PREFACE

The age of modern control theory was ushered in at the launching of the first sputnik in 1957. This achievement of Soviet technology focused attention of scientists and engineers in general, and the automatic-control community in particular, eastward toward the USSR. By worldwide consensus, Moscow was the appropriate location for the First Congress of the International Federation of Automatic Control in 1960.

In turning their attention to the Soviet Union, control system scientists and engineers discovered a different approach to control theory than the approach with which they were familiar. Differential equations replaced transfer functions for describing the dynamics of processes; stability was approached via the theory of Liapunov instead of the frequency-domain methods of Bode and Nyquist; optimization of system performance was studied by the special form of the calculus of variations developed by Pontryagin instead of by the Wiener-Hopf methods of an earlier era.

In a few years of frenzied effort, Western control theory had absorbed and mastered this new "state-space" approach to control system analysis and design, which has now become the basis of much of modern control theory.

State-space concepts have made an enormous impact on the thinking of those control scientists and engineers who work at the frontiers of technology. These concepts have also been used with notable success in a number of important high-technology projects—the U.S. Apollo project was a highly visible example. Nevertheless, the majority of control systems implemented at the present time are designed by methods of an earlier era.

Many control engineers schooled in the earlier methods have felt that the modern state-space approach is mathematically esoteric and more suited to advanced graduate research than to the design of practical control systems. I can sympathize with the plight of the engineer who has waded through a morass of mathematics with the hope of learning how to solve his practical problem only to return empty-handed; I have been there too. One thesis of this book is

xi

xii PREFACE

that state-space methods can be presented in a style that can be grasped by the engineer who is more interested in using the results than in proving them. Another thesis is that the results *are useful*. I would even go so far as to say that if one had to choose between the frequency-domain methods of the past and the state-space methods of the present, then the latter are the better choice. Fortunately, one does not need to make the choice: both methods are useful and complement each other. Testimony to my continued faith in frequency-domain analysis is a long chapter, Chap. 4, which presents some of the basic methods of that approach, as a review and for those readers who may not be knowledgeable in these methods.

This book is addressed not only to students but also to a general audience of engineers and scientists (e.g., physicists, applied mathematicians) who are interested in becoming familiar with state-space methods either for direct application to control system design or as a background for reading the periodical literature. Since parts of the book may already be familiar to some of these readers, I have tried, at the expense of redundancy, to keep the chapters reasonably independent and to use customary symbols wherever practical. It was impossible, of course, to eliminate all backward references, but I hope the reader will find them tolerable.

Vectors and matrices are the very language of state-space methods; there is no way they can be avoided. Since they are also important in many other branches of technology, most contemporary engineering curricula include them. For the reader's convenience, however, a summary of those facts about vectors and matrices that are used in the book is presented in the Appendix.

Design is an interplay of science and art—the instinct of using exactly the right methods and resources that the application requires. It would be presumptuous to claim that one could learn control system design by reading this book. The most one could claim is to have presented examples of how state-space methods could be used to advantage in several representative applications. I have attempted to do this by selecting fifteen or so examples and weaving them into the fabric of the text and the homework problems. Several of the examples are started in Chap. 2 or 3 and taken up again and again later in the book. (This is one area where backward references are used extensively.) To help the reader follow each example on its course through the book, an applications index is furnished (pages 503 to 505). Many of the examples are drawn from fields I am best acquainted with: aerospace and inertial instrumentation. Many other applications of state-space methods have been studied and implemented: chemical process control, maritime operations, robotics, energy systems, etc. To demonstrate the wide applicability of state-space methods, I have included examples from some of these fields, using dynamic models and data selected from the periodical literature. While not personally familiar with these applications, I have endeavored to emphasize some of their realistic aspects.

The emphasis on application has also motivated the selection of topics. Most of the attention is given to those topics that I believe have the most practical utility. A number of topics of great intrinsic interest do not, in my judgment, have the practical payoff commensurate with the effort needed to learn them. Such topics have received minimal attention. Some important concepts are really quite simple and do not need much explaining. Other concepts, although of lesser importance, require more elaborate exposition. It is easy to fall into the trap of dwelling on subjects in inverse proportion to their significance. I have tried to avoid this by confining the discussion of secondary topics to notes at the end of each chapter, with references to the original sources, or to the homework problems.

Much of practical engineering design is accomplished with the aid of computers. Control systems are no exception. Not only are computers used for on-line, real-time implementation of feedback control laws-in applications as diverse as aircraft autopilots and chemical process controls-but they are also used extensively to perform the design calculations. Indeed, one of the major advantages of state-space design methods over frequency-domain methods is that the former are better suited to implementation by digital computers. Computer-aided design, however, creates a dilemma for the author. On the one hand, he wants to make the concepts understandable to a reader who doesn't have a computer. On the other hand the full power of the method is revealed only through applications that require the use of a computer. My decision has been a compromise. I have tried to keep the examples in the text simple enough to be followed by the reader, at least part of the way, without recourse to a computer for numerical calculation. There are a number of homework problems, however, some of which continue examples from the text, for which a computer is all but essential.

The reader is certainly not expected to write the software needed to perform the numerical calculations. During the past several years a number of organizations have developed software packages for computer-aided control system design (CACSD). Such software is available for mainframes and personal computers at prices to suit almost any budget and with capabilities to match. Several of these packages would be adequate for working the homework problems that require a computer and for other applications. Anyone with more than a casual interest in state-space methods would be well advised to consider acquiring and maintaining such software.

The education of most engineers ends with the bachelor's or master's degree. Hence, if state-space methods are to be widely used by practicing engineers, they must be included in the undergraduate or first-year graduate curriculum—they must not be relegated to advanced graduate courses. In support of my commitment to state-space methods as a useful tool for practicing engineers, I have endeavored to teach them as such. A number of years ago I presented some introductory after-hours lectures on this subject to fellow employees at the Kearfott Division of The Singer Company. These lectures served as the basis of an undergraduate elective I have been teaching at the Polytechnic Institute of New York. For want of a more suitable textbook, I have been distributing hard copies of the overhead transparencies used in the lectures. It occurred to me that

xiv PREFACE

the material I had assembled in these overhead transparencies was the nucleus of the book I had needed but had been unable to locate. And so I embarked upon this project.

It is a pleasure to acknowledge the contributions made by a number of individuals to this project. Most of the manuscript was patiently and expertly typed by Win Griessemer. Additional typing and editorial assistance, not to mention moral support, was provided when needed most by my wife and daughters, to whom this book is dedicated. My associates at The Singer Company, Dave Haessig, Appa Madiwale, Jack Richman, and Doug Williams between them read most of the manuscript, found many errors large and small, and offered a number of helpful suggestions. A preliminary version of this book was used as a text for my undergraduate course at the Polytechnic Institute of New York and for a similar course, taught by Professor Nan K. Loh, at Oakland University (Michigan). The students in these courses provided additional feedback used in the preparation of the final manuscript.

The vision of this book has long been in my mind's eye. To all those named above, and others not named but not forgotten, who have helped me realize this vision, my gratitude is boundless.

Bernard Friedland

CHAPTER ONE

82700

MOULINA

FEEDBACK CONTROL

1.1 THE MECHANISM OF FEEDBACK

No mechanism in nature or technology is more pervasive than the mechanism of feedback.

By the mechanism of feedback a mammal maintains its body temperature constant to within a fraction of a degree even when the ambient temperature fluctuates by a hundred degrees or more.

Through feedback the temperature in an oven or in a building is kept to within a fraction of a degree of a desired setting even though the outside temperature fluctuates by 20 or 30 degrees in one day.

An aircraft can maintain its heading and altitude and can even land, all without human intervention, through feedback.

Feedback is the mechanism that makes it possible for a biped to stand erect on two legs and to walk without falling.

When the Federal Reserve Bank exercises its controls in the interest of stabilizing the national economy, it is attempting to use feedback.

When the Mayor of New York City asks, "How'm I doing?" he is invoking the mechanism of feedback.

Hardly a process occurring in nature or designed by man does not, in one way or another, entail feedback.

Because feedback is ubiquitous, it is taken for granted except when it is not working properly: when the volume control of a public address system in an auditorium is turned up too high and the system whistles; then everyone becomes aware of "feedback." Or when the thermostat in a building is not working properly and all the occupants are freezing, or roasting.



Figure 1.1 Open-loop control, lnput \bar{u} is selected to produce desired output \bar{v} .

To get an appreciation of the mechanism of feedback, suppose that there is a process H that we wish to control. Call the input to the process u and the output from the process y. Suppose that we have a complete description of the process: we know what the output y will be for any input. Suppose that there is one particular input, say \bar{u} , which corresponds to a specified, desired output, say $\bar{\nu}$. One way of controlling the process so that it produces the desired output \bar{y} is to supply it with the input \bar{u} . This is "open-loop control." (Fig. 1.1.) A billiard player uses this kind of control. With an instinctive or theoretical knowledge of the physics of rolling balls that bounce off resilient cushions, an expert player knows exactly how to hit the cue ball to make it follow the planned trajectory. The blow delivered by the cue stick is an open-loop control. In order for the ball to follow the desired trajectory, the player must not only calculate exactly how to impart that blow, but also to execute it faultlessly. Is it any wonder that not everyone is an expert? On the other hand, suppose one wants to cheat at billiards by putting some kind of sensor on the cue ball so that it can always "see" the target-a point on another ball or a cushion-and by some means can control its motion-"steer"-to the target. Finally, put a tiny radio in the ball so that the cheater can communicate the desired target to the cue ball. With such a magic cue ball the cheater cannot but win every game. He has a cue ball that uses the mechanism of feedback.

The magic cue ball has two of the characteristics common to every feedback system: a means of monitoring its own behavior ("How'm I doing") and a means of correcting any sensed deviation therefrom. These elements of a feedback control system are shown in Fig. 1.2. Instead of controlling the output of the process by picking the control signal \bar{u} which produces the desired \bar{y} , the control signal u is generated as a function of the "system error," defined as the difference between the desired output \bar{y} and the actual output y

$$e = \bar{y} - y \tag{1.1}$$



Figure 1.2 Feedback control system. Input u is proportional to difference between desired and actual output.

FEEDBACK CONTROL 3

NUBBRID PRIDULING

This error, suitably amplified, as shown by the output of the box labeled "amplifier," is the input to the process.

Suppose that the operation of the process under control can be represented by a simple algebraic relation

$$y = Hu \tag{1.2}$$

and that the amplifier can similarly be described

$$u = Ke \tag{1.3}$$

Combine (1.1), (1.2), and (1.3) into the single relation

$$y = HKe = HK(\bar{y} - y)$$

Solve for y and obtain

$$y = \frac{HK}{1 + HK} \bar{y} \tag{1.4}$$

Although the output y is not exactly equal to the desired output \bar{y} , if the amplifier "gain" K is large enough (i.e., $HK \gg 1$) then

$$y \simeq \bar{y}$$
 (1.5)

We can make the actual output y approach the desired output as closely as we wish simply by making the gain K large enough. Moreover, this result holds, for any desired output! We don't have to know \bar{y} in advance as we did in determining the open-loop control \bar{u} . And, even more remarkably, this result holds independent of the process—it doesn't matter what H is. In fact H can even change over the course of time without affecting the basic result. These are among the wonders of feedback and help to explain why it is so useful.

Unfortunately, nature is not as simple as the above analysis would suggest; if it were, there would be no need for this book. The problem is that the process whose input is u and whose output is y cannot be represented by an algebraic equation as simple as (1.2). Because of the process dynamics, the relationship between the output and the input is much more complex than (1.2).

The effect of dynamics on the behavior of a feedback system is easily illustrated by a simple example. Suppose that the output of system H is an exact replica of the input, except delayed by a small amount of time, say τ :

$$y(t) = u(t - \tau) \tag{1.6}$$

for any input u(t). (See Fig. 1.3.) We assume that (1.1) and (1.2) continue to hold for every time t. Then

$$u(t-\tau) = Ke(t-\tau) = K[\bar{y}(t-\tau) - y(t-\tau)]$$
(1.7)

Substitute (1.7) into (1.6) to obtain

$$y(t) = K[\bar{y}(t-\tau) - y(t-\tau)]$$
(1.8)



Figure 1.3 Example of input and output of a process in which the output is an exact but delayed replica of the input.

This is an example of a "difference equation" and describes how y(t) evolves as time t increases. Difference equations are a common way of describing the dynamic behavior of discrete-time (sampled-data) systems but they are not studied extensively in this book. This equation, however, is so simple that it can be solved without any theory.

Suppose the desired output is a "unit step" as shown in Fig. 1.4(a):

$$\bar{y}(t) = \begin{cases} 0 & \text{for } t < 0\\ 1 & \text{for } t > 0 \end{cases}$$
(1.9)

and also suppose that y(t) = 0 for t < 0. Then, by (1.8) and also by looking at Fig. 1.3, we can see that there is no output for the first τ units of time. But the input to the process

$$u(t) = K(1-0) = K$$
 for $0 < t < \tau$

After an interval of τ units of time the output starts to appear as shown in Fig. 1.4(b)

$$y(t) = K \qquad \text{for } \tau < t < 2\tau$$

For the next τ units of time the input to the process is

$$u(t) = K(1-K) = K - K^2$$
 for $\tau < t < 2\tau$

This is the value of y(t) for the next τ units of time, i.e., for $2\tau < t < 3\tau$. Proceeding in this fashion we see that

$$y(t) = K - K^{2} + K^{3} + \dots + (-1)^{n-1} K^{n-1} \qquad \text{for } n\tau < t < (n+1)\tau \quad (1.10)$$

FEEDBACK CONTROL 5

N VI627 20

LAND PLAN



Figure 1.4 Response of feedback control system to $\bar{y} = 1$ when output is a replica of input delayed by τ with gain $K = \frac{1}{2}$. (a) Desired output $\bar{y}(t)$; (b) Input to process u(t); (c) Output from process y(t).

If K is less than 1, then (1.10) implies that y(t) will eventually converge to a limit:

$$\lim_{t \to \infty} y(t) = K - K^2 + K^3 - K^4 + \dots = \frac{K}{1 + K}$$
(1.11)

If K is exactly equal to 1, the output y(t) will flip between 0 and 1 indefinitely. And if K > 1, the output will flip between positive and negative values, ever increasing in amplitude, and ultimately become infinite.

Thus we see that the amplification factor (or gain) K of the amplifier cannot be made as large as 1 if we want the output to stabilize. (Also, as K approaches 1, the output is only half the value of the input. This can be corrected, however, by multiplying the desired output by (1 + K) before comparing it with the actual output.)

Our earlier discussion suggests that we would like an amplifier gain approaching infinity, but here we see that we cannot even make the gain as large as unity without causing the system to break into unstable oscillation. All because the output of the process is delayed by a small amount of time—an *arbitrarily small* amount of time. In every real process there is always some delay. Does this mean that feedback control cannot be used in any real process? The answer of course is no. And the reason is twofold. First, while it is true that there is some amount of delay in any physical process, the output is rarely simply a delayed replica of the input. The output will also not look exactly like the input. The time-distortion of the output is a benefit for control system design. Second, the black box which we called an amplifier, with gain K, is usually more than just an amplifier. It also changes the shape of the signal that passes through it. The amplifier is a "compensator," which the control system engineer, knowing the dynamic characteristics of the process H, designs to achieve favorable operation.

By proper design of the compensator it is generally possible to achieve satisfactory closed-loop performance for complex, even nasty processes. For example, it is possible to "close the loop" around a process H, which is itself unstable, in such a way that the closed-loop system not only is stable, but that the output y faithfully tracks the desired output \bar{y} .

1.2 FEEDBACK CONTROL ENGINEERING

Feedback control engineering may be regarded as the conscious, intentional use of the mechanism of feedback to control the behavior of a dynamic process.

The course that a typical feedback system design follows is exemplified by the hypothetical magic cue ball of the previous section. Suppose one has a client who comes prepared to pay the expense of the design and construction of such a technical marvel. How would one proceed?

The performance requirements are easy to imagine. To escape detection, the entire system must fit inside a hollowed-out ball and its weight and inertia must exactly equal those of the material removed. If the cue ball is to be able to home-in on its target, it should be able to sense its position relative to the target. How can this be accomplished? Perhaps a miniature infrared sensor? Will the sensor be able to discriminate between the actual target and another cue ball that resembles the target? Perhaps the billiard table can have a hidden means of generating an electric or magnetic field that is altered by the presence of the balls and this information can be transmitted to the cue ball.

Suppose we have tentatively solved the problem of sensing the motion of the cue ball. Next we need some means to alter its trajectory. Can we use tiny, barely visible gas jets? Perhaps we can use a movable weight inside the cue ball which will displace the center of mass from the geometric center and hence, with the aid of gravity, create moments which, when combined with friction, can

N 11627 20

change the trajectory. Maybe we can use tiny, almost imperceptible bumps on the surface that can be moved to change the course of the ball.

Conception of the means of measuring the behavior of the process—the cue ball—and affecting or altering its behavior is the first stage of control system design. Without a doubt, this is the stage that requires the greatest degree of inventiveness and understanding of what can be achieved at the current level of technology and at what price.

If the project has not been abandoned at the first stage for want of suitable technological means, the next step is to acquire or design the sensors that have been chosen to measure the motion of the vehicle relative to the target and the actuation means that have been selected to alter the motion of the cue ball.

After the hardware is all selected, the final stage of the design is begun. This is the stage in which it is decided how the feedback loop or loops are to be closed: how the data from the sensor or sensors are to be processed before being sent to the actuator. It is at this stage that the designer decides what the block box labeled "amplifier" must really do in order for the closed loop system to operate properly. This step is the design of the "control law" or "control algorithm."

This last stage of control system design is the entire content of control theory. By the time control theory enters the picture, the system concept has already been established and the control hardware has already been selected. The whole apparatus of control theory, it would appear, deals with only a small, insignificant fraction of the overall problem. In this light, the effort devoted to the development of control theory—the subject matter of this book—hardly seems worth the effort.

The magic cue ball design problem, however, does not represent the typical design problem. Although it is true enough that the control concept must be defined and the hardware must be selected for every control system design, not every design requires such inventiveness. In most cases, the process to be controlled is only slightly different from yesterday's. Today's control hardware is only slightly different from yesterday's, probably better (more accurate, cheaper, and more reliable). Hence the first design steps are taken almost unconsciously. The engineer, not without justification, forgets about the first two steps and believes that the control system design begins at the point that it is almost over.

If today's process and control hardware are not changed much from yesterday's, why can't one simply use yesterday's control law? Oftentimes, one can. Most control laws are probably designed by this very method: Take yesterday's control law and modify its parameters to account for the difference between yesterday's hardware and today's.

But the procedure is not always satisfactory. The new process may not be sufficiently similar to the old one. The new control hardware, although improved (say digital instead of analog) may have different characteristics that cannot be overlooked. And finally, the customer may demand a higher level of performance than yesterday's system was able to deliver.

1.3 CONTROL THEORY BACKGROUND

This book is concerned with the third and final stage of control system engineering—the stage in which the dynamic characteristics of the compensator are designed, after the control concept has been established, after the hardware (sensors and actuators) have been selected, after the performance requirements have been determined.

This aspect of control system engineering is generally called control "*theory*." The term "theory" is appropriate for several reasons. First, it is essentially mathematical in content, and mathematics is often equated to theory. Second, it deals not with the actual devices but with their idealized (theoretical, i.e., mathematical) models. Third, it constitutes a systematic body of knowledge: theorems, design algorithms, graphical methods, and the like which can be applied to control systems independent of the specific technology used in the practical implementation.

The history of control theory can be conveniently divided into three periods. The first, starting in prehistory and ending in the early 1940s, may be termed the *primitive* period. This was followed by a *classical* period, lasting scarcely 20 years, and finally came the *modern* period which includes the content of this book.

The term *primitive* is used here not in a pejorative sense, but rather in the sense that the theory consisted of a collection of analyses of specific processes by mathematical methods appropriate to, and often invented to deal with, the specific processes, rather than an organized body of knowledge that characterizes the classical and the modern period.

Although feedback principles can be recognized in the technology of the Middle Ages and earlier, the intentional use of feedback to improve the performance of dynamic systems was started at around the beginning of the industrial revolution in the late 18th and early 19th centuries. The benchmark development was the ball-governor invented by James Watt to control the speed of his steam engine. Throughout the first half of the 19th century, engineers and "mechanics" were inventing improved governors. The theoretical principles that describe their operation were studied by such luminaries of 18th and 19th century mathematical physics as Huygens,[1] Hooke,[2] Airy,[3] and Maxwell.[4] By the mid 19th century it was understood that the stability of a dynamic system was determined by the location of the roots of the algebraic characteristic equation. Routh[5] in his Adams Prize Essay of 1877 invented the stability algorithm that bears his name.

Mathematical problems that had arisen in the stability of feedback control systems (as well as in other dynamic systems including celestial mechanics) occupied the attention of early 20th century mathematicians Poincaré and Liapunov, both of whom made important contributions that have yet to be superseded.

Development of the gyroscope as a practical navigation instrument during the first quarter of the 20th century led to the development of a variety of

NU62720

MADUL MA

autopilots for aircraft (and also for ships). Theoretical problems of stabilizing these systems and improving their performance engaged various mathematicians of the period. Notable among them was N. Minorsky[6] whose mimeographed notes on nonlinear systems was virtually the only text on the subject before 1950.

The classical period of control theory begins during World War II in the Radiation Laboratory of the Massachusetts Institute of Technology. (See Note [.1.) The personnel of the Radiation Laboratory included a number of engineers, physicists, and mathematicians concerned with solving engineering problems that arose in the war effort, including radar and advanced fire control systems. The laboratory that was assigned problems in control systems included individuals knowledgeable in the frequency response methods, developed by people such as Nyquist and Bode for communication systems, as well as by engineers familiar with other techniques. Working together, they evolved a systematic control theory which is not tied to any particular application. Use of frequency-domain (Laplace transform) methods made possible the representation of a process by its transfer function and thus permitted a visualization of the interaction of the various subsystems in a complex system by the interconnection of the transfer functions in the block diagram. The block diagram contributed perhaps as much as any other factor to the development of control theory as a distinct discipline. Now it was possible to study the dynamic behavior of a hypothetical system by manipulating and combining the black boxes in the block diagram without having to know what goes on inside the boxes.

The classical period of control theory, characterized by frequency-domain analysis, is still going strong, and is now in a "neoclassical" phase—with the development of various sophisticated techniques for multivariable systems. But concurrent with it is the *modern* period, which began in the late 1950s and early 1960s.

State-space methods are the cornerstone of modern control theory. The essential feature of state-space methods is the characterization of the processes of interest by differential equations instead of transfer functions. This may seem like a throwback to the earlier, primitive, period where differential equations also constituted the means of representing the behavior of dynamic processes. But in the earlier period the processes were simple enough to be characterized by a *single* differential equation of fairly low order. In the modern approach the processes are characterized by systems of coupled, first-order differential equations. In principle there is no limit to the order (i.e., the number of independent first-order differential equations) and in practice the only limit to the order is the availability of computer software capable of performing the required calculations reliably.

Although the roots of modern control theory have their origins in the early 20th century, in actuality they are intertwined with the concurrent development of computers. A digital computer is all but essential for performing the calculations that must be done in a typical application. Only in the simplest examples can the calculations be performed without a digital computer. The

fact that calculations for simple applications can be done manually can sometimes be misleading, because the design for such simple applications can usually be achieved more efficiently by classical frequency-domain methods. State-space methods prove their mettle in applications which are intractable by classical methods.

Digital computers of even modest capability can crunch out the numerical calculations of the design for a complicated system in a few seconds or minutes. It is thus very easy to arrive at a design which is correct numerically but not practical. (There is no inherent reason why this can't also happen with a design based on classical methods. But because of the labor entailed in achieving the design, the engineer is more likely to check intermediate results for reasonability rather than to wait for the final design to emerge as a unit.) The realization that there may be practical problems with a computer-aided design ought to make the designer especially cautious: both in making certain that the computer has good data to begin with, i.e., a proper model of the process to be controlled, and in testing the proposed design by all appropriate means including simulation.

1.4 SCOPE AND ORGANIZATION OF THIS BOOK

The vision of the early pioneers of modern control theory was that it would provide a single, unified framework for all feedback control systems: linear and nonlinear, continuous-time, and discrete-time, fixed and time-varying. That vision is a chimera. A few results of broad generality have been achieved, but for the most part the vaunted general theory has been achieved only for linear systems, and furthermore, the required calculations can be performed only for time-invariant, linear systems. This is nevertheless no mean accomplishment, because the theory that does exist is still able to cope with any design problem that the classical theory can cope with, because the frequency-domain approach is entirely predicated on linear, time-invariant models.

Being an introduction to state-space methods, this book does not go beyond systems that can be characterized by linear, time-invariant models. (The sole exception is a missile guidance system which has time-varying dynamics that are so simple that they can easily be handled without the need for any special theory.)

The first few chapters are intended as an introduction to the use of state-space methods for characterizing the behavior of dynamic systems. In particular, in Chap. 2, we learn how linear state-space models can be set up for various kinds of physical processes, and in Chap. 3 we study the basic properties of such models: such things as the state-transition matrix, the resolvent, the characteristic equation. Although the properties of linear, time-invariant systems can be gleaned without use of the Laplace transform, they are more readily obtained through its use. Since most readers of this book are familiar with the basic theory of Laplace transforms, we see no reason for not

making use of them. We also see no reason for abandoning classical, frequencydomain methods and the insights they provide. Hence, in Chap. 4, we provide a review of frequency-domain analysis, emphasizing where possible the connection with state-space methods. Notwithstanding the length of Chap. 4, it is still only an overview; the reader is assumed to be already somewhat familiar with the material or prepared to consult one of the standard textbooks in the field to gain a more comprehensive understanding.

Controllability and observability theory, one of the earliest unique achievements of modern control theory, is the subject of Chap. 5.

The first five chapters set the stage for the use of state-space methods for control system design. These are followed by three which show how state-space methods can be used in design. Chapter 6 is concerned with design of controllers that use "full-state" feedback, i.e., design under the assumption that all of the state variables are accessible to measurement, if needed for the control law. This is an unrealistic assumption, and Chap. 7 shows how to design observers which are dynamic systems, the inputs to which are the measured inputs and outputs of the process under control. The state of the observer is an estimate of the state of the process under control. Chapter 8, which concludes the three introductory chapters on design, shows how the full-state feedback control of Chap. 6 can be combined with the observer of Chap. 7, to finally provide the design of a compensator which is typically the goal of the control system designer. Chapter 8 is also concerned with the robustness of compensators designed by the methods of these three chapters: it addresses the question of how well the compensator will work if the mathematical model used in the design is not exactly matched to the actual physical process.

Although the methods of Chaps. 6 through 8 constitute a set of procedures for designing compensators for controllable and observable processes, they do not of themselves arrive at *optimum* designs. Optimization of these designs is the subject of Chaps. 9 through 11.

In Chap. 9, we learn how to optimize the gain matrix of the full-state control law by choosing it to minimize a quadratic integral performance criterion. The weighting matrices in the integral are putatively chosen to correspond, at least approximately, to physical performance requirements. Computing the gain matrix is shown to entail solving for the appropriate (matrix) root of a matrix quadratic equation which has come to be known as the *algebraic Riccati equation*. Numerical solution of this equation is a job for the digital computer.

The selection of the optimum gain for the observer is formulated as a statistical problem: to find the observer gain matrix that minimizes the estimation error variance under the hypothesis that the process is excited by white noise with a known spectral density matrix and that the observations are corrupted with white noise also with known spectral density. The resulting observer gain matrix is also the solution to an algebraic Riccati equation which has a structure quite similar to that of the algebraic Riccati equation for the optimum controller. The theory for the optimum observer, also known as a

Kalman filter, is developed in Chap. 11. A minimal overview of the statistical prerequisites to Chap. 11 is presented in Chap. 10.

Matrices and vectors are the very language of state-space methods. By now they are so commonplace in every branch of technology, that we can hardly imagine a reader unfamiliar with them. Nevertheless we have included an appendix in which the basic facts about matrices are summarized, without an attempt at proofs. If the reader wants proofs, there are innumerable texts available for that purpose.

One of the objectives of this book has been to illustrate the use of state-space methods in various aspects of system analysis and design by means of examples that have some relationship to real-world applications. In line with that objective a number of "running examples" are provided. Each example occurs in several places in the text: to exemplify development of the model of a system an example may appear in Chap. 2. The same example may appear again to illustrate the calculation of open-loop response, and in various aspects of control system design. References to earlier and later appearances of the same example are given each time an example reappears. In addition, an index of the examples is given at the end of the book. By use of this index, the reader should be able to locate all references to an example and thereby trace the course of its development through the book. Thus each example constitutes something of a "case study." Some of the examples are the subject of homework problems: the reader thereby actively participates in the development of the case study.

NOTES

Note 1.1 Historical antecedents

Before World War II, feedback control systems were largely mechanical. The feedback paths, not generally identified as such, were implemented by means of ingenious combinations of springs, dashpots, pneumatic devices, and similar gadgets. The electrical components that were used were magnets and perhaps resistors. Almost every new control system represented a genuine invention and many were in fact patented. Nothwithstanding the ingenuity that these inventions required[7, 8] the variety of functions that could be achieved with these devices was (and still is) extremely limited. Thus a mathematical theory of the function that a feedback compensator must perform would have been of little practical value, since no means of implementing the function was available. Electronic technology of the era was represented by large vacuum tubes enclosed in fragile glass envelopes, massive inductors and capacitors, and similar bulky and unreliable hardware. A few electrical components were used in the 1920s and 1930s,[9] but a control system would very likely have been the object of ridicule. Before World War II, the only industry with any serious interest in electronics was the communications industry—radio and telephony.

Electronic technology underwent a major transformation during the war. Electronic components (i.e., tubes) became smaller and more reliable, and the functions that electronic systems were able to perform became more sophisticated as a result of concerted efforts by scientists and engineers and mathematicians working together in the war effort. A notable wartime development, among others, was radar.

FEEDBACK CONTROL 13

At the Radiation Laboratory, established at the Massachusetts Institute of Technology in 1940 to aid in the war effort, one of the technical groups was concerned with control— "servomechanism"—problems. Members of the group included physicists, communication engineers, mechanical engineers, mathematicians, and technicians. Each member brought different insights to the problems assigned to the group. The resulting collaboration laid the groundwork for the second, *classical* phase of control technology, in which the frequency-domain methods of communication engineering (transfer function analysis, Bode and Nyquist diagrams, and the like) were applied to the analysis and design of control systems.

After the war, the results of the work at the Radiation Laboratory were published in a multivolume series. The volume by James, Nichols, and Phillips[10] constituted the exposition of the classical frequency-domain methodology developed at that laboratory.

With the war concluded, research in control theory was continued along these lines at a number of universities. One of the centers of research in control theory was at Columbia University. Under the leadership of John R. Ragazzini, much of the classical (i.e., Z transform) theory for sampled-data systems was developed there during the decade of the 1950s. Into the hospitable environment that Ragazzini fostered at Columbia was welcomed an iconoclastic young graduate student who preached against the frequency-domain methods and taught a new doctrine: state-space. That student was Rudolf E. Kalman.

Kalman argued with increasing success that the frequency-domain methods developed for communication systems were not the most appropriate for control systems: the methods were not readily adaptable to time-varying and nonlinear systems and even for linear, time-invariant systems they dealt with the wrong problems. Moreover, Kalman taught, the classical methods of analysis and design obscured the physical nature of the dynamic variables which the state-space methods preserved.

The ranks of adherents to Kalman's state-space approach swelled during the decade of the 1960s and the modern era of control theory thus became firmly established. But not everyone was persuaded that frequency-domain methods had been superseded. Debates, sometimes acrimonious, over the merits of the two approaches, which started then, continue unto the present time.

REFERENCES

- 1. Huygens, G., Horologium Oscillatorium (1673) in Oeuvres, complètes, Nijhoff, Amsterdam, vol. 17-18, 1932.
- 2. Hooke, R., "Lampas, or Description of Some Mechanical Improvements of Lamps and Waterpones Together with Some Other Physical and Mechanical Discoveries," *Proc. Royal Society (London)*, vol. 8, 1831.
- 3. Airy, G. B., "On the Regulator of the Clockwork for Effecting Uniform Movement of the Equatoreals," *Memoirs, Royal Astronomical Society*, vol. 11, 1840, pp. 249-267.
- 4. Maxwell, J. C., "On Governors," Philosophical Magazine, vol. 35, 1868, pp. 385-398.
- 5. Routh, E. J., A Treatise on the Stability of a Given State of Motion, Macmillan & Co., London, 1877.
- 6. Minorsky, N., Nonlinear Oscillations, D. Van Nostrand, New York, 1962.
- 7. Fuller, A. T., "The Early Development of Control Theory," Trans. ASME (J. Dynamic Systems, Measurement & Control), vol. 98G, no. 2, June 1976, pp. 109-118.
- 8. Fuller, A. T., "The Early Development of Control Theory, II," Trans. ASME (J. Dynamic Systems, Measurement & Control), vol. 98G, no. 3, September 1976, pp. 224-235.
- Oppelt, W., "A Historical Review of Autopilot Development, Research, and Theory in Germany," Trans. ASME (J. Dynamic Systems, Measurement & Control), vol. 98G, no. 3, September 1976, pp. 215-223.
- 10. James, H. M., Nichols, N. B., and Phillips, R. S., *Theory of Servomechanisms* (MIT Radiation Laboratory Series, vol. 25), McGraw-Hill Book Co., New York, 1947.

CHAPTER **TWO**

STATE-SPACE REPRESENTATION OF DYNAMIC SYSTEMS

2.1 MATHEMATICAL MODELS

The most important task confronting the control system analyst is developing a mathematical model of the process of interest. In many situations the essence of the analytical design problem is in the modeling: once that is done the rest of the analysis falls quickly into place.

The control system engineer is often required to deal with a system having a number of subsystems the physical principles of which depend on entirely different types of physical laws. A chemical process, for example, may comprise a chemical reactor, the dynamics of which are the subject of chemical kinetic theory, a heat exchanger which is governed by thermodynamic principles, and various valves and motors the dynamics of which depend on the physics of mechanical and electrical systems. The control of a typical aircraft entails an understanding of the interaction between the airframe governed by principles of aerodynamics and structural dynamics, the actuators which are frequently hydraulic or electrical, and the sensors (gyroscopes and accelerometers) which operate under laws of rigid body dynamics. And, if the human pilot of the aircraft is to be considered, aspects of physiology and psychology enter into the analysis.

One of the attractions of control system engineering is its interdisciplinary content. The control system engineer sees the "big picture" in the challenge to harmonize the operation of a number of interconnected subsystems, each of which operates under a different set of laws. But at the same time the control system engineer is almost totally dependent on the other disciplines. It is simply impossible to gain a sufficient understanding of the details of each of the

14

subsystems in a typical control process without the assistance of individuals having an intimate understanding of these subsystems. These individuals often have the knowledge that the control system analyst requires, but are not accustomed to expressing it in the form that the analyst would like to have it. The analyst must be able to translate the information he receives from others into the form he needs for his work.

The analyst needs *mathematical models* of the processes in the system under study: equations and formulas that predict how the various devices will behave in response to the inputs to these devices. From the viewpoint of the systems analyst each device is the proverbial "black box," whose operation is governed by appropriate mathematical models. The behavior of the overall process is studied and controlled by studying the interaction of these black boxes.

There are two modeling and analysis approaches in customary use for linear systems: the transfer-function or frequency-domain approach, to be discussed in Chap. 4, and the state-space approach which is the subject of the present chapter.

The feature of the state-space approach that sets it apart from the frequency-domain approach is the representation of the processes under examination by systems of first-order differential equations. This method of representation may appear novel to the engineer who has become accustomed to thinking in terms of transfer functions, but it is not at all a new way of looking at dynamic systems. The state-space is the mode of representation of a dynamic system that would be most natural to the mathematician or the physicist. Were it not that much of classical control theory was developed by electrical engineers, it is arguable that the state-space approach would have been in use much sooner.

State-space methods were introduced to the United States engineering community through the efforts of a small number of mathematically oriented engineers and applied mathematicians during the late 1950s and early 1960s. The spiritual father of much of this activity was Professor Solomon Lefschetz who organized a mathematical systems research group at the Research Institute of Advanced Studies (RIAS) in Baltimore, Md. Lefschetz, already a worldfamous mathematician, brought together a number of exceptionally talented engineers and mathematicians committed to the development of mathematical control theory. At Columbia University another group, under the aegis of Professor J. R. Ragazzini, and including R. E. Kalman and J. E. Bertram among others, was also at work developing the foundations of modern control theory.

In the Soviet Union there was less of an emphasis on transfer functions than on differential equations. Accordingly, many of the earliest uses of state-space methods were made by investigators in the Soviet Union. Much of the activity in the United States during the late 1950s entailed translation of the latest Russian papers into English. The Moscow location of the First Congress of the International Federation of Automatic Control (IFAC) in 1960 was entirely appropriate, and provided the first major opportunity for investigators from all over the world to meet and exchange ideas. Although the IFAC



congress was concerned with components and applications as well as with control theory, much of the interest of the meeting was on the newest theoretical developments.

2.2 PHYSICAL NOTION OF SYSTEM STATE

The notion of the state of a dynamic system is a fundamental notion in physics. The basic premise of newtonian dynamics is that the future evolution of a dynamic process is entirely determined by its present state. Indeed we might consider this premise as the basis of an abstract definition of the state of a dynamic system:

The state of a dynamic system is a set of physical quantities, the specification of which (in the absence of external excitation) completely determines the evolution of the system.

The difficulty with this definition, as well as its major advantage, is that the *specific* physical quantities that define the system state are not unique, although their number (called the system *order*) is unique. In many situations there is an obvious choice of the variables (*state variables*) to define the system state, but there are also many cases in which the choice of state variables is by no means obvious.

Newton invented calculus as a means of characterizing the behavior of dynamic systems, and his method continues in use to this very day. In particular, behavior of dynamic systems is represented by systems of ordinary differential equations. The differential equations are said to constitute a *mathematical model* of the physical process. We can predict how the physical process will behave by solving the differential equations that are used to model the process.

In order to obtain a solution to a system of ordinary differential equations, it is necessary to specify a set of *initial conditions*. The number of initial conditions that must be specified defines the *order* of the system. When the differential equations constitute the mathematical model of a physical system, the initial conditions needed to solve the differential equations correspond to physical quantities needed to predict the future behavior of the system. It thus follows that the initial conditions and physical state variables are equal in number.

In analysis of dynamic systems such as mechanical systems, electric networks, etc. the differential equations typically relate the dynamic variables and their time derivatives of various orders. In the state-space approach, all the differential equations in the mathematical model of a system are *first-order* equations: only the dynamic variables and their first derivatives (with respect to time) appear in the differential equations. Since only one initial condition is needed to specify the solution of a first-order equation, it follows that the

number of first-order differential equations in the mathematical model is equal to the order of the corresponding system.

The dynamic variables that appear in the system of first-order equations are called the *state variables*. From the foregoing discussion, it should be clear that the *number* of state variables in the model of a physical process is unique, although the identity of these variables may not be unique. A few familiar examples serve to illustrate these points.

Example 2A Mass acted upon by friction and spring forces The mechanical system consisting of a mass which is acted upon by the forces of friction and a spring is a paradigm of a second-order dynamic process which one encounters time and again in control processes.

Consider an object of mass M moving in a line. In accordance with Newton's law of motion, the acceleration of the object is the total force f acting on the object divided by the mass.

$$\frac{d^2x}{dt^2} = \frac{f}{M} \tag{2A.1}$$

where the direction of f is in the direction of x. We assume that the force f is the sum of two forces, namely a friction force f_1 and a spring force f_2 . Both of these forces physically tend to resist the motion of the object. The friction force tends to resist the velocity: there is no friction force unless the velocity is nonzero. The spring force, on the other hand, is proportional to the amount that the spring has been compressed, which is equal to the amount that the object has been displaced. Thus

where

$$f_1 = -\beta \left(\frac{dx}{dt}\right)$$
$$f_2 = -\kappa(x)$$

 $f = f_1 + f_2$

Thus

$$\frac{d^2x}{dt^2} = -\left[\beta\left(\frac{dx}{dt}\right) + \kappa(x)\right] / M$$
(2A.2)

A more familiar form of (2A.2) is the second-order differential equation

$$M\frac{d^2x}{dt^2} + \beta\left(\frac{dx}{dt}\right) + \kappa(x) = 0$$
(2A.3)

But (2A.2) is a form more appropriate for the state-space representation. Differential equation (2A.2) or its equivalent (2A.3) is a second-order differential equation and its solution requires two initial conditions: x_0 , the initial position, and \dot{x}_0 , the initial velocity.

To obtain a state-space representation, we need two state variables in terms of which the dynamics of (2A.2) can be expressed as two first-order differential equations. The obvious choice of variables in this case are the displacement x and the velocity v = dx/dt. The two first-order equations for the process in this case are the equation by which velocity is defined

$$\frac{dx}{dt} = v \tag{2A.4}$$

and (2A.2) expressed in terms of x and v. Since $d^2x/dt^2 = dv/dt$, (2A.2) becomes

$$\frac{dv}{dt} = -[\beta(v) + \kappa(x)]/M$$
(2A.5)



Thus (2A.4) and (2A.5) constitute a system of two first-order differential equations in terms of the state variables x and v_s

If we wish to control the motion of the object we would include an additional force f_0 external to the system which would be added to the right-hand side of (2A,5)

$$\frac{dv}{dt} = -[\beta(v) + \kappa(x)]/M + f_0/M$$
(2A.6)

How such a control force would be produced is a matter of concern to the control system designer. But it is not considered in the present example.

In a practical system both the friction force and the spring force are nonlinear functions of their respective variables and a realistic prediction of the system behavior would entail solution of (2A,4) and (2A,5) in which $\beta(v)$ and $\kappa(x)$ are *nonlinear* functions of their arguments. As an approximation, however, it may be permissible to treat these functions as being linear

$$\beta\left(\frac{dx}{dt}\right) \approx B\frac{dx}{dt}$$
$$\kappa(x) \approx Kx$$

where B and K are constants. Often $\beta()$ and $\kappa()$ are treated as linear functions for purposes of control system design, but the accurate nonlinear functions are used in evaluating how the design performs.

A block diagram representation of the differential equations (2A.4) and (2A.6), in accordance with the discussion of Sec. 2.3, is shown in Fig. 2.1.

Example 2B Electric motor with inertia load One of the most common uses of feedback control is to position an inertia load using an electric motor. (See Fig. 2.2.) The inertia load may consist of a very large, massive object such as a radar antenna or a small object such as a precision instrument. An important aspect of the control system design is the selection of a suitable motor, capable of achieving the desired dynamic response and suited to the objective in cost, size, weight, etc. An electric motor is a device that converts electrical energy (input) to mechanical energy (output). The electro-mechanical energy transducer relations are idealizations of Faraday's law of induction and Ampere's law for the force produced on a conductor moving in a magnetic field. In particular, under ideal circumstances the torque developed at the shaft of a motor is proportional to the input current to the motor; the induced emf'v ("back emf") is proportional to the speed ω of rotation

τυ

$$r = K_1 i \tag{2B.1}$$

$$=K_2\omega$$
 (2B.2)



Figure 2.1 Block diagram representing motion of mass with friction and spring reaction forces.



Figure 2.2 DC motor driving inertia load.

The electrical power p_e input to the motor is the product of the current and the induced emf

$$p_e = vi = K_2 \omega \tau / K_1 \tag{2B.3}$$

The mechanical output power is the product of the torque and the angular velocity

 $p_m = \omega \tau$

Thus, from (2B.3)

$$p_e = \frac{K_2}{K_1} p_m$$

If the energy conversion is 100 percent efficient, then

 $K_1 = K_2 = K$

If the energy-conversion efficiency is less than 100 percent then $K_2/K_1 > 1$.

To completely specify the behavior of the system we need the relationships between the input voltage e and the induced emf, and between the torque and the angular velocity of the motor. These are given by

$$e - v = Ri$$
 (Ohm's law) (2B.4)

where R is the electrical resistance of the motor armature, and

$$\tau = J \frac{d\omega}{dt} \tag{2B.5}$$

where J is the inertia of the load. From (2B.1), (2B.5), and (2B.4)

$$J\frac{d\omega}{dt} = K_1 i = \frac{K_1}{R} (e - v)$$
(2B.6)

On using (2B.2) this becomes

ог

$$J\frac{d\omega}{dt} = \frac{K_1}{R}e - \frac{K_1K_2}{R}\omega$$
$$\frac{d\omega}{dt} = -\frac{K_1K_2}{JR}\omega + \frac{K_1}{JR}e$$
(2B.7)

which is a first-order equation with the angular velocity ω as the state variable and with *e* serving as the external control input.

The first-order model of (2B.7) is suitable for control of the speed of the shaft rotation. When the position θ of the shaft carrying the inertia J is also of concern, we must add the differential equation

$$\frac{d\theta}{dt} = \omega \tag{2B.8}$$

This and (2B.7) together constitute a second-order system.





Figure 2.3 Block diagram representing dynamics of dc motor driving inertia load.

Equations (2B.7) and (2B.8) can be arranged in the vector-matrix form

$$\frac{d}{dt} \begin{bmatrix} \theta \\ \omega \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & -K_1 K_2 / JR \end{bmatrix} \begin{bmatrix} \theta \\ \omega \end{bmatrix} + \begin{bmatrix} 0 \\ K_1 / JR \end{bmatrix} e$$

A block-diagram representation of the differential equations that represent this system is given in Fig. 2.3.

Example 2C Electrical network and its thermal analog It is not generally required to design feedback control systems for electrical networks comprising resistors, capacitors, and inductors. But such networks often are mathematically analogous to mechanical systems which one does desire to control, and an engineer experienced in the analysis of electrical networks might be more comfortable with the latter than with the mechanical systems they represent.

One class of mechanical system which is analogous to an electrical network is a thermal conduction system. Electrical voltages are analogous to temperatures and currents are analogous to heat flow rates. The paths of conduction of heat between various points in the system are represented by resistors; the mass storage of heat in various bodies is represented by capacitances; the input of heat by current sources; and fixed temperatures at the boundaries of the system by voltage sources.

Table 2C.1 summarizes the thermal quantities and their electrical analogs.

As an illustration of the use of electrical analogs of thermal systems, consider the system shown in Fig. 2.4 consisting of two masses of temperatures T_1 and T_2 embedded in a thermally

Table 2C.1 Electrical analogs of thermal systems

Thermal system			Electrical system		
Quantity	Symbol	Unit	Quantity	Symbol	Unit
Temperature	Т	deg	Voltage	υ	volt
Heat flux	9	cal/s	Current	i	ampere
Thermal resistivity	Ŕ	deg · s/cal	Resistance	R	ohm
Thermal capacity	С	cal/deg	Capacitance	С	farad
Conduction equation	$q = \frac{1}{R} \left(T_2 - T_1 \right)$			$i=\frac{1}{R}(v_2$	$-v_{1})$
Storage equation	$\frac{dT}{dt} = \frac{q}{C}$			$\frac{dv}{dt} = \frac{1}{C} i$	





NUMBRICO PERIOULINA

insulating medium contained in a metal container which, because of its high thermal conductivity, may be assumed to have a constant temperature T_0 . The temperatures T_1 and T_2 of the masses are to be controlled by controlling the temperature T_0 of the container.

An electrical analog of the system is shown in Fig. 2.5. The capacitors C_1 and C_2 represent the heat capacities of the masses; the resistor R_3 represents the path of heat flow from mass 1 to mass 2; R_1 and R_2 represent the heat flow path from these masses to the metal container.

The differential equations governing the thermal dynamics of the mechanical system are the same as the differential equations of the electrical system, which can be obtained by various standard methods. By use of nodal analysis, for example, it is determined that

$$C_{1} \frac{dv_{1}}{dt} + \left(\frac{1}{R_{1}} + \frac{1}{R_{3}}\right)v_{1} - \frac{1}{R_{3}}v_{2} - \frac{1}{R_{1}}e_{0} = 0$$

$$C_{2} \frac{dv_{2}}{dt} + \left(\frac{1}{R_{2}} + \frac{1}{R_{3}}\right)v_{2} - \frac{1}{R_{3}}v_{1} - \frac{1}{R_{2}}e_{0} = 0$$
(2C.1)

The appropriate state variables for the process are the capacitor voltages v_1 and v_2 . The temperature of the case is represented by a voltage source e_0 which is the input variable to the process. Thus the differential equations of the process are

$$\frac{dv_1}{dt} = -\frac{1}{C_1} \left(\frac{1}{R_1} + \frac{1}{R_3} \right) v_1 + \frac{1}{C_1 R_3} v_2 + \frac{1}{R_1 C_1} e_0$$

$$\frac{dv_2}{dt} = \frac{1}{C_2 R_2} v_1 - \frac{1}{C_2} \left(\frac{1}{R_2} + \frac{1}{R_3} \right) v_2 + \frac{1}{R_2 C_2} e_0$$
(2C.2)



Figure 2.5 Electrical analog of thermal system of Fig. 2.4.

The foregoing examples are typical of the general form of the dynamic equations of a dynamic process. The state variables of a process of order k are designated by x_1, x_2, \ldots, x_k and the external inputs by u_1, u_2, \ldots, u_l

$$\dot{x}_{1} = \frac{dx_{1}}{dt} = f_{1}(x_{1}, x_{2}, \dots, x_{k}, u_{1}, u_{2}, \dots, u_{k}, t)$$

$$\dot{x}_{2} = \frac{dx_{2}}{dt} = f_{2}(x_{1}, x_{2}, \dots, x_{k}, u_{1}, u_{2}, \dots, u_{k}, t)$$

$$\vdots$$

$$\dot{x}_{k} = \frac{dx_{k}}{dt} = f_{k}(x_{1}, x_{2}, \dots, x_{k}, u_{1}, u_{2}, \dots, u_{k}, t)$$
(2.1)

These equations express the time-derivatives of each of the state variables as general functions of all the state variables, inputs, and (possibly) time. The dot over a variable is Newton's notation for the derivative with respect to time.

To simplify the notation the state variables x_1, x_2, \ldots, x_k and control variables u_1, u_2, \ldots, u_l are collected in vectors

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_k \end{bmatrix} \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_l \end{bmatrix}$$
(2.2)

called the *state vector* and the *input vector*, respectively. These are vectors in the mathematical sense and not necessarily in the physical sense. The components of a *physical* vector are usually projections of a physical quantity (e.g., force, velocity) along a set of reference axes. But the components of the state vector of a dynamic system generally do not have this interpretation and need not even represent the same kind of physical quantities: As our examples show, position and velocity are typical components of a mathematical state vector.

In some books the state vector is printed in a special typeface such as boldface \mathbf{x} , to distinguish it from a scalar variable x. We have chosen not to use any special typeface for the state vector since there is rarely any possibility of confusing the entire state vector x with one of its components x_i (always written with a subscript). In subsequent chapters we will make use of a boldface symbol \mathbf{x} to denote the *metastate* of a system, which is the vector comprising the state (or error) vector, concatenated with the exogenous state vector x_0 as explained in Chap. 5 and later.

Using vector notation, the set of differential equations (2.1) that defines a general process can be written compactly as the single vector differential equation

$$\dot{x} = \frac{dx}{dt} = f(x, u, t)$$
(2.3)

where f(x, u, t) is understood to be a k-dimensional vector-valued function of k + l + 1 arguments. When time t does not appear explicitly in any of the functions f_i in (2.1), i.e., in the vector f of (2.3), the system is said to be *time-invariant*. If (2.3) is an accurate model of a physical process, we would expect it to be time-invariant, since we do not have physical laws that change with time. In many situations, however, the differential equations represented by (2.3) are only an approximate model of the physical world, either because a more accurate model is not known, or because it is too complicated to be useful in the intended application. Very often such approximate models are time-varying.

An exact model of a physical process is usually nonlinear. But fortunately many processes can be adequately approximated by linear models over a significant range of operation. In the state-space model of a linear process, the general differential equations of (2.1) take the special form:

$$\dot{x}_{1} = \frac{dx_{1}}{dt} = a_{11}(t)x_{1} + \dots + a_{1k}(t)x_{k} + b_{11}(t)u_{1} + \dots + b_{1l}(t)u_{l}$$

$$\dot{x}_{2} = \frac{dx_{2}}{dt} = a_{21}(t)x_{1} + \dots + a_{2k}(t)x_{k} + b_{21}(t)u_{1} + \dots + b_{2l}(t)u_{l}$$

$$\dot{x}_{k} = \frac{dx_{k}}{dt} = a_{k1}(t)x_{1} + \dots + a_{kk}(t)x_{k} + b_{k1}(t)u_{1} + \dots + b_{kl}(t)u_{l}$$
(2.4)

In vector notation, using the definitions of the state and control vectors as defined in (2.2), the linear dynamic model of (2.4) is written

$$\dot{x} = \frac{dx}{dt} = A(t)x + B(t)u \tag{2.5}$$

where A(t) and B(t) are matrices given by

$$A(t) = \begin{bmatrix} a_{11}(t) \cdots a_{1k}(t) \\ a_{21}(t) \cdots a_{2k}(t) \\ \vdots \\ a_{k1}(t) \cdots a_{kk}(t) \end{bmatrix} \qquad B(t) = \begin{bmatrix} b_{11}(t) \cdots b_{1l}(t) \\ b_{21}(t) \cdots b_{2l}(t) \\ \vdots \\ b_{k1}(t) \cdots b_{kl}(t) \end{bmatrix}$$
(2.6)

It is noted that the matrix A(t) is always a square (k by k) matrix, but that the matrix B(t) need not be square. In most processes of interest the number lof inputs is smaller than the number of state variables: B(t) is a tall, thin matrix. Often there is only one input and the matrix B(t) is only one column wide.

When the system is time-invariant, none of the elements in the matrices A and B depend upon time. Most of this book is concerned with linear, time-

invariant processes, having the dynamic equations

$$\dot{x} = Ax + Bu \tag{2.7}$$

where A and B are constant matrices.

Although the concept of the state of a system is fundamental, there are many situations in which one is not interested in the state directly, but only in its effect on the system *output* vector y(t)

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{bmatrix}$$
(2.8)

for a system having m outputs. In a linear system the output vector is assumed to be a linear combination of the state and the input

$$y(t) = C(t)x(t) + D(t)u(t)$$
(2.9)

where C(t) is an $m \times k$ matrix and D(t) is an $m \times l$ matrix. If the system is time-invariant, C(t) and D(t) are constant matrices.

The outputs of a system are generally those quantities which can be *observed*, i.e., measured by means of suitable sensors. Accordingly, the output vector is called the *observation vector* and (2.9) is called the observation equation.

The presence of the matrix D in (2.9) means that there is a direct connection between the input u(t) and the output y(t), without the intervention of the state x(t). Although there is no general reason for the matrix D to be absent in a practical application, it turns out that it *is* absent in the overwhelming majority of applications. This is fortunate, because the presence of D increases the complexity of much of the theory. Thus most of our development will rest on the assumption that D = 0.

The input vector u in (2.7) represents the assemblage of all physical quantities that affect the behavior of the state. From the control system design standpoint, however, the inputs are of two types:

Control inputs, produced intentionally by the operation of the control system, and

"Exogenous" inputs, present in the environment and not subject to control within the system.

It is customary to reserve the symbol u for the control inputs and to use another symbol for the exogenous inputs. (The word "exogenous," widely used in the field of economics and other social sciences, is gaining currency in the field of control theory.) In this book we shall find it convenient to represent the exogenous inputs by the vector x_0 . The use of the letter "x" suggests that the exogenous inputs are state variables and so they may be regarded: x_0 may be

162720

BLO UL M

regarded as the state of the environment. (Later in the book we shall concatenate the state x of the system to be controlled with the state x_0 of the environment into a metastate of the overall process.)

Thus, separating the input u of (2.7) into a control input and an exogenous input, (2.7) becomes

$$\dot{x} = Ax + Bu + Ex_0 \tag{2.10}$$

which, together with (2.9) will serve as the general representation of a linear system.

2.3 BLOCK-DIAGRAM REPRESENTATIONS

System engineers often find it helpful to visualize the relationships between dynamic variables and subsystems of a system by means of block diagrams. Each subsystem is represented by a geometric figure (such as a rectangle, a circle, a triangle, etc.) and lines with arrows on them show the inputs and the outputs. For many systems, these block diagrams are more expressive than the mathematical equations to which they correspond.

The relationships between the variables in a linear system (2.4) can be expressed using only three kinds of elementary subsystems:

Integrators, represented by triangles

Summers, represented by circles, and

Gain elements, represented by rectangular or square boxes as shown in Fig. 2.6.

An integrator is a block-diagram element whose output is the integral of the input; put in other words, it is the element whose input is the derivative of the output.



Figure 2.6 Elements used in block-diagram representation of linear systems. (a) Integrator; (b) Summer; (c) Gain element.

A summer is a block-diagram element whose output is the sum of all its inputs.

A gain element is a block-diagram element whose output is proportional to its input. The constant of proportionality, which may be time-varying, is placed inside the box (when space permits) or adjacent to it.

Note that the integrator and the gain element are single-input elements; the summer, on the other hand, always has at least two inputs.

A general block diagram for a second-order system (k = 2) with two external inputs u_1 and u_2 is shown in Fig. 2.7. Two integrators are needed, the outputs of which are x_1 and x_2 , and the inputs to which are \dot{x}_1 and \dot{x}_2 , respectively. From the general form of the differential equations (2.4) these are given by

$$\dot{x}_1 = a_{11}x_1 + a_{12}x_2 + b_{11}u_1 + b_{12}u_2$$
$$\dot{x}_2 = a_{21}x_1 + a_{22}x_2 + b_{21}u_1 + b_{22}u_2$$

which are the relationships expressed by the outputs of the two summers shown in Fig. 2.7.

The same technique applies in higher-order systems. If the A matrix has many nonzero terms, the diagram can look like a plate of spaghetti and meatballs. In most practical cases, however, the A matrix is fairly sparse, and



Figure 2.7 Block diagram of general second-order linear system.

ANDRETCO MARDIE MA

12 16

with some attention to layout it is possible to draw a block diagram with a minimum of crossed lines.

To simplify the appearance of the block-diagram it is sometimes convenient to use redundant summers. This is shown in Fig. 2.7. Instead of using two summers, one feeding another, in front of each integrator we could have drawn the diagram with only one summer with four inputs in front of each integrator. But the diagram as shown has a neater appearance. Another technique to simplify the appearance of a block diagram is to show a sign reversal by means of a minus sign adjacent to the arrow leading into a summer instead of a gain element with a gain of -1. This usage is illustrated in Figs. 2.1 and 2.3 of the foregoing examples.

Although there are several international standards for block-diagram symbols, these standards are rarely adhered to in technical papers and books. The differences between the symbols used by various authors, however, are not large and are not likely to cause the reader any confusion.

The following examples illustrate the use of matrices and block diagrams to represent the dynamics of various processes.

Often it is convenient to express relationships between vector quantities by means of block diagrams. The block-diagram symbols of Fig. 2.6 can also serve to designate operations on vectors. In particular, when the input to an integrator of Fig. 2.6(a) is a vector quantity, the output is a vector each component of which is the integral of the corresponding input. The summer of Fig. 2.6(c) represents a vector summer, and the gain element box of Fig. 2.6(c) represents a matrix. In the last case, the matrix need not be square and the dimension of the vector of outputs from the box need not equal the dimension of the vector



Figure 2.8 Block-diagram representation of general linear system.

of inputs. Using this mode of representation, the block diagram of Fig. 2.8 represents the general system given by (2.9) and (2.10).

Example 2D Hydraulically actuated tank gun turret The control of a hydraulically actuated gun turret in an experimental tank has been studied by Loh, Cheok, and Beck.[1] The linearized dynamic model they used for each axis (elevation, azimuth) is given by

$$\theta = \omega$$

$$\dot{\omega} = p + d_{\tau}$$

$$\dot{p} = -\Omega_m p + \frac{K_m}{J} q - \frac{K_m}{J} \omega + d_p$$

$$\dot{q} = -K_v L_v q - K_v K_{\Delta \nu} J p + K_v u + d_q$$
(2D.1)

where $x_1 = \theta =$ turret angle

 $x_2 = \omega =$ turret angular rate

 $x_3 = p$ = angular acceleration produced by hydraulic drive

 $x_4 = q =$ hydraulic servo valve displacement

...

u =control input to servo valve

 $K_m =$ servo motor gain

J = turret inertia

 $\Omega_m = motor natural frequency$

 K_v = servo valve gain

 $K_{\Delta p}$ = differential pressure feedback coefficient

The quantities d_r , d_p , and d_q represent disturbances, including effects of nonlinearities not accounted for by the linearized model (2D.1).

With the state variable definitions given above, the matrices of this process are

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & K_m/J & -\Omega_m & -K_m/J \\ 0 & 0 & -K_v K_{\Delta p}J & -K_v L_v \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 0 \\ 0 \\ K_v \end{bmatrix}$$



Figure 2.9 Dynamic model of hydraulically actuated tank gun turret.

162720

PARD UL M

Table 2D.1 Numerical values of parameters in tank turret control

	Numerical value		
Parameter	Azimuth	Elevation	
K _v	94.3	94.3	
L_{ν}	1.00	1.07	
J (ft-lb \cdot s ²)	7900.	2070.	
Km	8.46×10^{6}	1.96×10^{6}	
ω_m (rad/s)	45.9	17.3	
$K_{\Delta p}$	6.33×10^{-6}	3.86×10^{-5}	

Numerical data for a specific tank were found by Loh, Cheok, and Beck to be as given in Table 2D.1

A block-diagram representation of the dynamics represented by (2D.1) is shown in Fig. 2.9.

2.4 LAGRANGE'S EQUATIONS

The equations governing the motion of a complicated mechanical system, such as a robot manipulator, can be expressed very efficiently through the use of a method developed by the eighteenth-century French mathematician Lagrange. The differential equations that result from use of this method are known as *Lagrange's equations* and are derived from Newton's laws of motion in most textbooks on advanced dynamics.[2, 3]

Lagrange's equations are particularly advantageous in that they automatically incorporate the constraints that exist by virtue of the different parts of a system being connected to each other, and thereby eliminate the need for substituting one set of equations into another to eliminate forces and torques of constraint. Since they deal with scalar quantities (potential and kinetic energy) rather than with vectors (forces and torques) they also minimize the need for complicated vector diagrams that are usually required to define and resolve the vector quantities in the proper coordinate system. The advantages of Lagrange's equations may also turn out to be disadvantages, because it is necessary to identify the generalized coordinates correctly at the very beginning of the analysis of a specific system. An error made at this point may result in a set of differential equations that look correct but do not constitute the correct model of the physical system under investigation.

The fundamental principle of Lagrange's equations is the representation of the system by a set of generalized coordinates q_i (i = 1, 2, ..., r), one for each independent degree of freedom of the system, which completely incorporate the constraints unique to that system, i.e., the interconnections between the parts of the system. After having defined the generalized coordinates, the kinetic energy T is expressed in terms of these coordinates and their derivatives, and the

potential energy V is expressed in terms of the generalized coordinates. (The potential energy is a function of only the generalized coordinates and *not* their derivatives.) Next, the *lagrangian* function

$$L = T(q_1,\ldots,q_r,\dot{q}_1,\ldots,\dot{q}_r) - V(q_1,\ldots,q_r)$$

is formed. And finally the desired equations of motion are derived using Lagrange's equations

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) - \frac{\partial L}{\partial q_i} = Q_i \qquad i = 1, 2, \dots, r$$
(2.11)

where Q_i denotes generalized forces (i.e., forces and torques) that are external to the system or not derivable from a scalar potential function.

Each of the differential equations in the set (2.11) will be a second-order differential equation, so a dynamic system with r degrees of freedom will be represented by r second-order differential equations. If one *state variable* is assigned to each generalized coordinate and another to the corresponding derivative, we end up with 2r equations. Thus a system with r degrees of freedom is of order 2r.

Example 2E Inverted pendulum on moving cart A typical application of Lagrange's equations is to define the motion of a collection of bodies that are connected together in some manner such as the inverted pendulum on a cart illustrated in Fig. 2.10.

It is observed that the motion of the system is uniquely defined by the displacement of the cart from some reference point, and the angle that the pendulum rod makes with respect to the vertical. Instead of using θ , we could use the horizontal displacement, say y_1 , of the bob relative to the pivot point, or the vertical height z_2 of the bob. But, whatever variables are used, it is essential to know that the system has only two degrees of freedom, and that the dynamics must be expressed in terms of the corresponding generalized coordinates.

The kinetic energy of the system is the sum of the kinetic energy of each mass. The cart is confined to move in the horizontal direction so its kinetic energy is

 $T_1 = \frac{1}{2}M\dot{y}^2$



Figure 2.10 Inverted pendulum on moving cart.

STATE-SPACE REPRESENTATION OF DYN/

The bob can move in the horizontal and in the vertical direction so

 $T_2 = \frac{1}{2}m(\dot{y}_2^2 + \dot{z}_2^2)$

But the rigid rod constrains z_2 and y_2

 $y_2 = y + l \sin \theta \qquad \dot{y}_2 = \dot{y} + l \dot{\theta} \cos \theta$ $z_2 = l \cos \theta \qquad \dot{z}_2 = -l \dot{\theta} \sin \theta$

Thus

$$T = T_1 + T_2 = \frac{1}{2}M\dot{y}^2 + \frac{1}{2}m[(\dot{y} + l\dot{\theta}\cos\theta)^2 + l^2\dot{\theta}^2\sin^2\theta]$$
$$= \frac{1}{2}M\dot{y}^2 + \frac{1}{2}m[\dot{y}^2 + 2\dot{y}\dot{\theta}l\cos\theta + l^2\dot{\theta}^2]$$

The only potential energy is stored in the bob

 $V = mgz_2 = mgl\cos\theta$

Thus the lagrangian is

L

$$= T - V = \frac{1}{2}(M + m)\dot{y}^{2} + ml\cos\theta\dot{y}\dot{\theta} + \frac{1}{2}ml^{2}\dot{\theta}^{2} - mgl\cos\theta$$
(2E.1)

The generalized coordinates having been selected as (y, θ) , Lagrange's equations for this system are

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{y}} \right) - \frac{\partial L}{\partial y} = f$$

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\theta}} \right) - \frac{\partial L}{\partial \theta} = 0$$
(2E.2)

Now

$$\frac{\partial L}{\partial \dot{y}} = (M+m)\dot{y} + ml\cos\theta\dot{\theta}$$
$$\frac{\partial L}{\partial y} = 0$$
$$\frac{\partial L}{\partial \dot{\theta}} = ml\cos\theta\dot{y} + ml^2\dot{\theta}$$
$$\frac{\partial L}{\partial \theta} = mgl\sin\theta - ml\sin\theta\dot{y}\dot{\theta}$$

Thus (2E.2) become

 $(M+m)\ddot{y} + ml\cos\theta\ddot{\theta} - ml\dot{\theta}^{2}\sin\theta = f$ ml cos $\theta\ddot{y} + ml^{2}\ddot{\theta} - mgl\sin\theta = 0$ (2E.3)

These are the exact equations of motion of the inverted pendulum on a cart shown in Fig. 2.10. They are nonlinear owing to the presence of the trigonometric terms $\sin \theta$ and $\cos \theta$ and the quadratic terms $\dot{\theta}^2$ and $\dot{y}\dot{\theta}$. If the pendulum is stabilized, however, then θ will be kept small. This justifies the approximations

$$\cos \theta \simeq 1 \qquad \sin \theta \simeq \theta$$

We may also assume that $\dot{\theta}$ and \dot{y} will be kept small, so the quadratic terms are negligible. Using these approximations we obtain the linearized dynamic model

$$(M+m)\ddot{y} + ml\ddot{\theta} = f$$

$$m\ddot{y} + ml\ddot{\theta} - mg\theta = 0$$
 (2E.4)

A state-variable representation corresponding to (2E.4) is obtained by defining the state vector

$$x = [y, \theta, \dot{y}, \dot{\theta}]'$$

Then

$$\frac{dy}{dt} = \dot{y}$$

$$\frac{d\theta}{dt} = \dot{\theta}$$
(2E.5)

constitute the first two dynamic equations and on solving (2E.4) for \ddot{y} and $\ddot{\theta}$, we obtain two more equations

$$\frac{d}{dt}(\dot{y}) = \ddot{y} = \frac{f}{M} - \frac{mg}{M}\theta$$

$$\frac{d}{dt}(\dot{\theta}) = \ddot{\theta} = -\frac{f}{Ml} + \left(\frac{M+m}{Ml}\right)g\theta$$
(2E.6)

The four equations can be put into the standard matrix form

$$\dot{x} = Ax + Bu$$

with

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -mg/M & 0 & 0 \\ 0 & (M+m)g/Ml & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ 0 \\ 1/M \\ -1/Ml \end{bmatrix}$$

and

u = f = external force

A block-diagram representation of the dynamics (2E.5) and (2E.6) is shown in Fig. 2.11.



Figure 2.11 Block diagram of dynamics of inverted pendulum on moving cart.

2.5 RIGID BODY DYNAMICS

The motion of a single rigid body has six dynamic degrees of freedom: three of these define the location of a reference point (usually the center of mass) in the body, and three define the orientation (attitude) of the body. Since each of the six degrees of freedom takes two state variables (one position and one velocity) a total of 12 first-order differential equations are required to completely describe the motion of the body. In most applications, however, not all of these 12 state variables are of interest and not all the differential equations are needed. In a gyroscope, for example, only the orientation is of interest.

The motion of a rigid body is, of course, governed by the familiar newtonian laws of motion

$$\frac{d\vec{p}}{dt} = \vec{f} \tag{2.12}$$

$$\frac{d\vec{h}}{dt} = \vec{\tau} \tag{2.13}$$

where $\vec{p} = [p_x, p_y, p_z]'$ is the linear momentum of the body

 $\vec{h} = [h_x, h_y, h_z]'$ is the angular momentum of the body

 $\vec{f} = [f_x, f_y, f_z]'$ is force acting on the body

 $\vec{\tau} = [\tau_x, \tau_y, \tau_z]'$ is torque acting on the body

It is important to understand that (2.12) and (2.13) are valid only when the axes along which the motion is resolved are an inertial frame of reference, i.e., they are neither accelerating nor rotating. If the axes are accelerating linearly or rotating, then (2.12) and (2.13) must be modified to account for the motion of the reference axes.

The rotational dynamics of a rigid body are more complicated than the translational dynamics for several reasons: the mass M of a rigid body is a scalar, but the moment of inertia J is a 3×3 matrix. If the body axes are chosen to coincide with the "principal axes," the moment of inertia matrix is diagonal; otherwise the matrix J has off-diagonal terms. This is not the only complication, however, or even the main one. The main complication is in the description of the attitude or orientation of the body in space. To define the orientation of the body in space, we can define three axes (x_B, y_B, z_B) fixed in the body, as shown in Fig. 2.12. One way of defining the attitude of the body is to define the angles between the body axes and the inertial reference axes (x_I, y_I, z_I) . These angles are not shown in the diagram. Not only are they difficult to depict in a two-dimensional picture, but they are not always defined the same way. In texts on classical mechanics, the orientation of the body is defined by a set of three angles, called Euler angles, which describe the orientation of a set of nonorthogonal axes fixed in the body with respect to the inertial reference axes. In aircraft and space mechanics it is now customary to define the orientation of a set of orthogonal axes in the body (body axes) with respect to the inertial reference.







Suppose the body axes are initially aligned with the inertial reference axes. Then, the following sequence of rotations are made to bring the body axes into *general position*:

First, a rotation ψ (yaw) about the z axis Second, a rotation θ (pitch) about the resulting y axis Third, a rotation ϕ (roll) about the resulting x axis

By inspection of the diagrams of Fig. 2.13 we see that

$$\begin{bmatrix} x_{B1} \\ y_{B1} \\ z_{B1} \end{bmatrix} = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix}$$
(2.14)

$$\begin{vmatrix} x_{B2} \\ y_{B2} \\ z_{B2} \end{vmatrix} = \begin{vmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{vmatrix} \begin{vmatrix} x_{B1} \\ y_{B1} \\ z_{B1} \end{vmatrix}$$
(2.15)

$$\begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} = \begin{bmatrix} x_{B3} \\ y_{B3} \\ z_{B3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} x_{B2} \\ y_{B2} \\ z_{B2} \end{bmatrix}$$
(2.16)

Thus we see that

$$\begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} = T_{BI} \begin{bmatrix} x_I \\ y_I \\ z_I \end{bmatrix}$$



Figure 2.13 Sequence of rotations of body axes from reference to "general" orientation (z axis down in aircraft convention). (a) Axes in reference position; (b) First rotation about z axis—yaw (ψ); (c) Second rotation—about y axis—pitch (θ); (d) Third and final rotation—about x axis—roll (ϕ).

where T_{BI} is the matrix that rotates the body axes from reference position, and is the product of the three matrices in (2.14)-(2.16).

$$T_{BI} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi & \sin \phi \\ 0 & -\sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{bmatrix} \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
(2.17)

Each factor of T_{BI} is an orthogonal matrix and hence T_{BI} is orthogonal, i.e.,

$$T_{IB} = T_{BI}^{-1} = T_{BI}' \tag{2.18}$$

Note that $T_{BI}^{-1} = T_{IB}$ is the matrix that returns the body axes from the general position to the reference position.

Note that the order of rotations implicit in T_{BI} is important: the three matrices in (2.17) do not commute.

Since any vector in space can be resolved into its components in body axes or in inertial axes, we can use the transformation (2.17) to obtain the components of a vector in one set of axes, given its components in the other. In particular suppose \vec{a} is any vector in space. When it is resolved into components along an inertial reference we attach the subscript *I*; when it is resolved in body axes, we attach the subscript *B*

$$\vec{a}_{I} = \begin{bmatrix} a_{xI} \\ a_{yI} \\ a_{zI} \end{bmatrix} \qquad \vec{a}_{B} = \begin{bmatrix} a_{xB} \\ a_{yB} \\ a_{zB} \end{bmatrix}$$



Using (2.17) we obtain

$$\vec{a}_B = T_{BI}\vec{a}_I \tag{2.19}$$

This relationship can be applied to (2.13) for the angular motion of a rigid body and, as we shall see later, for describing the motion of an aircraft along rotating body axes.

In the case of a rigid body, the angular momentum vector is

$$\vec{h} = J\vec{\omega} \tag{2.20}$$

where J is the moment of inertia matrix and $\vec{\omega}$ is the angular velocity vector. If the axes along which \vec{h} is resolved are defined to be coincident with the *physical* principal axes of the body, then J is a diagonal matrix. Thus when \vec{h} is resolved along principal body axes, we get from (2.17)

$$\vec{h}_B = \begin{bmatrix} J_x \omega_x \\ J_y \omega_y \\ J_z \omega_z \end{bmatrix}$$
(2.21)

But (2.13) holds only when the vector \vec{h} is measured with respect to an inertial reference: In the notation established above

$$\frac{d\vec{h}_I}{dt} = \frac{d}{dt} \left(T_{IB} \vec{h}_B \right) = \vec{\tau}_I \tag{2.22}$$

The transformation T_{IB} , however, is not constant. Hence (2.22) must be written

$$T_{IB}\vec{h}_B + \dot{T}_{IB}\vec{h}_B = \vec{\tau}_I$$

or, multiplying both sides by $T_{BI} = T_{IB}^{-1}$:

$$\dot{\vec{h}}_B + T_{BI} \dot{T}_{IB} \vec{h}_B = T_{BI} \vec{\tau}_I = \vec{\tau}_B$$
 (2.23)

which, in component form can be written

$$\begin{bmatrix} J_x \dot{\omega}_{xB} \\ J_y \dot{\omega}_{yB} \\ J_z \dot{\omega}_{zB} \end{bmatrix} + T_{BI} \dot{T}_{IB} \begin{bmatrix} J_x \omega_{xB} \\ J_y \omega_{yB} \\ J_z \omega_{zB} \end{bmatrix} = \begin{bmatrix} \tau_{xB} \\ \tau_{yB} \\ \tau_{zB} \end{bmatrix}$$
(2.24)

These differential equations relate the components of the angular velocity vector, $\vec{\omega}$ projected onto rotating body axes

$$\vec{\omega}_B = [\omega_{xB}, \omega_{yB}, \omega_{zB}]'$$

to the torque vector also projected along body axes. To complete (2.24) we need the matrix $T_{BI}\dot{T}_{IB}$. It can be shown that

$$\dot{T}_{IB} = T_{IB} \begin{bmatrix} 0 & -\omega_{zB} & \omega_{yB} \\ \omega_{zB} & 0 & -\omega_{xB} \\ -\omega_{yB} & \omega_{xB} & 0 \end{bmatrix}$$
(2.25)

So that

$$T_{BI}\dot{T}_{IB} = \begin{bmatrix} 0 & -\omega_{zB} & \omega_{yB} \\ \omega_{zB} & 0 & -\omega_{xB} \\ -\omega_{yB} & \omega_{xB} & 0 \end{bmatrix}$$
(2.26)

(See Note 2.1.)

Hence (2.24) becomes

$$J_{x}\dot{\omega}_{xB} + (J_{z} - J_{y})\omega_{yB}\omega_{zB} = \tau_{xB}$$

$$J_{y}\dot{\omega}_{yB} + (J_{x} - J_{z})\omega_{xB}\omega_{zB} = \tau_{yB}$$

$$J_{z}\dot{\omega}_{zB} + (J_{y} - J_{x})\omega_{xB}\omega_{yB} = \tau_{zB}$$
(2.27)

These are the famous *Euler equations* that describe how the body-axis components of the angular velocity vector evolve in time, in response to torque components in body axes.

In order to completely define the attitude (orientation), we need to relate the rotation angles ϕ , θ , and ψ to the angular velocity components. One way—not the easiest, however—of obtaining the required relations is via (2.17) and (2.25). It can be shown that

$$\dot{\phi} = \omega_x + (\omega_y \sin \phi + \omega_z \cos \phi) \tan \theta$$

$$\dot{\theta} = \omega_y \cos \phi - \omega_z \sin \phi$$

$$\dot{\psi} = (\omega_x \sin \phi + \omega_y \cos \phi) / \cos \theta$$
(2.28)

These relations, also nonlinear, complete the description of the rigid body dynamics.

Example 2F The gyroscope One of the most interesting applications of Euler's equations is to the study of the gyroscope. This device (also the spinning top) has fascinated mathematicians and physicists for over a century. (See Note 2.2.) And the gyroscope is an extremely useful sensor of aircraft and spacecraft motion. Its design and control has been an important technological problem for half a century.

In an ideal gyroscope the rotor, or "wheel," is kept spinning at a constant angular velocity. (A motor is provided to overcome the inevitable friction torques present even in the best of instruments. The precise control of wheel speed is another important control problem.) Suppose that the axis through the wheel is the body z axis. We assume that τ_{zB} is such that $\dot{\omega}_{zB} = 0$, i.e., that

$$H_z = J_z \omega_z = \text{const}$$
 (2F.1)

 $(J_z$ is called the "polar" moment of inertia in gyro parlance.) We can also assume that the gyroscope wheel is a "true" wheel: that the z axis is an axis of symmetry, and hence that

$$J_x = J_y = J_d$$
 (the "diametrical" moment of inertia)

The first two equations of (2.27) then become

$$\dot{\omega}_{xB} + \frac{H}{J_d} \omega_{yB} = \frac{\tau_x}{J_d}$$

$$\dot{\omega}_{yB} - \frac{H}{J_d} \omega_{xB} = \frac{\tau_y}{J_d}$$
(2F.2)





Figure 2.14 Two-degrees-of-freedom gyro wheel.

where

$$H = H_z \left(1 - \frac{J_d}{J_z} \right)$$

To use a gyro as a sensor, the wheel is mounted in an appropriate system of gimbals which permit it to move with respect to the outer case of the gyro. In a two-axis gyro, the wheel is permitted two degrees of freedom with respect to the case, as depicted in Fig. 2.14. The case of the gyro is rigidly attached to the body whose motion is to be measured.

The range of motion of the wheel about its x and y body axes relative to the gyro case is very small (usually a fraction of a degree). Hence the gyro must be "torqued" about the axes in the plane normal to the spin axis to make the wheel keep up with its case, and as we shall see shortly, the torque required to do this is a measure of the angular velocity of the case.

Since the motion of the wheel relative to the case is very small, we do not need equations like (2.27) to relate the angular displacements of the gyro wheel from its null positions in the case. We can write

$$\begin{split} \delta_{x} &= \omega_{xB} - \omega_{xE} \\ \delta_{y} &= \omega_{yB} - \omega_{yE} \end{split} \tag{2F.3}$$

where ω_{xE} and ω_{yE} are the external angular velocities that the gyro is to measure.

These equations, together with (2F.2), constitute the basic equations of an ideal gyro. A block-diagram representation of (2F.2) and (2F.3), and a closed-loop feedback system for controlling the gyro is shown in Fig. 2.15. The feedback system shows the control torques generated as functions of the displacements δ_x and δ_y . These displacements can be measured by means of "pick-offs"—small magnetic sensors located on the case and capable of measuring small tilts of the wheel. The control torque needed to drive the "pick-off angles" δ_x and δ_y to zero can also be generated magnetically. In some designs the pick-off and torquer functions can be combined in a single device. The control system is designed to drive the angular displacements δ_x and δ_y to zero. If this is accomplished

$$\omega_{xB} = \omega_{xE} \qquad \omega_{yB} = \omega_{yE} \tag{2F.4}$$





If the angular velocity components ω_{xB} and ω_{yB} are constant

$$\tau_{x} = H\omega_{yB} = H\omega_{yE}$$

$$\tau_{y} = -H\omega_{xB} = -H\omega_{xE}$$
(2F.5)

where H is a constant of the gyro. If this constant is accurately calibrated, and if the input torque to the gyro is accurately metered, then the steady state torques about the respective axes that keep the wheel from tilting relative to its case (i.e., "capture" the wheel) are proportional to the measured external angular velocity components.

The control system that keeps the wheel captured is an important part of every practical gyro. Some of the issues in the design of such a control system will be the subject of problems in later chapters.

The differential equations of (2F.2) are idealized to the point of being all but unrealistic. In addition to the control torques acting on the gyro, other torques, generated internal to the gyro, are also inevitably present. These include damping torques (possibly aerodynamic). And in a so-called tuned-rotor gyro, the gimbals are implemented by a special flexure hinge which produces small but not insignificant spring torques. When these torques are included, (2F.2) becomes

$$\begin{split} \dot{\omega}_{xB} &= \frac{H}{J_d} \, \omega_{yB} - \frac{B}{J_d} \left(\omega_{xB} - \omega_{xE} \right) - \frac{K_D}{J_d} \, \delta_x - \frac{K_Q}{J_d} \, \delta_y + \frac{\tau_x}{J_d} \\ \dot{\omega}_{yB} &= -\frac{H}{J_d} \, \omega_{xB} - \frac{B}{J_d} \left(\omega_{yB} - \omega_{yE} \right) - \frac{K_D}{J_d} \, \delta_y + \frac{K_Q}{J_d} \, \delta_x + \frac{\tau_y}{J_d} \end{split}$$



Note that the damping coefficients D in both axes are assumed equal and that the "spring" matrix

$$K = \begin{bmatrix} -K_D & -K_Q \\ K_Q & -K_D \end{bmatrix}$$

has a special kind of symmetry. This form of the matrix is justified by the physical characteristics of typical tuned-rotor gyros.

2.6 AERODYNAMICS

One of the most important applications of state-space methods is in the design of control systems for aircraft and missiles.

The forces (except for gravitation) and moments on such vehicles are produced by the motion of the vehicle through the air and are obtained, in principle, by integrating the aerodynamic pressure over the entire surface of the aircraft. Computer programs for actually performing this integration numerically are currently available. In an earlier era this was accomplished by approximate analysis done by skillful aerodynamicists, and verified by extensive wind-tunnel testing. (Wind-tunnel tests are performed to this day, notwithstanding the computer codes.)

Several textbooks, e.g., [4, 5], are available which give an exposition of the relevant aerodynamic facts of interest to the control system designer. The aerodynamic forces and moments are complicated, nonlinear functions of many variables and it is barely possible to scratch the surface of this subject here. The purpose of this section is to provide only enough of the principles as are needed to motivate the design examples to be found later on in the book.

The aerodynamic forces and moments depend on the velocity of the aircraft relative to the air mass. In still air (no winds) they depend on the velocity of the aircraft along its own body axes: the orientation of the aircraft is not relevant in determining the aerodynamic forces and moments. But, since the natural axes for resolving the aerodynamic forces and moments are moving (rotating and accelerating), it is necessary to formulate the equations of motion in the moving coordinate system.

The rotation motion of a general rigid body has been given in (2.24). In aircraft terminology the projections of the angular velocity vector on the body x, y, and z axes have standard symbols:

$$\omega_x = p$$
 (roll rate)
 $\omega_y = q$ (pitch rate) (2.29)
 $\omega_z = r$ (yaw rate)

(The logic of using three consecutive letters of the alphabet (p, q, r) to denote the projections of the angular velocity vector on the three consecutive body axes is unassailable. But the result is "amnemonic" (hard to remember): p does not represent pitch rate and r does not represent roll rate.)

Thus, assuming that the body axes are the principal axes of the aircraft, the rotational dynamics are expressed as

$$\dot{p} = \frac{L}{J_x} - \frac{J_z - J_y}{J_x} qr$$

$$\dot{q} = \frac{M}{J_y} - \frac{J_x - J_z}{J_y} pr$$

$$\dot{r} = \frac{N}{J_z} - \frac{J_y - J_x}{J_z} pq$$
(2.30)

where L, M, and N are the aerodynamic moments about the body x, y, and z axes respectively. Thus L is the rolling moment, M is the pitching moment, and N is the yawing moment. These are functions of various dynamic variables, as explained later.

To define the translational motion of an aircraft it is customary to project the velocity vector onto body fixed axes

$$\vec{v}_B = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$
(2.31)

where u, v, and w are the projections of the vehicle velocity vector onto the body x, y, and z axes. The linear momentum of the body, in an inertial frame, is

$$\vec{p} = m\vec{v}_I = mT_{IB}\vec{v}_B$$

Hence, the dynamic equations for translation are

$$\frac{d}{dt}(mT_{IB}\vec{v}_B) = m\left(T_{IB}\frac{d\vec{v}_B}{dt} + \dot{T}_{IB}\vec{v}_B\right) = \vec{f}_I$$
(2.32)

where \vec{f}_I are the external forces acting on the aircraft referred to an inertial frame. Proceeding as we did in developing (2.24) we find that

$$\frac{d\vec{v}_B}{dt} = -T_{BI}\dot{T}_{IB}\vec{v}_B + \frac{1}{m}\vec{f}_B$$
(2.33)

where $\vec{f}_B = T_{BL}\vec{f}_I$ is the force acting on the aircraft resolved along the body-fixed axes and

$$T_{BI}\dot{T}_{IB} = \begin{bmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{bmatrix}$$
(2.34)

as given by (2.26) but using the p, q, r notation defined in (2.29).

In component form (2.33) becomes

$$\dot{u} = rv - qw + \frac{1}{m} f_{xB}$$

$$\dot{v} = -ru + pw + \frac{1}{m} f_{yB}$$
(2.35)

$$\dot{w} = qu - pv + \frac{1}{m} f_{zB}$$

where f_{xB} , f_{yB} , and f_{zB} are the total forces (engine, aerodynamic, and gravitational) acting on the body. Since the aircraft axes are not in general in the direction of the gravity vector, each component f_{xB} , f_{yB} , and f_{zB} will have a term due to gravity. In addition to the force of gravity, there is the thrust force produced by the aircraft engine—generally assumed to act along the vehicle x axis—and the aerodynamic forces—the lift and drag forces. The acceleration terms rv, qw, etc., are Coriolis accelerations due to the rotation of the body axes.

Complete dynamic equations of the vehicle consist of (2.30) which give the angular accelerations, (2.35) which give the linear accelerations, (2.28) which give the angular orientation, and finally the equations for the vehicle position:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \dot{\vec{r}} = \vec{v}_I = T_{IB} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$
(2.36)

This system of 12 first-order differential equations, with the moments and forces evaluated as functions of whatever they depend upon constitute the complete six-degrees-of-freedom description of the aircraft behavior.

The aerodynamic forces and moments all depend on the dynamic pressure

 $Q = \frac{1}{2}\rho V^2 \tag{2.37}$

where ρ is the air density and

$$V = (u^2 + v^2 + w^2)^{1/2}$$

is the speed of the aircraft. (Dynamic pressure has the dimension of force per unit area.) Thus the aerodynamic forces and moments can be expressed in the form

$$f_{xA} = QAC_x$$

$$f_{yA} = QAC_y$$

$$f_{zA} = QAC_z$$

$$L = lQAC_L$$

$$M = lQAC_M$$

$$N = lQAC_N$$
(2.38)

where C_x , C_y , C_z , C_L , C_M , C_N , are dimensionless aerodynamic "coefficients," A

is a reference area (usually the frontal area of the vehicle), and l is a reference length. (In some treatments different reference lengths are used for roll, pitch, and yaw.)

The aerodynamic coefficients in turn are functions of the vehicle velocity (linear and angular) components, and, for movable control surfaces, also functions of the deflections of the surfaces from their positions of reference. The variables of greatest influence on the coefficients are the vehicle speed (or, more precisely, the Mach number), the angle-of-attack α and the side-slip angle β . These, respectively, define the direction of the velocity vector relative to the vehicle body axes; α is the angle that the velocity vector makes with respect to the longitudinal axis in the pitch direction and β is the angle it makes with respect to the longitudinal axis in the yaw direction. (See Fig. 2.16.) From the figure

$$\alpha = \tan^{-1} \left(\frac{w}{(u^2 + v^2)^{1/2}} \right) \approx \frac{w}{u}$$

$$\beta = \tan^{-1} \left(\frac{v}{u} \right) \approx \frac{v}{u}$$
(2.39)

with the approximate expressions being valid for small angles.

For purposes of control system design, the aircraft dynamics are frequently linearized about some operating condition or "flight regime," in which it is assumed that the aircraft velocity and attitude are constant. The control surfaces and engine thrust are set, or "trimmed," to these conditions and the control system is designed to maintain them, i.e., to force any perturbations from these conditions to zero.

If the forward speed is approximately constant, then the angle of attack and angle of side slip can be used as state variables instead of w and v, respectively.



Figure 2.16 Definitions of angle-of-attack α and side-slip angle β .

	Longitudinal	Lateral
Rates	α : angle of attack q: pitch rate Δu : change in speed	 β: side slip angle p: roll rate r: yaw rate
Positions	θ: pitch z: altitude	 φ: roll angle ψ: yaw angle x: forward displacement y: cross-track displacement
Controls	δ_E : elevator deflection	δ_A : aileron deflection δ_R : rudder deflection

Table 2.1 Aerodynamic variables

Also in studying small perturbations from trim conditions it is customary to separate the longitudinal motion from the lateral motion. In many cases the lateral and longitudinal dynamics are only lightly coupled, and the control system can be designed for each channel without regard to the other. The variables are grouped as shown in Table 2.1.





The aircraft pitch motion is typically controlled by a control surface called the *elevator* (or by canards in the front of the vchicle). The roll is controlled by a pair of ailerons, and the yaw is controlled by a rudder. These are also shown in Table 2.1.

The function of most control system designs is to regulate *small* motions rather than to control absolute position (x, y, and z). Thus the inertial position is frequently not included in the state equations. This leaves nine equations, four in the longitudinal channel and five in the lateral channel. These can be written in the following form:

Longitudinal dynamics (See Fig. 2.17)

$$\Delta \dot{u} = X_u \Delta u + X_\alpha \alpha - g\theta + X_E \delta_E$$

$$\dot{\alpha} = \frac{Z_u}{V} \Delta u + \frac{Z_\alpha}{V} \alpha + q + \frac{Z_E}{V} \delta_E$$

$$\dot{q} = M_u \Delta u + M_\alpha \alpha + M_q q + M_E \delta_E$$

$$\dot{\theta} = q$$
(2.40)

Lateral dynamics (See Fig. 2.18)

$$\dot{\beta} = \frac{Y_{\beta}}{V}\beta + \frac{Y_{p}}{V}p + \left(\frac{Y_{r}}{V} - 1\right)r + \frac{g}{V}\phi + \frac{Y_{A}}{V}\delta_{A} + \frac{Y_{R}}{V}\delta_{R}$$

$$\dot{p} = L_{\beta}\beta + L_{p}p + L_{r}r + L_{A}\delta_{A} + L_{R}\delta_{R}$$

$$\dot{r} = N_{\beta}\beta + N_{p}p + N_{r}r + N_{A}\delta_{A} + N_{R}\delta_{R}$$

$$\dot{\phi} = p$$

$$\dot{\psi} = r$$
(2.41)

The symbols X, Y, Z, L, M, and N, with subscripts have become fairly standardized in the field of aircraft and missile control, although the sign conventions often differ from one user to another, which can often cause consternation. The symbols with the capital-letter subscripts, E, A, and R (for elevator, ailerons, and rudder), however, are not standard. It is customary to use cumbersome double subscript notation for these quantities.

2.7 CHEMICAL AND ENERGY PROCESSES

It is often necessary to control large industrial processes which involve heat exchangers, chemical reactors, evaporators, furnaces, boilers, driers, and the like.

Because of their large physical size, such processes have very slow dynamic behavior—measured on a scale of minutes or hours rather than seconds as in the case for aircraft and instrument controls. Such processes are often slow





Figure 2.18 Aircraft lateral dynamics.

enough to be controlled manually: an experienced plant operator can monitor the instruments in the control room and (remotely) open and close the valves to maintain a satisfactory equilibrium condition. But slow as such processes are, they are not necessarily stable. If the operator is not constantly monitoring and actively intervening in its operation the process may run away with itself. The "Three-Mile Island" nuclear plant accident (See Note 2.3), is perhaps the most dramatic episode of this kind in recent years, but other episodes, less dramatic than Three-Mile Island, occur with regrettable frequency.

Gross failures of the type of Three-Mile Island are probably not traceable to inadequate dynamic performance of the control system, but rather to failures in hardware that inadequate procedures and training permit to go without prompt repair. The considerations that apply to design of proper procedures and personnel practices are outside the scope of this book. Here we are concerned with the design of systems for normal, closed-loop operation, i.e., under the assumption that the sensors and actuators are maintained in good

working order. Often so much of the engineering effort is spent in selecting suitable hardware—and this effort is totally justified—that little time is left to consider efficient operation under normal conditions. Large industrial processes are costly to operate, however, and even small percentage improvements (such as in reduced energy consumption) can be worth a considerable sum. There is a celebrated design (See Note 2.4) in the paper industry in which a small improvement in the product quality ("base weight" of the paper) returned many times the cost of installing a computer control system.

One of the difficulties in working with large industrial processes is that they involve subsystems the behavior of which are not readily characterized by simple mathematical models. The physics and chemistry of devices like evaporators, heat exchangers, driers, and the like, are not as amenable to mathematical representation as are the physics of simple electromechanical systems, or even of aircraft. It is often necessary to work with empirical models obtained by fitting curves to test data. And test data is often hard to come by because the processes are slow and there is considerable reluctance to shut them down long enough to amass a sufficient quantity of data with which to construct an empirical model.

Still another difficulty in dealing with industrial processes is the large number of dynamic variables that must be considered. Unless suitable simplifications are made, the number of variables can run into the hundreds. Although the methods described in this book can be used for designing control systems for very high order processes, the insights that the engineer often can develop using low-order models will be lacking.

To show how state-space methods can be applied to industrial processes we have selected several examples that have been described in the literature and are actually in operation. These examples show that it is possible to deal effectively with processes of considerable complexity using models of reasonably low order.

Example 2G Distillation column A distillation column is a complex process. A large number of variables (upward of 100) are needed to accurately model its dynamic behavior.

In the interest of applying modern control techniques to the design of a control system for a distillation column, Gilles and Retzbach[6, 7, 8] manage ingeniously to reduce the number of state variables to only 4. Their study deals with an extractive column intended for separating isopropanol from a mixture with water, using glycol as an extractant. A schematic diagram of the column is shown in Fig. 2.19. The mixture of water and isopropanol is introduced at the feed stage F_A and the glycol extractant is introduced near the top of the column. A controlled amount of heating steam is introduced near the bottom of the column where the *bottom product*—the extractant, glycol, is drawn off. In addition, the *vapor side stream* flow rate can be controlled by another valve. The objective of the process is to produce nearly pure isopropanol at the top of the column.

The key to the simplified model of the distillation column developed by Gilles and Retzbach are the profiles of concentration and temperature in the column, sketched in Fig. 2.19. There are two vertical locations in the tower at which the principal physical changes occur: z_1 at which there is an interphase change between water and isopropanol, and a second location z_2 where there is an interphase change between the water and the glycol extractant. At each of these locations there is a sharp temperature gradient.







By varying the flow rates of the water-isopropanol mixture, the heating steam, and the vapor side stream, the positions of these loci or "fronts" can be moved up and down, but the shapes of the distributions are otherwise hardly changed. Thus by controlling the positions of these fronts, the distribution of temperature and concentration can be controlled throughout the column. This property of the distributions motivated Gilles and Retzbach to use the positions z_1 and z_2 as state variables that can adequately represent the behavior of this complex process.

In addition to these state variables, other state variables needed to represent the steam boiler are included in the overall model.

The boiler dynamics are represented by

$$\Delta Q_l = a_{11} \Delta Q_l + b_{11} \Delta u_1$$

$$\Delta \dot{V} = a_{21} \Delta Q_l + a_{22} \Delta V_l$$
(2G.1)

where ΔQ_l = heat flow to reboiler "holdup")

 $\Delta V_i = \text{vapor flow rate}$ changes from equilibrium $\Delta u_1 = \text{steam flow rate}$

Gilles and Retzbach in [6] show that rates of change $\Delta \dot{z}_1$ and $\Delta \dot{z}_2$ in the position of the interphase loci (fronts) are linearly related to the various flow rates:

$$\Delta \dot{z}_1 = b_{32}\Delta S + f_{31}\Delta x_{FA1} + f_{32}\Delta F_A$$

$$\Delta \dot{z}_2 = b_{42}\Delta S + f_{42}\Delta F_A$$
(2G.2)

where ΔS = flow rate of vapor side stream Δx_{FA1} = feed composition ΔF_A = feed flow rate ΔF_A = feed flow rate

As noted, the steam flow rate and the flow rate of the vapor side-stream are control variables. Changes in the feed composition and flow rate are disturbances that the control system is to be designed to counteract.

The positions of the fronts are determined in this process by measuring the temperatures with thermomocouples located near the desired positions of the fronts. It has been found that the temperature changes are approximately proportional to the front position changes:

$$\Delta T_1 = c_{13} \Delta z_1 \tag{2G.3}$$

$$\Delta T_2 = c_{24} \Delta z_2$$

The state, input, disturbance, and observation vectors are defined respectively by

$$x = \begin{bmatrix} \Delta Q_1 \\ \Delta V_i \\ \Delta z_1 \\ \Delta z_2 \end{bmatrix} \qquad u = \begin{bmatrix} \Delta u_1 \\ \Delta S \end{bmatrix} \qquad x_0 = \begin{bmatrix} \Delta x_{FA1} \\ \Delta F_A \end{bmatrix} \qquad y = \begin{bmatrix} \Delta T_1 \\ \Delta T_2 \end{bmatrix}$$

In terms of these variables, the process has the standard state-space representation

F107

$$\dot{x} = Ax + Bu + Ex_0$$

$$y = Cx$$
(2G.4)

with

$$A = \begin{bmatrix} a_{11} & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 \\ 0 & a_{32} & 0 & 0 \\ 0 & a_{42} & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & 0 \\ 0 & 0 \\ 0 & b_{32} \\ 0 & b_{42} \end{bmatrix} \qquad E = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ f_{31} & f_{32} \\ 0 & f_{42} \end{bmatrix}$$
(2G.5)
$$C = \begin{bmatrix} 0 & 0 & c_{13} & 0 \\ 0 & 0 & c_{13} & 0 \\ 0 & 0 & 0 & c_{24} \end{bmatrix}$$

Numerical data for a specific process considered by Gilles and Retzbach are as follows:

Time is measured in hours, and temperature in degrees Celsius.

Example 2H Double effect evaporator Over a period of several years in the mid 1970s a group of chemical engineering faculty members and students at the University of Alberta developed a laboratory pilot plant which could be used to test various concepts and control system design techniques. The results of some of these studies have been published in a number of technical journals and reprinted as a case study [9].

The pilot plant is a double-effect evaporator shown schematically in Fig. 2.20. According to Professors D. G. Fisher and D. E. Seborg, leaders of the project and authors of the case study: "The first effect is a short-tube vertical calandria-type unit with natural circulation. The 9-in diameter unit has an operating holdup of 2 to 4 gallons, and its 32 stainless steel tubes, $\frac{3}{4}$ -in o.d. by 18 in. long, provide approximately 10 square feet of heat transfer surface altogether.







"The second stage is a long-tube vertical effect setup for either natural or forced circulation. It has a heat transfer area of 5 square feet and is made up of three 6-ft long 1-in o.d. tubes. Capacity of the circulating system is about 3 gallons."[10]

The inputs to the plant are steam and a concentrated solution of triethylene glycol. The outputs are glycol, the concentration and flow rate of which is to be controlled, and the condensate.

The system is a relatively complicated dynamic process requiring many state variables for its accurate description. A number of studies, reported in [11], were undertaken aimed at developing a model that represents a reasonable compromise between fidelity to the real process and amenability to control system designs. On the basis of such considerations the investigators found that a fifth-order model is in close agreement with a tenth-order model, the latter fitting the pilot plant test data very well.

The fifth-order model uses the state variables

$x_1 = W_1 = $ first-effect "holdup"	
$x_2 = C_1 = $ first-effect concentration	
$x_3 = H_1 = $ first-effect enthalpy	(2H.1)
$x_4 = W_2 =$ second-effect "holdup"	
$x_5 = C_2$ = second-effect concentration	

and control variables

 $u_1 = S_1 = \text{first-effect steam flow rate}$ $u_2 = B_1 = \text{first-effect "bottoms" flow rate}$ (2H.2) $u_3 = B_2 = \text{second-effect "bottoms" flow rate}$

In addition to the state and control variables there are also disturbance inputs to the process

$$d_1 = F_1 =$$
variations in feed flow rate
 $d_2 = C_{F1} =$ variations in feed concentration (2H.3)
 $d_3 = H_{F1} =$ variations in feed enthalpy



Figure 2.21 Fifth-order evaporator dynamic model.



The linearized differential equations for this process have been developed by Newell and Fisher[9] and are in the standard state space form

$$\dot{x} = Ax + Bu + Ex_0 \tag{2H.4}$$

For one particular configuration of the system, the numerical values of the matrices were found to be [11]

$$A = \begin{bmatrix} 0 & -.00156 & -.0711 & 0 & 0 \\ 0 & -.1419 & .0711 & 0 & 0 \\ 0 & -.00875 & -1.102 & 0 & 0 \\ 0 & -.00128 & -.1489 & 0 & -.0013 \\ 0 & .0605 & .1489 & 0 & -.0591 \end{bmatrix}$$
(2H.5)
$$B = \begin{bmatrix} 0 & -.143 & 0 \\ 0 & 0 & 0 \\ .392 & 0 & 0 \\ 0 & .108 & -.0592 \\ 0 & -.0486 & 0 \end{bmatrix}$$
(2H.6)
$$E = \begin{bmatrix} .2174 & 0 & 0 \\ -.074 & .1434 & 0 \\ -.036 & 0 & .1814 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
(2H.7)

A block-diagram representation of the system, using the structure implied by (2H.5)-(2H.7) is shown in Fig. 2.21. Over the period that the process was in operation various changes were made that result in changes in numerical values in the matrices,[9] but the structure of Fig. 2.21 did not change.

PROBLEMS

Problem 2.1 Motor-driven cart with inverted pendulum

The cart carrying the inverted pendulum of Example 2E is driven by an electric motor having the characteristics described in Example 2B. Assume that the motor drives one pair of wheels of the cart, so that the whole cart, pendulum and all, becomes the "load" on the motor. Show that the differential equations that describe the entire system can be written

$$\ddot{x} + \frac{k^2}{Mr^2R} \dot{x} + \frac{mg}{M} \theta = \frac{k}{MRr} e$$
$$\dot{x} - \left(\frac{M+m}{Ml}\right)g\theta - \frac{k^2}{Mr^2Rl} \dot{x} = -\frac{k}{MRrl} e$$

where k is the motor torque constant, R is the motor resistance (both as described in Example 2B), r is the ratio of motor torque to linear force applied to the cart ($\tau = rf$), and e is the voltage applied to the motor.

Problem 2.2 Motor-driven inverted pendulum

Derive the dynamic model for an inverted pendulum pivoted at its lower end and driven by an electric motor, as shown in Fig. 6.3.



Figure P2.1 Inverted pendulum on cart.

Show that the dynamic equations of the inverted pendulum on a cart of Prob. 2.1 reduce to that of a pendulum fixed at its lower end as the mass of the cart becomes infinite.

Problem 2.3 Three-capacitance thermal system

A conducting bar (Fig. P2.3(a)) is insulated along its length but exposed to the ambient temperature at one end, and heated at the other end. An approximate electrical equivalent, based on "lumping" the bar into three finite lengths, is shown in Fig. P2.3(b).

Write the differential equations for the system using as state variables v_1 , v_2 , and v_3 , the capacitor voltages. The input u is the temperature e_0 at the heated end, and the output y is the temperature v_3 at point 3 on the rod, as would be determined by a thermocouple, for example.



Figure P2.3 Three-capacitance thermal system. (a) Cross-sectional view; (b) Electrical analog.

Problem 2.4 Spring-coupled masses

Use Lagrange's equations (Sec. 2.4) to derive the dynamic equations of a pair of masses connected by a spring as shown in Fig. P2.4.

As the state variables use





Figure P2.4 Spring-coupled masses (two-car train).

Problem 2.5 Two-car train

An idealized two-car train consists of a pair of masses coupled by a spring, as shown in Fig. P2.4. The wheels of each car are independently driven by an electric motor such as described in Example 2B. (Also see Prob 2.1.)

(a) Express the differential equations of the system in state-space form. (Find the A and B matrices.) Assume R is the motor resistance, K is the spring constant, k is the motor torque constant, and $r = \tau/f$ is the ratio of the motor torque to the linear force applied to the car. Use the following state and control variables

 $x_1 = z_1$ $x_2 = z_2$ $x_3 = \dot{z_1}$ $x_4 = \dot{z_2}$ $u_1 = e_1$ $u_2 = e_2$

where e_1 and e_2 are the voltages on the drive motors.

(b) Draw the block diagram of the system.

Problem 2.6 Missile guidance dynamics

The geometry of a missile and target, both confined to move in a plane, is depicted in Fig. P2.6. The target moves in a straight line at constant velocity V_T and the missile moves at constant speed V_M but the direction of the velocity vector can be controlled by the use of an acceleration a which is assumed *perpendicular* to the relative velocity vector $\vec{V} = \vec{V}_M - \vec{V}_T$.

(a) Using a coordinate system that is "attached to the target" show that the dynamics of relative motion are

$$r = -V \cos \sigma$$

$$\dot{\lambda} = V \sin \sigma / r$$

$$\dot{\sigma} = V \sin \sigma / r + a / V$$
(P2.6(a))



Figure P2.6 Missile dynamics.

where r is the range to the target, λ is the inertial line-of-sight angle, σ is the angle subtended at the missile by the velocity vector and the line of sight, and a is the applied acceleration.

(b) Let z be the "distance of closest approach" of the missile to the target, under the assumption that the missile continues in a straight line without any further acceleration. (Sometimes z is called the *projected miss distance*.) Show that

$$z = r \sin \sigma$$

and, using P2.6(a)

$$\dot{z} = (r \cos \sigma / V)a \tag{P2.6(b)}$$

(c) Assume that σ is a small angle. Then $\dot{r} \simeq -V$. Thus $r(t) = r_0 - Vt$, then $r/V \approx T_0 - t = \bar{T}$ where $T_0 = r_0/V$: \bar{T} is often called "time-to-go." Show that the following equations represent the approximate dynamics

$$\dot{\lambda} = \frac{1}{V\bar{T}^2} z \tag{P2.6(c)}$$
$$\dot{z} = \bar{T}a$$

These equations are studied further in Chap. 3, Example 3B.

NOTES

Note 2.1 Rigid body dynamics

The representation of the motion of a rigid body in a noninertial coordinate system (i.e., a coordinate system in which the reference axes rotate and accelerate linearly) is treated in most standard textbooks in classical dynamics, such as Goldstein.[2] The basic relationship with regard to axes fixed in a rotating body are expressed by

$$\left. \frac{d\vec{\mu}}{dt} \right|_{\text{inertial}} = \left. \frac{d\vec{\mu}}{dt} \right|_{\text{body}} + \vec{\omega} \times \vec{\mu}$$

where $\vec{\mu}$ is any (three-component) vector. The symbol \times denotes the vector cross product and "body" means that the derivatives are taken as if the body axes were inertially fixed; $\vec{\omega}$ is the angular velocity of the body axes. Thus (2.27) can be expressed as

$$\vec{\tau} = \frac{d\vec{h}}{dt}\Big|_{\text{inertial}} = \frac{d\vec{h}}{dt}\Big|_{\text{body}} + \vec{\omega} \times \vec{h}$$

Since $\vec{h} = J\vec{\omega}$, this becomes

$$\vec{\tau} = J \frac{d\vec{\omega}}{dt} + \vec{\omega} \times J\vec{\omega}$$

which is the form in which the "Euler equation" appears in many books.

Note 2.2 The gyroscope

The gyroscope is one of the two basic components of all inertial navigation systems. (The other is an accelerometer.) Since the 1920s, gyroscopes (or "gyros" as they are now known) have been used in navigation systems, first in gyro compasses and more recently (i.e., since about 1950) in complete inertial navigation systems. In addition to being used in navigation systems, gyros are also used as motion sensors for stabilizing the motion of ships, aircraft, and other mechanical systems. The inner ear of a human has a vestibular system that includes three gyroscopic sensors known as "semicircular canals" which are important in the biological feedback system that maintains the human body upright.



The remarkable properties of spinning bodies have always been a source of fascination, not only to children, but also to mathematicians and physicists. The renowned mathematician Felix Klein, one of the founders of the field of topology, also wrote a famous treatise on the theory of tops. [12]

Note 2.3 Three-Mile Island

The near disaster caused by the sequence of failures at the Three-Mile Island (Pennsylvania) nuclear plant in 1979, has a number of valuable, if costly, lessons. The failure was not due to use of novel, untested design concepts nor to new state-of-the-art hardware having been insufficiently tested. Neither the design nor the hardware were flawed in principle. The combination of misfortune, lack of training, and deficient critical judgment were in part responsible for the accident that may well have spelled doom for the nuclear industry in the United States. An outstanding account of the Three-Mile Island incident and its implications were presented in the November 1979 issue of the IEEE Spectrum.[13]

Note 2.4 Swedish papermaking industry

The benefits of using modern control concepts in the field of process control were vividly demonstrated by Karl J. Åström, now a professor at the Lund Institute of Technology in Sweden. During the late 1950s and early 1960s, Åström, under sponsorship of IBM, in association with a group of investigators (including R. E. Kalman, J. E. Bertram) at Columbia University, initiated the investigation of the use of state-space methods for improved process control design, particularly in papermaking. After his return to Sweden he succeeded in persuading the management of a paper company that the improved performance using modern methods, and implemented by means of a digital computer, would more than justify the cost of the new installation. With the cooperation of the plant management, he performed the tests needed to get the required dynamic characteristics of the plant and then installed the new computer control. The results were outstandingly successful; within a few years much of the Swedish paper industry adopted the new control system design approach. A technical account of Åström's work is found in Chap. 6 of [14].

REFERENCES

- Loh, N. K., Cheok, K. C., and Beck, R. R., "Modern Control Design for Gun-Turret Control System," Southcon/85 Convention Record, Atlanta GA, Paper no. 1/5, 1983.
- 2. Goldstein, H., Classical Mechanics, Addison-Wesley Publishing Co., Reading MA, 1953.
- 3. Synge, J. L., and Griffith, B. A., Principles of Mechanics, McGraw-Hill Book Co., New York, 1949.
- 4. Etkin, B., Dynamics of Flight, John Wiley & Sons, New York, 1959.
- 5. Seckel, E., Stability and Control of Airplanes and Helicopters, Academic Press, New York, 1964.
- Gilles, E. D., and Retzbach, B., "Reduced Models and Control of Distillation Columns with Sharp Temperature Profiles," *IEEE Trans. on Automatic Control*, vol. AC-28, no. 5, May 1983, pp. 628-630.
- 7. Gilles, E. D., Retzbach, B., and Silberberger, F., "Modeling, Simulation, and Control of an Extractive Distillation Column," *Computer Applications to Chem. Eng.* (ACS Symp. ser. no. 124), 1980, pp. 481-492.
- 8. Retzbach, B., "Einsatz von systemtechnische Methoden am Beispeil einer Mehrstoffdestillation," Chem.-Ing.-Techn., vol. 55, no. 3, 1983, p. 235.
- 9. Fisher, D. G., and Seborg, D. E. (eds.), Multivariable Computer Control: A Case Study, North-Holland Publishing Co., Amsterdam, 1976.
- Fisher, D. G., and Seborg, D. E., "Advanced Computer Control Improves Process Performance," Instrumentation Technology, vol. 20, no. 9, September 1973, pp. 71-77.

- 11. Fisher, D. G., and Seborg, D. E., "Model Development, Reduction, and Experimental Evalu-ation for an Evaporator," Ind. Eng. Chem. Process Design and Development, vol. 11, no. 2, February 1972, pp. 213-221.
- 12. Klein, F., and Sommerfeld, A., Über die Theorie des Kreisels (2 vol.), Teubner, Leipzig, 1897.
- Rich, J., and Sommerrein, P., Oser the Informe des Kreisers (2 vol.), Teubner, Leipzig, 1897.
 Special Issue: "Three-Mile Island and the Future of Nuclear Power," *IEEE Spectrum*, vol. 16, 16, 170. no. 11, November 1979.
- 14. Åström, K. J., Introduction to Stochastic Control Theory, Academic Press, New York, 1970.

