CHAPTER 3

MODELS OF LINEAR SYSTEMS

The nonlinear system models that are the topic of this book are all generalizations, in one way or another, of linear models. Thus, to understand nonlinear models, it is first necessary to appreciate the structure and behavior of the linear system models from which they evolved. Consequently, this chapter will review a variety of mathematical models used for linear systems.

The fundamental properties of linear systems will be defined first and then the most commonly used models for linear systems will be introduced. Methods for identifying models of linear systems will be presented in Chapter 5.

3.1 LINEAR SYSTEMS

Linear systems must obey both the principles of proportionality (1.5) and superposition (1.6). Thus, if $\mathbf{N}$ is a linear system, then

$$\mathbf{N}(k_1 u_1(t) + k_2 u_2(t)) = k_1 y_1(t) + k_2 y_2(t)$$

where k_1 and k_2 are any two scalar constants and

$$y_1(t) = \mathbf{N}(u_1(t))$$
$$y_2(t) = \mathbf{N}(u_2(t))$$

are any two input–output pairs.

Identification of Nonlinear Physiological Systems, By David T. Westwick and Robert E. Kearney
ISBN 0-471-27456-9

3.2 NONPARAMETRIC MODELS

The principles of superposition and proportionality may be used to develop nonparametric models of linear systems. Conceptually, the idea is to select a set of basis functions capable of describing any input signal. The set of the system's responses to these basis functions can then be used to predict the response to any input; that is, it provides a model of the system.

The decomposition of an arbitrary input into a set of scaled basis functions is illustrated in Figure 3.1; in this case the basis functions consist of a series of delayed pulses, shown in the left column. The input, shown at the top of the figure, is projected onto the delayed pulses to yield the expansion consisting of scaled copies of the basis elements, shown in the right column. The input signal may be reconstructed by summing the elements of the expansion.

Knowing the system's response to each basis function, the response to each scaled basis function can be determined using proportionality (1.5). Then, by superposition (1.6), the individual response components may be summed to produce the overall response. This is illustrated in Figure 3.2, where the scaled basis functions and the corresponding responses of a linear system are shown in the first three rows. The bottom row shows that summing these generates the original input signal and the resulting response.

Note that this approach will work with any set of basis functions that spans the set of input signals. However, it is desirable to select a set of basis functions that are orthogonal since this leads to the most concise signal representations and best conditioning of identification algorithms.

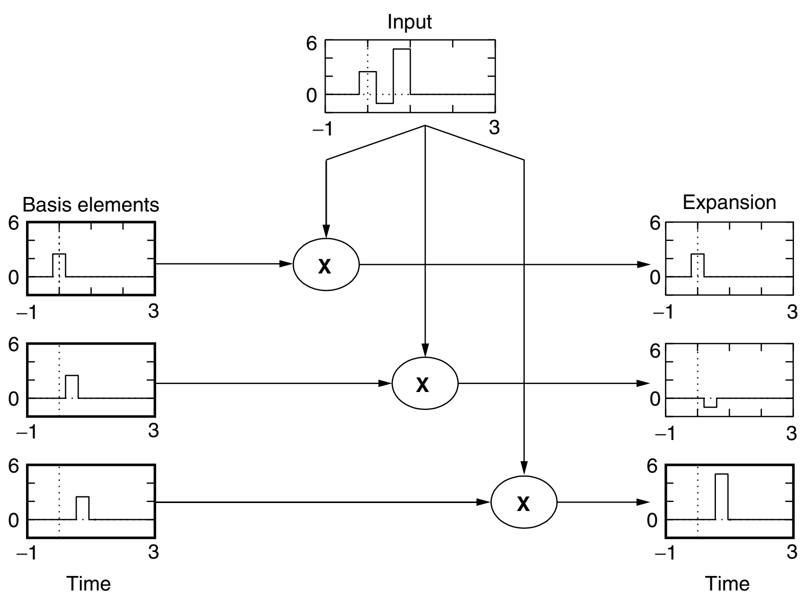

Figure 3.1 Expansion of a signal onto a basis of delayed pulses. The input signal (top) is projected onto a series of delayed pulses (left), resulting in the decomposition shown on the right.

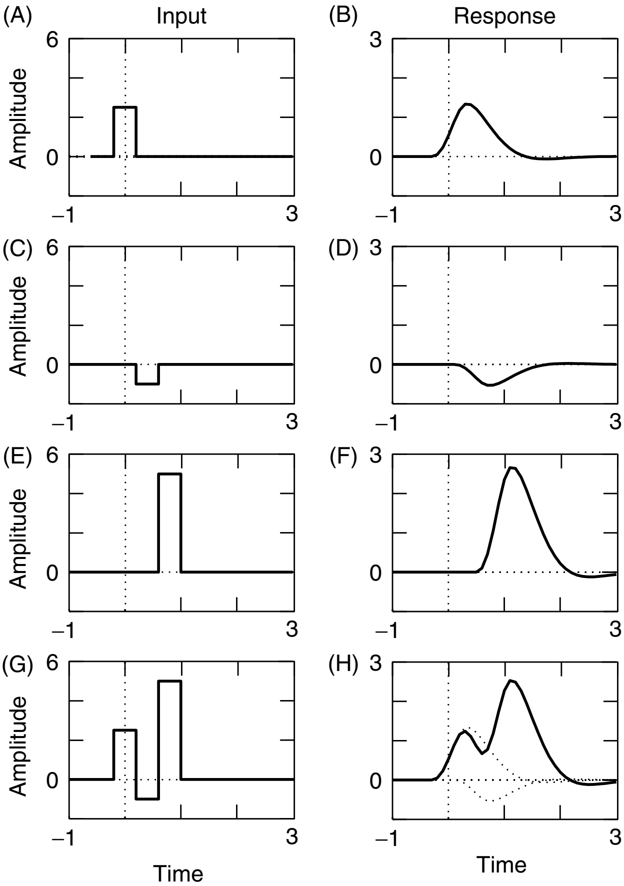

Figure 3.2 Linear systems and superposition. (A, C, E) Scaled and delayed unit pulse inputs. (B, D, F) Responses of a linear, time-invariant system to the scaled pulses. (G) The sum of the three scaled, delayed pulses. (H) The system's response to the input G. By superposition, this is the sum of the responses to the individual pulses.

3.2.1 Time Domain Models

Using the set of delayed pulses as basis functions leads to a nonparametric time domain model. Consider a pulse of width Δ_t and of unit area

$$d(t, \Delta_t) = \begin{cases} \left(\dfrac{1}{\Delta_t}\right) & \text{for } |t| < \dfrac{\Delta_t}{2} \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

as illustrated in Figure 3.2A. Let the response of the linear system, $\mathbf{N}$, to a single pulse be

$$\mathbf{N}(d(t, \Delta_t)) = h(t, \Delta_t) \tag{3.2}$$

Next, consider an input signal consisting of a weighted sum of delayed pulses,

$$u(t) = \sum_{k=-\infty}^{\infty} u_k\, d(t - k\Delta_t, \Delta_t) \tag{3.3}$$

By superposition, the output generated by this "staircase" input will be

$$\begin{aligned} y(t) &= \mathbf{N}(u(t)) \\ &= \sum_{k=-\infty}^{\infty} u_k \mathbf{N}(d(t - k\Delta_t, \Delta_t)) \\ &= \sum_{k=-\infty}^{\infty} u_k h(t - k\Delta_t, \Delta_t) \end{aligned} \tag{3.4}$$

The relationships among these signals are illustrated in Figure 3.2. Thus, if the input can be represented as the weighted sum of pulses, then the output can be written as the equivalently weighted sum of the pulse responses.

Now, consider what happens in the limit, as $\Delta_t \to 0$, and the unit-area pulses become impulses,

$$\lim_{\Delta_t \to 0} d(t, \Delta_t) = \delta(t) \tag{3.5}$$

The pulse response becomes the impulse response, $h(t)$, and the sum of equation (3.4) becomes the *convolution integral*,

$$\boxed{y(t) = \int_{-\infty}^{\infty} h(\tau) u(t - \tau)\, d\tau \qquad (3.6)}$$

Thus, the response of the system to an arbitrary input can be determined by convolving it with the system's *impulse response*. Hence, the impulse response function (IRF) provides a complete model of a system's dynamic response in the *time domain*.*

Theoretically, the limits of the integration in (3.6) extend from $\tau = -\infty$ to $\tau = \infty$. However, in practice, the impulse response will usually be of finite duration so that

$$h(\tau) = 0 \qquad \text{for } \tau \leq T_1 \text{ or } \tau \geq T_2 \tag{3.7}$$

The value of T_1, the lower integration limit in the convolution, determines whether the system is causal or noncausal. If $T_1 \geq 0$, the IRF starts at the same time or after the impulse. In contrast, if $T_1 < 0$, the IRF starts before the impulse is applied and the system is *noncausal*. Any physically realizable system must be causal but, as discussed in Section 1.2.2, there are important, practical situations where noncausal responses are observed. These include behavioral experiments with predictable inputs, signals measured from inside feedback loops, and situations where the roles of the system input and output are deliberately reversed for analysis purposes.

The value of T_2, the upper integration limit in the convolution (3.6), determines the memory length of the system. This defines how long the response to a single impulse

*The convolution operation, (3.6), is often abbreviated using an centered asterisk—that is, $y(t) = h(\tau) * u(t)$.

lasts or, conversely, how long a "history" must be considered when computing the current output.

Thus, for a finite memory, causal system, the convolution integral simplifies to

$$y(t) = \int_0^T h(\tau)u(t-\tau)\,d\tau \tag{3.8}$$

If the sampling rate is adequate, the convolution integral can be converted to discrete time using rectangular integration, to give the sum:

$$\boxed{y(t) = \sum_{\tau=0}^{T-1} h(\tau)u(t-\tau)\Delta_t} \tag{3.9}$$

where Δ_t is the sampling interval. Note that in this formulation the time variable, t, and time lag, τ, are discrete variables that are integer multiples of the sampling interval. Notice that although the upper limit of the summation is $T-1$, the memory length is described as "T," the number of samples between 0 and $T-1$ inclusively.

In practice, the impulse response is often scaled to incorporate the effect of the sampling increment, Δ_t. Thus, equation (3.9) is often written as

$$y(t) = \sum_{\tau=0}^{T-1} g(\tau)u(t-\tau)$$

where $g(\tau) = \Delta_t h(\tau)$.

The linear IRF will be generalized, in Sections 4.1 and 4.2, to produce the Volterra and Wiener series models for nonlinear systems. Furthermore, the IRF is the basic building block for block-structured models and their generalizations, to be discussed in Sections 4.3–4.5.

3.2.1.1 Example: Human Ankle Compliance Model—Impulse Response

Throughout this chapter, a running example will be used to illustrate the linear system models. The system used in this running example is the dynamic compliance of the human ankle, which defines the dynamic relationship between torque, $T_q(t)$, and position, $\Theta(t)$. Under some conditions, this relation is well modeled by a causal, linear, time-invariant system (Kearney and Hunter, 1990). The example used in this chapter is based on a transfer function model (see Section 3.3.1 below) of ankle compliance using parameter values from a review article on human joint dynamics (Kearney and Hunter, 1990).

Figure 3.3 shows a typical ankle compliance IRF. Inspection of this IRF provides several insights into the system. First, the largest peak is negative, indicating that the output position, at least at low frequencies, will have the opposite sign from the input. Second, the system is causal with a memory of less than 100 ms. Finally, the decaying oscillations in the IRF suggest that the system is somewhat resonant.

3.2.2 Frequency Domain Models

Sinusoids are another set of basis functions commonly used to model signals—in this case in the frequency domain. The superposition and scaling properties of linear systems

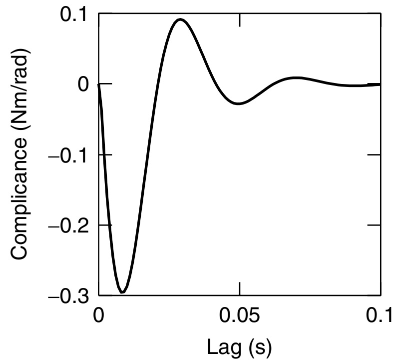

Figure 3.3 Impulse response of the dynamic compliance of the human ankle.

guarantee that the steady-state response to a sinusoidal input will be a sinusoid at the same frequency, but with a different amplitude and phase. Thus, the steady-state response of a linear system may be fully characterized in terms of how the amplitude and phase of its sinusoidal response change with the input frequency. Normally, these are expressed as the complex-valued *frequency response* of the system, $H(j\omega)$. The magnitude, $|H(j\omega)|$, describes how the amplitude of the input is scaled, whereas the argument (a.k.a. angle or phase), $\phi(H(j\omega))$, defines the phase shift. Thus, the output generated by the sinusoidal input, $u(t) = \sin(\omega t)$, is given by

$$y(t) = |H(j\omega)| \sin(\omega t + \phi(H(j\omega))) \tag{3.10}$$

The *Fourier transform*

$$U(j\omega) = \int_{-\infty}^{\infty} u(t)e^{-j\omega t}\, dt \tag{3.11}$$

expands time domain signals onto an infinite basis of sinusoids. It provides a convenient tool for using the frequency response to predict the response to an arbitrary input, $u(t)$. Thus, the Fourier transform of the input, $U(j\omega)$, is computed and then multiplied by the frequency response, $H(j\omega)$, to give the Fourier transform of the output,

$$\boxed{Y(j\omega) = H(j\omega)U(j\omega) \qquad (3.12)}$$

Taking the inverse Fourier transform of $Y(j\omega)$ gives the time domain response, $y(t)$.

Methods for estimating the frequency response will be discussed in Section 5.3. The nonlinear generalization of the frequency response, along with methods for its identification, will be presented in Section 6.1.3.

3.2.2.1 Example: Human Ankle Compliance — Frequency Response

Figure 3.4 shows the frequency response, the gain and phase plotted as functions of frequency, of the human ankle compliance dynamics corresponding to the IRF presented in Figure 3.3. The gain plot shows that at low frequency the response is almost constant at about −50 dB, there is a slight resonance at about 25 Hz, and the response falls off rapidly at higher frequencies.

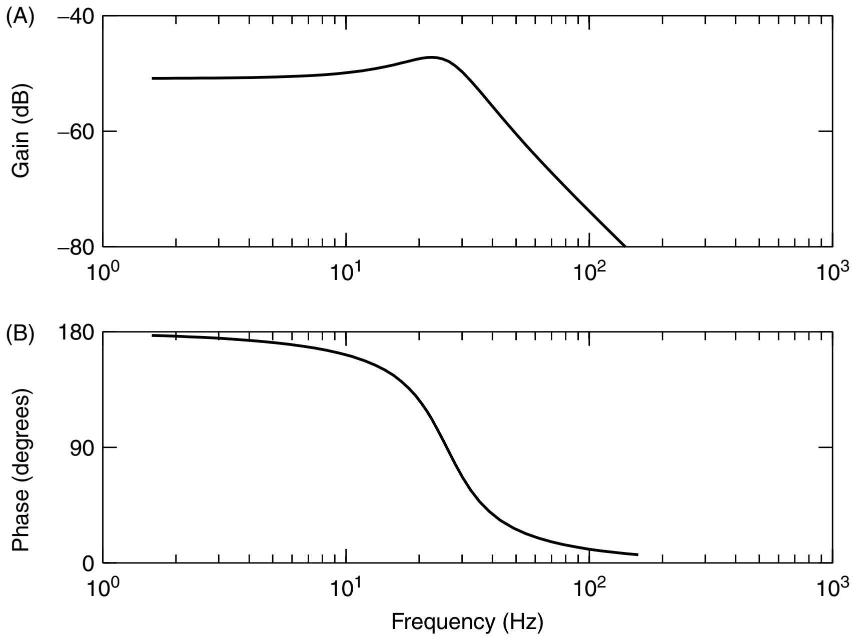

Figure 3.4 Frequency response of the dynamic compliance of the human ankle. (A) Transfer function magnitude $|H(j\omega)|$. (B) Transfer function phase $\phi(H(j\omega))$. The system's impulse response is shown in Figure 3.3.

The phase plot shows that at low frequencies the output has the opposite sign to the input. In contrast, at higher frequencies, the output is in phase with the input, although greatly attenuated.

This behavior is consistent with a physical interpretation of a system having elastic, viscous, and inertial properties.

3.2.2.2 Relationship to the Impulse Response Consider the Fourier transform (3.11) of the convolution integral (3.6),

$$\begin{aligned}
Y(j\omega) &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(\tau)u(t-\tau)\,d\tau \right) e^{-j\omega t}\,dt \\
&= \int_{-\infty}^{\infty} h(\tau) \int_{-\infty}^{\infty} u(t-\tau)e^{-j\omega t}\,dt\,d\tau \\
&= U(j\omega) \int_{-\infty}^{\infty} h(\tau)e^{-j\omega\tau}\,d\tau \\
&= U(j\omega)H(j\omega)
\end{aligned}$$

where $H(j\omega)$ is the Fourier transform of the impulse response, $h(\tau)$.

This relation is identical to equation (3.12), demonstrating that the frequency response, $H(j\omega)$, is equal to the Fourier Transform of the impulse response, $h(\tau)$. Conversely, the inverse Fourier transform of the frequency response is equal to the impulse response.

3.3 PARAMETRIC MODELS

Any causal, linear, time-invariant, continuous-time system can be described by a differential equation with constant coefficients of the form

$$\boxed{\begin{aligned}\frac{d^n y(t)}{dt^n} &+ a_{n-1}\frac{d^{n-1}y(t)}{dt^{n-1}} + \cdots + a_1\frac{dy(t)}{dt} + a_0 y(t)\\ &= b_m\frac{d^m u(t)}{dt^m} + b_{m-1}\frac{d^{m-1}u(t)}{dt^{m-1}} + \cdots + b_1\frac{du(t)}{dt} + b_0 u(t) \quad (3.13)\end{aligned}}$$

where $n \geq m$. Equation (3.13) is often abbreviated as

$$A(D)y = B(D)u$$

where A and B are polynomials in the differential operator, $D = d/dt$.

3.3.1 Parametric Frequency Domain Models

Taking the Laplace transform of (3.13) gives

$$\begin{aligned}(s^n + a_{n-1}s^{n-1} &+ \cdots + a_1 s + a_0)Y(s)\\ &= (b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0)U(s)\end{aligned}$$

which may be manipulated to obtain the transfer function,

$$H(s) = \frac{Y(s)}{U(s)}$$

$$\boxed{H(s) = \frac{b_m s^m + \cdots + b_1 s + b_0}{s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0} = \frac{B(s)}{A(s)} \qquad (3.14)}$$

Note that equation (3.14) can be factored to give

$$H(s) = K\frac{(s - z_1)\ldots(s - z_m)}{(s - p_1)\ldots(s - p_n)} \tag{3.15}$$

where the zeros (z_i) and poles (p_i) of the transfer function may be real, zero, or complex (if complex they come as conjugate pairs). Furthermore, for a physically realizable system, $n \geq m$, that is, the system cannot have more zeros than poles.

It is straightforward to convert a transfer function into the equivalent nonparametric model. Thus, the impulse response function is determined from $Y(s) = H(s)U(s)$, by setting $U(s) = \mathcal{L}(\delta) = 1$ and taking the inverse Laplace transform. The frequency response may be determined by setting $U(s) = \omega/(s^2 + \omega^2)$, the Laplace transform

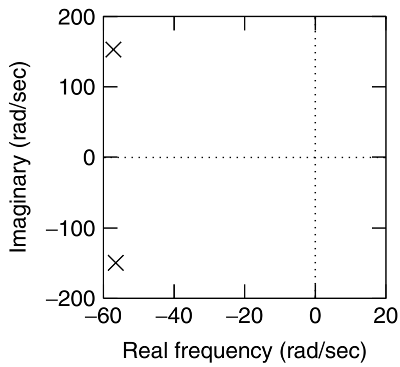

Figure 3.5 Pole-zero map of a continuous-time, parametric model of the human ankle compliance. Notice that the system has two poles and no zeros.

of $u(t) = \sin(\omega t)$, and then taking the inverse Laplace transform. Operationally, this corresponds to making the substitution $j\omega = s$ in the transfer function.

3.3.1.1 Example: Human Ankle Compliance — Transfer Function Figure 3.5 shows a pole-zero map of a parametric model of the human ankle compliance. The transfer function is given by

$$H(s) = \frac{\Theta(s)}{T(s)} \approx \frac{-76}{s^2 + 114s + 26{,}700}$$

This model has the same form as that of a mass, spring and damper

$$H(s) = \frac{\Theta(s)}{T(s)} \approx \frac{1}{Is^2 + Bs + K}$$

Consequently, it is possible to obtain some physical insight from the parameter values. For example, a second-order transfer function is often written as

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

where ω_n is the undamped natural frequency of the system, G is its DC gain, and ζ is the damping factor. For this model we have

$$\begin{aligned} \omega_n &= 163.4 \text{ rad/s} = 26 \text{ Hz} \\ G &= -0.0028 \\ \zeta &= 0.35 \end{aligned}$$

In fact, this model was constructed using values for these parameters obtained from a review on human joint dynamics (Kearney and Hunter, 1990). This transfer function was used to generate all the compliance models presented in this chapter.

3.3.2 Discrete-Time Parametric Models

Linear systems may also be modeled in discrete time using difference equations. In this case, the fundamental operators are the forward and backward shift operators, defined as*

$$zu(t) = u(t+1) \tag{3.16}$$

$$z^{-1}u(t) = u(t-1) \tag{3.17}$$

The forward difference is defined as

$$[z-1]u(t) = zu(t) - u(t) = u(t+1) - u(t)$$

Even though there is an analogy between the derivative and the forward difference, it is the backward shift operator, z^{-1}, that is most commonly employed in difference equation models. Thus, by analogy to equation (3.13), any deterministic, linear, time-invariant, discrete-time system can be modeled using a difference equation of the form

$$\boxed{A(z^{-1})y(t) = B(z^{-1})u(t) \tag{3.18}}$$

It must be remembered that the transformation from discrete to continuous time is *not* achieved by simply replacing derivatives with forward differences. A variety of more complex transformations are used including the bilinear transform, Padé approximations, and the impulse invariant transform. These transformations differ chiefly in the assumptions made regarding the behavior of the signals between sampling instants. For example, a zero-order hold assumes that the signals are "staircases" that remain constant between samples. This is equivalent to using rectangular (Euler) integration. In addition, a discrete-time model's parameters will depend on the sampling rate; as it increases there will be less change between samples and so the backward differences will be smaller. The interested reader should consult a text on digital signal processing, such as Oppenheim and Schafer (1989), for a more detailed treatment of these issues.

Adding a white noise component, $w(t)$, to the output of a difference equation model (3.18) gives the *output error* (OE) model.

$$A(z^{-1})y(t) = B(z^{-1})u(t)$$
$$z(t) = y(t) + w(t)$$

This is usually written more compactly as

$$\boxed{z(t) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + w(t) \tag{3.19}}$$

The difference equation (3.18) and output error (3.19) models are sometimes referred to as autoregressive moving average (ARMA) models, or as ARMA models with additive

*Note that in this context, z and z^{-1} are operators that modify signals. Throughout this text, the symbol $z(t)$ is used in both discrete and continuous time to denote the measured output of a system. The context should make it apparent which meaning is intended.

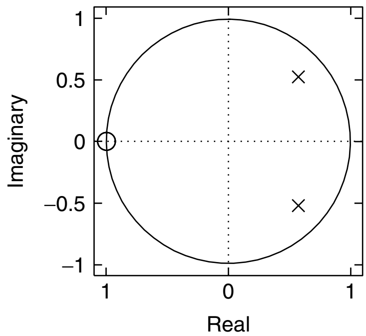

Figure 3.6 Pole-zero map of the discrete-time, parametric frequency domain representation dynamic ankle compliance, at a sampling frequency of 200 Hz. Note the presence of two zeros at $z = -1$.

noise in the case of an OE model. However, the term ARMA model is more correctly used to describe the stochastic signal model described in the next section.

Finally, note that if $A(z^{-1}) = 1$, the difference equation reduces to a finite impulse response (FIR) filter,

$$\begin{aligned} y(t) &= B(z^{-1})u(t) \\ &= b_0u(t) + b_1u(t-1) + \cdots + b_mu(t-m) \end{aligned} \tag{3.20}$$

3.3.2.1 Example: Human Ankle Compliance — Discrete-Time Transfer Function A z-domain transfer function description of human ankle compliance is given by

$$H(z) = \frac{\Theta(z)}{T(z)} \approx \frac{(-3.28z^2 - 6.56z - 3.28) \times 10^{-4}}{z^2 - 1.15z + 0.606}$$

for a sampling frequency of 200 Hz. Figure 3.6 shows the corresponding pole-zero map.

3.3.2.2 Parametric Models of Stochastic Signals Linear difference equations may also be used to model stochastic signals as the outputs of linear dynamic systems driven by unobserved white Gaussian noise sequences. The simplest of these models is the autoregressive (AR) model,

$$\boxed{A(z^{-1})y(t) = w(t)} \tag{3.21}$$

where $w(t)$ is filtered by the all-pole filter, $1/A(z^{-1})$. In a systems context, $w(t)$ would be termed a process disturbance since it excites the dynamics of a system. The origin of the term autoregressive can best be understood by examining the difference equation of an AR model:

$$y(t) = w(t) - a_1y(t-1) - \cdots - a_ny(t-n)$$

which shows that the current value of the output depends on the current noise sample, $w(t)$, and n past output values.

A more general signal description extends the filter to include both poles and zeros,

$$A(z^{-1})y(t) = C(z^{-1})w(t) \tag{3.22}$$

This structure is known as an autoregressive moving average (ARMA) model and has the same form as the deterministic difference equation, (3.18). However, in this case the input is an unmeasured, white-noise, process disturbance so the ARMA representation is a stochastic model of the output signal. In contrast, equation (3.18) describes the deterministic relationship between measured input and output signals.

3.3.2.3 Parametric Models of Stochastic Systems Parametric discrete-time models can also represent systems having both stochastic and deterministic components. For example, the autoregressive exogenous input (ARX) model given by

$$A(z^{-1})y(t) = B(z^{-1})u(t) + w(t) \tag{3.23}$$

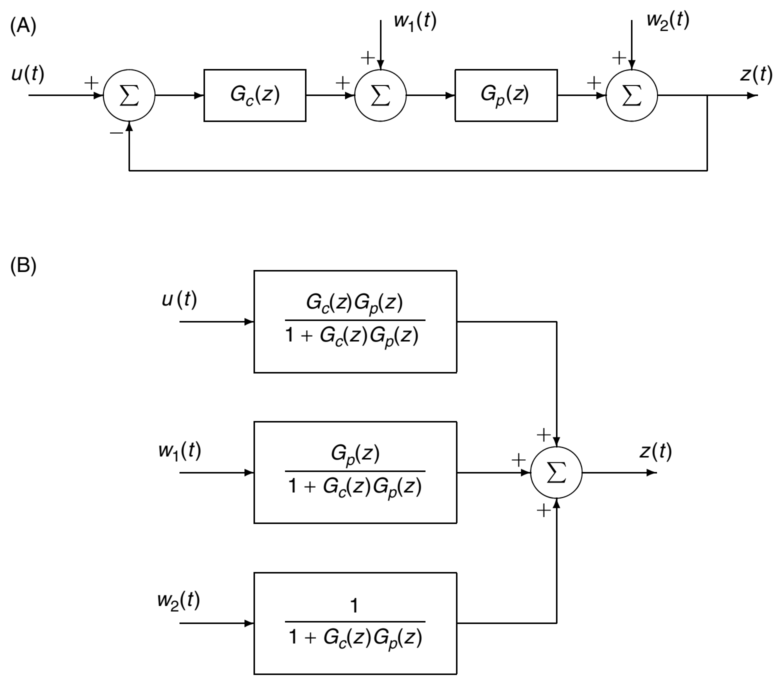

Figure 3.7 Two equivalent representations of a closed-loop control system with controlled input, $u(t)$, and two process disturbances, $w_1(t)$ and $w_2(t)$. (A) Block diagram explicitly showing the feedback loop including the plant $G_p(z)$ and controller $G_c(z)$. (B) Equivalent representation comprising three open-loop transfer functions, one for each input. Note that the three open-loop transfer functions all share the same denominator.

incorporates terms accounting for both a measured (exogenous) input, $u(t)$, and an unobserved process disturbance, $w(t)$. Consequently, its output will contain both deterministic and stochastic components.

The ARMA model can be generalized in a similar manner to give the ARMAX model,

$$\boxed{A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})w(t) \qquad (3.24)}$$

which can be written as

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + \frac{C(z^{-1})}{A(z^{-1})}w(t)$$

This makes it evident that the deterministic input, $u(t)$, and the noise, $w(t)$, are filtered by the same dynamics, the roots of $A(z^{-1})$. This is appropriate if the noise is a process disturbance. For example, consider the feedback control system illustrated in Figure 3.7. Regardless of where the disturbance (or control signal) enters the closed-loop system, the denominator of the transfer function will be $1+G_c(z)G_p(z)$. Thus, both the deterministic and noise models will have the same denominators, and an ARX or ARMAX model structure will be appropriate.

However, if the noise source is outside the feedback loop, or if the process generating the disturbance input contains additional poles not found in the closed-loop dynamics, the ARMAX structure is not appropriate. The more general Box–Jenkins structure

$$\boxed{y(t) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + \frac{C(z^{-1})}{D(z^{-1})}w(t) \qquad (3.25)}$$

addresses this problem. Here the deterministic and stochastic inputs are filtered by separate dynamics, so the effects of process and measurement noise may be combined into the single term, $w(t)$. The Box–Jenkins model is the most general parametric linear system model; all other linear parametric models are special cases of the Box–Jenkins model. This is illustrated in Figure 3.8 as follows:

1. The output error model Figure (3.8B) is obtained by removing the noise filter from the Box–Jenkins model, so that $C(z^{-1}) = D(z^{-1}) = 1$.
2. The ARMAX model Figure (3.8C), is obtained by equating the denominator polynomials in the deterministic and noise models (i.e., $A(z^{-1}) = D(z^{-1})$).
3. Setting the numerator of the noise model to 1, $C(z^{-1}) = 1$, reduces the ARMAX model to an ARX structure Figure (3.8D).
4. The ARMA model (Figure 3.8E) can be thought of as a Box–Jenkins (or ARMAX) structure with no deterministic component.
5. Setting the numerator of the noise model to 1, $C(z^{-1}) = 1$, reduces the ARMA model to an AR model Figure (3.8F)

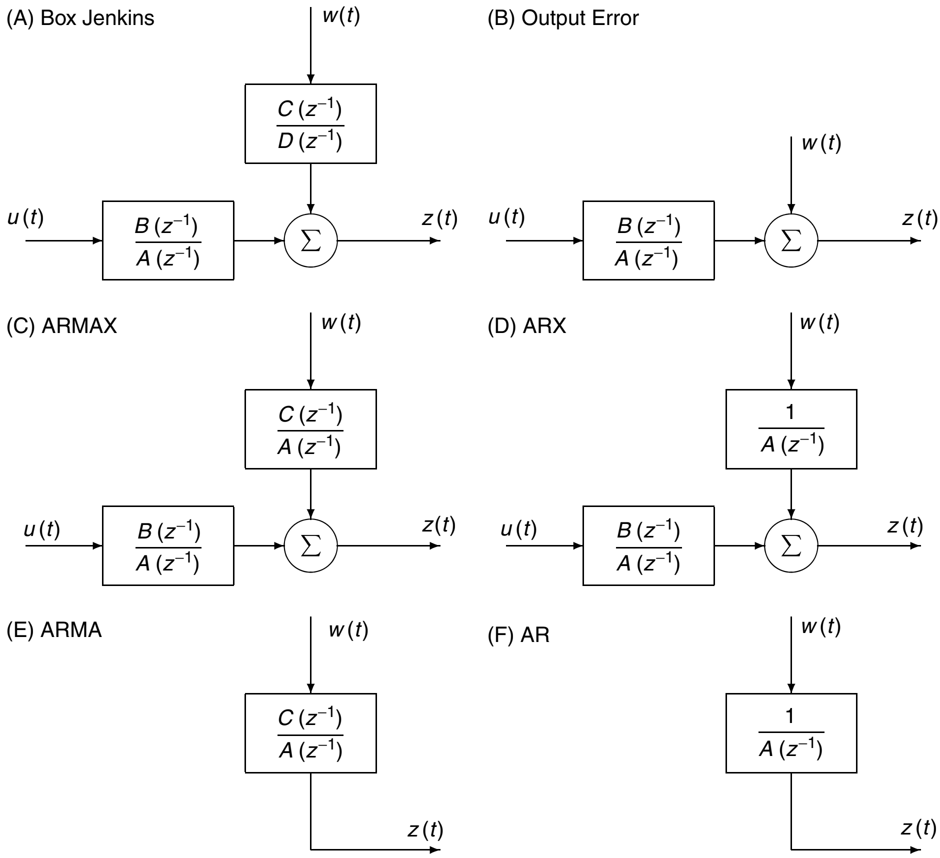

Figure 3.8 Taxonomy of the different parametric difference equation-type linear system models. The input, $u(t)$, and output, $z(t)$, are assumed to be measured signals. The disturbance, $w(t)$, is an unobserved white noise process. (A) The Box–Jenkins model is the most general structure. (B–F) Models that are special cases of the Box–Jenkins structure.

3.4 STATE-SPACE MODELS

An nth-order differential equation describing a linear time-invariant system (3.13) can also be expressed as a system of n coupled first-order differential equations. These are expressed conveniently in matrix form as

$$\boxed{\begin{aligned} \dot{\mathbf{x}}(\mathbf{t}) &= \mathbf{A}\mathbf{x}(\mathbf{t}) + \mathbf{B}\mathbf{u}(\mathbf{t}) \\ \mathbf{y}(\mathbf{t}) &= \mathbf{C}\mathbf{x}(\mathbf{t}) + \mathbf{D}\mathbf{u}(\mathbf{t}) \end{aligned} \qquad (3.26)}$$

This representation is known as a state-space model where the state of the system is defined by an n element vector, $\mathbf{x}(\mathbf{t})$. The $n \times n$ matrix $\mathbf{A}$ maps the n-dimensional state onto its time derivative. For a single-input single-output (SISO) system, $\mathbf{B}$ is an $n \times 1$ (column) vector that maps the input onto the derivative of the n-dimensional state.

Similarly, $\mathbf{C}$ will be a $1 \times n$ (row) vector, and D will be a scalar. Multiple-input multiple-output (MIMO) systems may be modeled with the same structure simply by increasing the dimensions of the matrices $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$. The matrices $\mathbf{B}$ and $\mathbf{D}$ will have one column per input, while $\mathbf{C}$ and $\mathbf{D}$ will have one row per output.

In discrete time, the state-space model becomes a set of n coupled difference equations.

$$\boxed{\begin{aligned} \mathbf{x(t+1)} &= \mathbf{Ax(t) + Bu(t)} \\ \mathbf{y(t)} &= \mathbf{Cx(t) + Du(t)} \end{aligned} \qquad (3.27)}$$

where $\mathbf{x(t)}$ is an n-dimensional vector, and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are time-invariant matrices of appropriate dimensions.

State-space models may be extended to include the effects of both process and measurement noise, as follows:

$$\begin{aligned} \mathbf{x(t+1)} &= \mathbf{Ax(t) + Bu(t) + w(t)} \\ \mathbf{y(t)} &= \mathbf{Cx(t) + Du(t) + v(t)} \end{aligned} \tag{3.28}$$

Note that the terms representing the process disturbance, $\mathbf{w(t)}$, and measurement noise, $\mathbf{v(t)}$, remain distinct. This contrasts with the Box–Jenkins model, where they are combined into a single disturbance term.

The impulse response of a discrete-time, state-space model may be generated by setting the initial state, $\mathbf{x(0)}$, to zero, applying a discrete impulse input, and solving equations (3.27) for successive values of t. This gives

$$\mathbf{h} = \begin{bmatrix} \mathbf{D} & \mathbf{CB} & \mathbf{CAB} & \mathbf{CA^2B} & \dots \end{bmatrix}^T \tag{3.29}$$

where $\mathbf{h}$ is a vector containing the impulse response, $h(\tau)$.

Now, consider the effects of transforming the matrices of the state-space system as follows

$$\begin{aligned} \mathbf{A_T} &= \mathbf{T^{-1}AT}, & \mathbf{B_T} &= \mathbf{T^{-1}B} \\ \mathbf{C_T} &= \mathbf{CT}, & \mathbf{D_T} &= \mathbf{D} \end{aligned} \tag{3.30}$$

where $\mathbf{T}$ is an invertible matrix. The impulse response of this transformed system will be

$$\begin{aligned} h_T(k) &= \begin{bmatrix} \mathbf{D_T} & \mathbf{C_TB_T} & \mathbf{C_TA_TB_T} & \dots \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{D} & \mathbf{CTT^{-1}B} & \mathbf{CTT^{-1}ATT^{-1}B} & \dots \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{D} & \mathbf{CB} & \mathbf{CAB} & \dots \end{bmatrix}^T \end{aligned}$$

which is the same as the original system. Hence, $\mathbf{T}$ is a similarity transformation that does not alter the input–output behavior of the system. Thus, a system's state-space model will have many different realizations with identical input–output behaviors. Consequently, it is possible to choose the realization that best suits a particular problem. One approach is to seek a *balanced* realization where all states have the same order of magnitude; this is best suited to fixed precision arithmetic applications. An alternative approach is to seek

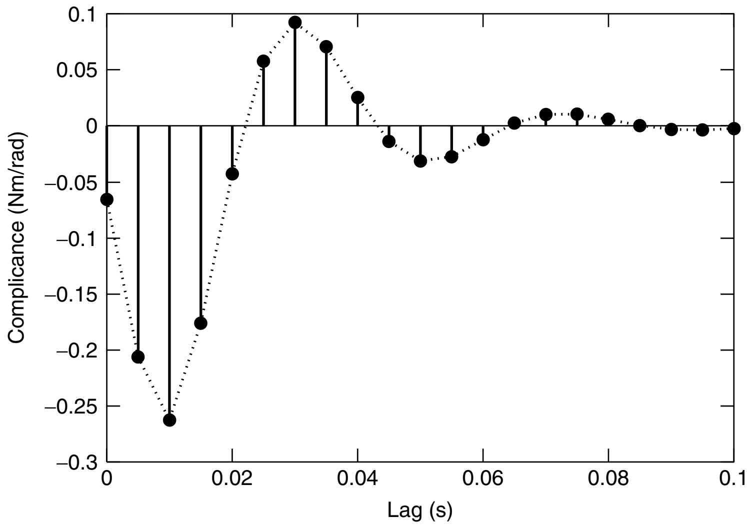

Figure 3.9 Impulse response computed from the discrete state-space model of the dynamic compliance of the human ankle.

a *minimal* realization, one that minimizes the number of free parameters; this is most efficient for computation since it minimizes the number of nonzero matrix elements.

3.4.1 Example: Human Ankle Compliance — Discrete-Time, State-Space Model

One discrete-time, state-space realization of the ankle-compliance model, considered throughout this chapter, is

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1.15 & -0.606 \\ 1 & 0 \end{bmatrix}, & \mathbf{B} &= \begin{bmatrix} 0.0313 \\ 0 \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} -0.0330 & -0.00413 \end{bmatrix}, & \mathbf{D} &= -3.28 \times 10^{-4} \end{aligned} \tag{3.31}$$

In this case $x_2(t) = x_1(t-1)$, and the input drives only the first state, $x_1(t)$. Thus, in this realization the states act as a sort of delay line.

Figure 3.9 shows the discrete impulse response obtained from the state-space realization shown above. Compare this with the continuous time IRF shown in Figure 3.3. Notice that the continuous-time IRF is smoother than the discrete-time version. This jagged appearance is largely due to the behavior of the IRF between sample points.

3.5 NOTES AND REFERENCES

1. Bendat and Piersol (1986) is a good reference for nonparametric models of linear systems, in both the time and frequency domains.

2. The relationships between continuous- and discrete-time signals and systems are dealt with in texts on digital signal processing, such as Oppenheim and Schafer (1989), and on digital control systems, such as Ogata (1995).
3. There are several texts dealing with discrete-time, stochastic, parametric models. Ljung (1999), and Söderström and Stoica (1989) in particular are recommended.
4. For more information concerning state-space systems, the reader should consult Kailath (1980).

3.6 THEORETICAL PROBLEMS

1. Suppose that the output of system is given by

$$y(t) = \int_{\tau=0}^{T} h(t,\tau)u(t-\tau)\,d\tau$$

- Is this system linear?
- Is this system causal?
- Is it time-invariant?

Justify your answers.

2. Compute and plot the frequency response of the following transfer function,

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

What does this system do? Which frequencies does it pass, which does it stop? Compute its complex conjugate, $H^*(s)$. What is special about $H(j\omega)H^*(j\omega)$?

3. Draw a pole-zero map for the deterministic, discrete-time parametric model,

$$y(t) = u(t) - 0.25u(t-1) + 0.5y(t-1)$$

Next, compute its impulse response. How long does it take for the impulse response to decay to 10% of its peak value? How long does it take before it reaches less than 1%? How many lags would be required for a reasonable FIR approximation to this system?

4. Consider the following state-space system

$$\begin{aligned}
\mathbf{x(t+1)} &= \mathbf{Ax(t)} + \mathbf{Bu(t)}\\
\mathbf{w(t-1)} &= \mathbf{Gw(t)} + \mathbf{Hu(t)}\\
\mathbf{y(t)} &= \mathbf{Cx(t)} + \mathbf{Ew(t)} + \mathbf{Du(t)}
\end{aligned}$$

Is this system linear? time-invariant? causal? deterministic? Compute the system's response to a unit impulse. Can you express the output as a convolution?

3.7 COMPUTER EXERCISES

1. Create a linear system object. Check superposition. Is the system causal?

2. Transform a linear system object from TF to IRF to SS.

3. Excite a linear system with a cosine—look at the first few points. Why isn't the response a perfect sinusoid?

4. Generate two linear system objects, G and H. Compute $G(H(u))$ and $H(G(u))$. Are they equivalent. Explain why/why not.