**Chapter Two**

**System Modeling**

... I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, “How many arbitrary parameters did you use for your calculations?” I thought for a moment about our cut-off procedures and said, “Four.” He said, “I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk.”

Freeman Dyson on describing the predictions of his model for meson-proton scattering to Enrico Fermi in 1953 [Dys04].

A model is a precise representation of a system’s dynamics used to answer questions via analysis and simulation. The model we choose depends on the questions we wish to answer, and so there may be multiple models for a single dynamical system, with different levels of fidelity depending on the phenomena of interest. In this chapter we provide an introduction to the concept of modeling and present some basic material on two specific methods commonly used in feedback and control systems: differential equations and difference equations.

### 2.1 Modeling Concepts

A *model* is a mathematical representation of a physical, biological or information system. Models allow us to reason about a system and make predictions about how a system will behave. In this text, we will mainly be interested in models of dynamical systems describing the input/output behavior of systems, and we will often work in “state space” form.

Roughly speaking, a dynamical system is one in which the effects of actions do not occur immediately. For example, the velocity of a car does not change immediately when the gas pedal is pushed nor does the temperature in a room rise instantaneously when a heater is switched on. Similarly, a headache does not vanish right after an aspirin is taken, requiring time for it to take effect. In business systems, increased funding for a development project does not increase revenues in the short term, although it may do so in the long term (if it was a good investment). All of these are examples of dynamical systems, in which the behavior of the system evolves with time.

In the remainder of this section we provide an overview of some of the key concepts in modeling. The mathematical details introduced here are explored more fully in the remainder of the chapter.
The Heritage of Mechanics

The study of dynamics originated in attempts to describe planetary motion. The basis was detailed observations of the planets by Tycho Brahe and the results of Kepler, who found empirically that the orbits of the planets could be well described by ellipses. Newton embarked on an ambitious program to try to explain why the planets move in ellipses, and he found that the motion could be explained by his law of gravitation and the formula stating that force equals mass times acceleration. In the process he also invented calculus and differential equations.

One of the triumphs of Newton’s mechanics was the observation that the motion of the planets could be predicted based on the current positions and velocities of all planets. It was not necessary to know the past motion. The state of a dynamical system is a collection of variables that completely characterizes the motion of a system for the purpose of predicting future motion. For a system of planets the state is simply the positions and the velocities of the planets. We call the set of all possible states the state space.

A common class of mathematical models for dynamical systems is ordinary differential equations (ODEs). In mechanics, one of the simplest such differential equations is that of a spring–mass system with damping:

\[ m\ddot{q} + c(\dot{q}) + kq = 0. \]  

This system is illustrated in Figure 2.1. The variable \( q \in \mathbb{R} \) represents the position of the mass \( m \) with respect to its rest position. We use the notation \( \dot{q} \) to denote the derivative of \( q \) with respect to time (i.e., the velocity of the mass) and \( \ddot{q} \) to represent the second derivative (acceleration). The spring is assumed to satisfy Hooke’s law, which says that the force is proportional to the displacement. The friction element (damper) is taken as a nonlinear function \( c(\dot{q}) \), which can model effects such as stiction and viscous drag. The position \( q \) and velocity \( \dot{q} \) represent the instantaneous state of the system. We say that this system is a second-order system since the dynamics depend on the first two derivatives of \( q \).

The evolution of the position and velocity can be described using either a time plot or a phase portrait, both of which are shown in Figure 2.2. The time plot, on
2.1. MODELING CONCEPTS

The Heritage of Electrical Engineering

A different view of dynamics emerged from electrical engineering, where the design of electronic amplifiers led to a focus on input/output behavior. A system was considered a device that transforms inputs to outputs, as illustrated in Figure 2.3. Conceptually an input/output model can be viewed as a giant table of inputs and
outputs. Given an input signal $u(t)$ over some interval of time, the model should produce the resulting output $y(t)$.

The input/output framework is used in many engineering disciplines since it allows us to decompose a system into individual components connected through their inputs and outputs. Thus, we can take a complicated system such as a radio or a television and break it down into manageable pieces such as the receiver, demodulator, amplifier and speakers. Each of these pieces has a set of inputs and outputs and, through proper design, these components can be interconnected to form the entire system.

The input/output view is particularly useful for the special class of linear time-invariant systems. This term will be defined more carefully later in this chapter, but roughly speaking a system is linear if the superposition (addition) of two inputs yields an output that is the sum of the outputs that would correspond to individual inputs being applied separately. A system is time-invariant if the output response for a given input does not depend on when that input is applied.

Many electrical engineering systems can be modeled by linear time-invariant systems, and hence a large number of tools have been developed to analyze them. One such tool is the step response, which describes the relationship between an input that changes from zero to a constant value abruptly (a step input) and the corresponding output. As we shall see later in the text, the step response is very useful in characterizing the performance of a dynamical system, and it is often used to specify the desired dynamics. A sample step response is shown in Figure 2.4a.

Another way to describe a linear time-invariant system is to represent it by its response to sinusoidal input signals. This is called the frequency response, and a rich, powerful theory with many concepts and strong, useful results has emerged. The results are based on the theory of complex variables and Laplace transforms. The basic idea behind frequency response is that we can completely characterize the behavior of a system by its steady-state response to sinusoidal inputs. Roughly
speaking, this is done by decomposing any arbitrary signal into a linear combination of sinusoids (e.g., by using
the Fourier transform) and then using linearity to compute the output by combining the response to the individual
frequencies. A sample frequency response is shown in Figure 2.4b.

The input/output view lends itself naturally to experimental determination of system dynamics, where a system is characterized by recording its response to particular inputs, e.g., a step or a set of sinusoids over a range of frequencies.

The Control View

When control theory emerged as a discipline in the 1940s, the approach to dynamics was strongly influenced by the electrical engineering (input/output) view. A second wave of developments in control, starting in the late 1950s, was inspired by mechanics, where the state space perspective was used. The emergence of space flight is a typical example, where precise control of the orbit of a spacecraft is essential. These two points of view gradually merged into what is today the state space representation of input/output systems.

The development of state space models involved modifying the models from mechanics to include external actuators and sensors and utilizing more general forms of equations. In control, the model given by equation (2.2) was replaced by

\[ \frac{dx}{dt} = f(x, u), \quad y = h(x, u), \]  

where \( x \) is a vector of state variables, \( u \) is a vector of control signals and \( y \) is a vector of measurements. The term \( dx/dt \) represents the derivative of \( x \) with respect to time, now considered a vector, and \( f \) and \( h \) are (possibly nonlinear) mappings of their arguments to vectors of the appropriate dimension. For mechanical systems, the state consists of the position and velocity of the system, so that \( x = (q, \dot{q}) \) in the case of a damped spring–mass system. Note that in the control formulation we model
dynamics as first-order differential equations, but we will see that this can capture
the dynamics of higher-order differential equations by appropriate definition of the
state and the maps \( f \) and \( h \).

Adding inputs and outputs has increased the richness of the classical problems
and led to many new concepts. For example, it is natural to ask if possible states \( x \)
can be reached with the proper choice of \( u \) (reachability) and if the measurement \( y \)
contains enough information to reconstruct the state (observability). These topics
will be addressed in greater detail in Chapters 6 and 7.

A final development in building the control point of view was the emergence of
disturbances and model uncertainty as critical elements in the theory. The simple
way of modeling disturbances as deterministic signals like steps and sinusoids has
the drawback that such signals cannot be predicted precisely. A more realistic ap-
proach is to model disturbances as random signals. This viewpoint gives a natural
connection between prediction and control. The dual views of input/output repre-
sentations and state space representations are particularly useful when modeling
uncertainty since state models are convenient to describe a nominal model but un-
certainties are easier to describe using input/output models (often via a frequency
response description). Uncertainty will be a constant theme throughout the text and
will be studied in particular detail in Chapter 12.

An interesting observation in the design of control systems is that feedback
systems can often be analyzed and designed based on comparatively simple models.
The reason for this is the inherent robustness of feedback systems. However, other
uses of models may require more complexity and more accuracy. One example is
feedforward control strategies, where one uses a model to precompute the inputs
that cause the system to respond in a certain way. Another area is system validation,
where one wishes to verify that the detailed response of the system performs as it
was designed. Because of these different uses of models, it is common to use a
hierarchy of models having different complexity and fidelity.

Multidomain Modeling

Modeling is an essential element of many disciplines, but traditions and methods
from individual disciplines can differ from each other, as illustrated by the previous
discussion of mechanical and electrical engineering. A difficulty in systems engi-
eering is that it is frequently necessary to deal with heterogeneous systems from
many different domains, including chemical, electrical, mechanical and informa-
tion systems.

To model such multidomain systems, we start by partitioning a system into
smaller subsystems. Each subsystem is represented by balance equations for mass,
energy and momentum, or by appropriate descriptions of information processing
in the subsystem. The behavior at the interfaces is captured by describing how the
variables of the subsystem behave when the subsystems are interconnected. These
interfaces act by constraining variables within the individual subsystems to be equal
(such as mass, energy or momentum fluxes). The complete model is then obtained
by combining the descriptions of the subsystems and the interfaces.
2.1. MODELING CONCEPTS

Using this methodology it is possible to build up libraries of subsystems that correspond to physical, chemical and informational components. The procedure mimics the engineering approach where systems are built from subsystems that are themselves built from smaller components. As experience is gained, the components and their interfaces can be standardized and collected in model libraries. In practice, it takes several iterations to obtain a good library that can be reused for many applications.

State models or ordinary differential equations are not suitable for component-based modeling of this form because states may disappear when components are connected. This implies that the internal description of a component may change when it is connected to other components. As an illustration we consider two capacitors in an electrical circuit. Each capacitor has a state corresponding to the voltage across the capacitors, but one of the states will disappear if the capacitors are connected in parallel. A similar situation happens with two rotating inertias, each of which is individually modeled using the angle of rotation and the angular velocity. Two states will disappear when the inertias are joined by a rigid shaft.

This difficulty can be avoided by replacing differential equations by differential algebraic equations, which have the form

\[ F(z, \dot{z}) = 0, \]

where \( z \in \mathbb{R}^n \). A simple special case is

\[ \dot{x} = f(x, y), \quad g(x, y) = 0, \quad (2.4) \]

where \( z = (x, y) \) and \( F = (\dot{x} - f(x, y), g(x, y)) \). The key property is that the derivative \( \dot{z} \) is not given explicitly and there may be pure algebraic relations between the components of the vector \( z \).

The model (2.4) captures the examples of the parallel capacitors and the linked rotating inertias. For example, when two capacitors are connected, we simply add the algebraic equation expressing that the voltages across the capacitors are the same.

Modelica is a language that has been developed to support component-based modeling. Differential algebraic equations are used as the basic description, and object-oriented programming is used to structure the models. Modelica is used to model the dynamics of technical systems in domains such as mechanical, electrical, thermal, hydraulic, thermofluid and control subsystems. Modelica is intended to serve as a standard format so that models arising in different domains can be exchanged between tools and users. A large set of free and commercial Modelica component libraries are available and are used by a growing number of people in industry, research and academia. For further information about Modelica, see http://www.modelica.org or Tiller [Til01].
2.2 State Space Models

In this section we introduce the two primary forms of models that we use in this text: differential equations and difference equations. Both make use of the notions of state, inputs, outputs and dynamics to describe the behavior of a system.

**Ordinary Differential Equations**

The state of a system is a collection of variables that summarize the past of a system for the purpose of predicting the future. For a physical system the state is composed of the variables required to account for storage of mass, momentum and energy. A key issue in modeling is to decide how accurately this storage has to be represented. The state variables are gathered in a vector $x \in \mathbb{R}^n$ called the state vector. The control variables are represented by another vector $u \in \mathbb{R}^p$, and the measured signal by the vector $y \in \mathbb{R}^q$. A system can then be represented by the differential equation

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

(2.5)

where $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ and $h : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^q$ are smooth mappings. We call this model a state space model.

The dimension of the state vector is called the order of the system. The system (2.5) is called time-invariant because the functions $f$ and $h$ do not depend explicitly on time $t$; there are more general time-varying systems where the functions do depend on time. The model consists of two functions: the function $f$ gives the rate of change of the state vector as a function of state $x$ and control $u$, and the function $h$ gives the measured values as functions of state $x$ and control $u$.

A system is called a linear state space system if the functions $f$ and $h$ are linear in $x$ and $u$. A linear state space system can thus be represented by

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

(2.6)

where $A$, $B$, $C$ and $D$ are constant matrices. Such a system is said to be linear and time-invariant, or LTI for short. The matrix $A$ is called the dynamics matrix, the matrix $B$ is called the control matrix, the matrix $C$ is called the sensor matrix and the matrix $D$ is called the direct term. Frequently systems will not have a direct term, indicating that the control signal does not influence the output directly.

A different form of linear differential equations, generalizing the second-order dynamics from mechanics, is an equation of the form

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

(2.7)

where $t$ is the independent (time) variable, $y(t)$ is the dependent (output) variable and $u(t)$ is the input. The notation $d^k y/dt^k$ is used to denote the $k$th derivative of $y$ with respect to $t$, sometimes also written as $y^{(k)}$. The controlled differential equation (2.7) is said to be an $n$th-order system. This system can be converted into
state space form by defining

\[
\begin{pmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_{n-1} \\
    x_n
\end{pmatrix}
= 
\begin{pmatrix}
    d^{n-1}y/dt^{n-1} \\
    d^{n-2}y/dt^{n-2} \\
    \vdots \\
    dy/dt \\
y
\end{pmatrix},
\]

and the state space equations become

\[
\frac{d}{dt}
\begin{pmatrix}
    x_1 \\
    x_2 \\
    \vdots \\
    x_{n-1} \\
    x_n
\end{pmatrix}
= 
\begin{pmatrix}
    -a_1x_1 - \cdots - a_nx_n \\
    x_1 \\
    \vdots \\
    x_{n-2} \\
    x_{n-1}
\end{pmatrix}
+ 
\begin{pmatrix}
    u \\
    0 \\
    \vdots \\
    0 \\
    0
\end{pmatrix},
\]

\[y = x_n.\]

With the appropriate definitions of \(A, B, C\) and \(D\), this equation is in linear state space form. An even more general system is obtained by letting the output be a linear combination of the states of the system, i.e.,

\[y = b_1x_1 + b_2x_2 + \cdots + b_nx_n + du.\]

This system can be modeled in state space as

\[
\frac{d}{dt}
\begin{pmatrix}
    x_1 \\
    x_2 \\
    x_3 \\
    \vdots \\
    x_n
\end{pmatrix}
= 
\begin{pmatrix}
    -a_1 & -a_2 & \cdots & -a_{n-1} & -a_n \\
    1 & 0 & \cdots & 0 & 0 \\
    0 & 1 & \cdots & 0 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    0 & 0 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
    x \\
    0 \\
    0 \\
    \vdots \\
    0
\end{pmatrix}
+ 
\begin{pmatrix}
    1 \\
    0 \\
    0 \\
    \vdots \\
    0
\end{pmatrix}u,
\]

\[y = \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix} x + du.\]

This particular form of a linear state space system is called reachable canonical form and will be studied in more detail in later chapters.

**Example 2.1 Balance systems**

An example of a type of system that can be modeled using ordinary differential equations is the class of balance systems. A balance system is a mechanical system in which the center of mass is balanced above a pivot point. Some common examples of balance systems are shown in Figure 2.5. The Segway® Personal Transporter (Figure 2.5a) uses a motorized platform to stabilize a person standing on top of it. When the rider leans forward, the transportation device propels itself along the ground but maintains its upright position. Another example is a rocket (Figure 2.5b), in which a gimbaled nozzle at the bottom of the rocket is used to stabilize the body of the rocket above it. Other examples of balance systems include humans or other animals standing upright or a person balancing a stick on their hand.
Balance systems are a generalization of the spring–mass system we saw earlier. We can write the dynamics for a mechanical system in the general form

\[ M(q) \ddot{q} + C(q, \dot{q}) + K(q) = B(q)u, \]

where \( M(q) \) is the inertia matrix for the system, \( C(q, \dot{q}) \) represents the Coriolis forces as well as the damping, \( K(q) \) gives the forces due to potential energy and \( B(q) \) describes how the external applied forces couple into the dynamics. The specific form of the equations can be derived using Newtonian mechanics. Note that each of the terms depends on the configuration of the system \( q \) and that these terms are often nonlinear in the configuration variables.

Figure 2.5c shows a simplified diagram for a balance system consisting of an inverted pendulum on a cart. To model this system, we choose state variables that represent the position and velocity of the base of the system, \( p \) and \( \dot{p} \), and the angle and angular rate of the structure above the base, \( \theta \) and \( \dot{\theta} \). We let \( F \) represent the force applied at the base of the system, assumed to be in the horizontal direction (aligned with \( p \)), and choose the position and angle of the system as outputs. With this set of definitions, the dynamics of the system can be computed using Newtonian mechanics and have the form

\[
\begin{pmatrix}
(M + m) & -ml \cos \theta \\
-ml \cos \theta & (I + ml^2)
\end{pmatrix}
\begin{pmatrix}
\ddot{p} \\
\ddot{\theta}
\end{pmatrix}
+ \begin{pmatrix}
c \dot{\theta} + ml \sin \theta \dot{\theta}^2 \\
c \dot{\theta} - mgl \sin \theta
\end{pmatrix}
= \begin{pmatrix}
F \\
0
\end{pmatrix}.
\tag{2.9}
\]

where \( M \) is the mass of the base, \( m \) and \( J \) are the mass and moment of inertia of the system to be balanced, \( l \) is the distance from the base to the center of mass of the balanced body, \( c \) and \( \gamma \) are coefficients of viscous friction and \( g \) is the acceleration due to gravity.

We can rewrite the dynamics of the system in state space form by defining the state as \( x = (p, \theta, \dot{p}, \dot{\theta}) \), the input as \( u = F \) and the output as \( y = (p, \theta) \). If we
define the total mass and total inertia as
\[ M_t = M + m, \quad J_t = J + ml^2, \]
the equations of motion then become
\[
\frac{d}{dt} \begin{bmatrix} p \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & m^2l^2g/\mu & -cJ_t/\mu & -\gamma lm/\mu \\ 0 & M_mgl/\mu & -clm/\mu & -\gamma M_t/\mu \end{bmatrix} \begin{bmatrix} p \\ \dot{\theta} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ J_t/\mu \\ lm/\mu \end{bmatrix} u, \\
y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} x,
\]
where \( M = M_tJ_t - m^2l^2 \).

**Example 2.2 Inverted pendulum**
A variation of the previous example is one in which the location of the base \( p \) does not need to be controlled. This happens, for example, if we are interested only in stabilizing a rocket’s upright orientation without worrying about the location of base of the rocket. The dynamics of this simplified system are given by
\[
\frac{d}{dt} \begin{bmatrix} \theta \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \frac{mgl}{J_t} \sin \theta - \frac{\gamma}{J_t} \theta + \frac{l}{J_t} \cos \theta u \\ \dot{\theta} \end{bmatrix}, \quad y = \theta, \quad (2.10)
\]
where \( \gamma \) is the coefficient of rotational friction, \( J_t = J + ml^2 \) and \( u \) is the force applied at the base. This system is referred to as an inverted pendulum.

**Difference Equations**
In some circumstances, it is more natural to describe the evolution of a system at discrete instants of time rather than continuously in time. If we refer to each of
these times by an integer $k = 0, 1, 2, \ldots$, then we can ask how the state of the system changes for each $k$. Just as in the case of differential equations, we define the state to be those sets of variables that summarize the past of the system for the purpose of predicting its future. Systems described in this manner are referred to as *discrete-time systems*.

The evolution of a discrete-time system can be written in the form

$$
x[k + 1] = f(x[k], u[k]), \quad y[k] = h(x[k], u[k]),
$$

where $x[k] \in \mathbb{R}^n$ is the state of the system at time $k$ (an integer), $u[k] \in \mathbb{R}^p$ is the input and $y[k] \in \mathbb{R}^q$ is the output. As before, $f$ and $h$ are smooth mappings of the appropriate dimension. We call equation (2.11) a *difference equation* since it tells us how $x[k + 1]$ differs from $x[k]$. The state $x[k]$ can be either a scalar- or a vector-valued quantity; in the case of the latter we write $x_j[k]$ for the value of the $j$th state at time $k$.

Just as in the case of differential equations, it is often the case that the equations are linear in the state and input, in which case we can describe the system by

$$
x[k + 1] = Ax[k] + Bu[k], \quad y[k] = Cx[k] + Du[k].
$$

As before, we refer to the matrices $A$, $B$, $C$ and $D$ as the dynamics matrix, the control matrix, the sensor matrix and the direct term. The solution of a linear difference equation with initial condition $x[0]$ and input $u[0], \ldots, u[T]$ is given by

$$
x[k] = A^k x_0 + \sum_{j=0}^{k-1} A^{k-j-1} Bu[j],
$$

$$
y[k] = CA^k x_0 + \sum_{j=0}^{k-1} CA^{k-j-1} Bu[j] + Du[k],
$$

$k > 0$, (2.12)

Difference equations are also useful as an approximation of differential equations, as we will show later.

**Example 2.3 Predator–prey**

As an example of a discrete-time system, consider a simple model for a predator–prey system. The predator–prey problem refers to an ecological system in which we have two species, one of which feeds on the other. This type of system has been studied for decades and is known to exhibit interesting dynamics. Figure 2.6 shows a historical record taken over 90 years for a population of lynxes versus a population of hares [Mac37]. As can been seen from the graph, the annual records of the populations of each species are oscillatory in nature.

A simple model for this situation can be constructed using a discrete-time model by keeping track of the rate of births and deaths of each species. Letting $H$ represent the population of hares and $L$ represent the population of lynxes, we can describe the state in terms of the populations at discrete periods of time. Letting $k$ be the
Figure 2.6: Predator versus prey. The photograph on the left shows a Canadian lynx and a snowshoe hare, the lynx’s primary prey. The graph on the right shows the populations of hares and lynxes between 1845 and 1935 in a section of the Canadian Rockies [Mac37]. The data were collected on an annual basis over a period of 90 years. (Photograph copyright Tom and Pat Leeson.)

discrete-time index (e.g., the day or month number), we can write

\[ H[k + 1] = H[k] + b_r(u)H[k] - aL[k]H[k], \]
\[ L[k + 1] = L[k] + cL[k]H[k] - d_fL[k], \]

(2.13)

where \( b_r(u) \) is the hare birth rate per unit period and as a function of the food supply \( u \), \( d_f \) is the lynx mortality rate and \( a \) and \( c \) are the interaction coefficients.

The interaction term \( aL[k]H[k] \) models the rate of predation, which is assumed to be proportional to the rate at which predators and prey meet and is hence given by the product of the population sizes. The interaction term \( cL[k]H[k] \) in the lynx dynamics has a similar form and represents the rate of growth of the lynx population. This model makes many simplifying assumptions—such as the fact that hares decrease in number only through predation by lynxes—but it often is sufficient to answer basic questions about the system.

To illustrate the use of this system, we can compute the number of lynxes and hares at each time point from some initial population. This is done by starting with \( x[0] = (H_0, L_0) \) and then using equation (2.13) to compute the populations in the following period. By iterating this procedure, we can generate the population over time. The output of this process for a specific choice of parameters and initial conditions is shown in Figure 2.7. While the details of the simulation are different from the experimental data (to be expected given the simplicity of our assumptions), we see qualitatively similar trends and hence we can use the model to help explore the dynamics of the system.

Example 2.4 E-mail server

The IBM Lotus server is an collaborative software system that administers users’ e-mail, documents and notes. Client machines interact with end users to provide access to data and applications. The server also handles other administrative tasks. In the early development of the system it was observed that the performance was poor when the central processing unit (CPU) was overloaded because of too many service requests, and mechanisms to control the load were therefore introduced.

The interaction between the client and the server is in the form of remote pro-
Procedure calls (RPCs). The server maintains a log of statistics of completed requests. The total number of requests being served, called Ris (RPCs in server), is also measured. The load on the server is controlled by a parameter called MaxUsers, which sets the total number of client connections to the server. This parameter is controlled by the system administrator. The server can be regarded as a dynamic system with MaxUsers as the input and Ris as the output. The relationship between input and output was first investigated by exploring the steady-state performance and was found to be linear.

In [HDPT04] a dynamic model in the form of a first-order difference equation is used to capture the dynamic behavior of this system. Using system identification techniques, they construct a model of the form

$$y[k + 1] = ay[k] + bu[k],$$

where $u = \text{MaxUsers} - \text{MaxUsers}$ and $y = \text{Ris} - \text{Ris}$. The parameters $a = 0.43$ and $b = 0.47$ are parameters that describe the dynamics of the system around the operating point, and $\text{MaxUsers} = 165$ and $\text{Ris} = 135$ represent the nominal operating point of the system. The number of requests was averaged over a sampling period of 60 s.

Simulation and Analysis

State space models can be used to answer many questions. One of the most common, as we have seen in the previous examples, involves predicting the evolution of the system state from a given initial condition. While for simple models this can be done in closed form, more often it is accomplished through computer simulation. One can also use state space models to analyze the overall behavior of the system without making direct use of simulation.

Consider again the damped spring–mass system from Section 2.1, but this time with an external force applied, as shown in Figure 2.8. We wish to predict the motion of the system for a periodic forcing function, with a given initial condition, and determine the amplitude, frequency and decay rate of the resulting motion.
2.2. STATE SPACE MODELS

We choose to model the system with a linear ordinary differential equation. Using Hooke’s law to model the spring and assuming that the damper exerts a force that is proportional to the velocity of the system, we have

\[ m\ddot{q} + c\dot{q} + kq = u, \quad (2.14) \]

where \( m \) is the mass, \( q \) is the displacement of the mass, \( c \) is the coefficient of viscous friction, \( k \) is the spring constant and \( u \) is the applied force. In state space form, using \( x = (q, \dot{q}) \) as the state and choosing \( y = q \) as the output, we have

\[
\begin{align*}
\frac{dx}{dt} = \begin{bmatrix}
x_2 \\
-\frac{c}{m}x_2 - \frac{k}{m}x_1 + \frac{u}{m}
\end{bmatrix}, \\
y = x_1.
\end{align*}
\]

We see that this is a linear second-order differential equation with one input \( u \) and one output \( y \).

We now wish to compute the response of the system to an input of the form \( u = A \sin \omega t \). Although it is possible to solve for the response analytically, we instead make use of a computational approach that does not rely on the specific form of this system. Consider the general state space system

\[
\frac{dx}{dt} = f(x, u).
\]

Given the state \( x \) at time \( t \), we can approximate the value of the state at a short time \( h > 0 \) later by assuming that the rate of change of \( f(x, u) \) is constant over the interval \( t \) to \( t + h \). This gives

\[
x(t + h) = x(t) + hf(x(t), u(t)). \quad (2.15)
\]

Iterating this equation, we can thus solve for \( x \) as a function of time. This approximation is known as Euler integration and is in fact a difference equation if we let \( h \) represent the time increment and write \( x[k] = x(kh) \). Although modern simulation tools such as MATLAB and Mathematica use more accurate methods than Euler integration, they still have some of the same basic trade-offs.

Returning to our specific example, Figure 2.9 shows the results of computing \( x(t) \) using equation (2.15), along with the analytical computation. We see that as
Figure 2.9: Simulation of the forced spring–mass system with different simulation time constants. The dashed line represents the analytical solution. The solid lines represent the approximate solution via the method of Euler integration, using decreasing step sizes.

$h$ gets smaller, the computed solution converges to the exact solution. The form of the solution is also worth noticing: after an initial transient, the system settles into a periodic motion. The portion of the response after the transient is called the steady-state response to the input.

In addition to generating simulations, models can also be used to answer other types of questions. Two that are central to the methods described in this text concern the stability of an equilibrium point and the input/output frequency response. We illustrate these two computations through the examples below and return to the general computations in later chapters.

Returning to the damped spring–mass system, the equations of motion with no input forcing are given by

\[
\frac{dx}{dt} = \begin{bmatrix}
  x_2 \\
  \frac{c}{m} x_2 - \frac{k}{m} x_1
\end{bmatrix},
\]  

(2.16)

where $x_1$ is the position of the mass (relative to the rest position) and $x_2$ is its velocity. We wish to show that if the initial state of the system is away from the rest position, the system will return to the rest position eventually (we will later define this situation to mean that the rest position is asymptotically stable). While we could heuristically show this by simulating many, many initial conditions, we seek instead to prove that this is true for any initial condition.

To do so, we construct a function $V : \mathbb{R}^n \to \mathbb{R}$ that maps the system state to a positive real number. For mechanical systems, a convenient choice is the energy of the system,

\[
V(x) = \frac{1}{2} k x_1^2 + \frac{1}{2} m x_2^2.
\]  

(2.17)

If we look at the time derivative of the energy function, we see that

\[
\frac{dV}{dt} = k x_1 \dot{x}_1 + m x_2 \dot{x}_2 = k x_1 x_2 + m x_2 \left( -\frac{c}{m} x_2 - \frac{k}{m} x_1 \right) = -c x_2^2,
\]

which is always either negative or zero. Hence $V(x(t))$ is never increasing and,
using a bit of analysis that we will see formally later, the individual states must remain bounded.

If we wish to show that the states eventually return to the origin, we must use a slightly more detailed analysis. Intuitively, we can reason as follows: suppose that for some period of time, \( V(x(t)) \) stops decreasing. Then it must be true that \( \dot{V}(x(t)) = 0 \), which in turn implies that \( x_2(t) = 0 \) for that same period. In that case, \( \dot{x}_2(t) = 0 \), and we can substitute into the second line of equation (2.16) to obtain

\[
0 = \dot{x}_2 = -\frac{c}{m}x_2 - \frac{k}{m}x_1 = \frac{k}{m}x_1.
\]

Thus we must have that \( x_1 \) also equals zero, and so the only time that \( V(x(t)) \) can stop decreasing is if the state is at the origin (and hence this system is at its rest position). Since we know that \( V(x(t)) \) is never increasing (because \( \dot{V} \leq 0 \)), we therefore conclude that the origin is stable (for any initial condition).

This type of analysis, called Lyapunov stability analysis, is considered in detail in Chapter 4. It shows some of the power of using models for the analysis of system properties.

Another type of analysis that we can perform with models is to compute the output of a system to a sinusoidal input. We again consider the spring–mass system, but this time keeping the input and leaving the system in its original form:

\[
m\ddot{q} + c\dot{q} + kq = u.
\]

We wish to understand how the system responds to a sinusoidal input of the form

\[
u(t) = A \sin \omega t.
\]

We will see how to do this analytically in Chapter 6, but for now we make use of simulations to compute the answer.

We first begin with the observation that if \( q(t) \) is the solution to equation (2.18) with input \( u(t) \), then applying an input \( 2u(t) \) will give a solution \( 2q(t) \) (this is easily verified by substitution). Hence it suffices to look at an input with unit magnitude, \( A = 1 \). A second observation, which we will prove in Chapter 5, is that the long-term response of the system to a sinusoidal input is itself a sinusoid at the same frequency, and so the output has the form

\[
q(t) = g(\omega) \sin(\omega t + \varphi(\omega)),
\]

where \( g(\omega) \) is called the gain of the system and \( \varphi(\omega) \) is called the phase (or phase offset).

To compute the frequency response numerically, we can simulate the system at a set of frequencies \( \omega_1, \ldots, \omega_N \) and plot the gain and phase at each of these frequencies. An example of this type of computation is shown in Figure 2.10.
Figure 2.10: A frequency response (gain only) computed by measuring the response of individual sinusoids. The figure on the left shows the response of the system as a function of time to a number of different unit magnitude inputs (at different frequencies). The figure on the right shows this same data in a different way, with the magnitude of the response plotted as a function of the input frequency. The filled circles correspond to the particular frequencies shown in the time responses.

2.3 Modeling Methodology

To deal with large, complex systems, it is useful to have different representations of the system that capture the essential features and hide irrelevant details. In all branches of science and engineering it is common practice to use some graphical description of systems, called schematic diagrams. They can range from stylistic pictures to drastically simplified standard symbols. These pictures make it possible to get an overall view of the system and to identify the individual components. Examples of such diagrams are shown in Figure 2.11. Schematic diagrams are useful because they give an overall picture of a system, showing different subprocesses and their interconnection and indicating variables that can be manipulated and signals that can be measured.

Block Diagrams

A special graphical representation called a block diagram has been developed in control engineering. The purpose of a block diagram is to emphasize the information flow and to hide details of the system. In a block diagram, different process elements are shown as boxes, and each box has inputs denoted by lines with arrows pointing toward the box and outputs denoted by lines with arrows going out of the box. The inputs denote the variables that influence a process, and the outputs denote the signals that we are interested in or signals that influence other subsystems. Block diagrams can also be organized in hierarchies, where individual blocks may themselves contain more detailed block diagrams.

Figure 2.12 shows some of the notation that we use for block diagrams. Signals are represented as lines, with arrows to indicate inputs and outputs. The first diagram is the representation for a summation of two signals. An input/output response is represented as a rectangle with the system name (or mathematical description) in
2.3. MODELING METHODOLOGY

Figure 2.11: Schematic diagrams for different disciplines. Each diagram is used to illustrate the dynamics of a feedback system: (a) electrical schematics for a power system [Kun93], (b) a biological circuit diagram for a synthetic clock circuit [ASMN03], (c) a process diagram for a distillation column [SEM04] and (d) a Petri net description of a communication protocol.

Figure 2.12: Standard block diagram elements. The arrows indicate the the inputs and outputs of each element, with the mathematical operation corresponding to the blocked labeled at the output. The system block (f) represents the full input/output response of a dynamical system.
Figure 2.13: A block diagram representation of the flight control system for an insect flying against the wind. The mechanical portion of the model consists of the rigid-body dynamics of the fly, the drag due to flying through the air and the forces generated by the wings. The motion of the body causes the visual environment of the fly to change, and this information is then used to control the motion of the wings (through the sensory motor system), closing the loop.

The block. Two special cases are a proportional gain, which scales the input by a multiplicative factor, and an integrator, which outputs the integral of the input signal.

Figure 2.13 illustrates the use of a block diagram, in this case for modeling the flight response of a fly. The flight dynamics of an insect are incredibly intricate, involving careful coordination of the muscles within the fly to maintain stable flight in response to external stimuli. One known characteristic of flies is their ability to fly upwind by making use of the optical flow in their compound eyes as a feedback mechanism. Roughly speaking, the fly controls its orientation so that the point of contraction of the visual field is centered in its visual field.

To understand this complex behavior, we can decompose the overall dynamics of the system into a series of interconnected subsystems (or blocks). Referring to Figure 2.13, we can model the insect navigation system through an interconnection of five blocks. The sensory motor system (a) takes the information from the visual system (e) and generates muscle commands that attempt to steer the fly so that the point of contraction is centered. These muscle commands are converted into forces through the flapping of the wings (b) and the resulting aerodynamic forces that are produced. The forces from the wings are combined with the drag on the fly (d) to produce a net force on the body of the fly. The wind velocity enters through the drag aerodynamics. Finally, the body dynamics (c) describe how the fly translates and rotates as a function of the net forces that are applied to it. The insect position, speed and orientation are fed back to the drag aerodynamics and vision system blocks as inputs.

Each of the blocks in the diagram can itself be a complicated subsystem. For example, the visual system of a fruit fly consists of two complicated compound eyes (with about 700 elements per eye), and the sensory motor system has about 200,000
neurons that are used to process information. A more detailed block diagram of
the insect flight control system would show the interconnections between these
elements, but here we have used one block to represent how the motion of the fly
affects the output of the visual system, and a second block to represent how the visual
field is processed by the fly’s brain to generate muscle commands. The choice of the
level of detail of the blocks and what elements to separate into different blocks often
depends on experience and the questions that one wants to answer using the model.
One of the powerful features of block diagrams is their ability to hide information
about the details of a system that may not be needed to gain an understanding of
the essential dynamics of the system.

Modeling from Experiments

Since control systems are provided with sensors and actuators, it is also possible to
obtain models of system dynamics from experiments on the process. The models
are restricted to input/output models since only these signals are accessible to
experiments, but modeling from experiments can also be combined with modeling
from physics through the use of feedback and interconnection.

A simple way to determine a system’s dynamics is to observe the response to a
step change in the control signal. Such an experiment begins by setting the control
signal to a constant value; then when steady state is established, the control signal is
changed quickly to a new level and the output is observed. The experiment gives the
step response of the system, and the shape of the response gives useful information
about the dynamics. It immediately gives an indication of the response time, and it
tells if the system is oscillatory or if the response is monotone.

Example 2.5 Spring–mass system

Consider the spring–mass system from Section 2.1, whose dynamics are given by

\[ m\ddot{q} + c\dot{q} + kq = u. \]  

(2.19)

We wish to determine the constants \( m, c \) and \( k \) by measuring the response of the
system to a step input of magnitude \( F_0 \).

We will show in Chapter 6 that when \( c^2 < 4km \), the step response for this
system from the rest configuration is given by

\[ q(t) = \frac{F_0}{k} \left( 1 - \exp\left( -\frac{ct}{2m} \right) \sin(\omega_d t + \varphi) \right), \quad \omega_d = \frac{\sqrt{4km - c^2}}{2m}, \quad \varphi = \tan^{-1} \left( \frac{\sqrt{4km - c^2}}{c} \right). \]

From the form of the solution, we see that the form of the response is determined
by the parameters of the system. Hence, by measuring certain features of the step
response we can determine the parameter values.

Figure 2.14 shows the response of the system to a step of magnitude \( F_0 = 20 \)
N, along with some measurements. We start by noting that the steady-state position
of the mass (after the oscillations die down) is a function of the spring constant \( k \):

\[
q(\infty) = \frac{F_0}{k},
\]

(2.20)

where \( F_0 \) is the magnitude of the applied force (\( F_0 = 1 \) for a unit step input). The parameter \( 1/k \) is called the gain of the system. The period of the oscillation can be measured between two peaks and must satisfy

\[
\frac{2\pi}{T} = \sqrt{\frac{4km - c^2}{2m}}.
\]

(2.21)

Finally, the rate of decay of the oscillations is given by the exponential factor in the solution. Measuring the amount of decay between two peaks, we have

\[
\log\left(\frac{q(t_1) - F_0}{k}\right) - \log\left(\frac{q(t_2) - F_0}{k}\right) = \frac{c}{2m}(t_2 - t_1).
\]

(2.22)

Using this set of three equations, we can solve for the parameters and determine that for the step response in Figure 2.14 we have \( m \approx 250 \text{ kg}, c \approx 60 \text{ N s/m} \) and \( k \approx 40 \text{ N/m} \).

Modeling from experiments can also be done using many other signals. Sinusoidal signals are commonly used (particularly for systems with fast dynamics) and precise measurements can be obtained by exploiting correlation techniques. An indication of nonlinearities can be obtained by repeating experiments with input signals having different amplitudes.

**Normalization and Scaling**

Having obtained a model, it is often useful to scale the variables by introducing dimension-free variables. Such a procedure can often simplify the equations for a system by reducing the number of parameters and reveal interesting properties of
the model. Scaling can also improve the numerical conditioning of the model to allow faster and more accurate simulations.

The procedure of scaling is straightforward: choose units for each independent variable and introduce new variables by dividing the variables by the chosen normalization unit. We illustrate the procedure with two examples.

Example 2.6 Spring–mass system
Consider again the spring–mass system introduced earlier. Neglecting the damping, the system is described by

\[ m \ddot{q} + kq = u. \]

The model has two parameters \( m \) and \( k \). To normalize the model we introduce dimension-free variables \( x = q/l \) and \( \tau = \omega_0 t \), where \( \omega_0 = \sqrt{k/m} \) and \( l \) is the chosen length scale. We scale force by \( m\omega_0^2 \) and introduce \( v = u/(m\omega_0^2) \). The scaled equation then becomes

\[
\frac{d^2x}{d\tau^2} = \frac{d^2q}{d(\omega_0 t)^2} = \frac{1}{m\omega_0^2}(-kq + u) = -x + v,
\]

which is the normalized undamped spring–mass system. Notice that the normalized model has no parameters, while the original model had two parameters \( m \) and \( k \).

Introducing the scaled, dimension-free state variables \( z_1 = x = q/l \) and \( z_2 = dx/d\tau = \dot{q}/(l\omega_0) \), the model can be written as

\[
\frac{d}{dt} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} 0 \\ v \end{bmatrix}.
\]

This simple linear equation describes the dynamics of any spring–mass system, independent of the particular parameters, and hence gives us insight into the fundamental dynamics of this oscillatory system. To recover the physical frequency of oscillation or its magnitude, we must invert the scaling we have applied. \( \nabla \)

Example 2.7 Balance system
Consider the balance system described in Section 2.1. Neglecting damping by putting \( c = 0 \) and \( \gamma = 0 \) in equation (2.9), the model can be written as

\[
(M + m) \frac{d^2q}{dt^2} - ml \cos \theta \frac{d^2\theta}{dt^2} + ml \sin \theta \left( \frac{dq}{dt} \right)^2 = F,
\]

\[
-ml \cos \theta \frac{d^2q}{dt^2} + (J + ml^2) \frac{d^2\theta}{dt^2} - mgl \sin \theta = 0.
\]

Let \( \omega_0 = \sqrt{mgl/(J + ml^2)} \), choose the length scale as \( l \), let the time scale be \( 1/\omega_0 \), choose the force scale as \( (M + m)\omega_0^2 \) and introduce the scaled variables \( \tau = \omega_0 t \), \( x = q/l \) and \( u = F/((M + m)\omega_0^2) \). The equations then become

\[
\frac{d^2x}{d\tau^2} - \alpha \cos \theta \frac{d^2\theta}{d\tau^2} + \alpha \sin \theta \left( \frac{d\theta}{d\tau} \right)^2 = u,
\]

\[
-\beta \cos \theta \frac{d^2x}{d\tau^2} + \frac{d^2\theta}{d\tau^2} - \sin \theta = 0,
\]

where \( \alpha = m/(M + m) \) and \( \beta = ml^2/(J + ml^2) \). Notice that the original model has five parameters \( m, M, J, l \) and \( g \) but the normalized model has only two parameters
\[ M \gg m \quad \text{and} \quad ml^2 \gg J \], we get \( \alpha \approx 0 \) and \( \beta \approx 1 \) and the model can be approximated by

\[
\frac{d^2 x}{d\tau^2} = u, \quad \frac{d^2 \theta}{d\tau^2} - \sin \theta = u \cos \theta.
\]

The model can be interpreted as a mass combined with an inverted pendulum driven by the same input.

\[ \nabla \]

**Model Uncertainty**

Reducing uncertainty is one of the main reasons for using feedback, and it is therefore important to characterize uncertainty. When making measurements, there is a good tradition to assign both a nominal value and a measure of uncertainty. It is useful to apply the same principle to modeling, but unfortunately it is often difficult to express the uncertainty of a model quantitatively.

For a static system whose input/output relation can be characterized by a function, uncertainty can be expressed by an uncertainty band as illustrated in Figure 2.15a. At low signal levels there are uncertainties due to sensor resolution, friction and quantization. Some models for queuing systems or cells are based on averages that exhibit significant variations for small populations. At large signal levels there are saturations or even system failures. The signal ranges where a model is reasonably accurate vary dramatically between applications, but it is rare to find models that are accurate for signal ranges larger than \( 10^4 \).

Characterization of the uncertainty of a dynamic model is much more difficult. We can try to capture uncertainties by assigning uncertainties to parameters of the model, but this is often not sufficient. There may be errors due to phenomena that have been neglected, e.g., small time delays. In control the ultimate test is how well a control system based on the model performs, and time delays can be important. There is also a frequency aspect. There are slow phenomena, such as aging, that
2.4. Modeling Examples

In this section we introduce additional examples that illustrate some of the different types of systems for which one can develop differential equation and difference equation models. These examples are specifically chosen from a range of different fields to highlight the broad variety of systems to which feedback and control concepts can be applied. A more detailed set of applications that serve as running examples throughout the text are given in the next chapter.

**Motion Control Systems**

Motion control systems involve the use of computation and feedback to control the movement of a mechanical system. Motion control systems range from nanopositioning systems (atomic force microscopes, adaptive optics), to control systems for the read/write heads in a disk drive of a CD player, to manufacturing systems (transfer machines and industrial robots), to automotive control systems (anti-lock brakes, suspension control, traction control), to air and space flight control systems (airplanes, satellites, rockets and planetary rovers).

**Example 2.8 Vehicle steering—the bicycle model**

A common problem in motion control is to control the trajectory of a vehicle through an actuator that causes a change in the orientation. A steering wheel on an automobile and the front wheel of a bicycle are two examples, but similar dynamics occur in the steering of ships or control of the pitch dynamics of an aircraft. In many cases, we can understand the basic behavior of these systems through the use of a simple model that captures the basic kinematics of the system.

Consider a vehicle with two wheels as shown in Figure 2.16. For the purpose of steering we are interested in a model that describes how the velocity of the vehicle
depends on the steering angle \( \delta \). To be specific, consider the velocity \( v \) at the center of mass, a distance \( a \) from the rear wheel, and let \( b \) be the wheel base, as shown in Figure 2.16. Let \( x \) and \( y \) be the coordinates of the center of mass, \( \theta \) the heading angle and \( \alpha \) the angle between the velocity vector \( v \) and the centerline of the vehicle. Since \( b = r_a \tan \delta \) and \( a = r_a \tan \alpha \), it follows that \( \tan \alpha = \frac{a}{b} \tan \delta \) and we get the following relation between \( \alpha \) and the steering angle \( \delta \):

\[
\alpha(\delta) = \arctan \left( \frac{a \tan \delta}{b} \right). \tag{2.23}
\]

Assume that the wheels are rolling without slip and that the velocity of the rear wheel is \( v_0 \). The vehicle speed at its center of mass is \( v = v_0 / \cos \alpha \), and we find that the motion of this point is given by

\[
\begin{align*}
\frac{dx}{dt} &= v \cos (\alpha + \theta) = \frac{v_0 \cos (\alpha + \theta)}{\cos \alpha}, \\
\frac{dy}{dt} &= v \sin (\alpha + \theta) = \frac{v_0 \sin (\alpha + \theta)}{\cos \alpha}.
\end{align*} \tag{2.24}
\]

To see how the angle \( \theta \) is influenced by the steering angle, we observe from Figure 2.16 that the vehicle rotates with the angular velocity \( v_0 / r_a \) around the point \( O \). Hence

\[
\frac{d\theta}{dt} = \frac{v_0}{r_a} = \frac{v_0}{b} \tan \delta. \tag{2.25}
\]

Equations (2.23)–(2.25) can be used to model an automobile under the assumptions that there is no slip between the wheels and the road and that the two front wheels can be approximated by a single wheel at the center of the car. The assumption of no slip can be relaxed by adding an extra state variable, giving a more realistic model. Such a model also describes the steering dynamics of ships as well.
as the pitch dynamics of aircraft and missiles. It is also possible to choose coordinates so that the reference point is at the rear wheels (corresponding to setting $\alpha = 0$), a model often referred to as the Dubins car [Dub57].

Figure 2.16 represents the situation when the vehicle moves forward and has front-wheel steering. The case when the vehicle reverses is obtained by changing the sign of the velocity, which is equivalent to a vehicle with rear-wheel steering.

\[\n\]

**Example 2.9 Vectored thrust aircraft**

Consider the motion of vectored thrust aircraft, such as the Harrier “jump jet” shown Figure 2.17a. The Harrier is capable of vertical takeoff by redirecting its thrust downward and through the use of smaller maneuvering thrusters located on its wings. A simplified model of the Harrier is shown in Figure 2.17b, where we focus on the motion of the vehicle in a vertical plane through the wings of the aircraft. We resolve the forces generated by the main downward thruster and the maneuvering thrusters as a pair of forces $F_1$ and $F_2$ acting at a distance $r$ below the aircraft (determined by the geometry of the thrusters).

Let $(x, y, \theta)$ denote the position and orientation of the center of mass of the aircraft. Let $m$ be the mass of the vehicle, $J$ the moment of inertia, $g$ the gravitational constant and $c$ the damping coefficient. Then the equations of motion for the vehicle are given by

\[
\begin{align*}
m \ddot{x} &= F_1 \cos \theta - F_2 \sin \theta - c \dot{x}, \\
m \ddot{y} &= F_1 \sin \theta + F_2 \cos \theta - mg - c \dot{y}, \\
J \ddot{\theta} &= r F_1.
\end{align*}
\]

(2.26)

It is convenient to redefine the inputs so that the origin is an equilibrium point of the
Figure 2.18: Schematic diagram of a queuing system. Messages arrive at rate $\lambda$ and are stored in a queue. Messages are processed and removed from the queue at rate $\mu$. The average size of the queue is given by $x \in \mathbb{R}$.

These equations describe the motion of the vehicle as a set of three coupled second-order differential equations.

Information Systems

Information systems range from communication systems like the Internet to software systems that manipulate data or manage enterprise-wide resources. Feedback is present in all these systems, and designing strategies for routing, flow control and buffer management is a typical problem. Many results in queuing theory emerged from design of telecommunication systems and later from development of the Internet and computer communication systems [BG87, Kle75, Sch87]. Management of queues to avoid congestion is a central problem and we will therefore start by discussing the modeling of queuing systems.

Example 2.10 Queuing systems

A schematic picture of a simple queue is shown in Figure 2.18. Requests arrive and are then queued and processed. There can be large variations in arrival rates and service rates, and the queue length builds up when the arrival rate is larger than the service rate. When the queue becomes too large, service is denied using an admission control policy.

The system can be modeled in many different ways. One way is to model each incoming request, which leads to an event-based model where the state is an integer that represents the queue length. The queue changes when a request arrives or a request is serviced. The statistics of arrival and servicing are typically modeled as random processes. In many cases it is possible to determine statistics of quantities like queue length and service time, but the computations can be quite complicated.

A significant simplification can be obtained by using a flow model. Instead of keeping track of each request we instead view service and requests as flows, similar to what is done when replacing molecules by a continuum when analyzing...
2.4. MODELING EXAMPLES

Figure 2.19: Queueing dynamics. (a) The steady-state queue length as a function of $\lambda / \mu_{\text{max}}$. (b) The behavior of the queue length when there is a temporary overload in the system. The solid line shows a realization of an event-based simulation, and the dashed line shows the behavior of the flow model (2.29).

Fluids. Assuming that the average queue length $x$ is a continuous variable and that arrivals and services are flows with rates $\lambda$ and $\mu$, the system can be modeled by the first-order differential equation

$$\frac{dx}{dt} = \lambda - \mu = \lambda - \mu_{\text{max}} f(x), \quad x \geq 0,$$

(2.28)

where $\mu_{\text{max}}$ is the maximum service rate and $f(x)$ is a number between 0 and 1 that describes the effective service rate as a function of the queue length.

It is natural to assume that the effective service rate depends on the queue length because larger queues require more resources. In steady state we have $f(x) = \lambda / \mu_{\text{max}}$, and we assume that the queue length goes to zero when $\lambda / \mu_{\text{max}}$ goes to zero and that it goes to infinity when $\lambda / \mu_{\text{max}}$ goes to 1. This implies that $f(0) = 0$ and that $f(\infty) = 1$. In addition, if we assume that the effective service rate deteriorates monotonically with queue length, then the function $f(x)$ is monotone and concave.

A simple function that satisfies the basic requirements is $f(x) = x / (1 + x)$, which gives the model

$$\frac{dx}{dt} = \lambda - \mu_{\text{max}} \frac{x}{x + 1}.$$

(2.29)

This model was proposed by Agnew [Agn76]. It can be shown that if arrival and service processes are Poisson processes, the average queue length is given by equation (2.29) and that equation (2.29) is a good approximation even for short queue lengths; see Tipper [TS90].

To explore the properties of the model (2.29) we will first investigate the equilibrium value of the queue length when the arrival rate $\lambda$ is constant. Setting the derivative $dx/dt$ to zero in equation (2.29) and solving for $x$, we find that the queue length $x$ approaches the steady-state value

$$x_e = \frac{\lambda}{\mu_{\text{max}} - \lambda}.$$

(2.30)

Figure 2.19a shows the steady-state queue length as a function of $\lambda / \mu_{\text{max}}$, the effective service rate excess. Notice that the queue length increases rapidly as $\lambda$
many complex systems use discrete control actions. Such systems can be modeled by characterizing the situations that correspond to each control action, as illustrated in the following example.

**Example 2.11 Virtual memory paging control**

An early example of the use of feedback in computer systems was applied in the operating system OS/VS for the IBM 370 [BG68, Cro75]. The system used virtual memory, which allows programs to address more memory than is physically available as fast memory. Data in current fast memory (random access memory, RAM) is accessed directly, but data that resides in slower memory (disk) is automatically loaded into fast memory. The system is implemented in such a way that it appears to the programmer as a single large section of memory. The system performed very well in many situations, but very long execution times were encountered in overload situations, as shown by the open circles in Figure 2.20a. The difficulty was resolved with a simple discrete feedback system. The load of the central processing
Example 2.12 Consensus protocols in sensor networks

Sensor networks are used in a variety of applications where we want to collect and aggregate information over a region of space using multiple sensors that are connected together via a communications network. Examples include monitoring environmental conditions in a geographical area (or inside a building), monitoring the movement of animals or vehicles and monitoring the resource loading across a group of computers. In many sensor networks the computational resources are distributed along with the sensors, and it can be important for the set of distributed agents to reach a consensus about a certain property, such as the average temperature in a region or the average computational load among a set of computers.

We model the connectivity of the sensor network using a graph, with nodes corresponding to the sensors and edges corresponding to the existence of a direct communications link between two nodes. We use the notation $\mathcal{N}_i$ to represent the set of neighbors of a node $i$. For example, in the network shown in Figure 2.21a $\mathcal{N}_2 = \{1, 3, 4, 5\}$ and $\mathcal{N}_3 = \{2, 4\}$.

To solve the consensus problem, let $x_i$ be the state of the $i$th sensor, corresponding to that sensor’s estimate of the average value that we are trying to compute. We initialize the state to the value of the quantity measured by the individual sensor.
The consensus protocol (algorithm) can now be realized as a local update law

\[ x_i[k + 1] = x_i[k] + \gamma \sum_{j \in N_i} (x_j[k] - x_i[k]). \]  

(2.31)

This protocol attempts to compute the average by updating the local state of each agent based on the value of its neighbors. The combined dynamics of all agents can be written in the form

\[ x[k + 1] = x[k] - \gamma (D - A)x[k], \]  

(2.32)

where \( A \) is the adjacency matrix and \( D \) is a diagonal matrix with entries corresponding to the number of neighbors of each node. The constant \( \gamma \) describes the rate at which the estimate of the average is updated based on information from neighboring nodes. The matrix \( L := D - A \) is called the Laplacian of the graph.

The equilibrium points of equation (2.32) are the set of states such that \( x_i[k + 1] = x_i[k] \). It can be shown that \( x_e = (\alpha, \alpha, \ldots, \alpha) \) is an equilibrium state for the system, corresponding to each sensor having an identical estimate \( \alpha \) for the average. Furthermore, we can show that \( \alpha \) is indeed the average value of the initial states. Since there can be cycles in the graph, it is possible that the state of the system could enter into an infinite loop and never converge to the desired consensus state. A formal analysis requires tools that will be introduced later in the text, but it can be shown that for any connected graph we can always find a \( \gamma \) such that the states of the individual agents converge to the average. A simulation demonstrating this property is shown in Figure 2.21b.

\[ \nabla \]

**Biological Systems**

Biological systems provide perhaps the richest source of feedback and control examples. The basic problem of homeostasis, in which a quantity such as temperature or blood sugar level is regulated to a fixed value, is but one of the many types of complex feedback interactions that can occur in molecular machines, cells, organisms and ecosystems.

**Example 2.13 Transcriptional regulation**

Transcription is the process by which messenger RNA (mRNA) is generated from a segment of DNA. The promoter region of a gene allows transcription to be controlled by the presence of other proteins, which bind to the promoter region and either repress or activate RNA polymerase, the enzyme that produces an mRNA transcript from DNA. The mRNA is then translated into a protein according to its nucleotide sequence. This process is illustrated in Figure 2.22.

A simple model of the transcriptional regulation process is through the use of a Hill function [dJ02, Mur04]. Consider the regulation of a protein \( A \) with a concentration given by \( p_a \) and a corresponding mRNA concentration \( m_a \). Let \( B \) be a second protein with concentration \( p_b \) that represses the production of protein \( A \) through transcriptional regulation. The resulting dynamics of \( p_a \) and \( m_a \) can be
2.4. MODELING EXAMPLES

Figure 2.22: Biological circuitry. The cell on the left is a bovine pulmonary cell, stained so that the nucleus, actin and chromatin are visible. The figure on the right gives an overview of the process by which proteins in the cell are made. RNA is transcribed from DNA by an RNA polymerase enzyme. The RNA is then translated into a protein by an organelle called a ribosome.

written as

\[
\frac{dm_a}{dt} = \frac{\alpha_{ab} k_{ab} p_b^n_{ab}}{1 + k_{ab} p_b^{n_{ab}}} + \alpha_{a0} - \gamma_a m_a, \quad \frac{dp_a}{dt} = \beta_a m_a - \delta_a p_a, \tag{2.33}
\]

where \(\alpha_{ab} + \alpha_{a0}\) is the unregulated transcription rate, \(\gamma_a\) represents the rate of degradation of mRNA, \(\alpha_{ab}, k_{ab}\) and \(n_{ab}\) are parameters that describe how B represses A, \(\beta_a\) represents the rate of production of the protein from its corresponding mRNA and \(\delta_a\) represents the rate of degradation of the protein A. The parameter \(\alpha_{a0}\) describes the “leakiness” of the promoter, and \(n_{ab}\) is called the Hill coefficient and relates to the cooperativity of the promoter.

A similar model can be used when a protein activates the production of another protein rather than repressing it. In this case, the equations have the form

\[
\frac{dm_a}{dt} = \frac{\alpha_{ab} k_{ab} p_b^n_{ab}}{1 + k_{ab} p_b^{n_{ab}}} + \alpha_{a0} - \gamma_a m_a, \quad \frac{dp_a}{dt} = \beta_a m_a - \delta_a p_a, \tag{2.34}
\]

where the variables are the same as described previously. Note that in the case of the activator, if \(p_b\) is zero, then the production rate is \(\alpha_{a0}\) (versus \(\alpha_{ab} + \alpha_{a0}\) for the repressor). As \(p_b\) gets large, the first term in the expression for \(m_a\) approaches 1 and the transcription rate becomes \(\alpha_{ab} + \alpha_{a0}\) (versus \(\alpha_{a0}\) for the repressor). Thus we see that the activator and repressor act in opposite fashion from each other.

As an example of how these models can be used, we consider the model of a “repressilator,” originally due to Elowitz and Leibler [EL00]. The repressilator is a synthetic circuit in which three proteins each repress another in a cycle. This is shown schematically in Figure 2.23a, where the three proteins are TetR, \(\lambda\) cl and LacI. The basic idea of the repressilator is that if TetR is present, then it represses the production of \(\lambda\) cl. If \(\lambda\) cl is absent, then LacI is produced (at the unregulated transcription rate), which in turn represses TetR. Once TetR is repressed, then \(\lambda\) cl is no longer repressed, and so on. If the dynamics of the circuit are designed properly, the resulting protein concentrations will oscillate.

We can model this system using three copies of equation (2.33), with A and B replaced by the appropriate combination of TetR, cl and LacI. The state of the
system is then given by \( x = (m_{\text{TetR}}, p_{\text{TetR}}, m_{\text{cI}}, p_{\text{cI}}, m_{\text{LacI}}, p_{\text{LacI}}) \). Figure 2.23b shows the traces of the three protein concentrations for parameters \( n = 2, \alpha = 0.5, k = 6.25 \times 10^{-4}, \alpha_0 = 5 \times 10^{-4}, \gamma = 5.8 \times 10^{-3}, \beta = 0.12 \) and \( \delta = 1.2 \times 10^{-3} \) with initial conditions \( x(0) = (1, 0, 0, 200, 0, 0) \) (following [EL00]).

**Example 2.14 Wave propagation in neuronal networks**

The dynamics of the membrane potential in a cell are a fundamental mechanism in understanding signaling in cells, particularly in neurons and muscle cells. The Hodgkin–Huxley equations give a simple model for studying propagation waves in networks of neurons. The model for a single neuron has the form

\[
C \frac{dV}{dt} = -I_{\text{Na}} - I_{\text{K}} - I_{\text{leak}} + I_{\text{input}},
\]

where \( V \) is the membrane potential, \( C \) is the capacitance, \( I_{\text{Na}} \) and \( I_{\text{K}} \) are the current caused by the transport of sodium and potassium across the cell membrane, \( I_{\text{leak}} \) is a leakage current and \( I_{\text{input}} \) is the external stimulation of the cell. Each current obeys Ohm’s law, i.e.,

\[
I = g(V - E),
\]

where \( g \) is the conductance and \( E \) is the equilibrium voltage. The equilibrium voltage is given by Nernst’s law,

\[
E = \frac{RT}{nF} \log \frac{c_e}{c_i},
\]

where \( R \) is Boltzmann’s constant, \( T \) is the absolute temperature, \( F \) is Faraday’s constant, \( n \) is the charge (or valence) of the ion and \( c_i \) and \( c_e \) are the ion concentrations inside the cell and in the external fluid. At 20 °C we have \( RT/F = 20 \text{ mV} \).

The Hodgkin–Huxley model was originally developed as a means to predict the quantitative behavior of the squid giant axon [HH52]. Hodgkin and Huxley shared the 1963 Nobel Prize in Physiology (along with J. C. Eccles) for analysis of the
electrical and chemical events in nerve cell discharges. The voltage clamp described in Section 1.3 was a key element in Hodgkin and Huxley’s experiments.

2.5 Further Reading

Modeling is ubiquitous in engineering and science and has a long history in applied mathematics. For example, the Fourier series was introduced by Fourier when he modeled heat conduction in solids [Fou07]. Models of dynamics have been developed in many different fields, including mechanics [Arn78, Gol53], heat conduction [CJ59], fluids [BRS60], vehicles [Abk69, Bla91, Ell94], robotics [MLS94, SV89], circuits [Gui63], power systems [Kun93], acoustics [Ber54] and micromechanical systems [Sen01]. Control theory requires modeling from many different domains, and most control theory texts contain several chapters on modeling using ordinary differential equations and difference equations (see, for example, [FPEN05]). A classic book on the modeling of physical systems, especially mechanical, electrical and thermofluid systems, is Cannon [Can03]. The book by Aris [Ari94] is highly original and has a detailed discussion of the use of dimension-free variables. Two of the authors’ favorite books on modeling of biological systems are J. D. Murray [Mur04] and Wilson [Wil99].

Exercises

2.1 (Chain of integrators form) Consider the linear ordinary differential equation (2.7). Show that by choosing a state space representation with \( x_1 = y \), the dynamics can be written as

\[
A = \begin{bmatrix}
0 & 1 & 0 \\
0 & \ddots & \ddots & 0 \\
0 & \cdots & 0 & 1 \\
-a_n & -a_{n-1} & \cdots & -a_1
\end{bmatrix}, \quad B = \begin{bmatrix}
0 \\
0 \\
\vdots \\
1
\end{bmatrix}, \quad C = \begin{bmatrix} 1 & \cdots & 0 & 0 \end{bmatrix}.
\]

This canonical form is called the chain of integrators form.

2.2 (Inverted pendulum) Use the equations of motion for a balance system to derive a dynamic model for the inverted pendulum described in Example 2.2 and verify that for small \( \theta \) the dynamics are approximated by equation (2.10).

2.3 (Discrete-time dynamics) Consider the following discrete-time system

\[
x[k + 1] = Ax[k] + Bu[k], \quad y[k] = Cx[k],
\]

where

\[
x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad A = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \end{bmatrix}.
\]
In this problem, we will explore some of the properties of this discrete-time system as a function of the parameters, the initial conditions and the inputs.

(a) For the case when \( a_{12} = 0 \) and \( u = 0 \), give a closed form expression for the output of the system.

(b) A discrete system is in equilibrium when \( x[k + 1] = x[k] \) for all \( k \). Let \( u = r \) be a constant input and compute the resulting equilibrium point for the system. Show that if \( |a_{ii}| < 1 \) for all \( i \), all initial conditions give solutions that converge to the equilibrium point.

(c) Write a computer program to plot the output of the system in response to a unit step input, \( u[k] = 1, k \geq 0 \). Plot the response of your system with \( x[0] = 0 \) and \( A \) given by \( a_{11} = 0.5, a_{12} = 1 \) and \( a_{22} = 0.25 \).

2.4 (Keynesian economics) Keynes’ simple model for an economy is given by

\[
Y[k] = C[k] + I[k] + G[k],
\]

where \( Y, C, I \) and \( G \) are gross national product (GNP), consumption, investment and government expenditure for year \( k \). Consumption and investment are modeled by difference equations of the form

\[
C[k + 1] = aY[k], \quad I[k + 1] = b(C[k + 1] - C[k]),
\]

where \( a \) and \( b \) are parameters. The first equation implies that consumption increases with GNP but that the effect is delayed. The second equation implies that investment is proportional to the rate of change of consumption.

Show that the equilibrium value of the GNP is given by

\[
Y_e = \frac{1}{1 - a} (I_e + G_e),
\]

where the parameter \( 1/(1 - a) \) is the Keynes multiplier (the gain from \( I \) or \( G \) to \( Y \)). With \( a = 0.25 \) an increase of government expenditure will result in a fourfold increase of GNP. Also show that the model can be written as the following discrete-time state model:

\[
\begin{pmatrix}
C[k + 1] \\
I[k + 1]
\end{pmatrix} = \begin{pmatrix}
a & a \\
ab - b & ab
\end{pmatrix} \begin{pmatrix}
C[k] \\
I[k]
\end{pmatrix} + \begin{pmatrix}
a \\
ab
\end{pmatrix} G[k],
\]

\[
Y[k] = C[k] + I[k] + G[k].
\]

2.5 (Least squares system identification) Consider a nonlinear differential equation that can be written in the form

\[
\frac{dx}{dt} = \sum_{i=1}^{M} a_i f_i(x),
\]

where \( f_i(x) \) are known nonlinear functions and \( a_i \) are unknown, but constant, parameters. Suppose that we have measurements (or estimates) of the full state \( x \)
at time instants $t_1, t_2, \ldots, t_N$, with $N > M$. Show that the parameters $\alpha_i$ can be determined by finding the least squares solution to a linear equation of the form
\[ H\alpha = b, \]
where $\alpha \in \mathbb{R}^M$ is the vector of all parameters and $H \in \mathbb{R}^{N \times M}$ and $b \in \mathbb{R}^N$ are appropriately defined.

2.6 (Normalized oscillator dynamics) Consider a damped spring–mass system with dynamics
\[ m\ddot{q} + c\dot{q} + kq = F. \]
Let $\omega_0 = \sqrt{k/m}$ be the natural frequency and $\zeta = c/(2\sqrt{km})$ be the damping ratio.

(a) Show that by rescaling the equations, we can write the dynamics in the form
\[ \ddot{q} + 2\zeta\omega_0\dot{q} + \omega_0^2 q = \omega_0^2 u, \tag{2.35} \]
where $u = F/k$. This form of the dynamics is that of a linear oscillator with natural frequency $\omega_0$ and damping ratio $\zeta$.

(b) Show that the system can be further normalized and written in the form
\[ \frac{dz_1}{d\tau} = z_2, \quad \frac{dz_2}{d\tau} = -z_1 - 2\zeta z_2 + v. \tag{2.36} \]
The essential dynamics of the system are governed by a single damping parameter $\zeta$. The $Q$-value defined as $Q = 1/2\zeta$ is sometimes used instead of $\zeta$.

2.7 (Electric generator) An electric generator connected to a strong power grid can be modeled by a momentum balance for the rotor of the generator:
\[ J\frac{d^2\phi}{dt^2} = P_m - P_e = P_m - \frac{EV}{X} \sin \phi, \]
where $J$ is the effective moment of inertia of the generator, $\phi$ the angle of rotation, $P_m$ the mechanical power that drives the generator, $P_e$ is the active electrical power, $E$ the generator voltage, $V$ the grid voltage and $X$ the reactance of the line. Assuming that the line dynamics are much faster than the rotor dynamics, $P_e = VI = (EV/X) \sin \phi$, where $I$ is the current component in phase with the voltage $E$ and $\phi$ is the phase angle between voltages $E$ and $V$. Show that the dynamics of the electric generator has a normalized form that is similar to the dynamics of a pendulum with forcing at the pivot.

2.8 (Admission control for a queue) Consider the queuing system described in Example 2.10. The long delays created by temporary overloads can be reduced by rejecting requests when the queue gets large. This allows requests that are accepted to be serviced quickly and requests that cannot be accommodated to receive a rejection quickly so that they can try another server. Consider an admission control system described by
\[ \frac{dx}{dt} = \lambda u - \mu_{\text{max}} \frac{x}{x + 1}, \quad u = \text{sat}_{(0,1)}(k(r - x)), \tag{2.37} \]
where the controller is a simple proportional control with saturation (sat\(_{(a,b)}\)) is defined by equation (3.9)) and \( r \) is the desired (reference) queue length. Use a simulation to show that this controller reduces the rush-hour effect and explain how the choice of \( r \) affects the system dynamics.

2.9 (Biological switch) A genetic switch can be formed by connecting two repressors together in a cycle as shown below.

Using the models from Example 2.13—assuming that the parameters are the same for both genes and that the mRNA concentrations reach steady state quickly—show that the dynamics can be written in normalized coordinates as

\[
\frac{dz_1}{d\tau} = \frac{\mu}{1 + z_2^\theta} - z_1 - v_1, \quad \frac{dz_2}{d\tau} = \frac{\mu}{1 + z_1^\theta} - z_2 - v_2,
\]

where \( z_1 \) and \( z_2 \) are scaled versions of the protein concentrations and the time scale has also been changed. Show that \( \mu \approx 200 \) using the parameters in Example 2.13, and use simulations to demonstrate the switch-like behavior of the system.

2.10 (Motor drive) Consider a system consisting of a motor driving two masses that are connected by a torsional spring, as shown in the diagram below.

This system can represent a motor with a flexible shaft that drives a load. Assuming that the motor delivers a torque that is proportional to the current, the dynamics of the system can be described by the equations

\[
J_1 \frac{d^2 \varphi_1}{dt^2} + c \left( \frac{d\varphi_1}{dt} - \frac{d\varphi_2}{dt} \right) + k(\varphi_1 - \varphi_2) = k_I I,
\]

\[
J_2 \frac{d^2 \varphi_2}{dt^2} + c \left( \frac{d\varphi_2}{dt} - \frac{d\varphi_1}{dt} \right) + k(\varphi_2 - \varphi_1) = T_d.
\]

Similar equations are obtained for a robot with flexible arms and for the arms of DVD and optical disk drives.

Derive a state space model for the system by introducing the (normalized) state variables \( x_1 = \varphi_1, x_2 = \varphi_2, x_3 = \omega_1/\omega_0, \) and \( x_4 = \omega_2/\omega_0, \) where \( \omega_0 = \sqrt{k(J_1 + J_2)/(J_1 J_2)} \) is the undamped natural frequency of the system when the control signal is zero.