Chapter Eight
Transfer Functions

> *The typical regulator system can frequently be described, in essentials, by differential equations of no more than perhaps the second, third or fourth order. …In contrast, the order of the set of differential equations describing the typical negative feedback amplifier used in telephony is likely to be very much greater. As a matter of idle curiosity, I once counted to find out what the order of the set of equations in an amplifier I had just designed would have been, if I had worked with the differential equations directly. It turned out to be 55.*
>
> Henrik Bode, 1960 [Bod60].

This chapter introduces the concept of the *transfer function*, which is a compact description of the input/output relation for a linear system. Combining transfer functions with block diagrams gives a powerful method for dealing with complex linear systems. The relationship between transfer functions and other descriptions of system dynamics is also discussed.

8.1 Frequency Domain Modeling

Figure 8.1 is a block diagram for a typical control system, consisting of a process to be controlled and a controller that combines feedback and feedforward. We saw in the previous two chapters how to analyze and design such systems using state space descriptions of the blocks. As mentioned in Chapter 2, an alternative approach is to focus on the input/output characteristics of the system. Since it is the inputs and outputs that are used to connect the systems, one could expect that this point of

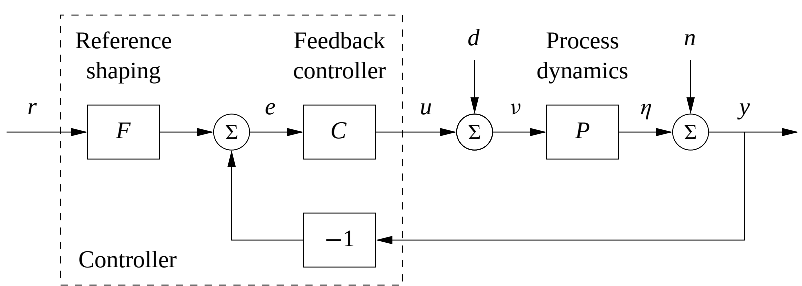


Figure 8.1: A block diagram for a feedback control system. The reference signal r is fed through a reference shaping block, which produces the signal that will be tracked. The error between this signal and the output is fed to a controller, which produces the input to the process. Disturbances and noise are included as external signals at the input and output of the process dynamics.

view would allow an understanding of the overall behavior of the system. Transfer functions are the main tool in implementing this point of view for linear systems.

The basic idea of the transfer function comes from looking at the frequency response of a system. Suppose that we have an input signal that is periodic. Then we can decompose this signal into the sum of a set of sines and cosines,

$$u(t) = \sum_{k=0}^{\infty} a_k \sin(k\omega t) + b_k \cos(k\omega t),$$

where ω is the fundamental frequency of the periodic input. Each of the terms in this input generates a corresponding sinusoidal output (in steady state), with possibly shifted magnitude and phase. The gain and phase at each frequency are determined by the frequency response given in equation (5.24):

$$G(s) = C(sI - A)^{-1}B + D, \tag{8.1}$$

where we set $s = i(k\omega)$ for each $k = 1, \dots, \infty$ and $i = \sqrt{-1}$. If we know the steady-state frequency response $G(s)$, we can thus compute the response to any (periodic) signal using superposition.

The transfer function generalizes this notion to allow a broader class of input signals besides periodic ones. As we shall see in the next section, the transfer function represents the response of the system to an *exponential input*, $u = e^{st}$. It turns out that the form of the transfer function is precisely the same as that of equation (8.1). This should not be surprising since we derived equation (8.1) by writing sinusoids as sums of complex exponentials. Formally, the transfer function is the ratio of the Laplace transforms of output and input, although one does not have to understand the details of Laplace transforms in order to make use of transfer functions.

Modeling a system through its response to sinusoidal and exponential signals is known as *frequency domain modeling*. This terminology stems from the fact that we represent the dynamics of the system in terms of the generalized frequency s rather than the time domain variable t. The transfer function provides a complete representation of a linear system in the frequency domain.

The power of transfer functions is that they provide a particularly convenient representation in manipulating and analyzing complex linear feedback systems. As we shall see, there are many graphical representations of transfer functions that capture interesting properties of the underlying dynamics. Transfer functions also make it possible to express the changes in a system because of modeling error, which is essential when considering sensitivity to process variations of the sort discussed in Chapter 12. More specifically, using transfer functions, it is possible to analyze what happens when dynamic models are approximated by static models or when high-order models are approximated by low-order models. One consequence is that we can introduce concepts that express the degree of stability of a system.

While many of the concepts for state space modeling and analysis apply directly to nonlinear systems, frequency domain analysis applies primarily to linear systems. The notions of gain and phase can be generalized to nonlinear systems and, in particular, propagation of sinusoidal signals through a nonlinear system can approximately be captured by an analog of the frequency response called the describing function. These extensions of frequency response will be discussed in Section 9.5.

8.2 Derivation of the Transfer Function

As we have seen in previous chapters, the input/output dynamics of a linear system have two components: the initial condition response and the forced response. In addition, we can speak of the transient properties of the system and its steady-state response to an input. The transfer function focuses on the steady-state forced response to a given input and provides a mapping between inputs and their corresponding outputs. In this section, we will derive the transfer function in terms of the exponential response of a linear system.

Transmission of Exponential Signals

To formally compute the transfer function of a system, we will make use of a special type of signal, called an *exponential signal,* of the form e^{st}, where $s = \sigma + i\omega$ is a complex number. Exponential signals play an important role in linear systems. They appear in the solution of differential equations and in the impulse response of linear systems, and many signals can be represented as exponentials or sums of exponentials. For example, a constant signal is simply $e^{\alpha t}$ with $\alpha = 0$. Damped sine and cosine signals can be represented by

$$e^{(\sigma+i\omega)t} = e^{\sigma t}e^{i\omega t} = e^{\sigma t}(\cos\omega t + i\sin\omega t),$$

where $\sigma < 0$ determines the decay rate. Figure 8.2 gives examples of signals that can be represented by complex exponentials; many other signals can be represented by linear combinations of these signals. As in the case of sinusoidal signals, we will allow complex-valued signals in the derivation that follows, although in practice we always add together combinations of signals that result in real-valued functions.

To investigate how a linear system responds to an exponential input $u(t) = e^{st}$ we consider the state space system

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx + Du. \tag{8.2}$$

Let the input signal be $u(t) = e^{st}$ and assume that $s \neq \lambda_j(A)$, $j = 1, \dots, n$, where $\lambda_j(A)$ is the jth eigenvalue of A. The state is then given by

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}Be^{s\tau}\,d\tau = e^{At}x(0) + e^{At}\int_0^t e^{(sI-A)\tau}B\,d\tau.$$

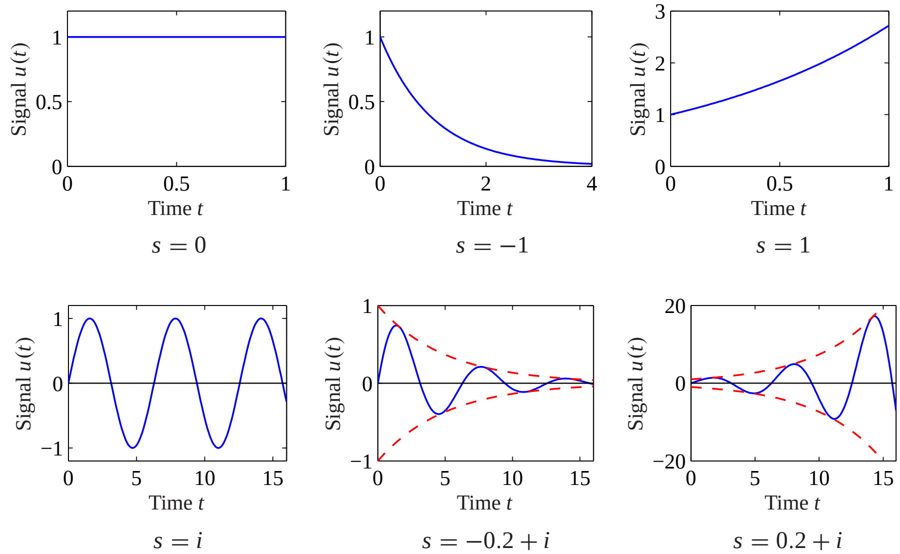


Figure 8.2: Examples of exponential signals. The top row corresponds to exponential signals with a real exponent, and the bottom row corresponds to those with complex exponents. The dashed line in the last two cases denotes the bounding envelope for the oscillatory signals. In each case, if the real part of the exponent is negative then the signal decays, while if the real part is positive then it grows.

As we saw in Section 5.3, if $s \neq \lambda(A)$, the integral can be evaluated and we get

$$\begin{aligned} x(t) &= e^{At}x(0) + e^{At}(sI - A)^{-1}\Big(e^{(sI-A)t} - I\Big)B \\ &= e^{At}\Big(x(0) - (sI - A)^{-1}B\Big) + (sI - A)^{-1}Be^{st}. \end{aligned}$$

The output of equation (8.2) is thus

$$\begin{aligned} y(t) &= Cx(t) + Du(t) \\ &= Ce^{At}\Big(x(0) - (sI - A)^{-1}B\Big) + \Big(C(sI - A)^{-1}B + D\Big)e^{st}, \end{aligned} \tag{8.3}$$

a linear combination of the exponential functions e^{st} and e^{At}. The first term in equation (8.3) is the transient response of the system. Recall that e^{At} can be written in terms of the eigenvalues of A (using the Jordan form in the case of repeated eigenvalues), and hence the transient response is a linear combination of terms of the form $e^{\lambda_j t}$, where λ_j are eigenvalues of A. If the system is stable, then $e^{At} \to 0$ as $t \to \infty$ and this term dies away.

The second term of the output (8.3) is proportional to the input $u(t) = e^{st}$. This term is called the *pure exponential response*. If the initial state is chosen as

$$x(0) = (sI - A)^{-1}B,$$

then the output consists of only the pure exponential response and both the state and the output are proportional to the input:

$$x(t) = (sI - A)^{-1}Be^{st} = (sI - A)^{-1}Bu(t),$$
$$y(t) = \big(C(sI - A)^{-1}B + D\big)e^{st} = \big(C(sI - A)^{-1}B + D\big)u(t).$$

This is also the output we see in steady state, when the transients represented by the first term in equation (8.3) have died out. The map from the input to the output,

$$G_{yu}(s) = C(sI - A)^{-1}B + D, \tag{8.4}$$

is the *transfer function* from u to y for the system (8.2), and we can write $y(t) = G_{yu}(s)u(t)$ for the case that $u(t) = e^{st}$. Compare with the definition of frequency response given by equation (5.24).

An important point in the derivation of the transfer function is the fact that we have restricted s so that $s \neq \lambda_j(A)$, the eigenvalues of A. At those values of s, we see that the response of the system is singular (since $sI - A$ will fail to be invertible). If $s = \lambda_j(A)$, the response of the system to the exponential input $u = e^{\lambda_j t}$ is $y = p(t)e^{\lambda_j t}$, where $p(t)$ is a polynomial of degree less than or equal to the multiplicity of the eigenvalue λ_j (see Exercise 8.2).

Example 8.1 Damped oscillator

Consider the response of a damped linear oscillator, whose state space dynamics were studied in Section 6.3:

$$\frac{dx}{dt} = \begin{bmatrix} 0 & \omega_0 \\ -\omega_0 & -2\zeta\omega_0 \end{bmatrix} x + \begin{bmatrix} 0 \\ k\omega_0 \end{bmatrix} u, \qquad y = \begin{bmatrix} 1 & 0 \end{bmatrix} x. \tag{8.5}$$

This system is stable if $\zeta > 0$, and so we can look at the steady-state response to an input $u = e^{st}$,

$$\begin{aligned} G_{yu}(s) = C(sI - A)^{-1}B &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s & -\omega_0 \\ \omega_0 & s + 2\zeta\omega_0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ k\omega_0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\frac{1}{s^2 + 2\zeta\omega_0 s + \omega_0^2} \begin{bmatrix} s & \omega_0 \\ -\omega_0 & s + 2\zeta\omega_0 \end{bmatrix} \right) \begin{bmatrix} 0 \\ k\omega_0 \end{bmatrix} \\ &= \frac{k\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}. \end{aligned} \tag{8.6}$$

To compute the steady-state response to a step function, we set $s = 0$ and we see that

$$u = 1 \qquad \Longrightarrow \qquad y = G_{yu}(0)u = k.$$

If we wish to compute the steady-state response to a sinusoid, we write

$$u = \sin \omega t = \frac{1}{2}\left(ie^{-i\omega t} - ie^{i\omega t}\right),$$
$$y = \frac{1}{2}\left(iG_{yu}(-i\omega)e^{-i\omega t} - iG_{yu}(i\omega)e^{i\omega t}\right).$$

We can now write $G(i\omega)$ in terms of its magnitude and phase,

$$G(i\omega) = \frac{k\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2} = Me^{i\theta},$$

where the magnitude (or gain) M and phase θ are given by

$$M = \frac{k\omega_0^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + (2\zeta\omega_0\omega)^2}}, \qquad \frac{\sin\theta}{\cos\theta} = \frac{-2\zeta\omega_0\omega}{\omega_0^2 - \omega^2}.$$

We can also make use of the fact that $G(-i\omega)$ is given by its complex conjugate $G^*(i\omega)$, and it follows that $G(-i\omega) = Me^{-i\theta}$. Substituting these expressions into our output equation, we obtain

$$\begin{aligned} y &= \frac{1}{2}\left(i(Me^{-i\theta})e^{-i\omega t} - i(Me^{i\theta})e^{i\omega t}\right) \\ &= M \cdot \frac{1}{2}\left(ie^{-i(\omega t+\theta)} - ie^{i(\omega t+\theta)}\right) = M\sin(\omega t + \theta). \end{aligned}$$

The responses to other signals can be computed by writing the input as an appropriate combination of exponential responses and using linearity. ∇

Coordinate Changes

The matrices A, B and C in equation (8.2) depend on the choice of coordinate system for the states. Since the transfer function relates input to outputs, it should be invariant to coordinate changes in the state space. To show this, consider the model (8.2) and introduce new coordinates z by the transformation $z = Tx$, where T is a nonsingular matrix. The system is then described by

$$\begin{aligned} \frac{dz}{dt} &= T(Ax + Bu) = TAT^{-1}z + TBu =: \tilde{A}z + \tilde{B}u, \\ y &= Cx + DU = CT^{-1}z + Du =: \tilde{C}z + Du. \end{aligned}$$

This system has the same form as equation (8.2), but the matrices A, B and C are different:

$$\tilde{A} = TAT^{-1}, \qquad \tilde{B} = TB, \qquad \tilde{C} = CT^{-1}. \tag{8.7}$$

Computing the transfer function of the transformed model, we get

$$\begin{aligned} \tilde{G}(s) &= \tilde{C}(sI - \tilde{A})^{-1}\tilde{B} + \tilde{D} = CT^{-1}(sI - TAT^{-1})^{-1}TB + D \\ &= C\left(T^{-1}(sI - TAT^{-1})T\right)^{-1}B + D = C(sI - A)^{-1}B + D = G(s), \end{aligned}$$

which is identical to the transfer function (8.4) computed from the system description (8.2). The transfer function is thus invariant to changes of the coordinates in the state space.

Another property of the transfer function is that it corresponds to the portion of the state space dynamics that is both reachable and observable. In particular, if we make use of the Kalman decomposition (Section 7.5), then the transfer function depends only on the dynamics in the reachable and observable subspace Σ_{ro} (Exercise 8.7).

Transfer Functions for Linear Systems

Consider a linear input/output system described by the controlled differential equation

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_0 \frac{d^m u}{dt^m} + b_1 \frac{d^{m-1} u}{dt^{m-1}} + \cdots + b_m u, \tag{8.8}$$

where u is the input and y is the output. This type of description arises in many applications, as described briefly in Section 2.2; bicycle dynamics and AFM modeling are two specific examples. Note that here we have generalized our previous system description to allow both the input and its derivatives to appear.

To determine the transfer function of the system (8.8), let the input be $u(t) = e^{st}$. Since the system is linear, there is an output of the system that is also an exponential function $y(t) = y_0 e^{st}$. Inserting the signals into equation (8.8), we find

$$(s^n + a_1 s^{n-1} + \cdots + a_n) y_0 e^{st} = (b_0 s^m + b_1 s^{m-1} \cdots + b_m) e^{-st},$$

and the response of the system can be completely described by two polynomials

$$a(s) = s^n + a_1 s^{n-1} + \cdots + a_n, \qquad b(s) = b_0 s^m + b_1 s^{m-1} + \cdots + b_m. \tag{8.9}$$

The polynomial $a(s)$ is the characteristic polynomial of the ordinary differential equation. If $a(s) \neq 0$, it follows that

$$y(t) = y_0 e^{st} = \frac{b(s)}{a(s)} e^{st}. \tag{8.10}$$

The transfer function of the system (8.8) is thus the rational function

$$G(s) = \frac{b(s)}{a(s)}, \tag{8.11}$$

where the polynomials $a(s)$ and $b(s)$ are given by equation (8.9). Notice that the transfer function for the system (8.8) can be obtained by inspection since the coefficients of $a(s)$ and $b(s)$ are precisely the coefficients of the derivatives of u and y. The *order* of the transfer function is defined as the order of the denominator polynomial.

Equations (8.8)–(8.11) can be used to compute the transfer functions of many simple ordinary differential equations. Table 8.1 gives some of the more common forms. The first five of these follow directly from the analysis above. For the proportional-integral-derivative (PID) controller, we make use of the fact that the integral of an exponential input is given by $(1/s)e^{st}$.

The last entry in Table 8.1 is for a pure time delay, in which the output is identical to the input at an earlier time. Time delays appear in many systems: typical examples are delays in nerve propagation, communication and mass transport. A system with

Table 8.1: Transfer functions for some common ordinary differential equations.

Type	ODE	Transfer Function
Integrator	$\dot{y} = u$	$\frac{1}{s}$
Differentiator	$y = \dot{u}$	s
First-order system	$\dot{y} + ay = u$	$\frac{1}{s+a}$
Double integrator	$\ddot{y} = u$	$\frac{1}{s^2}$
Damped oscillator	$\ddot{y} + 2\zeta\omega_0\dot{y} + \omega_0^2 y = u$	$\frac{1}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$
PID controller	$y = k_p u + k_d \dot{u} + k_i \int u$	$k_p + k_d s + \frac{k_i}{s}$
Time delay	$y(t) = u(t-\tau)$	$e^{-\tau s}$

a time delay has the input/output relation

$$y(t) = u(t-\tau). \tag{8.12}$$

As before, let the input be $u(t) = e^{st}$. Assuming that there is an output of the form $y(t) = y_0 e^{st}$ and inserting into equation (8.12), we get

$$y(t) = y_0 e^{st} = e^{s(t-\tau)} = e^{-s\tau} e^{st} = e^{-s\tau} u(t).$$

The transfer function of a time delay is thus $G(s) = e^{-s\tau}$, which is not a rational function but is analytic except at infinity. (A complex function is *analytic* in a region if it has no singularities in the region.)

Example 8.2 Electrical circuit elements

Modeling of electrical circuits is a common use of transfer functions. Consider, for example, a resistor modeled by Ohm's law $V = IR$, where V is the voltage across the resister, I is the current through the resistor and R is the resistance value. If we consider current to be the input and voltage to be the output, the resistor has the transfer function $Z(s) = R$. $Z(s)$ is also called the *impedance* of the circuit element.

Next we consider an inductor whose input/output characteristic is given by

$$L\frac{dI}{dt} = V.$$

Letting the current be $I(t) = e^{st}$, we find that the voltage is $V(t) = Lse^{st}$ and the transfer function of an inductor is thus $Z(s) = Ls$. A capacitor is characterized by

$$C\frac{dV}{dt} = I,$$

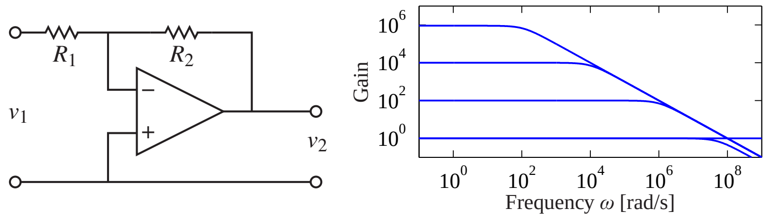


Figure 8.3: Stable amplifier based on negative feedback around an operational amplifier. The block diagram on the left shows a typical amplifier with low-frequency gain R_2/R_1. If we model the dynamic response of the op amp as $G(s) = ak/(s+a)$, then the gain falls off at frequency $\omega = a$, as shown in the gain curves on the right. The frequency response is computed for $k = 10^7$, $a = 100$ rad/s, $R_2 = 10^6$ Ω, and $R_1 = 1$, 10^2, 10^4 and 10^6 Ω.

and a similar analysis gives a transfer function from current to voltage of $Z(s) = 1/(Cs)$. Using transfer functions, complex electrical circuits can be analyzed algebraically by using the complex impedance $Z(s)$ just as one would use the resistance value in a resistor network. ∇

Example 8.3 Operational amplifier circuit

To further illustrate the use of exponential signals, we consider the operational amplifier circuit introduced in Section 3.3 and reproduced in Figure 8.3a. The model introduced in Section 3.3 is a simplification because the linear behavior of the amplifier was modeled as a constant gain. In reality there are significant dynamics in the amplifier, and the static model $v_{\text{out}} = -kv$ (equation (3.10)) should therefore be replaced by a dynamic model. In the linear range of the amplifier, we can model the operational amplifier as having a steady-state frequency response

$$\frac{v_{\text{out}}}{v} = -\frac{ak}{s+a} =: -G(s). \tag{8.13}$$

This response corresponds to a first-order system with time constant $1/a$. The parameter k is called the *open loop gain*, and the product ak is called the *gain-bandwidth product*; typical values for these parameters are $k = 10^7$ and $ak = 10^7$–10^9 rad/s.

Since all of the elements of the circuit are modeled as being linear, if we drive the input v_1 with an exponential signal e^{st}, then in steady state all signals will be exponentials of the same form. This allows us to manipulate the equations describing the system in an algebraic fashion. Hence we can write

$$\frac{v_1 - v}{R_1} = \frac{v - v_2}{R_2} \quad \text{and} \quad v_2 = -G(s)v, \tag{8.14}$$

using the fact that the current into the amplifier is very small, as we did in Section 3.3. Eliminating v between these equations gives the following transfer function of the system

$$\frac{v_2}{v_1} = \frac{-R_2 G(s)}{R_1 + R_2 + R_1 G(s)} = \frac{-R_2 ak}{R_1 ak + (R_1 + R_2)(s+a)}.$$

The low-frequency gain is obtained by setting $s = 0$, hence

$$G_{v_2 v_1}(0) = \frac{-kR_2}{(k+1)R_1 + R_2} \approx -\frac{R_2}{R_1},$$

which is the result given by (3.11) in Section 3.3. The bandwidth of the amplifier circuit is

$$\omega_b = a\frac{R_1(k+1) + R_2}{R_1 + R_2} \approx a\frac{R_1 k}{R_2},$$

where the approximation holds for $R_2/R_1 \gg 1$. The gain of the closed loop system drops off at high frequencies as $R_2 k/(\omega(R_1 + R_2))$. The frequency response of the transfer function is shown in Figure 8.3b for $k = 10^7$, $a = 100$ rad/s, $R_2 = 10^6$ Ω and $R_1 = 1$, 10^2, 10^4 and 10^6 Ω.

Note that in solving this example, we bypassed explicitly writing the signals as $v = v_0 e^{st}$ and instead worked directly with v, assuming it was an exponential. This shortcut is handy in solving problems of this sort and when manipulating block diagrams. A comparison with Section 3.3, where we made the same calculation when $G(s)$ was a constant, shows analysis of systems using transfer functions is as easy as using static systems. The calculations are the same if the resistances R_1 and R_2 are replaced by impedances, as discussed in Example 8.2. ∇

Although we have focused thus far on ordinary differential equations, transfer functions can also be used for other types of linear systems. We illustrate this via an example of a transfer function for a partial differential equation.

Example 8.4 Heat propagation
Consider the problem of one-dimensional heat propagation in a semi-infinite metal rod. Assume that the input is the temperature at one end and that the output is the temperature at a point along the rod. Let $\theta(x, t)$ be the temperature at position x and time t. With a proper choice of length scales and units, heat propagation is described by the partial differential equation

$$\frac{\partial \theta}{\partial t} = \frac{\partial^2 \theta}{\partial^2 x}, \tag{8.15}$$

and the point of interest can be assumed to have $x = 1$. The boundary condition for the partial differential equation is

$$\theta(0, t) = u(t).$$

To determine the transfer function we choose the input as $u(t) = e^{st}$. Assume that there is a solution to the partial differential equation of the form $\theta(x, t) = \psi(x)e^{st}$ and insert this into equation (8.15) to obtain

$$s\psi(x) = \frac{d^2\psi}{dx^2},$$

with boundary condition $\psi(0) = e^{st}$. This ordinary differential equation (with independent variable x) has the solution

$$\psi(x) = Ae^{x\sqrt{s}} + Be^{-x\sqrt{s}}.$$

Matching the boundary conditions gives $A = 0$ and $B = e^{st}$, so the solution is

$$y(t) = \theta(1, t) = \psi(1)e^{st} = e^{-\sqrt{s}}e^{st} = e^{-\sqrt{s}}u(t).$$

The system thus has the transfer function $G(s) = e^{-\sqrt{s}}$. As in the case of a time delay, the transfer function is not a rational function but is an analytic function. ∇

Gains, Poles and Zeros

The transfer function has many useful interpretations and the features of a transfer function are often associated with important system properties. Three of the most important features are the gain and the locations of the poles and zeros.

The *zero frequency gain* of a system is given by the magnitude of the transfer function at $s = 0$. It represents the ratio of the steady-state value of the output with respect to a step input (which can be represented as $u = e^{st}$ with $s = 0$). For a state space system, we computed the zero frequency gain in equation (5.22):

$$G(0) = D - CA^{-1}B.$$

For a system written as a linear differential equation

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1}y}{dt^{n-1}} + \cdots + a_n y = b_0 \frac{d^m u}{dt^m} + b_1 \frac{d^{m-1}u}{dt^{m-1}} + \cdots + b_m u,$$

if we assume that the input and output of the system are constants y_0 and u_0, then we find that $a_n y_0 = b_m u_0$. Hence the zero frequency gain is

$$G(0) = \frac{y_0}{u_0} = \frac{b_m}{a_n}. \tag{8.16}$$

Next consider a linear system with the rational transfer function

$$G(s) = \frac{b(s)}{a(s)}.$$

The roots of the polynomial $a(s)$ are called the *poles* of the system, and the roots of $b(s)$ are called the *zeros* of the system. If p is a pole, it follows that $y(t) = e^{pt}$ is a solution of equation (8.8) with $u = 0$ (the homogeneous solution). A pole p corresponds to a *mode* of the system with corresponding modal solution e^{pt}. The unforced motion of the system after an arbitrary excitation is a weighted sum of modes.

Zeros have a different interpretation. Since the pure exponential output corresponding to the input $u(t) = e^{st}$ with $a(s) \neq 0$ is $G(s)e^{st}$, it follows that the pure exponential output is zero if $b(s) = 0$. Zeros of the transfer function thus block transmission of the corresponding exponential signals.

For a state space system with transfer function $G(s) = C(sI - A)^{-1}B + D$, the poles of the transfer function are the eigenvalues of the matrix A in the state space

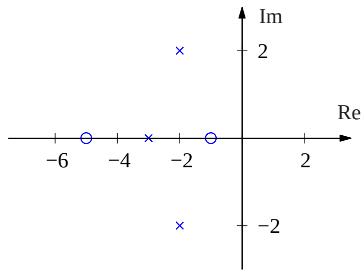


Figure 8.4: A pole zero diagram for a transfer function with zeros at -5 and -1 and poles at -3 and $-2 \pm 2j$. The circles represent the locations of the zeros, and the crosses the locations of the poles. A complete characterization requires we also specify the gain of the system.

model. One easy way to see this is to notice that the value of $G(s)$ is unbounded when s is an eigenvalue of a system since this is precisely the set of points where the characteristic polynomial $\lambda(s) = \det(sI - A) = 0$ (and hence $sI - A$ is noninvertible). It follows that the poles of a state space system depend only on the matrix A, which represents the intrinsic dynamics of the system. We say that a transfer function is stable if all of its poles have negative real part.

To find the zeros of a state space system, we observe that the zeros are complex numbers s such that the input $u(t) = u_0 e^{st}$ gives zero output. Inserting the pure exponential response $x(t) = x_0 e^{st}$ and $y(t) = 0$ in equation (8.2) gives

$$se^{st}x_0 = Ax_0e^{st} + Bu_0e^{st} \qquad 0 = Ce^{st}x_0 + De^{st}u_0,$$

which can be written as

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \begin{bmatrix} x_0 \\ u_0 \end{bmatrix} e^{st} = 0.$$

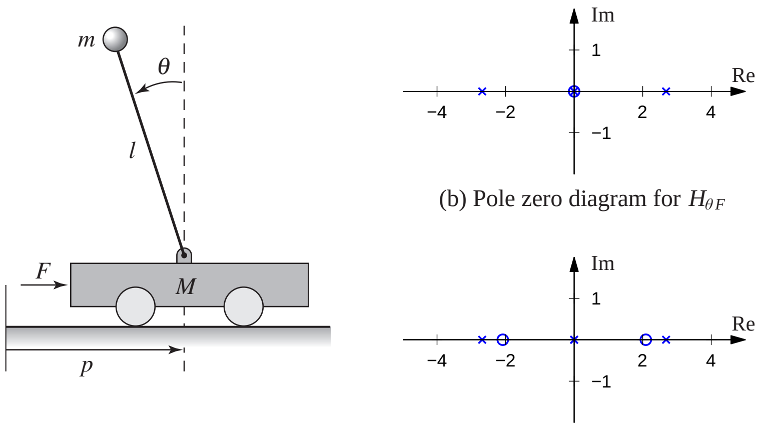
This equation has a solution with nonzero x_0, u_0 only if the matrix on the left does not have full rank. The zeros are thus the values s such that the matrix

$$\begin{bmatrix} A - sI & B \\ C & D \end{bmatrix} \tag{8.17}$$

looses rank.

Since the zeros depend on A, B, C and D, they therefore depend on how the inputs and outputs are coupled to the states. Notice in particular that if the matrix B has full row rank, then the matrix in equation (8.17) has n linearly independent rows for all values of s. Similarly there are n linearly independent columns if the matrix C has full column rank. This implies that systems where the matrix B or C is square and full rank do not have zeros. In particular it means that a system has no zeros if it is fully actuated (each state can be controlled independently) or if the full state is measured.

A convenient way to view the poles and zeros of a transfer function is through a *pole zero diagram*, as shown in Figure 8.4. In this diagram, each pole is marked with a cross, and each zero with a circle. If there are multiple poles or zeros at a



(a) Cart–pendulum system

(b) Pole zero diagram for $H_{\theta F}$

(c) Pole zero diagram for H_{pF}

Figure 8.5: Poles and zeros for a balance system. The balance system (a) can be modeled around its vertical equilibrium point by a fourth order linear system. The poles and zeros for the transfer functions $H_{\theta F}$ and H_{pF} are shown in (b) and (c), respectively.

fixed location, these are often indicated with overlapping crosses or circles (or other annotations). Poles in the left half-plane correspond to stable modes of the system, and poles in the right half-plane correspond to unstable modes. We thus call a pole in the left-half plane a *stable pole* and a pole in the right-half plane an *unstable pole*. A similar terminology is used for zeros, even though the zeros do not directly related to stability or instability of the system. Notice that the gain must also be given to have a complete description of the transfer function.

Example 8.5 Balance system

Consider the dynamics for a balance system, shown in Figure 8.5. The transfer function for a balance system can be derived directly from the second-order equations, given in Example 2.1:

$$M_t\frac{d^2p}{dt^2} - ml\frac{d^2\theta}{dt^2}\cos\theta + c\frac{dp}{dt} + ml\sin\theta\Big(\frac{d\theta}{dt}\Big)^2 = F,$$
$$-ml\cos\theta\frac{d^2p}{dt^2} + J_t\frac{d^2\theta}{dt^2} - mgl\sin\theta + \gamma\dot\theta = 0.$$

If we assume that θ and $\dot\theta$ are small, we can approximate this nonlinear system by a set of linear second-order differential equations,

$$M_t\frac{d^2p}{dt^2} - ml\frac{d^2\theta}{dt^2} + c\frac{dp}{dt} = F,$$
$$-ml\frac{d^2p}{dt^2} + J_t\frac{d^2\theta}{dt^2} + \gamma\frac{d\theta}{dt} - mgl\theta = 0.$$

If we let F be an exponential signal, the resulting response satisfies

$$M_t s^2\, p - mls^2\,\theta + cs\, p = F,$$
$$J_t s^2\,\theta - mls^2\, p + \gamma\, s\,\theta - mgl\,\theta = 0,$$

where all signals are exponential signals. The resulting transfer functions for the position of the cart and the orientation of the pendulum are given by solving for p and θ in terms of F to obtain

$$H_{\theta F} = \frac{mls}{(M_t J_t - m^2l^2)s^4 + (\gamma\, M_t + c\, J_t)s^2 + (c\gamma\, - M_t mgl)s - mglc},$$
$$H_{pF} = \frac{J_t s^2 + \gamma\, s - mgl}{(M_t J_t - m^2l^2)s^4 + (\gamma\, M_t + c\, J_t)s^3 + (c\gamma\, - M_t mgl)s^2 - mglcs},$$

where each of the coefficients is positive. The pole zero diagrams for these two transfer functions are shown in Figure 8.5 using the parameters from Example 6.7.

If we assume the damping is small and set $c = 0$ and $\gamma\, = 0$, we obtain

$$H_{\theta F} = \frac{ml}{(M_t J_t - m^2l^2)s^2 - M_t mgl},$$
$$H_{pF} = \frac{J_t s^2 - mgl}{s^2\big((M_t J_t - m^2l^2)s^2 - M_t mgl\big)}.$$

This gives nonzero poles and zeros at

$$p = \pm\sqrt{\frac{mgl M_t}{M_t J_t - m^2l^2}} \approx \pm 2.68, \qquad z = \pm\sqrt{\frac{mgl}{J_t}} \approx \pm 2.09.$$

We see that these are quite close to the pole and zero locations in Figure 8.5. ∇

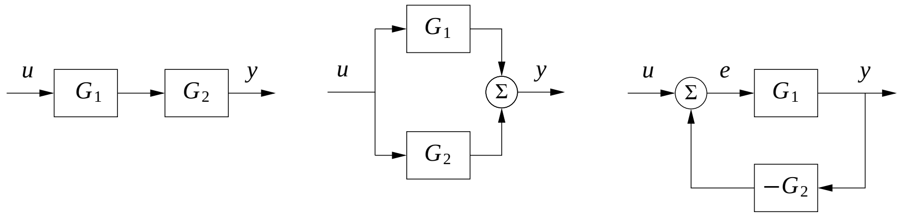
8.3 Block Diagrams and Transfer Functions

The combination of block diagrams and transfer functions is a powerful way to represent control systems. Transfer functions relating different signals in the system can be derived by purely algebraic manipulations of the transfer functions of the blocks using *block diagram algebra*. To show how this can be done, we will begin with simple combinations of systems.

Consider a system that is a cascade combination of systems with the transfer functions $G_1(s)$ and $G_2(s)$, as shown in Figure 8.6a. Let the input of the system be $u = e^{st}$. The pure exponential output of the first block is the exponential signal $G_1 u$, which is also the input to the second system. The pure exponential output of the second system is

$$y = G_2(G_1 u) = (G_2 G_1)u.$$

The transfer function of the series connection is thus $G = G_2 G_1$, i.e., the product of the transfer functions. The order of the individual transfer functions is due to the fact that we place the input signal on the right-hand side of this expression,



(a) $G_{yu} = G_2G_1$ (b) $G_{yu} = G_1 + G_2$ (c) $G_{yu} = \dfrac{G_1}{1 + G_1G_2}$

Figure 8.6: Interconnections of linear systems. Series (a), parallel (b) and feedback (c) connections are shown. The transfer functions for the composite systems can be derived by algebraic manipulations assuming exponential functions for all signals.

hence we first multiply by G_1 and then by G_2. Unfortunately, this has the opposite ordering from the diagrams that we use, where we typically have the signal flow from left to right, so one needs to be careful. The ordering is important if either G_1 or G_2 is a vector-valued transfer function, as we shall see in some examples.

Consider next a parallel connection of systems with the transfer functions G_1 and G_2, as shown in Figure 8.6b. Letting $u = e^{st}$ be the input to the system, the pure exponential output of the first system is then $y_1 = G_1 u$ and the output of the second system is $y_2 = G_2 u$. The pure exponential output of the parallel connection is thus

$$y = G_1 u + G_2 u = (G_1 + G_2)u,$$

and the transfer function for a parallel connection is $G = G_1 + G_2$.

Finally, consider a feedback connection of systems with the transfer functions G_1 and G_2, as shown in Figure 8.6c. Let $u = e^{st}$ be the input to the system, y be the pure exponential output, and e be the pure exponential part of the intermediate signal given by the sum of u and the output of the second block. Writing the relations for the different blocks and the summation unit, we find

$$y = G_1 e, \qquad e = u - G_2 y.$$

Elimination of e gives

$$y = G_1(u - G_2 y) \quad \Longrightarrow \quad (1 + G_1G_2)y = G_1 u \quad \Longrightarrow \quad y = \frac{G_1}{1 + G_1G_2}u.$$

The transfer function of the feedback connection is thus

$$G = \frac{G_1}{1 + G_1G_2}.$$

These three basic interconnections can be used as the basis for computing transfer functions for more complicated systems.

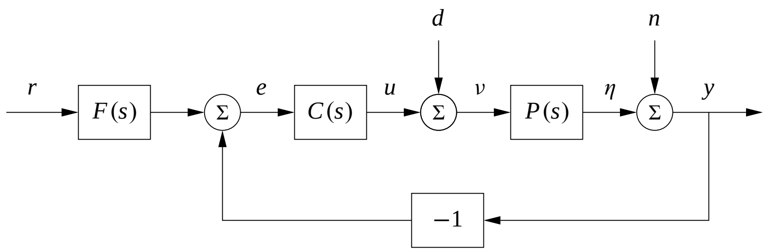


Figure 8.7: Block diagram of a feedback system. The inputs to the system are the reference signal r, the process disturbance d and the measurement noise n. The remaining signals in the system can all be chosen as possible outputs, and transfer functions can be used to relate the system inputs to the other labeled signals.

Control System Transfer Functions

Consider the system in Figure 8.7, which was given at the beginning of the chapter. The system has three blocks representing a process P, a feedback controller C and a feedforward controller F. Together, C and F define the *control law* for the system. There are three external signals: the reference (or command signal) r, the load disturbance d and the measurement noise n. A typical problem is to find out how the error e is related to the signals r, d and n.

To derive the relevant transfer functions we assume that all signals are exponential signals, drop the arguments of signals and transfer functions and trace the signals around the loop. We begin with the signal in which we are interested, in this case the control error e, given by

$$e = Fr - y.$$

The signal y is the sum of n and η, where η is the output of the process:

$$y = n + \eta, \qquad \eta = P(d + u), \qquad u = Ce.$$

Combining these equations gives

$$\begin{aligned} e &= Fr - y = Fr - (n + \eta) = Fr - \bigl(n + P(d + u)\bigr) \\ &= Fr - \bigl(n + P(d + Ce)\bigr), \end{aligned}$$

and hence

$$e = Fr - n - Pd - PCe.$$

Finally, solving this equation for e gives

$$e = \frac{F}{1 + PC} r - \frac{1}{1 + PC} n - \frac{P}{1 + PC} d = G_{er} r + G_{en} n + G_{ed} d, \qquad (8.18)$$

and the error is thus the sum of three terms, depending on the reference r, the measurement noise n and the load disturbance d. The functions

$$G_{er} = \frac{F}{1 + PC}, \qquad G_{en} = \frac{-1}{1 + PC}, \qquad G_{ed} = \frac{-P}{1 + PC} \qquad (8.19)$$

are transfer functions from reference r, noise n and disturbance d to the error e.

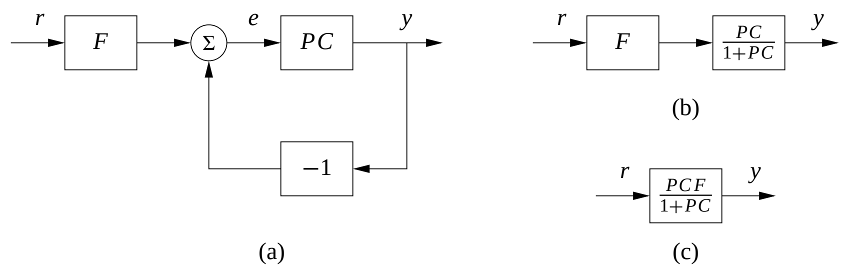


Figure 8.8: Example of block diagram algebra. The results from multiplying the process and controller transfer functions (from Figure 8.7) are shown in (a). Replacing the feedback loop with its transfer function equivalent yields (b), and finally multiplying the two remaining blocks gives the reference to output representation in (c).

We can also derive transfer functions by manipulating the block diagrams directly, as illustrated in Figure 8.8. Suppose we wish to compute the transfer function between the reference r and the output y. We begin by combining the process and controller blocks in Figure 8.7 to obtain the diagram in Figure 8.8a. We can now eliminate the feedback loop using the algebra for a feedback interconnection (Figure 8.8b) and then use the series interconnection rule to obtain

$$G_{yr} = \frac{PCF}{1 + PC}. \tag{8.20}$$

Similar manipulations can be used to obtain the other transfer functions (Exercise 8.8).

The derivation illustrates an effective way to manipulate the equations to obtain the relations between inputs and outputs in a feedback system. The general idea is to start with the signal of interest and to trace signals around the feedback loop until coming back to the signal we started with. With some practice, equations (8.18) and (8.19) can be written directly by inspection of the block diagram. Notice, for example, that all terms in equation (8.19) have the same denominators and that the numerators are the blocks that one passes through when going directly from input to output (ignoring the feedback). This type of rule can be used to compute transfer functions by inspection, although for systems with multiple feedback loops it can be tricky to compute them without writing down the algebra explicitly.

Example 8.6 Vehicle steering

Consider the linearized model for vehicle steering introduced in Example 5.12. In Examples 6.4 and 7.3 we designed a state feedback compensator and state estimator for the system. A block diagram for the resulting control system is given in Figure 8.9. Note that we have split the estimator into two components, $G_{\hat{x}u}(s)$ and $G_{\hat{x}y}(s)$, corresponding to its inputs u and y. The controller can be described as the sum of two (open loop) transfer functions

$$u = G_{uy}(s)y + G_{ur}(s)r.$$

The first transfer function, $G_{uy}(s)$, describes the feedback term and the second, $G_{ur}(s)$, describes the feedforward term. We call these *open loop* transfer functions

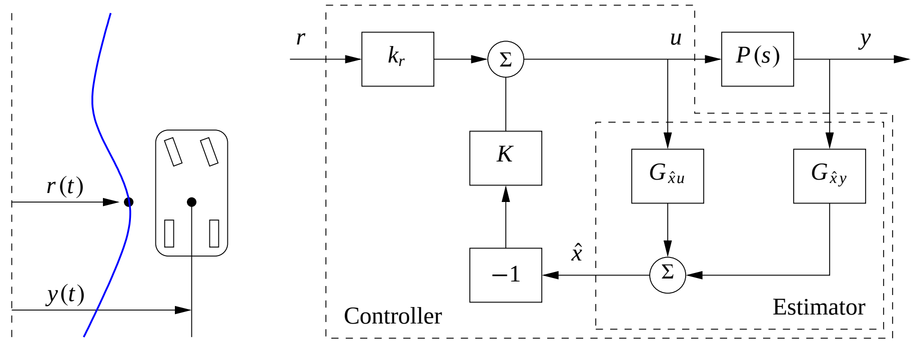


Figure 8.9: Block diagram for a steering control system. The control system is designed to maintain the lateral position of the vehicle along a reference curve (left). The structure of the control system is shown on the right as a block diagram of transfer functions. The estimator consists of two components that compute the estimated state $\hat{x}$ from the combination of the input u and output y of the process. The estimated state is fed through a state feedback controller and combined with a reference gain to obtain the commanded steering angle u.

because they represent the relationships between the signals without considering the dynamics of the process (e.g., removing $P(s)$ from the system description). To derive these functions, we compute the transfer functions for each block and then use block diagram algebra.

We begin with the estimator, which takes u and y as its inputs and produces an estimate $\hat{x}$. The dynamics for this process were derived in Example 7.3 and are given by

$$\frac{d\hat{x}}{dt} = (A - LC)\hat{x} + Ly + Bu,$$
$$\hat{x} = \underbrace{\left(sI - (A - LC)\right)^{-1}B}_{G_{\hat{x}u}}\,u + \underbrace{\left(sI - (A - LC)\right)^{-1}L}_{G_{\hat{x}y}}\,y.$$

Using the expressions for A, B, C and L from Example 7.3, we obtain

$$G_{\hat{x}u}(s) = \begin{pmatrix} \dfrac{\gamma s + 1}{s^2 + l_1 s + l_2} \\[2ex] \dfrac{s + l_1 - \gamma l_2}{s^2 + l_1 s + l_2} \end{pmatrix}, \qquad G_{\hat{x}y}(s) = \begin{pmatrix} \dfrac{l_1 s + l_2}{s^2 + l_1 s + l_2} \\[2ex] \dfrac{l_2 s}{s^2 + l_1 s + l_2} \end{pmatrix},$$

where l_1 and l_2 are the observer gains and γ is the scaled position of the center of mass from the rear wheels. The controller was a state feedback compensator, which can be viewed as a constant, multi-input, single-output transfer function of the form $u = -K\hat{x}$.

We can now proceed to compute the transfer function for the overall control system. Using block diagram algebra, we have

$$G_{uy}(s) = \frac{-KG_{\hat{x}y}(s)}{1 + KG_{\hat{x}u}(s)} = -\frac{s(k_1 l_1 + k_2 l_2) + k_1 l_2}{s^2 + s(\gamma k_1 + k_2 + l_1) + k_1 + l_2 + k_2 l_1 - \gamma k_2 l_2}$$

and

$$G_{ur}(s) = \frac{k_r}{1 + KG_{\hat{x}u}(s)} = \frac{k_1(s^2 + l_1 s + l_2)}{s^2 + s(\gamma k_1 + k_2 + l_1) + k_1 + l_2 + k_2 l_1 - \gamma k_2 l_2},$$

where k_1 and k_2 are the controller gains.

Finally, we compute the full closed loop dynamics. We begin by deriving the transfer function for the process $P(s)$. We can compute this directly from the state space description of the dynamics, which was given in Example 5.12. Using that description, we have

$$P(s) = G_{yu}(s) = C(sI - A)^{-1}B + D = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} s & -1 \\ 0 & s \end{bmatrix}^{-1} \begin{bmatrix} \gamma \\ 1 \end{bmatrix} = \frac{\gamma s + 1}{s^2}.$$

The transfer function for the full closed loop system between the input r and the output y is then given by

$$G_{yr} = \frac{k_r P(s)}{1 + P(s)G_{uy}(s)} = \frac{k_1(\gamma s + 1)}{s^2 + (k_1\gamma + k_2)s + k_1}.$$

Note that the observer gains l_1 and l_2 do not appear in this equation. This is because we are considering steady-state analysis and, in steady state, the estimated state exactly tracks the state of the system assuming perfect models. We will return to this example in Chapter 12 to study the robustness of this particular approach. ∇

Pole/Zero Cancellations

Because transfer functions are often polynomials in s, it can sometimes happen that the numerator and denominator have a common factor, which can be canceled. Sometimes these cancellations are simply algebraic simplifications, but in other situations they can mask potential fragilities in the model. In particular, if a pole/zero cancellation occurs because terms in separate blocks that just happen to coincide, the cancellation may not occur if one of the systems is slightly perturbed. In some situations this can result in severe differences between the expected behavior and the actual behavior.

To illustrate when we can have pole/zero cancellations, consider the block diagram in Figure 8.7 with $F = 1$ (no feedforward compensation) and C and P given by

$$C(s) = \frac{n_c(s)}{d_c(s)}, \qquad P(s) = \frac{n_p(s)}{d_p(s)}.$$

The transfer function from r to e is then given by

$$G_{er}(s) = \frac{1}{1 + PC} = \frac{d_c(s)d_p(s)}{d_c(s)d_p(s) + n_c(s)n_p(s)}.$$

If there are common factors in the numerator and denominator polynomials, then these terms can be factored out and eliminated from both the numerator and denominator. For example, if the controller has a zero at $s = -a$ and the process has a pole at $s = -a$, then we will have

$$G_{er}(s) = \frac{(s+a)d_c'(s)d_p(s)}{(s+a)d_c(s)d_p'(s) + (s+a)n_c'(s)n_p(s)} = \frac{d_c'(s)d_p(s)}{d_c(s)d_p'(s) + n_c'(s)n_p(s)},$$

where $n_c'(s)$ and $d_p'(s)$ represent the relevant polynomials with the term $s + a$ factored out. In the case when $a < 0$ (so that the zero or pole is in the right half-plane), we see that there is no impact on the transfer function G_{er}.

Suppose instead that we compute the transfer function from d to e, which represents the effect of a disturbance on the error between the reference and the output. This transfer function is given by

$$G_{ed}(s) = \frac{d_c'(s)n_p(s)}{(s+a)d_c(s)d_p'(s) + (s+a)n_c'(s)n_p(s)}.$$

Notice that if $a < 0$, then the pole is in the right half-plane and the transfer function G_{ed} is *unstable*. Hence, even though the transfer function from r to e appears to be okay (assuming a perfect pole/zero cancellation), the transfer function from d to e can exhibit unbounded behavior. This unwanted behavior is typical of an *unstable pole/zero cancellation*.

It turns out that the cancellation of a pole with a zero can also be understood in terms of the state space representation of the systems. Reachability or observability is lost when there are cancellations of poles and zeros (Exercise 8.11). A consequence is that the transfer function represents the dynamics only in the reachable and observable subspace of a system (see Section 7.5).

Example 8.7 Cruise control

The input/output response from throttle to velocity for the linearized model for a car has the transfer function $G(s) = b/(s-a)$, $a < 0$. A simple (but not necessarily good) way to design a PI controller is to choose the parameters of the PI controller so that the controller zero at $s = -k_i/k_p$ cancels the process pole at $s = a$. The transfer function from reference to velocity is $G_{vr}(s) = bk_p/(s+bk_p)$, and control design is simply a matter of choosing the gain k_p. The closed loop system dynamics are of first order with the time constant $1/bk_p$.

Figure 8.10 shows the velocity error when the car encounters an increase in the road slope. A comparison with the controller used in Figure 3.3b (reproduced in dashed curves) shows that the controller based on pole/zero cancellation has very poor performance. The velocity error is larger, and it takes a long time to settle.

Notice that the control signal remains practically constant after $t = 15$ even if the error is large after that time. To understand what happens we will analyze the system. The parameters of the system are $a = -0.0101$ and $b = 1.32$, and the controller parameters are $k_p = 0.5$ and $k_i = 0.0051$. The closed loop time constant is $1/(bk_p) = 2.5$ s, and we would expect that the error would settle in about 10 s (4 time constants). The transfer functions from road slope to velocity and control

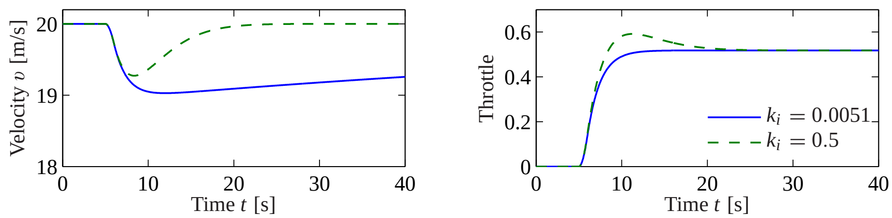


Figure 8.10: Car with PI cruise control encountering a sloping road. The velocity error is shown on the left and the throttle is shown on the right. Results with a PI controller with $k_p = 0.5$ and $k_i = 0.0051$, where the process pole $s = -0.0101$, is shown by solid lines, and a controller with $k_p = 0.5$ and $k_i = 0.5$ is shown by dashed lines. Compare with Figure 3.3b.

signals are

$$G_{v\theta}(s) = \frac{b_g k_p s}{(s-a)(s+bk_p)}, \qquad G_{u\theta}(s) = \frac{bk_p}{s+bk_p}.$$

Notice that the canceled mode $s = a = -0.0101$ appears in $G_{v\theta}$ but not in $G_{u\theta}$. The reason why the control signal remains constant is that the controller has a zero at $s = -0.0101$, which cancels the slowly decaying process mode. Notice that the error would diverge if the canceled pole was unstable. ∇

The lesson we can learn from this example is that it is a bad idea to try to cancel unstable or slow process poles. A more detailed discussion of pole/zero cancellations is given in Section 12.4.

Algebraic Loops

When analyzing or simulating a system described by a block diagram, it is necessary to form the differential equations that describe the complete system. In many cases the equations can be obtained by combining the differential equations that describe each subsystem and substituting variables. This simple procedure cannot be used when there are closed loops of subsystems that all have a direct connection between inputs and outputs, known as an *algebraic loop*.

To see what can happen, consider a system with two blocks, a first-order nonlinear system,

$$\frac{dx}{dt} = f(x,u), \qquad y = h(x), \tag{8.21}$$

and a proportional controller described by $u = -ky$. There is no direct term since the function h does not depend on u. In that case we can obtain the equation for the closed loop system simply by replacing u by $-ky$ in (8.21) to give

$$\frac{dx}{dt} = f(x,-ky), \qquad y = h(x).$$

Such a procedure can easily be automated using simple formula manipulation.

The situation is more complicated if there is a direct term. If $y = h(x, u)$, then replacing u by $-ky$ gives

$$\frac{dx}{dt} = f(x, -ky), \qquad y = h(x, -ky).$$

To obtain a differential equation for x, the algebraic equation $y = h(x, -ky)$ must be solved to give $y = \alpha(x)$, which in general is a complicated task.

When algebraic loops are present, it is necessary to solve algebraic equations to obtain the differential equations for the complete system. Resolving algebraic loops is a nontrivial problem because it requires the symbolic solution of algebraic equations. Most block diagram-oriented modeling languages cannot handle algebraic loops, and they simply give a diagnosis that such loops are present. In the era of analog computing, algebraic loops were eliminated by introducing fast dynamics between the loops. This created differential equations with fast and slow modes that are difficult to solve numerically. Advanced modeling languages like Modelica use several sophisticated methods to resolve algebraic loops.

8.4 The Bode Plot

The frequency response of a linear system can be computed from its transfer function by setting $s = i\omega$, corresponding to a complex exponential

$$u(t) = e^{i\omega t} = \cos(\omega t) + i \sin(\omega t).$$

The resulting output has the form

$$y(t) = G(i\omega)e^{i\omega t} = Me^{i(\omega t+\varphi)} = M\cos(\omega t + \varphi) + iM\sin(\omega t + \varphi),$$

where M and φ are the gain and phase of G:

$$M = |G(i\omega)|, \qquad \varphi = \arctan\frac{\operatorname{Im} G(i\omega)}{\operatorname{Re} G(i\omega)}.$$

The phase of G is also called the *argument* of G, a term that comes from the theory of complex variables.

It follows from linearity that the response to a single sinusoid (sin or cos) is amplified by M and phase-shifted by φ. Note that $-\pi < \varphi \le \pi$, so the arctangent must be taken respecting the signs of the numerator and denominator. It will often be convenient to represent the phase in degrees rather than radians. We will use the notation $\angle G(i\omega)$ for the phase in degrees and $\arg G(i\omega)$ for the phase in radians. In addition, while we always take $\arg G(i\omega)$ to be in the range $(-\pi, \pi]$, we will take $\angle G(i\omega)$ to be continuous, so that it can take on values outside the range of -180° to 180°.

The frequency response $G(i\omega)$ can thus be represented by two curves: the gain curve and the phase curve. The *gain curve* gives $|G(i\omega)|$ as a function of frequency ω, and the *phase curve* gives $\angle G(i\omega)$. One particularly useful way of drawing these

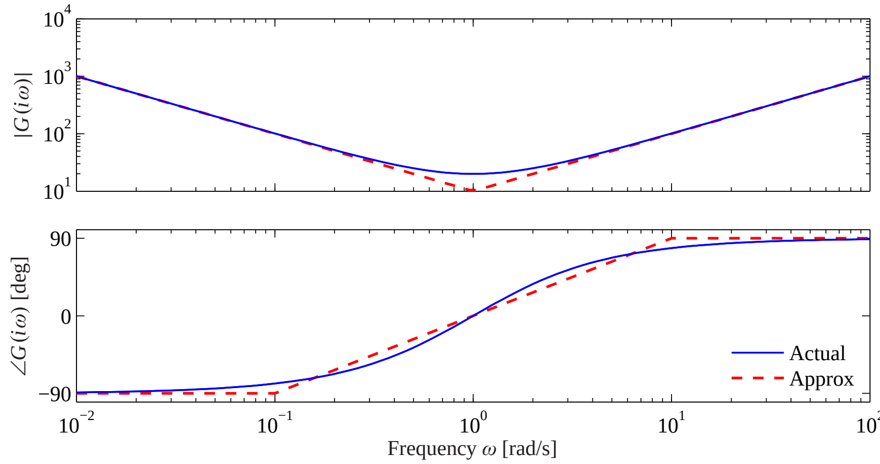


Figure 8.11: Bode plot of the transfer function $C(s) = 20 + 10/s + 10s$ corresponding to an ideal PID controller. The top plot is the gain curve and the bottom plot is the phase curve. The dashed lines show straight-line approximations of the gain curve and the corresponding phase curve.

curves is to use a log/log scale for the gain plot and a log/linear scale for the phase plot. This type of plot is called a *Bode plot* and is shown in Figure 8.11.

Sketching and Interpreting Bode Plots

Part of the popularity of Bode plots is that they are easy to sketch and interpret. Since the frequency scale is logarithmic, they cover the behavior of a linear system over a wide frequency range.

Consider a transfer function that is a rational function of the form

$$G(s) = \frac{b_1(s)b_2(s)}{a_1(s)a_2(s)}.$$

We have

$$\log |G(s)| = \log |b_1(s)| + \log |b_2(s)| - \log |a_1(s)| - \log |a_2(s)|,$$

and hence we can compute the gain curve by simply adding and subtracting gains corresponding to terms in the numerator and denominator. Similarly,

$$\angle G(s) = \angle b_1(s) + \angle b_2(s) - \angle a_1(s) - \angle a_2(s),$$

and so the phase curve can be determined in an analogous fashion. Since a polynomial can be written as a product of terms of the type

$$k, \quad s, \quad s + a, \quad s^2 + 2\zeta\omega_0 s + \omega_0^2,$$

it suffices to be able to sketch Bode diagrams for these terms. The Bode plot of a complex system is then obtained by adding the gains and phases of the terms.

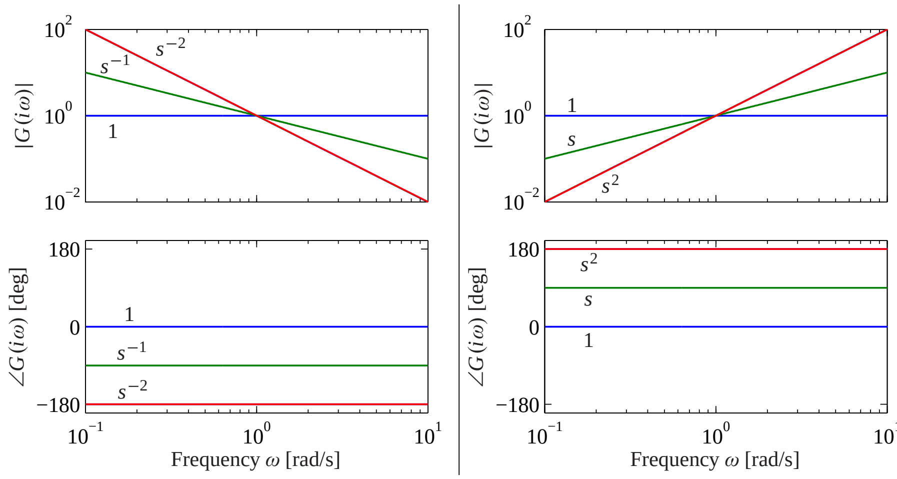


Figure 8.12: Bode plots of the transfer functions $G(s) = s^k$ for $k = -2, -1, 0, 1, 2$. On a log-log scale, the gain curve is a straight line with slope k. Using a log-linear scale, the phase curves for the transfer functions are constants, with phase equal to $90^\circ \times k$

.

The simplest term in a transfer function is one of the form s^k, where $k > 0$ if the term appears in the numerator and $k < 0$ if the term is in the denominator. The gain and phase of the term are given by

$$\log |G(i\omega)| = k \log \omega, \quad \angle G(i\omega) = 90k.$$

The gain curve is thus a straight line with slope k, and the phase curve is a constant at $90^\circ \times k$. The case when $k = 1$ corresponds to a differentiator and has slope 1 with phase 90°. The case when $k = -1$ corresponds to an integrator and has slope -1 with phase -90°. Bode plots of the various powers of k are shown in Figure 8.12.

Consider next the transfer function of a first-order system, given by

$$G(s) = \frac{a}{s + a}.$$

We have

$$|G(s)| = \frac{|a|}{|s + a|}, \qquad \angle G(s) = \angle(a) - \angle(s + a),$$

and hence

$$\log |G(i\omega)| = \log a - \frac{1}{2} \log (\omega^2 + a^2), \qquad \angle G(i\omega) = -\frac{180}{\pi} \arctan \frac{\omega}{a}.$$

The Bode plot is shown in Figure 8.13a, with the magnitude normalized by the zero frequency gain. Both the gain curve and the phase curve can be approximated by

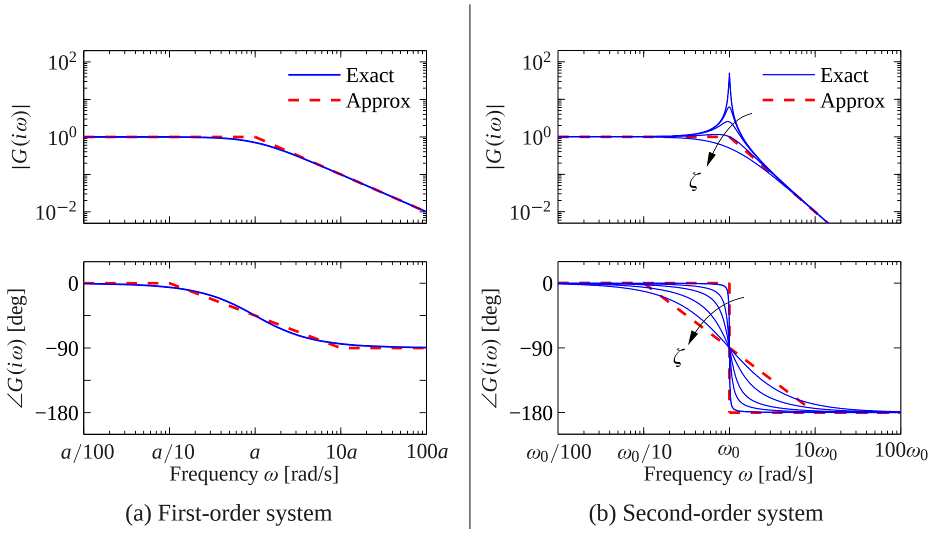


Figure 8.13: Bode plots for first- and second-order systems. (a) The first-order system $G(s) = a/(s + a)$ can be approximated by asymptotic curves (dashed) in both the gain and the frequency, with the breakpoint in the gain curve at $\omega = a$ and the phase decreasing by 90° over a factor of 100 in frequency. (b) The second-order system $G(s) = \omega_0^2/(s^2+2\zeta\omega_0 s+\omega_0^2)$ has a peak at frequency a and then a slope of −2 beyond the peak; the phase decreases from 0° to −180°. The height of the peak and the rate of change of phase depending on the damping ratio ζ ($\zeta = 0.02$, 0.1, 0.2, 0.5 and 1.0 shown).

the following straight lines

$$\log |G(i\omega)| \approx \begin{cases} 0 & \text{if } \omega < a \\ \log a - \log \omega & \text{if } \omega > a, \end{cases}$$

$$\angle G(i\omega) \approx \begin{cases} 0 & \text{if } \omega < a/10 \\ -45 - 45(\log \omega - \log a) & a/10 < \omega < 10a \\ -90 & \text{if } \omega > 10a. \end{cases}$$

The approximate gain curve consists of a horizontal line up to frequency $\omega = a$, called the *breakpoint* or *corner frequency*, after which the curve is a line of slope −1 (on a log-log scale). The phase curve is zero up to frequency $a/10$ and then decreases linearly by 45°/decade up to frequency $10a$, at which point it remains constant at 90°. Notice that a first-order system behaves like a constant for low frequencies and like an integrator for high frequencies; compare with the Bode plot in Figure 8.12.

Finally, consider the transfer function for a second-order system,

$$G(s) = \frac{\omega_0^2}{s^2 + 2\omega_0\zeta s + \omega_0^2},$$

for which we have

$$\log |G(i\omega)| = 2\log\omega_0 - \frac{1}{2}\log\left(\omega^4 + 2\omega_0^2\omega^2(2\zeta^2 - 1) + \omega_0^4\right),$$
$$\angle G(i\omega) = -\frac{180}{\pi}\arctan\frac{2\zeta\omega_0\omega}{\omega_0^2 - \omega^2}.$$

The gain curve has an asymptote with zero slope for $\omega \ll \omega_0$. For large values of ω the gain curve has an asymptote with slope -2. The largest gain $Q = \max_\omega |G(i\omega)| \approx 1/(2\zeta)$, called the *Q-value*, is obtained for $\omega \approx \omega_0$. The phase is zero for low frequencies and approaches 180° for large frequencies. The curves can be approximated with the following piecewise linear expressions

$$\log |G(i\omega)| \approx \begin{cases} 0 & \text{if } \omega \ll \omega_0 \\ 2\log\omega_0 - 2\log\omega & \text{if } \omega \gg \omega_0, \end{cases}$$
$$\angle G(i\omega) \approx \begin{cases} 0 & \text{if } \omega \ll \omega_0 \\ -180 & \text{if } \omega \gg \omega_0. \end{cases}$$

The Bode plot is shown in Figure 8.13b. Note that the asymptotic approximation is poor near $\omega = \omega_0$ and that the Bode plot depends strongly on ζ near this frequency.

Given the Bode plots of the basic functions, we can now sketch the frequency response for a more general system. The following example illustrates the basic idea.

Example 8.8 Asymptotic approximation for a transfer function
Consider the transfer function given by

$$G(s) = \frac{k(s+b)}{(s+a)(s^2 + 2\zeta\omega_0 s + \omega_0^2)}, \qquad a \ll b \ll \omega_0.$$

The Bode plot for this transfer function appears in Figure 8.14, with the complete transfer function shown as a solid line and the asymptotic approximation shown as a dashed line.

We begin with the gain curve. At low frequency, the magnitude is given by

$$G(0) = \frac{kb}{a\omega_0^2}.$$

When we reach $\omega = a$, the effect of the pole begins and the gain decreases with slope -1. At $\omega = b$, the zero comes into play and we increase the slope by 1, leaving the asymptote with net slope 0. This slope is used until the effect of the second-order pole is seen at $\omega = \omega_c$, at which point the asymptote changes to slope -2. We see that the gain curve is fairly accurate except in the region of the peak due to the second-order pole (since for this case ζ is reasonably small).

The phase curve is more complicated since the effect of the phase stretches out much further. The effect of the pole begins at $\omega = a/10$, at which point we change from phase 0 to a slope of $-45°$/decade. The zero begins to affect the phase at $\omega = b/10$, producing a flat section in the phase. At $\omega = 10a$ the phase

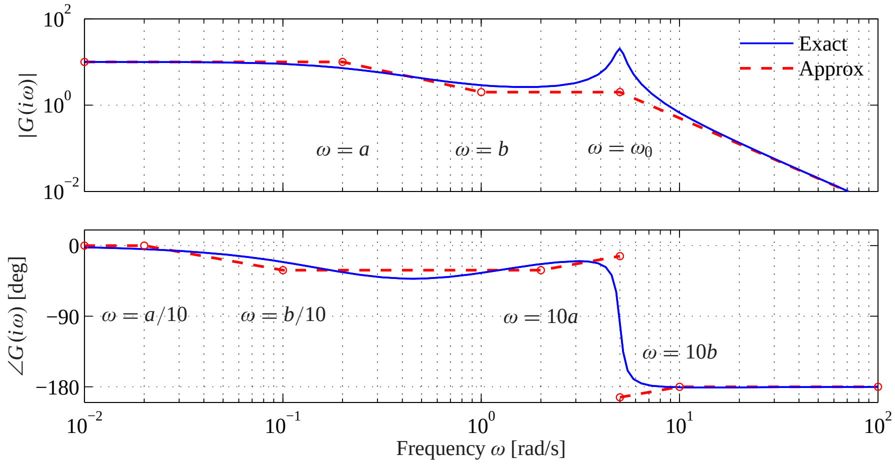


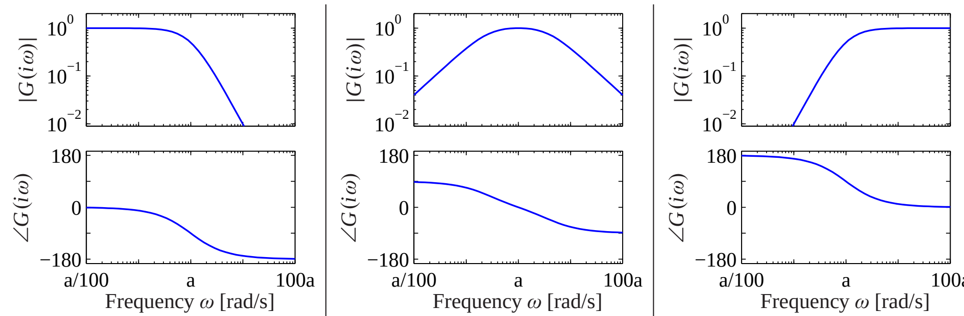
Figure 8.14: Asymptotic approximation to a Bode plot. The thin line is the Bode plot for the transfer function $G(s) = k(s+b)/(s+a)(s^2 + 2\zeta\omega_0 s + \omega_0^2)$, where $a \ll b \ll \omega_0$. Each segment in the gain and phase curves represents a separate portion of the approximation, where either a pole or a zero begins to have effect. Each segment of the approximation is a straight line between these points at a slope given by the rules for computing the effects of poles and zeros.

contributions from the pole end, and we are left with a slope of +45°/decade (from the zero). At the location of the second-order pole, $s \approx i\omega_c$, we get a jump in phase of -180°. Finally, at $\omega = 10b$ the phase contributions of the zero end, and we are left with a phase of -180 degrees. We see that the straight-line approximation for the phase is not as accurate as it was for the gain curve, but it does capture the basic features of the phase changes as a function of frequency. ∇

The Bode plot gives a quick overview of a system. Since any signal can be decomposed into a sum of sinusoids, it is possible to visualize the behavior of a system for different frequency ranges. The system can be viewed as a filter that can change the amplitude (and phase) of the input signals according to the frequency response. For example, if there are frequency ranges where the gain curve has constant slope and the phase is close to zero, the action of the system for signals with these frequencies can be interpreted as a pure gain. Similarly, for frequencies where the slope is +1 and the phase close to 90°, the action of the system can be interpreted as a differentiator, as shown in Figure 8.12.

Three common types of frequency responses are shown in Figure 8.15. The system in Figure 8.15a is called a *low-pass filter* because the gain is constant for low frequencies and drops for high frequencies. Notice that the phase is zero for low frequencies and -180° for high frequencies. The systems in Figure 8.15b and c are called a *band-pass filter* and *high-pass filter* for similar reasons.

To illustrate how different system behaviors can be read from the Bode plots we consider the band-pass filter in Figure 8.15b. For frequencies around $\omega = \omega_0$, the signal is passed through with no change in gain. However, for frequencies well



$$G(s) = \frac{\omega_0^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$

(a) Low-pass filter

$$G(s) = \frac{2\zeta\omega_0 s}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$

(b) Band-pass filter

$$G(s) = \frac{s^2}{s^2 + 2\zeta\omega_0 s + \omega_0^2}$$

(c) High-pass filter

Figure 8.15: Bode plots for low-pass, band-pass and high-pass filters. The top plots are the gain curves and the bottom plots are the phase curves. Each system passes frequencies in a different range and attenuates frequencies outside of that range.

below or well above ω_0, the signal is attenuated. The phase of the signal is also affected by the filter, as shown in the phase curve. For frequencies below $a/100$ there is a phase lead of 90°, and for frequencies above $100a$ there is a phase lag of 90°. These actions correspond to differentiation and integration of the signal in these frequency ranges.

Example 8.9 Transcriptional regulation

Consider a genetic circuit consisting of a single gene. We wish to study the response of the protein concentration to fluctuations in the mRNA dynamics. We consider two cases: a *constitutive promoter* (no regulation) and self-repression (negative feedback), illustrated in Figure 8.16. The dynamics of the system are given by

$$\frac{dm}{dt} = \alpha(p) - \gamma m - u, \qquad \frac{dp}{dt} = \beta m - \delta p,$$

where u is a disturbance term that affects mRNA transcription.

For the case of no feedback we have $\alpha(p) = \alpha_0$, and the system has an equilibrium point at $m_e = \alpha_0/\gamma$, $p_e = \beta\alpha_0/(\delta\gamma)$. The transfer function from v to p is given by

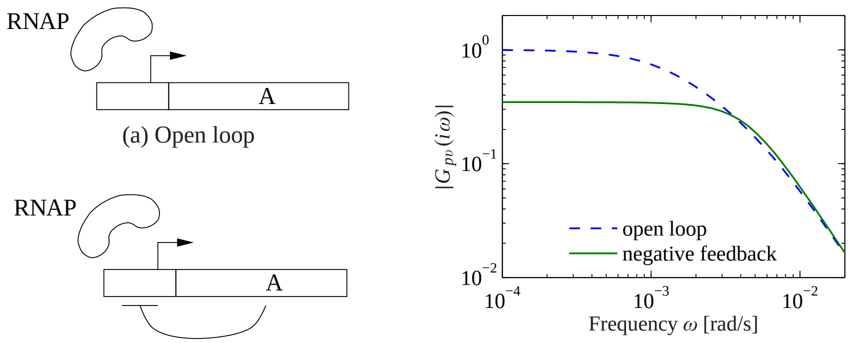
$$G^{\mathrm{ol}}_{pv}(s) = \frac{-\beta}{(s+\gamma)(s+\delta)}.$$

For the case of negative regulation, we have

$$\alpha(p) = \frac{\alpha_1}{1 + kp^n} + \alpha_0,$$

and the equilibrium points satisfy

$$m_e = \frac{\delta}{\beta} p_e, \qquad \frac{\alpha}{1 + kp_e^n} + \alpha_0 = \gamma m_e = \frac{\gamma\delta}{\beta} p_e.$$



(a) Open loop (b) Negative feedback (c) Frequency response

Figure 8.16: Noise attenuation in a genetic circuit. The open loop system (a) consists of a constitutive promoter, while the closed loop circuit (b) is self-regulated with negative feedback (repressor). The frequency response for each circuit is shown in (c).

The resulting transfer function is given by

$$G_{pv}^{\text{cl}}(s) = \frac{\beta}{(s+\gamma)(s+\delta)+\beta\sigma}, \qquad \sigma = \frac{n\alpha_1 k p_e^{n-1}}{(1+kp_e^n)^2}.$$

Figure 8.16c shows the frequency response for the two circuits. We see that the feedback circuit attenuates the response of the system to disturbances with low-frequency content but slightly amplifies disturbances at high frequency (compared to the open loop system). Notice that these curves are very similar to the frequency response curves for the op amp shown in Figure 8.3b. ∇

Transfer Functions from Experiments

The transfer function of a system provides a summary of the input/output response and is very useful for analysis and design. However, modeling from first principles can be difficult and time-consuming. Fortunately, we can often build an input/output model for a given application by directly measuring the frequency response and fitting a transfer function to it. To do so, we perturb the input to the system using a sinusoidal signal at a fixed frequency. When steady state is reached, the amplitude ratio and the phase lag give the frequency response for the excitation frequency. The complete frequency response is obtained by sweeping over a range of frequencies.

By using correlation techniques it is possible to determine the frequency response very accurately, and an analytic transfer function can be obtained from the frequency response by curve fitting. The success of this approach has led to instruments and software that automate this process, called *spectrum analyzers*. We illustrate the basic concept through two examples.

Example 8.10 Atomic force microscope

To illustrate the utility of spectrum analysis, we consider the dynamics of the atomic force microscope, introduced in Section 3.5. Experimental determination of the

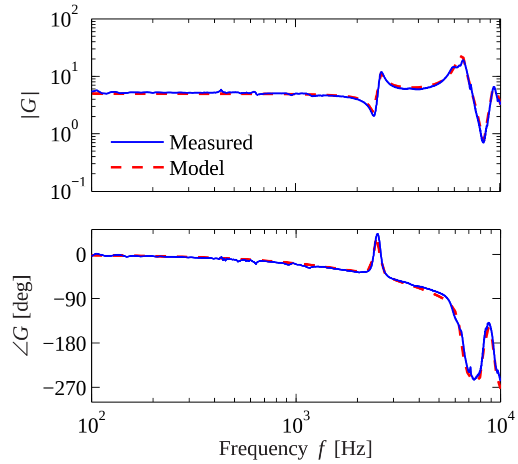


Figure 8.17: Frequency response of a preloaded piezoelectric drive for an atomic force microscope. The Bode plot shows the response of the measured transfer function (solid) and the fitted transfer function (dashed).

frequency response is particularly attractive for this system because its dynamics are very fast and hence experiments can be done quickly. A typical example is given in Figure 8.17, which shows an experimentally determined frequency response (solid line). In this case the frequency response was obtained in less than a second. The transfer function

$$G(s) = \frac{k\omega_2^2\omega_3^2\omega_5^2(s^2 + 2\zeta_1\omega_1 s + \omega_1^2)(s^2 + 2\zeta_4\omega_4 s + \omega_4^2)e^{-s\tau}}{\omega_1^2\omega_4^2(s^2 + 2\zeta_2\omega_2 s + \omega_2^2)(s^2 + 2\zeta_3\omega_3 s + \omega_3^2)(s^2 + 2\zeta_5\omega_5 s + \omega_5^2)},$$

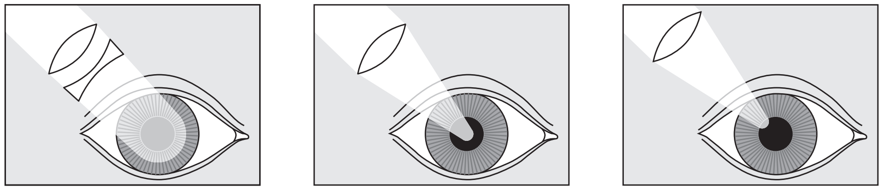
with $\omega_k = 2\pi f_k$ and $f_1 = 2.42$ kHz, $\zeta_1 = 0.03$, $f_2 = 2.55$ kHz, $\zeta_2 = 0.03$, $f_3 = 6.45$ kHz, $\zeta_3 = 0.042$, $f_4 = 8.25$ kHz, $\zeta_4 = 0.025$, $f_5 = 9.3$ kHz, $\zeta_5 = 0.032$, $\tau = 10^{-4}$ s and $k = 5$, was fit to the data (dashed line). The frequencies associated with the zeros are located where the gain curve has minima, and the frequencies associated with the poles are located where the gain curve has local maxima. The relative damping ratios are adjusted to give a good fit to maxima and minima. When a good fit to the gain curve is obtained, the time delay is adjusted to give a good fit to the phase curve. The piezo drive is preloaded, and a simple model of its dynamics is derived in Exercise 3.7. The pole at 2.42 kHz corresponds to the trampoline mode derived in the exercise; the other resonances are higher modes.

∇

Example 8.11 Pupillary light reflex dynamics

The human eye is an organ that is easily accessible for experiments. It has a control system that adjusts the pupil opening to regulate the light intensity at the retina.

This control system was explored extensively by Stark in the 1960s [Sta68]. To determine the dynamics, light intensity on the eye was varied sinusoidally and the pupil opening was measured. A fundamental difficulty is that the closed loop system is insensitive to internal system parameters, so analysis of a closed loop



(a) Closed loop (b) Open loop (c) High gain

Figure 8.18: Light stimulation of the eye. In (a) the light beam is so large that it always covers the whole pupil, giving closed loop dynamics. In (b) the light is focused into a beam which is so narrow that it is not influenced by the pupil opening, giving open loop dynamics. In (c) the light beam is focused on the edge of the pupil opening, which has the effect of increasing the gain of the system since small changes in the pupil opening have a large effect on the amount of light entering the eye. From Stark [Sta68].

system thus gives little information about the internal properties of the system. Stark used a clever experimental technique that allowed him to investigate both open and closed loop dynamics. He excited the system by varying the intensity of a light beam focused on the eye and measured pupil area, as illustrated in Figure 8.18. By using a wide light beam that covers the whole pupil, the measurement gives the closed loop dynamics. The open loop dynamics were obtained by using a narrow beam, which is small enough that it is not influenced by the pupil opening. The result of one experiment for determining open loop dynamics is given in Figure 8.19. Fitting a transfer function to the gain curve gives a good fit for $G(s) = 0.17/(1+0.08s)^3$. This curve gives a poor fit to the phase curve as shown by the dashed curve in Figure 8.19. The fit to the phase curve is improved by adding a time delay, which leaves the gain curve unchanged while substantially modifying the phase curve. The final fit gives the model

$$G(s) = \frac{0.17}{(1+0.08s)^3} e^{-0.2s}.$$

The Bode plot of this is shown with solid curves in Figure 8.19. Modeling of the pupillary reflex from first principles is discussed in detail in [KS01]. ∇

Notice that for both the AFM drive and pupillary dynamics it is not easy to derive appropriate models from first principles. In practice, it is often fruitful to use a combination of analytical modeling and experimental identification of parameters. Experimental determination of frequency response is less attractive for systems with slow dynamics because the experiment takes a long time.

8.5 Laplace Transforms

Transfer functions are conventionally introduced using Laplace transforms, and in this section we derive the transfer function using this formalism. We assume basic familiarity with Laplace transforms; students who are not familiar with them can safely skip this section. A good reference for the mathematical material in this

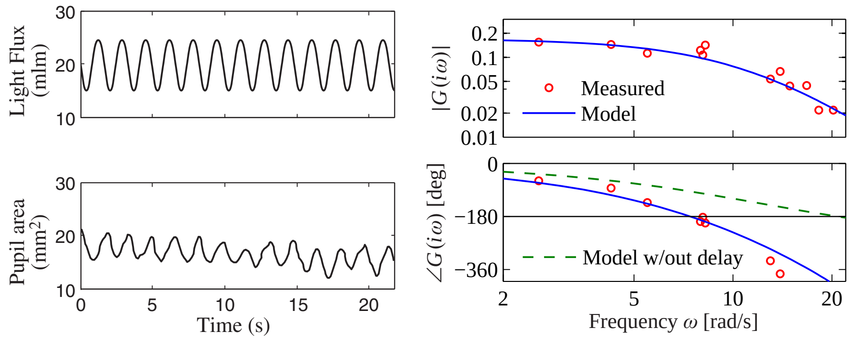


Figure 8.19: Sample curves from an open loop frequency response of the eye (left) and a Bode plot for the open loop dynamics (right). The solid curve shows a fit of the data using a third-order transfer function with time delay. The dashed curve in the Bode plot is the phase of the system without time delay, showing that the delay is needed to properly capture the phase. (Figure redrawn from the data of Stark [Sta68].)

section is the classic book by Widder [Wid41].

Traditionally, Laplace transforms were used to compute responses of linear systems to different stimuli. Today we can easily generate the responses using computers. Only a few elementary properties are needed for basic control applications. There is, however, a beautiful theory for Laplace transforms that makes it possible to use many powerful tools from the theory of functions of a complex variable to get deep insights into the behavior of systems.

Consider a function $f(t)$, $f : \mathbb{R}^+ \to \mathbb{R}$, that is integrable and grows no faster than $e^{s_0 t}$ for some finite $s_0 \in \mathbb{R}$ and large t. The Laplace transform maps f to a function $F = \mathcal{L} f : \mathbb{C} \to \mathbb{C}$ of a complex variable. It is defined by

$$F(s) = \int_0^\infty e^{-st} f(t)\, dt, \quad \operatorname{Re} s > s_0. \tag{8.22}$$

The transform has some properties that makes it well suited to deal with linear systems.

First we observe that the transform is linear because

$$\begin{aligned} \mathcal{L}(af + bg) &= \int_0^\infty e^{-st}(af(t) + bg(t))\, dt \\ &= a\int_0^\infty e^{-st} f(t)\, dt + b\int_0^\infty e^{-st} g(t)\, dt = a\mathcal{L}f + b\mathcal{L}g. \end{aligned} \tag{8.23}$$

Next we calculate the Laplace transform of the derivative of a function. We have

$$\mathcal{L}\frac{df}{dt} = \int_0^\infty e^{-st} f'(t)\, dt = e^{-st} f(t)\Big|_0^\infty + s\int_0^\infty e^{-st} f(t)\, dt = -f(0) + s\mathcal{L}f,$$

where the second equality is obtained using integration by parts. We thus obtain

$$\mathcal{L}\frac{df}{dt} = s\mathcal{L}f - f(0) = sF(s) - f(0). \tag{8.24}$$

This formula is particularly simple if the initial conditions are zero because it follows that differentiation of a function corresponds to multiplication of the transform by s.

Since differentiation corresponds to multiplication by s, we can expect that integration corresponds to division by s. This is true, as can be seen by calculating the Laplace transform of an integral. Using integration by parts, we get

$$\begin{aligned}\mathcal{L}\int_0^t f(\tau)\,d\tau &= \int_0^\infty \Big(e^{-st}\int_0^t f(\tau)\,d\tau\Big)\,dt \\ &= -\frac{e^{-st}}{s}\int_0^t f(\tau)\,d\tau\Big|_0^\infty + \int_0^\infty \frac{e^{-s\tau}}{s} f(\tau)\,d\tau = \frac{1}{s}\int_0^\infty e^{-s\tau} f(\tau)\,d\tau,\end{aligned}$$

hence

$$\mathcal{L}\int_0^t f(\tau)\,d\tau = \frac{1}{s}\mathcal{L}f = \frac{1}{s}F(s). \tag{8.25}$$

Next consider a linear time-invariant system with zero initial state. We saw in Section 5.3 that the relation between the input u and the output y is given by the convolution integral

$$y(t) = \int_0^\infty h(t-\tau)u(\tau)\,d\tau,$$

where $h(t)$ is the impulse response for the system. Taking the Laplace transform of this expression, we have

$$\begin{aligned}Y(s) &= \int_0^\infty e^{-st}y(t)\,dt = \int_0^\infty e^{-st}\int_0^\infty h(t-\tau)u(\tau)\,d\tau\,dt \\ &= \int_0^\infty \int_0^t e^{-s(t-\tau)}e^{-s\tau}h(t-\tau)u(\tau)\,d\tau\,dt \\ &= \int_0^\infty e^{-s\tau}u(\tau)\,d\tau \int_0^\infty e^{-st}h(t)\,dt = H(s)U(s).\end{aligned}$$

Thus, the input/output response is given by $Y(s) = H(s)U(s)$, where H, U and Y are the Laplace transforms of h, u and y. The system theoretic interpretation is that the Laplace transform of the output of a linear system is a product of two terms, the Laplace transform of the input $U(s)$ and the Laplace transform of the impulse response of the system $H(s)$. A mathematical interpretation is that the Laplace transform of a convolution is the product of the transforms of the functions that are convolved. The fact that the formula $Y(s) = H(s)U(s)$ is much simpler than a convolution is one reason why Laplace transforms have become popular in engineering.

We can also use the Laplace transform to derive the transfer function for a state space system. Consider, for example, a linear state space system described by

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx + Du.$$

Taking Laplace transforms *under the assumption that all initial values are zero* gives

$$sX(s) = AX(s) + BU(s) \qquad Y(s) = CX(s) + DU(s).$$

Elimination of $X(s)$ gives

$$Y(s) = \Big(C(sI - A)^{-1}B + D\Big)U(s). \tag{8.26}$$

The transfer function is $G(s) = C(sI - A)^{-1}B + D$ (compare with equation (8.4)).

8.6 Further Reading

The idea of characterizing a linear system by its steady-state response to sinusoids was introduced by Fourier in his investigation of heat conduction in solids [Fou07]. Much later, it was used by the electrical engineer Steinmetz who introduced the $i\omega$ method for analyzing electrical circuits. Transfer functions were introduced via the Laplace transform by Gardner Barnes [GB42], who also used them to calculate the response of linear systems. The Laplace transform was very important in the early phase of control because it made it possible to find transients via tables (see, e.g., [JNP47]). Combined with block diagrams, transfer functions and Laplace transforms provided powerful techniques for dealing with complex systems. Calculation of responses based on Laplace transforms is less important today, when responses of linear systems can easily be generated using computers. There are many excellent books on the use of Laplace transforms and transfer functions for modeling and analysis of linear input/output systems. Traditional texts on control such as [DB04], [FPEN05] and [Oga01] are representative examples. Pole/zero cancellation was one of the mysteries of early control theory. It is clear that common factors can be canceled in a rational function, but cancellations have system theoretical consequences that were not clearly understood until Kalman's decomposition of a linear system was introduced [KHN63]. In the following chapters, we will use transfer functions extensively to analyze stability and to describe model uncertainty.

Exercises

8.1 Let $G(s)$ be the transfer function for a linear system. Show that if we apply an input $u(t) = A\sin(\omega t)$, then the steady-state output is given by $y(t) = |G(i\omega)|A\sin(\omega t + \arg G(i\omega))$. (Hint: Start by showing that the real part of a complex number is a linear operation and then use this fact.)

8.2 Consider the system

$$\frac{dx}{dt} = ax + u.$$

Compute the exponential response of the system and use this to derive the transfer function from u to y. Show that when $s = a$, a pole of the transfer function, the response to the exponential input $u(t) = e^{st}$ is $x(t) = e^{at}x(0) + te^{at}$.

8.3 (Inverted pendulum) A model for an inverted pendulum was introduced in Example 2.2. Neglecting damping and linearizing the pendulum around the upright position gives a linear system characterized by the matrices

$$A = \begin{pmatrix} 0 & 1 \\ mgl/J_t & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 1/J_t \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad D = 0.$$

Determine the transfer function of the system.

8.4 (Solutions corresponding to poles and zeros) Consider the differential equation

$$\frac{d^n y}{dt^n} + a_1 \frac{d^{n-1} y}{dt^{n-1}} + \cdots + a_n y = b_1 \frac{d^{n-1} u}{dt^{n-1}} + b_2 \frac{d^{n-2} u}{dt^{n-2}} + \cdots + b_n u.$$

(a) Let λ be a root of the characteristic polynomial

$$s^n + a_1 s^{n-1} + \cdots + a_n = 0.$$

Show that if $u(t) = 0$, the differential equation has the solution $y(t) = e^{\lambda t}$.

(b) Let κ be a zero of the polynomial

$$b(s) = b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n.$$

Show that if the input is $u(t) = e^{\kappa t}$, then there is a solution to the differential equation that is identically zero.

8.5 (Operational amplifier) Consider the operational amplifier introduced in Section 3.3 and analyzed in Example 8.3. A PI controller can be constructed using an op amp by replacing the resistor R_2 with a resistor and capacitor in series, as shown in Figure 3.10. The resulting transfer function of the circuit is given by

$$G(s) = -\left(R_2 + \frac{1}{Cs}\right) \cdot \left(\frac{kCs}{(kR_1C + R_2C)s + 1}\right),$$

where k is the gain of the op amp, R_1 and R_2 are the resistances in the compensation network and C is the capacitance.

(a) Sketch the Bode plot for the system under the assumption that $k \gg R_2 > R_1$. You should label the key features in your plot, including the gain and phase at low frequency, the slopes of the gain curve, the frequencies at which the gain changes slope, etc.

(b) Suppose now that we include some dynamics in the amplifier, as outlined in Example 8.1. This would involve replacing the gain k with the transfer function

$$H(s) = \frac{k}{1 + sT}.$$

Compute the resulting transfer function for the system (i.e., replace k with $H(s)$) and find the poles and zeros assuming the following parameter values

$$\frac{R_2}{R_1} = 100, \qquad k = 10^6, \qquad R_2C = 1, \qquad T = 0.01.$$

(c) Sketch the Bode plot for the transfer function in part (b) using straight line approximations and compare this to the exact plot of the transfer function (using MATLAB). Make sure to label the important features in your plot.

8.6 (Transfer function for state space system) Consider the linear state space system

$$\frac{dx}{dt} = Ax + Bu, \qquad y = Cx.$$

(a) Show that the transfer function is

$$G(s) = \frac{b_1 s^{n-1} + b_2 s^{n-2} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n},$$

where

$$b_1 = CB, \quad b_2 = CAB + a_1 CB, \quad \ldots, \quad b_n = CA^{n-1}B + a_1 CA^{n-1}B + \cdots + a_{n-1}CB$$

and $\lambda(s) = s^n + a_1 s^{n-1} + \cdots + a_n$ is the characteristic polynomial for A.

(b) Compute the transfer function for a linear system in reachable canonical form and show that it matches the transfer function given above.

8.7 (Kalman decomposition) Show that the transfer function of a system depends only on the dynamics in the reachable and observable subspace of the Kalman decomposition. (Hint: Consider the representation given by equation (7.27).)

8.8 Using block diagram algebra, show that the transfer functions from d to y and n to y in Figure 8.7 are given by

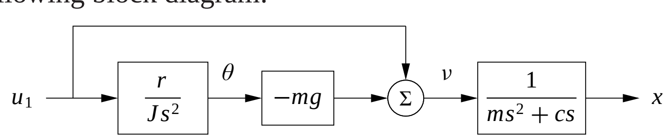
$$G_{yd} = \frac{P}{1 + PC} \qquad G_{yd} = \frac{1}{1 + PC}.$$

8.9 (Bode plot for a simple zero) Show that the Bode plot for transfer function $G(s) = (s + a)/a$ can be approximated by

$$\log |G(i\omega)| \approx \begin{cases} 0 & \text{if } \omega < a \\ \log\omega - \log a & \text{if } \omega > a, \end{cases}$$

$$\angle G(i\omega) \approx \begin{cases} 0 & \text{if } \omega < a/10 \\ 45 + 45(\log\omega - \log a) & a/10 < \omega < 10a \\ 90 & \text{if } \omega > 10a. \end{cases}$$

8.10 (Vectored thrust aircraft) Consider the lateral dynamics of a vectored thrust aircraft as described in Example 2.9. Show that the dynamics can be described using the following block diagram:



Use this block diagram to compute the transfer functions from u_1 to θ and x and show that they satisfy

$$H_{\theta u_1} = \frac{r}{Js^2}, \qquad H_{xu_1} = \frac{Js^2 - mgr}{Js^2(ms^2 + cs)}.$$

8.11 (Common poles) Consider a closed loop system of the form of Figure 8.7, with $F = 1$ and P and C having a pole/zero cancellation. Show that if each system is written in state space form, the resulting closed loop system is not reachable and not observable.

8.12 (Congestion control) Consider the congestion control model described in Section 3.4. Let w represent the individual window size for a set of N identical sources, q represent the end-to-end probability of a dropped packet, b represent the number of packets in the router's buffer and p represent the probability that that a packet is dropped by the router. We write $\bar{w} = Nw$ to represent the total number of packets being received from all N sources. Show that the linearized model can be described by the transfer functions

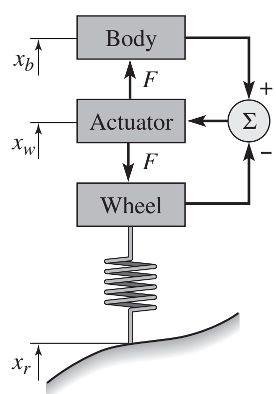
$$G_{b\bar{w}}(s) = \frac{e^{-\tau_f s}}{\tau_e s + e^{-\tau_f s}}, \qquad G_{\bar{w}q}(s) = -\frac{N}{q_e(\tau_e s + q_e w_e)}, \qquad G_{pb}(s) = \rho,$$

where (w_e, b_e) is the equilibrium point for the system, τ_e is the steady-state round-trip time and τ_f is the forward propagation time.

8.13 (Inverted pendulum with PD control) Consider the normalized inverted pendulum system, whose transfer function is given by $P(s) = 1/(s^2 - 1)$ (Exercise 8.3). A proportional-derivative control law for this system has transfer function $C(s) = k_p + k_d s$ (see Table 8.1). Suppose that we choose $C(s) = \alpha(s - 1)$. Compute the closed loop dynamics and show that the system has good tracking of reference signals but does not have good disturbance rejection properties.

8.14 (Vehicle suspension [HB90]) Active and passive damping are used in cars to give a smooth ride on a bumpy road. A schematic diagram of a car with a damping system in shown in the figure below.

(Porter Class I race car driven by Todd Cuffaro)



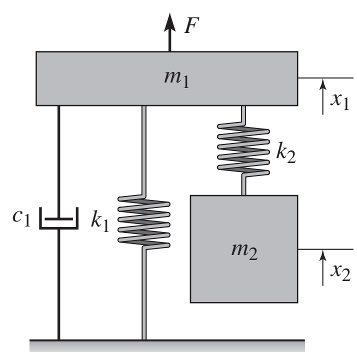
This model is called a *quarter car model*, and the car is approximated with two masses, one representing one fourth of the car body and the other a wheel. The actuator exerts a force F between the wheel and the body based on feedback from the distance between the body and the center of the wheel (the *rattle space*).

Let x_b, x_w and x_r represent the heights of body, wheel and road measured from their equilibria. A simple model of the system is given by Newton's equations for the body and the wheel,

$$m_b\ddot{x}_b = F, \qquad m_w\ddot{x}_w = -F + k_t(x_r - x_w),$$

where m_b is a quarter of the body mass, m_w is the effective mass of the wheel including brakes and part of the suspension system (the *unsprung mass*) and k_t is the tire stiffness. For a conventional damper consisting of a spring and a damper, we have $F = k(x_w - x_b) + c(\dot{x}_w - \dot{x}_b)$. For an active damper the force F can be more general and can also depend on riding conditions. Rider comfort can be characterized by the transfer function G_{ax_r} from road height x_r to body acceleration $a = \ddot{x}_b$. Show that this transfer function has the property $G_{ax_r}(i\omega_t) = k_t/m_b$, where $\omega_t = \sqrt{k_t/m_w}$ (the *tire hop frequency*). The equation implies that there are fundamental limitations to the comfort that can be achieved with any damper.

8.15 (Vibration absorber) Damping vibrations is a common engineering problem. A schematic diagram of a damper is shown below:



The disturbing vibration is a sinusoidal force acting on mass m_1, and the damper consists of the mass m_2 and the spring k_2. Show that the transfer function from disturbance force to height x_1 of the mass m_1 is

$$G_{x_1F} = \frac{m_2s^2 + k_2}{m_1m_2s^4 + m_2c_1s^3 + (m_1k_2 + m_2(k_1 + k_2))s^2 + k_2c_1s + k_1k_2}.$$

How should the mass m_2 and the stiffness k_2 be chosen to eliminate a sinusoidal oscillation with frequency ω_0. (More details are vibration absorbers is given in the classic text by Den Hartog [DH85, pp. 87–93].)