TRANSFER FUNCTIONS

This article considers continuous time systems based on differential equations and discrete time systems based on difference equations. In the following sections, we will look at continuous time transfer function models for the resistor, inductor, and capacitor elements using the Laplace transform. We will use these models to construct first-order and second-order transfer function models and will discuss the construction of complex transfer functions from these subsystems. We will examine the properties of transfer functions and introduce special transfer functions like sensitivity and complementary sensitivity. We will also show how the single-input, single-output (SISO) transfer function can be extended to the multiinput, multioutput (MIMO) transfer function matrix. Also, we will take an abbreviated look at state-space representations of these systems. Finally, we will examine discrete time transfer functions and the Z transform and we will introduce model identification and parameter estimation.

A transfer function is a method for representing a dynamic mathematical model of a system. It is an algebraic expression that models the outputs of a system as a function of the system inputs. The input/output system is defined by the user of the model. Typically, a transfer function is used to model a *physical* system, which is something that can be described by the laws of physics. For example, consider the physical system defined by the resistor shown in Fig. 1(a). We define the system input to be the voltage, V, across the resistor, and we define the system output to be the electrical current, I, through the resistor. For systems with only one input and one output we can express the transfer function model as the ratio of the output divided by the input, which is the slope of the line shown in Fig. 1(b).

$$\frac{I}{V} = \frac{1}{R} \tag{1}$$

By definition, R is the resistance of the resistor. Transfer functions have a number of assumptions associated with them. For example, we have assumed that the resistance is constant in Eq. (1). As current passes through the resistor, it could cause self-heating from power dissipation (I^2R). This would cause a change in the temperature and the resistance of the resistor, violating the constant resistance assumption and possibly causing a modeling error. Thus the assumptions are an essential part of the model.

The model of the resistor described by Eq. (1) is a static, nondynamic transfer function that is a trivial example. A transfer function typically represents the dynamic characteristics of a system by parameterizing the transfer function with an operator that is indicative of the dynamics. The operator is usually the Laplace variable s, which results from con-

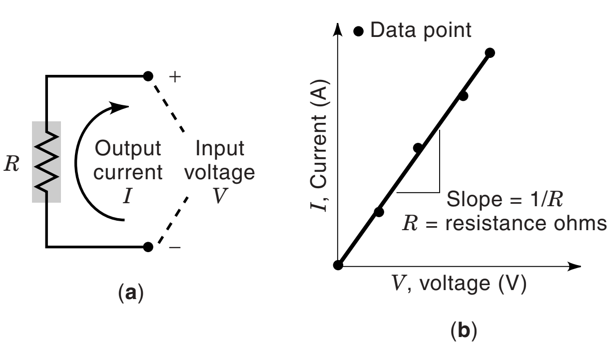
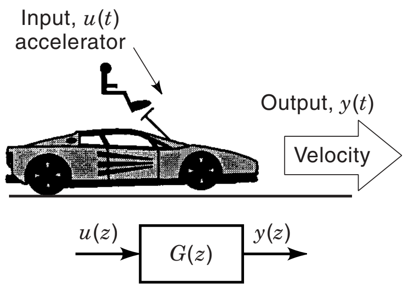
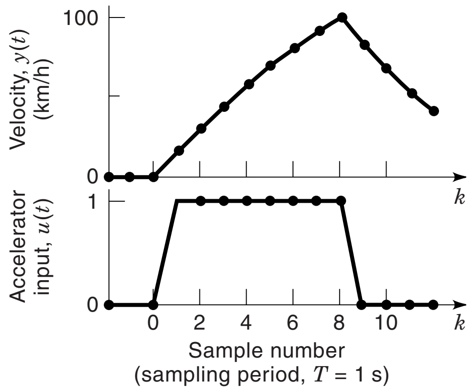


Figure 1. (a) Electrical resistor model showing components of a simple, static, transfer function with input voltage and output current. (b) Nondynamic, linear relationship between input voltage and output current in resistor model.



(a)



(b)

Figure 2. (a) Model of an automotive system with accelerator input and velocity output. (b) Dynamic velocity response of the model of the automotive system to a step input in the accelerator.

verting a differential equation to an algebraic equation using the Laplace transform, or the operator is the z variable, which results from converting a difference equation to an algebraic equation using the Z transform. We will use the automotive system sketched in Fig. 2(a) to introduce the dynamic transfer function. We can model the automobile with the accelerator as the system input and the velocity as the system output. We can then use the mathematical model to calculate the velocity as a function of the accelerator value and time. We can use this response to calculate how long it will take the vehicle to accelerate to some velocity from zero velocity after the driver provides an accelerator input, thus predicting the automobile acceleration performance. We will use a difference equation for this model with a fixed sampling timestep, T, equal to one second. The difference equation calculates the velocity at time = kT based on the accelerator position at time = kT and the previous value of the velocity at time = $(k-1)T$ and is shown in Eq. (2):

$$y(k) = 0.93 \cdot y(k-1) + 16 \cdot u(k) \tag{2}$$

where u is the accelerator input ranging from zero to one (0 to 100%); y is the automobile velocity in kilometers per hour; k is the integer index where time = kT; and $y(k-1)$ represents the velocity at the previous timestep. Table 1 shows the values for Eq. (1), starting at time equal to zero ($k = 0$) and the initial velocity equal to zero. The data in this table are plotted in Fig. 2(b). The top plot in Fig. 2(b) shows the exponential velocity response, and the bottom figure shows the step input to the accelerator. If we tune and validate the model in Eq. (2) against data taken from a real automobile, then we can use this model to represent the velocity response of that automobile. We could use this model in the design of a cruise control system for the automobile. The difference equation in Eq. (2) can be converted to an algebraic equation using the Z transform. A detailed discussion of the Z transform is outside the scope of this article, but it just requires a simple modification to Eq. (2) in this example. The transformation results in Eq. (3), which is in the z domain.

$$y(z) = 0.93 \cdot z^{-1} y(z) + 16 \cdot u(z) \tag{3}$$

where $u(z)$ is the Z transform of $u(k)$; $y(z)$ is the Z transform of $y(k)$; and $z^{-1}y(z)$ is the Z transform of $y(k-1)$, where z^{-1} is the unit time delay operator. Note that Eq. (3) is an algebraic equation parameterized by z. The only variables in Eq. (3) are the input, $u(z)$, and the output, $y(z)$, so Eq. (3) can be manipulated to obtain the output over input ratio:

$$\frac{y(z)}{u(z)} = G(z) = \frac{16}{1 - 0.93 \cdot z^{-1}} = \frac{16z}{z - 0.93} \tag{4}$$

Table 1. Automotive Discrete Model Velocity Response as a Function of Time to a Step Input in the Accelerator

Sample Index, k and Time = kT	Accelerator Input $u(k)$ $(0-1)$	Vehicle Velocity $y(k)$ km/hour	One Step Delayed Vehicle Velocity $y(k-1)$ km/hour
0	0	0	0
1	1	16	0
2	1	30.9	16
3	1	44.7	30.9
4	1	57.6	44.7
5	1	69.6	57.6
6	1	80.7	69.6
7	1	91.0	80.7
8	1	100.7	91.0
9	0	93.6	100.7
10	0	87.0	93.6
11	0	81.0	87.0
12	0	75.3	81.0

and $G(z)$ defines the discrete time, SISO dynamic transfer function model of the automotive system with a fixed sampling period, $T = 1$ second. Once we have validated the model by comparing it to the response of the real automobile, then we can use the model in place of the real system to perform the desired numerical studies, subject to the model assumptions. (An example of an assumption for a specific make and model of automobile might be the number of passengers and weight of the cargo that the vehicle was carrying.)

Note that $G(z)$ in Eq. (4) is comprised of a numerator and a denominator polynomial in the operator variable, z. The roots of the numerator polynomial are called system zeros, and the roots of the denominator polynomial are called system poles. For Eq. (4), there is one zero, $z = 0$, and one pole, $z = 0.93$. The pole and zero locations convey characteristics about the transfer function model.

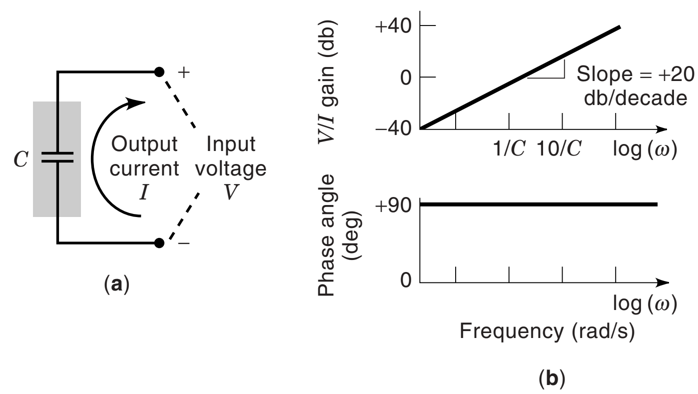


Figure 4. (a) Dynamic model of an electrical Capacitor with input voltage and output current. (b) Bode diagram frequency response of an electrical capacitor showing 20 db/decade slope of gain and 90 degree phase lead.

CONTINUOUS TIME TRANSFER FUNCTIONS

Three basic electrical components are the resistor, inductor, and capacitor. We have already seen the static transfer function model for the resistor in Eq. (1). We now consider continuous time dynamic transfer functions models for the inductor and capacitor using differential equations and the Laplace transform.

Continuous Time Transfer Function Model of an Inductor

Consider the integral equation of the system defined by the inductor shown in Fig. 3(a). The current through the inductor is a function of the initial current, $I(0)$, L, the inductance, and the time integral of the voltage across the inductor:

$$I(t) = \frac{1}{L}\int V(t)\,dt + I(0) \tag{5}$$

This equation is a dynamic model of the inductor. By using the Laplace transform, the operation of integration with respect to time can be replaced by the operation of division by the Laplace variable, s. Assuming that the initial current is zero, $I(0) = 0$, we can compute the output (current) over input

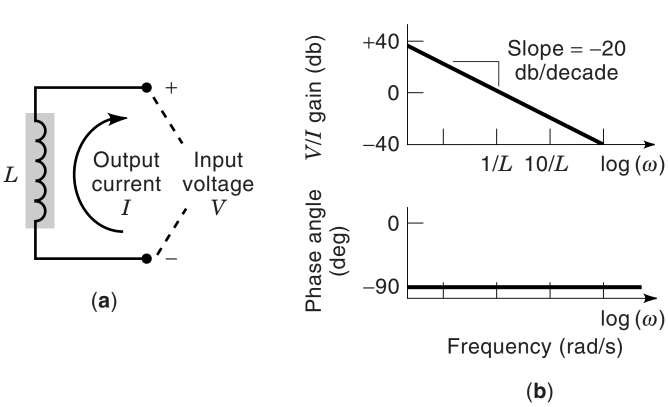


Figure 3. (a) Dynamic model of an electrical inductor with voltage input and current output. (b) Bode diagram frequency response of an electrical inductor showing −20 db/decade slope of gain and −90 degree phase lag.

(voltage) ratio resulting in the *sinusoidal* transfer function:

$$\frac{I(s)}{V(s)} = G_L(s) = \frac{1}{Ls} \tag{6}$$

Division by s, $1/s$, is indicative of an integration with respect to time. Equation (6) is a sinusoidal transfer function because the properties of the Laplace transform allow us to replace s with the complex term $s = j\omega$ to calculate the frequency response of the model. The term ω is the frequency in radians per second of a sinusoidal input to the system, and j is the complex variable used to represent $\sqrt{(-1)}$. Thus $G_L(s = j\omega)$ is a complex function composed of real and imaginary terms. Note that $I(s)$ is the Laplace transform of $I(t)$. Similarly, $V(s)$ is the Laplace transform of $V(t)$. $G_L(s = j\omega)$ is a complex function of frequency that we can represent with phasor notation. The phasor notation transforms the real and imaginary terms of a complex number to an amplitude (gain) and phase angle. At a specific frequency, ω_o, the complex value $G_L(j\omega_o)$ can be converted to an amplitude ratio (gain from input to output) and a phase angle between the input and the output. The frequency response of the dynamic transfer function model of the inductor is shown in Fig. 3(b) using a Bode diagram. The amplitude ratio has a −20 db/decade slope because of the pure integration in the model. Conversion of the amplitude to decibels is obtained by taking the log-based 10 of the gain and multiplying by 20 (20 $\log_{10}$(gain). A decade is an order of magnitude change in frequency from ω to 10ω. The negative phase angle shown in Fig. 3(b) implies that the current output lags the input voltage, so if the input voltage were $V(t) = \cos(\omega t)$, then the output current would be $I(t) = \cos(\omega t - 90)/(L\omega)$.

Continuous Time Transfer Function Model of a Capacitor

We can also use the Laplace transform to obtain a dynamic transfer function model of a capacitor, as shown in Fig. 4(a). The current passing through a capacitor is the time derivative of the input voltage. The dynamic model of the capacitor is shown as a linear ordinary differential equation:

$$I(t) = C\frac{dV(t)}{dt} \tag{7}$$

Using the Laplace transform, Eq. (7) can be converted into an algebraic equation and then the output over the input ratio can be computed, resulting in the following:

$$\frac{I(s)}{V(s)} = G_c(s) = Cs \tag{8}$$

where s implies a time derivative operator. The resulting frequency response of this dynamic model of the capacitor is plotted in Fig. 4(b) using a Bode diagram to plot the input-output phasor information as an amplitude ratio (gain) and a phase angle between the input and the output. The positive phase angle implies that the current output leads the input voltage, so if the input voltage were $V(t) = \cos(\omega t)$, then the output current would be $I(t) = C\omega \cos(\omega t + 90)$.

Continuous Time Transfer Function Model of a General System

The transfer functions of individual components can be used to model interconnected devices. This offers a convenient way to construct models out of tested subsystems. The resulting single-input, single-output model may have many terms in the numerator and denominator polynomials, as shown in the general transfer function in Eq. (9):

$$\frac{y(s)}{u(s)} = G(s) = \frac{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0} \tag{9}$$

where m is the highest order of the numerator polynomial and n is the highest order of the denominator polynomial. Causal models of physical systems require $n \geq m$, where causality means that only present inputs and past information are required to calculate the model output. Once there is a model of the system in the form $G(s)$, it can be used to estimate $y(t)$ given $u(t)$. For example, if a functional relationship is known for $u(t)$ (a step input, for example), it can be converted to $u(s)$ using the Laplace transform or a table of Laplace transformations. Then $y(s)$ can be calculated using the product $y(s) = G(s)u(s)$. Finally, $y(t)$ can be calculated from $y(s)$ using the inverse Laplace transform or a partial fraction expansion and the Laplace transformation tables. The section on Laplace transforms will present more details on this type of calculation.

TRANSFER FUNCTION MODELS OF FIRST- AND SECOND-ORDER LINEAR SYSTEMS

In Eq. (9) the mth order numerator polynomial has m roots and the nth order denominator polynomial has n roots. These numerator and denominator polynomials can be factored into the product of first-order and second-order polynomials as shown for a general case:

$$\frac{y(s)}{u(s)} = G(s) = \frac{K(s-z_1)(s-z_2)\ldots(s^2 + 2\zeta_{n1}w_{nn1}s + w_{nn1}^2)}{(s-p_1)(s-p_2)\ldots(s^2 + 2\zeta_{d1}w_{nd1}s + w_{nd1}^2)} \tag{10}$$

The first-order polynomial contains only a single real root. The second-order polynomial can contain two real roots or a complex pair of roots. The response of complex, high-order transfer functions can be obtained by adding up the contributions of all the first- and second-order polynomials in the frequency domain. To gain an understanding of how first- and second-order polynomials affect the response of a transfer function model, we will look at first- and second-order transfer functions.

First-Order Transfer Functions

Consider the electrical circuit of a low-pass filter comprising a resistor and capacitor, as shown in Fig. 5(a). A transfer function can be written from the current, I, to the output voltage V_o, and from the input voltage, V_i, to the current, I, as shown in Eq. (8) and in the following equations:

$$\frac{I(s)}{V_i(s)} = G_a(s) = \frac{1}{\left(R + \dfrac{1}{Cs}\right)} = \frac{Cs}{RCs+1} \tag{11}$$

$$\frac{V_o(s)}{I(s)} = G_b(s) = \frac{1}{Cs} \tag{12}$$

These equations assume that no current is required to obtain the measurement of V_o [$I_o = 0$ in Fig. 5(a)]. We can combine Eqs. (11) and (12) and eliminate $I(s)$. The result is only a function of $V_o(s)$ and $V_i(s)$, which can be rewritten as the transfer function from the input voltage V_i to the output voltage V_o. The dynamic transfer function model of the low-pass filter is then

$$\begin{aligned}\frac{V_o(s)}{V_i(s)} &= G_{1st}(s) = \frac{I(s)}{V_i(s)}\frac{V_o(s)}{I(s)} \\ &= G_a(s)G_b(s) = \frac{1}{\tau s + 1}, \text{ where } \tau = RC\end{aligned} \tag{13}$$

By combining $G_a(s)$ and $G_b(s)$, we have demonstrated the multiplicative property of transfer functions. The frequency

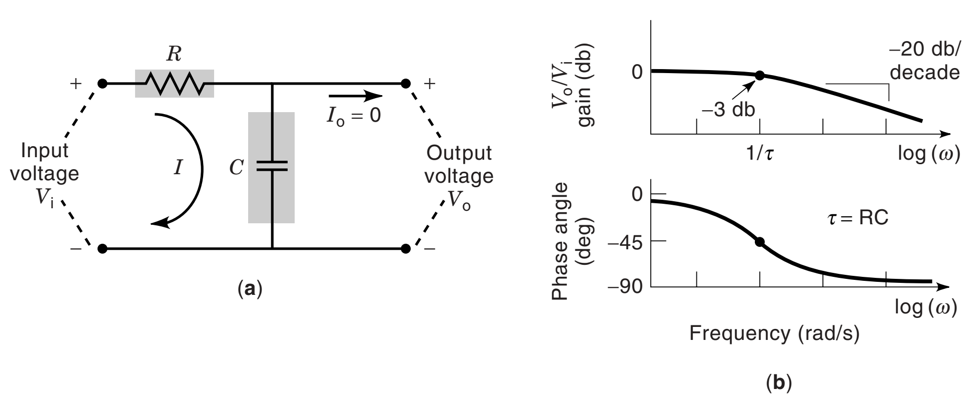


Figure 5. (a) Electrical first-order model of a passive low-pass filter comprised of a resistor and capacitor. (b) Bode diagram frequency response of a low-pass filter showing attenuation of gain at a frequencies greater than $1/\tau$. Gain is -3 dB and phase is -45 degrees at a frequency of $1/\tau$ rads/s.

response of this dynamic model is shown in Fig. 5(b) as a Bode diagram. The amplitude ratio is “flat” or approximately equal to 1 (0 dB) at low frequencies, and the amplitude decreases for frequencies greater than $1/\tau$. The gain equals -3 dB at a frequency $\omega = 1/\tau$. At higher frequencies the amplitude ratio decreases at the rate of -20 dB per decade (a decade is a range of frequency from ω to 10ω). This appears as a straight line on the plot of amplitude in decibels versus log frequency. Thus the system is called a low-pass filter because it allows low frequencies to pass but attenuates high frequencies. The phase angle starts at zero degrees, passes through -45 degrees at $\omega = 1/\tau$, and progresses to a phase angle of -90 degrees at high frequency. Note that the negative phase angle means that the output lags the input. The root of the polynomial in the denominator of this transfer function is $s = -1/\tau$. This value is the pole of the first-order transfer function, and it conveys information regarding the speed of response of the system. The time constant, τ is equal to the time it takes the system to respond to 63.2% of the final value when commanded with a step input.

Second-Order Transfer Functions

Figure 6(a) shows a second-order system comprised of a resistor, inductor, and a capacitor. Again we can write a transfer function from the input voltage to the current passing through all the components, shown in Eq. (14), and from the current to the output voltage, shown in Eq. (15).

$$\frac{I(s)}{V_{\mathrm{i}}(s)} = G_1(s) = \frac{1}{\left(R + Ls + \dfrac{1}{Cs}\right)} = \frac{Cs}{LCs^2 + RCs + 1} \tag{14}$$

$$\frac{V_{\mathrm{o}}(s)}{I(s)} = G_2(s) = \frac{1}{Cs} \tag{15}$$

By combining $G_1(s)$ and $G_2(s)$, eliminating the current, I, and rearranging the terms, we obtain the transfer function from the input voltage to the output voltage:

$$\frac{V_{\mathrm{o}}(s)}{V_{\mathrm{i}}(s)} = G_{2\mathrm{nd}}(s) = G_1(s)G_2(s) = \frac{1}{\dfrac{s^2}{w_{\mathrm{n}}^2} + \dfrac{2\zeta s}{w_{\mathrm{n}}} + 1} \tag{16}$$

where $\omega_{\mathrm{n}} = 1/\sqrt{LC}$ is the system natural frequency in radians per second and $\zeta = (R/2)\sqrt{C/L}$ is the system nondimensional damping coefficient. Note that Eq. (16) is a standard second-order transfer function. Depending on the value of ζ, the roots of the denominator polynomial can be real or both complex. Figure 6(b) shows the plot for $\zeta \approx 0.2$, which causes the amplitude of the frequency response to peak at a frequency near ω_{n}. There is an entire family of curves for the frequency response of a second-order system that vary with the value of ζ. When $\zeta > 1$, both of the roots of the denominator polynomial are real and the system is said to be overdamped. The amplitude response does not have a peak. When $\zeta = 1$, the system is said to be critically damped and the roots of the denominator polynomial are both real and repeated or identical. When $\zeta < 1$, the system is said to be underdamped. The two roots of the polynomial are a complex pair, $(a + jb)$ and $(a - jb)$, where a is the real part and b is the complex part of the root. As ζ approaches 0 from 1, the response becomes more oscillatory and the magnitude of the complex portion of the root increases. Figure 7 shows the second-order step response for four values of ζ when $\omega = 10$ rad/s. When $\zeta = 3$ the system is overdamped and responds slowly. When $\zeta = 1$ the system is critically damped and responds without an overshoot. When $\zeta = 0.4$ the system is underdamped and overshoots before settling in on the final value of 1.0 for a unit step response. When $\zeta = 0.1$ the system is oscillatory, and it takes several seconds for the oscillations to die out. When $\zeta = 0$ the system is said to be undamped and it will oscillate continuously because it is marginally stable (on the borderline between the mathematical definitions of stability and instability). We do not consider the case for $\zeta < 0$ because it implies a negative coefficient in the denominator polynomial, which indicates that the system is unstable from Routh stability criterion.

CASCADING TRANSFER FUNCTIONS AND THE LOADING ASSUMPTION

We have shown how transfer functions can be multiplied. But there is an assumption associated with this. When transfer functions are cascaded together, there can be energy trans-

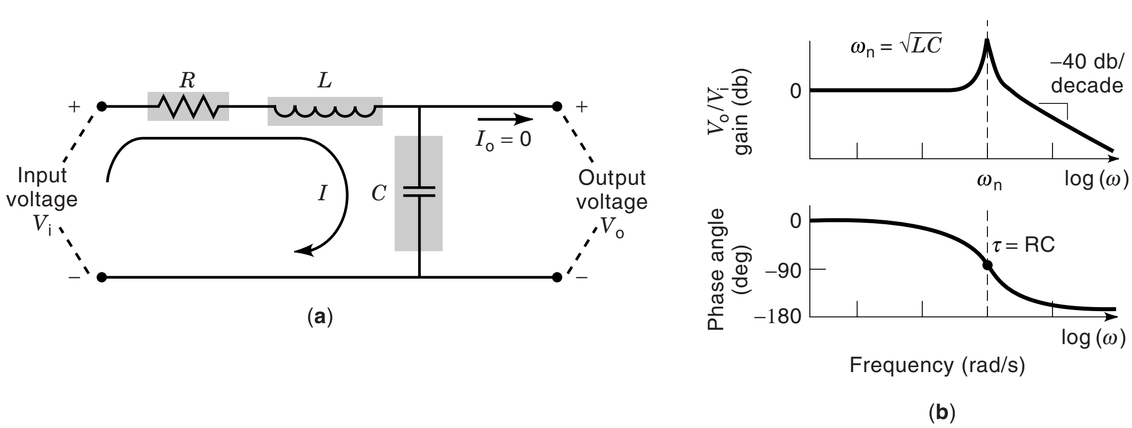


Figure 6. (a) Electrical second-order model comprised of inductor, resistor, and capacitor. (b) Bode diagram frequency response of a second-order system showing peak at natural frequency for system with low damping and high frequency attenuation of gain. The gain at ω_{n} depends on the damping coefficient, $\zeta = (R/2)\sqrt{C/L}$. The gain at ω_{n} is large when ζ is small (underdamped) and the gain is smaller when ζ is large (overdamped).

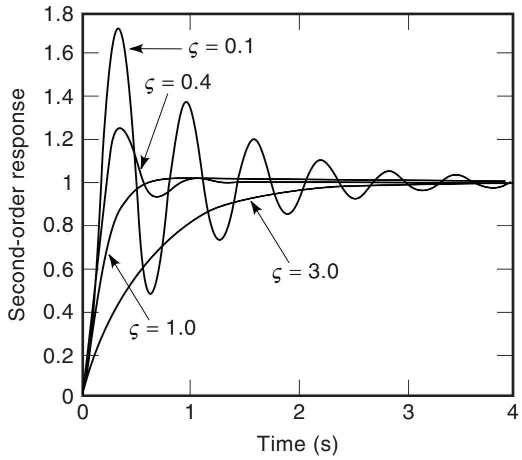


Figure 7. Time history response of second-order system with ω_n = 10 rads/s to a step input showing variation of the response with a range of damping coefficients, ζ. Note that the step response of the second-order system is sluggish for large values of ζ and oscillatory for small values of ζ.

ferred between the two systems represented by the transfer functions. By cascading transfer functions it is assumed that the energy extracted from one system does not significantly impact the response of that system. This energy flow is called loading. If there is significant loading from one system to the next, then multiplying the transfer functions violates an assumption and can lead to erroneous results, and the systems must be reanalyzed. Consider a system comprised of two cascaded low-pass first-order passive filters that were introduced in Fig. 5(a). If we multiply the two low-pass filter transfer functions from Eq. (13), we get

$$\frac{V_o(s)}{V_i(s)} = G_{\text{cascade}}(s) = G_1(s)G_2(s) = \frac{1}{\tau_1 s + 1} \cdot \frac{1}{\tau_2 s + 1} \tag{17}$$

where $\tau_1 = R_1C_1$ and $\tau_2 = R_2C_2$ are the time constants of the two cascaded filters. Equation (17) can be written as a second-order transfer function:

$$G_{\text{cascade}}(s) = \frac{1}{\tau_1\tau_2 s^2 + (\tau_1 + \tau_2)s + 1} \tag{18}$$

Note that by cascading two of the systems depicted in Fig. 5(a), we may have violated the assumptions that the output current draw is zero, ($I_{o1} = 0$). Reanalyzing the new system shown in Fig. 8 without this assumption results in the following transfer function for the two cascaded first-order filters:

$$G_{\text{cascade}}(s) = \frac{1}{\tau_1\tau_2 s^2 + (\tau_1 + R_1C_2 + \tau_2)s + 1} \tag{19}$$

The difference between Eq. (18) and Eq. (19) is the R_1C_2 term. So if R_1C_2 is small relative to τ_1 and τ_2, then cascading these two transfer functions is a good approximation. If these filters were active instead of passive filters, then the input impedance of the second filter would be high and the output impedance of the first filter would be low, so there would be little current drawn, and no loading effect.

BLOCK DIAGRAMS

Block diagrams and signal flow graphs are methods for visualizing systems constructed from subsystems, including transfer functions. We will concentrate on block diagrams because most of the computer-aided control system design software uses block diagrams for model construction. Block diagrams have a set of rules for manipulating the blocks in the diagram. These rules are identical to the rules of manipulating transfer functions. Figure 9(a) shows the product of two blocks representing the cascading of two transfer functions considered in the previous section in Eq. (17). The block diagram multiplication in Fig. 9(a) assumes a no-loading condition, and readers should be aware of this assumption when using the computer-aided simulation tools. Figure 9(b) shows the block diagram addition of two transfer functions.

Closed-Loop Block Diagrams

Figure 10 shows the concept of negative feedback in a closed-loop system using transfer functions $K(s)$ and $G(s)$. $K(s)$ represents a control system, and $G(s)$ represents a plant or controlled system. This figure represents the servomechanism control problem. There are a number of important transfer functions that will be examined in the next section using the closed-loop block diagram in Fig. 10.

PROPERTIES OF TRANSFER FUNCTIONS

Thus far in this article we have introduced the transfer function models for the resistor, inductor, and capacitor. It is important to note that these electrical elements have analogies in mechanical, thermal, and fluid systems. Transfer functions have applicability in a wide class of scientific and engineering fields. Equation (9) represents the Laplace transform of a constant coefficient, linear differential equation. Constant coefficient, linear differential equations adhere to the principle of superposition. Superposition states that the linear system input signal can be broken up into a sum of signals and the

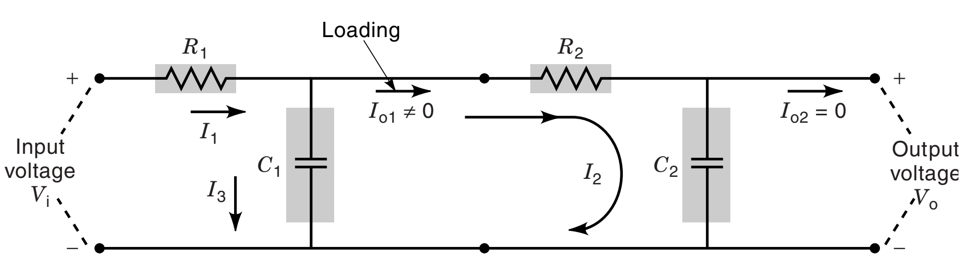


Figure 8. Cascaded first-order systems used to show the possible violation of a modeling assumption due to loading.

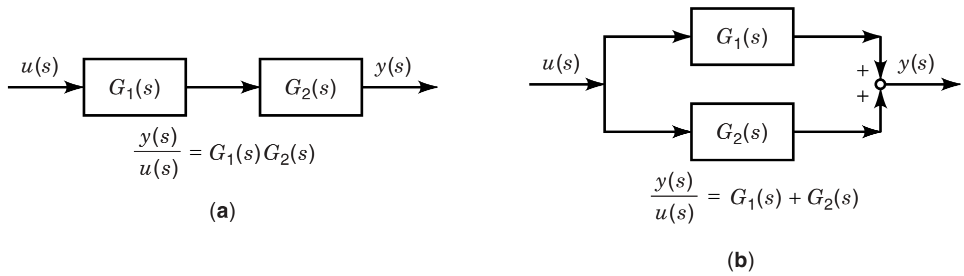


Figure 9. (a) Block diagram and transfer function multiplication. If $G_1(s)$ and $G_2(s)$ are transfer functions and not transfer function matrices, then the process of multiplication is commutative. (b) Block diagram and transfer function addition.

system output can be expressed as the sum of the system responses to each of the individual input signals. Consider the system $y(s) = G(s)u(s)$. We want to know what $y(t)$ is given by $u(t) = 6 + \cos(2\pi t) + \sin(3\pi t)$. We can define $u(t) = u_1(t) + u_2(t) + u_3(t) + u_4(t)$, where $u_1(t) = 3$, $u_2(t) = 3$, $u_3(t) = \cos(2\pi t)$, and $u_4(t) = \sin(3\pi t)$. Substituting for $u(t)$, we have

$$y(s) = G(s)u_1(s) + G(s)u_2(s) + G(s)u_3(s) + G(s)u_4(s) \tag{20}$$

where

$$y_1(s) = G(s)u_1(s), y_2(s) = G(s)u_2(s)$$
$$y_3(s) = G(s)u_3(s), y_4(s) = G(s)u_4(s) \tag{21}$$
$$y(s) = y_1(s) + y_2(s) + y_3(s) + y_4(s) \tag{22}$$

where

$$y_1(s) = G(s)u_1(s), y_2(s) = G(s)u_2(s)$$
$$y_3(s) = G(s)u_3(s), y_4(s) = G(s)u_4(s)$$

Using the inverse Laplace transform yields

$$y(t) = y_1(t) + y_2(t) + y_3(t) + y_4(t) \tag{23}$$

where $y_1(t)$, $y_2(t)$, $y_3(t)$, and $y_4(t)$, are the inverse Laplace transforms of $y_1(s)$, $y_2(s)$, $y_3(s)$, and $y_4(s)$, respectively. Thus the system output is equal to the sum of the system outputs corresponding to the individual portions of the input. Note that the numerical value of 6 was broken up into 3 and 3. This implies that scale factors pass through undisturbed, and Eq. (23) could be written as

$$y(t) = 2(y_1(t)) + y_2(t) + y_3(t) \tag{24}$$

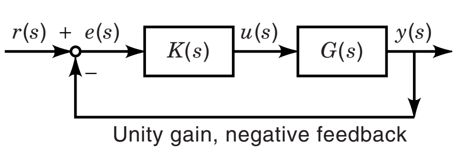


Figure 10. Block diagram of a unity gain, negative feedback system showing the control system transfer function, $K(s)$, and the controlled system transfer function, $G(s)$. The Laplace transform of the respective command signal, error, input, and output are $r(s)$, $e(s)$, $u(s)$, and $y(s)$. This block diagram is used to calculate the closed-loop, sensitivity, and complementary sensitivity transfer functions.

This is not true in general for nonlinear systems. The property of superposition holds for transfer functions because they are transformed from systems of linear, constant coefficient equations. Another property of transfer functions comes from the fact that the Laplace transform of a unit impulse is equal to one. Thus the unit impulse response of a system is just equal to the system transfer function since $y(s) = G(s)1 = G(s)$. Thus the inverse Laplace transform of the transfer function is the system time response to a unit impulse input.

Stability

We have not discussed the stability of the transfer functions because stability is covered elsewhere in this encyclopedia. We will just briefly mention that continuous time transfer functions in the Laplace domain are unstable if any denominator root, or transfer function pole, has a positive real portion. Thus, if the pole lies in the right-hand side of the y axis when plotted in the complex s plane, then the transfer function is unstable, as shown in Fig. 11(a). For discrete transfer functions parameterized with the Z transform variable, the transfer function is unstable if the complex pole lies outside the unit circle in the z plane, as shown in Fig. 11(b). (See Z TRANSFORMS for details.)

Sensitivity Transfer Function

Let us use transfer function algebra to solve for various transfer functions between variables. Consider the closed-loop block diagram shown in Fig. 10, where $G(s)$ is a transfer function of the system to be controlled (the plant) and $K(s)$ is a transfer function of the control system (the controller). The relationship from the commanded input, $r(s)$ to the controller error, $e(s)$, is

$$y(s) = G(s)u(s) = G(s)K(s)e(s) \tag{25}$$
$$e(s) = r(s) - y(s) \tag{26}$$

Substituting for $y(s)$ in Eq. (26) from Eq. (25) results in the following:

$$e(s) = r(s) - G(s)K(s)e(s) \tag{27}$$
$$e(s) = \frac{1}{1 + G(s)K(s)} r(s) \tag{28}$$
$$\frac{e(s)}{r(s)} = S(s) = \frac{1}{1 + G(s)K(s)} \tag{29}$$

$S(s)$ is called the sensitivity function, and it shows that as long as the product $G(s)K(s)$ is large relative to one, then the

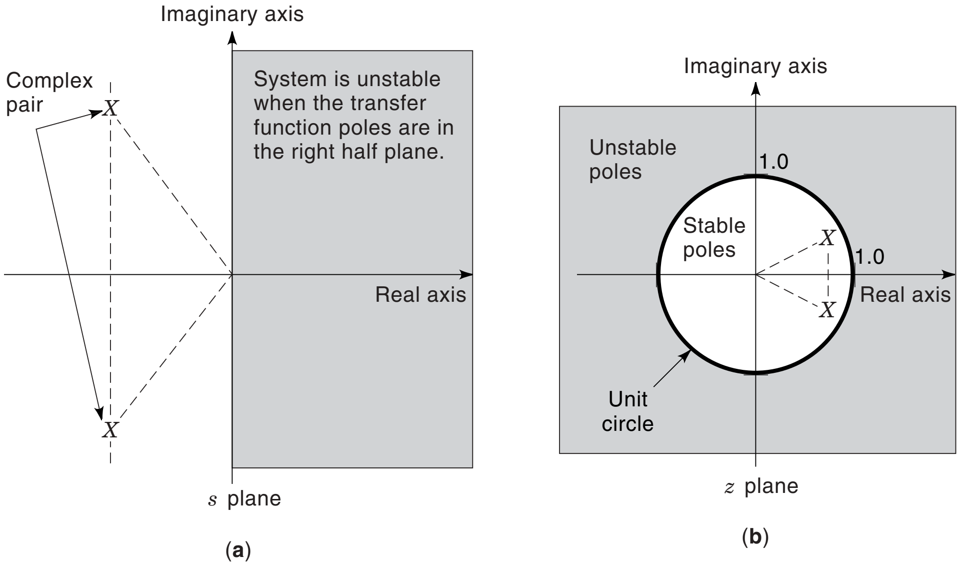


Figure 11. (a) An s domain plot showing the location of unstable poles for a continuous transfer function in the Laplace s variable. Two complex pair roots from a continuous second-order transfer function are shown. (b) A z domain plot showing the location of the unstable poles for a discrete transfer function in the z variable. Two complex pair roots from a discrete second-order transfer function are shown. (a) Stable poles of s domain transfer function are in the right half plane. (b) Stable poles of the z domain transfer function are within the unit circle.

error, $e(s)$, will be small. For large values of the product $G(s)K(s)$, Eq. (29) can be approximated as follows:

$$\text{when } G(s)K(s) \gg 1, \text{ then } S(s) \approx \frac{1}{G(s)K(s)} \tag{30}$$

The reason that Eq. (29) is called the sensitivity function will become apparent later in this discussion.

Complementary Sensitivity Transfer Function

In Fig. 10, consider the closed-loop transfer function relationship from the commanded input, $r(s)$, to the controlled output, $y(s)$. Using Eqs. (25) and (26) but substituting for $e(s)$ in Eq. (25) from Eq. (26) results in the following:

$$y(s) = G(s)K(s)r(s) - G(s)K(s)y(s) \tag{31}$$

$$y(s) = \frac{G(s)K(s)}{1 + G(s)K(s)}\, r(s) \tag{32}$$

$$\frac{y(s)}{r(s)} = T(s) = \frac{G(s)K(s)}{1 + G(s)K(s)} \tag{33}$$

$T(s)$ is called the complementary sensitivity transfer function. It is the transfer function from the commanded input, $r(s)$, to the controlled output, $y(s)$, and complements the sensitivity function because of the relationship between $T(s)$ and $S(s)$, shown as follows:

$$T(s) + S(s) = \frac{y(s)}{r(s)} + \frac{e(s)}{r(s)} = \frac{y(s) + (r(s) - y(s))}{r(s)} = 1 \tag{34}$$

So the sum of the sensitivity function and the complementary sensitivity function is equal to one.

The primary purpose of feedback is to reduce the sensitivity of the system to parameter variations and unwanted disturbances. Let us consider the block diagram in Fig. 10 without the feedback path. This would be an open-loop control system resulting in the following model:

$$y(s) = G(s)K(s)r(s) \tag{35}$$

where $G(s)$ represents a model of the system, which is just an approximation. If $G(s)$ is inaccurate, then the true system might be represented by

$$\begin{aligned} y(s) &= (G(s) + \Delta G(s))K(s)r(s) \\ &= G(s)K(s)r(s) + \Delta G(s)K(s)r(s) \end{aligned} \tag{36}$$

where $\Delta G(s)$ represents the modeling error. The resulting error in the output $y(s)$ is directly proportional to the modeling error. The closed-loop block diagram in Fig. 10 results in the transfer function in Eq. (33). If we introduce the plant uncertainty, $\Delta G(s)$, into Eq. (33) we have

$$\frac{y(s)}{r(s)} = \frac{(G(s) + \Delta G(s))K(s)}{1 + (G(s) + \Delta G(s))K(s)} = \frac{G(s)K(s) + \Delta G(s)K(s)}{1 + G(s)K(s) + \Delta G(s)K(s)} \tag{37}$$

Assuming that the model variation is $\Delta G(s)$ is small relative to $G(s)$, then Eq. (37) can be approximated as

$$\frac{y(s)}{r(s)} \approx \frac{G(s)K(s) + \Delta G(s)K(s)}{1 + G(s)K(s)} = T(s) + \frac{\Delta G(s)K(s)}{1 + G(s)K(s)} \tag{38}$$

So the change in the closed-loop, input/output transfer function, $T(s)$, due to the change in the open-loop transfer function, $G(s)$, is equal to $\Delta T(s)$, defined as follows:

$$\Delta T(s) = \frac{\Delta G(s)K(s)}{1 + G(s)K(s)} \tag{39}$$

Note that compared to the change in the open-loop equation in Eq. (36), the change in the closed-loop response is scaled by the denominator $(1 + G(s)K(s))$. Thus closing the loop with negative feedback reduces the affect of variations in the system plant. Also note that by the sensitivity function, $S(s)$ is essentially the reduction factor between Eqs. (36) and (39).

Control Sensitivity Transfer Function

Consider the relationship from the commanded input, $r(s)$, to the controller output or plant input, $u(s)$. Using Eqs. (25) and

(26) but substituting for $e(s)$ in Eq. (25) from Eq. (26), and using both pieces of Eq. (25), results in the following:

$$\begin{aligned} u(s) &= K(s)e(s) = K(s)(r(s) - y(s)) \\ &= K(s)r(s) - K(s)G(s)u(s) \end{aligned} \tag{40}$$

$$\frac{u(s)}{r(s)} = \frac{K(s)}{1 + K(s)G(s)} \tag{41}$$

Equation (41) is important in control system design because it gives the actuator response in a closed loop design. This allows the designer to take actuator rate and range limits into account by limiting the control sensitivity within the design procedure.

STATE-SPACE METHODS

The standard state-space representation is a set of four matrices, A, B, C, D, that make up a set of ordinary differential equations as follows:

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + Du \end{aligned} \tag{42}$$

where A is the system matrix, B is the input matrix, C is the output matrix, D is the feedforward matrix, x is a vector comprised of state variables, and $\dot{x}$ is the time derivative of the state vector. As before, u is the scalar input and y is the scalar output. State-space methods, state feedback, and state estimation are topics covered in other articles of this encyclopedia. We will discuss state-space representation briefly by saying that a transfer function can be converted to a state-space representation and vice versa. A state-space representation is not unique since the state of any representations can be transformed to another equivalent input-output representation and a new set of state variables. There are state-space representations and block diagrams that are standard representations of block diagrams. These standard forms represent specific state variable formulations. These forms are called the control, observer, and modal canonical forms, and they use only isolated integrators and gains as dynamic elements. The control and observer canonical forms are related to the concepts of observerability and controllability, which are discussed elsewhere in this encyclopedia. The following discussion holds for the transfer function of any single-input, single-output system. We will use a third-order system with a third-order numerator as an example, as follows:

$$\frac{y(s)}{u(s)} = G(s) = \frac{b_3 s^3 + b_2 s^2 + b_1 s + b_0}{s^3 + a_2 s^2 + a_1 s + a_0} \tag{43}$$

Control Canonical Form

The control canonical block diagram is shown in Fig. 12. The control canonical state-space representation is as follows:

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} u \\ y &= [b_0 - b_3 a_0 \quad b_1 - b_3 a_1 \quad b_2 - b_3 a_2] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + [b_3]u \end{aligned} \tag{44}$$

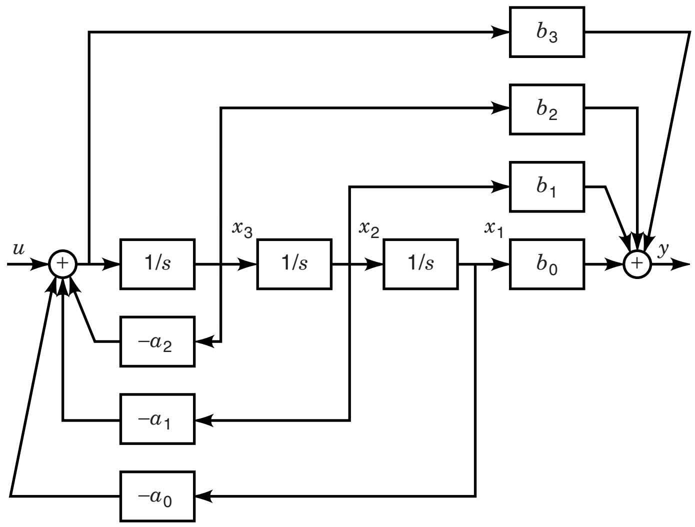


Figure 12. Block diagram showing the structure of the control canonical form.

Note that the block coefficients in Fig. 12 and the matrix scalar elements in Eq. (44) are the coefficients of the denominator in the transfer function in Eq. (43). Also, three new variables and the time derivatives of these variables were defined, x_1, x_2, x_3. These three variables make up the vector, $\boldsymbol{x}$, which is called the system state vector. This state variable is not unique, as we will see in a moment. The format in Eq. (44) and Fig. 12 is also called the phase variable form in some references.

Observer Canonical Form

The observer canonical block diagram is shown in Fig. 13. The observer canonical state space representation is as follows:

$$\begin{aligned} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} &= \begin{bmatrix} -a_2 & 1 & 0 \\ -a_1 & 0 & 1 \\ -a_0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + \begin{bmatrix} b_2 - a_2 b_3 \\ b_1 - a_1 b_3 \\ b_0 - a_0 b_3 \end{bmatrix} u \\ y &= [1 \quad 0 \quad 0] \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} + [b_3]u \end{aligned} \tag{45}$$

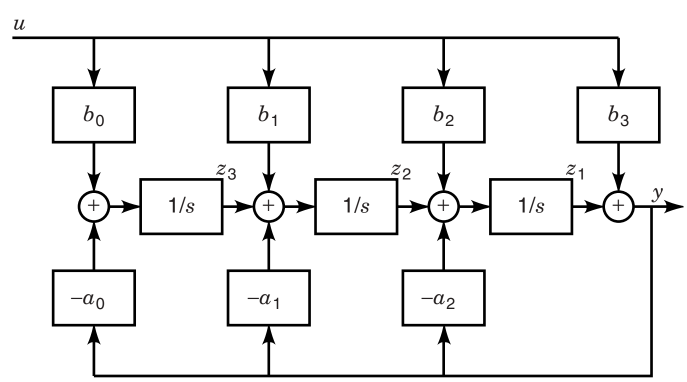


Figure 13. Block diagram showing the structure of the observer canonical form.

Note that the block coefficients in Fig. 13 and the matrix scalar elements in Eq. (45) are the coefficients of the denominator in the transfer function in Eq. (43). Also note that the state vector, $\boldsymbol{z}$, used in Eq. (45) is not the same state vector, $\boldsymbol{x}$, used in Eq. (44), even though the input, u, and output, y, variables are the same. The state vectors, $\boldsymbol{x}$ and $\boldsymbol{z}$, differ by a coordinate transformation. See the article in this encyclopedia on state space for more details. The form shown in Fig. 13 and in Eq. (45) is also called the rectangular form in some references because of the shape of the block diagram.

Modal Canonical Form

The block diagram for modal canonical form requires a discussion of residues and repeated roots and is outside of the scope of this article. We will just say that the modal canonical form results in a system matrix that is diagonal. The elements on the diagonal are made up of the roots of the denominator polynomial of the transfer function. For the modal canonical form, the elements on the diagonal of the A matrix could be complex, but there are methods for representing this A matrix with real values using a block diagonal A matrix.

Multivariable Systems

The advantage of the state-space system is that it can easily be extended to multivariable systems. If the system has nu inputs, nx state variables, and ny outputs, then $\boldsymbol{u}$ and $\boldsymbol{y}$ are nx by 1 and ny by 1 column vectors. x is the nx by 1 state vector. A, B, C, D are matrices of appropriate dimensions. A is a square nx by nx matrix. The dimensions of the B, C, and D matrices are nx by nu, ny by nx, ny by nu, respectively. Equation (46) shows the general format.

$$
\begin{aligned}
\begin{bmatrix} \dot{x}_1 \\ \vdots \\ \dot{x}_{nx} \end{bmatrix} &= \begin{bmatrix} a_{1,1} & \cdots & a_{1,nx} \\ \vdots & \cdots & \vdots \\ a_{nx,1} & \cdots & a_{nx,n} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{nx} \end{bmatrix} + \begin{bmatrix} b_{1,1} & \cdots & b_{1,nu} \\ \vdots & \cdots & \vdots \\ b_{nx,1} & \cdots & b_{nx,nu} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_{nu} \end{bmatrix} \\
\begin{bmatrix} y_1 \\ \vdots \\ y_{ny} \end{bmatrix} &= \begin{bmatrix} c_{1,1} & \cdots & c_{1,nx} \\ \vdots & \cdots & \vdots \\ c_{ny,1} & \cdots & c_{ny,nx} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{nx} \end{bmatrix} + \begin{bmatrix} d_{1,1} & \cdots & d_{1,nu} \\ \vdots & \cdots & \vdots \\ d_{ny,1} & \cdots & d_{ny,nu} \end{bmatrix} \begin{bmatrix} u_1 \\ \vdots \\ u_{nu} \end{bmatrix}
\end{aligned} \tag{46}
$$

There exists a transfer function for each input and output pair shown. The result is a transfer function matrix. We can no longer obtain the output over input ratio of $\boldsymbol{y}(s)/\boldsymbol{u}(s)$, since $\boldsymbol{u}(s)$ and $\boldsymbol{y}(s)$ are no longer scalars. They are column vectors. We can calculate the output over input ratio of each input-output pair. The transfer function matrix can be obtained as described for single-input, single-output systems. Take the Laplace transform that results in the replacement of $\dot{\boldsymbol{x}}$ with $s\boldsymbol{x}$, and solve for $\boldsymbol{y}(s)$ as a function of $\boldsymbol{u}(s)$. The following equation results:

$$\boldsymbol{y}(s) = [C[sI - A]^{-1}B + D]\boldsymbol{u}(s) \tag{47}$$

where I is an identity matrix of the appropriate dimension and $[sI - A]^{-1}$ is the matrix inverse of $[sI - A]$. For A matrices larger than 3 by 3, this is a difficult inverse to perform symbolically and is normally only performed numerically at discrete values of frequency ω, after substituting for $s = j\omega$. The result is a matrix of transfer functions:

$$
\boldsymbol{y}(s) = \begin{bmatrix} \dfrac{y_1(s)}{u_1(s)} & \cdots & \dfrac{y_1(s)}{u_i(s)} & \cdots & \dfrac{y_1(s)}{u_{nu}(s)} \\ \vdots & & & & \vdots \\ \dfrac{y_j(s)}{u_1(s)} & \cdots & \dfrac{y_j(s)}{u_i(s)} & \cdots & \dfrac{y_j(s)}{u_{nu}(s)} \\ \vdots & & & & \vdots \\ \dfrac{y_{ny}(s)}{u_1(s)} & \cdots & \dfrac{y_{ny}(s)}{u_i(s)} & \cdots & \dfrac{y_{ny}(s)}{u_{nu}(s)} \end{bmatrix} \boldsymbol{u}(s) \tag{48}
$$

where each element of the matrix is a transfer function.

SIMULATION OF LINEAR DYNAMIC SYSTEMS

A state-space system in Eq. (46) can be simulated with the addition of a numerical integration routine. The general nonlinear case is shown in Eq. (49). Equation (46) is obtained through the multivariable Taylor's series expansion of Eq. (49).

$$
\begin{aligned}
\dot{x} &= F(x, u, t) \\
y &= G(x, u, t)
\end{aligned} \tag{49}
$$

Using a numerical integration scheme, the value of x can be obtained at each timestep. The timestep has to be selected small enough such that the system dynamics are properly represented and the simulation plus integration is numerically stable. Typical numerical integration routines are the Euler and Runge–Kutta routines.

CONTROL SYSTEM DESIGN AND ANALYSIS

It is the pole and zero locations along with the gain that determine the transient response of any transfer function. Control system design basically results in the manipulation of the poles, zeros, and gain, although we are not much concerned with their location as we are with receiving the desired response and system robustness. Design procedures have been built around both the Nyquist and Bode plots, which represent the frequency response of a transfer function. Both phase and gain margin are used to check the robustness of single-input, single-output systems, but a discussion of phase and gain margin is outside the scope of this article. The topic on control system design should be examined for more details.

EXPERIMENTAL IDENTIFICATION OF DISCRETE TRANSFER FUNCTIONS

System models can be obtained using various methods, but the two primary methods are (1) to model the system physics using differential equations, and (2) to identify the system dynamic using dynamic data measured from the system. The first method addresses the physical relationships between all the components that make up the system. The second approach takes measured data from an existing system, assumes a model structure, and optimizes the model parameters to fit that measured data. The identified model is typically a discrete time model since the data are typically sampled, but a continuous time model could also be derived from the data. In this section we are interested in the identification of discrete models. These models are sometimes called autoregressive (AR) and autoregressive moving average (ARMA) models, but there are other names depending on the model structure and parameter optimization scheme used. You will see this type of model in the articles on finite impulse response (FIR) digital and adaptive filters and linear systems. Equation (50) is an example of a general linear, constant coefficient difference equation.

$$a_0 y(k-d) + a_1 y(k-d-1) + \cdots + a_n y(k-d-n)$$
$$= b_0 u(k) + b_1 u(k-1) + \cdots + b_m u(k-m)$$
$$\sum_{i=0}^{n} a_i y(k-d-i) = \sum_{j=0}^{m} b_j u(k-j) \tag{50}$$

Equation (2) is an example of a first-order difference equation with $n = 1$, $d = 0$, $a_0 = 1$, $a_1 = 0.-92$, and $m = 0$, and $b_0 = 16$. When identifying a discrete model of the type in Eq. (50), there are generally three steps. The first step is to analyze the system response to a step input to ascertain if there is any time delay between the input and when the effect of that input appears in the output. This time delay is represented by the value of d in Eq. (50). The second step is the selection of the structure for the model. The structure is defined by the values of n and m, which correspond to the order of the two polynomials in the coefficients a_i and b_j. The third step is parameter estimation or optimization. Parameter estimation solves for the parameters (coefficients in the polynomials) in order to reduce the error between the response of the real system and the discrete model. There are various optimization routines for solving for the parameters. We will briefly examine each of these three steps in the following sections.

Delay Estimation

The time delay, d, represented in Eq. (50) can be obtained from the system step response. Figure 14 repeats the automotive velocity model step response but includes a time delay of 2.5 samples. With experimental step response data you can measure d. However, d has to be an integer number, so you have two choices: either approximate d as either 2 or 3, or increase the sampling rate to 2 Hz (decrease the sampling period, $T = 0.5$ s). One of the difficulties with time delays is that when they occur, they are typically not constant. You are more likely to see a time delay in a fluid or thermal system than in a mechanical or electrical system.

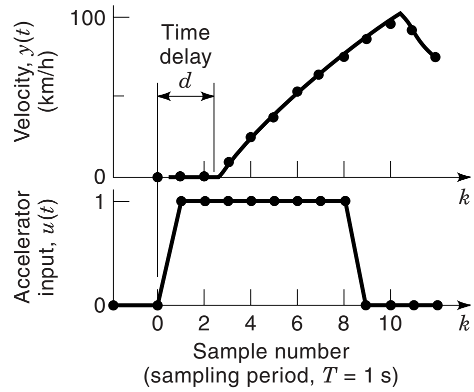


Figure 14. Automotive velocity response showing an example of how the discrete time delay would be estimated.

Model Structure Selection

Model structure section comes down to selecting the mathematical representation that will be used in the parameter estimation scheme. Most parameter estimation schemes are based on model structures that are linear in the parameters or coefficients. The model variables do not have to be linear, but in this section on transfer functions, we only consider linear models of the type shown in Eq. (50). Given this restriction, model structure selection comes down to selecting the order of the polynomials, the values for n and m in Eq. (50). Typically this is done based on experience and trial and error for simple systems. A model order is selected and then the parameter estimation scheme is executed. The model order can be increased until there is no improvement in the average error between the model and the data used to fit the system. There have been theories developed on model order selection, but these theories involve the concepts of probability and random processes, which are beyond the scope of this article. The Akaike's Baysian information criterion is one such example. In the following section on parameter estimation we will assume that the model order has already been selected.

Parameter Estimation

There are articles in this encyclopedia on self-tuning regulators, adaptive control, recursive filters, parameter estimation, least squares approximation, and recursive estimation, so we will just refer the reader to those articles. Basically parameter estimation comes down to an optimization scheme to solve for the coefficients in the model shown in Eq. (50) based on some criterion. There are various criteria that account for the various parameter optimization schemes. Recursive methods were developed for on-line parameter estimation. Recursive schemes require less computation and less memory compared to their nonrecursive counterparts, but they require computations at each timestep. The recursive scheme came about with the advent of adaptive control. Adaptive control attempts to address slow variations in a plant's dynamic by identifying a model on-line. *Slow* in this instance is relative to the system characteristics. For example, in the automotive acceleration model described earlier, a slow variation might be the difference caused by adding passengers or a load to the vehicle. This load variation does not occur at the same rate as the change in velocity.

Research in the Area of Transfer Functions. Transfer functions are just a modeling technique for single-input, single-output systems, and they are based on the linear, constant coefficient differential and difference equations and the Laplace and Z transforms. These mathematical concepts are well defined and mature, so there are no new developments in the area of transfer functions themselves. There is research in many areas of linear systems that use transfer functions concepts. We will just mention a few areas of research.

We have mentioned the use of least squares as an optimization method for estimating the parameters that make up a discrete time transfer function identified model of a system. Least squares is one optimization method. There have been many developments in different optimization approaches to parameter estimation that offer performance improvements for a particular application. Some of these approaches are stochastic in nature and they consider the measured variables as random variable and random processes (see PROBABILITY). The research areas of linear system and linear control system design have been active with robust and μ-synthesis control design techniques and the linear matrix inequalities (LMI) approach to solving optimization problems in linear systems. Much of the systems and control research has moved beyond the restriction of linear, constant coefficient systems. For example, one of the newer approaches to system identification uses genetic programming to solve for the system structure and a nonlinear ordinary differential system (see GENETIC ALGORITHMS). Also there have been developments in the area of system modeling based on chaos and wavelets (see WAVELETS).

BIBLIOGRAPHY

K. Astrom, Theory and applications of adaptive control—A survey, *Automatica,* **19**: 471–486, 1983.

O. H. Bosgra and H. Kwakernaak, Design Methods for Control Systems, course notes of the Dutch Institute of Systems and Control, Winter term, 1996–1997, Portable Document File (PDF) Internet Resource: http://www.math.utwente.nl/disc/dmcs/, April 26, 1998.

E. O. Doebelin, *System Dynamics Modeling and Response,* Columbus, OH: Charles E. Merrill, 1972.

G. Franklin, J. Powel, and A. Emami-Naeini, *Feedback Control of Dynamic Systems,* 3rd ed., Reading, MA: Addison-Wesley, 1994.

G. C. Goodwin and K. S. Sin, *Adaptive Filtering, Prediction and Control,* Englewood Cliffs, NJ: Prentice Hall, 1984.

R. Isermann, Parameter adaptive control algorithms—A tutorial, *Automatica,* 513–528, 1982.

R. Isermann, Practical aspects of process identification, *Automatica,* **16**: 575–587, 1982.

K. Ogata, *Modern Control Engineering,* Englewood Cliffs, NJ: Prentice Hall, 1970.

A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing,* Englewood Cliffs, NJ: Prentice Hall, 1975.

D. Ridgely and S. Banda, Introduction to Robust Multivariable Control, US Air Force Wright Aeronautical Lab, Wright-Patterson Air Force Base, Dayton, OH 45433, AFWAL-TR-85-3102, 1986.

R. Rosenberg and D. Karnopp, *Introduction to Physical System Dynamics,* New York: McGraw-Hill, 1983.

R. E. Skelton, *Dynamic Systems Control: Linear Systems Analysis and Synthesis,* New York: Wiley, 1988.

R. E. Skelton, *A Unified Algebraic Approach to Linear Control Design,* New York: Wiley, 1988.

B. Wittenmark and K. Astrom, Practical issues in the implementation of self-tuning control, *Automatica,* **20** (8): 595–605, 1984.

DUANE MATTERN
Mattern Engineering, Controls and Hardware

TRANSFERRED-ELECTRON DEVICES. See GUNN OR TRANSFERRED-ELECTRON DEVICES.

TRANSFORMABLE COMPUTING. See CONFIGURABLE COMPUTING.

TRANSFORMATIONS, GRAPHICS 2-D. See GRAPHICS TRANSFORMATIONS IN 2-D.

TRANSFORMER, DC. See DC TRANSFORMER.