak time, should the poles be moved vertically *from or toward* the real axis?

- The overshoot is related to the angle of the poles through Eqs. (23) and (25):

$$OS = \exp\left(-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}\right) = \exp\left(-\frac{\pi}{\tan\theta}\right). \tag{28}$$

  This means poles at the same direction from the origin provide the same amount of overshoot (see Figure 7(c)). This is why radial lines (i.e., lines radiating from the origin) on the s plane are called *overshoot lines* or *damping ratio lines*.

> Quiz 6
>
> To *decrease* the overshoot, should θ be *increased or decreased*?

> **Quiz 7**
>
> By the polar representation of the poles in Eq. (25), what do all the poles distributed on the same circle centered at the origin have in common?

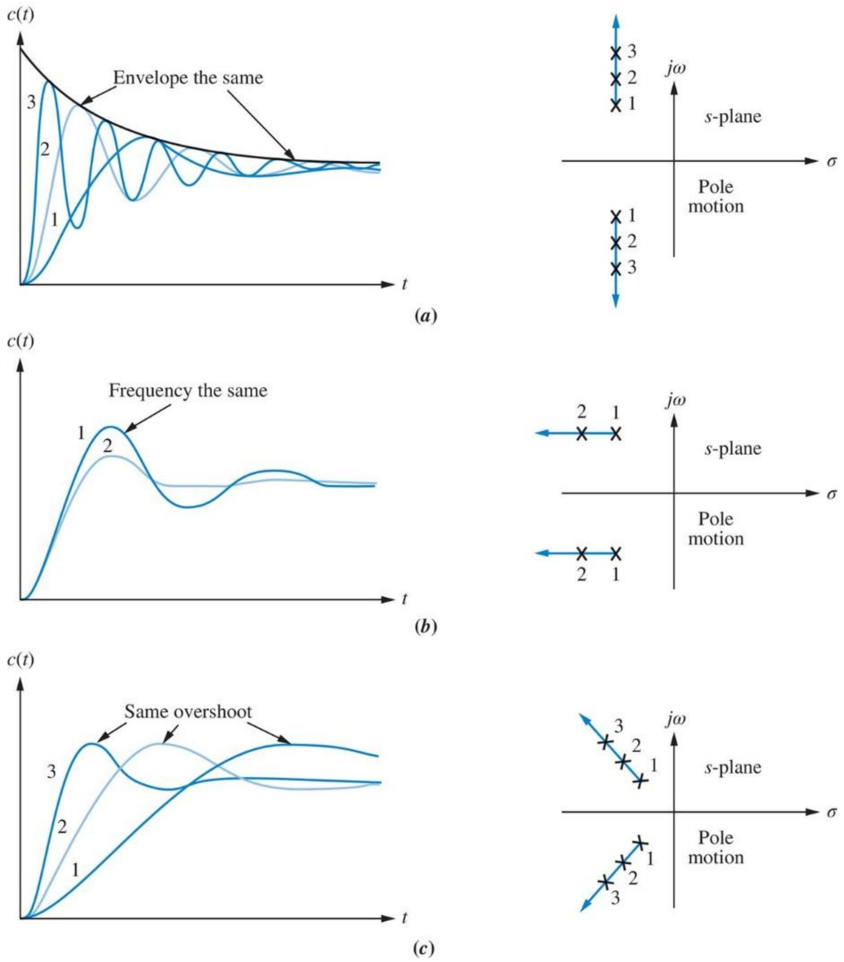


Figure 7: Step responses of second-order underdamped systems as poles move (a) vertically, (b) horizontally, (c) radially in the s plane.

When designing a controller,

1. It is customary to start with a desired percent overshoot, which determines the overshoot line (equivalently, damping ratio line) on the s plane. There are actually two overshoot lines, that are symmetrical about the real axis (see Figure 7(c)), but we usually just talk about the one above the real axis; the one below is implied.

2. Then on the overshoot line, the designer picks a pole that gives a satisfactory settling time; the other pole is implied. The further away the poles are from the origin, the shorter the settling time and peak time. However, the poles cannot be arbitrarily far away, because correspondingly these require arbitrarily large controller gains, which saturate the actuators earlier, use more energy, cause more wear and tear, and may in the worst case destabilize the system.

Figure 8 shows now that we can calculate the locations of closed-loop poles, we are close to being able to design controllers.

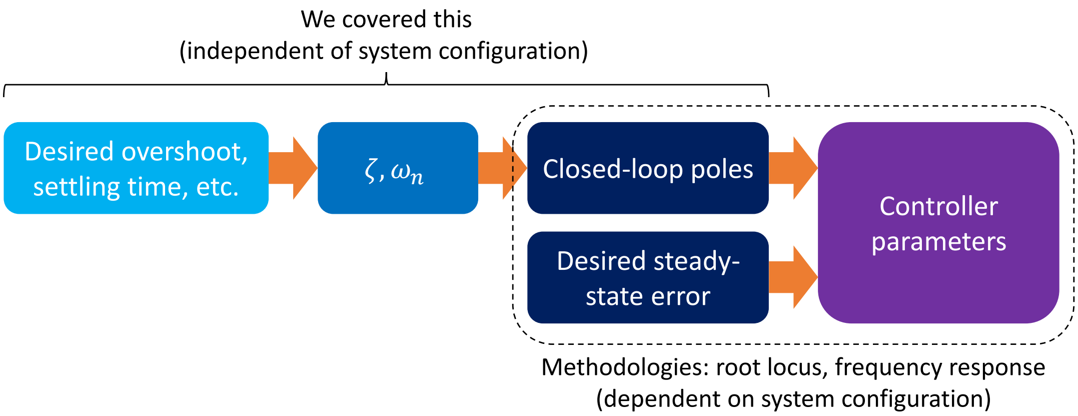


Figure 8: Controller design starts with the specification of desired transient response characteristics.

Example 4

This example is taken from [DB11, Example 5.1]. Suppose for the system

$$G(s) = \frac{K}{s^2 + as + K},$$

we need to select the gain K and parameter a such that the step response has

- an overshoot of less than $e^{-\pi} \approx 4.3214\%$; and
- a settling time of less than 4 seconds.

Solution: The damping ratio, ζ, for an overshoot of less than $e^{-\pi}$ must satisfy

$$\zeta > -\frac{\ln(e^{-\pi})}{\sqrt{\pi^2 + \ln^2(e^{-\pi})}} = \frac{\pi}{\sqrt{2\pi^2}} = \frac{1}{\sqrt{2}}.$$

Since the settling time $t_s \approx \frac{4}{\zeta\omega_n}$,

$$\frac{4}{\zeta\omega_n} < 4 \implies \zeta\omega_n > 1 \implies \omega_n > \frac{1}{\zeta}.$$

Graphically, the poles of $G(s)$ must lie in the shaded region in Figure 9. Consequently, we must have

$$a = 2\zeta\omega_n > 2, \text{ and } K = \omega_n^2 > \frac{1}{\zeta^2}, \text{ for any } \zeta > \frac{1}{\sqrt{2}}.$$

5 System responses with zeros

So far in the discussion of second-order systems, we have not considered the effect of zeros. Given system response $Y_o(s)$, where the subscript "o" denotes "original", how will it change if we add a zero to the system, as such: $(s + a)Y_o(s)$? Observe the modified system response:

$$Y(s) = (s + a)Y_o(s) = sY_o(s) + aY_o(s). \tag{29}$$

- $sY_o(s)$ is the Laplace transform of $\dot{y}(t)$, i.e., Laplace transform of the rate of change of $y_o(t)$.

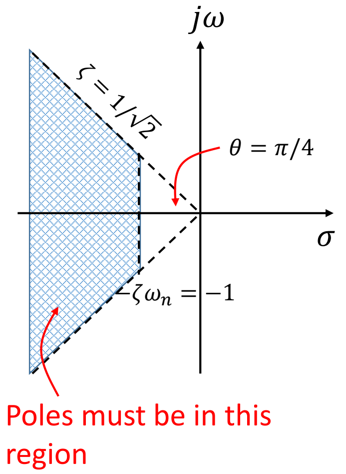


Figure 9: Region in which the poles of Example 4 must lie.

- If a is large, $Y(s) \approx aY_o(s)$, i.e., $Y(s)$ is approximately an a-multiple of $Y_o(s)$ (see Figure 10(a)).
- If a is small, $Y(s) \approx sY_o(s)$, i.e., $Y(s)$ is a differentiated version of $Y_o(s)$, and consequently has more overshoot than $Y_o(s)$ (see Figure 10(b)).

The case where $a < 0$ is equivalent to having a zero in the open right-half plane.

Definition: Nonminimum-phase system [DB11, p. 570]

A system with one or more zeros in the open right-half plane is called a **nonminimum-phases system**.

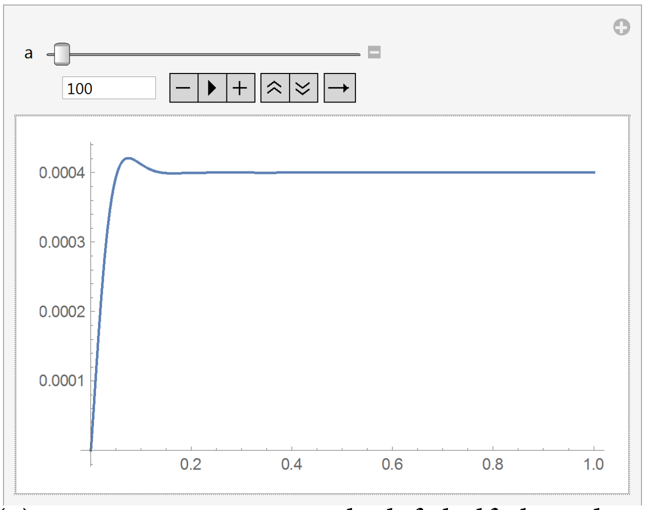
Nonminimum-phase systems are so-called because they have more phase than systems with only left-half-plane zeros; this will become evident once we study frequency response later. *Undershoot* is characteristic of nonminimum-phase systems. Figs. 10(c) and 10(d) show the closer the right-half-plane zero is to the imaginary axis, the larger is the undershoot. Undershoot is problematic because the system response always starts by going in the wrong direction before recovering in the right direction, rendering the system inefficient. The inherent performance deficiency of nonminimum-phase systems makes them undesirable, and even unfit for some control schemes such as feedforward control and feedback linearization, but their unavoidable existence in the real world means we should develop some good understanding of these systems. As a start, let us study a classic example of a real-world nonminimum-phase system in Example 5. We will encounter nonminimum-phase systems in later lectures again.

Example 5

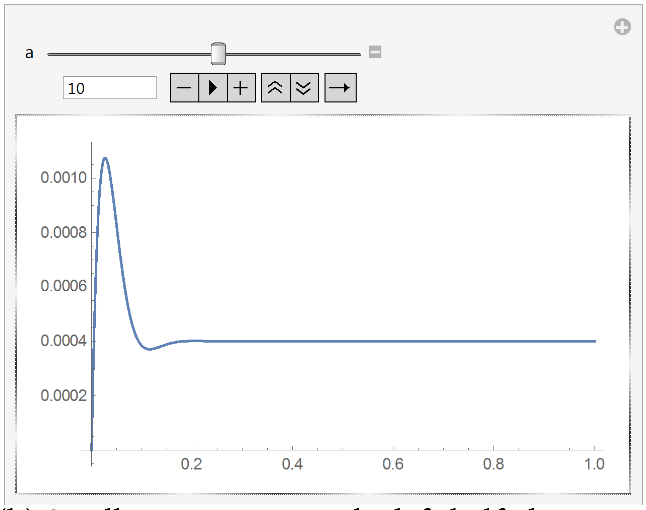
Consider the aircraft in Figure 11, which is initially cruising at an altitude of $h = h_0$. To increase its altitude, the aircraft rotates the elevator by an angle of E, generating a small aerodynamic force of L_E on the elevator, and thus a torque about the center of gravity CG. The torque rotates the aircraft by α about CG. The lift force applied to the wings is proportional to α, i.e.,

$$L_W = C_{ZW}\alpha, \tag{30}$$

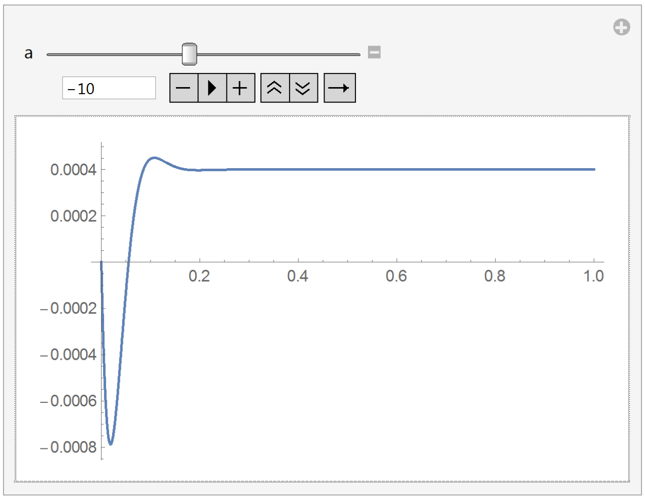
where C_{ZW} is the lift coefficient of the wing. Similarly, L_E is proportional to the angle between



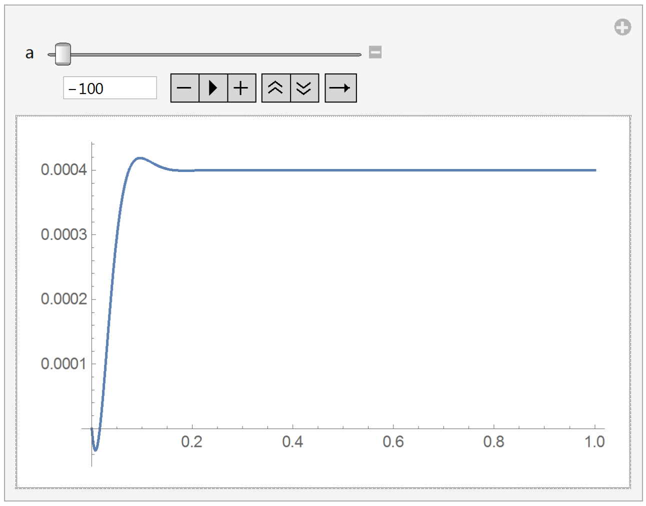
(a) Large system zero on the left-half plane does not significantly change the original overshoot.



(b) Small system zero on the left-half plane amplifies the overshoot.



(c) Small system zero on the right-half plane causes large undershoot.



(d) Large system zero on the right-half plane causes small undershoot.

Figure 10: Plots of the step response of the system $\dfrac{s+a}{s^2+2\zeta\omega_n s+\omega_n^2}$ subjected to step input $\dfrac{1}{as}$, where $\zeta = 1/\sqrt{2}, \omega_n = 50$.

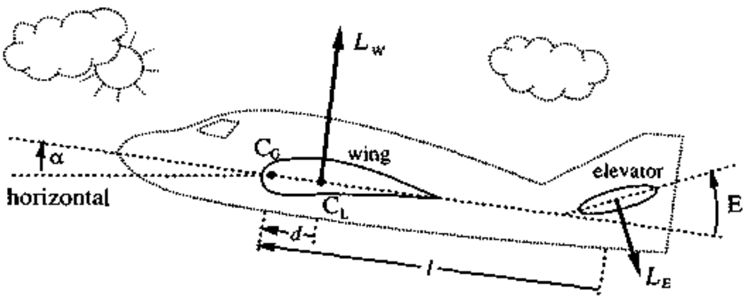


Figure 11: The aircraft altitude control system is a classic example of a nonminimum-phase [SL91, Example 6.5].

the horizontal and elevator, i.e.,

$$L_E = C_{ZE}(E - \alpha), \tag{31}$$

where C_{ZE} is the lift coefficient of the elevator. Furthermore, various aerodynamic forces create friction torques proportional to $\dot{\alpha}$, of the form $b\dot{\alpha}$. Applying Newton's second law, we can thus write

$$\begin{aligned} m\ddot{h} &= L_W \cos(\alpha) - L_E \cos(E-\alpha) = C_{ZW}\alpha \cos(\alpha) - C_{ZE}(E-\alpha)\cos(E-\alpha) \\ &\approx C_{ZW}\alpha - C_{ZE}(E-\alpha) \qquad \text{(assuming angles } \alpha \text{ and } E-\alpha \text{ are small)} \\ &= (C_{ZW}+C_{ZE})\alpha - C_{ZE}E. \end{aligned} \tag{32}$$

Applying the rotational version of Newton's second law, we can also write

$$\begin{aligned} & J\ddot{\alpha} + b\dot{\alpha} = lC_{ZE}(E-\alpha) - dC_{ZW}\alpha \\ \implies & J\ddot{\alpha} + b\dot{\alpha} + (C_{ZE}l + C_{ZW}d)\alpha = C_{ZE}lE \\ \implies & (Js^2 + bs + C_{ZE}l + C_{ZW}d)\alpha(s) = C_{ZE}lE(s) \\ \implies & \alpha(s) = \frac{C_{ZE}l}{Js^2 + bs + C_{ZE}l + C_{ZW}d}E(s). \end{aligned} \tag{33}$$

Substituting Eq. (33) into the Laplace transform of Eq. (32), we get

$$\begin{aligned} & ms^2H(s) = (C_{ZW}+C_{ZE})\alpha(s) - C_{ZE}E(s) \\ \implies & ms^2H(s) = \left[\frac{(C_{ZW}+C_{ZE})C_{ZE}l}{Js^2 + bs + C_{ZE}l + C_{ZW}d} - C_{ZE}\right]E(s) \\ \implies & \frac{H(s)}{E(s)} = -\frac{C_{ZE}Js^2 + C_{ZE}bs - C_{ZW}C_{ZE}(l-d)}{ms^2(Js^2 + bs + C_{ZE}l + C_{ZW}d)}. \end{aligned} \tag{34}$$

Judging by the signs of the coefficients of the numerator $(C_{ZE}Js^2 + C_{ZE}bs - C_{ZW}C_{ZE}(l-d))$, we can conclude that $H(s)/E(s)$ has one positive zero and one negative zero. Thus $H(s)/E(s)$ is nonminimum-phase.

The nonminimum phase of $H(s)/E(s)$ is consistent with the common observation that the initial effect of a step change in E is an instantaneous downward force on the elevator, leading to an initial downward acceleration of the aircraft's CG. The step change in E also creates a torque about the CG, which builds up the pitch angle α, thereby creating an increasing upward lift force on the wings and body. This lift eventually takes over the downward force on the elevator. Clearly, it is important for pilots to recognize such nonminimum-phase behavior, especially when flying at low altitudes.

6 Summary

- To design a stable and responsive control system, we need to know how the system parameters are related to the transient response characteristics, e.g., the rise time, peak time, settling time and overshoot.
- A higher-order system can often be expressed as a sum of first- and second-order subsystems; and its transient response characteristic can be approximated by those of the dominant second-order subsystem. "Dominant" means having poles closest to the imaginary axis.
- Many real-world processes can be modeled as a first-order system.

- Eqs. (7) and (8) can be used to estimate rise time and settling time of the step response of such a system.
  - The time constant and magnitude of a step response can be used to estimate the parameters of a first-order system (see Example 1).

- A second-order system is characterized by its damping ratio ζ and natural frequency ω_n.
  - ζ determines whether a system is undamped, underdamped, critically damped or overdamped; the underdamped case is of interest because of shorter settling time.
  - Eqs. (19), (21) and (23) can be used to calculate the settling time, peak time and overshoot.
  - Damping ratio determines the overshoot; the absolute real part of the system poles determines the settling time; the absolute imaginary part of the system poles determines the peak time.
  - Specifying the desired overshoot and settling time (alternatively, peak time) is equivalent to specifying the desired dominant closed-loop poles — this is the first step of a typical controller design process.

- System zeros affect transient response. The nearer they are to the imaginary axis, the worse they affect the performance (see Figure 10). Positive zeros make the system nonminimum-phase, and the nearer they are to the imaginary axis, the worse the undershoot becomes.

References

[Bol13] W. Bolton. *Mechatronics: Electronic control systems in mechanical and electrical engineering.* Prentice Hall, 5th edition, 2013.

[DB11] R.C. Dorf and R.H. Bishop. *Modern Control Systems.* Pearson, 12th edition, 2011.

[Nis15] Norman S. Nise. *Control Systems Engineering.* Wiley, 7th edition, 2015.

[SL91] Jean-Jacques E. Slotine and Weiping Li. *Applied Nonlinear Control.* Prentice Hall, 1991.