١٤٧٢

١١٣١٢١

کتابخانه مرکزی دانشگاه تربیت مدرس

٢.٤٥

Principles and Techniques of Vibrations

Leonard Meirovitch

College of Engineering
Virginia Polytechnic Institute & State University

١٤٠٩

PRENTICE-HALL INTERNATIONAL, INC.

QC
235
.M48
1997

This edition may be sold only in those countries to which it is consigned by Prentice-Hall International. It is not to be re-exported, and it is not for sale in the USA, Mexico, or Canada.

© 1997 by Prentice-Hall, Inc.

A Simon & Schuster Company
Upper Saddle River, New Jersey 07458

The author and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and formulas to determine their effectiveness. The author and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these formulas.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-270430-7

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Simon & Schuster Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*
Prentice Hall, Upper Saddle River, *New Jersey*

Contents

Preface

Vibration phenomena are omnipresent: many occur in nature, and many in man-made devices and structures. Quite often vibration is not desirable and the interest lies in reducing it. Before such a task can be undertaken, it is necessary to have a good understanding of the vibration characteristics. This requires the construction of a mathematical model acting as a surrogate for the actual system, where models range from simple single-degree-of-freedom systems to multi-degree-of-freedom lumped systems and continuous systems. The term mathematical model implies a mathematical formulation, which consists of ordinary differential equations of motion for lumped, or discrete systems and boundary-value problems for continuous, or distributed systems. Of course, to study the vibration phenomena, it is necessary to solve the equations of motion, or the boundary-value problem, for the system response.

The area of vibrations, and in particular analytical vibrations, has experienced significant progress in the last few decades. A great deal of this progress can be attributed to the development of increasingly powerful digital computers. Indeed, this rapidly increasing power of solving complex problems has stimulated the development of many numerical methods, which in turn has encouraged the development of more refined theories.

For many years, the theory of vibrations was a subject for physicists and mathematicians. As vibration problems became more and more critical, the interest in the subject among engineers grew rapidly. In addressing these challenging problems, engineers have created a modern approach to vibrations by developing a variety of techniques for solving complex problems. In the process, they have not been merely building on an existing mathematical foundation but also extending this foundation, incorporating the best of the old into the new. It is in this spirit that this book has been written.

The main objective of the book is to present a mathematically rigorous approach to vibrations, one that not only permits efficient formulations and solutions to problems but also enhances the understanding of the physics of the problem. To this end, various principles and techniques for formulating and solving vibration problems are presented. To enhance understanding, a broad view is adopted whereby emphasis is placed on the similarity of the dynamic characteristics exhibited by various types of vibrating systems, a similarity that may not be obvious to the uninitiated. This similarity has significant practical implications, as solutions for different types of systems can be obtained in analogous manner. As with any profession, tasks are made appreciably easier by working with the right tools. In this regard, this book strives to strike a proper balance between problem formulations and solutions by providing effective methods for both. For the derivation of the equations of motion or boundary-value problems for complex systems, the principles and techniques of analytical dynamics have few peers. For solving ordinary differential equations of motion for lumped, or discrete systems, concepts from linear system theory and linear algebra prove indispensable. In treating boundary-value problems for continuous, or distributed systems, operator notation is very useful, as it permits the treatment of whole classes of systems, instead of individual cases. Closed-form solutions for systems with distributed parameters are not plentiful, so that there is a great deal of interest in approximate solutions. This invariably amounts to approximating distributed systems by discrete ones, which further amounts to replacing partial differential equations by sets of ordinary differential equations, a task that can be carried out conveniently by variational techniques.

A chapter-by-chapter review of the book should help develop a better appreciation for some of the statements made above. The review is as follows:

Chapter 1 is devoted to concepts and techniques from linear system theory. It is concerned with the relation between excitation and response and includes selected topics from the theory of ordinary differential equations and matrix theory. Concepts such as the superposition principle, frequency response, impulse response, convolution integral, state equations, transition matrix and discrete-time systems are presented.

Chapter 2 provides a review of Newtonian mechanics and a comprehensive discussion of analytical dynamics. Concepts such as linear and angular momentum, work and energy, generalized coordinates and degrees of freedom, are reviewed and topics such as the virtual work principle, d'Alembert's principle, Hamilton's principle, Lagrange's equations, Hamilton's equations and conservation laws are presented in great detail.

Chapter 3 contains a discussion of single-degree-of-freedom systems, typical of a first course on vibrations. It represents an application of the developments of Chapter 1 to the vibration of simple systems. Many of the results obtained in this chapter are to be used in later chapters.

Chapter 4 is concerned with the vibration of discrete systems. A geometric description of the motion in the state space is used to explain concepts such as equilibrium points, motion stability and linearization about equilibrium points. It is here that the algebraic eigenvalue problem is first introduced. The solution of this problem is pivotal to the modal analysis for system response. Derivation of the system response is carried out both in continuous and in discrete time. Procedures for the numerical solution of nonlinear differential equations are presented for the case in which linearization is not suitable.

Chapter 5 begins the process of building a mathematical foundation for modern vibrations. Ironically, some of the material can be traced to developments over a century old. Topics considered are the geometric interpretation of the eigenvalue problem, Rayleigh's quotient and its stationarity, Rayleigh's principle, the Courant-Fischer maximin theorem and the separation theorem all helping characterize the eigenvalues of natural and gyroscopic conservative systems. Some of these concepts are essential to the variational approach to the eigenvalue problem. Also included is a perturbation of the eigenvalue problem, a subject somewhat apart from the rest of the chapter.

Chapter 6 contains a variety of computational algorithms for the algebraic eigenvalue problem, both for conservative and nonconservative systems, which translates into symmetric and nonsymmetric problems. Some of the algorithms, such as Gaussian elimination, Cholesky decomposition, matrix tridiagonalization and reduction to Hessenberg form, play a supporting role only. Actual algorithms for solving the symmetric eigenvalue problem include the power method, the Jacobi method, Givens method, the QR method, inverse iteration, the Rayleigh quotient iteration and simulataneous iteration. Algorithms for the nonsymmetric eigenvalue problem include the power method, the QR method and inverse iteration, all suitably modified to accommodate complex eigensolutions. Although some of the algorithms can be found in computer software packages, such as IMSL, Matlab, etc., the reader can only benefit from a closer exposure to the methods presented in this chapters.

Chapter 7 is devoted entirely to formulations and solutions of boundary-value problems for distributed-parameter systems. This chapter contains a great deal of information, some of it basic and some more specialized in nature. Fundamental material includes the derivation of boundary-value problems for strings, rods, shafts and beams using the extended Hamilton's principle, Lagrange's equation for distributed systems, the differential eigenvalue problem, closed-form solutions of the eigenvalue problem, membrane and plate vibration, variational and integral formulations of the eigenvalue problem and system response. More specialized material (that can be omitted at a first reading) includes extensions of Lagrange's equations, generalization of the differential eigenvalue problem, systems with boundary conditions depending on the eigenvalue, Timoshenko beam and systems with nonhomogeneous boundary conditions.

Chapter 8 is concerned with techniques for the approximate solution of eigenvalue problems for distributed systems. All techniques approximate differential eigenvalue problems by means of algebraic ones, some through lumping the system parameters at discrete point and others by assuming a solution in the form of a finite series of trial functions. Methods of the first type included in this chapter are the lumped-parameter method using influence coefficients, Holzer's method and Myklestad's method. Methods of the second type are Rayleigh's energy method, the Rayleigh-Ritz method and methods of weighted residuals, such as Galerkins' method and the collocation method. Two other closely related series discretization procedures, component-mode synthesis and substructure synthesis, extend the Rayleigh-Ritz method to structures in the form of assemblages of substructures.

Chapter 9 is devoted exclusively to the most versatile and widely used of the series discretization techniques, the finite element method. The chapter begins by demonstrating that the finite element method is a potent version of the Rayleigh-Ritz method. Coverage includes a matrix approach for the generation of interpolation functions for strings (and hence for rods and shafts), beams, membranes and plates, the derivation of element mass and stiffness matrices for these systems and the assembly process for producing global mass and stiffness matrices. Also discussed are an estimation of errors in the eigenvalues and eigenfunctions, the hierarchical version of the finite element method and system response.

Appendix A presents a brief introduction to the Laplace transformation method including the basic tools for deriving the response of linear systems with constant coefficients. Appendix B provides elements from linear algebra of interest in a modern treatment of vibrations, and in particular vector spaces and matrices.

The book is intended as a text for a one-year course on vibrations at the graduate level and as a reference for practicing engineers. It is also suitable for self-study. A first course can be based on material culled from the following: Ch. 1, Secs. 2.1-2.11, Ch. 3, a selection of topics from Ch.4, Secs. 5.1 and 5.2, Secs. 6.1-6.4 and Secs. 7.1-7.7. The material for a second course is more difficult to prescribe, but should include a review of material from the first course, a sampling of algorithms from Secs. 6.5-6.12, a selection from Secs. 7.8-7.18, Secs. 8.4-8.8, 8.10, and perhaps a sampling from Secs. 8.1-8.3, and Ch. 9. The book contains an ample selection of homework problems. On occasions, several problems cover the same subject so as to permit changes in the homework assignment from year to year.

Inevitably, comparisons will be made between this book and my first book, Analytical Methods in Vibrations (Macmillan, 1967). The idea was to combine the best of the old, material that stood the test of time, with the new. Some of the old material included has undergone various improvements, such as a broader and more rigorous treatment of the principles of analytical dynamics, an expanded and enhanced discussion of the qualitative aspects of the eigenvalue problem and a more complete discussion of boundary-value problems for distributed systems. The new material reflects the significant progress made in the last three decades in the treatment of discrete systems and in approximate techniques for distributed systems. The most noteworthy of the latter is the inclusion of a chapter on the finite element method.

A great deal of credit for this project belongs to David E. Johnstone, Senior Editor, the Macmillan Co., a man of great charm and dedication, well liked by all who knew him. Sadly, Dave was not able to see the fruits of his efforts, cut down unnecessarily in the prime of his life and his career. Fortunately, another man of great ability and dedication, and one with whom I have been enjoying an excellent professional relation, Bill Stenquist, Executive Editor, Prentice Hall, Inc., has taken charge of the project, seeing it to the successful completion Dave would have loved to see. Many thanks go to both of them.

I would like to take this opportunity to thank Jan Hjelmgren, Virginia Polytechnic Institute and State University, Kent L. Lawrence, University of Texas, Arlington, Fai Ma, University of California, Berkeley, and Harold D. Nelson, Texas Christian University, for reviewing the manuscript and making valuable suggestions, as well as to Sungsoo Na, Virginia Polytechnic Institute and State University, for reviewing an early version of the manuscript. Special thanks are due to Timothy J. Stemple, Virginia Polytechnic Institute and State University, for producing numerical solutions to many of the example problems and for generating the computer plots. Last but not least, I wish to thank Norma B. Guynn for her excellent typing of the manuscript.

Leonard Meirovitch

1

CONCEPTS AND TECHNIQUES FROM LINEAR SYSTEM THEORY

A *system* is defined as an assemblage of parts or components acting together as a whole. When acted upon by a given *excitation*, or *input*, a system exhibits a certain *response*, or *output*. Systems can be divided into two major classes, linear and nonlinear, depending on how a system responds to excitations. Broadly speaking, linear system theory is concerned with the behavior of linear systems, and in particular with the excitation-response relation. The theory is of fundamental importance to the study of vibrations and control. It includes selected topics from the theory of ordinary differential equations and matrix theory.

For the most part, the interest in vibration theory lies in *linear time-invariant systems*, i.e., systems described by linear differential equations with constant coefficients. The problem of deriving the response of linear time-invariant systems to given excitations can be treated in the frequency domain or in the time domain. Frequency-domain techniques are most indicated when the excitation is harmonic, periodic, or random, while the time-domain approach is recommended for initial and arbitrary excitations. Due to the nature of the applications presented throughout this text, the emphasis in this chapter is on time-domain techniques, with some frequency-domain methods being introduced as demand arises. For low-order systems, the excitation-response relation can be expressed as a scalar single-input, single-output relation. For higher-order systems, the excitation-response is more conveniently expressed as a matrix multi-input, multi-output relation. For integration purposes, first-order differential equations have in general a distinct advantage over second-order equations, which implies a formulation in terms of state variables.

This chapter is devoted largely to linear time-invariant systems. Concepts and techniques from linear system theory fundamental to vibration theory are introduced by means of some generic differential equations. Specific differential equations of motion for various vibrating systems are derived throughout this text, at which point these concepts and techniques are used to solve them. Hence, in many ways, this chapter serves as a reference source.

1.1 PROBLEM FORMULATION

The motion of simple mechanical systems, such as a single mass particle acted upon by forces, is governed by *Newton's second law*. The law is commonly expressed in terms of a differential equation, which can be written in the generic form

$$a_0 \frac{d^2x(t)}{dt^2} = r\left(x(t), \frac{dx(t)}{dt}, t\right) \tag{1.1}$$

where $x(t)$ is the displacement at time t and $r(x(t),\ dx(t)/dt,\ t)$ is a function representing the resultant force. We assume that the force function can be expressed more explicitly as

$$r\left(x(t), \frac{dx(t)}{dt}, t\right) = -a_1 \frac{dx(t)}{dt} - a_2 x(t) + f(t) \tag{1.2}$$

where $-a_1\, dx(t)/dt$ and $-a_2 x(t)$ are internal forces and $f(t)$ is an external force. We shall examine the nature of the forces later in this text. Introducing Eq. (1.2) into Eq. (1.1), we can write the differential equation of motion in the form

$$a_0 \frac{d^2x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + a_2 x(t) = f(t) \tag{1.3}$$

In vibrations and controls, x is known as the *response* and f as the *excitation*, or as *output* and *input*, respectively. The relation between the excitation and response for a given system is given schematically in Fig. 1.1, in which the system is represented by a box, the excitation by an incoming arrow and the response by an outgoing arrow. A diagram of the type shown in Fig. 1.1 is known as a *block diagram*. It states simply that when the system is acted upon by the excitation $f(t)$ it exhibits the response $x(t)$.

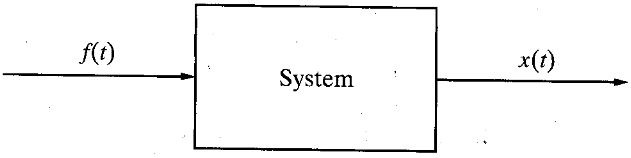


Figure 1.1 Block diagram relating the output to the input in the time domain

The differential equation, Eq. (1.3), can be written in the compact operator form

$$Dx(t) = f(t) \tag{1.4}$$

where

$$D = a_0 \frac{d^2}{dt^2} + a_1 \frac{d}{dt} + a_2 \tag{1.5}$$

represents a *differential operator*. Consistent with this, the block diagram of Fig. 1.1 can be replaced by the block diagram of Fig. 1.2, in which the system is represented by the operator D. This is quite appropriate because the operator provides a full description of the system characteristics, as can be concluded from Eq. (1.5).

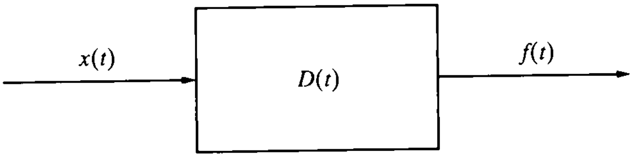


Figure 1.2 Block diagram relating the input to the output through the operator D

The study of vibrations is concerned with all aspects of the excitation–response, or input–output relation. Clearly, the first task is to establish this relation, which consists of the derivation of the equation of motion. Then the problem reduces to the determination of the response, which amounts to solving the differential equation. Before an attempt to solve the differential equation is made, however, it is important to establish the characteristics of the system, as these characteristics determine how the system responds to excitations. This, in turn, affects the choice of methodology to be used in producing the solution. In particular, as we shall see in the next section, it is very important to establish whether the system is linear or nonlinear and whether it is time-invariant or time-varying. The choice of methodology depends also on the type of excitation. In this chapter, we discuss various aspects of the input–output relation by means of the generic differential equation presented above, and in Chapter 2 we specialize the discussion to basic vibrating systems.

1.2 SYSTEM CLASSIFICATION. THE SUPERPOSITION PRINCIPLE

As pointed out in Sec. 1.1, in the study of vibrations it is important to verify early the characteristics of a system, as these characteristics determine the manner in which the system responds to excitations. This, in turn, permits a judicious choice of methodology best suited for deriving the response. In Sec. 1.1, we were conspicuously silent about the terms a_0, a_1 and a_2 in Eq. (1.3). To explore the nature of the system, and how it affects the response, we must now break this silence.

Let us consider the system described by Eq. (1.4), where $f(t)$ is the excitation, or the input, $x(t)$ is the response, or the output, and D is a differential operator. The system is said to be *linear* if it satisfies the two conditions:

1. The response to $\alpha f(t)$ is $\alpha x(t)$, where α is a constant.
2. The response to $f_1(t) + f_2(t)$ is $x_1(t) + x_2(t)$, where $x_1(t)$ is the response to $f_1(t)$ and $x_2(t)$ is the response to $f_2(t)$.

The first condition is satisfied if the operator D is such that

$$D[\alpha x(t)] = \alpha D x(t) \tag{1.6}$$

In this case, the operator D, and hence the system, is said to possess the *homogeneity* property. The second condition is satisfied if

$$D\left[x_1(t) + x_2(t)\right] = Dx_1(t) + Dx_2(t) \tag{1.7}$$

In this case, the operator D, and hence the system, possesses the *additivity* property. A differential operator D satisfying Eqs. (1.6) and (1.7) is referred to as a *linear homogeneous differential operator.* If an operator D does not possess the homogeneity and additivity properties, the system is *nonlinear.*

The question of whether a system is linear or nonlinear has enormous implications in vibrations, so that a simple criterion enabling us to classify a system is highly desirable. To this end, we consider the case in which the differential operator D is defined by

$$Dx(t) = a_0(t)\frac{d^2x(t)}{dt^2} + a_1(t)\frac{dx(t)}{dt} + a_2(t)\left[1 - \epsilon x^2(t)\right]x(t) \tag{1.8}$$

where ϵ is a constant, and propose to verify of homogeneity and additivity by means of Eqs. (1.6) and (1.7), respectively. Replacing $x(t)$ by $\alpha x(t)$ in Eq. (1.8), we obtain

$$\begin{aligned}
D\left[\alpha x(t)\right] &= a_0(t)\frac{d^2[\alpha x(t)]}{dt^2} + a_1(t)\frac{d[\alpha x(t)]}{dt} + a_2(t)\left[1 - \epsilon\alpha^2 x^2(t)\right]\alpha x(t) \\
&= \alpha\left[a_0(t)\frac{d^2x(t)}{dt^2} + a_1(t)\frac{dx(t)}{dt} + a_2(t)x(t)\right] - a_2(t)\epsilon\alpha^3x^3(t) \\
&\neq \alpha D(t)
\end{aligned} \tag{1.9}$$

Similarly, letting $x(t) = x_1(t) + x_2(t)$ in Eq. (1.8), we can write

$$\begin{aligned}
D\left[x_1(t) + x_2(t)\right] &= a_0(t)\frac{d^2x_1(t)}{dt^2} + a_1(t)\frac{dx_1(t)}{dt} + a_2(t)\left[1 - \epsilon x_1^2(t)\right]x_1(t) \\
&\quad + a_0(t)\frac{d^2x_2(t)}{dt^2} + a_1(t)\frac{dx_2(t)}{dt} + a_2(t)\left[1 - \epsilon x_2^2(t)\right]x_2(t) \\
&\quad - 3a_2(t)\epsilon\left[x_1^2(t)x_2(t) + x_1(t)x_2^2(t)\right] \\
&\neq Dx_1(t) + Dx_2(t)
\end{aligned} \tag{1.10}$$

It is easy to see from Eqs. (1.9) and (1.10) that the operator D possesses neither the homogeneity nor the additivity property, so that the system is nonlinear.

A cursory examination of Eqs. (1.9) and (1.10) reveals that the nonlinearity of the system is caused by the term containing ϵ. In the case in which ϵ is equal to zero, Eq. (1.8) reduces to

$$Dx(t) = a_0(t)\frac{d^2x(t)}{dt^2} + a_1(t)\frac{dx(t)}{dt} + a_2(t)x(t) \tag{1.11}$$

It is not difficult to verify that the system defined by Eq. (1.11) possesses both the homogeneity and additivity properties, so that the system is indeed linear.

The above example permits us to make the following observations:

1. A system is linear if the function $x(t)$ and its derivatives appear to the first (or zero) power only; otherwise, the system is nonlinear.
2. A system is linear if a_0, a_1 and a_2 depend on time alone, or they are constant.

Next, we consider another important system characteristic affecting the input-output relation. In particular, we ask the question as to how the relation is affected by a shift in time, as this question has significant implications in the derivation of the system response. To answer this question, we return to the system described by Eq. (1.4) and shift the time scale in both the input $f(t)$ and output $x(t)$. We denote the input and output delayed by the amount τ by $f(t-\tau)$ and $x(t-\tau)$, respectively. Then, if the delayed input and output satisfy the equation

$$Dx(t-\tau) = f(t-\tau) \tag{1.12}$$

the system is said to be *time-invariant*. Otherwise, the system is *time-varying*. As an example, we consider

$$Dx(t) = a_0\frac{d^2x(t)}{dt^2} + a_1\frac{dx(t)}{dt} + a_2(1-\epsilon\sin t)x(t) = f(t) \tag{1.13}$$

where a_0, a_1, a_2 and ϵ are constant. Shifting the input and output, we obtain

$$Dx(t-\tau) = a_0\frac{d^2x(t-\tau)}{dt^2} + a_1\frac{dx(t-\tau)}{dt} + a_2(1-\epsilon\sin t)x(t-\tau) \neq f(t-\tau) \tag{1.14}$$

Because Eq. (1.14) contradicts Eq. (1.12), the system is time-varying. Clearly, the term preventing the system from being time invariant is $-a_2\epsilon\sin t$, which is the only term in Eq. (1.13) in which the time t appears explicitly. This permits us to draw the conclusion that *if at least one of the coefficients of the differential equation depends explicitly on time, the system is time-varying*. Time-varying systems are also known as systems with *time-dependent coefficients*. Similarly, *if the coefficients of the differential equation are constant, the system is time-invariant*, more commonly known as a *system with constant coefficients*. The coefficients represent the system parameters. In future discussions, we will refer to the property reflected by Eq. (1.12) as the *time-invariance property*. For the most part, the classification of a system can be carried out by mere inspection of the differential equation. In vibrations, there is considerable interest in linear time-invariant systems.

At this point, we propose to extend our discussion to linear systems of nth order. To this end, we consider the differential equation (1.4), where

$$D = a_0(t)\frac{d^n}{dt^n} + a_1(t)\frac{d^{n-1}}{dt^{n-1}} + \dots + a_{n-1}(t)\frac{d}{dt} + a_n(t) \tag{1.15}$$

is a linear homogeneous differential operator of order n. The assumption of linearity implies that the excitation and response can be written in the form of the linear combinations

$$f(t) = \sum_{j=1}^{m}\alpha_j f_j(t)\,,\quad x(t) = \sum_{j=1}^{m}\alpha_j x_j(t) \tag{1.16a,b}$$

where $x_j(t)$ represents the response of the system described by Eq. (1.4) to the excitation $f_j(t) (j = 1, 2, \ldots, m)$. Inserting Eqs. (1.16) into Eq. (1.4), considering Eq. (1.15) and carrying out the indicated operations, we can separate the result into the independent set of equations

$$Dx_j(t) = f_j(t)\ ,\ j = 1, 2, \ldots, m \tag{1.17}$$

Equations (1.16) and (1.17) permit us to make a very important statement, namely, *if a linear system is acted upon by a linear combination of individual excitations, the individual responses can be obtained separately and then combined linearly.* This statement is known as the *superposition principle* and it *applies to linear systems alone.* It makes the determination of the response of linear systems to complex inputs considerably easier than for nonlinear systems. Indeed, quite often it is possible to decompose a complex input into a linear combination of simple inputs, thus reducing the response problem to a series of simple problems. The superposition principle is fundamental to linear system theory. This text is devoted largely to linear systems, and in particular to linear systems with constant coefficients.

Before leaving this subject, we should point out that in many cases the distinction between linear and nonlinear systems is not as sharp as the foregoing discussion may have implied. Indeed, quite often the same system can behave both as linear and nonlinear, although not at the same time. As an example, for sufficiently small values of x that $\epsilon x^2 \ll 1$, the term ϵx^2 can be ignored in Eq. (1.8), so that the system can be approximated by a linear system. As x increase s, the term ϵx^2 increases in value relative to 1, so that it can no longer be ignored. Clearly, there is no sharp point at which the system changes from linear into nonlinear, and the shift is very gradual.

1.3 EXCITATION AND RESPONSE CHARACTERIZATION

The response of a system depends on the system and on the excitation characteristics. In Sec. 1.2, we explored ways of ascertaining the system characteristics by examining the coefficients of the differential equation. In this section, we turn our attention to the characterization of the excitation, as various types of excitations call for different approaches to the derivation of the response. It should be pointed out here that nonlinear systems require different methodology from linear systems, although some of the methodology for nonlinear systems relies on that for linear systems. Unless otherwise stated, over the balance of this chapter, we consider linear systems only, and in particular time-invariant linear systems. Moreover, we confine our discussion to second-order systems. Hence, we consider the differential equation

$$a_0 \frac{d^2x(t)}{dt^2} + a_1 \frac{dx(t)}{dt} + a_2 x(t) = f(t) \tag{1.18}$$

in which the coefficients $a_i (i = 0, 1, 2)$ are constant. For the moment, we keep the discussion fairly general, in the sense that we do not enter into details as to how the response to a given excitation is derived.

The solution of differential equations consists of two parts, namely, the *homogeneous solution* and the *particular solution.* With reference to Eq. (1.18), the homogeneous solution corresponds to the case in which the external force is zero, $f(t) \equiv 0$. Hence, the homogeneous solution is the solution to initial conditions alone. Because the order of the system is two, the solution requires two initial conditions, $x(0)$ and $dx(t)/dt\big|_{t=0}$. In the case of a mechanical system, the initial conditions can be identified as the initial displacement and velocity, respectively. On the other hand, the particular solution is the response to the external excitation alone, which implies the solution in the absence of the initial conditions. The homogeneous solution and the particular solution complement each other, so that the total solution is the sum of the homogeneous solution and the particular solution. This is a direct reflection of the superposition principle, according to which the response of linear systems to different excitations can be first obtained separately and then combined linearly. At this point, it is appropriate to inject a note of caution, as the above distinction between initial excitations and external excitations is not as airtight at it may appear. Indeed, in some cases the distinction is artificial, as some initial conditions are generated by external forces. We will have the opportunity to verify this statement in Sec. 3.5.

As pointed out earlier, the response depends not only on the system characteristics but also on the type of excitation. We begin our discussion with the response to external excitations. To this end, it is convenient to distinguish between *steady-state response* and *transient response*. The steady-state response can be defined broadly as a long-term response, which implies a response that persists long after short-term effects have disappeared. This implies further a steady-state excitation, such as constant, harmonic, or periodic. In describing the steady-state response, time becomes incidental. In fact, the steady-state response can be defined as the response of a system as time approaches infinity. On the other hand, the transient response depends strongly on time, and its presence is closely linked to the presence of the excitation. The excitation can be any nonperiodic function of time, which excludes steady-state functions, such as constant, harmonic and periodic functions. It is common practice to regard the response to initial excitations as transient, even when the effect persists indefinitely. Consistent with this, it is meaningless to add the response to initial excitations to a steady-state response. Indeed, the steady-state response is not materially affected by a shift in the time scale, while the response to initial excitations implies a certain initial time.

The various excitation functions discussed above have one thing in common, namely, their value at any given time can be specified in advance. Such functions are said to be *deterministic*. Yet there are cases in which it is not possible to specify the value of a function in advance. Indeed, in some cases there are so many unknown contributing factors that the function tends to acquire a certain randomness. Such functions are known as *nondeterministic*, or *random*; they are also known as *stochastic*. An example of a random function is runway roughness. Nondeterministic functions are described in terms of statistics, such as expected value, mean square value, etc., rather than time. If the excitation is nondeterministic, so is the response.

It is clear from the above that the system can be subjected to a large variety of inputs. Hence, it should come as no surprise that, to obtain the system response, it is necessary to employ a variety of approaches. Some of these approaches are discussed in this chapter and many others are presented in later chapters.

1.4 RESPONSE TO INITIAL EXCITATIONS

In Sec. 1.3, we established that the response of linear systems to initial excitations and external excitations can be first obtained separately and then combined linearly. In this section, we propose to derive a general expression for the response to initial excitations, and in subsequent sections we consider the response to a variety of external excitations.

Our interest lies in the case of linear, time-invariant systems. In the absence of external excitations, $f(t) \equiv 0$, Eq. (1.18) reduces to the homogeneous equation

$$Dx(t) = 0 \tag{1.19}$$

where

$$D = a_0 \frac{d^2}{dt^2} + a_1 \frac{d}{dt} + a_2 \tag{1.20}$$

is a homogeneous differential operator, in which the coefficients $a_i (i = 0, 1, 2)$ are constant. The solution $x(t)$ is subject to the initial conditions

$$x(0) = x_0, \qquad \left.\frac{dx(t)}{dt}\right|_{t=0} = v_0 \tag{1.21}$$

in which x_0 is the initial displacement and v_0 is the initial velocity.

Because Eq. (1.19) is homogeneous and possesses constant coefficients, its solution can be expressed in the exponential form

$$x(t) = Ae^{st} \tag{1.22}$$

where A and s are constants yet to be determined. Inserting Eq. (1.22) into Eq. (1.19), we obtain

$$DAe^{st} = ADe^{st} = 0 \tag{1.23}$$

Recalling that D involves derivatives with respect to time and observing that

$$\frac{d^r}{dt^r} e^{st} = s^r e^{st}, \quad r = 1, 2 \tag{1.24}$$

we can rewrite Eq. (1.23) in the form

$$ADe^{st} = AZ(s)e^{st} = 0 \tag{1.25}$$

where $Z(s)$ is a polynomial in s known as the *generalized impedance*. It is obtained from D by replacing derivatives with respect to time by s raised to a power equal to the derivative order, in conformity with Eq. (1.24). Recognizing that A and e^{st} cannot be zero, we conclude from Eq. (1.25) that

$$Z(s) = a_0 s^2 + a_1 s + a_2 = 0 \tag{1.26}$$

Equation (1.26) is known as the *characteristic equation*, and it has in general as many roots as the order of Z, where the roots are called *characteristic values*. In the case at hand, the system is of order two, so that there are two roots, s_1 and s_2. Introducing these roots into Eq. (1.22), we can write the solution of Eq. (1.19) as

$$x(t) = A_1 e^{s_1 t} + A_2 e^{s_2 t} \tag{1.27}$$

The exponents s_1 and s_2 depend on the system parameters, so that they represent inherent characteristics of the system. On the other hand, A_1 and A_2 are constants of integration, and they depend on the initial conditions. Hence, their values are determined by factors external to the system. To evaluate the constants of integration, we use Eqs. (1.21) in conjunction with Eq. (1.27) and write

$$x(0) = A_1 + A_2 = x_0, \qquad \left.\frac{dx(t)}{dt}\right|_{t=0} = s_1 A_1 + s_2 A_2 = v_0 \tag{1.28}$$

Equations (1.28) represent two nonhomogeneous algebraic equations in the unknowns A_1 and A_2 and have the solution

$$A_1 = \frac{\begin{vmatrix} x_0 & 1 \\ v_0 & s_2 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ s_1 & s_2 \end{vmatrix}} = \frac{s_2 x_0 - v_0}{s_2 - s_1}, \qquad A_2 = \frac{\begin{vmatrix} 1 & x_0 \\ s_1 & v_0 \end{vmatrix}}{\begin{vmatrix} 1 & 1 \\ s_1 & s_2 \end{vmatrix}} = \frac{-s_1 x_0 + v_0}{s_2 - s_1} \tag{1.29}$$

Finally, inserting Eqs. (1.29) into Eq. (1.27) and rearranging, we obtain the response of our second-order system to the initial excitations x_0 and v_0 in the form

$$x(t) = \frac{1}{s_2 - s_1}\left[\left(s_2 e^{s_1 t} - s_1 e^{s_2 t}\right) x_0 - \left(e^{s_1 t} - e^{s_2 t}\right) v_0\right] \tag{1.30}$$

and we recall that the exponents s_1 and s_2 are determined by solving the characteristic equation, Eq. (1.26). Hence, s_1 and s_2 depend on a_0, a_1 and a_2 and represent in general complex numbers.

Example 1.1

Obtain the response of the system described by the differential equation

$$\frac{d^2x(t)}{dt^2} + a\frac{dx(t)}{dt} + bx(t) = 0 \tag{a}$$

to the initial excitations $x(0) = x_0, dx(t)/dt|_{t=0} = v_0$. The coefficients a and b are constant.

Using Eq. (1.20), the system differential operator can be written as

$$D = \frac{d^2}{dt^2} + a\frac{d}{dt} + b \tag{b}$$

so that

$$De^{st} = \left(\frac{d^2}{dt^2} + a\frac{d}{dt} + b\right)e^{st} = (s^2 + as + b)e^{st} \tag{c}$$

Hence, using Eq. (1.26), the characteristic equation is

$$Z(s) = s^2 + as + b = 0 \tag{d}$$

where $Z(s)$ is the generalized impedance. Equation (d) represents a quadratic equation and its roots are the characteristic values

$$\begin{matrix} s_1 \\ s_2 \end{matrix} = -\frac{a}{2} \pm \frac{1}{2}\sqrt{a^2 - 4b} \tag{e}$$

The general expression for the response of a second-order system to initial excitations is given by Eq. (1.30). Hence, inserting Eqs. (e) into Eq. (1.30), we obtain

$$\begin{aligned} x(t) &= \frac{1}{s_2 - s_1}\left[\left(s_2 e^{s_1 t} - s_1 e^{s_2 t}\right) x_0 - \left(e^{s_1 t} - e^{s_2 t}\right) v_0\right] \\ &= -\frac{1}{\sqrt{a^2 - 4b}}\left\{\left[\left(-\frac{a}{2} - \frac{1}{2}\sqrt{a^2 - 4b}\right) e^{\left(-\frac{a}{2}+\frac{1}{2}\sqrt{a^2-4b}\right)t}\right.\right. \\ &\quad \left. - \left(-\frac{a}{2} + \frac{1}{2}\sqrt{a^2 - 4b}\right) e^{\left(-\frac{a}{2}-\frac{1}{2}\sqrt{a^2-4b}\right)t}\right] x_0 \\ &\quad \left. - \left[e^{\left(-\frac{a}{2}+\frac{1}{2}\sqrt{a^2-4b}\right)t} - e^{\left(-\frac{a}{2}-\frac{1}{2}\sqrt{a^2-4b}\right)t}\right] v_0\right\} \\ &= \frac{e^{-at/2}}{\sqrt{a^2 - 4b}}\left[\left(a \sinh \frac{1}{2}\sqrt{a^2 - 4b}\, t + \sqrt{a^2 - 4b} \cosh \frac{1}{2}\sqrt{a^2 - 4b}\, t\right) x_0\right. \\ &\quad \left. + 2v_0 \sinh \frac{1}{2}\sqrt{a^2 - 4b}\, t\right] \end{aligned} \tag{f}$$

It is not difficult to verify that the response given by Eq. (f) satisfies both initial conditions.

1.5 RESPONSE TO HARMONIC EXCITATIONS. FREQUENCY RESPONSE

The interest lies in the response of linear time-invariant systems to harmonic excitations. In Sec. 1.1, we have shown that the differential equation can be written in the operator form

$$Dx(t) = f(t) \tag{1.31}$$

where, from Eq. (1.15), the operator D for a linear time-invariant system of order n reduces to

$$D = a_0 \frac{d^n}{dt^n} + a_1 \frac{d^{n-1}}{dt^{n-1}} + \ldots + a_{n-1}\frac{d}{dt} + a_n \tag{1.32}$$

in which a_i $(i = 0, 1, \ldots, n)$ are constant coefficients.

The relation between the excitation and response can be displayed schematically by the block diagram of Fig. 1.2. Actually, Fig. 1.2 does not represent a genuine block diagram, because the flow is in the wrong direction. Indeed, in a genuine block diagram the flow must be in the opposite direction to that in Fig. 1.2, reflecting the idea that when a system is acted upon by a given input it exhibits a certain output. Consistent with this idea, Eq. (1.31) should be rewritten in the form

$$x(t) = D^{-1} f(t) \tag{1.33}$$

and the corresponding block diagram should be as in Fig. 1.3. The idea expressed by Eq. (1.33), or by the block diagram of Fig. 1.3, is that the response can be obtained by operating on the excitation with the operator D^{-1}, where D^{-1} can be interpreted as the inverse of the operator D. But, because D is a differential operator, D^{-1} must be an integral operator. Still, in spite of the fact that the direction of the flow is correct, Fig. 1.3 represents only a symbolic block diagram, because in a genuine block diagram the output is equal to the input multiplied by the quantity in the box and not merely to the input operated on by the quantity in the box. Moreover, whereas the idea embodied by Eq. (1.33) has esthetic appeal, it is not very helpful in producing a solution, because no method for generating D^{-1} explicitly exists. This should not cause undue pessimism, however, as there are various methods permitting one to achieve the same result in an implicit fashion. The various methods differ in details, depending on the type of excitation. In this section we consider an approach suitable for harmonic excitations, i.e., excitations in the form of sinusoidal functions of time. Approaches suitable for other types of excitations are discussed in subsequent sections.

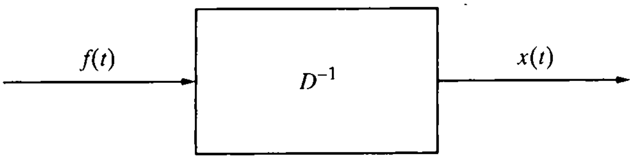


Figure 1.3 Symbolic block diagram relating the output to the input through the inverse operator D^{-1}

We consider the case in which the system described by Eq. (1.31) is subjected to a harmonic force, which can be expressed in one of the two forms

$$f(t) = f_0 \cos \omega t \tag{1.34a}$$

or

$$f(t) = f_0 \sin \omega t \tag{1.34b}$$

where f_0 is the *amplitude* and ω is the *excitation frequency*, or *driving frequency*. We note that the units of f are pounds (lb) or newtons (N) and those of ω are radians per second (rad/s).

The response of a system to harmonic excitations can be derived more expeditiously by using complex variables. This approach also permits treating the two forms of harmonic excitation simultaneously. To this end, we consider the complex unit vector given by

$$e^{i\omega t} = \cos \omega t + i \sin \omega t, \ i^2 = -1 \tag{1.35}$$

The vector is plotted in the complex plane shown in Fig. 1.4, and we note that the vector has unit magnitude and makes the angle ωt with the real axis. The projections of the unit vector on the real and imaginary axes are

$$\text{Re}\, e^{i\omega t} = \cos \omega t, \ \text{Im}\, e^{i\omega t} = \sin \omega t \tag{1.36}$$

As time unfolds, the angle ωt increases linearly, causing the vector to rotate counterclockwise in the complex plane with the angular velocity ω. In the process, the two projections vary harmonically with time, as can also be concluded from Eqs. (1.36).

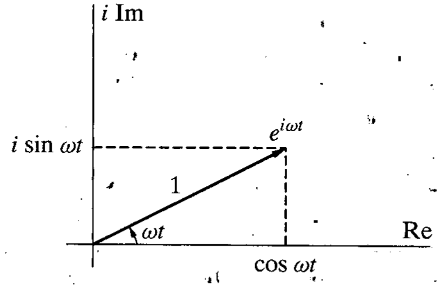


Figure 1.4 Unit vector $e^{i\omega t}$ rotating in the complex plane

In view of the above, we shall find it convenient to express the harmonic excitation in the complex form

$$f(t) = f_0 e^{i\omega t} = Aa_n e^{i\omega t} \tag{1.37}$$

where A is a constant having the same units as $x(t)$ and a_n is the coefficient of $x(t)$ in $Dx(t)$. Then, inserting Eq. (1.37) into Eq. (1.31), we have

$$Dx(t) = Aa_n e^{i\omega t} \tag{1.38}$$

with the understanding that, if the excitation is given by Eq. (1.34a), we must retain Re $x(t)$ as the response. Similarly, if the excitation is given by Eq. (1.34b), we must retain Im $x(t)$ as the response. Observing that

$$\frac{d^r}{dt^r} e^{i\omega t} = (i\omega)^r e^{i\omega t}, \qquad r = 1, 2, \ldots, n \tag{1.39}$$

it is not difficult to verify that the solution of Eq. (1.38) can be expressed as

$$x(t) = X(i\omega) e^{i\omega t} \tag{1.40}$$

Indeed, if we insert Eq. (1.40) into Eq. (1.38) and consider Eqs. (1.39), we can write

$$Dx(t) = X(i\omega) D e^{i\omega t} = X(i\omega) Z(i\omega) e^{i\omega t} = Aa_n e^{i\omega t} \tag{1.41}$$

where

$$Z(i\omega) = a_0 (i\omega)^n + a_1 (i\omega)^{n-1} + \ldots + a_n \tag{1.42}$$

is known as the system *impedance*. The concept is analogous to the generalized impedance $Z(s)$ introduced in Sec. 1.4, except that here s is replaced by $i\omega$. Equation (1.41) yields simply

$$X(i\omega) = \frac{Aa_n}{Z(i\omega)} \tag{1.43}$$

۱۱۲۱۲۱

Inserting Eq. (1.43) into Eq. (1.40) and considering Eq. (1.42), we can write the solution in the form

$$x(t) = AG(i\omega)e^{i\omega t} \tag{1.44}$$

where

$$G(i\omega) = \frac{a_n}{Z(i\omega)} = \frac{a_n}{a_0(i\omega)^n + a_1(i\omega)^{n-1} + \ldots + a_n} \tag{1.45}$$

is known as the system *admittance*, or the *frequency response*, a dimensionless quantity.

The input-output relation represented by Eq. (1.44) is displayed schematically in the block diagram of Fig. 1.5. In contrast with Fig. 1.3, however, Fig. 1.5 represents a genuine block diagram, as the output is obtained by merely multiplying the input (divided by a_n) by the frequency response, an algebraic operation. Like the operator D, the frequency response $G(i\omega)$ contains all the system dynamic characteristics. Unlike D, however, G represents an algebraic expression in the frequency domain rather than a differential operator in the time domain. Of course, the input-output relation in the frequency domain is valid only for harmonic excitations.

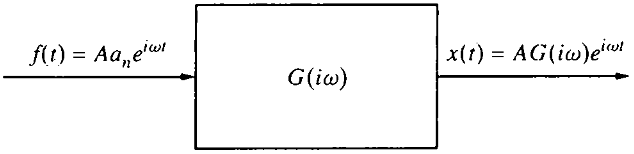

$f(t) = Aa_n e^{i\omega t}$
$G(i\omega)$
$x(t) = AG(i\omega)e^{i\omega t}$

Figure 1.5 Block diagram relating the output to the input through the frequency response function

Equation (1.44) is a complex expression, and we know that the response is a real quantity. To resolve this apparent conflict, we recall that we must retain either Re $x(t)$ or Im $x(t)$ from the solution given by Eq. (1.44), depending on whether the excitation is given by Eq. (1.34a) or by Eq. (1.34b) with f_0 replaced by Aa_n. This task is made easier by writing

$$G(i\omega) = |G(i\omega)|\, e^{-i\phi(\omega)} \tag{1.46}$$

where

$$|G(i\omega)| = \left\{[\text{Re}\, G(i\omega)]^2 + [\text{Im}\, G(i\omega)]^2\right\}^{1/2} = [G(i\omega)\overline{G}(i\omega)]^{1/2} \tag{1.47}$$

is the *magnitude* of the frequency response, in which $\overline{G}(i\omega)$ is the complex conjugate of $G(i\omega)$, and

$$\phi(\omega) = \tan^{-1}\frac{-\text{Im}\, G(i\omega)}{\text{Re}\, G(i\omega)} \tag{1.48}$$

is the *phase angle* of the frequency response. Inserting Eq. (1.46) into Eq. (1.44), we obtain

$$x(t) = A\,|G(i\omega)|\, e^{i(\omega t - \phi)} \tag{1.49}$$

Hence, retaining the real part in Eq. (1.49), we conclude that the response to the harmonic excitation given by $f(t) = Aa_n \cos \omega t$ is

$$x(t) = A\,|G(i\omega)| \cos(\omega t - \phi) \tag{1.50a}$$

Similarly, the response to $f(t) = Aa_n \sin \omega t$ is simply

$$x(t) = A\,|G(i\omega)| \sin(\omega t - \phi) \tag{1.50b}$$

It is possible, to plot the response $x(t)$ given by Eqs. (1.50a) and (1.50b) as a function of time, but this would not be very informative. Indeed, a great deal more information can be extracted from plots $|G(i\omega)|$ versus ω and $\phi(\omega)$ versus ω; they are known as *frequency response plots*. The concept of frequency response is very important in vibrations and control, and in Sec. 3.2 we will see how frequency response plots can be used to study the behavior of mechanical systems subjected to harmonic excitations.

Example 1.2

Obtain the steady-state response of the system described by the differential equation

$$\frac{dx(t)}{dt} + ax(t) = Aa \sin \omega t \tag{a}$$

where a and A are constants. Use the complex notation.

The desired steady-state response is given by Eq. (1.50b), or

$$x(t) = A\,|G(i\omega)| \sin(\omega t - \phi) \tag{b}$$

where $|G(i\omega)|$ is the magnitude and ϕ the phase angle of the frequency response. To derive an explicit expression for the frequency response for the case at hand, we first consider the impedance, Eq. (1.42), which for the system described by Eq. (a) reduces to

$$Z(i\omega) = i\omega + a \tag{c}$$

Then, using Eq. (1.45), we can write the frequency response

$$G(i\omega) = \frac{a}{Z(i\omega)} = \frac{a}{i\omega + a} = \frac{1}{1 + i\omega\tau} = \frac{1 - i\omega\tau}{1 + (\omega\tau)^2}, \; \tau = \frac{1}{a} \tag{d}$$

where τ is known as the *time constant* of the system. Moreover, using Eq. (1.47), the magnitude of the frequency response can be written as

$$\begin{aligned} |G(i\omega)| &= \{[\mathrm{Re}\, G(i\omega)]^2 + [\mathrm{Im}\, G(i\omega)]^2\}^{1/2} \\ &= [G(i\omega)\overline{G}(i\omega)]^{1/2} = \frac{1}{[1 + (\omega\tau)^2]^{1/2}} \end{aligned} \tag{e}$$

in which $\overline{G}(i\omega)$ is the complex conjugate of the frequency response. Moreover, using Eq. (1.48), we obtain the phase angle

$$\phi(\omega) = \tan^{-1}\frac{-\mathrm{Im}\, G(i\omega)}{\mathrm{Re}\, G(i\omega)} = \tan^{-1}\omega\tau \tag{f}$$

The generic system given by Eq. (a) is shown in Sec. 3.2 to describe the response of a simple mechanical system to harmonic excitations. In fact, Figs. 3.7 and 3.8 in Sec. 3.2 represent the frequency response plots $|G(i\omega)|$ versus $\omega\tau$ and $\phi(\omega)$ versus $\omega\tau$ corresponding to Eqs. (e) and (f), respectively.

1.6 RESPONSE TO ARBITRARY EXCITATIONS BY THE LAPLACE TRANSFORMATION

The problem of deriving the response of linear time-invariant systems to arbitrary excitations can be approached in several ways. In this section, we propose to use the approach most commonly used in linear system theory, namely, via the Laplace transformation method. As always, the general idea behind a solution by means of a transformation is to take a difficult problem, transform it into a relatively simple problem, solve the simple problem and finally carry out an inverse transformation to recover the solution to the original problem. In many cases, the inverse transformation can cause considerable difficulty, and this can be the case with the inverse Laplace transformation. More often than not, however, these difficulties can be avoided, which makes the Laplace transformation a preferred tool in linear system theory.

Let us consider once again the linear, time-invariant system described by Eq. (1.31), in which the operator D has the form given by Eq. (1.32). We confine ourselves to the case in which the input $f(t)$ is defined for $t > 0$ only. Moreover, in this text, we are concerned exclusively with systems for which the past affects the future, but not conversely. This implies that the system cannot react to inputs that have not yet been initiated, so that the output $x(t)$ must also be defined for $t > 0$ only. Such systems are said to be *causal* or *nonanticipatory*. For simplicity, we assume that the initial conditions are zero, so that the interest lies in the particular solution alone. In this regard, it should be recalled that, by virtue of the superposition principle, in the case of linear systems, the homogeneous solution and the particular solution can be derived separately and combined linearly to obtain the complete solution.

We propose to solve Eq. (1.31) by means of the Laplace transformation. To this end, we transform Eq. (1.31) into an algebraic equation, so that the problem of solving a differential equation for the actual response is replaced by the simpler problem of solving an algebraic equation for the transformed response. Finally, carrying out an inverse transformation of the transformed response, we obtain the actual response. It should be pointed out here that the Laplace transformation is perfectly capable of producing the response to the initial conditions at the same time (see Appendix A). Because this is not essential to the concepts to be introduced here, we postpone this task to a later time.

The unilateral Laplace transformation, or transform of the output is defined as

$$X(s) = \mathcal{L}x(t) = \int_0^\infty x(t)e^{-st}dt \tag{1.51}$$

where s is a complex variable. In general, the inverse Laplace transform involves the evaluation of the integral

$$x(t) = \mathcal{L}^{-1}X(s) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} X(s)e^{st}ds \tag{1.52}$$

where the path of integration is a line in the complex s-plane parallel to the imaginary axis, crossing the real axis at Re $s = \gamma$ and extending from $-\infty$ to $+\infty$. Under certain circumstances, the line integral can be replaced by a contour integral, which can be evaluated by means of the residue theorem (see Sec. A.5). In most cases of interest, however, evaluation of the inverse transform by such elaborate means is not really necessary, and the inversion can be carried out by resolving $X(s)$ into the sum of simpler functions and using tables of Laplace transforms (Appendix A). By analogy with Eq. (1.51), we can define the Laplace transform of the input as

$$F(s) = \mathcal{L}f(t) = \int_0^\infty f(t)e^{-st}dt \tag{1.53}$$

Before we proceed with the solution of Eq. (1.31) by means of the Laplace transformation method, we must address the question of transforming time derivatives. This question is addressed in Appendix A, from which we can write

$$\mathcal{L}\frac{d^r x(t)}{dt^r} = s^r X(s), \qquad r = 1, 2, \ldots, n \tag{1.54}$$

in which it is assumed that the initial conditions are zero. Carrying out the Laplace transformation on both sides of Eq. (1.31) and considering Eqs. (1.32), (1.51), (1.53) and (1.54), we obtain the algebraic expression

$$Z(s)X(s) = F(s) \tag{1.55}$$

where the polynomial

$$Z(s) = a_0 s^n + a_1 s^{n-1} + \ldots + a_n \tag{1.56}$$

represents the *generalized impedance*, encountered in Sec. 1.4 in connection with the response to initial excitations obtained by a classical approach. Equation (1.55) can be rewritten as

$$X(s) = G(s)F(s) \tag{1.57}$$

in which

$$G(s) = \frac{1}{Z(s)} = \frac{1}{a_0 s^n + a_1 s^{n-1} + \ldots + a_n} \tag{1.58}$$

is known as the *transfer function* of the system, a concept of fundamental importance in linear system theory. Contrasting Eq. (1.58) with Eq. (1.45), we conclude that the transfer function can be obtained from the frequency response $G(i\omega)$ by replacing $i\omega$ by s. Because s is a complex variable, while $i\omega$ is merely an imaginary variable, it is possible to interpret the transfer function as a generalization of the frequency response. This interpretation is quite appropriate in view of the fact that the frequency response applies to harmonic inputs only and the transfer function applies to arbitrary inputs.

Equation (1.57) represents an algebraic relation in the s-domain, which is also known as the Laplace domain. This relation can be displayed in the block diagram of Fig. 1.6. This diagram represents in the Laplace domain the same relation as the block diagram of Fig. 1.3 represents in the time domain. But, whereas Fig. 1.3 represents only a symbolic block diagram, Fig. 1.6 represents a genuine block diagram with practical implications. Indeed, the block diagram of Fig. 1.6 states that the transformed output $X(s)$ can be obtained by merely multiplying the transformed input $F(s)$ by the transfer function $G(s)$. This basic idea is used widely in linear system theory.

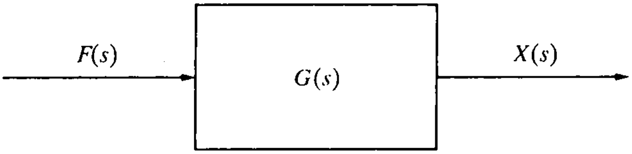


Figure 1.6 Block diagram relating the transformed output to the transformed input through the transfer function

Finally, inserting Eq. (1.57) into Eq. (1.52), we obtain a general expression for the response in the form of the inverse Laplace transform

$$x(t) = \mathcal{L}^{-1}X(s) = \mathcal{L}^{-1}G(s)F(s) \tag{1.59}$$

Before addressing the problem of deriving the general response of linear, time-invariant systems to arbitrary inputs by means of Eq. (1.59), in the next section we consider several inputs of significant importance in linear system theory in general and vibration theory in particular.

1.7 GENERALIZED FUNCTIONS. RESPONSE TO GENERALIZED FUNCTIONS

In the time-domain analysis of linear systems, and in particular in vibrations and controls, there is a family of functions of special significance. These functions are known as *generalized functions*,[1] or *singularity functions* and they are characterized by the fact that every generalized function and all its derivatives are continuous, except at a given value of the argument. Another characteristic of the generalized functions is that they can be obtained from one another through successive differentiation or integration. The generalized functions are based on a mathematical theory known as *distribution theory*.[2] The importance of the generalized functions can be attributed to the fact that a large variety of complicated excitations can be expressed as linear combinations of such functions. But for many linear time-invariant systems of interest, the response to a given generalized function is known. In fact, the response represents a system characteristic. Hence, invoking the superposition principle, the response of linear time-invariant systems to these linear combinations of generalized functions can be derived with relative ease.

The most important and widely used of the generalized functions is the *unit impulse*, or the *Dirac delta function*. The mathematical definition of the unit impulse is

$$\delta(t-a) = 0,\ t \neq a \tag{1.60a}$$

$$\int_{-\infty}^{\infty} \delta(t-a)dt = 1 \tag{1.60b}$$

[1] Lighthill, M. J., *Introduction to Fourier Analysis and Generalized Functions*, Cambridge University Press, New York, 1958.

[2] Zemanian, A. H., *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications*, McGraw-Hill, New York, 1965.

The unit impulse is plotted in Fig. 1.7 as a function of time. It can be regarded as a thin rectangle of width ϵ and height $1/\epsilon$ in the neighborhood of $t = a$, where ϵ is a small time increment. In the limit, as the increment ϵ, and hence the width of the rectangle, approaches zero, the height approaches infinity in a way that the area under the curve remains constant and equal to unity, which explains the term "unit" impulse. The units of the unit impulse are seconds^{-1}(s^{-1}).

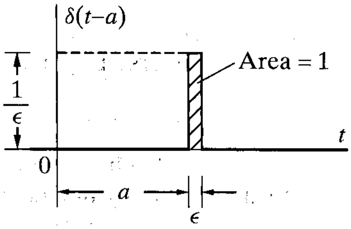


Figure 1.7 Unit impulse initiated at $t = a$

The *impulse response*, denoted by $g(t)$, is defined as the response of a linear time-invariant system to a unit impulse applied at $t = 0$, with the initial conditions being equal to zero. In this case, the input is $f(t) = \delta(t)$ and the output is $x(t) = g(t)$. Using Eqs. (1.53) and (1.60), we obtain the Laplace transform of the unit impulse in the form

$$\Delta(s) = \mathcal{L}\delta(t) = \int_0^\infty \delta(t)e^{-st}dt = e^{-st}\Big|_{t=0}\int_0^\infty \delta(t)dt = 1 \tag{1.61}$$

Then, inserting Eq. (1.61) into Eq. (1.59), we obtain

$$g(t) = \mathcal{L}^{-1}G(s)\Delta(s) = \mathcal{L}^{-1}G(s) \tag{1.62}$$

from which we conclude that *the impulse response is equal to the inverse Laplace transform of the transfer function*, or *the impulse response and the transfer function form a Laplace transform pair.*

Another very important generalized function in vibrations and controls is the *unit step function*, defined mathematically as

$$\omega(t - a) = 0, \qquad t < a \tag{1.63a}$$

$$\omega(t - a) = 1, \qquad t > a \tag{1.63b}$$

The unit step function is displayed in Fig. 1.8. Clearly, the function is defined for all times, except at $t = a$, where it experiences a discontinuity. The unit step function is dimensionless. It is easy to verify that the unit step function can be obtained by integrating the unit impulse, or

$$\omega(t - a) = \int_0^t \delta(\tau - a)d\tau \tag{1.64}$$

Conversely, the unit impulse can be obtained by differentiating the unit step function, or

$$\delta(t - a) = \frac{d\mathcal{u}(t - a)}{dt} \tag{1.65}$$

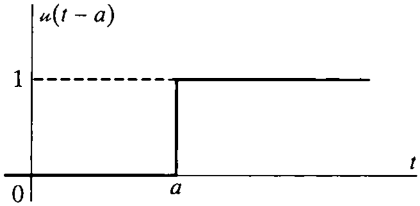


Figure 1.8 Unit step function initiated at $t = a$

The *step response*, denoted by $\mathcal{s}(t)$, is defined as the response of a linear time-invariant system to a unit step function applied at $t = 0$, with the initial conditions being equal to zero. Letting $f(t) = \mathcal{u}(t)$ in Eq. (1.53) and considering Eq. (1.63b), we obtain the Laplace transform of the unit step function in the form

$$\mathcal{U}(s) = \mathcal{L}\mathcal{u}(t) = \int_0^\infty \mathcal{u}(t)e^{-st}dt = \int_0^\infty e^{-st}dt = \left.\frac{e^{-st}}{-s}\right|_0^\infty = \frac{1}{s} \tag{1.66}$$

Then, letting $x(t) = \mathcal{s}(t)$ in Eq. (1.59) and considering Eq. (1.66), we can write

$$\mathcal{s}(t) = \mathcal{L}^{-1}G(s)\mathcal{U}(s) = \mathcal{L}^{-1}\frac{G(s)}{s} \tag{1.67}$$

or the step response is equal to the inverse Laplace transform of the transfer function divided by s.

Yet another generalized function of interest in vibrations and control is the *unit ramp function*, described by

$$r(t - a) = 0, \; t < a \tag{1.68a}$$

$$r(t - a) = t - a, \; t > a \tag{1.68b}$$

The function is shown in Fig. 1.9. At $t = a$, the slope of $r(t - a)$ experiences a discontinuity. The function has units of seconds (s). The unit ramp function can be obtained by integrating the unit step function, i.e.,

$$r(t - a) = \int_0^t \mathcal{u}(\tau - a)d\tau \tag{1.69}$$

Consistent with this, the unit step function is equal to the time derivative of the unit ramp function, or

$$\mathcal{u}(t - a) = \frac{dr(t - a)}{dt} \tag{1.70}$$

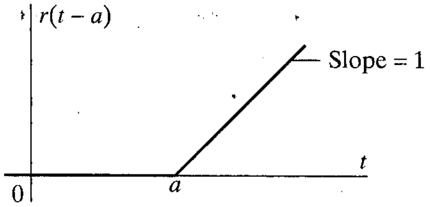


Figure 1.9 Unit ramp function initiated at $t = a$

The *ramp response* $\mathscr{r}(t)$ is defined as the response of a linear time-invariant system to a unit ramp function applied at $t = 0$, with zero initial conditions. Integrating by parts, we obtain the Laplace transform of the unit ramp function as follows:

$$\begin{aligned} R(s) &= \mathcal{L}r(t) = \int_0^\infty r(t)e^{-st}dt = \int_0^\infty te^{-st}dt \\ &= t\frac{e^{-st}}{-s}\bigg|_0^\infty + \frac{1}{s}\int_0^\infty e^{-st}dt = \frac{1}{s^2} \end{aligned} \tag{1.71}$$

so that, using Eq. (1.59), we can write the ramp response in the form

$$\mathscr{r}(t) = \mathcal{L}^{-1}G(s)R(s) = \mathcal{L}^{-1}\frac{G(s)}{s^2} \tag{1.72}$$

or the ramp response is equal to the inverse Laplace transform of the transfer function divided by s^2.

As indicated in the beginning of this section, and verified by Eqs. (1.64) and (1.69), generalized functions can be obtained from one another through successive integration. This property carries over to the response to generalized functions as well. To prove this statement, we write the relation between the impulse response and the unit impulse for a linear time-invariant system in the operator form

$$Dg(t) = \delta(t) \tag{1.73}$$

where D is a linear homogeneous differential operator with constant coefficients of the type given by Eq. (1.32). Next, we integrate Eq. (1.73) with respect to time, assume that the integration and differentiation processes are interchangeable, consider Eq. (1.64) and write

$$\int_0^t Dg(\tau)d\tau = D\int_0^t g(\tau)d\tau = \int_0^t \delta(\tau)d\tau = \mathscr{u}(t) \tag{1.74}$$

In view of the fact that the relation between the step response and the unit step function has the operator form

$$D\mathscr{s}(t) = \mathscr{u}(t) \tag{1.75}$$

we conclude from Eq. (1.74) that

$$\mathscr{s}(t) = \int_0^t g(\tau)d\tau \tag{1.76}$$

or the step response is equal to the integral of the impulse response. Integrating Eq. (1.75), repeating the preceding process and considering Eq. (1.69), we obtain

$$\int_0^t D\mathscr{s}(\tau)d\tau = D\int_0^t \mathscr{s}(\tau)d\tau = \int_0^t \mathscr{u}(\tau)d\tau = r(t) \tag{1.77}$$

with the obvious conclusion that

$$\mathscr{r}(t) = \int_0^t \mathscr{s}(\tau)d\tau \tag{1.78}$$

or the ramp response is equal to the integral of the step response. Equations (1.76) and (1.78) can provide at times an expedient way of deriving the response to some important generalized functions.

The unit step function can be used to express certain functions in a compact manner. The unit ramp function given by Eqs. (1.68) is a case in point. Indeed, because the ramp function begins at $t = a$, and is identically zero for $t < a$, it is necessary to describe the function by means of two expressions, one for $t < a$ and one for $t > a$. But, recalling the definition of the unit step function $\mathscr{u}(t - a)$, we conclude that the effect of multiplying an arbitrary function $f(t)$ by $\mathscr{u}(t - a)$ is to annihilate the portion of $f(t)$ corresponding to $t < a$ and leave unaffected the portion for which $t > a$. In view of this, the unit ramp function can be expressed in the compact form

$$r(t - a) = (t - a)\mathscr{u}(t - a) \tag{1.79}$$

which is valid for all times.

As pointed out earlier in this section, the response of linear time-invariant systems to some complicated excitations can be obtained with relative ease by working with generalized functions. As an example, we consider the trapezoidal pulse shown in Fig. 1.10. The function can be described in terms of four expressions, one for $t < 0$, one for $0 < t < t_1$, one for $t_1 < t < t_2$ and one for $t > t_2$. It can also be described by the linear combination

$$f(t) = \frac{f_0}{t_1}\left[t\mathscr{u}(t) - (t - t_1)\mathscr{u}(t - t_1) - t_1\mathscr{u}(t - t_2)\right] \tag{1.80}$$

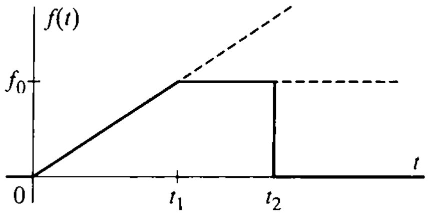


Figure 1.10 A trapezoidal pulse

which is significantly more convenient for deriving the response than the more traditional description mentioned above. Indeed, in view of the various definitions introduced earlier, we can invoke the superposition principle and the time-invariance property and write the response to $f(t)$ in the general form

$$x(t) = \frac{f_0}{t_1}\left[\mathscr{r}(t) - \mathscr{r}(t - t_1) - t_1\mathscr{s}(t - t_2)\right] \tag{1.81}$$

and we should point out that, unlike Eq. (1.80), it is not necessary to multiply the various responses in Eq. (1.81) by unit step functions, because this effect is already included in $\mathscr{r}(t), \mathscr{r}(t - t_1)$ and $\mathscr{s}(t - t_2)$.

Example 1.3

Derive the impulse response of the second-order system described by

$$\frac{d^2x(t)}{dt^2} + a\frac{dx(t)}{dt} + bx(t) = f(t) \tag{a}$$

From Eq. (1.62), the impulse response is given by the inverse Laplace transform

$$g(t) = \mathcal{L}^{-1}G(s) \tag{b}$$

where $G(s)$ is the transfer function. Using Eq. (1.58), the transfer function for the system at hand is

$$G(s) = \frac{1}{s^2 + as + b} \tag{c}$$

To carry out the inversion indicated by Eq. (b), it is advisable to decompose $G(s)$ into partial fractions as follows:

$$G(s) = \frac{1}{(s - s_1)(s - s_2)} = \frac{A}{s - s_1} + \frac{B}{s - s_2} \tag{d}$$

where

$$\begin{matrix} s_1 \\ s_2 \end{matrix} = -\frac{a}{2} \pm \frac{1}{2}\sqrt{a^2 - 4b} \tag{e}$$

are the roots of the characteristic equation and A and B are coefficients yet to be determined. Bringing the right side of Eq. (d) to a common denominator, we obtain

$$G(s) = \frac{A(s - s_2) + B(s - s_1)}{(s - s_1)(s - s_2)} \tag{f}$$

so that, comparing Eqs. (d) and (f), we conclude that A and B must solve the two equations

$$A + B = 0, \qquad -As_2 - Bs_1 = 1 \tag{g}$$

which yield

$$A = -B = \frac{1}{s_1 - s_2} \tag{h}$$

Inserting Eq. (h) into Eq. (d), we obtain the transfer function

$$G(s) = \frac{1}{s_1 - s_2}\left(\frac{1}{s - s_1} - \frac{1}{s - s_2}\right) \tag{i}$$

so that, introducing Eq. (i) into Eq. (b), we can rewrite the impulse response as

$$g(t) = \mathcal{L}^{-1}\frac{1}{s_1 - s_2}\left(\frac{1}{s - s_1} - \frac{1}{s - s_2}\right) \tag{j}$$

To carry out the inverse transformation indicated by Eq. (j), we turn to the table of Laplace transforms of Appendix A, where we find

$$\mathcal{L}^{-1}\frac{1}{s-\omega} = e^{\omega t} \tag{k}$$

so that, considering Eqs. (e), we obtain the desired impulse response

$$\begin{aligned} g(t) &= \frac{1}{s_1 - s_2}\left(e^{s_1 t} - e^{s_2 t}\right)\mathscr{u}(t) \\ &= \frac{e^{-at/2}}{\sqrt{a^2-4b}}\left(e^{\frac{1}{2}\sqrt{a^2-4b}\,t} - e^{-\frac{1}{2}\sqrt{a^2-4b}\,t}\right)\mathscr{u}(t) \\ &= \frac{2e^{-at/2}}{\sqrt{a^2-4b}}\sinh\frac{1}{2}\sqrt{a^2-4b}\,t\,\mathscr{u}(t) \end{aligned} \tag{l}$$

where we multiplied the result by $\mathscr{u}(t)$ in recognition of the fact that the impulse response is zero for $t < 0$.

At this point, we observe that the impulse response given by Eq. (l) is identical to the response to the initial velocity $v_0 = 1$ obtained in Example 1.1. This is no coincidence, and in Sec. 3.5 we discuss the relation between impulsive excitations and initial conditions.

Example 1.4

Derive the response of the first-order system

$$\frac{dx(t)}{dt} + ax(t) = f(t) \tag{a}$$

for the case in which $f(t)$ has the form of the trapezoidal pulse shown in Fig. 1.10.

Equation (1.80) expresses the function $f(t)$ as a linear combination of two ramp functions and one step function. Consistent with this, Eq. (1.81) expresses the response $x(t)$ as a corresponding linear combination of two ramp responses and one step response. Both types of response require the transfer function, as can be concluded from Eqs. (1.67) and (1.72). From Eq. (1.58), the transfer function for the system described by Eq. (a) is simply

$$G(s) = \frac{1}{s+a} \tag{b}$$

Hence, using Eq. (1.67), the step response is

$$\mathscr{s}(t) = \mathcal{L}^{-1}\frac{G(s)}{s} = \mathcal{L}^{-1}\frac{1}{s(s+a)} \tag{c}$$

As in Example 1.3, we decompose the function on the right side of Eq. (c) into partial fractions and carry out the inverse transformation to obtain the step response

$$\mathscr{s}(t) = \mathcal{L}^{-1}\frac{1}{s(s+a)} = \mathcal{L}^{-1}\frac{1}{a}\left(\frac{1}{s} - \frac{1}{s+a}\right) = \frac{1}{a}\left(1 - e^{-at}\right)\mathscr{u}(t) \tag{d}$$

Moreover, using Eq. (1.72), the ramp response is

$$\mathscr{r}(t) = \mathcal{L}^{-1}\frac{G(s)}{s^2} = \mathcal{L}^{-1}\frac{1}{s^2(s+a)} \tag{e}$$

Because the characteristic equation has a double root at $s = 0$, the partial fractions expansion has the form

$$\frac{1}{s^2(s+a)} = \frac{A}{s^2} + \frac{B}{s} + \frac{C}{s+a} \tag{f}$$

Bringing the right side of Eq. (f) to a common denominator and comparing with the left side, we conclude that

$$A = \frac{1}{a},\ B = -\frac{A}{a} = -\frac{1}{a^2},\ C = -B = \frac{1}{a^2} \tag{g}$$

Then, introducing Eqs. (f) and (g) into Eq. (e) and referring to the table of Laplace transforms in Appendix A, we obtain the ramp response

$$\begin{aligned} \mathscr{r}(t) &= \mathcal{L}^{-1}\frac{1}{s^2(s+a)} \\ &= \mathcal{L}^{-1}\frac{1}{a^2}\left(\frac{a}{s^2} - \frac{1}{s} + \frac{1}{s+a}\right) \\ &= \frac{1}{a^2}\left(at - 1 + e^{-at}\right)\mathscr{u}(t) \end{aligned} \tag{h}$$

Finally, inserting Eqs. (d) and (h) into Eq. (1.81), the response to the trapezoidal pulse given by Eq. (1.80) can be written in the explicit form

$$\begin{aligned} x(t) = \frac{f_0}{t_1}\Bigg\{ &\frac{1}{a^2}(at - 1 + e^{-at})\mathscr{u}(t) - \frac{1}{a^2}\left[a(t-t_1) - 1 + e^{-a(t-t_1)}\right]\mathscr{u}(t-t_1) \\ &-\frac{t_1}{a}\left(1 - e^{-a(t-t_2)}\right)\mathscr{u}(t-t_2)\Bigg\} \end{aligned} \tag{i}$$

It should be pointed out here that the ramp response, Eq. (h), could have been obtained more expeditiously by integrating the step response, Eq. (d), with respect to time.

1.8 RESPONSE TO ARBITRARY EXCITATIONS BY THE CONVOLUTION INTEGRAL

Invoking the superposition principle and the time-invariance property, we were able in Sec. 1.7 to derive the response of a linear time-invariant system to some complicated excitation with relative ease by decomposing the excitation and response into linear combinations of simple excitations and responses, respectively. But the power of this approach extends well beyond such examples. Indeed, the superposition principle and the time-invariance property can be used to derive a general formula for the response of linear systems to any arbitrary excitation.

We consider an arbitrary excitation $f(t)$, such as the one depicted in Fig. 1.11. Without loss of generality, we assume that $f(t)$ is defined for $t > 0$ only. Focusing our attention on an incremental area, identified as the shaded area in Fig. 1.11, and assuming that the width $\Delta\tau$ is relatively small, the area can be regarded as an increment of force representing an impulse of magnitude $f(\tau)\Delta\tau$ acting at $t = \tau$, or

$$\Delta f(t,\tau) = f(\tau)\Delta\tau\delta(t-\tau) \tag{1.82}$$

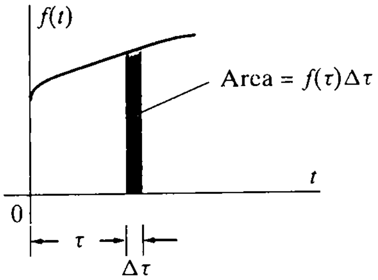


Figure 1.11 An arbitrary excitation function

In view of this, the function $f(t)$ can be expressed as a linear combination of impulses of the form

$$f(t) = \sum \Delta f(t, \tau) = \sum f(\tau)\Delta\tau\delta(t - \tau) \tag{1.83}$$

But, invoking the homogeneity and time-invariance properties (Sec. 1.2), the response of a linear time-invariant system to an impulse of a given magnitude and acting at $t = \tau$ is an impulse response of the same magnitude and shifted by the amount $t = \tau$. Hence, the increment in the system response corresponding to the excitation increment given by Eq. (1.82) is simply

$$\Delta x(t, \tau) = f(\tau)\Delta\tau g(t - \tau) \tag{1.84}$$

where $g(t)$ is the impulse response defined in Sec. 1.7. Then, invoking the superposition principle, the response of a linear system to the linear combination of impulses given by Eq. (1.83) is a linear combination of impulse responses of the form

$$x(t) = \sum f(\tau)\Delta\tau g(t - \tau) \tag{1.85}$$

In the limit, as $\Delta\tau \to 0$, the response of a linear system to an arbitrary excitation can be expressed as

$$x(t) = \int_0^\infty f(\tau)g(t - \tau)d\tau \tag{1.86}$$

But, $g(t - \tau)$ is equal to zero for $t - \tau < 0$, which is the same as $\tau > t$. Hence, the upper limit of the integral can be replaced by t, or

$$x(t) = \int_0^t f(\tau)g(t - \tau)d\tau \tag{1.87}$$

The right side of Eq. (1.87) is known as the *convolution integral* or, quite appropriately, the *superposition integral.*

Next, we consider the change of variables

$$\begin{gathered} t - \tau = \sigma,\ \tau = t - \sigma,\ d\tau = -d\sigma \\ \tau = 0 \to \sigma = t,\ \tau = t \to \sigma = 0 \end{gathered} \tag{1.88}$$

Inserting Eqs. (1.88) into Eq. (1.87), we obtain

$$x(t) = \int_t^0 f(t - \sigma)g(\sigma)(-d\sigma) = \int_0^t f(t - \sigma)g(\sigma)d\sigma \tag{1.89}$$

Then, recognizing that σ is a mere dummy variable of integration, and replacing σ by τ, we can combine Eqs. (1.87) a nd (1.89) into

$$x(t) = \int_0^t f(\tau)g(t - \tau)d\tau = \int_0^t f(t - \tau)g(\tau)d\tau \tag{1.90}$$

so that the convolution integral is symmetric in f and g, with the implication that it does not matter whether we shift the excitation or the impulse response. Clearly, it is more advantageous to shift the simpler of the two functions.

The above derivation of the convolution integral was based on physical considerations, and it relates the response to the excitation and the dynamical system, the latter being represented by the impulse response. In Appendix A, we derive the convolution integral more abstractly as the inverse Laplace transformation of a product of two arbitrary Laplace transforms, not necessarily those of the excitation and of the impulse response.

The convolution integral can be given an interesting and sometimes useful geometric interpretation. To this end, we propose to examine the various operations involved in the convolution integral. For the purpose of this discussion, we shift the impulse response rather than the excitation, so that we work with the convolution integral in the form given by Eq. (1.87). We begin by considering a typical excitation $f(\tau)$ and a typical impulse response $g(\tau)$, shown in Figs. 1.12a and 1.12b, respectively. Shifting $g(\tau)$ backward by the amount $\tau = t$ results in the function $g(\tau + t)$ shown in Fig. 1.12c. Taking the mirror image of the function $g(\tau + t)$ about the vertical axis, which amounts to replacing τ by $-\tau$, yields the function $g(t - \tau)$ depicted in Fig. 1.12d. The next operation is the multiplication of $f(\tau)$ by $g(t - \tau)$, resulting in the function $f(\tau)g(t - \tau)$ shown in Fig. 1.12e. Then the integral of the product $f(\tau)g(t-\tau)$ over τ from 0 to t, which is equal to the area under the curve, represents the response $x(t)$ corresponding a given value t of time, as shown in Fig. 1.12f. The complete response is obtained by letting t vary from zero to any desired value.

If the excitation $f(t)$ is a smooth function of time, the preceding geometric interpretation is primarily of academic interest. However, if the excitation $f(t)$ is only sectionally smooth, such as the function depicted in Fig. 1.10, then the geometric interpretation is essential to the successful evaluation of the convolution integral. We will verify this statement in Example 1.6.

Example 1.5

Derive the general response of the first-order system

$$\frac{dx(t)}{dt} + ax(t) = f(t) \tag{a}$$

by means of the convolution integral. Then use this expression to obtain the ramp response.

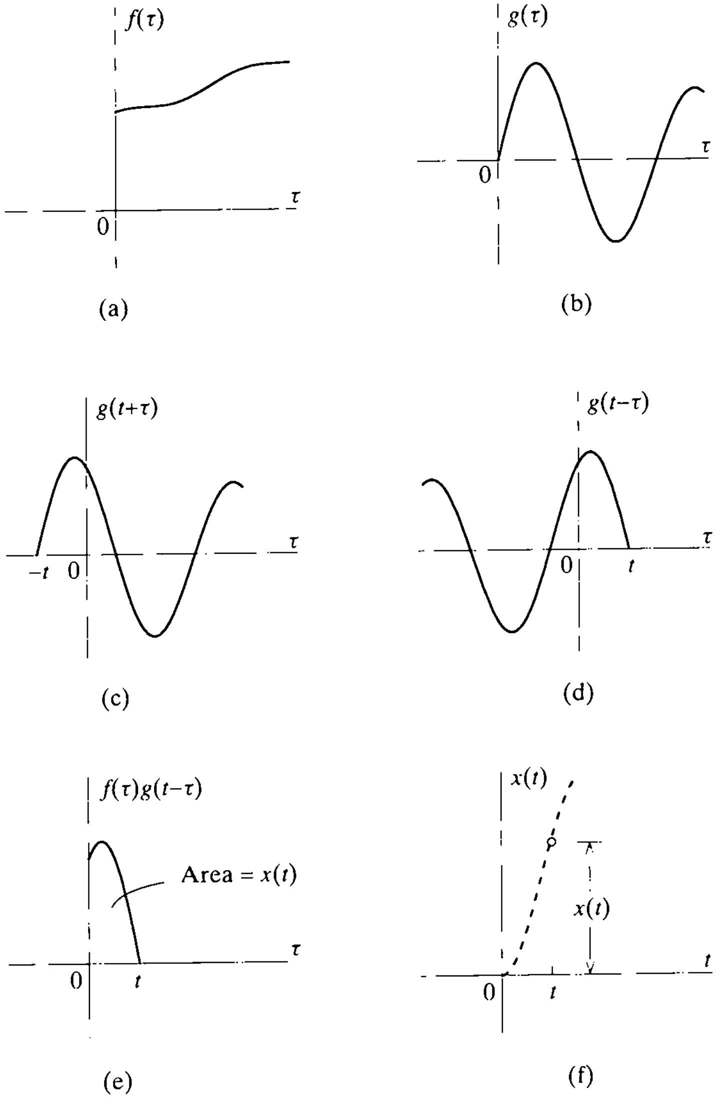


Figure 1.12 Geometric interpretation of the convolution integral **(a)** Excitation function **(b)** Impulse response **(c)** Impulse response shifted backward **(d)** Mirror image of the impulse response shifted backward **(e)** Multiplication of the excitation by the mirror image of the shifted impulse response **(f)** The response at time t resulting from the convolution integral

From Example 1.4, the system transfer function is

$$G(s) = \frac{1}{s+a} \tag{b}$$

so that, according to Eq. (1.62), the impulse response is given by

$$g(t) = \mathcal{L}^{-1}G(s) = \mathcal{L}^{-1}\frac{1}{s+a} = e^{-at}u(t) \tag{c}$$

Hence, using the first form of the convolution integral in Eq. (1.90), we obtain the general response

$$x(t) = \int_0^t f(\tau)g(t-\tau)d\tau = \int_0^t f(\tau)e^{-a(t-\tau)}u(t-\tau)d\tau$$

$$= \int_0^t f(\tau)e^{-a(t-\tau)}d\tau \tag{d}$$

The unit ramp function has the expression

$$f(t) = r(t) = tu(t) \tag{e}$$

so that, inserting Eq. (e) into Eq. (d), we obtain the desired ramp response

$$\mathscr{r}(t) = \int_0^t r(\tau)g(t-\tau)d\tau = \int_0^t \tau u(\tau)e^{-a(t-\tau)}d\tau$$

$$= e^{-at}\int_0^t \tau e^{a\tau}d\tau = e^{-at}\frac{1}{a^2}\left[1-(1-at)\,e^{at}\right]u(t)$$

$$= \frac{1}{a^2}\left(at-1+e^{-at}\right)u(t) \tag{f}$$

where we multiplied the result by $u(t)$ in recognition of the fact that the ramp response is zero for $t < 0$. It is easy to verify that the result obtained here is identical to that obtained in Example 1.4.

Example 1.6

In Example 1.4, we derived the response of the first-order system

$$\frac{dx(t)}{dt} + ax(t) = f(t) \tag{a}$$

for the case in which $f(t)$ has the form of the trapezoidal pulse shown in Fig. 1.10, by regarding the pulse as a combination of two ramp functions and one step function. In this example, we propose to solve the same problem by means of the convolution integral as given by Eq. (1.87).

From Example 1.5, we recall that the impulse response of the first-order system described by Eq. (a) is given by

$$g(t) = e^{-at}u(t) \tag{b}$$

The excitation, the impulse response and the impulse response shifted and folded are shown in Figs. 1.13a, 1.13b and 1.13c, respectively. To evaluate the convolution integral, we express the excitation as follows:

$$f(\tau) = \begin{cases} \dfrac{f_0}{t_1}\tau, & 0 < \tau < t_1 \\ f_0, & t_1 < \tau < t_2 \\ 0, & \tau > t_2 \end{cases} \tag{c}$$

Hence, the function $f(\tau)$ entering into the convolution integral and the limits of integration depend on the amount of shift t in the impulse response. Consistent with this, we determine three expressions for the response corresponding to $0 < t < t_1$, $t_1 < t < t_2$ and $t > t_2$.

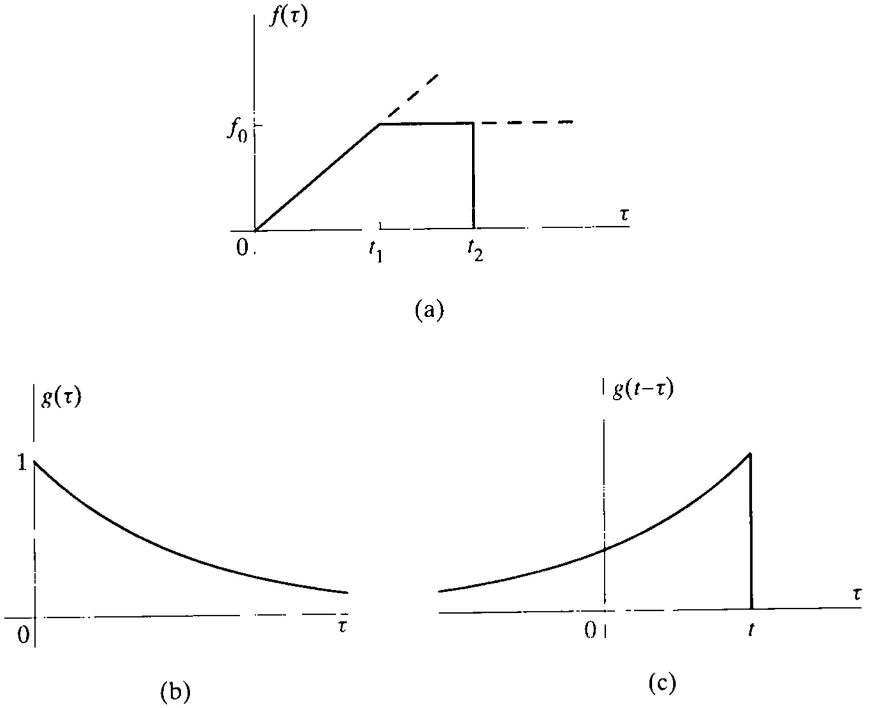


Figure 1.13 The convolution process **(a)** Excitation in the form of a trapezoidal pulse **(b)** Impulse response for a first-order system **(c)** Mirror image of the impulse response shifted backward

Inserting Eqs. (b) and (c) into Eq. (1.87) and considering Figs. 1.13a and 1.13c, we obtain for $0 < t < t_1$

$$x(t) = \frac{f_0}{t_1}\int_0^t \tau e^{-a(t-\tau)}d\tau = \frac{f_0}{t_1}\left[\tau\frac{e^{-a(t-\tau)}}{a}\bigg|_0^t - \frac{1}{a}\int_0^t e^{-a(t-\tau)}d\tau\right]$$

$$= \frac{f_0}{t_1 a^2}\left(at - 1 + e^{-at}\right) \qquad (d)$$

For $t_1 < t < t_2$, we observe that the integration extends over time intervals in which $f(\tau)$ has different expressions. Hence, for $t_1 < t < t_2$, we obtain the response

$$x(t) = \frac{f_0}{t_1}\int_0^{t_1} \tau e^{-a(t-\tau)}d\tau + f_0\int_{t_1}^t e^{-a(t-\tau)}d\tau$$

$$= \frac{f_0}{t_1}\left[\tau\frac{e^{-a(t-\tau)}}{a} - \frac{e^{-a(t-\tau)}}{a^2}\right]\Bigg|_0^{t_1} + f_0\frac{e^{-a(t-\tau)}}{a}\bigg|_{t_1}^t$$

$$= \frac{f_0}{t_1 a^2}\left[at_1 + e^{-at} - e^{-a(t-t_1)}\right] \qquad (e)$$

Finally, for $t > t_2$, the response is

$$\begin{aligned} x(t) &= \frac{f_0}{t_1}\int_0^{t_1} \tau e^{-a(t-\tau)}d\tau + f_0\int_{t_1}^{t_2} e^{-a(t-\tau)}d\tau \\ &= \frac{f_0}{t_1}\left[\tau\frac{e^{-a(t-\tau)}}{a} - \frac{e^{-a(t-\tau)}}{a^2}\right]\Bigg|_0^{t_1} + f_0\frac{e^{-a(t-\tau)}}{a}\Bigg|_{t_1}^{t_2} \\ &= \frac{f_0}{t_1a^2}\left[e^{-at} - e^{-a(t-t_1)} + at_1e^{-a(t-t_2)}\right] \qquad \text{(f)} \end{aligned}$$

It is not difficult to verify that Eqs. (d), (e) and (f) are equivalent to Eq. (i) obtained in Example 1.4.

This example makes it clear that extreme care must be exercised in using the convolution integral to determine the response to discontinuous excitations. To this end, the geometric interpretation of the operations involved in the convolution integral, and in particular Fig. 1.13c, proves indispensable. Clearly, as the time t increases, the folded impulse response moves to the right, which makes the task of choosing the proper expressions for $f(\tau)$ and the proper integration limits relatively easy.

1.9 STATE EQUATIONS. RESPONSE BY THE TRANSITION MATRIX

In our previous discussions, we were concerned with systems for which the excitation and response were characterized by a single function of time each. Such systems are commonly known as *single-input, single-output systems.* In linear system theory, there is considerable interest in systems with several inputs and several outputs, referred to as *multi-input, multi-output systems.* In particular, we wish to consider a linear time-invariant system described by the set of simultaneous first-order differential equations

$$\dot{x}_i(t) = \sum_{j=1}^{n} a_{ij}x_j(t) + \sum_{j=1}^{r} b_{ij}f_j(t), \quad i = 1, 2, \ldots, n \qquad (1.91)$$

where we used the customary overdot to denote a derivative with respect to time, $\dot{x}_i = dx_i/dt$ $(i = 1, 2, \ldots, n)$. Equations (1.91) are known as *state equations* and can be written in the compact form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{f}(t) \qquad (1.92)$$

in which

$$\mathbf{x}(t) = [x_1(t)\; x_2(t) \ldots x_n(t)]^T, \qquad \mathbf{f}(t) = [f_1(t)\; f_2(t) \ldots f_r(t)]^T \qquad (1.93a,b)$$

represent an n-dimensional *state vector* and an r-dimensional *excitation vector*, respectively, where the symbol T denotes a transposed quantity (see Appendix B); the components $x_i(t)$ of $\mathbf{x}(t)$ are called *state variables*. Moreover,

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}, \qquad B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1r} \\ b_{21} & b_{22} & \ldots & b_{2r} \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & \ldots & b_{nr} \end{bmatrix} \qquad (1.94a,b)$$

are $n \times n$ and $n \times r$ *coefficient matrices*, respectively. Equation (1.92) defines an nth-order system, and it represents an extension and generalization of the first-order system described by Eq. (a) of Example 1.4. We observe, however, that for the analogy to be more complete the term $ax(t)$ in the first-order system given by Eq. (a) of Example 1.4 should have had a minus sign.

Our interest lies in deriving a general formula for the response of the nth-order system described by Eq. (1.92). Unlike earlier derivations, here we propose to derive the homogeneous and particular solutions at the same time. To this end, we first multiply both sides of Eq. (1.92) on the left by the yet to be determined matrix $K(t)$, so that

$$K(t)\dot{\mathbf{x}}(t) = K(t)A\mathbf{x}(t) + K(t)B\mathbf{f}(t) \tag{1.95}$$

Then we consider

$$\frac{d}{dt}[K(t)\mathbf{x}(t)] = \dot{K}(t)\mathbf{x}(t) + K(t)\dot{\mathbf{x}}(t) \tag{1.96}$$

Inserting Eq. (1.96) into Eq. (1.95), we obtain

$$\frac{d}{dt}[K(t)\mathbf{x}(t)] - \dot{K}(t)\mathbf{x}(t) = K(t)A\mathbf{x}(t) + K(t)B\mathbf{f}(t) \tag{1.97}$$

Next, we require that the matrix $K(t)$ satisfy

$$\dot{K}(t) = -AK(t) \tag{1.98}$$

which has the solution

$$K(t) = e^{-At}K(0) \tag{1.99}$$

where

$$e^{-At} = I - tA + \frac{t^2}{2!}A^2 - \frac{t^3}{3!}A^3 + \dots \tag{1.100}$$

represents a matrix series, in which I is the identity matrix, and $K(0)$ is the initial value of $K(t)$. For convenience, we choose

$$K(0) = I \tag{1.101}$$

so that

$$K(t) = e^{-At} \tag{1.102}$$

It is easy to verify that the matrices $K(t)$ and A commute, or

$$AK(t) = K(t)A \tag{1.103}$$

so that Eq. (1.98) can also be written as

$$\dot{K}(t) = -K(t)A \tag{1.104}$$

In view of Eq. (1.104), Eq. (1.97) reduces to

$$\frac{d}{dt}[K(t)\mathbf{x}(t)] = K(t)B\mathbf{f}(t) \tag{1.105}$$

which can be integrated readily. Taking into consideration Eq. (1.101), the result is

$$\begin{aligned} K(t)\mathbf{x}(t) &= K(0)\mathbf{x}(0) + \int_0^t K(\tau)B\mathbf{f}(\tau)d\tau \\ &= \mathbf{x}(0) + \int_0^t K(\tau)B\mathbf{f}(\tau)d\tau \end{aligned} \tag{1.106}$$

Premultiplying Eq. (1.106) through by $K^{-1}(t)$ and considering Eq. (1.102), we obtain the desired response

$$\begin{aligned} \mathbf{x}(t) &= K^{-1}(t)\mathbf{x}(0) + K^{-1}(t)\int_0^t K(\tau)B\mathbf{f}(\tau)d\tau \\ &= \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t-\tau)B\mathbf{f}(\tau)d\tau \end{aligned} \tag{1.107}$$

where $\mathbf{x}(0)$ is the initial state. Moreover, the matrix

$$\Phi(t-\tau) = e^{A(t-\tau)} = I + (t-\tau)A + \frac{(t-\tau)^2}{2!}A^2 + \frac{(t-\tau)^3}{3!}A^3 + \dots \tag{1.108}$$

is known as the *transition matrix*. The series converges always, but the number of terms in the series required for a given accuracy depends on the time interval $t-\tau$ and the eigenvalue of A of largest modulus. At times, the transition matrix can be derived more conveniently by means of an approach based on the Laplace transformation. To this end, we consider the homogeneous part of the Eq. (1.92), namely,

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) \tag{1.109}$$

which has the solution

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) = \Phi(t)\mathbf{x}(0) \tag{1.110}$$

The Laplace transform of Eq. (1.109) is simply

$$s\bar{\mathbf{x}}(s) - \mathbf{x}(0) = A\bar{\mathbf{x}}(s) \tag{1.111}$$

where $\bar{\mathbf{x}}(s) = \mathcal{L}\mathbf{x}(t)$, so that the transformed state vector has the form

$$\bar{\mathbf{x}}(s) = (sI - A)^{-1}\mathbf{x}(0) \tag{1.112}$$

Inverse transforming both sides of Eq. (1.112), we obtain the homogeneous solution

$$\mathbf{x}(t) = \mathcal{L}^{-1}(sI - A)^{-1}\mathbf{x}(0) \tag{1.113}$$

so that, comparing Eqs. (1.110) and (1.113), we conclude that the transition matrix can also be obtained by writing

$$\Phi(t) = \mathcal{L}^{-1}(sI - A)^{-1} \tag{1.114}$$

The inverse Laplace transform in Eq. (1.114) can be carried out entry by entry. A more detailed discussion of the transition matrix, including an algorithm for its computation, is presented in Sec. 4.10.

Clearly, the first term on the right side of Eq. (1.107) represents the response to the initial excitation, or the homogeneous solution of Eq. (1.92), and the second term represents the response to the external excitation, or the particular solution. The latter has the form of a convolution integral, and it represents an extension of the convolution integral of Sec. 1.8 to multi-input, multi-output systems. Using the same procedure as in Sec. 1.8, it is not difficult to verify that the value of the convolution integral in Eq. (1.107) does not change if **f** is shifted instead of Φ, so that

$$\begin{aligned}\mathbf{x}(t) &= \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t-\tau)B\mathbf{f}(\tau)d\tau \\ &= \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(\tau)B\mathbf{f}(t-\tau)d\tau \end{aligned} \tag{1.115}$$

Equation (1.92) represents an nth-order system described by n simultaneous first-order differential equations. In Sec. 1.2, however, we encountered a different form of an nth-order system, namely, one described by a single equation, Eq. (1.4), in which the highest derivative in D was of order n, as stated by the accompanying equation, Eq. (1.15). The question arises whether one form can be reduced from the other, and in particular whether Eq. (1.4), in conjunction with Eq. (1.15), can be cast in state form so as to permit a solution of the type given by Eq. (1.107). Indeed, this is possible, and to carry out the conversion we first rewrite Eq. (1.4), in conjunction with Eq. (1.15), in the form

$$\frac{d^n x(t)}{dt^n} = -\frac{a_1}{a_0}\frac{d^{n-1}x(t)}{dt^{n-1}} - \frac{a_2}{a_0}\frac{d^{n-2}(t)}{dt^{n-2}} - \dots - \frac{a_{n-1}}{a_0}\frac{dx(t)}{dt} - \frac{a_n}{a_0}x(t) + \frac{1}{a_0}f(t) \tag{1.116}$$

Then, introducing the transformation

$$\begin{aligned} x(t) &= x_1(t) \\ \frac{dx(t)}{dt} &= \frac{dx_1(t)}{dt} = x_2(t) \\ \frac{d^2x(t)}{dt^2} &= \frac{dx_2(t)}{dt} = x_3(t) \\ &\dots\dots\dots\dots\dots\dots \\ \frac{d^{n-1}x(t)}{dt^{n-1}} &= \frac{dx_{n-1}(t)}{dt} = x_n(t) \\ \frac{d^n x(t)}{dt^n} &= \frac{dx_n(t)}{dt} \end{aligned} \tag{1.117}$$

we can write the desired state equations as follows:

$$\begin{aligned}
\dot{x}_1(t) &= x_2(t) \\
\dot{x}_2(t) &= x_3(t) \\
&\cdots\cdots\cdots\cdots \\
\dot{x}_{n-1}(t) &= x_n(t) \\
\dot{x}_n(t) &= -\frac{a_n}{a_0}x_1(t) - \frac{a_{n-1}}{a_0}x_2(t) - \ldots - \frac{a_2}{a_0}x_{n-1}(t) - \frac{a_1}{a_0}x_n(t) + \frac{1}{a_0}f(t)
\end{aligned} \tag{1.118}$$

Equations (1.118) can be expressed in the special form of Eq. (1.92) given by

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}f(t) \tag{1.119}$$

where the state vector is as given by Eq. (1.93a), the excitation vector is a mere scalar, $\mathbf{f}(t) = f(t)$, and the coefficient matrices have the form

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ \multicolumn{7}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1 \\ -\dfrac{a_n}{a_0} & -\dfrac{a_{n-1}}{a_0} & -\dfrac{a_{n-2}}{a_0} & -\dfrac{a_{n-3}}{a_0} & \ldots & -\dfrac{a_2}{a_0} & -\dfrac{a_1}{a_0} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdot \\ 0 \\ \dfrac{1}{a_0} \end{bmatrix} \tag{1.120a,b}$$

where the second is merely a column matrix, i.e., a vector. Consistent with this, the response given by Eq. (1.107) reduces to

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t-\tau)\mathbf{b}f(\tau)d\tau \tag{1.121}$$

More often than not, in vibrations the system is likely to be described by n simultaneous second-order linear equations with constant coefficients, rather than by a single equation of order n. In this case, using the analogy with Eq. (1.18), we can write the governing equations in the form

$$A_0\ddot{\mathbf{q}}(t) + A_1\dot{\mathbf{q}}(t) + A_2\mathbf{q}(t) = \mathbf{Q}(t) \tag{1.122}$$

where

$$\mathbf{q}(t) = \begin{bmatrix} q_1(t) \\ q_2(t) \\ \vdots \\ q_n(t) \end{bmatrix}, \quad \mathbf{Q}(t) = \begin{bmatrix} Q_1(t) \\ Q_2(t) \\ \vdots \\ Q_n(t) \end{bmatrix} \tag{1.123a,b}$$

are n-dimensional output and input vectors, respectively, and A_0, A_1 and A_2 are $n \times n$ coefficient matrices. It should be noted that the notation for the input and output in Eq. (1.122) differs from that in Eq. (1.18) so as not to conflict with the notation for the state equations, Eq. (1.92). Then, adjoining the identity

$$\dot{\mathbf{q}}(t) \equiv \dot{\mathbf{q}}(t) \tag{1.124}$$

and rewriting Eq. (1.122) as

$$\ddot{\mathbf{q}}(t) = -A_0^{-1}A_1\dot{\mathbf{q}}(t) - A_0^{-1}A_2\mathbf{q}(t) + A_0^{-1}\mathbf{Q}(t) \tag{1.125}$$

where A_0^{-1} is the inverse of A_0, it is not difficult to verify that Eqs. (1.124) and (1.125) can be expressed in the state form (1.92) in which the state vector and the corresponding excitation vector are given by

$$\mathbf{x}(t) = \begin{bmatrix} \mathbf{q}(t) \\ \dot{\mathbf{q}}(t) \end{bmatrix}, \quad \mathbf{f}(t) = \mathbf{Q}(t) \tag{1.126a,b}$$

respectively, and the coefficient matrices have the form

$$A = \begin{bmatrix} 0 & I \\ -A_0^{-1}A_2 & -A_0^{-1}A_1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 \\ A_0^{-1} \end{bmatrix} \tag{1.127a,b}$$

Example 1.7

Recast the second-order differential equation

$$\ddot{q}(t) + bq(t) = Q(t) \tag{a}$$

in state form and derive the step response by means of the transition matrix.

The state equations for the system at hand has the form (1.119), where the state vector is simply

$$\mathbf{x}(t) = [q(t) \quad \dot{q}(t)]^T \tag{b}$$

and the excitation is given by

$$f(t) = Q(t) = u(t) \tag{c}$$

in which $u(t)$ is the unit step function. Moreover, using Eqs. (1.120), the coefficient matrices are

$$A = \begin{bmatrix} 0 & 1 \\ -b & 0 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{d}$$

Using Eq. (1.108), we obtain the transition matrix

$$\begin{aligned}\Phi(t) &= e^{At} = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \dots \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + t\begin{bmatrix} 0 & 1 \\ -b & 0 \end{bmatrix} - \frac{t^2b}{2!}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{t^3b}{3!}\begin{bmatrix} 0 & 1 \\ -b & 0 \end{bmatrix} \\ &\quad + \frac{t^4b^2}{4!}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \dots \\ &= \begin{bmatrix} 1 - \frac{1}{2!}t^2b + \frac{1}{4!}t^4b^2 - \dots & t - \frac{1}{3!}t^3b + \frac{1}{5!}t^5b^2 - \dots \\ -tb + \frac{1}{3!}t^3b^2 - \frac{1}{5!}t^5b^3 + \dots & 1 - \frac{1}{2!}t^2b + \frac{1}{4!}t^4b^2 - \dots \end{bmatrix}\end{aligned}$$

$$= \begin{bmatrix} \cos\sqrt{b}\,t & \dfrac{1}{\sqrt{b}}\sin\sqrt{b}\,t \\ -\sqrt{b}\sin\sqrt{b}\,t & \cos\sqrt{b}\,t \end{bmatrix} \tag{e}$$

Letting $\mathbf{x}(0) = \mathbf{0}$ and inserting Eqs. (c)–(e) into Eq. (1.121), we obtain

$$\begin{aligned} \mathbf{x}(t) &= \int_0^t e^{A(t-\tau)}\mathbf{b} f(\tau)d\tau \\ &= \int_0^t \begin{bmatrix} \cos\sqrt{b}\,(t-\tau) & \dfrac{1}{\sqrt{b}}\sin\sqrt{b}\,(t-\tau) \\ -\sqrt{b}\sin\sqrt{b}\,(t-\tau) & \cos\sqrt{b}\,(t-\tau) \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \mathscr{u}(\tau)d\tau \\ &= \begin{bmatrix} \dfrac{1}{b}\cos\sqrt{b}\,(\tau - t) \\ \dfrac{1}{\sqrt{b}}\sin\sqrt{b}\,(\tau - t) \end{bmatrix}\Bigg|_0^t \mathscr{u}(t) = \begin{bmatrix} \dfrac{1}{b}(1 - \cos\sqrt{b}\,t) \\ \dfrac{1}{\sqrt{b}}\sin\sqrt{b}\,t \end{bmatrix} \mathscr{u}(t) \end{aligned} \tag{f}$$

The step response is simply the top component of $\mathbf{x}(t)$, or

$$\mathscr{s}(t) = \frac{1}{b}(1 - \cos\sqrt{b}\,t)\mathscr{u}(t) \tag{g}$$

where we multiplied the result by $\mathscr{u}(t)$ to account for the fact that $\mathscr{s}(t) = 0$ for $t < 0$.

Example 1.8

Use Eq. (1.114) to derive the transition matrix for the second-order system of Example 1.7.

Using the first of Eqs. (d) of Example 1.7, we can write

$$(sI - A)^{-1} = \begin{bmatrix} s & -1 \\ b & s \end{bmatrix}^{-1} = \frac{1}{s^2 + b}\begin{bmatrix} s & 1 \\ -b & s \end{bmatrix} \tag{a}$$

Hence, using the Laplace transform tables of Appendix A, and considering Eq. (1.114), we obtain the entries of the desired transition matrix Φ as follows:

$$\begin{aligned} \phi_{11} &= \phi_{22} = \mathcal{L}^{-1}\frac{s}{s^2 + b} = \cos\sqrt{b}t \\ \phi_{12} &= \mathcal{L}^{-1}\frac{1}{s^2 + b} = \frac{1}{\sqrt{b}}\sin\sqrt{b}t, \quad \phi_{21} = \mathcal{L}^{-1}\frac{-b}{s^2 + b} = -\sqrt{b}\sin\sqrt{b}t \end{aligned} \tag{b}$$

Clearly, the results agree with those obtained in Example 1.7, and we observe that they were obtained here with less effort. This seems to be true for low-order systems, but not for high-order ones. In fact, for high-order systems the transition matrix must be obtained numerically.

1.10 DISCRETE-TIME SYSTEMS

In Secs. 1.8 and 1.9, we discussed methods for deriving the response of linear, time-invariant systems to arbitrary excitations. In particular, we demonstrated that the response to external excitations can be expressed in the form of a convolution integral. In the case of single-input, single-output systems, the convolution integral involves the impulse response and in the case of multi-input, multi-output systems, it involves the transition matrix. Only in very few cases, evaluation of convolution integrals can be carried out in closed form. Indeed, in the vast majority of cases, evaluation of convolution integrals must be carried out numerically by means of a digital computer.

In processing numerical solutions on a digital computer, it is necessary to provide the information in a form acceptable to the computer. To this end, it is desirable to introduce a number of concepts. Functions of time such as the excitation and response, or input and output, are referred to in system theory as *signals*. Previous discussions have been concerned with inputs and outputs in the form of continuous functions of time. We refer to such functions as *continuous-time signals*. The concept of a continuous signal is somewhat broader than that of a continuous function, in the sense that a discontinuous function with a denumerable number of discontinuities represents a continuous signal. Digital computers cannot work with continuous-time signals but with signals defined for discrete values of time only. Such signals are known as *discrete-time signals*. A system involving continuous-time signals is called a *continuous-time system*, and one involving discrete-time signals is referred to as a *discrete-time system*.

Discrete-time functions can arise naturally. As an example, the amount of currency in a bank at the close of each business day can be regarded as a discrete function of time. The amount represents a sequence of numbers. In our case, the interest in discrete-time systems is due to our desire to process information on a digital computer. To this end, we must convert continuous-time systems into discrete-time systems, and vice versa.

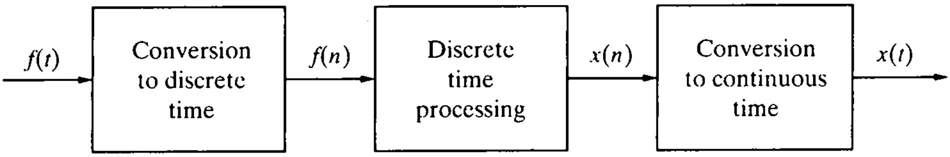


Figure 1.14 The process of computing the response of continuous-time systems in discrete time

The process of computing the response on a digital computer consists of three cascaded operations, as shown in Fig. 1.14. The first operation represents a conversion of the input signal from continuous time to discrete time. This operation is carried out by means of a *sampler*, which converts a continuous-time signal into a sequence of numbers corresponding to the value of the input signal at the sampling instances t_n $(n = 0, 1, 2, \ldots)$. The sampler can be modeled as a switch, where the switch is open for all times except at the sampling instances t_n, when it closes instantaneously to let the signal pass through. Normally, the sampling instances t_n are spaced uniformly in time, so that $t_n = nT$ $(n = 0, 1, 2, \ldots)$, where T is the *sampling period*. The second operation shown in Fig. 1.14 is the discrete-time processing, which implies the computation on a digital computer. The resulting output is a discrete-time signal. The third operation consists of reconstruction of a continuous-time output signal from the discrete-time signal. This operation can be carried out by a *data hold circuit*. The simplest and most frequently used hold is the *zero-order hold*, which maintains the discrete-time signal at the same level until the next sample arrives. The zero-order hold is defined mathematically by

$$x(t) = x(t_n) = x(nT) = x(n), \quad nT \leq t \leq nT + T \tag{1.128}$$

where for simplicity of notation we omitted the sampling period T from the argument. The zero-order hold generates a signal in the form of a *staircase*, as shown in Fig. 1.15. It should be pointed out here that the staircase is continuous when regarded as a signal but discontinuous when regarded as a function. Clearly, the jumps tend to disappear as the sampling period T becomes very small.

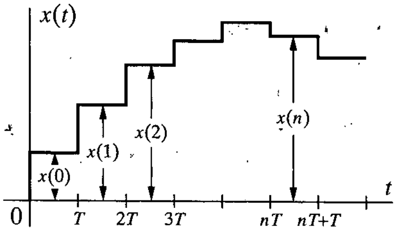


Figure 1.15 Reconstruction of a continuous-time signal from a discrete-time signal by means of a zero-order hold

Strictly speaking, the conversion of continuous-time information to discrete is not sufficient for processing on a digital computer, and some intermediate steps are necessary. To discuss these extra steps, we must introduce additional concepts. An *analog signal* is a signal whose amplitude is not restricted in any particular way. In contrast, a *digital signal* is a signal whose amplitude is restricted to a given set of values. Clearly, both continuous-time and discrete-time signals are analog signals. But digital computers can accept only digital signals, ordinarily encoded in a binary code. Hence, to use a digital computer, we must change the format of the signals from discrete analog to discrete digital, a task carried out by an *analog-to-digital converter*. Of course, after the computations on a digital computer have been completed, digital signals must be converted back to analog signals, which is done by a *digital-to-analog converter*. Hence, in a more complete description of the computational process, the single operation corresponding to the discrete-time processing in Fig. 1.14 is replaced by three cascaded operations, discrete analog-to-discrete digital conversion, digital processing and discrete digital-to-discrete analog conversion. The conversion from discrete analog signals to discrete digital signals involves certain *quantization*, which implies that the analog signal is rounded so as to coincide with the closest value from the restricted set. Hence, the quantization introduces errors, which depend on the number of quantization levels. This number depends on the number of bits of the binary word used by the digital computer. For practical purposes, these errors are mathematically insignificant. In view of this, in subsequent discussions, we make no particular distinction between discrete analog and discrete digital signals and refer to them simply as discrete-time signals.

Next, we turn our attention to the mathematical formalism for computing the response in discrete time. Our interest lies in discrete-time systems obtained through discretization of continuous-time systems. Hence, all discrete signals represent sequences of sample values resulting from sampling continuous-time signals. Assuming that the continuous-time signal shown in Fig. 1.11 is sampled every T seconds beginning at $t = 0$, the discrete-time signal $f(nT) = f(n)$ consists of the sequence $f(0), f(1), f(2), \ldots$. To describe this sequence mathematically, it is convenient to introduce the *discrete-time unit impulse*, or *unit sample*, as the discrete-time Kronecker delta

$$\delta(n-k) = \begin{cases} 1, & n = k \\ 0, & n \neq k \end{cases} \tag{1.129}$$

The unit impulse is shown in Fig. 1.16. Then the discrete-time signal $f(n)$ can be expressed mathematically in the form

$$f(n) = \sum_{k=0}^{\infty} f(k)\delta(n-k) \tag{1.130}$$

The discrete-time signal $f(n)$ is shown in Fig. 1.17.

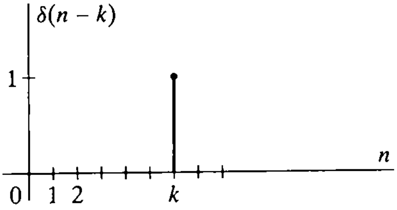


Figure 1.16 Discrete-time unit impulse

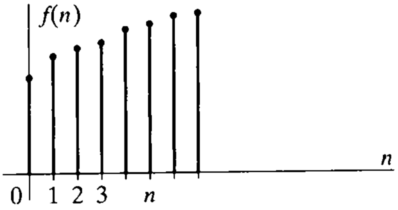


Figure 1.17 Discrete-time signal

Next, we propose to derive the response of a discrete-time system to the excitation $f(n)$ given by Eq. (1.130). To this end, it is important to recognize that a linear time-invariant system in continuous time remains a linear time-invariant system in discrete time, i.e., the various inherent system properties are not affected by the discretization process. Hence, invoking the analogy with continuous-time systems, we can define the *discrete-time impulse response* $g(n)$ as the response of a linear time-invariant discrete-time system to a unit sample $\delta(n)$ applied at $k = 0$, with all the initial conditions being equal to zero. Due to the time-invariance property, the response to $\delta(n-k)$ is $g(n-k)$ and, due to the homogeneity property, the response to $f(k)\delta(n-k)$ is $f(k)g(n-k)$. It follows that the response of a linear time-invariant discrete-time system to the excitation given by Eq. (1.130) is simply

$$x(n) = \sum_{k=0}^{\infty} f(k)g(n-k) = \sum_{k=0}^{n} f(k)g(n-k) \tag{1.131}$$

and we note that we replaced the upper limit in the series by n in recognition of the fact that $g(n-k) = 0$ for $n-k<0$, which is the same as $k>n$. Equation (1.131) expresses the response of linear time-invariant discrete-time systems in the form of a *convolution sum*, and it represents the discrete-time counterpart of the convolution integral given by Eq. (1.87). Here too it can be shown that it does not matter which of the two discrete-time signals, $f(n)$ or $g(n)$, is shifted.

Equation (1.131) represents a discrete-time signal, i.e., a sequence of numbers $x(0), x(1), x(2), \ldots$. Using a hold, the sequence of numbers can be used to generate a continuous-time signal, thus completing the task of deriving the system response.

Computation of the response by means of the convolution sum has several drawbacks. In the first place, the approach is confined to single-input, single-output systems, whereby the computation of every number $x(n)$ in the response sequence is carried out independently of the previously computed numbers $x(0), x(1), \ldots, x(n-1)$, and it requires all the values of $f(k)$ and $g(k)$ up to that instant $(k = 0, 1, 2, \ldots, n-1)$. Hence, one must save the excitation and impulse response sequences until the last number in the response sequence has been computed. The reason for this is that the computational process is not recursive. In a recursive process the computation of $x(n)$ requires only the value of $x(n-1)$ and $f(n-1)$, so that the previous values can be discarded. Moreover, the computation of $x(n)$ by the convolution sum becomes progressively longer, as the sum involves $n+1$ products of $f(k)$ and $g(n-k)(k = 0, 1, \ldots, n)$. By contrast, in a recursive process the computational effort is the same for each term in the response sequence. In view of the above, we propose to develop a recursive algorithm for the computation of the response in discrete time. The algorithm is based on the transition matrix approach, so that the response consists of a sequence of state vectors. In this regard, it should be pointed out that a state vector contains velocities, which may not be required. Clearly, the approach is suitable for multi-input, multi-output systems.

The differential equation for a linear time-invariant continuous-time system is

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{f}(t) \tag{1.132}$$

where $\mathbf{x}(t)$ and $\mathbf{f}(t)$ are n-dimensional state vector and r-dimensional input vector, respectively, and A and B are $n \times n$ and $n \times r$ constant coefficient matrices, respectively. Inserting Eq. (1.108) into Eq. (1.107), we can write the solution of Eq. (1.132) as

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) + \int_0^t e^{A(t-\tau)}B\mathbf{f}(\tau)d\tau \tag{1.133}$$

where $\mathbf{x}(0)$ is the initial state vector and $\exp[A(t-\tau)]$ is the transition matrix. Letting $t = kT$ in Eq. (1.133), where T is the sampling period, we obtain the state at that particular sampling time in the form

$$\mathbf{x}(kT) = e^{AkT}\mathbf{x}(0) + \int_0^{kT} e^{A(kT-\tau)}B\mathbf{f}(\tau)d\tau \tag{1.134}$$

At the next sampling, the state is given by

$$\begin{aligned}\mathbf{x}(kT+T) &= e^{A(kT+T)}\mathbf{x}(0) + \int_0^{kT+T} e^{A(kT+T-\tau)}B\mathbf{f}(\tau)d\tau \\ &= e^{AT}\left[e^{AkT}\mathbf{x}(0) + \int_0^{kT} e^{A(kT-\tau)}B\mathbf{f}(\tau)d\tau\right] \\ &\quad + \int_{kT}^{kT+T} e^{A(kT+T-\tau)}B\mathbf{f}(\tau)d\tau\end{aligned} \tag{1.135}$$

Assuming that the sampling period T is sufficiently small that the input vector $\mathbf{f}(t)$ can be regarded as constant over the time interval $kT < t < kT + T$ ($k = 0, 1, 2, \ldots$), we can write

$$\int_{kT}^{kT+T} e^{A(kT+T-\tau)}B\mathbf{f}(\tau)d\tau \cong \left[\int_{kT}^{kT+T} e^{A(kT+T-\tau)}d\tau\right]B\mathbf{f}(kT) \tag{1.136}$$

Moreover, using the change of variables $kT + T - \tau = \sigma$, the integral on the right side of Eq. (1.136) can be reduced to

$$\begin{aligned}\int_{kT}^{kT+T} e^{A(kT+T-\tau)}d\tau &= \int_T^0 e^{A\sigma}(-d\sigma) = \int_0^T e^{A\sigma}d\sigma \\ &= \int_0^T \left(I + A\sigma + \frac{A^2\sigma^2}{2!} + \ldots\right)d\sigma \\ &= IT + \frac{AT^2}{2!} + \frac{A^2T^3}{3!} + \ldots \\ &= A^{-1}\left(AT + \frac{A^2T^2}{2!} + \frac{A^3T^3}{3!} + \ldots\right) \\ &= A^{-1}(e^{AT} - I)\end{aligned} \tag{1.137}$$

Then, introducing the notation

$$e^{AT} = \Phi, \quad \int_0^T e^{A\sigma}d\sigma B = A^{-1}\left(e^{AT} - I\right)B = \Gamma \tag{1.138a,b}$$

dropping T from the argument of $\mathbf{x}$ and $\mathbf{f}$ and combining Eqs. (1.134)– (1.137), we obtain the discrete-time state vector sequence

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \Gamma\mathbf{f}(k), \quad k = 0, 1, 2, \ldots \tag{1.139}$$

where Φ is known as the *discrete-time transition matrix*. Its value can be computed by simply replacing $t - \tau$ by T in Eq. (1.108). Clearly, Eqs. (1.139) represent a recursive algorithm, lending itself to easy programming on a digital computer. Note that Φ and Γ are constant matrices and that $\mathbf{x}(k)$ and $\mathbf{f}(k)$ can be discarded as soon as $\mathbf{x}(k+1)$ has been computed. Moreover, it is easy to see that each state vector requires the same computational effort.

In the case of a single-degree-of-freedom system, we can use the analogy with Eq. (1.119) and rewrite Eq. (1.139) in the single-input form

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \boldsymbol{\gamma} f(k), \quad k = 0, 1, 2, \ldots \tag{1.140}$$

where $\mathbf{x}(k)$ is the two-dimensional discrete-time state vector and $f(k)$ is the scalar input. Moreover, Φ is the 2×2 transition matrix and

$$\boldsymbol{\gamma} = A^{-1}\left(e^{AT} - I\right)\mathbf{b} \tag{1.141}$$

is a two-dimensional coefficient vector, in which $\mathbf{b} = \begin{bmatrix}0 & a_0^{-1}\end{bmatrix}^T$.

Example 1.9

Use the convolution sum to obtain the discrete-time step response of the second-order system of Example 1.7 for $b = 4\ \mathrm{s}^{-2}$. Use $T = 0.05$ s as the sampling period.

By analogy with continuous-time systems, the discrete-time step response $\mathscr{s}(n)$ is defined as the response to the *unit step* sequence

$$\mathscr{u}(n) = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases} \tag{a}$$

Moreover, it is shown in Example 1.10 that the discrete-time impulse response is

$$g(n) = \frac{T}{\sqrt{b}} \sin n\sqrt{b}T, \quad n = 0, 1, 2, \ldots \tag{b}$$

Using Eq. (1.131), the discrete-time step response has the general form

$$\mathscr{s}(n) = \sum_{k=0}^{n} \mathscr{u}(k) g(n-k) \tag{c}$$

Hence, inserting Eqs. (a) and (b) into Eq. (c), we obtain the discrete-time response sequence

$$\begin{aligned}
\mathscr{s}(0) &= \mathscr{u}(0)g(0) = 0 \\
\mathscr{s}(1) &= \sum_{k=0}^{1} \mathscr{u}(k)g(1-k) = \frac{T}{\sqrt{b}} \sin\sqrt{b}T = 0.002496 \\
\mathscr{s}(2) &= \sum_{k=0}^{2} \mathscr{u}(k)g(2-k) = \frac{T}{\sqrt{b}}\left(\sin 2\sqrt{b}T + \sin\sqrt{b}T\right) = 0.007463 \\
\mathscr{s}(3) &= \sum_{k=0}^{3} \mathscr{u}(k)g(3-k) \\
&= \frac{T}{\sqrt{b}}\left(\sin 3\sqrt{b}T + \sin 2\sqrt{b}T + \sin\sqrt{b}T\right) = 0.014851 \\
\mathscr{s}(4) &= \sum_{k=0}^{4} \mathscr{u}(k)g(4-k)
\end{aligned} \tag{d}$$

$$= \frac{T}{\sqrt{b}}\left(\sin 4\sqrt{b}T + \sin 3\sqrt{b}T + \sin 2\sqrt{b}T + \sin\sqrt{b}T\right) = 0.024586$$

. . . .

The response sequence is plotted in Fig. 1.18.

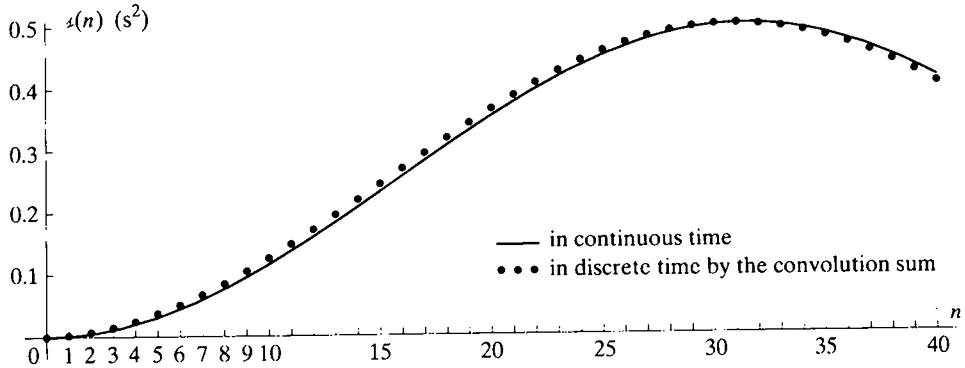


Figure 1.18 Step response of a second-order system in continuous time and in discrete time by the convolution sum.

Example 1.10

Use the transition matrix approach to derive the discrete-time impulse response for the system of Example 1.7.

The discrete-time response is given by Eqs. (1.140), in which the discrete-time transition matrix Φ can be obtained from Eq. (e) of Example 1.7 in the form

$$\Phi = e^{AT} = \begin{bmatrix} \cos\sqrt{b}T & \dfrac{1}{\sqrt{b}}\sin\sqrt{b}T \\ -\sqrt{b}\sin\sqrt{b}T & \cos\sqrt{b}T \end{bmatrix} \tag{a}$$

Moreover, recognizing that $a_0 = 1$, we can use Eq. (1.141) and write

$$\begin{aligned}\gamma &= \frac{1}{b}\begin{bmatrix} 0 & -1 \\ b & 0 \end{bmatrix}\begin{bmatrix} \cos\sqrt{b}T - 1 & \dfrac{1}{\sqrt{b}}\sin\sqrt{b}T \\ -\sqrt{b}\sin\sqrt{b}T & \cos\sqrt{b}T - 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{b}\begin{bmatrix} 1 - \cos\sqrt{b}T \\ \sqrt{b}\sin\sqrt{b}T \end{bmatrix} \end{aligned} \tag{b}$$

In addition, the initial state vector is zero and the discrete-time excitation has the form of the scalar sequence

$$f(0) = 1, \qquad f(n) = 0, \quad n = 1, 2, \ldots \tag{c}$$

Hence, inserting Eqs. (a)–(c) into Eq. (1.140), we obtain the discrete-time response sequence

$$\mathbf{x}(1) = \gamma f(0) = \frac{1}{b}\begin{bmatrix} 1 - \cos\sqrt{b}T \\ \sqrt{b}\sin\sqrt{T} \end{bmatrix}\cdot 1 = \frac{1}{b}\begin{bmatrix} 1 - \cos\sqrt{b}T \\ \sqrt{b}\sin\sqrt{b}T \end{bmatrix}$$

$$\begin{aligned}\mathbf{x}(2) &= \Phi\mathbf{x}(1) + \gamma f(1)\\ &= \frac{1}{b}\begin{bmatrix}\cos\sqrt{b}T & \frac{1}{\sqrt{b}}\sin\sqrt{b}T\\ -\sqrt{b}\sin\sqrt{b}T & \cos\sqrt{b}T\end{bmatrix}\begin{bmatrix}1-\cos\sqrt{b}T\\ \sqrt{b}\sin\sqrt{b}T\end{bmatrix}\\ &= \frac{1}{b}\begin{bmatrix}\cos\sqrt{b}T - \cos 2\sqrt{b}T\\ -\sqrt{b}(\sin\sqrt{b}T - \sin 2\sqrt{b}T)\end{bmatrix}\end{aligned} \tag{d}$$

$$\begin{aligned}\mathbf{x}(3) &= \Phi\mathbf{x}(2) + \gamma f(2)\\ &= \frac{1}{b}\begin{bmatrix}\cos\sqrt{b}T & \frac{1}{\sqrt{b}}\sin\sqrt{b}T\\ -\sqrt{b}\sin\sqrt{b}T & \cos\sqrt{b}T\end{bmatrix}\begin{bmatrix}\cos\sqrt{b}T - \cos 2\sqrt{b}T\\ -\sqrt{b}(\sin\sqrt{b}T - \sin 2\sqrt{b}T)\end{bmatrix}\\ &= \frac{1}{b}\begin{bmatrix}\cos 2\sqrt{b}T - \cos 3\sqrt{b}T\\ -\sqrt{b}(\sin 2\sqrt{b}T - \sin 3\sqrt{b}T)\end{bmatrix}\end{aligned}$$

. . . .

The impulse response is the upper entry in $\mathbf{x}(n)$. By induction, we can write

$$g(n) = \frac{1}{b}\left[\cos(n-1)\sqrt{b}T - \cos n\sqrt{b}T\right] \tag{e}$$

Then, recalling that T is relatively small, we can make the approximation

$$\cos(n-1)\sqrt{b}T \cong \cos n\sqrt{b}T + \sqrt{b}T\sin n\sqrt{b}T \tag{f}$$

so that, introducing Eq. (f) into Eq. (e), the desired discrete-time impulse response is simply

$$g(n) = \frac{T}{\sqrt{b}}\sin n\sqrt{b}T \tag{g}$$

1.11 SYNOPSIS

The basic question in the study of vibrations is how systems respond to given excitations. Depending on the excitation-response characteristics, systems can be classified broadly as linear or nonlinear. Linear system theory is concerned with the excitation-response characteristics, or input-output characteristics, of linear systems. Low-order systems can often be described by a single differential equation, but higher-order systems require a set of simultaneous equations, which can be conveniently cast in matrix form. Consistent with this, linear system theory represents a selection of material from the theory of ordinary differential equations and matrix theory. Of particular interest in vibrations are linear time-invariant systems, or systems with constant coefficients, for which a time shift in the input causes a like time shift in the output.

The solution of differential equations for the response of linear time-invariant systems can be obtained in a variety of ways. Our interest lies in methods of solution capable of extracting the largest amount of information possible about the system behavior. Of course, the choice of methodology depends largely on the nature of the excitation. Indeed, there is a significant difference between the approach to steady-state problems and the approach to transient problems, where the first include harmonic, and in general periodic excitations, and the second include initial and nonperiodic excitations. As indicated in Sec. 1.5, in the case of harmonic excitations, frequency response plots, i.e., plots of the magnitude and phase angle of the response as functions of the frequency, are significantly more informative than time descriptions of the response. On the other hand, in the case of transient excitations time descriptions of the response are more indicated. The transient response can be obtained conveniently by means of the Laplace transformation method, which leads to a convolution integral. The concept of transfer function and the intimately related impulse response play a pivotal role in the derivation of the convolution integral. In the case of multi-input, multi-output systems, it is often advisable to cast the equations in state form. Then, the response to arbitrary excitations can be conveniently produced by means of a matrix form of the convolution integral using the transition matrix.

In most cases of practical interest, it is necessary to evaluate the response on a digital computer. But the excitation and response are continuous functions of time, and digital computers cannot handle continuous-time variables. Hence, to process the solution on a digital computer, the problem must be discretized in time. In Sec. 1.10, the formalism for converting systems from continuous time to discrete time, computing in discrete time and converting back to continuous time is presented under the title of discrete-time systems.

This chapter represents a selection of topics from linear system theory of particular interest in the study of vibrations. The various concepts and techniques are introduced by means of generic differential equations, without any attempt to derive these equations. Methods for deriving differential equations of motion for general dynamical systems are presented in great detail in Chapter 2. Then, the concepts and techniques presented here are applied in Chapters 3 and 4 to study the excitation-response characteristics of vibrating systems.

PROBLEMS

1.1 Determine whether the system described by the differential equation

$$t^2\frac{d^2x}{dt^2} + t\frac{dx}{dt} + \left(t^2 - c^2\right)x = 0$$

is linear or nonlinear by checking the homogeneity and the additivity properties.

1.2 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + \sin 3t = 0$$

1.3 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + (\sin 3t)\,x = 0$$

1.4 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + \sin 3x = 0$$

1.5 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + (\sin x)\,x = 0$$

1.6 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} + \left(t^2 - c^2x^2\right)x = 0$$

1.7 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + (3 + 2\cos t)\,x = 0$$

1.8 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + (3 + 2\cos x)\,x = 0$$

1.9 Repeat Problem 1.1 for the differential equation

$$\frac{d^2x}{dt^2} + \epsilon\left(x^2 - 1\right)\frac{dx}{dt} + x = 0$$

1.10 Determine whether or not the system described by the differential equation

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} + (3 + 5\cos 2t)\,x = 0$$

is time-invariant.

1.11 Repeat Problem 1.10 for the system of Problem 1.3.

1.12 Repeat Problem 1.10 for the system of Problem 1.4.

1.13 Repeat Problem 1.10 for the system of Problem 1.6.

1.14 Repeat Problem 1.10 for the system of Problem 1.9.

1.15 Derive and plot the response of the system described by the differential equation

$$\frac{dx}{dt} + x = 0$$

to the initial condition $x(0) = 2$.

1.16 Derive and plot the response of the system described by the differential equation

$$\frac{d^2x}{dt^2} + 8\frac{dx}{dt} + 25x = 0$$

to the initial conditions $x(0) = 2,\ \dot{x}(0) = 3$.

1.17 A system is described by the differential equation

$$2\frac{dx}{dt} + 4x = 3\cos\omega t$$

Determine the impedance, the frequency response and the system response. Then plot the magnitude and the phase angle of the frequency response.

1.18 Repeat Problem 1.17 for the system

$$\frac{d^2x}{dt^2} + 0.4\frac{dx}{dt} + 4x = 5\sin\omega t$$

1.19 Determine the response of the system described by the differential equation

$$\frac{d^2x}{dt^2} + \frac{dx}{dt} + 25x = 5\sin(\omega t - \psi)$$

Draw conclusions concerning the effect of the phase angle ψ.

1.20 Determine the response of the system described by the differential equation

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 25x = \left(\frac{d}{dt} + 2\right)5e^{i\omega t}$$

1.21 Derive the transfer function for the system described by the differential equation

$$\frac{d^2x}{dt^2} + 0.4\frac{dx}{dt} + 4x = f$$

1.22 Repeat Problem 1.21 for the system

$$\frac{d^4x}{dt^2} + 2\frac{d^3x}{dt^3} + 3\frac{dx^2}{dt^2} + \frac{dx}{dt} + x = f$$

1.23 Repeat Problem 1.21 for the system

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 25x = \left(\frac{d}{dt} + 2\right)f$$

1.24 Derive the impulse response for the system described by the differential equation

$$\frac{dx}{dt} + 2x = f$$

1.25 Repeat Problem 1.24 for the system

$$\frac{d^3x}{dt^3} + \frac{d^2x}{dt^2} + 4\frac{dx}{dt} + 4x = f$$

1.26 Derive the step response for the system of Problem 1.24.

1.27 Derive the step response for the system of Problem 1.21.

1.28 Derive the step response for the system of Problem 1.25.

1.29 Derive the step response of the system of Example 1.3 by integrating the impulse response obtained in Example 1.3.

1.30 Derive the ramp response for the system of Problem 1.24.

1.31 Derive the ramp response for the system of Problem 1.24 by integrating the step response obtained in Problem 1.26 with respect to time.

1.32 Derive the ramp response of the system described by the differential equation

$$\frac{d^2x}{dt^2} + 4x = f$$

in two ways, first by means of Eq. (1.72) and then by integrating the step response.

1.33 Determine the response of the system of Problem 1.32 to the rectangular pulse shown in Fig. 1.19 by regarding it as a superposition of two step functions.

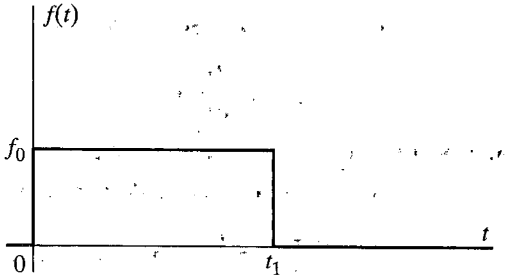


Figure 1.19 Rectangular pulse

1.34 Repeat Problem 1.33 for the system of Problem 1.21.

1.35 Determine the response of the system of Problem 1.32 to the trapezoidal pulse shown in Fig. 1.20 by regarding it as a superposition of one step function and two ramp functions.

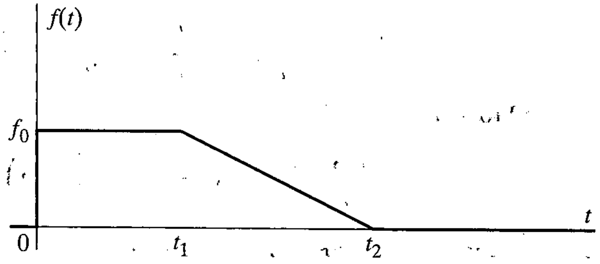


Figure 1.20 Trapezoidal pulse

1.36 Repeat Problem 1.35 for the system of Problem 1.21.

1.37 Derive the step response of the system of Problem 1.32 by means of the convolution integral.

1.38 Derive the step response of the system of Problem 1.21 by means of the convolution integral.

1.39 Solve Problem 1.33 by the convolution integral, as outlined in Example 1.6.

1.40 Solve Problem 1.35 by the convolution integral, as outlined in Example 1.6.

1.41 Derive the impulse response for the system described by the differential equation

$$\frac{d^2x}{dt^2} + 2\frac{dx}{dt} + 25x = f$$

by means of the approach based on the transition matrix.

1.42 Derive the impulse response for the system of Problem 1.21 by means of the approach based on the transition matrix.

1.43 Repeat Problem 1.42 for the system of Problem 1.25.

1.44 Derive the step response for the system of Problem 1.41 by means of the approach based on the transition matrix.

1.45 Repeat Problem 1.44 for the system of Problem 1.21.

1.46 Repeat Problem 1.44 for the system of Problem 1.25.

1.47 Derive the ramp response for the system of Problem 1.32 by means of the approach based on the transition matrix.

1.48 Repeat Problem 1.47 for the system of Problem 1.21.

1.49 Show that the discrete-time impulse response for the system of Problem 1.24 is

$$g(n) = Te^{-2nT}$$

Then use the convolution sum to derive the discrete-time response.

1.50 Determine the discrete-time response of the system of Problem 1.24 to the trapezoidal pulse of Fig. 1.20 by means of the convolution sum.

1.51 Use the convolution sum to derive the discrete-time response of the system of Example 1.8 to the rectangular pulse of Fig. 1.19 for the case in which $t_1 = 0.5$ s. Plot the response for $0 \leq n \leq 20$.

1.52 Repeat Problem 1.51 for the case in which the excitation is as shown in Fig. 1.20, where $t_1 = 0.5$ s, $t_2 = 1.0$ s. Plot the response for $0 \leq n \leq 30$.

1.53 Solve Problem 1.51 by the approach based on the discrete-time transition matrix. Compare results with those obtained in Problem 1.51 and draw conclusions.

1.54 Repeat Problem 1.54 for the system of Problem 1.52.

BIBLIOGRAPHY

1. Kailath, T., *Linear Systems*, Prentice Hall, Englewood Cliffs, NJ, 1980.
2. Meirovitch, L., *Introduction to Dynamics and Control*, Wiley, New York, 1985.
3. Oppenheim, A. V., Willsky, A. S. and Young, I. T., *Signals and Systems*, Prentice Hall, Englewood Cliffs, NJ, 1983.

2

PRINCIPLES OF NEWTONIAN AND ANALYTICAL DYNAMICS

Mechanics is the oldest and most fundamental part of physics. It is concerned with the equilibrium and motion of bodies. Mechanics has inspired the development of many elegant areas of mathematics. Problems of mechanics have aroused the interest of ancient Greek physicists, such as Aristotle, who lived in the fourth century B.C. The idea behind the principle of virtual work, which is concerned with the equilibrium of a body, is attributed to Aristotle. Also attributed to him is the concept of virtual displacements, although Aristotle based his ideas on virtual velocities. Aristotle was less successful in explaining the motion of bodies, as he advanced a law of motion that in modern terms implies that force is the product of mass and velocity.

The first step toward placing the study of dynamics on a firm scientific foundation was taken by Galileo about two millennia after Aristotle. Galileo developed the concepts of acceleration and inertia, or mass, as well as inertial space. He based his laws of motion on results of experiments with falling bodies, thus establishing experimental methods as an integral part of scientific research.

Toward the end of the seventeenth century, expanding on the ideas of Galileo, Newton enunciated his laws of motion (*Philosophiae Naturalis Principia Mathematica*, 1687). Newton's laws were formulated for single particles, and can be extended to systems of particles and rigid bodies. Like Galileo, he postulated the existence of an absolute space relative to which motions must be measured. Moreover, time is absolute and independent of space. In the case of a single particle, Newton's second law leads to a differential equation, which must be integrated to obtain the motion of the particle. In the case of a system of particles, a differential equation must be written for each particle. These equations contain constraint forces resulting from kinematical conditions, and solutions tend to be difficult to obtain. One of the difficulties can be traced to the presence of constraint forces in the equations of motion, although in most cases these forces present no particular interest. Basic concepts in *Newtonian mechanics* are displacement, force and momentum, all vector quantities. For this reason, Newtonian mechanics is also known as *vectorial mechanics*. Another important feature of Newtonian mechanics is that the motion is described in terms of physical coordinates.

Pioneering work on dynamics was carried out by Leibniz, a contemporary of Newton. He proposed an alternative to Newton's laws of motion in the form of a statement that the work performed by a force over a given path causes a change in the kinetic energy from its value at the initial point to its value at the final point. By replacing the concepts of force and momentum by work and kinetic energy, the approach permits an easy extension from single particles to systems of particles and rigid bodies. Indeed, this obviates the need for free-body diagrams, required of each particle in Newtonian mechanics, and eliminates constraint forces internal to the system automatically. It should be noted that force and momentum are vector quantities and work and kinetic energy are scalar quantities. Without a doubt, many of the concepts developed by Leibniz lie at the foundation of analytical mechanics.

The importance of d'Alembert's contributions to dynamics in the eighteenth century is acknowledged widely, although there is some controversy concerning the nature of these contributions. By creating the concept of inertia forces, he reduced the problems of dynamics to problems of equilibrium. This step by itself is of marginal value, as little is gained beyond what can be obtained by the Newtonian approach. What is important is the fact that this step permits the extension of the virtual work principle to problems of dynamics. Indeed, according to this approach, referred to as the *generalized principle of d'Alembert*, the derivation of the equations of motion is formulated as a variational problem, thus providing another element essential to analytical mechanics. The generalized d'Alembert's principle has the advantages over the Newtonian approach that it considers the system as a whole, without breaking it into components, and constraint forces performing no work are eliminated from the problem formulation. The controversy surrounding d'Alembert's contribution relates to the generalized principle, which is referred to by some as Lagrange's form of d'Alembert's principle.

Another serious contributor to analytical mechanics is Euler, a contemporary of d'Alembert. Euler's contributions lie in his fundamental research on variational problems known as isoperimetric problems. Of particular note is Euler's equation, an implicit solution to a large class of variational problems.

Whereas Leibniz, d'Alembert and Euler provided many of the ideas and developments, it is Lagrange, a mathematical genius who lived during the eighteenth century and the beginning of the nineteenth century, who must be considered as the real father of analytical mechanics. Indeed, it is Lagrange who used these ideas and developments, including many more of his own, to create a revolutionary approach to the field of dynamics (*Mécanique Analytique*, 1788), and one of extreme beauty. Lagrange recognized that the variational approach has the advantages that it treats the system as a whole, it eliminates constraint forces performing no work and it permits a formulation that is invariant to the coordinates used to describe the motion. In particular, the concept of coordinates is expanded to include the more abstract *generalized coordinates*, scalar coordinates not necessarily having physical meaning. In essence, he took dynamics from the physical, vectorial world of Newton to the abstract world of *analytical mechanics*, in which all the equations of motion of a generic dynamical system can be derived by means of *Lagrange's equations* using just two scalar expressions, the kinetic energy and the virtual work. In the process, he developed the necessary mathematical tools, such as the calculus of variations.

Another contributor to analytical dynamics is Hamilton, who lived in the nineteenth century. Among his contributions, *Hamilton's principle* and *Hamilton's equations* stand out. Lagrange's equations can be derived directly from the generalized principle of d'Alembert. They can also be derived, perhaps more conveniently, from the extended Hamilton's principle, an integral principle using the kinetic energy and virtual work. Lagrange's equations are second-order differential equations and tend to be cumbersome, particularly for nonlinear systems. Using the generalized momenta as auxiliary variables, it is possible to derive from Lagrange's equations a set of twice as many first-order differential equations called Hamilton's equations. The latter equations have the advantages that they are simpler and are in a form suitable for numerical integration.

In this chapter, we begin by presenting elements of Newtonian mechanics, thus providing the background for our real objective, a comprehensive study of analytical dynamics. Because of its fundamental nature, the material in this chapter occupies a special place in vibrations.

2.1 NEWTON'S SECOND LAW OF MOTION

Newtonian mechanics is based on three laws stated for the first time by Isaac Newton. Of the three laws, the second law is the most important and widely used. *Newton's second law* can be stated as follows: *A particle acted upon by a force moves so that the force vector is equal to the time rate of change of the linear momentum vector.* The *linear momentum vector* is defined as

$$\mathbf{p} = m\mathbf{v} \tag{2.1}$$

where m is the *mass* and $\mathbf{v}$ is the *velocity vector*. Hence, Newton's second law can be written mathematically as

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}(m\mathbf{v}) \tag{2.2}$$

where $\mathbf{F}$ is the *force vector*. The mass of the particle is defined as a positive quantity whose value does not depend on time.

In Newtonian mechanics, motions are measured relative to an inertial reference frame, i.e., a reference frame at rest or moving uniformly relative to an average position of "fixed stars." Quantities measured relative to an inertial frame are said to be absolute. Denoting by $\mathbf{r}$ the *absolute position vector* of the particle in an inertial frame, the *absolute velocity vector* is given by

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}} \tag{2.3}$$

in which overdots represent derivatives with respect to time, so that Newton's second law can be rewritten in the familiar form

$$\mathbf{F} = m\mathbf{a} \tag{2.4}$$

where

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \ddot{\mathbf{r}} \tag{2.5}$$

is the *absolute acceleration vector*. The various vector quantities are shown in Fig. 2.1, in which the rectangular coordinates x, y and z represent an inertial system. The units of force are pounds (lb) or newtons (N). The units of mass are pounds · second^2 per inch (lb · s^2/in) or kilograms (kg). Note that in SI units the kilogram is a basic unit and the newton is a derived unit, where one newton is equal to one kilogram · meter per second^2, 1 N = 1 kg · m/s^2.

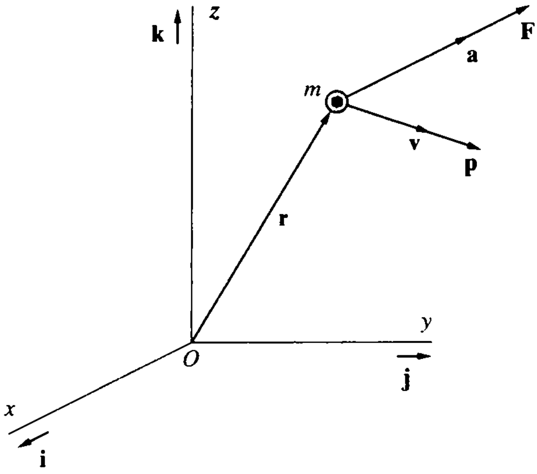


Figure 2.1 Motion of a particle relative to an inertial system

In the absence of forces acting upon the particle, $\mathbf{F} = \mathbf{0}$, Eq. (2.2) reduces to

$$m\mathbf{v} = \text{const} \tag{2.6}$$

which is the mathematical statement of the *conservation of linear momentum principle.*

Example 2.1

Use Newton's second law to derive the equation of motion for the simple pendulum shown in Fig. 2.2a.

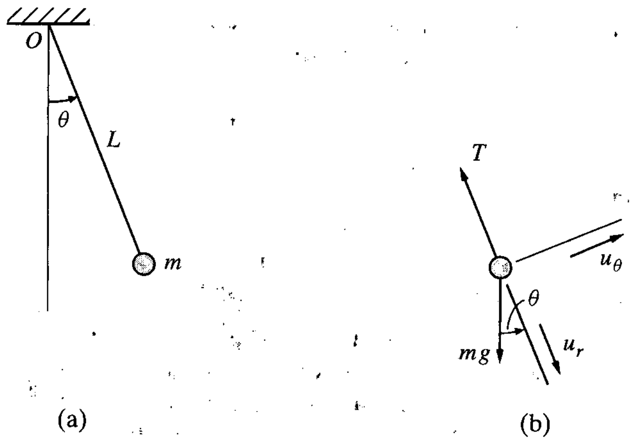


Figure 2.2 **(a)** Simple pendulum **(b)** Free-body diagram for the simple pendulum

The basic tool in deriving the equations of motion by means of Newton's second law is the free-body diagram. A free-body diagram for the simple pendulum is shown in Fig. 2.2b, from which we can write the force vector in terms of radial and transverse components in the form

$$\mathbf{F} = (mg\cos\theta - T)\,\mathbf{u}_r - mg\sin\theta\,\mathbf{u}_\theta \tag{a}$$

where m is the mass, g is the acceleration due to gravity, T is the tension force in the string and $\mathbf{u}_r$ and $\mathbf{u}_\theta$ are unit vectors in the radial and transverse direction, respectively. Moreover, from kinematics, the acceleration for the case at hand has the expression

$$\mathbf{a} = -L\dot{\theta}^2\mathbf{u}_r + L\ddot{\theta}\mathbf{u}_\theta \tag{b}$$

where L is the length of the pendulum. Inserting Eqs. (a) and (b) into Eq. (2.4) and equating the coefficients of $\mathbf{u}_r$ and $\mathbf{u}_\theta$ on both side, we obtain

$$\begin{aligned} mg\cos\theta - T &= -mL\dot{\theta}^2 \\ -mg\sin\theta &= mL\ddot{\theta} \end{aligned} \tag{c}$$

The second of Eqs. (c) yields the equation of motion for the simple pendulum

$$\ddot{\theta} + \frac{g}{L}\sin\theta = 0 \tag{d}$$

Equation (c) represents a nonlinear differential equation, which can be solved for the angle θ. Then, if so desired, from the first of Eqs. (c), we can determine the tension in the string

$$T = m\left(L\dot{\theta}^2 + g\cos\theta\right) \tag{e}$$

2.2 CHARACTERISTICS OF MECHANICAL COMPONENTS

The behavior of a large number of mechanical systems can be described by means of low-order ordinary differential equations. The corresponding mathematical models are referred to as *lumped-parameter* models. The parameters appear as coefficients in the differential equations and they reflect physical characteristics of various mechanical components. Hence, before we consider the problem of deriving the equations of motion, a brief review of mechanical components and their characteristics is in order.

Commonly encountered lumped mechanical components are *springs, dampers* and *masses* (Fig. 2.3). The spring shown in Fig. 2.3a is an elastic component, generally assumed to be massless, that elongates under tensile forces, and vice versa. Because the spring is massless, the force at both terminal points is the same. A typical force-elongation diagram is shown in Fig. 2.4. The elongation tends to be proportional to the force up to a certain point $\Delta x = \Delta x_\ell$, where the constant of proportionality is the *spring constant*, or the *spring stiffness* k. The units of k are pounds per inch (lb/in) or newtons per meter (N/m). Of course, if the force is compressive, then the spring contracts, as shown in Fig. 2.4. Hence, in the range defined by $-\Delta x_\ell < \Delta x < \Delta x_\ell$, called the *linear range*, the spring is said to be *linear* and the force-elongation relation has the form

$$f_s = k\Delta x = k(x_2 - x_1), \quad -\Delta x_\ell < \Delta x < \Delta x_\ell \tag{2.7}$$

where f_s denotes the spring force and x_1 and x_2 are the displacements of the terminal points. Outside the range $-\Delta x_\ell < \Delta x < \Delta x_\ell$, the elongation ceases to be proportional to the force, so that the spring behaves nonlinearly. If $f_s > k\Delta x$, the spring is known as a *hardening spring*, or a *stiffening spring* and if $f_s < k\Delta x$, the spring is referred to as a *softening spring*. In this text, we are concerned primarily with linear springs.

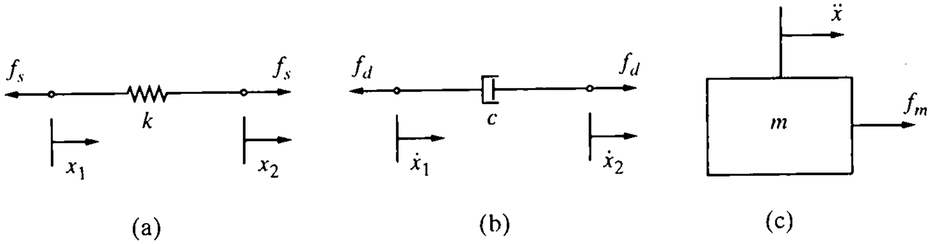


Figure 2.3 Lumped mechanical components **(a)** Elastic spring **(b)** Viscous damper **(c)** Lumped mass

The damper shown in Fig. 2.3b represents a *viscous damper*, or a *dashpot*. It consists of a piston fitting loosely in a cylinder filled with viscous fluid so that the viscous fluid can flow around the piston inside the cylinder. The damper is assumed to be massless, which implies that the force at the two terminal points is the same. If the force causes smooth shear in the viscous fluid, the force in the damper is proportional to the relative velocity of the terminal points, or

$$f_d = c(\dot{x}_2 - \dot{x}_1) \tag{2.8}$$

where the proportionality constant c is known as the *coefficient of viscous damping*, and we note that dots represent the usual derivatives with respect to time. The unit of c are pounds · second per inch (lb · s/in) or newtons · second per meter (N · s/m).

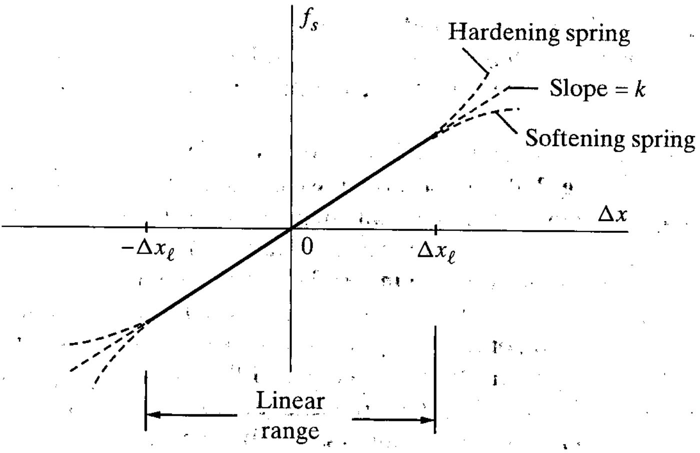


Figure 2.4 Force-elongation diagram for a spring

Finally, Fig. 2.3c shows the lumped mass. By Newton's second law, a force acting upon the mass causes an acceleration proportional to the force where the constant of proportionality is the mass m, or

$$f_m = m\ddot{x} \tag{2.9}$$

The units of m are pounds · second^2 per inch (lb · s^2/in) or newtons · second^2 per meter (N · s^2/m).

On certain occasions, distributed members can be treated as if they were lumped. As an example, we consider the torsion of the cylindrical shaft of circular cross section depicted in Fig. 2.5. The shaft is clamped at the left end and is acted upon by the torsional moment M at the right end. We assume that *the shaft is massless*. The torque M produces a torsional angle θ at the right end. From mechanics of materials, the relation between the torque and the torsional angle is

$$M = \frac{GJ}{L}\theta \tag{2.10}$$

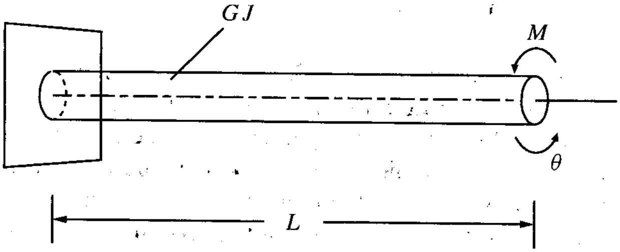


Figure 2.5 Massless shaft in torsion

where GJ is the torsional rigidity, in which G is the shear modulus and J is the cross-sectional area polar moment of inertia, and L is the length of the shaft. But, consistent with Eq. (2.7), the relation between the torque and the torsional angle can be written in the form

$$M = k_T\theta \tag{2.11}$$

where k_T is an *equivalent torsional spring constant.* Comparing Eqs. (2.10) and (2.11), we conclude that the distributed shaft can be treated as a lumped one with the equivalent spring constant

$$k_T = \frac{GJ}{L} \tag{2.12}$$

On other occasions, several springs appear in a given combination. In this regard, we distinguish between *springs in parallel* and *springs in series*, as shown in Figs. 2.6a and 2.6b, respectively. Consistent with the idea that the spring constant represents the ratio between the force and elongation, we wish to derive an equivalent spring constant for each of the two cases. From Fig. 2.6a, we observe that the springs k_1 and k_2 undergo the same elongation. Hence, we can write

$$f_{s1} = k_1(x_2 - x_1), \quad f_{s2} = k_2(x_2 - x_1) \tag{2.13}$$

where f_{s1} and f_{s2} are the forces in springs k_1 and k_2, respectively. But the force f_s at the terminal points must be the sum of the two. Hence, considering Eqs. (2.13), we can write

$$f_s = f_{s1} + f_{s2} = (k_1 + k_2)(x_2 - x_1) = k_{\text{eq}}(x_2 - x_1) \tag{2.14}$$

so that the *equivalent spring constant for springs in parallel* is

$$k_{\text{eq}} = k_1 + k_2 \tag{2.15}$$

Generalizing to the case in which there are n springs in parallel, we can write simply

$$k_{\text{eq}} = \sum_{i=1}^{n} k_i \tag{2.16}$$

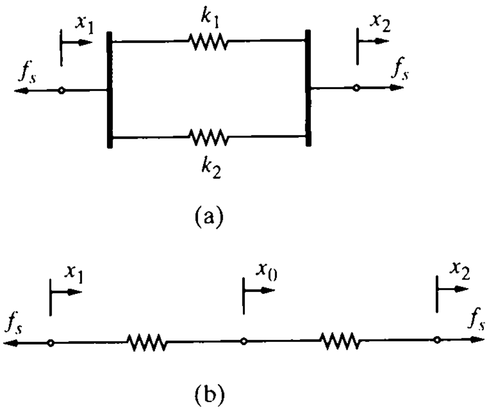


Figure 2.6 **(a)** Springs in parallel **(b)** Springs in series

In the case of springs in series, we observe from Fig. 2.6b that the same force f_s acts throughout both springs. Hence, we can write the relations

$$f_s = k_1(x_0 - x_1), \quad f_s = k_2(x_2 - x_0) \tag{2.17}$$

Eliminating the intermediate variable x_0 from Eqs. (2.17), we obtain

$$f_s = k_{\text{eq}}(x_2 - x_1) \tag{2.18}$$

where the *equivalent spring constant for springs in series* is

$$k_{\text{eq}} = \left(\frac{1}{k_1} + \frac{1}{k_2}\right)^{-1} \tag{2.19}$$

Generalizing, the equivalent constant for n springs in series is

$$k_{\text{eq}} = \left(\sum_{i=1}^{n} \frac{1}{k_i}\right)^{-1} \tag{2.20}$$

The above approach can be used to derive equivalent coefficients of viscous damping for dampers in parallel and dampers in series. In fact, the expressions are virtually identical, except that k_i are replaced by c_i.

2.3 DIFFERENTIAL EQUATIONS OF MOTION FOR FIRST-ORDER AND SECOND-ORDER SYSTEMS

One of the simplest mechanical systems is the *damper-spring system* shown in Fig. 2.7a. We propose to derive the differential equation of the system by a special case of Newton's second law, where the term "special" is in the sense that the mass is equal to zero. Referring to the free-body diagram of Fig. 2.7b, Newton's second law yields

$$\sum F_x = f(t) - f_d(t) - f_s(t) = m\ddot{x}(t) = 0 \tag{2.21}$$

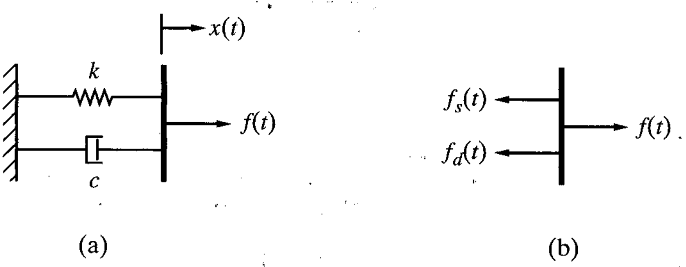


Figure 2.7 **(a)** Damper-spring system **(b)** Free-body diagram for a damper-spring system

where, from Eqs. (2.7) and (2.8), the spring and damper forces are

$$f_s(t) = kx(t), \quad f_d(t) = c\dot{x}(t) \tag{2.22}$$

Inserting Eqs. (2.22) into Eq. (2.21) and rearranging, we obtain the first-order differential equation of motion

$$c\dot{x}(t) + kx(t) = f(t) \tag{2.23}$$

Clearly, Eq. (2.23) represents a *first-order linear time-invariant system*, in which the coefficients c and k are the system parameters.

Next, we consider the *mass-damper-spring system* depicted in Fig. 2.8a. The corresponding free-body diagram is shown in Fig. 2.8b. Following the same approach as above, it is easy to verify that now the equation of motion has the form

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \tag{2.24}$$

which is a *second-order linear differential equation with constant coefficients*. The second-order system described by Eq. (2.24) is commonly known as a *single-degree-of-freedom system*.

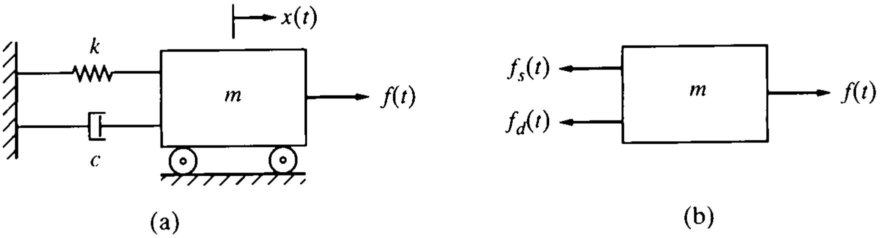


Figure 2.8 **(a)** Mass-damper-spring system **(b)** Free-body diagram for a mass-damper-spring system

Equation (2.24) may appear as describing some special type of systems, seldom encountered in the real world. As it turns out, the equation is representative of a large variety of systems, albeit in each case the parameters may be different. Moreover, the equations of motion of more complex linear multi-degree-of-freedom systems can be reduced to this forms, so that the importance of Eq. (2.24) is far greater than it may appear at this point. In Chapter 3, we study the behavior of first-order and second-order systems, paticularly the latter, based on solutions derived by the methods introduced in Chapter 1.

2.4 MOMENT OF A FORCE AND ANGULAR MOMENTUM

The *moment of momentum vector*, or *angular momentum vector*, of a particle m with respect to point O (Fig. 2.1) is defined as the cross product (vector product) of the radius vector $\mathbf{r}$ and the linear momentum vector $\mathbf{p}$, or

$$\mathbf{H}_O = \mathbf{r} \times \mathbf{p} = \mathbf{r} \times m\dot{\mathbf{r}} \tag{2.25}$$

In view of the fact that m is constant, the rate of change of the angular momentum vector is

$$\dot{\mathbf{H}}_O = \dot{\mathbf{r}} \times m\dot{\mathbf{r}} + \mathbf{r} \times m\ddot{\mathbf{r}} = \mathbf{r} \times m\ddot{\mathbf{r}} \tag{2.26}$$

By definition, however, the moment of a force about O is

$$\mathbf{M}_O = \mathbf{r} \times \mathbf{F} = \mathbf{r} \times m\mathbf{a} = \mathbf{r} \times m\ddot{\mathbf{r}} \tag{2.27}$$

in which use was made of Eqs. (2.4) and (2.5). Comparing Eqs. (2.26) and (2.27), we conclude that

$$\mathbf{M}_O = \dot{\mathbf{H}}_O \tag{2.28}$$

or *the moment of a force about a fixed point is equal to the time rate of change of the angular momentum about that point.*

In the case in which the moment about O is zero, $\mathbf{M}_O = \mathbf{0}$, Eq. (2.28) reduces to

$$\mathbf{H}_O = \text{const} \tag{2.29}$$

which represents the mathematical statement of the *conservation of angular momentum principle.* Note that for the moment to be zero it is not necessary that the force be zero. Indeed, from Eq. (2.27) we conclude that the moment is zero if the force $\mathbf{F}$ is aligned with the radius vector $\mathbf{r}$, i.e., if $\mathbf{F}$ passes through the fixed point O.

Example 2.2

Use the moment equation of motion, Eq. (2.28), to obtain the equation of motion for the simple pendulum of Example 2.1.

The moment equation about O is given by Eq. (2.28). In the case at hand, the radius vector is simply

$$\mathbf{r} = L\mathbf{u}_r \tag{a}$$

so that, inserting Eqs. (a) and (b) of Example 2.1 into Eq. (2.28) and recalling Eqs. (2.26) and (2.27), we obtain

$$\begin{aligned} \mathbf{M}_O &= L\mathbf{u}_r \times [(mg\cos\theta - T)\,\mathbf{u}_r - mg\sin\theta\,\mathbf{u}_\theta] \\ &= L\mathbf{u}_r \times m\left(-L\dot{\theta}^2\mathbf{u}_r + L\ddot{\theta}\mathbf{u}_\theta\right) \end{aligned} \tag{b}$$

Recognizing that $\mathbf{u}_r \times \mathbf{u}_r = \mathbf{0}$ and $\mathbf{u}_r \times \mathbf{u}_\theta = \mathbf{k}$, where $\mathbf{k}$ is a unit vector normal to both $\mathbf{u}_r$ and $\mathbf{u}_\theta$, and canceling $\mathbf{k}$ on both sides of the equation, we can write

$$-Lmg\sin\theta = mL^2\ddot{\theta} \tag{c}$$

Dividing through by mL^2 and rearranging, we obtain Eq. (d) of Example 2.1.

2.5 WORK AND ENERGY

We consider a particle m moving along a curve C under the action of a given force $\mathbf{F}$ (Fig. 2.9). By definition, the *increment of work* performed in moving m from position $\mathbf{r}$ to position $\mathbf{r} + d\mathbf{r}$ is given by the dot product (scalar product)

$$\overline{dW} = \mathbf{F} \cdot d\mathbf{r} \tag{2.30}$$

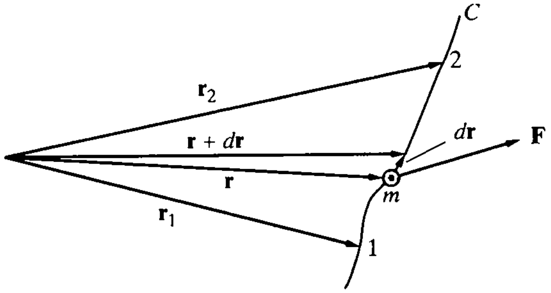


Figure 2.9 Particle moving under the action of a force

where the overbar indicates that $\overline{dW}$, which is a scalar, does not represent the true differential of a function W but simply an infinitesimal expression. Inserting Eqs. (2.3)–(2.5) into Eq. (2.30), we can write

$$\overline{dW} = m\ddot{\mathbf{r}}\cdot d\mathbf{r} = m\frac{d\dot{\mathbf{r}}}{dt}\cdot\dot{\mathbf{r}}dt = m\dot{\mathbf{r}}\cdot d\dot{\mathbf{r}} = d\left(\frac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}\right) = dT \qquad (2.31)$$

In contrast with $\overline{dW}$, however, the quantity on the right side of Eq. (2.31) does represent the differential of a function, namely, the *kinetic energy* given by the expression

$$T = \frac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}} \qquad (2.32)$$

and we note that the kinetic energy is a scalar function. Then, considering the work performed in carrying the particle from position $\mathbf{r}_1$ to position $\mathbf{r}_2$ and using Eqs. (2.30)–(2.32), we obtain

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2}\mathbf{F}\cdot d\mathbf{r} = \int_{\dot{\mathbf{r}}_1}^{\dot{\mathbf{r}}_2} d\left(\frac{1}{2}m\dot{\mathbf{r}}\cdot\dot{\mathbf{r}}\right) = \frac{1}{2}m\dot{\mathbf{r}}_2\cdot\dot{\mathbf{r}}_2 - \frac{1}{2}m\dot{\mathbf{r}}_1\cdot\dot{\mathbf{r}}_1 = T_2 - T_1 \qquad (2.33)$$

where the subscripts 1 and 2 denote quantities associated with positions $\mathbf{r}_1$ and $\mathbf{r}_2$, respectively. Hence, *the work performed in moving the particle from position* $\mathbf{r}_1$ *to position* $\mathbf{r}_2$ *is responsible for a change in the kinetic energy from* T_1 *to* T_2.

There exists one class of forces for which *the work performed in moving a particle from position* $\mathbf{r}_1$ *to position* $\mathbf{r}_2$ *depends only on* $\mathbf{r}_1$ *and* $\mathbf{r}_2$, and not on the path taken to go from $\mathbf{r}_1$ to $\mathbf{r}_2$. Considering two distinct paths I and II, as shown in Fig. 2.10, the preceding statement can be expressed mathematically in the form

$$\underset{\textit{path I}}{\int_{\mathbf{r}_1}^{\mathbf{r}_2}\mathbf{F}\cdot d\mathbf{r}} = \underset{\textit{path II}}{\int_{\mathbf{r}_1}^{\mathbf{r}_2}\mathbf{F}\cdot d\mathbf{r}} \qquad (2.34)$$

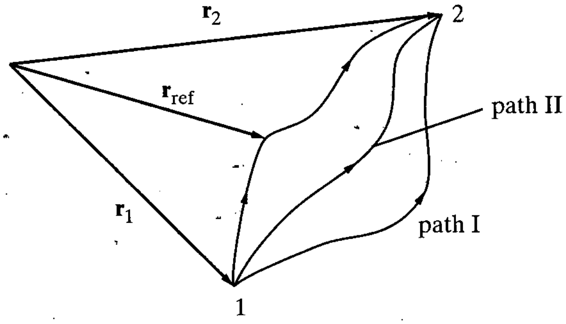


Figure 2.10 Different paths between two positions

Equation (2.34) implies that

$$\oint \mathbf{F} \cdot d\mathbf{r} = 0 \tag{2.35}$$

or *the work performed in moving a particle over a closed path* (beginning at a given point and returning to the same point) *is zero*. Equation (2.35) can be proved by considering a closed path going from $\mathbf{r}_1$ to $\mathbf{r}_2$ over path I and returning from $\mathbf{r}_2$ to $\mathbf{r}_1$ over path II. Forces for which the above statements are true for all possible paths are said to be *conservative*; they will be identified by the subscript c.

Next, we consider a conservative force $\mathbf{F}_c$ and choose a path from $\mathbf{r}_1$ to $\mathbf{r}_2$ passing through a *reference position* $\mathbf{r}_{\text{ref}}$ (Fig. 2.10). The associated work is simply

$$\begin{aligned}\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}_c \cdot d\mathbf{r} &= \int_{\mathbf{r}_1}^{\mathbf{r}_{\text{ref}}} \mathbf{F}_c \cdot d\mathbf{r} + \int_{\mathbf{r}_{\text{ref}}}^{\mathbf{r}_2} \mathbf{F}_c \cdot d\mathbf{r} \\ &= \int_{\mathbf{r}_1}^{\mathbf{r}_{\text{ref}}} \mathbf{F}_c \cdot d\mathbf{r} - \int_{\mathbf{r}_2}^{\mathbf{r}_{\text{ref}}} \mathbf{F}_c \cdot d\mathbf{r}\end{aligned} \tag{2.36}$$

At this point, we define the *potential energy* as *the work performed by a conservative force in moving a particle from an arbitrary position* $\mathbf{r}$ *to the reference position* $\mathbf{r}_{\text{ref}}$, or

$$V(\mathbf{r}) = \int_{\mathbf{r}}^{\mathbf{r}_{\text{ref}}} \mathbf{F}_c \cdot d\mathbf{r} \tag{2.37}$$

and we note that the potential energy, as the kinetic energy, is a scalar function. Inserting Eq. (2.37) into Eq. (2.36), we obtain

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}_c \cdot d\mathbf{r} = V_1 - V_2 = -(V_2 - V_1) \tag{2.38}$$

where $V_i = V(\mathbf{r}_i)$ $(i = 1, 2)$. Equation (2.38) states that *the work performed by a conservative force in moving a particle from* $\mathbf{r}_1$ *to* $\mathbf{r}_2$ *is equal to the negative of the change in the potential energy from* V_1 *and* V_2. It should be pointed out here that our interest lies primarily in changes in the potential energy, rather than in the potential energy itself, so that the reference position is immaterial. It should also be pointed out that in the case of conservative forces the increment of work does represent the differential of a function. Indeed, from Eq. (2.38), we can write

$$dW_c = \mathbf{F}_c \cdot d\mathbf{r} = -dV(\mathbf{r}) \tag{2.39}$$

where $W_c = -V$ is known as the *work function.*

In general, there are both conservative and nonconservative forces, so that the increment of work is simply

$$\mathbf{F} \cdot d\mathbf{r} = \mathbf{F}_c \cdot d\mathbf{r} + \mathbf{F}_{nc} \cdot d\mathbf{r} \tag{2.40}$$

where $\mathbf{F}_{nc}$ denotes the nonconservative force. Introducing Eqs. (2.30), (2.31) and (2.39) into Eq. (2.40) and rearranging, we obtain

$$\mathbf{F}_{nc} \cdot d\mathbf{r} = \mathbf{F} \cdot d\mathbf{r} - \mathbf{F}_c \cdot d\mathbf{r} = dT - (-dV) = d(T + V) = dE \tag{2.41}$$

where

$$E = T + V \tag{2.42}$$

is the *total energy*. Integrating Eq. (2.41) over a path from position $\mathbf{r}_1$ to position $\mathbf{r}_2$, we obtain

$$\int_{\mathbf{r}_1}^{\mathbf{r}_2} \mathbf{F}_{nc} \cdot d\mathbf{r} = \int_{E_1}^{E_2} dE = E_2 - E_1 \tag{2.43}$$

or *the work performed by the nonconservative forces in carrying a particle from position* $\mathbf{r}_1$ *to position* $\mathbf{r}_2$ *is responsible for a change in the total energy from* E_1 *to* E_2.

In the absence of nonconservative forces, i.e., in a conservative force field, $\mathbf{F}_{nc} = \mathbf{0}$, so that Eq. (2.41) yields

$$E = \text{constant} \tag{2.44}$$

Equation (2.44) represents the *principle of conservation of energy*, which explains why the force field defined by Eq. (2.39) is called conservative.

Before we close this section, it is appropriate to examine the potential energy of a very important component in mechanical systems, namely, the spring. Figure 2.11a shows a spring elongating under a force F from an initial length L, when the spring is unstretched, to some final length $L + \delta$. Figure 2.11b depicts the relation between force and elongation for a typical softening spring. In our particular case, the spring remains in the linear range, so that for a given elongation x the spring force is $-kx$, where we recognized that the spring force opposes the elongation. Note that k is the spring constant, which is equal to the slope of the curve F versus x. Hence, using Eq. (2.37) and taking $x = 0$ as the reference position, the potential energy corresponding to the elongation $x = \delta$ is simply

$$V(\delta) = \int_{\delta}^{0} F\,dx = -\int_{\delta}^{0} kx\,dx = \frac{1}{2}k\delta^2 \tag{2.45}$$

and we observe that the potential energy is equal to the shaded area under the curve F versus x. Equation (2.45) indicates that in the linear range the shaded area is triangular.

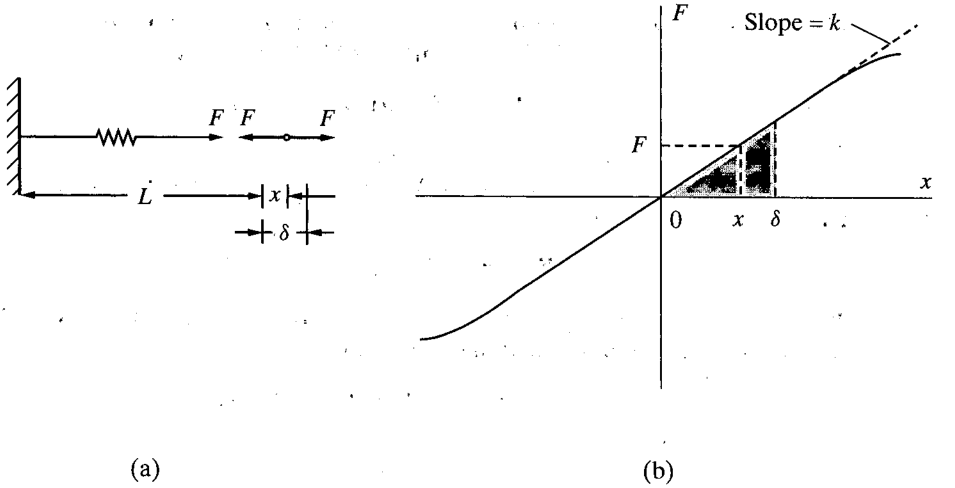


Figure 2.11 **(a)** Spring elongating under a force **(b)** Force-elongation diagram for a softening spring

2.6 SYSTEMS OF PARTICLES AND RIGID BODIES

Newton's laws of motion were formulated for single particles and can be extended to systems of particles and bodies of finite dimensions. To this end, we must recognize that the particles are subjected to two types of forces, *external* and *internal*. External forces are due to sources outside the system and internal forces are due to the interaction among the particles.

We consider a system of N particles of mass $m_i (i = 1, 2, \ldots, N)$, as shown in Fig. 2.12, in which $\mathbf{F}_i$ denote external forces and $\mathbf{f}_{ij}$ denote internal forces exerted by particles m_j on particle m_i ($j = 1, 2, \ldots, N;\ j \neq i$). The internal forces are subject to *Newton's third law*, which can be stated as follows: *The forces that two particles exert upon one another act along the line joining the particles and are equal in magnitude and opposite in direction.* Mathematically, Newton's third law reads

$$\mathbf{f}_{ij} = -\mathbf{f}_{ji}, \quad i, j = 1, 2, \ldots, N; \quad i \neq j \tag{2.46}$$

According to Newton's second law, the equation of motion for particle m_i is

$$\mathbf{F}_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{f}_{ij} = m_i \ddot{\mathbf{r}}_i = m_i \mathbf{a}_i \tag{2.47}$$

where $\ddot{\mathbf{r}}_i = \mathbf{a}_i$ is the acceleration of particle m_i relative to the inertial space xyz. The equation of motion for the system of particles is obtained by extending Eq. (2.47) over the entire system and summing up the corresponding equations. The result is

$$\sum_{i=1}^{N} \mathbf{F}_i + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{f}_{ij} = \sum_{i=1}^{N} m_i \ddot{\mathbf{r}}_i = \sum_{i=1}^{N} m_i \mathbf{a}_i \tag{2.48}$$

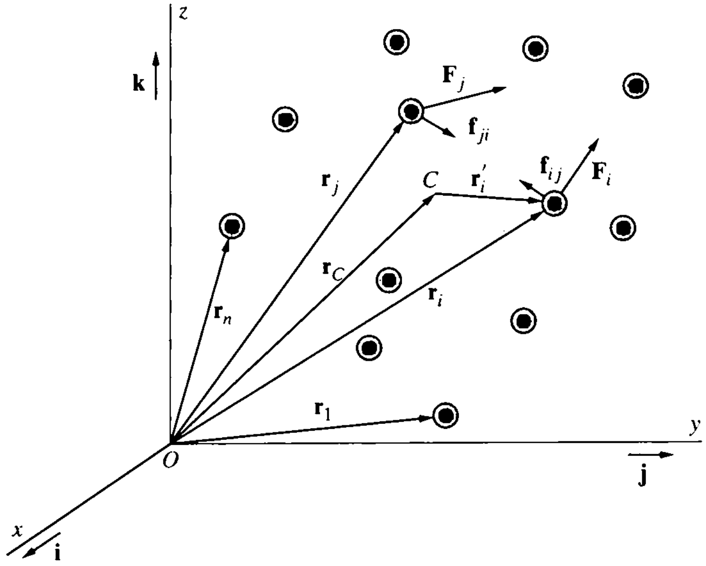


Figure 2.12 System of particles

But by Newton's third law, Eq. (2.46), the internal forces cancel out in pairs, or

$$\sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\mathbf{f}_{ij} = \mathbf{0} \tag{2.49}$$

Moreover, introducing the resultant of the external forces

$$\mathbf{F} = \sum_{i=1}^{N}\mathbf{F}_i \tag{2.50}$$

the equation of motion for the system of particles reduces to

$$\mathbf{F} = \sum_{i=1}^{N} m_i\ddot{\mathbf{r}}_i = \sum_{i=1}^{N} m_i\mathbf{a}_i \tag{2.51}$$

Equation (2.51) represents a relation between the resultant force and the motion of the individual particles in the system. In many cases, however, the interest lies not so much in the motion of the individual particles but in a measure of the motion of the system as a whole. To this end, we introduce the concept of *center of mass C*, defined as a point in space representing a *weighted average position of the system*, where the weighting factor for each particle is the mass of the particle. Denoting the radius vector from O to C by $\mathbf{r}_C$, the mathematical definition of mass center is given by

$$\mathbf{r}_C = \frac{1}{m}\sum_{i=1}^{N} m_i\mathbf{r}_i \tag{2.52}$$

where

$$m = \sum_{i=1}^{N} m_i \tag{2.53}$$

is the total mass of the system. Differentiating Eq. (2.52) twice with respect to time and introducing the result into Eq. (2.51), we obtain the desired equation in the form

$$\mathbf{F} = m\ddot{\mathbf{r}}_C = m\mathbf{a}_C \tag{2.54}$$

Equation (2.54) can be interpreted as describing the motion of a fictitious body equal in mass to the total mass of the system and moving with the acceleration $\mathbf{a}_C$ of the mass center while being acted upon by the resultant $\mathbf{F}$ of the external forces.

For future reference, we wish to examine a certain property of the mass center. To this end, we denote the radius vector from C to m_i by $\mathbf{r}_i'$ (Fig. 2.12), so that

$$\mathbf{r}_i = \mathbf{r}_C + \mathbf{r}_i' \tag{2.55}$$

Inserting Eq. (2.55) into Eq. (2.52) and recalling Eq. (2.53), we obtain

$$\sum_{i=1}^{N} m_i \mathbf{r}_i' = \mathbf{0} \tag{2.56}$$

which defines the mass center C as a point in space such that, if the position of each particle is measured relative to C, the weighted average position is zero.

Equation (2.54) can be written in a different form. To this end, we use the analogy with Eq. (2.1) and define the linear momentum of m_i as

$$\mathbf{p}_i = m_i \mathbf{v}_i \tag{2.57}$$

where $\mathbf{v}_i = \dot{\mathbf{r}}_i$ is the absolute velocity of m_i. Then, if we consider the time derivative of Eq. (2.52), we can write the linear momentum for the system of particles in the form

$$\mathbf{p} = \sum_{i=1}^{N} \mathbf{p}_i = \sum_{i=1}^{N} m_i \mathbf{v}_i = m\mathbf{v}_C \tag{2.58}$$

in which $\mathbf{v}_C$ is the velocity of the mass center C. Introducing the time derivative of Eq. (2.58) into Eq. (2.54), we obtain simply

$$\mathbf{F} = \dot{\mathbf{p}} \tag{2.59}$$

or *the resultant of the external forces acting on the system of particles is equal to the time rate of change of the system linear momentum.*

If the resultant of the external forces is equal to zero, $\mathbf{F} = \mathbf{0}$, Eq. (2.59) yields

$$\mathbf{p} = \text{const} \tag{2.60}$$

which represents the *principle of conservation of linear momentum for a system of particles.*

Next, we define the angular momentum about O of a system of particles as

$$\mathbf{H}_O = \sum_{i=1}^{N} \mathbf{H}_{Oi} = \sum_{i=1}^{N} \mathbf{r}_i \times m_i \mathbf{v}_i \tag{2.61}$$

and the resultant moment about O of the external forces acting on the system of particles as

$$\mathbf{M}_O = \sum_{i=1}^{N} \mathbf{r}_i \times \mathbf{F}_i \tag{2.62}$$

Then, if we consider Eq. (2.47) and recognize from Fig. 2.12 that the moments about O due to the internal forces cancel out in pairs, because $\mathbf{r}_i \times \mathbf{f}_{ij} = -\mathbf{r}_j \times \mathbf{f}_{ji}$, it is not difficult to prove that

$$\mathbf{M}_O = \dot{\mathbf{H}}_O \tag{2.63}$$

or *the moment about a fixed point O of the external forces acting on a system of particles is equal to the time rate of change of the system angular momentum about the same fixed point.*

If the resultant moment about O is zero, $\mathbf{M}_O = \mathbf{0}$, Eq. (2.63) yields

$$\mathbf{H}_O = \text{constant} \tag{2.64}$$

which is the mathematical statement of the *principle of conservation of angular momentum about a fixed point for a system of particles.*

Equation (2.63) represents a simple relation between the moment of the external forces about a fixed point O and the angular momentum of a system of particles about the same fixed point. The question arises whether such a simple relation exists for a moving point. The answer is that there is only one point for which this is true, namely, the mass center of the system of particles. To demonstrate this, we recall that the radius vector from the mass center C to m_i is denoted by $\mathbf{r}'_i$, so that the angular momentum of the system of particles is simply

$$\mathbf{H}_C = \sum_{i=1}^{N} \mathbf{H}_{Ci} = \sum_{i=1}^{N} \mathbf{r}'_i \times m_i \mathbf{v}_i \tag{2.65}$$

On the other hand, the moment of the external forces about C is

$$\mathbf{M}_C = \sum_{i=1}^{N} \mathbf{r}'_i \times \mathbf{F}_i \tag{2.66}$$

Taking the time derivative of Eq. (2.65), considering the time derivative of Eqs. (2.55) and (2.56), as well as Eq. (2.47), and recognizing that $\sum \sum \mathbf{r}'_i \times \mathbf{f}_{ij} = \mathbf{0}$, we obtain

$$\mathbf{M}_C = \dot{\mathbf{H}}_C \tag{2.67}$$

or *the moment of the external forces about the mass center C is equal to the time rate of change of the system angular momentum about C.* It should be reiterated here that Eq. (2.67) holds true only if the moving point is the mass center. If the reference point for the moment and angular momentum of a system of particles is an arbitrary moving point, then in general an extra term appears in the relation.

If the resultant moment about C is zero, $\mathbf{M}_C = \mathbf{0}$, Eq. (2.67) yields

$$\mathbf{H}_C = \text{constant} \tag{2.68}$$

which represents the *principle of conservation of angular momentum about the mass center C for a system of particles.*

The above results provide a clear indication of the useful properties of the mass center. The usefulness does not stop there, however. Indeed, the concept of mass center can be used to separate the kinetic energy into two simple forms. To this end, we write the kinetic energy for a system of particles in the form

$$T = \sum_{i=1}^{N} T_i = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i \tag{2.69}$$

Then, inserting the time derivative of Eq. (2.55) into Eq. (2.69) and considering the time derivative of Eq. (2.56), we obtain

$$\begin{aligned} T &= \frac{1}{2}\sum_{i=1}^{N} m_i \left(\dot{\mathbf{r}}_C + \dot{\mathbf{r}}_i'\right) \cdot \left(\dot{\mathbf{r}}_C + \dot{\mathbf{r}}_i'\right) \\ &= \frac{1}{2}\dot{\mathbf{r}}_C \cdot \dot{\mathbf{r}}_C \sum_{i=1}^{N} m_i + \dot{\mathbf{r}}_C \cdot \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' + \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' \cdot \dot{\mathbf{r}}_i' \\ &= \frac{1}{2} m \dot{\mathbf{r}}_C \cdot \dot{\mathbf{r}}_C + \frac{1}{2}\sum_{i=1}^{n} m_i \dot{\mathbf{r}}_i' \cdot \dot{\mathbf{r}}_i' \end{aligned} \tag{2.70}$$

Introducing the notation

$$T_{\text{tr}} = \frac{1}{2} m \dot{\mathbf{r}}_C \cdot \dot{\mathbf{r}}_C = \frac{1}{2} m \mathbf{v}_C \cdot \mathbf{v}_C. \tag{2.71a}$$

$$T_{\text{rel}} = \frac{1}{2}\sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i' \cdot \dot{\mathbf{r}}_i' = \frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v}_i' \cdot \mathbf{v}_i' \tag{2.71b}$$

we can rewrite Eq. (2.70) as

$$T = T_{\text{tr}} + T_{\text{rel}} \tag{2.72}$$

so that the kinetic energy can be separated into two parts, the first representing the kinetic energy as if all the particles were translating with the velocity of the mass center and the second representing the kinetic energy due to the motion of the particles relative to the mass center. If the reference point were an arbitrary point A, other than the mass center, then a term coupling the motion of A with the motion relative to A would appear, as can be concluded from Eq. (2.70).

In the study of vibrations, we sometimes encounter rigid bodies. But, rigid bodies can be regarded as systems of particles of a special type, namely, one in which the distance between any two particles is constant. It follows that the motion of a particle in the rigid body relative to another is due entirely to the rotation of the rigid body. The force equations of motion for systems of particles, Eq. (2.54), involve only the motion of the mass center, and are not affected by motions relative to the mass center. Hence, the force equations of motion for rigid bodies remain in the form given by Eq. (2.54). In fact, the moment equations of motion for rigid bodies also retain the same general form as for systems of particles, Eq. (2.63) or Eq. (2.67), as the case may be. However, in the case of rigid bodies it is possible to derive more explicit moment equations.

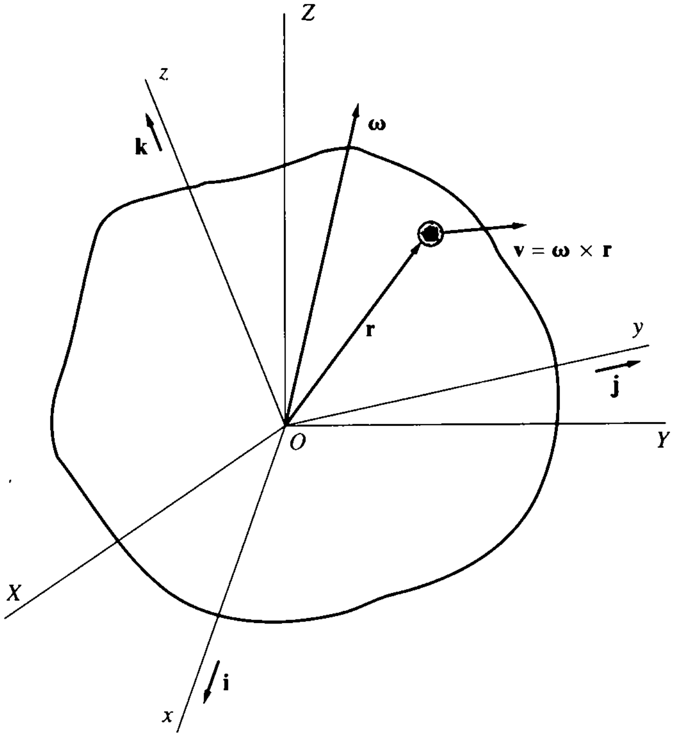


Figure 2.13 Rigid-body rotating relative to the inertial space

In general, it is more convenient to think of a rigid body as a continuous system rather than a system of particles. Hence, although the concepts and definitions presented in this section remain valid, some of the expressions require modification. In particular, the typical mass m_i is to be replaced by a differential of mass dm and the process of summation over the system of particles is to be replaced by integration over the rigid body. Moreover, it is convenient to refer the motion to an auxiliary reference frame embedded in the body and known as *body axes*. Consistent with this, we denote the inertial axes by XYZ and the body axes by xyz (Fig. 2.13). Then, by analogy with Eq. (2.61), the angular momentum about O is defined as

$$\mathbf{H}_O = \int_m \mathbf{r} \times \mathbf{v}\, dm \tag{2.73}$$

Moreover, denoting the angular velocity vector of the rigid body by $\boldsymbol{\omega}$, the velocity of a typical point in the body due to the rotation about O can be shown to be

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{r} \tag{2.74}$$

so that the angular momentum becomes

$$\mathbf{H}_O = \int_m \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r})\, dm \tag{2.75}$$

To obtain a more explicit expression for $\mathbf{H}_O$, we express $\mathbf{r}$ and $\boldsymbol{\omega}$ in terms of rectangular components as follows:

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{2.76}$$

and

$$\boldsymbol{\omega} = \omega_x\mathbf{i} + \omega_y\mathbf{j} + \omega_z\mathbf{k} \tag{2.77}$$

where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are unit vectors along axes x, y and z, respectively, and we recognize that the axes are rotating with the body. Introducing Eqs. (2.76) and (2.77) into Eq. (2.75) and carrying out the various vector operations, we can write the expression of the angular momentum in the compact matrix form

$$\mathbf{H}_O = I_O\boldsymbol{\omega} \tag{2.78}$$

where

$$\mathbf{H}_O = \left[H_{Ox}\ H_{Oy}\ H_{Oz}\right]^T, \quad \boldsymbol{\omega} = \left[\omega_x\ \omega_y\ \omega_z\right]^T \tag{2.79}$$

are column vectors of the angular momentum and angular velocity, respectively, and

$$I_O = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{xy} & I_{yy} & -I_{yz} \\ -I_{xz} & -I_{yz} & I_{zz} \end{bmatrix} \tag{2.80}$$

is the symmetric *inertia matrix*, in which

$$I_{xx} = \int_m (y^2 + z^2)\, dm, \quad I_{yy} = \int_m (x^2 + z^2)\, dm, \quad I_{zz} = \int_m (x^2 + y^2)\, dm \tag{2.81a}$$

are mass moments of inertia about the body axes and

$$I_{xy} = \int_m xy\, dm, \quad I_{xz} = \int_m xz\, dm, \quad I_{yz} = \int_m yz\, dm \tag{2.81b}$$

are mass products of inertia about the same axes.

The moment equation about O is given by Eq. (2.63). Because the vector $\mathbf{H}_O$ is in terms of components about rotating axes, if we recognize that (Ref. 4, Sec. 3.2)

$$\frac{d\mathbf{i}}{dt} = \boldsymbol{\omega} \times \mathbf{i}, \quad \frac{d\mathbf{j}}{dt} = \boldsymbol{\omega} \times \mathbf{j}, \quad \frac{d\mathbf{k}}{dt} = \boldsymbol{\omega} \times \mathbf{k} \tag{2.82}$$

then we can express the moment equation for a rigid body compactly as

$$\mathbf{M}_O = \left.\frac{d\mathbf{H}_O}{dt}\right|_{xyz=\text{fixed}} + \boldsymbol{\omega} \times \mathbf{H}_O = \dot{\mathbf{H}}'_O + \tilde{\omega}\mathbf{H}_O = I_O\dot{\boldsymbol{\omega}} + \tilde{\omega}I_O\boldsymbol{\omega} \tag{2.83}$$

where $\dot{\mathbf{H}}'_O$ is the time derivative of $\mathbf{H}_O$ regarding axes xyz as fixed, $\dot{\boldsymbol{\omega}}$ is the angular acceleration vector and $\tilde{\omega}$ is the skew symmetric matrix

$$\tilde{\omega} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix} \tag{2.84}$$

which is the matrix equivalent of the vector product $\boldsymbol{\omega}\times$. Note that body axes xyz have the advantage that the inertia matrix in terms of these axes is constant.

The angular momentum and moment about the mass center C have the same structure as Eqs. (2.78) and (2.83), respectively, except that the origin of the body axes is at C, so that the mass moments of inertia and mass products of inertia must be modified accordingly.

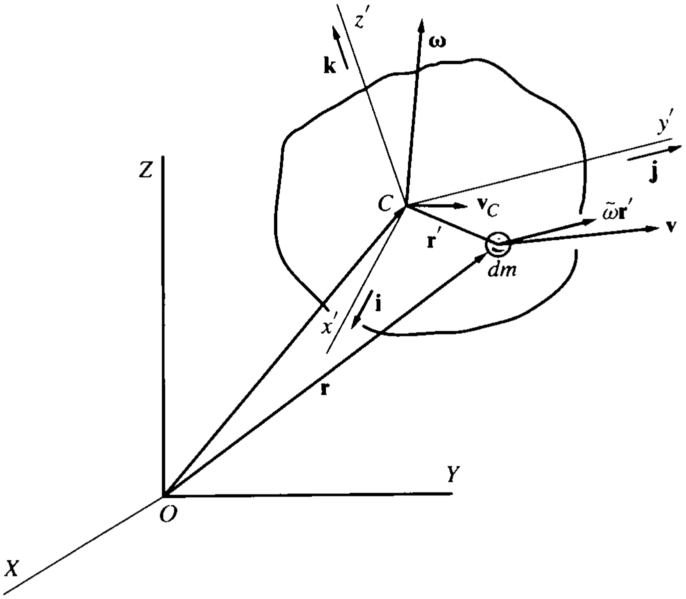


Figure 2.14 Rigid body translating and rotating relative to the inertial space

Finally, we wish to derive an expression for the kinetic energy for a rigid body translating and rotating in space, as shown in Fig. 2.14. To this end, we write the velocity of a typical point in the rigid body in the form

$$\mathbf{v} = \mathbf{v}_C + \tilde{\omega}\mathbf{r}' \tag{2.85}$$

Then, letting $\tilde{r}'$ be a skew symmetric matrix of the type given by Eq. (2.84), but with the rectangular components of $\mathbf{r}'$ replacing those of $\boldsymbol{\omega}$, the kinetic energy can be written as

$$\begin{aligned} T &= \frac{1}{2}\int_m \mathbf{v}^T\mathbf{v}\,dm = \frac{1}{2}\int_m \left(\mathbf{v}_C + \tilde{\omega}\mathbf{r}'\right)^T\left(\mathbf{v}_C + \tilde{\omega}\mathbf{r}'\right)dm \\ &= \frac{1}{2}\mathbf{v}_C^T\mathbf{v}_C\int_m dm + \mathbf{v}_C^T\tilde{\omega}\int_m \mathbf{r}'dm + \frac{1}{2}\int_m \left(\tilde{\omega}\mathbf{r}'\right)^T\tilde{\omega}\mathbf{r}'\,dm \\ &= \frac{1}{2}m\mathbf{v}_C^T\mathbf{v}_C + \frac{1}{2}\int_m \left(-\tilde{r}'\boldsymbol{\omega}\right)^T\left(-\tilde{r}'\boldsymbol{\omega}\right)dm \\ &= \frac{1}{2}m\mathbf{v}_C^T\mathbf{v}_C + \frac{1}{2}\boldsymbol{\omega}^T I_C\boldsymbol{\omega} = T_{\text{tr}} + T_{\text{rot}} \end{aligned} \tag{2.86}$$

where we recognized that $\int_m \mathbf{r}'\,dm = \mathbf{0}$ and, moreover, we introduced the notation

$$T_{\text{tr}} = \frac{1}{2}m\mathbf{v}_C^T\mathbf{v}_C \tag{2.87a}$$

$$T_{\text{rot}} = \frac{1}{2}\boldsymbol{\omega}^T I_C\boldsymbol{\omega} \tag{2.87b}$$

in which

$$I_C = \int_m \left(\tilde{r}'\right)^T\tilde{r}'\,dm \tag{2.88}$$

is the inertia matrix about the mass center C. It has the same form as Eq. (2.80), except that x, y and z are replaced by x', y' and z', respectively. Hence, the kinetic energy is the sum of the kinetic energy of the rigid body as if it were translating with the velocity of the mass center and the kinetic energy of rotation of the rigid body about the mass center. Note that Eq. (2.86) is entirely analogous to Eq. (2.72) for a system of particles, except that here the kinetic energy relative to C is demonstrated to be due entirely to rotation about C.

Example 2.3

The system shown in Fig. 2.15a consists of a uniform rigid bar suspended by a string. Derive the system equations of motion and the kinetic energy.

This example is concerned with an oscillatory system involving a rigid body. Because the motion is planar, all rotations are about the z-axis, whose direction is constant. Hence, in this case it is more advantageous to refer the motion to an inertial set of axes, such as axes xyz shown in Fig. 2.15a, rather than to body axes.

The force equations of motion are given by Eq. (2.54), which requires the acceleration of the mass center C. From Fig. 2.15a, the position vector is

$$\mathbf{r}_C = \left(L_1\sin\theta_1 + \frac{L_2}{2}\sin\theta_2\right)\mathbf{i} - \left(L_1\cos\theta_1 + \frac{L_2}{2}\cos\theta_2\right)\mathbf{j} \tag{a}$$

where $\mathbf{i}$ and $\mathbf{j}$ are constant unit vectors. Hence, the velocity vector is

$$\mathbf{v}_C = \left(L_1\dot{\theta}_1\cos\theta_1 + \frac{L_2}{2}\dot{\theta}_2\cos\theta_2\right)\mathbf{i} + \left(L_1\dot{\theta}_1\sin\theta_1 + \frac{L_2}{2}\dot{\theta}_2\sin\theta_2\right)\mathbf{j} \tag{b}$$

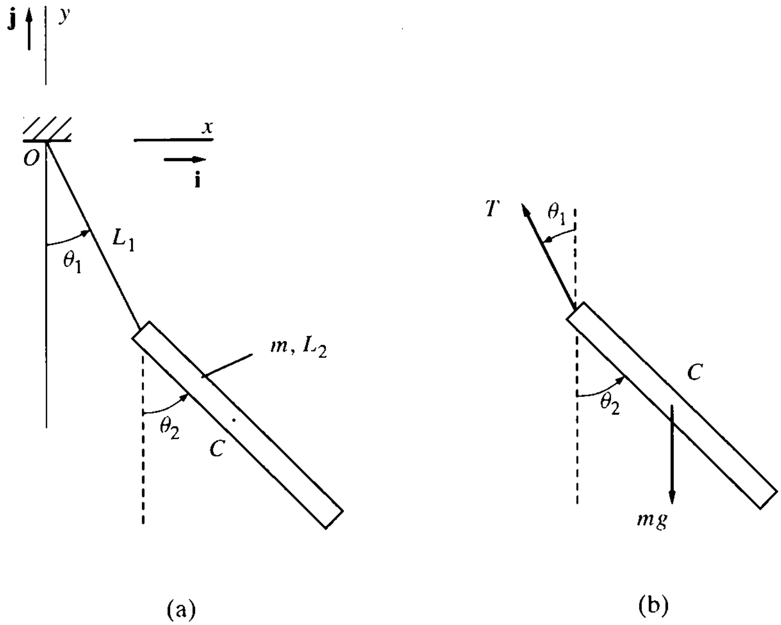


Figure 2.15 **(a)** Uniform rigid bar suspended by a string **(b)** Free-body diagram for the bar

and the acceleration vector has the expression

$$\mathbf{a}_C = \left[L_1\left(\ddot{\theta}_1\cos\theta_1 - \dot{\theta}_1^2\sin\theta_1\right) + \frac{L_2}{2}\left(\ddot{\theta}_2\cos\theta_2 - \dot{\theta}_2^2\sin\theta_2\right)\right]\mathbf{i}$$
$$+ \left[L_1\left(\ddot{\theta}_1\sin\theta_1 + \dot{\theta}_1^2\cos\theta_1\right) + \frac{L_2}{2}\left(\ddot{\theta}_2\sin\theta_2 + \dot{\theta}_2^2\cos\theta_2\right)\right]\mathbf{j} \qquad \text{(c)}$$

Moreover, from the free-body diagram of Fig. 2.15b, the force vector is

$$\mathbf{F} = -T\sin\theta_1\mathbf{i} + (T\cos\theta_1 - mg)\mathbf{j} \qquad \text{(d)}$$

Inserting Eqs. (c) and (d) into Eq. (2.54), the equations of motion in the x- and y-direction are

$$\begin{aligned} m\left[L_1\left(\ddot{\theta}_1\cos\theta_1 - \dot{\theta}_1^2\sin\theta_1\right) + \frac{L_2}{2}\left(\ddot{\theta}_2\cos\theta_2 - \dot{\theta}_2^2\sin\theta_2\right)\right] &= -T\sin\theta_1 \\ m\left[L_1\left(\ddot{\theta}_1\sin\theta_1 + \dot{\theta}_1^2\cos\theta_1\right) + \frac{L_2}{2}\left(\ddot{\theta}_2\sin\theta_2 + \dot{\theta}_2^2\cos\theta_2\right)\right] &= T\cos\theta_1 - mg \end{aligned} \qquad \text{(e)}$$

For bodies undergoing arbitrary motions, it is advantageous to write the moment equation about the mass center. Moreover, because the motion is planar, and measured relative to an inertial reference frame, the moment equation reduces to the scalar form

$$M_C = \dot{H}_C = I_C\alpha \qquad \text{(f)}$$

where I_C is the mass moment of inertia about C and α is the angular acceleration. For the case at hand, they have the values

$$I_C = \frac{mL_2^2}{12}, \quad \alpha = \ddot{\theta}_2 \tag{g}$$

From Fig. 2.15b, the moment about the mass center is

$$M_C = -T\frac{L_2}{2}\sin(\theta_2 - \theta_1) \tag{h}$$

so that, inserting Eqs. (g) and (h) into Eq. (f), the moment equation of motion about C is simply

$$\frac{mL_2^2}{12}\ddot{\theta}_2 = -T\frac{L_2}{2}\sin(\theta_2 - \theta_1) \tag{i}$$

The system equations of motion consist of Eqs. (e) and (i).

According to Eq. (2.86), the kinetic energy has the form

$$T = T_{\mathrm{tr}} + T_{\mathrm{rot}} \tag{j}$$

where, inserting Eq. (b) into Eq. (2.87a), the kinetic energy of translation is

$$\begin{aligned} T_{\mathrm{tr}} &= \frac{1}{2}m\mathbf{v}_C^T\mathbf{v}_C \\ &= \frac{1}{2}m\left[\left(L_1\dot{\theta}_1\cos\theta_1 + \frac{L_2}{2}\dot{\theta}_2\cos\theta_2\right)^2 + \left(L_1\dot{\theta}_1\sin\theta_1 + \frac{L_2}{2}\dot{\theta}_2\sin\theta_2\right)^2\right] \\ &= \frac{1}{2}m\left[L_1^2\dot{\theta}_1^2 + \frac{L_2^2}{4}\dot{\theta}_2^2 + L_1L_2\dot{\theta}_1\dot{\theta}_2\cos(\theta_2 - \theta_1)\right] \end{aligned} \tag{k}$$

Moreover, from Eq. (2.87b), the kinetic energy for planar rotation reduces to

$$T_{\mathrm{rot}} = \frac{1}{2}I_C\omega^2 = \frac{1}{2}\frac{mL_2^2}{12}\dot{\theta}_2^2 \tag{l}$$

2.7 GENERALIZED COORDINATES AND DEGREES OF FREEDOM

In Secs. 2.1–2.6, we introduced fundamental concepts from Newtonian mechanics, which is based on Newton's laws, and in particular on the second law. In Newtonian mechanics, the motion is described in terms of physical coordinates. In this section, we begin making the transition from Newtonian to Lagrangian mechanics, where in the latter the concept of coordinates is enlarged to include more abstract coordinates, not necessarily physical.

In Sec. 2.6, we considered the motion of a system of N particles of mass m_i, where the position of each particle is given by the radius vector $\mathbf{r}_i\,(i = 1, 2, \ldots, N)$. With reference to Fig. 2.12, these positions can be given in terms of rectangular coordinates as follows:

$$\mathbf{r}_i = x_i\mathbf{i} + y_i\mathbf{j} + z_i\mathbf{k}, \quad i = 1, 2, \ldots, N \tag{2.89}$$

The motion of the system is defined completely if the coordinates of all the particles are known functions of time, or

$$x_i = x_i(t), \quad y_i = y_i(t), \quad z_i = z_i(t), \quad i = 1, 2, \ldots, N \tag{2.90}$$

In many problems, the rectangular coordinates x_i, y_i, z_i $(i = 1, 2, \ldots, N)$ are not all independent. In such cases, it is advisable to describe the motion in terms of a different set of coordinates, denoted by $q_1, q_2, \ldots, q_n$, instead of the coordinates x_i, y_i, z_i. The relation between the coordinates $x_i, y_i, z_i (i = 1, 2, \ldots, N)$ and the new coordinates $q_k (k = 1, 2, \ldots, n)$ can be written in the general form

$$\begin{aligned} x_1 &= x_1(q_1, q_2, \ldots, q_n) \\ y_1 &= y_1(q_1, q_2, \ldots, q_n) \\ z_1 &= z_1(q_1, q_2, \ldots, q_n) \\ x_2 &= x_2(q_1, q_2, \ldots, q_n) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ z_N &= z_N(q_1, q_2, \ldots, q_n) \end{aligned} \tag{2.91}$$

Equations (2.91) represent a *coordinate transformation*. We propose to use this transformation to simplify the problem formulation. As a simple example, we consider the double pendulum of Fig. 2.16. The motion of the system can be described by the rectangular coordinates $x_1, y_1, z_1, x_2, y_2, z_2$. It is not difficult to see that $x_1, y_1, z_1, x_2, y_2, z_2$ are not independent, as they are related by the four equations

$$\begin{gathered} x_1^2 + y_1^2 = L_1^2 = \text{constant}, \ (x_2 - x_1)^2 + (y_2 - y_1)^2 = L_2^2 = \text{constant} \\ z_1 = 0, z_2 = 0 \end{gathered} \tag{2.92}$$

Equations (2.92) can be regarded as *constraint equations* reflecting the facts that the length of the strings does not change and that the motion is planar. Rather than

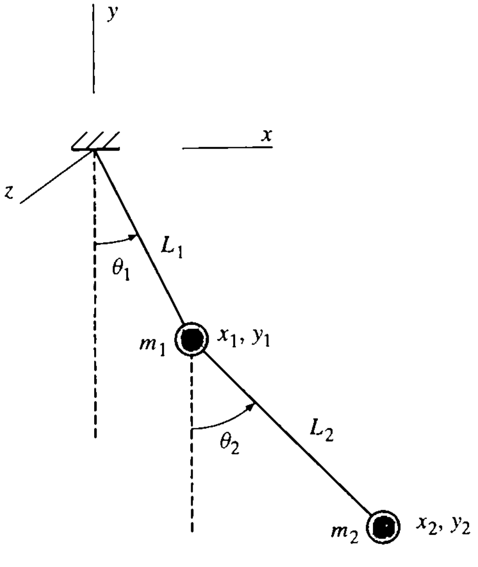


Figure 2.16 Double pendulum

working with rectangular coordinates subject to constraints, it is more convenient to describe the motion in terms of a smaller number of independent coordinates. To this end, we consider the angular displacements θ_1 and θ_2. Hence, letting $\theta_1 = q_1$ and $\theta_2 = q_2$, we can write

$$\begin{aligned} x_1 &= L_1 \sin q_1, \; y_1 = -L_1 \cos q_1, \; z_1 = 0 \\ x_2 &= L_1 \sin q_1 + L_2 \sin q_2, \; y_2 = -L_1 \cos q_1 - L_2 \cos q_2, \; z_2 = 0 \end{aligned} \tag{2.93}$$

Equations (2.93) represent the explicit version of Eqs. (2.91) for the case at hand.

Of course, it was never our intention to describe the planar motion of a double pendulum by means of six rectangular coordinates. In fact, Eqs. (2.92) and (2.93) can be bypassed from the start by formulating the equations of motion in terms of the two angular displacements θ_1 and θ_2 directly. Indeed, two judiciously chosen coordinates are sufficient to describe the motion of the system completely. In general, if a system of N particles moving in a three-dimensional space is subject to c kinematical constraint equations, such as those given by Eqs. (2.92), then the motion of the system can be described completely by n coordinates, where

$$n = 3N - c \tag{2.94}$$

is known as the *number of degrees of freedom* of the system. Hence, the number of degrees of freedom can be defined as the minimum number of coordinates required to describe the motion of a system completely. The n coordinates $q_1, q_2, \ldots, q_n$ capable of describing the motion of the system are called *generalized coordinates*. The concept of generalized coordinates enables us to expand our horizon by accepting as coordinates given functions of physical coordinates or even quantities devoid of physical meaning. The generalized coordinates are not necessarily unique. As an illustration, the motion of the double pendulum can also be described completely by the two angular displacements $\theta = \theta_1$ and $\phi = \theta_2 - \theta_1$. Moreover, later in this text we shall see that the coefficients in a series expansion can play the role of generalized coordinates. The use of generalized coordinates permits a shift in emphasis from the physical world of vectorial mechanics associated with Newton to the more mathematical world of analytical mechanics associated with Lagrange.

2.8 THE PRINCIPLE OF VIRTUAL WORK

The principle of virtual work is essentially a statement of the static equilibrium of a mechanical system. It represents the first variational principle of mechanics. Our interest in the principle is not for the purpose of solving problems of static equilibrium but as a means for effecting the transition from Newtonian to Lagrangian mechanics.

Before we can discuss the virtual work principle, it is necessary to introduce a new class of displacements known as *virtual displacements*. We recall from Sec. 2.6 that the position in space of a system of N particles m_i is defined by the vectors $\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$. Then, the virtual displacements represent *imagined infinitesimal changes* $\delta\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$ in these position vectors that are *consistent with the constraints of the system*, but are otherwise *arbitrary*. The virtual displacements are not true displacements but small *variations* in the system coordinates resulting from imagining the system in a slightly displaced position, a process that does not necessitate any corresponding change in time, so that the forces and constraints do not change during this process. Hence, the virtual displacements take place *instantaneously*. This is in direct contrast with actual displacements, which require a certain amount of time to evolve, during which time the forces and constraints may change. Lagrange introduced the special symbol δ to emphasize the virtual character of the instantaneous variations, as opposed to the symbol d designating actual differentials of positions taking place in the time interval dt. The virtual displacements, being infinitesimal, obey the rules of differential calculus. As an illustration, if a system of N particles is subject to a constraint in the form of the equation

$$f(x_1, y_1, z_1, x_2, \dots, z_N, t) = C \tag{2.95}$$

then the virtual displacements must be such that

$$f(x_1 + \delta x_1, y_1 + \delta y_1, z_1 + \delta z_1, \dots, z_N + \delta z_N, t) = C \tag{2.96}$$

where we note that the time has not been varied. Expanding Eq. (2.96) in a Taylor's series about the position $x_1, y_1, \dots, z_N$, we obtain

$$f(x_1, y_1, z_1, \dots, z_N, t) + \sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \delta x_i + \frac{\partial f}{\partial y_i} \delta y_i + \frac{\partial f}{\partial z_i} \delta z_i \right) + O\left(\delta^2\right) = C \tag{2.97}$$

where $O(\delta^2)$ denotes terms of order two and higher in the virtual displacements. Considering Eq. (2.95) and ignoring the higher-order terms as insignificantly small, we conclude that for the virtual displacements to be consistent with the constraint given by Eq. (2.95) they must satisfy

$$\sum_{i=1}^{N} \left(\frac{\partial f}{\partial x_i} \delta x_i + \frac{\partial f}{\partial y_i} \delta y_i + \frac{\partial f}{\partial z_i} \delta z_i \right) = 0 \tag{2.98}$$

so that only $3N - 1$ of the virtual displacements are arbitrary. In general, the number of arbitrary virtual displacements coincides with the number of degrees of freedom of the system.

Next, we assume that each of the N particles m_i is acted upon by a set of forces with resultant $\mathbf{R}_i$ $(i = 1, 2, \dots, N)$. For a system in equilibrium, the resultant force on each particle vanishes, $\mathbf{R}_i = \mathbf{0}$, so that

$$\overline{\delta W_i} = \mathbf{R}_i \cdot \delta \mathbf{r}_i = 0, \; i = 1, 2, \dots, N \tag{2.99}$$

where $\overline{\delta W_i}$ is the *virtual work* performed by the resultant force $\mathbf{R}_i$ over the virtual displacement $\delta \mathbf{r}_i$. Note that, consistent with Eq. (2.30), the overbar in $\overline{\delta W_i}$ indicates in general a mere infinitesimal expression and not the variation of a function W_i, because such a function does not exist in general; it exists only if the force field is conservative. Summing over the entire system of particles, we obtain simply

$$\overline{\delta W} = \sum_{i=1}^{N} \overline{\delta W_i} = \sum_{i=1}^{N} \mathbf{R}_i \cdot \delta \mathbf{r}_i = 0 \tag{2.100}$$

in which $\overline{\delta W}$ denotes the virtual work for the entire system.

Equation (2.100) appears quite trivial, and indeed it contains no new information. The situation is different, however, when the system is subject to constraints. In this case, we distinguish between *applied*, or *impressed forces* $\mathbf{F}_i$ and *constraint forces* $\mathbf{f}_i$, so that

$$\mathbf{R}_i = \mathbf{F}_i + \mathbf{f}_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{f}_{ij} = \mathbf{0}, \; i = 1, 2, \ldots, N \qquad (2.101)$$

where $\mathbf{f}_{ij}$ are internal forces exerted by particles m_j on particle m_i $(j = 1, 2, \ldots, n)$. Introducing Eqs. (2.101) into Eq. (2.100) and considering Eq. (2.49), we have

$$\overline{\delta W} = \sum_{i=1}^{N} \mathbf{F}_i \cdot \delta\mathbf{r}_i + \sum_{i=1}^{N} \mathbf{f}_i \cdot \delta\mathbf{r}_i + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{f}_{ij} \cdot \delta\mathbf{r}_i = \sum_{i=1}^{N} \mathbf{F}_i \cdot \delta\mathbf{r}_i + \sum_{i=1}^{N} \mathbf{f}_i \cdot \delta\mathbf{r}_i = 0 \qquad (2.102)$$

We confine ourselves to constraint forces that are normal to the virtual displacements. As an example, we consider a particle constrained to a perfectly smooth surface, in which case the constraint force is normal to the surface and the virtual displacements, to be consistent with the constraints, must be parallel to the surface. Note that this rules out reaction forces due to friction, such as those caused by sliding on a rough surface. It follows that the work of the constraint forces through virtual displacements compatible with the system constraints is zero, or

$$\sum_{i=1}^{N} \mathbf{f}_i \cdot \delta\mathbf{r}_i = 0 \qquad (2.103)$$

Virtual displacements for which Eq. (2.103) holds are referred to as *reversible*, because the constraint forces do not prevent replacement of $\delta\mathbf{r}_i$ by $-\delta\mathbf{r}_i$. Confining ourselves to reversible virtual displacements, Eq. (2.102) reduces to

$$\sum_{i=1}^{N} \mathbf{F}_i \cdot \delta\mathbf{r}_i = 0 \qquad (2.104)$$

or *the work performed by the applied forces in infinitesimal reversible virtual displacements compatible with the system constraints is zero*. This is the statement of the *principle of virtual work*.

Equation (2.104) represents a substantially new result. Indeed, because for systems with constraints the virtual displacements $\delta\mathbf{r}_i$ are not all independent, Eq. (2.104) cannot be interpreted as implying that $\mathbf{F}_i = \mathbf{0}$ $(i = 1, 2, \ldots, N)$. To derive the equilibrium equations for such cases, we first rewrite Eqs. (2.91) in the compact form

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \ldots, q_n), \; i = 1, 2, \ldots, N \qquad (2.105)$$

where $q_1, q_2, \ldots, q_n$ are independent generalized coordinates. Then, using Eqs. (2.105), we express the virtual displacement vectors $\delta\mathbf{r}_i (i = 1, 2, \ldots, N)$ in terms of the *generalized virtual displacements* $\delta q_k (k = 1, 2, \ldots, n)$ in the form

$$\delta \mathbf{r}_i = \sum_{k=1}^{n} \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k \, , \; i = 1, 2, \ldots, N \tag{2.106}$$

Inserting Eqs. (2.106) into Eq. (2.104), we obtain

$$\sum_{i=1}^{N} \mathbf{F}_i \cdot \delta \mathbf{r}_i = \sum_{i=1}^{N} \mathbf{F}_i \cdot \sum_{k=1}^{n} \frac{\partial \mathbf{r}_i}{\partial q_k} \delta q_k = \sum_{k=1}^{n} \left(\sum_{i=1}^{N} \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \right) \delta q_k = \sum_{k=1}^{n} Q_k \delta q_k = 0 \tag{2.107}$$

where

$$Q_k = \sum_{i=1}^{N} \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_k} \, , \; k = 1, 2, \ldots, n \tag{2.108}$$

are known as *generalized forces*. Because the generalized virtual displacement δq_k are all independent, and hence entirely arbitrary, Eq. (2.107) can be satisfied if and only if

$$Q_k = 0, \; k = 1, 2, \ldots, n \tag{2.109}$$

Equations (2.109) represent the equilibrium equations.

Example 2.4

A system consisting of a massless rigid bar of length L with both ends mounted on rollers is constrained to move as shown in Fig. 2.17. The left end is attached to a linear spring of stiffness k and the right end supports a hanging mass m. When the bar is horizontal the spring is unstretched. Use the virtual work principle to derive the equilibrium equation.

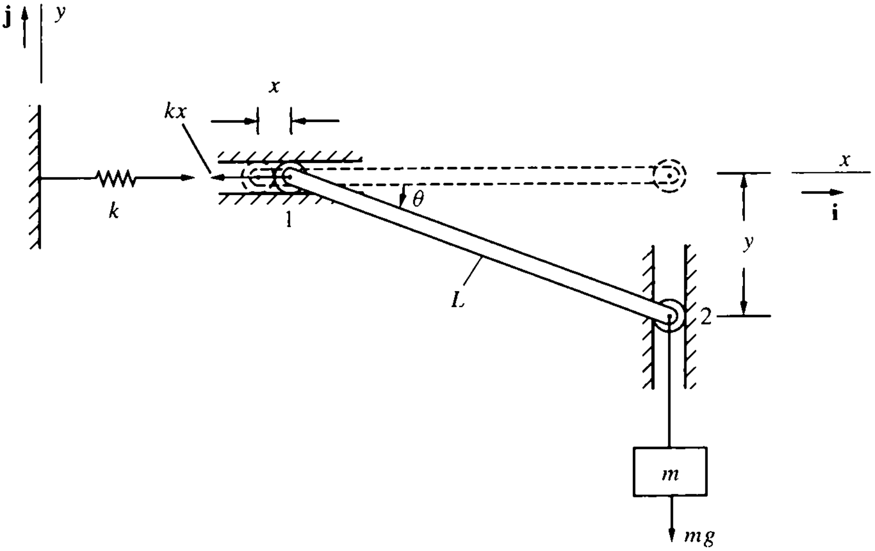


Figure 2.17 Constrained system in equilibrium

The virtual work requires vector expressions for the forces and the virtual displacements. In terms of rectangular components, the forces are given by

$$\mathbf{F}_1 = -kx\mathbf{i}, \quad \mathbf{F}_2 = -mg\mathbf{j} \tag{a}$$

and the displacements by

$$\mathbf{r}_1 = x\mathbf{i}, \quad \mathbf{r}_2 = -y\mathbf{j} \tag{b}$$

so that the virtual displacements have the form

$$\delta\mathbf{r}_1 = \delta x\mathbf{i}, \quad \delta\mathbf{r}_2 = -\delta y\mathbf{j} \tag{c}$$

Using Eqs. (a) and (c), the principle of virtual work for the problem at hand has the expression

$$\sum_{i=1}^{2}\mathbf{F}_i \cdot \delta\mathbf{r}_i = -kx\delta x + mg\delta y = 0 \tag{d}$$

In the present form the virtual work principle is incapable of yielding the equilibrium equation, because x and y are not independent. Indeed, we have used two coordinates to describe the problem, and this is a single-degree-of-freedom system, which requires a single coordinate. The most convenient coordinate is neither x nor y, but the angle θ. The relations between x and y on the one hand and θ on the other are

$$x = L(1 - \cos\theta), \quad y = L\sin\theta \tag{e}$$

so that the virtual displacements are related by

$$\delta x = L\sin\theta\,\delta\theta, \quad \delta y = L\cos\theta\,\delta\theta \tag{f}$$

Inserting Eqs. (e) and (f) into Eq. (d) and factoring out $\delta\theta$, we obtain

$$[-kL(1 - \cos\theta)L\sin\theta + mgL\cos\theta]\,\delta\theta = 0 \tag{g}$$

Because $\delta\theta$ is arbitrary, Eq. (g) can be satisfied for all $\delta\theta$ if and only if the coefficient of $\delta\theta$ is identically zero. Setting the coefficient of $\delta\theta$ equal to zero and rearranging, we obtain the desired equilibrium equation

$$(1 - \cos\theta)\tan\theta = \frac{mg}{kL} \tag{h}$$

Equation (h) is a transcendental equation, and its solution must be obtained numerically. It should be noted that, whereas Eq. (h) admits an infinity of solutions, physically there is only one solution, as θ cannot exceed $\pi/2$.

2.9 THE GENERALIZED PRINCIPLE OF D'ALEMBERT

The principle of virtual work provides a statement of the static equilibrium of a mechanical system. Using an ingenuous idea due to d'Alembert, the principle of virtual work can be extended to dynamical systems, thus paving the way for analytical mechanics.

Newton's second law for a particle of mass m_i can be written in the form

$$\mathbf{F}_i + \mathbf{f}_i + \sum_{\substack{j=1\\ j\neq i}}^{N}\mathbf{f}_{ij} - m_i\ddot{\mathbf{r}}_i = \mathbf{0} \tag{2.110}$$

where $\mathbf{F}_i$ is an applied force, $\mathbf{f}_i$ is a constraint force and $\mathbf{f}_{ij}$ $(j = 1, 2, \ldots, n)$ are internal forces. Equation (2.110) is often referred to as *d'Alembert's principle*. In this context, the term $-m_i\ddot{\mathbf{r}}_i$ is referred to as an *inertia force*.

At first sight, Eq. (2.110) does not seem to provide any new information. Indeed, if the object is to derive the equations of motion, then Eq. (2.110) does not offer anything beyond what Newton's second law does. Under these circumstances, one can question whether there is any justification for referring to Eq. (2.110) as a principle. Hence, we must seek the virtue of Eq. (2.110) by pursuing a different line of thought. Equation (2.110) can be regarded as a statement of dynamic equilibrium of particle m_i, and it is this interpretation that carries far-reaching implications. Indeed, now the principle of virtual work can be extended to dynamics, thus producing the first variational principle of dynamics. To this end, we refer to Eq. (2.110) and write

$$\left(\mathbf{F}_i + \mathbf{f}_i + \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{f}_{ij} - m_i\ddot{\mathbf{r}}_i\right) \cdot \delta\mathbf{r}_i = 0 \tag{2.111}$$

Then, considering a system of N particles and assuming that the virtual displacements $\delta\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$ are reversible, so that Eq. (2.103) holds, we can write for the system of particles

$$\sum_{i=1}^{N} (\mathbf{F}_i - m_i\ddot{\mathbf{r}}_i) \cdot \delta\mathbf{r}_i = 0 \tag{2.112}$$

where we recalled Eq. (2.49). Equation (2.112) embodies both the principle of virtual work of statics and d'Alembert's principle of dynamics and is referred to as the *generalized principle of d'Alembert.* The sum of the applied force and the inertia force, $\mathbf{F}_i - m_i\ddot{\mathbf{r}}_i$, is sometimes referred to as the *effective force*. This permits us to enunciate the generalized principle of d'Alembert as follows: *The virtual work performed by the effective forces through infinitesimal virtual displacements compatible with the system constraints is zero.*

D'Alembert's principle, Eq. (2.112), represents the most general formulation of the problems of dynamics. Its main advantage over Newton's second law is that it obviates the need for constraint forces. Still, Eq. (2.112) is not very convenient for the derivation of the equations of motion, particularly for more complex systems. The generalized principle of d'Alembert is a first variational principle of dynamics, and all other principles can be derived from it. In fact, our own interest in d'Alembert's principle can be traced to the fact that it permits the derivation of another principle, namely, Hamilton's principle.

The generalized principle of d'Alembert is still a vectorial approach using physical coordinates to describe the motion. One objection to the use of physical coordinates is that in many cases they are not independent. In fact, before we can obtain the equations of motion by means of d'Alembert's principle, it is necessary to convert the formulation from a vectorial one in terms of dependent physical coordinates subject to constraints to a scalar one in terms of independent generalized coordinates in a manner similar to the one in Sec. 2.8.

Example 2.5

Derive the equation of motion for the systems of Example 2.4 by means of the generalized d'Alembert's principle.

From Eq. (2.112), the generalized d'Alembert's principle for the case at hand has the form

$$\mathbf{F}_1 \cdot \delta\mathbf{r}_1 + (\mathbf{F}_2 - m_2\ddot{\mathbf{r}}_2) \cdot \delta\mathbf{r}_2 = 0 \tag{a}$$

where $m_2 = m$ and

$$\ddot{\mathbf{r}}_2 = -\ddot{y}\mathbf{j} \tag{b}$$

Hence, inserting Eq. (b) from above and Eqs. (a) and (c) of Example 2.4 into Eq. (a), we obtain

$$-kx\delta x + (mg - m\ddot{y})\,\delta y = 0 \tag{c}$$

As in Example 2.4, we conclude that, to derive the equation of motion for this single-degree-of-freedom system, we must transform Eq. (c) into one in terms of a single generalized coordinate. We choose once again to work with the angle θ as the sole generalized coordinate, so that Eqs. (e) and (f) of Example 2.4 apply here as well and, moreover, we can write

$$\ddot{y} = L\left(\ddot{\theta}\cos\theta - \dot{\theta}^2\sin\theta\right) \tag{d}$$

Introducing Eq. (d) from above and Eqs. (e) and (f) of Example 2.4 into Eq. (c) and factoring out $\delta\theta$, we obtain

$$\left[-kL\,(1 - \cos\theta)\,L\sin\theta + mgL\cos\theta - mL\left(\ddot{\theta}\cos\theta - \dot{\theta}^2\sin\theta\right)L\cos\theta\right]\delta\theta = 0 \tag{e}$$

Due to the arbitrariness of $\delta\theta$, Eq. (e) can be satisfied for all $\delta\theta$ if and only if the coefficient of $\delta\theta$ is identically zero. This yields the desired equation of motion

$$mL\left(\ddot{\theta}\cos\theta - \dot{\theta}^2\sin\theta\right) + kL\,(1 - \cos\theta)\tan\theta - mg = 0 \tag{f}$$

Note that the equilibrium equation can be obtained from Eq. (f) by letting $\dot{\theta} = \ddot{\theta} = 0$.

2.10 HAMILTON'S PRINCIPLE

Although d'Alembert's principle is capable of yielding a complete formulation of the problems of dynamics, Eq. (2.112) is not very convenient, particularly when the virtual displacements are not independent. Indeed, a formulation in terms of generalized coordinates would be more satisfactory. In this regard, we recall from Example 2.5 that, before we could derive the equation of motion, it was necessary to convert the formulation obtained by means of d'Alembert's principle from one in terms of the dependent rectangular coordinates x and y to one in terms of the generalized coordinate θ. Our object is to derive a formulation capable of working directly with generalized coordinates, thus obviating the problem of coordinate transformations. One such formulation is an integral principle known as *Hamilton's principle*. Actually, we will derive a more general version of the principle, referred to as the *extended Hamilton's principle* and containing Hamilton's principle as a special case. The extended Hamilton's principle is not only independent of the coordinates used but also permits the derivation of the equations of motion from a definite integral involving a scalar function, the kinetic energy, and an infinitesimal scalar expression, the virtual work performed by the applied forces.

We propose to derive the extended Hamilton's principle from the generalized principle of d'Alembert, which for a system of N particles is given by Eq. (2.112), in which $\mathbf{F}_i$ are applied forces and $\delta\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$ are virtual displacements. We consider the case in which all $\mathbf{r}_i(t)$ are independent, so that $\delta\mathbf{r}_i$ are entirely arbitrary. Referring to Eq. (2.112), we first recognize that

$$\sum_{i=1}^{N} \mathbf{F}_i \cdot \delta\mathbf{r}_i = \overline{\delta W} \tag{2.113}$$

is the virtual work performed by the applied forces. Then, we assume that the mass m_i is constant and consider the following:

$$\begin{aligned}\frac{d}{dt}(m_i\dot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i) &= m_i\ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i + m_i\dot{\mathbf{r}}_i \cdot \delta\dot{\mathbf{r}}_i = m_i\ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i + \delta\left(\frac{1}{2}m_i\dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i\right) \\ &= m_i\ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i + \delta T_i \end{aligned} \tag{2.114}$$

where T_i is the kinetic energy of particle m_i. Summing over the entire system of particles and rearranging, we obtain

$$\sum_{i=1}^{N} \frac{d}{dt}(m_i\dot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i) = \sum_{i=1}^{N} m_i\ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i + \delta T \tag{2.115}$$

where T is the kinetic energy of the entire system of particles. Inserting Eqs. (2.113) and (2.115) into Eq. (2.112), we can write

$$\delta T + \overline{\delta W} = \sum_{i=1}^{N} \frac{d}{dt}(m_i\dot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i) \tag{2.116}$$

The next step is the integration of Eq. (2.116) with respect to time. Before we take this step, we must introduce some additional concepts. As indicated in Sec. 2.7, the motion of a system of N particles is defined by the position vectors $\mathbf{r}_i(t)$, which represent the solution of the dynamical problem and can be written in terms of rectangular components in the form

$$\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}\,, \quad i = 1, 2, \ldots, N \tag{2.117}$$

We can conceive of a $3N$-dimensional space with the axes x_i, y_i, z_i and represent the position of the system of particles in that space and at any time t as the position of a *representative point* P with the coordinates $x_i(t), y_i(t), z_i(t)$ $(i = 1, 2, \ldots, N)$; the $3N$-dimensional space is known as the *configuration space*. As time unfolds, the representative point P traces a curve in the configuration space called the *true path*, or the *Newtonian path*, or the *dynamical path*. At the same time, we can envision a different representative point P' resulting from imagining the system in a slightly different position defined by the virtual displacements $\delta\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$. As time changes, the point P' traces a curve in the configuration space known as the *varied path*. Multiplying Eq. (2.116) by dt and integrating between the times t_1 and t_2, we obtain

$$\int_{t_1}^{t_2} \left(\delta T + \overline{\delta W}\right) dt = \int_{t_1}^{t_2} \sum_{i=1}^{N} \frac{d}{dt} (m_i \dot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i)\, dt = \sum_{i=1}^{N} m_i \dot{\mathbf{r}}_i \cdot \delta \mathbf{r}_i \Big|_{t_1}^{t_2} \tag{2.118}$$

Of all the possible varied paths, we now consider only those that coincide with the true path at the two instants t_1 and t_2, as shown in Fig. 2.18, so that Eq. (2.118) reduces to

$$\int_{t_1}^{t_2} \left(\delta T + \overline{\delta W}\right) dt = 0, \quad \delta \mathbf{r}_i (t_1) = \delta \mathbf{r}_i (t_2) = \mathbf{0}, \quad i = 1, 2, \ldots, N \tag{2.119}$$

We refer to Eq. (2.119) as the *extended Hamilton's principle*.

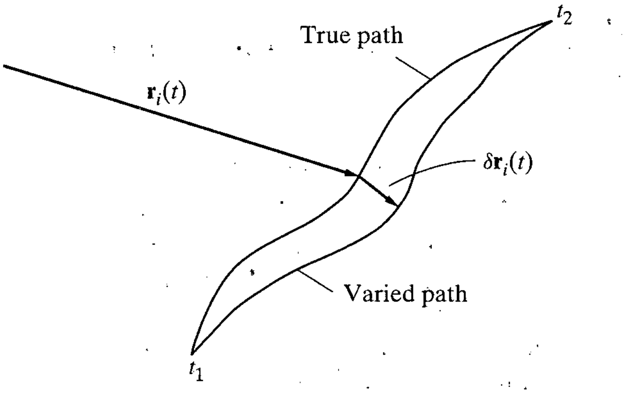


Figure 2.18 True path and varied path in the configuration space

The derivation of the extended Hamilton's principle, Eq. (2.119), was carried out in terms of the physical coordinates $\mathbf{r}_i$ $(i = 1, 2, \ldots, N)$. In many cases, however, it is more desirable to work with the generalized coordinates q_k $(k = 1, 2, \ldots, n)$. In this regard, we recall that, in using the generalized d'Alembert's principle to derive the equations of motion, the transformation from $\delta \mathbf{r}_i$ $(i = 1, 2, \ldots, N)$ to δq_k $(k = 1, 2, \ldots, n)$ must be carried out explicitly, as can be concluded from Example 2.4. In contrast, no such explicit transformation is required here, as the principle has the same form regardless of the coordinates used to express δT and $\overline{\delta W}$. In view of this, we can express δT and $\overline{\delta W}$ directly in terms of independent generalized coordinates, and the same can be said about the conditions on the virtual displacements at t_1 and t_2. Hence, the extended Hamilton's principle can be stated in the form

$$\int_{t_1}^{t_2} \left(\delta T + \overline{\delta W}\right) dt = 0, \quad \delta q_k(t_1) = \delta q_k(t_2) = 0, \quad k = 1, 2, \ldots, n \tag{2.120}$$

where n is the number of degrees of freedom of the system.

The extended Hamilton's principle, Eq. (2.120), is quite general and can be used to derive the equations of motion for a large variety of systems. In fact, although it was derived for a system of particles, the principle is equally valid for rigid bodies, as we shall see in a following example, and for distributed-parameter systems, as we shall have the opportunity to verify in Chapter 7. The only limitation is that the virtual displacements must be reversible, which implies that the constraint forces must perform no work. Hence, the principle cannot be used for systems with friction forces.

In general the virtual work $\overline{\delta W}$ includes contributions from both conservative and nonconservative forces, or

$$\overline{\delta W} = \overline{\delta W_c} + \overline{\delta W}_{nc} \tag{2.121}$$

where the subscripts c and nc denote conservative and nonconservative virtual work, respectively. Using the analogy with Eq. (2.39), however, we conclude that the virtual work performed by the conservative forces can be expressed in the form

$$\overline{\delta W}_c = \delta W_c = -\delta V \tag{2.122}$$

in which $V = V(q_1, q_2, \ldots, q_n)$ is the potential energy. The implication is that the virtual work of the conservative forces represents the variation of the work function W_c, where the work function is the negative of the potential energy V. Then, introducing the *Lagrangian*

$$L = T - V \tag{2.123}$$

a scalar function, and considering Eqs. (2.121) and (2.122), we can rewrite Eq. (2.120) in the equivalent form

$$\int_{t_1}^{t_2} \left(\delta L + \overline{\delta W}_{nc}\right) dt = 0, \ \delta q_k(t_1) = \delta q_k(t_2) = 0, \ k = 1, 2, \ldots, n \tag{2.124}$$

In the special case in which there are no nonconservative forces, so that $\overline{\delta W}_{nc} = 0$, Eq. (2.124) reduces to the *Hamilton's principle for conservative systems*

$$\int_{t_1}^{t_2} \delta L \, dt = 0, \ \delta q_k(t_1) = \delta q_k(t_2) = 0, \ k = 1, 2, \ldots, n \tag{2.125}$$

A system for which the constraint equations represent relations between coordinates alone, such as Eqs. (2.92), is known as *holonomic*. For holonomic systems the varied path is a possible path. If the constraint equations involve velocities, and the equations cannot be integrated to yield relations between coordinates alone, the system is said to be *nonholonomic*. For nonholonomic systems the varied path is in general not a possible path. In the case of holonomic systems, the variation and integration processes are interchangeable, so that Eq. (2.125) can be replaced by

$$\delta I = \delta \int_{t_1}^{t_2} L \, dt = 0, \ \delta q_k(t_1) = \delta q_k(t_2) = 0, \ k = 1, 2, \ldots, n \tag{2.126}$$

where

$$I = \int_{t_1}^{t_2} L \, dt \tag{2.127}$$

Equation (2.126) represents *Hamilton's principle* in its most familiar form. The principle can be stated as follows: *The actual path in the configuration space renders the value of the definite integral $I = \int_{t_1}^{t_2} L\,dt$ stationary with respect to all arbitrary variations of the path between two instants t_1 and t_2, provided that the path variations vanish at these two instants.*

Although Eq. (2.126) permits a nice mathematical interpretation of Hamilton's principle, in deriving the equations of motion for conservative systems we actually make use of Eq. (2.125).

Example 2.6

The system shown in Fig. 2.19 consists of a mass M connected to a spring of stiffness k and a uniform link of mass m and length L hinged to M at the upper end and subjected to a horizontal force at the lower end. Derive the equations of motion by means of the extended Hamilton's principle.

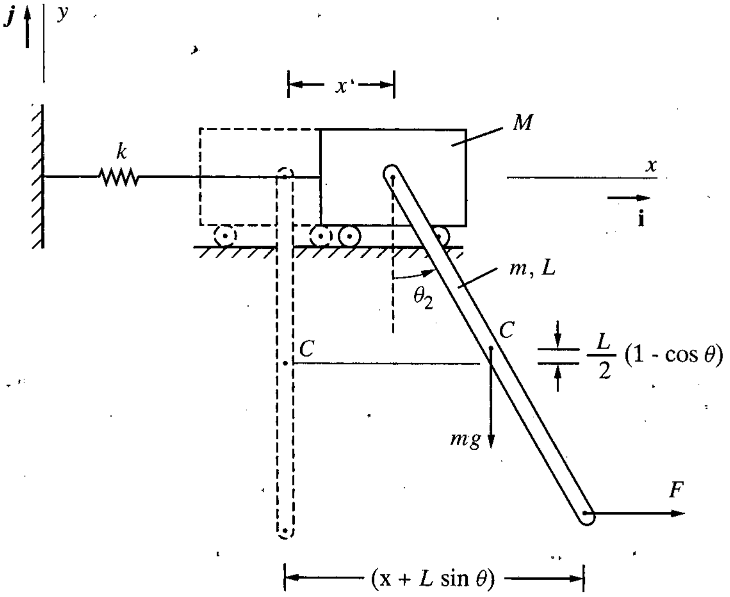


Figure 2.19 System consisting of a mass and a link

From Fig. 2.19, we conclude that the motion can be described fully by means of the translation x of the mass M and the rotation θ of the link. Hence, we use as generalized coordinates

$$q_1 = x, \quad q_2 = \theta \tag{a}$$

Before we write the kinetic energy expression, we wish to derive the velocity of the mass center C of the link. The position vector of point C is

$$\mathbf{r}_C = \left(x + \frac{L}{2}\sin\theta\right)\mathbf{i} - \frac{L}{2}\cos\theta\mathbf{j} \tag{b}$$

so that the velocity vector is simply

$$\mathbf{v}_C = \left(\dot{x} + \frac{L}{2}\dot{\theta}\cos\theta\right)\mathbf{i} + \frac{L}{2}\dot{\theta}\sin\theta\mathbf{j} \tag{c}$$

Hence, the kinetic energy has the expression

$$\begin{aligned} T &= \frac{1}{2}M\dot{x}^2 + \frac{1}{2}m\mathbf{v}_C \cdot \mathbf{v}_C + \frac{1}{2}I_C\dot{\theta}^2 \\ &= \frac{1}{2}M\dot{x}^2 + \frac{1}{2}m\left[\left(\dot{x} + \frac{L}{2}\dot{\theta}\cos\theta\right)^2 + \left(\frac{L}{2}\dot{\theta}\sin\theta\right)^2\right] + \frac{1}{2}\frac{mL^2}{12}\dot{\theta}^2 \\ &= \frac{1}{2}\left[(M+m)\dot{x}^2 + mL\dot{x}\dot{\theta}\cos\theta + \frac{1}{3}mL^2\dot{\theta}^2\right] \end{aligned} \tag{d}$$

The potential energy is the sum of the elastic potential energy of the spring and the gravitational potential energy of the link, or

$$V = \frac{1}{2}kx^2 + mg\frac{L}{2}(1 - \cos\theta) \tag{e}$$

so that the Lagrangian is simply

$$\begin{aligned} L &= T - V \\ &= \frac{1}{2}\left[(M+m)\dot{x}^2 + mL\dot{x}\dot{\theta}\cos\theta + \frac{1}{3}mL^2\dot{\theta}^2\right] - \frac{1}{2}kx^2 - mg\frac{L}{2}(1 - \cos\theta) \end{aligned} \tag{f}$$

Hence, the variation in the Lagrangian has the form

$$\begin{aligned} \delta L &= (M+m)\dot{x}\,\delta\dot{x} + \frac{1}{2}mL(\dot{\theta}\cos\theta\,\delta\dot{x} + \dot{x}\cos\theta\,\delta\dot{\theta} - \dot{x}\dot{\theta}\sin\theta\,\delta\theta) \\ &\quad + \frac{1}{3}mL^2\dot{\theta}\,\delta\dot{\theta} - kx\,\delta x - mg\frac{L}{2}\sin\theta\,\delta\theta \\ &= \left[(M+m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta\right]\delta\dot{x} \\ &\quad + \frac{1}{6}mL(3\dot{x}\cos\theta + 2L\dot{\theta})\,\delta\dot{\theta} - kx\,\delta x - \frac{1}{2}mL(\dot{x}\dot{\theta} + g)\sin\theta\,\delta\theta \end{aligned} \tag{g}$$

Moreover, the nonconservative virtual work of the horizontal force F is simply

$$\overline{\delta W}_{nc} = F\,\delta(x + L\sin\theta) = F\,\delta x + FL\cos\theta\,\delta\theta \tag{h}$$

Inserting Eqs. (g) and (h) into Eq. (2.124), we obtain

$$\begin{aligned} &\int_{t_1}^{t_2}(\delta L + \overline{\delta W}_{nc})\,dt \\ &\quad = \int_{t_1}^{t_2}\left\{\left[(M+m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta\right]\delta\dot{x} + \frac{1}{6}mL(3\dot{x}\cos\theta + 2L\dot{\theta})\,\delta\dot{\theta}\right. \\ &\qquad \left. - kx\,\delta x - \frac{1}{2}mL(\dot{x}\dot{\theta} + g)\sin\theta\,\delta\theta + F\,\delta x + FL\cos\theta\,\delta\theta\right\}dt = 0, \\ &\qquad \delta x = 0,\ \delta\theta = 0 \text{ at } t = t_1, t_2 \end{aligned} \tag{i}$$

The integrand in Eq. (i) contains $\delta\dot{x}$ and $\delta\dot{\theta}$. Before we can obtain the equations of motion by means of the extended Hamilton's principle, Eq. (i), it is necessary to carry out integrations by parts of the type

$$\begin{aligned}\int_{t_1}^{t_2} f_k(\mathbf{q}, \dot{\mathbf{q}})\, \delta\dot{q}_k\, dt &= \int_{t_1}^{t_2} f_k(\mathbf{q}, \dot{\mathbf{q}}) \frac{d}{dt}\delta q_k\, dt \\ &= f_k(\mathbf{q}, \dot{\mathbf{q}})\, \delta q_k \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{df_k(\mathbf{q}, \dot{\mathbf{q}})}{dt}\delta q_k\, dt \\ &= -\int_{t_1}^{t_2} \frac{df_k(\mathbf{q}, \dot{\mathbf{q}})}{dt}\delta q_k\, dt \end{aligned} \tag{j}$$

where we took into account the fact that $\delta q_k(t_1) = \delta q_k(t_2) = 0$. Hence, considering Eq. (j), Eq. (i) can be rewritten as

$$\begin{aligned}&\int_{t_1}^{t_2} \left\langle \left\{ -\frac{d}{dt}\left[(M + m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta\right] - kx + F \right\} \delta x \right. \\ &\quad \left. + \left[-\frac{1}{6}mL\frac{d}{dt}(3\dot{x}\cos\theta + 2L\dot{\theta}) - \frac{1}{2}mL(\dot{x}\dot{\theta} + g)\sin\theta + FL\cos\theta\right]\delta\theta \right\rangle dt \\ &= 0 \end{aligned} \tag{k}$$

But the generalized coordinates x and θ are independent, so that the virtual displacements δx and $\delta\theta$ are entirely arbitrary. It follows that the integral can be zero for all δx and $\delta\theta$ if and only if the coefficients of δx and $\delta\theta$ are identically zero, which yields the equations of motion

$$\begin{aligned}\frac{d}{dt}\left[(M + m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta\right] + kx &= F \\ \frac{1}{6}mL\frac{d}{dt}(3\dot{x}\cos\theta + 2L\dot{\theta}) + \frac{1}{2}mL(\dot{x}\dot{\theta} + g)\sin\theta &= FL\cos\theta \end{aligned} \tag{l}$$

We should observe at this point that the right side of Eqs. (l) represents the generalized nonconservative forces

$$Q_{1nc} = X = F, \quad Q_{2nc} = \Theta = FL\cos\theta \tag{m}$$

and we note that Q_{2nc} is really a moment, which is consistent with the fact that q_2 is an angle. Indeed, the product $Q_{2nc}\,\delta q_2$ represents an increment of work.

2.11 LAGRANGE'S EQUATIONS OF MOTION

Lagrange's equations occupy a special place in analytical mechanics. They represent equations of motion in terms of generalized coordinates and can be obtained solely from two scalar expressions, the kinetic energy and the virtual work, a feature shared with Hamilton's principle. There are several ways in which Lagrange's equations can be derived, directly from the generalized principle of d'Alembert, or by means of Hamilton's principle. We choose the latter approach.

In deriving the equations of motion by means of Hamilton's principle, there are two steps that must be carried out repeatedly, namely, eliminating the generalized virtual velocities from the formulation through integrations by parts, thus obtaining an integral in terms of the generalized virtual displacements alone, and then invoking the arbitrariness of the generalized virtual displacements to obtain the equations of motion by setting the coefficients of the generalized virtual displacements equal to zero. Lagrange's equations can be derived in a natural manner by carrying out the two steps indicated above for a generic dynamical system, instead of deriving them for every specific example.

The kinetic energy for a system of particles can be expressed in the general form

$$T = T(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N, \dot{\mathbf{r}}_1, \dot{\mathbf{r}}_2, \ldots, \dot{\mathbf{r}}_N) \tag{2.128}$$

where $\mathbf{r}_i$ is the displacement vector and $\dot{\mathbf{r}}_i$ the velocity vector of a typical particle of mass m_i $(i = 1, 2, \ldots, N)$. Our interest, however, is in a formulation in terms of the generalized coordinates q_k and generalized velocities $\dot{q}_k$ $(k = 1, 2, \ldots, n)$ and not $\mathbf{r}_i$ and $\dot{\mathbf{r}}_i (i = 1, 2, \ldots, N)$. We recall that the relation between $\mathbf{r}_i$ and q_k is given by Eqs. (2.105). Moreover, using the analogy with Eqs. (2.106), we can write

$$\dot{\mathbf{r}}_i = \sum_{k=1}^{n} \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k \, , \; i = 1, 2, \ldots, N \tag{2.129}$$

Introducing Eqs. (2.105) and (2.129) into Eq. (2.128), we can express the kinetic energy in terms of generalized displacements and velocities as follows:

$$T = T(q_1, q_2, \ldots, q_n, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n) \tag{2.130}$$

Hence, the variation in the kinetic energy is simply

$$\delta T = \sum_{k=1}^{n} \left(\frac{\partial T}{\partial q_k} \delta q_k + \frac{\partial T}{\partial \dot{q}_k} \delta \dot{q}_k \right) \tag{2.131}$$

Moreover, from Eq. (2.107), the virtual work performed by the applied forces can be written in terms of generalized forces and virtual displacements in the form

$$\overline{\delta W} = \sum_{k=1}^{n} Q_k \delta q_k \tag{2.132}$$

where the generalized forces Q_k $(k = 1, 2, \ldots, n)$ are as given by Eqs. (2.108). Introducing Eqs. (2.131) and (2.132) into the extended Hamilton's principle, Eq. (2.120), we can write

$$\int_{t_1}^{t_2} (\delta T + \overline{\delta W})\, dt = \int_{t_1}^{t_2} \sum_{k=1}^{n} \left[\frac{\partial T}{\partial \dot{q}_k} \delta \dot{q}_k + \left(\frac{\partial T}{\partial q_k} + Q_k \right) \delta q_k \right] dt = 0,$$

$$\delta q_k (t_1) = \delta q_k, (t_2) = 0, \; k = 1, 2, \ldots, n \tag{2.133}$$

The terms $\delta \dot{q}_k$ stand in the way of the derivation of the equations of motion. To eliminate them, we carry out an integration by parts, consider the end conditions and obtain

$$\int_{t_1}^{t_2} \frac{\partial T}{\partial \dot{q}_k}\delta \dot{q}_k \, dt = \int_{t_1}^{t_2} \frac{\partial T}{\partial \dot{q}_k}\frac{d}{dt}\delta q_k \, dt = \left.\frac{\partial T}{\partial \dot{q}_k}\delta q_k\right|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right)\delta q_k \, dt$$

$$= - \int_{t_1}^{t_2} \frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right)\delta q_k \, dt, \; k = 1, 2, \ldots, n \qquad (2.134)$$

Introducing Eqs. (2.134) into Eq. (2.133), we have

$$\int_{t_1}^{t_2} \sum_{k=1}^{n} \left[-\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right) + \frac{\partial T}{\partial q_k} + Q_k \right] \delta q_k \, dt = 0 \qquad (2.135)$$

Then, invoking the arbitrariness of the virtual displacements, we conclude that Eq. (2.135) is satisfied for all δq_k provided

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right) - \frac{\partial T}{\partial q_k} = Q_k, \; k = 1, 2, \ldots, n \qquad (2.136)$$

Equations (2.136) represent the celebrated *Lagrange's equations of motion* in their most general form, and we note that Q_k include both conservative and nonconservative generalized forces.

It is common practice to distinguish between conservative and nonconservative forces, or

$$Q_k = Q_{kc} + Q_{knc}, \; k = 1, 2, \ldots, n \qquad (2.137)$$

But, using Eq. (2.122) and recalling that the potential energy depends on coordinates alone, we can write

$$\delta W_c = -\delta V = -\sum_{k=1}^{n} \frac{\partial V}{\partial q_k}\delta q_k = \sum_{k=1}^{n} Q_{kc}\,\delta q_k \qquad (2.138)$$

so that the conservative generalized forces have the form

$$Q_{kc} = -\frac{\partial V}{\partial q_k}, \; k = 1, 2, \ldots, n \qquad (2.139)$$

Hence, introducing Eqs. (2.137) and (2.139) into Eq. (2.136), we obtain

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{q}_k}\right) - \frac{\partial T}{\partial q_k} + \frac{\partial V}{\partial q_k} = Q_{knc}, \; k = 1, 2, \ldots, n \qquad (2.140)$$

Finally, because the potential energy does not depend on velocities, Eqs. (2.140) can be rewritten as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = Q_{knc}, \; k = 1, 2, \ldots, n \qquad (2.141)$$

where $L = T - V$ is the Lagrangian.

Lagrange's equations can be used for any discrete system whose motion lends itself to a description in terms of generalized coordinates, which includes rigid bodies, in the same manner as Hamilton's principle can. They can be extended to distributed-parameter systems, but for such systems they are not as versatile as the extended Hamilton's principle, as we shall see in Chapter 7.

Example 2.7

Derive the equations of motion for the system of Example 2.6 by means of Lagrange's equations.

Letting $q_1 = x, q_2 = \theta$, $Q_{1nc} = X$ and $Q_{2nc} = \Theta$ in Eqs. (2.141), Lagrange's equations for this example take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = X$$
$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = \Theta \tag{a}$$

From Example 2.6, we obtain the Lagrangian

$$L = \frac{1}{2}\left[(M+m)\dot{x}^2 + mL\dot{x}\dot{\theta}\cos\theta + \frac{1}{3}mL^2\dot{\theta}^2\right] - \frac{1}{2}kx^2 - mg\frac{L}{2}(1-\cos\theta) \tag{b}$$

so that

$$\frac{\partial L}{\partial \dot{x}} = (M+m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta$$
$$\frac{\partial L}{\partial x} = -kx$$
$$\frac{\partial L}{\partial \dot{\theta}} = \frac{1}{2}mL\dot{x}\cos\theta + \frac{1}{3}mL^2\dot{\theta} = \frac{1}{6}mL(\dot{x}\cos\theta + L\dot{\theta}) \tag{c}$$
$$\frac{\partial L}{\partial \theta} = -\frac{1}{2}mL\dot{x}\dot{\theta}\sin\theta - mgL\sin\theta$$

Moreover, the generalized nonconservative forces X and Θ can also be obtained from Example 2.6. Hence, inserting Eqs. (c) above and Eqs. (m) of Example 2.6 into Eqs. (a), we obtain the explicit Lagrange's equations of motion

$$\frac{d}{dt}\left[(M+m)\dot{x} + \frac{1}{2}mL\dot{\theta}\cos\theta\right] + kx = F$$
$$\frac{1}{6}mL\frac{d}{dt}(3\dot{x}\cos\theta + 2L\dot{\theta}) + \frac{1}{2}mL\dot{x}\dot{\theta}\sin\theta + \frac{1}{2}mgL\sin\theta = FL\cos\theta \tag{d}$$

which are identical to Eqs. (l) of Example 2.6, as is to be expected.

2.12 HAMILTON'S EQUATIONS

Lagrange's equations, Eqs. (2.141), constitute a set of n simultaneous second-order differential equations. On occasion, it is more desirable to work with first-order differential equations rather than second-order equations, particularly when the object is integration of the equations. But second-order equations can be transformed into first-order ones. To this end, it is common practice to introduce the generalized velocities as auxiliary variables and replace the n second-order differential equations by $2n$ first-order differential equations. Of the $2n$ equations, n are purely kinematical in nature, stating that the time derivative of the coordinates is equal to the velocities, and the remaining n are the original equations expressed in first-order form by replacing accelerations by time derivatives of velocities. The resulting first-order equations are known as *state equations*, encountered for the first time in Sec. 1.9. In this section, we consider a somewhat different formulation, namely, one in which the auxiliary variables are momenta instead of velocities.

The *generalized momenta* associated with the generalized coordinates

$$q_k \quad (k = 1, 2, \dots, n)$$

are defined as

$$p_k = \frac{\partial L}{\partial \dot{q}_k}, \quad k = 1, 2, \dots, n \tag{2.142}$$

where

$$L = L(q_1, q_2, \dots, q_n, \dot{q}_1, \dot{q}_2, \dots, \dot{q}_n, t) \tag{2.143}$$

is the Lagrangian. Moreover, the *Hamiltonian function* is defined as follows:

$$\mathcal{H} = \sum_{k=1}^{n} \frac{\partial L}{\partial \dot{q}_k}\dot{q}_k - L = \sum_{k=1}^{n} p_k\dot{q}_k - L \tag{2.144}$$

If the generalized velocities are replaced by the generalized momenta, the Hamiltonian can be written in the general functional form

$$\mathcal{H} = \mathcal{H}(q_1, q_2, \dots, q_n, p_1, p_2, \dots, p_n, t) \tag{2.145}$$

Next, we take the variation of $\mathcal{H}$ in both Eqs. (2.144) and (2.145), consider Eqs. (2.142) and (2.143) and write

$$\begin{aligned} \delta\mathcal{H} &= \sum_{k=1}^{n}\left(p_k\delta\dot{q}_k + \dot{q}_k\delta p_k - \frac{\partial L}{\partial q_k}\delta q_k - \frac{\partial L}{\partial \dot{q}_k}\delta\dot{q}_k\right) \\ &= \sum_{k=1}^{n}\left(\dot{q}_k\delta p_k - \frac{\partial L}{\partial q_k}\delta q_k\right) = \sum_{k=1}^{n}\left(\frac{\partial\mathcal{H}}{\partial q_k}\delta q_k + \frac{\partial\mathcal{H}}{\partial p_k}\delta p_k\right) \end{aligned} \tag{2.146}$$

from which it follows that

$$\dot{q}_k = \frac{\partial\mathcal{H}}{\partial p_k}, \qquad -\frac{\partial L}{\partial q_k} = \frac{\partial\mathcal{H}}{\partial q_k}, \quad k = 1, 2, \dots, n \tag{2.147a, b}$$

Using Eqs. (2.141) and (2.142), however, we can write

$$\dot{p}_k = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) = \frac{\partial L}{\partial q_k} + Q_{knc}, \quad k = 1, 2, \dots, n \tag{2.148}$$

where Q_{knc} are nonconservative generalized forces. Then, inserting Eqs. (2.148) into Eqs. (2.147b), we can rewrite Eqs. (2.147) in the form

$$\dot{q}_k = \frac{\partial \mathcal{H}}{\partial p_k}, \quad k = 1, 2, \ldots, n \tag{2.149a}$$

$$\dot{p}_k = -\frac{\partial \mathcal{H}}{\partial q_k} + Q_{knc}, \quad k = 1, 2, \ldots, n \tag{2.149b}$$

Equations (2.149) represent the desired set of $2n$ first-order differential equations. They are known as *Hamilton's equations*. Note that, if we differentiate Eqs. (2.144) and (2.145) with respect to time and consider Eqs. (2.142), (2.147) and (2.149a), we conclude that

$$\frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial L}{\partial t} \tag{2.150}$$

Clearly, if the Lagrangian does not depend explicitly on time, neither does the Hamiltonian.

The Hamiltonian function and the virtual work define the motion of the system fully, as all the differential equations of motion can be derived from these two expressions. The clear advantage of Hamilton's equations over Lagrange's equations is that in Hamilton's equations the time derivative of the variables, coordinates and momenta, appear on the left side of the equations only and they are first-order derivatives, which makes Hamilton's equations suitable for numerical integration. Another advantage is that Hamilton's equations permit a geometric interpretation of the solution, as discussed in Sec. 4.2.

At this point, we wish to relate the Hamiltonian to the system kinetic and potential energy. To this end, we consider the case in which the kinetic energy of an n-degree-of-freedom system can be written in the form

$$T = T_2 + T_1 + T_0 \tag{2.151}$$

where

$$T_2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\dot{q}_i\dot{q}_j \tag{2.152}$$

is quadratic in the generalized velocities, in which $m_{ij} = m_{ji}$ are symmetric coefficients depending on the generalized coordinates, $m_{ij} = m_{ij}(q_1, q_2, \ldots, q_n)$,

$$T_1 = \sum_{j=1}^{n} f_j\dot{q}_j \tag{2.153}$$

is linear in the generalized velocities, in which f_j are coefficients depending on the generalized coordinates, $f_j = f_j(q_1, q_2, \ldots, q_n)$, and

$$T_0 = T_0(q_1, q_2, \ldots, q_n) \tag{2.154}$$

is a function of the generalized coordinates alone and contains no generalized velocities. In the case in which the kinetic energy is given by Eq. (2.151) the system is said to be *nonnatural*. It should be noted here that nonnatural systems are most commonly associated with motions relative to rotating reference frames, in which case T_1 gives rise to forces of the *Coriolis* type, or *gyroscopic forces*, and the term T_0 leads to so-called *centrifugal forces*. Introducing Eq. (2.151) into Eq. (2.123), we can write the Lagrangian for nonnatural systems as

$$L = T - V = T_2 + T_1 + T_0 - V = T_2 + T_1 - U \tag{2.155}$$

where

$$U = V - T_0 = U(q_1, q_2, \ldots, q_n) \tag{2.156}$$

represents a modified potential energy known as the *dynamic potential*. It is generally a nonlinear function of the generalized coordinates. Then, inserting Eq. (2.155) into Eq. (2.144) and considering Euler's theorem on homogeneous functions, we obtain the Hamiltonian for *nonnatural systems* in the explicit form

$$\mathcal{H} = \sum_{k=1}^{n} \frac{\partial L}{\partial \dot{q}_k}\dot{q}_k - L = 2T_2 + T_1 - (T_2 + T_1 - U) = T_2 + U \tag{2.157}$$

When $T_2 = T$, $T_1 = T_0 = 0$, the system is said to be *natural* and the Hamiltonian reduces to

$$\mathcal{H} = T + V = E \tag{2.158}$$

where E is recognized as the *total energy*, first encountered in Sec. 2.5.

Example 2.8

Derive Hamilton's equation for the spherical pendulum shown in Fig. 2.20.

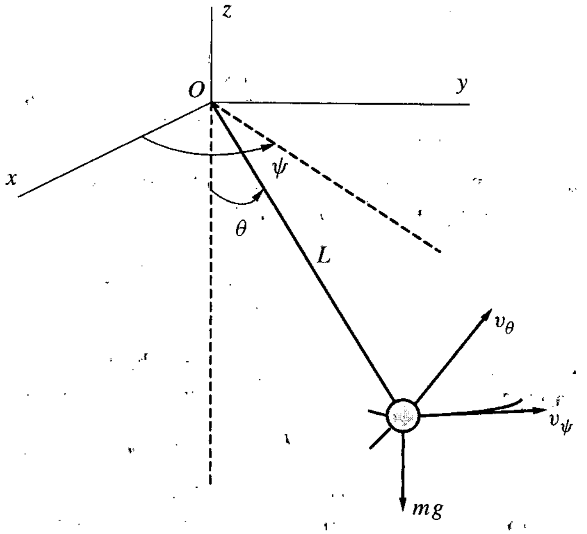


Figure 2.20 Spherical pendulum

Considering Fig. 2.20, we conclude that a suitable set of generalized coordinates for the problem at hand has the form

$$q_1 = \theta,\ q_2 = \psi \tag{a}$$

Moreover, from Fig. 2.20, we obtain the velocities

$$v_\theta = L\dot{\theta},\ v_\psi = L\dot{\psi}\sin\theta \tag{b}$$

so that the kinetic energy is simply

$$T = \frac{1}{2}m\left(v_\theta^2 + v_\psi^2\right) = \frac{1}{2}mL^2\left(\dot{\theta}^2 + \dot{\psi}^2\sin^2\theta\right) \tag{c}$$

The potential energy is due to gravity alone and has the form

$$V = mgL(1 - \cos\theta) \tag{d}$$

Hence, the Lagrangian has the expression

$$L = T - V = \frac{1}{2}mL^2\left(\dot{\theta}^2 + \dot{\psi}^2\sin^2\theta\right) - mgL(1 - \cos\theta) \tag{e}$$

Using Eqs. (2.142), we obtain the generalized momenta

$$\begin{aligned} p_\theta &= \frac{\partial L}{\partial\dot{\theta}} = mL^2\dot{\theta} \\ p_\psi &= \frac{\partial L}{\partial\dot{\psi}} = mL^2\dot{\psi}\sin^2\theta \end{aligned} \tag{f}$$

Because this is a natural system, we insert Eqs. (f) into Eq. (2.158) and write the Hamiltonian in the form

$$\mathcal{H} = T + V = \frac{1}{2mL^2}\left(p_\theta^2 + \frac{p_\psi^2}{\sin^2\theta}\right) + mgL(1 - \cos\theta) \tag{g}$$

Then, using Eqs. (2.149) and recognizing that there are no nonconservative forces present, we obtain the desired Hamilton's equations

$$\begin{aligned} \dot{\theta} &= \frac{\partial\mathcal{H}}{\partial p_\theta} = \frac{p_\theta}{mL^2} \\ \dot{\psi} &= \frac{\partial\mathcal{H}}{\partial p_\psi} = \frac{p_\psi}{mL^2\sin^2\theta} \\ \dot{p}_\theta &= -\frac{\partial\mathcal{H}}{\partial\theta} = \frac{2p_\psi^2\cos\theta}{mL^2\sin^3\theta} - mgL\sin\theta \\ \dot{p}_\psi &= -\frac{\partial\mathcal{H}}{\partial\psi} = 0 \end{aligned} \tag{h}$$

and we observe that the first two of Eqs. (h) could have been obtained more directly from Eqs. (f).

2.13 CONSERVATION LAWS

From Secs. 2.11 and 2.12, we conclude that the motion of an n-degree-of-freedom system can be described by n second-order Lagrange's equations or $2n$ first-order Hamilton's equations. A complete solution of the system of differential equations of motion requires $2n$ constants of integration, which ordinarily consist of the initial values of the n coordinates q_k and n velocities $\dot{q}_k$, or of the n coordinates q_k and n momenta p_k. Closed-form solutions to $2n$-order systems of equations can be obtained in special cases only, primarily for linear time-invariant systems (see Sec. 1.2). In more general cases, closed-form solutions prove elusive. Under given circumstances, however, certain integrals of motion can be obtained. These integrals can be used at times to lower the number of degrees of freedom. Moreover, they can provide some insights into the system behavior.

Let us consider the case in which a given generalized coordinate, say q_s, is absent from the Lagrangian, and hence from the Hamiltonian, although the generalized velocity $\dot{q}_s$ does appear in the Lagrangian. A coordinate that does not appear explicitly in the Lagrangian is said to be *ignorable*, or *cyclic*. In addition, we assume that the nonconservative force corresponding to q_s is zero, $Q_{snc} = 0$. Under these circumstances, we conclude from Eqs. (2.149b) that the system admits a *first integral of motion* having the form

$$p_s = \text{constant} \tag{2.159}$$

Hence, *the generalized momentum associated with an ignorable coordinate is conserved.* Equation (2.159) represents the *conservation of momentum principle* and can be regarded as a generalization of the more physical conservation of linear and angular momentum principles encountered in Secs. 2.1 and 2.4, respectively.

Next, we consider the case in which the Lagrangian does not depend explicitly on time, differentiate Eq. (2.144) with respect to time, use Eqs. (2.141) and obtain

$$\begin{aligned}\frac{d\mathcal{H}}{dt} &= \sum_{k=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right)\dot{q}_k + \frac{\partial L}{\partial \dot{q}_k}\ddot{q}_k\right] - \sum_{k=1}^{n}\left(\frac{\partial L}{\partial q_k}\dot{q}_k + \frac{\partial L}{\partial \dot{q}_k}\ddot{q}_k\right)\\ &= \sum_{k=1}^{n}\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k}\right]\dot{q}_k = \sum_{k=1}^{n} Q_{knc}\dot{q}_k\end{aligned} \tag{2.160}$$

But, by analogy with Eq. (2.132), the right side of Eq. (2.160) can be identified as the time rate of work of the nonconservative forces, which represents the *power* developed by the nonconservative forces. It follows that, in the case in which the Lagrangian does not depend explicitly on time, the time derivative of the Hamiltonian is equal to the power developed by the nonconservative forces.

In the case of a holonomic conservative system $Q_{knc} = 0$ $(k = 1, 2, \ldots, n)$ and Eq. (2.160) reduces to

$$\frac{d\mathcal{H}}{dt} = 0 \tag{2.161}$$

which can be integrated immediately to obtain

$$\mathcal{H} = \text{constant} \tag{2.162}$$

Equation (2.162) represents another kind of first integral of motion known as a *Jacobi integral*. The conservation principle embodied by Eq. (2.162) can be stated as follows: *In the case in which the Lagrangian does not depend explicitly on time and all the nonconservative forces are zero the Hamiltonian is conserved.*

Recalling Eq. (2.157), we conclude that *the conservation of the Hamiltonian for a nonnatural system* can be expressed in the explicit form

$$\mathcal{H} = T_2 + U = \text{constant} \tag{2.163}$$

Moreover, from Eq. (2.158), *the conservation of the Hamiltonian for a natural system* can be expressed as

$$\mathcal{H} = E = T + V = \text{constant} \tag{2.164}$$

which is recognized as the *conservation of energy principle*, first encountered in Sec. 2.5.

Example 2.9

Consider the spherical pendulum of Example 2.8 and determine whether there are any integrals of motion.

From Example 2.8, we obtain the kinetic energy

$$T = \frac{1}{2}mL^2\left(\dot{\theta}^2 + \dot{\psi}^2\sin^2\theta\right) \tag{a}$$

and the potential energy

$$V = mgL(1 - \cos\theta) \tag{b}$$

so that the Lagrangian has the expression

$$L = T - V = \frac{1}{2}mL^2\left(\dot{\theta}^2 + \dot{\psi}^2\sin^2\theta\right) - mgL(1 - \cos\theta) \tag{c}$$

and we observe that the coordinate ψ is absent. It follows that ψ is an ignorable coordinate and the associated momentum is conserved, or

$$p_\psi = \frac{\partial L}{\partial\dot{\psi}} = mL^2\dot{\psi}\sin^2\theta = \text{constant} \tag{d}$$

Moreover, there are no nonconservative forces and the Lagrangian does not depend explicitly on time, so that the Hamiltonian is conserved. This being a natural system, we insert Eqs. (a) and (b) into Eq. (2.164) and obtain a second integral of motion in the form of the total energy

$$\mathcal{H} = E = T + V = \frac{1}{2}mL^2\left(\dot{\theta}^2 + \dot{\psi}^2\sin^2\theta\right) + mgL(1 - \cos\theta) = \text{constant} \tag{e}$$

The presence of an ignorable coordinate permits a reduction in the number of degrees of freedom from two to one. Indeed, Eq. (d) can be used to write

$$\dot{\psi} = \frac{p_{\psi 0}}{mL^2\sin^2\theta} \tag{f}$$

where $p_{\psi 0}$ is the initial value of p_ψ. Then, introducing Eq. (f) into Eq. (e), we obtain

$$\frac{1}{2}mL^2\dot{\theta}^2 + \frac{1}{2}\frac{p_{\psi 0}^2}{mL^2\sin^2\theta} + mgL(1 - \cos\theta) = \text{constant} \tag{g}$$

Hence, the spherical pendulum can be regarded as an equivalent single-degree-of-freedom system characterized by the kinetic energy

$$T_{\text{equiv}} = \frac{1}{2} m L^2 \dot{\theta}^2 \tag{h}$$

and the potential energy

$$V_{\text{equiv}} = \frac{1}{2} \frac{p_{\psi 0}^2}{m L^2 \sin^2 \theta} + mgL\,(1 - \cos\theta) \tag{i}$$

and we note that, whereas the kinetic energy of the equivalent system is the same as the kinetic energy of a simple pendulum, the potential energy is not. In fact, the potential energy has been modified so as to include the centrifugal effect caused by the rotation $\dot{\psi}$ of the spherical pendulum about the z-axis.

2.14 SYNOPSIS

In Chapter 1, we introduced many concepts and techniques from linear system theory by means of generic differential equations. To study the behavior of vibrating systems, it is necessary to work with actual differential equations of motion for such systems. As the equations of motion for the variety of systems considered in this text are not readily available, it is incumbent upon us to derive them. For simple systems, characterized by a limited number of degrees of freedom, the methods of Newtonian mechanics are quite adequate. For more complex systems, such as multi-degree-of-freedom discrete systems and distributed-parameter systems, the methods of analytical mechanics are indispensable.

This chapter contains a selection of topics from Newtonian mechanics and analytical mechanics essential to a serious study of vibrations. The principles and techniques presented in this chapter are useful not only in the derivation of the equations of motion for multi-degree-of-freedom systems but also for testing the stability of such systems, as demonstrated in Chapter 4. The versatility of analytical mechanics is not confined to discrete systems alone. Indeed, this versatility is amply demonstrated in Chapter 7, in which methods of analytical mechanics are used to derive boundary-value problems associated with distributed-parameter systems. In this regard, the extended Hamilton's principle proves without equal, producing results where other methods fail.

This chapter is undoubtedly the most fundamental in the entire book. It is also the most satisfying.

PROBLEMS

2.1 The system shown in Fig. 2.21 consists of a smooth massless hoop of radius R rotating with the constant angular velocity Ω about a vertical axis. A bead of mass m can slide freely around the hoop. Use Newton's second law to derive the equation of motion.

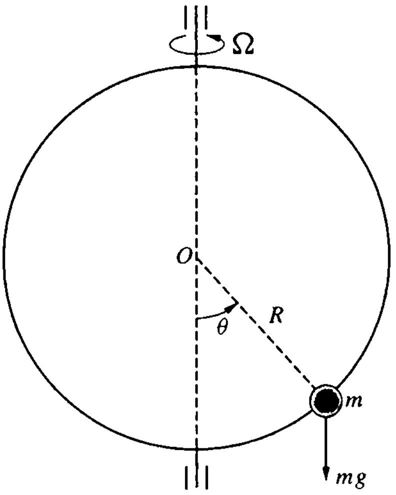


Figure 2.21 Mass sliding on a rotating hoop

2.2 Two masses, m_1 and m_2, suspended on a massless string, are vibrating in a vertical plane, as shown in Fig. 2.22. The displacements are sufficiently small that the slope of the string at any point remains small and the tension T in the string remains constant at all times. Derive the equations of motion by means of Newton's second law.

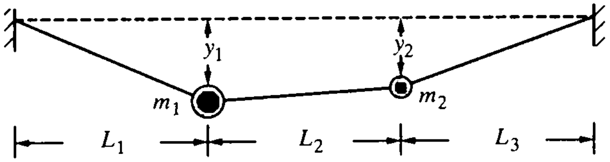


Figure 2.22 Two masses on a string

2.3 A system consisting of a uniform rigid bar of mass m and length L, and with both ends mounted on rollers, is constrained to move as shown in Fig. 2.23. The left and right ends of the bar are attached to linear springs of stiffnesses k_1 and k_2, respectively. When the bar is horizontal the springs are unstretched. Use the Newtonian approach of Sec. 2.5 to derive two equations for the translation of the mass center C and one equation for the rotation about C. The angle θ can be arbitrarily large.

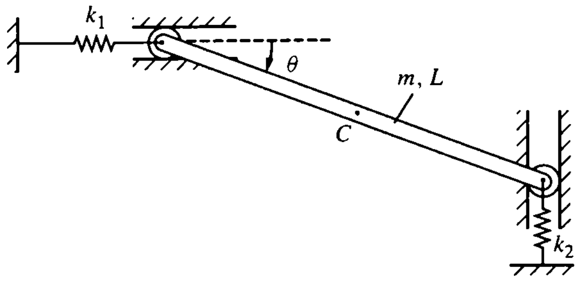


Figure 2.23 Rigid bar in constrained motion

2.4 Derive the equations of motion for the system of Example 2.6 by means of the Newtonian approach. Then eliminate the constraint forces between the two bodies so as to obtain two equations for x and θ, thus verifying the results of Example 2.6.

2.5 Derive the equations of motion for the double pendulum shown in Fig. 2.16 by means of Newton's second law.

2.6 Derive the equilibrium equations of the system of Problem 2.2 by means of the virtual work principle.

2.7 Derive the equilibrium equation for the system of Problem 2.3 by means of the virtual work principle.

2.8 The system shown in Fig. 2.24 consists of two uniform rigid links of mass m and length L, a massless roller free to move horizontally and a linear spring of stiffness k. The links are hinged at both ends and when they are horizontal the spring is unstretched. Derive the equilibrium equation by means of the virtual work principle. The angle θ can be arbitrarily large.

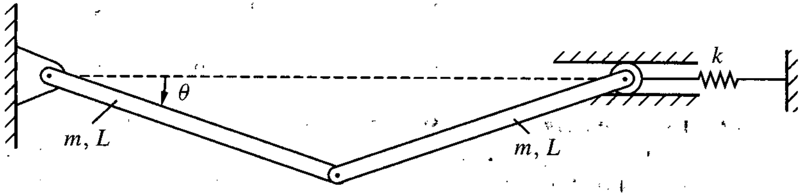


Figure 2.24 System consisting of two links

2.9 The system shown in Fig. 2.25 is similar to that of Problem 2.7, except that the spring is replaced by the mass M and two springs of stiffness k_1 and k_2. Derive the equilibrium equations by means of the virtual work principle. The angle θ can be arbitrarily large.

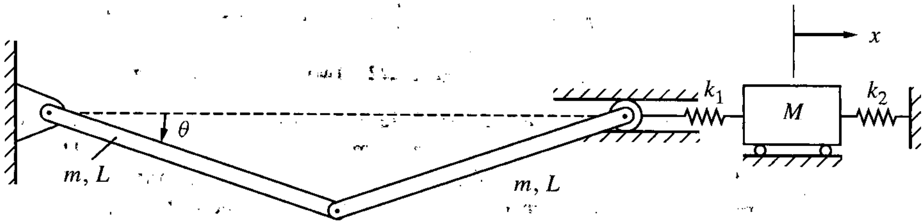


Figure 2.25 System consisting of two links and a mass

2.10 Derive the equations of motion for the double pendulum shown in Fig. 2.16 by means of the generalized principle of d'Alembert.

2.11 Derive the equation of motion for the system of Problem 2.8 by means of the generalized d'Alembert's principle. Hint: Treat the links as continuous by the approach of Sec. 2.6, whereby summation over systems of particles is replaced by integration over rigid bodies.

2.12 Derive the equation of motion for the system of Problem 2.1 by means of Hamilton's principle.

2.13 Derive the equation of motion for the system of Problem 2.3 by means of Hamilton's principle.

2.14 Derive the equation of motion for the system of Problem 2.8 by means of Hamilton's principle.

2.15 The system of Fig. 2.26 consists of a uniform rigid link of mass m and length L hinged at the upper end to a linear spring of stiffness k. Use Hamilton's principle to derive the equations of motion.

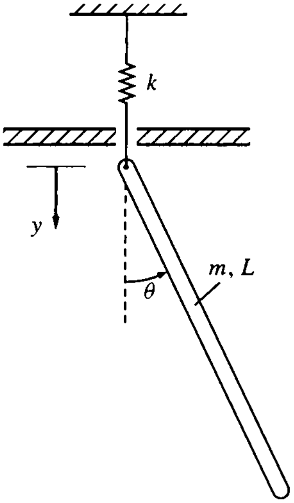


Figure 2.26 Link suspended on a spring

2.16 Derive the equations of motion for the system of Problem 2.9 by means of Hamilton's principle.

2.17 Derive the equations of motion for the system of Problem 2.5 by means of Hamilton's principle.

2.18 The system of Example 2.3 is acted upon at the lower end of the bar by a horizontal force F. Derive the equations of motion by means of the extended Hamilton's principle.

2.19 Derive the Lagrange equation of motion for the system of Problem 2.1.

2.20 Derive the Lagrange equations of motion for the system of Problem 2.2.

2.21 Derive the Lagrange equation of motion for the system of Problem 2.3.

2.22 Derive the Lagrange equation of motion for the system of Problem 2.8.

2.23 Derive the Lagrange equations of motion for the system of Problem 2.9.

2.24 Derive the Lagrange equations of motion for the system of Problem 2.15.

2.25 Derive the Lagrange equations of motion for the system of Problem 2.5.

2.26 Derive the Lagrange equations of motion for the system of Problem 2.18.

2.27 Derive Hamilton's equations for the system of Problem 2.1.

2.28 Derive Hamilton's equations for the system of Problem 2.2.

2.29 Derive Hamilton's equations for the system of Problem 2.8.

2.30 Derive Hamilton's equations for the system of Problem 2.9.

2.31 Derive Hamilton's equations for the system of Problem 2.15.

2.32 Derive Hamilton's equations for the system of Problem 2.18.

2.33 Consider the case in which the Lagrangian does not depend explicitly on time and prove Eq. (2.160) beginning with Eq. (2.145).

2.34 Determine whether the system of Problem 2.1 possesses any integrals of motion.

2.35 The system shown in Fig. 2.27 is similar to that in Fig. 2.21, except that the hoop is acted upon by the torque $M(t)$ about the vertical axis and the angular velocity about the vertical axis is no longer constant. Determine whether there are any integrals of motion.

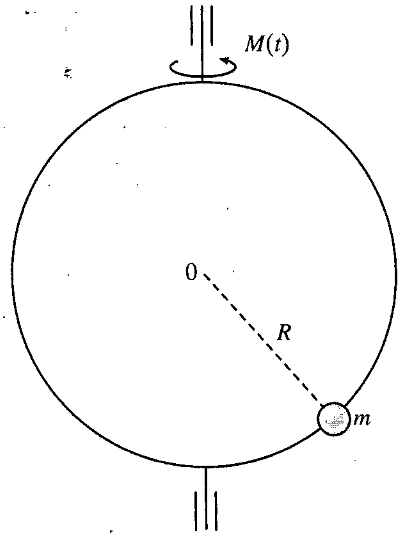


Figure 2.27 Mass sliding on a hoop rotating under the action of a torque

2.36 Determine whether the system of Problem 2.18 possesses any integrals of motion.

BIBLIOGRAPHY

1. Goldstein, H., *Classical Mechanics*, 2nd ed., Addison-Wesley, Reading, MA, 1980.
2. Lanczos, C., *The Variational Principles of Mechanics*, 4th ed., Dover, New York, 1986.
3. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan, New York, 1967.
4. Meirovitch, L., *Methods of Analytical Dynamics*, McGraw-Hill, New York, 1970.
5. Meirovitch, L., *Introduction to Dynamics and Control*, Wiley, New York, 1985.
6. Pars, C. A., *A Treatise on Analytical Dynamics*, Heinemann, London, 1979.
7. Whittaker, E. T., *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge University Press, London, 1937.

3

SINGLE-DEGREE-OF-FREEDOM SYSTEMS

In Chapter 1, we introduced a variety of concepts from linear system theory regarding the excitation-response, or input-output relation. For the most part, the discussion was concerned with linear time-invariant systems, also known as systems with constant coefficients. Such systems are of considerable importance in vibrations, as most vibrating systems can be modeled as linear time-invariant systems.

This chapter marks the real beginning of our study of vibrations. Vibrating mechanical systems represent a subclass of the all-encompassing class of dynamical systems characterized by the presence of restoring forces. These forces can arise from a variety of sources, but our interest lies primarily in restoring forces due to elasticity, i.e., forces caused by the tendency of elastic systems to return to the original undeformed state when disturbed.

The dynamic behavior of mechanical systems is governed by Newton's second law. In Sec. 2.3, we derived equations of motion for simple dynamical systems, such as first-order and second-order systems. Our interest lies primarily in second-order systems, more commonly known as single-degree-of-freedom systems. Before solutions to the differential equation of motion can be produced, it is necessary to specify the nature of the excitation. There are basically two types of excitations, steady-state and transient. Harmonic and periodic excitations fall in the first category and initial and nonperiodic excitations fall in the second. The response to harmonic and periodic excitations is conveniently derived by means of frequency-domain techniques. On the other hand, the response to transient excitations is more conveniently obtained by time-domain methods. Among these, the Laplace transformation proves well suited for linear time-invariant systems, as it is capable to produce the response to both initial and nonperiodic excitations at the same time. The latter has the form of a convolution integral. The response can also be obtained by a state-space technique involving a vector form of the convolution integral using the transition matrix. If the transient response is to be evaluated numerically on a digital computer, then discrete-time techniques are particularly effective. The convolution integral in continuous time is replaced by the convolution sum in discrete time. Moreover, the state space approach can be extended to discrete time by replacing the vector form of the convolution integral by some recursive equations using the discrete time transition matrix. All these subjects are discussed in this chapter.

3.1 RESPONSE TO INITIAL EXCITATIONS

In Sec. 1.4, we derived the response of linear time-invariant systems to initial excitations by first assuming an exponential solution and then adjusting the solution so as to match the initial conditions. Then, in Sec. 1.6, we indicated that the response to initial excitations can also be obtained by the Laplace transformation method. This latter approach is particularly suited for linear time-invariant systems, so that we propose to use it here.

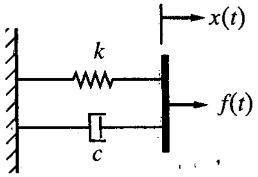


Figure 3.1 Damper-spring system

Figure 3.1 shows the simplest of mechanical systems, namely, a *damper-spring system*. The behavior of this system was shown in Sec. 2.3 to be governed by the single first-order differential equation

$$c\dot{x}(t) + kx(t) = f(t) \tag{3.1}$$

where $x(t)$ is the displacement, c the coefficient of viscous damping, k the spring constant and $f(t)$ the external excitation. In the absence of external excitations, $f(t) = 0$, and after dividing through by c, Eq. (3.1) reduces to

$$\dot{x}(t) + ax(t) = 0, \qquad a = k/c \tag{3.2}$$

This being a first-order system , the solution of Eq. (3.2) is subject to a single initial condition, namely,

$$x(0) = x_0 \tag{3.3}$$

where x_0 represents the *initial displacement.* Before we proceed with the solution of Eq. (3.2), we note from Appendix A that the Laplace transform of time derivatives of functions are given by

$$\mathcal{L}\frac{d^r x(t)}{dt^r} = s^r X(s) - s^{r-1}x(0) - s^{r-2}\frac{dx(t)}{dt}\bigg|_{t=0} - \dots$$

$$- s\frac{d^{r-2}x(t)}{dt^{r-2}}\bigg|_{t=0} - \frac{d^{r-1}x(t)}{dt^{r-1}}\bigg|_{t=0}, \qquad r = 1, 2, \dots, n \tag{3.4}$$

where $X(s)$ is the Laplace transform of $x(t)$, and note that Eq. (3.4) is a generalization of Eq. (1.54), in the sense that now the terms due to the initial conditions are included. Laplace transforming both sides of Eq. (3.2) and using Eq. (3.4), we obtain

$$sX(s) - x(0) + aX(s) = 0 \tag{3.5}$$

Then, considering the initial condition, Eq. (3.3), and rearranging, we have

$$X(s) = \frac{x_0}{s + a} \tag{3.6}$$

Finally, from the table of Laplace transforms in Appendix A, we obtain the response to the initial displacement x_0 in the form of the inverse Laplace transformation

$$x(t) = \mathcal{L}^{-1}\, X(s) = x_0 e^{-at} \tag{3.7}$$

Quite frequently, the response of a first-order system is expressed in the form

$$x(t) = x_0 e^{-t/\tau} \tag{3.8}$$

where

$$\tau = 1/a = c/k \tag{3.9}$$

is the time constant of the system, first introduced in Example 1.2.

Equation (3.8) indicates that the response of a first-order system to an initial displacement decays exponentially with time, with the decay rate depending on the time constant τ. Figure 3.2 presents several plots of $x(t)$ versus t, with τ playing the role of a parameter. It is easy to see that the rate of decay decreases as the time constant increases.

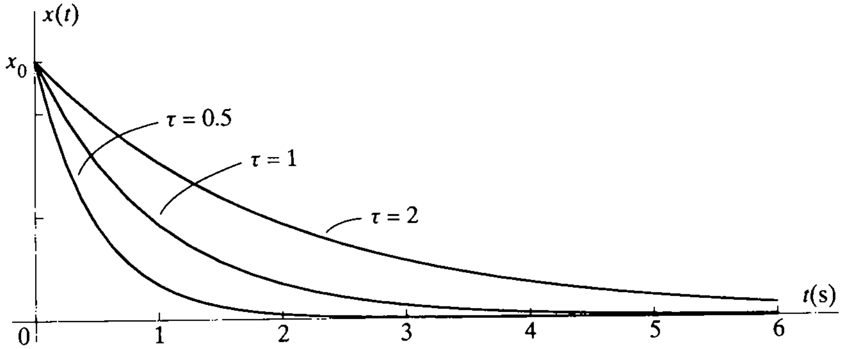


Figure 3.2 Free response of a damper-spring system with the time constant as a parameter

Next we turn our attention to the response of a *mass-damper-spring system* to initial excitations. Such a system is shown in Fig. 3.3 and it represents a second-order system, commonly referred to as a *single-degree-of-freedom system*. Its differential equation of motion was derived in Sec. 2.3 in the form

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \tag{3.10}$$

where m is the mass. The remaining quantities are as defined for the first-order system discussed above. The simplest second-order system is the *undamped system*; it represents a mathematical idealization seldom encountered in practice. Quite often, however, damping is sufficiently small that it can be ignored for all practical purposes. The undamped single-degree-of-freedom system occupies a very important place in vibrations, so that a separate discussion is fully justified.

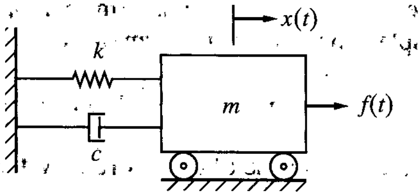


Figure 3.3 Damped single-degree-of-freedom system

Letting $c = 0$ and $f(t) = 0$ in Eq. (3.10) and dividing through by m, we can write the differential equation for the *free vibration* of a typical undamped single-degree-of-freedom system in the form

$$\ddot{x}(t) + \omega_n^2 x(t) = 0 \tag{3.11}$$

where

$$\omega_n = \sqrt{k/m} \tag{3.12}$$

is known as the *natural frequency*. Its units are radians per second (rad/s). Because Eq. (3.11) represents a second-order system, the solution $x(t)$ is subject to two initial conditions, namely,

$$x(0) = x_0, \qquad \dot{x}(t) = v_0 \tag{3.13}$$

where x_0 and v_0 are the *initial displacement and velocity*, respectively.

We propose to solve Eq. (3.11) by the Laplace transformation. Hence, using Eq. (3.4), the Laplace transform of Eq. (3.11) can be written in the form

$$s^2 X(s) - sx(0) - \dot{x}(0) + \omega_n^2 X(s) = 0 \tag{3.14}$$

Then, considering the initial conditions, Eqs. (3.13), and solving for $X(s)$, we obtain

$$X(s) = \frac{s}{s^2 + \omega_n^2} x_0 + \frac{1}{s^2 + \omega_n^2} v_0 \tag{3.15}$$

Both functions of s on the right side of Eq. (3.15) can be found in the table of Laplace transforms in Appendix A, which permits us to obtain the response in the form of the inverse Laplace transform

$$x(t) = \mathcal{L}^{-1}X(s) = x_0 \cos\omega_n t + \frac{v_0}{\omega_n}\sin\omega_n t \tag{3.16}$$

It is customary to express the response in the form

$$x(t) = A\cos(\omega_n t - \phi) \tag{3.17}$$

where

$$A = \sqrt{x_0^2 + (v_0/\omega_n)^2} \tag{3.18}$$

is known as the *amplitude* and

$$\phi = \tan^{-1}\frac{v_0}{x_0\omega_n} \tag{3.19}$$

is called the *phase angle*. Figure 3.4 shows a plot of $x(t)$ versus t for given natural frequency and initial conditions.

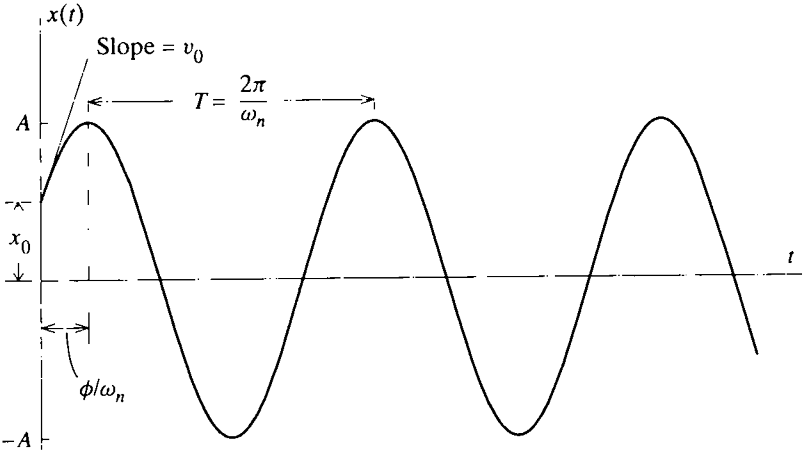


Figure 3.4 Free response of an undamped single-degree-of-freedom system

Equation (3.17) states that the system executes *simple harmonic oscillation* with the frequency ω_n. For this reason, Eq. (3.11) is said to represent a *harmonic oscillator*. Regardless of how the motion is initiated, in the absence of external excitations, an undamped single-degree-of-freedom system always oscillates at the same frequency, which explains why ω_n is called the natural frequency. The initial excitations affect only the amplitude and phase angle. All these quantities are displayed in Fig. 3.4, and we note that the effect of the phase angle is to shift the curve $A\cos\omega_n t$ to the right by an amount of time equal to ϕ/ω_n. Also from Fig. 3.4, we can identify the *period* of oscillation

$$T = \frac{2\pi}{\omega_n} \tag{3.20}$$

defined as the amount of time between two consecutive points on the curve having equal displacement and velocity both in magnitude and sign, such as the time between two consecutive peaks. The period has units of seconds (s). There exists another definition of the natural frequency, namely,

$$f_n = \frac{1}{2\pi}\omega_n = \frac{1}{T} \tag{3.21}$$

and we note that f_n is measured in cycles per second, where one cycle per second is commonly known as one hertz (Hz).

The amplitude A is a well-defined quantity. It represents the maximum value of the response. On the other hand, the phase angle ϕ does not have the same degree of physical meaning. It can be interpreted as a measure of the time necessary for the response to reach the maximum value. When the initial velocity is zero, the phase angle is zero, so that the response has the maximum value at $t = 0$, as well as at $t_i = iT$ $(i = 1, 2, \ldots)$; it has the minimum value at $t_i = (2i - 1)/2$ $(i = 1, 2, \ldots)$.

A surprisingly large number of systems, from a variety of areas, behave like harmonic oscillators. Many of these systems do so only when confined to small motions about an equilibrium position. As an example, the simple pendulum of Example 1.1 behaves as a harmonic oscillator with the natural frequency $\omega_n = \sqrt{g/L}$ only in the neighborhood of $\theta = 0$. This neighborhood covers the region in which $\sin\theta \cong \theta$, which is approximately true for values of θ reaching 30° and beyond, depending on the accuracy required. For larger amplitudes, the motion of the pendulum is periodic but not harmonic.

At this point, we consider the response of *damped single-degree-of-freedom systems* to initial excitations. To this end, we let $f(t) = 0$ in Eq. (3.10) and divide through by m to obtain the differential equation

$$\ddot{x}(t) + 2\zeta\omega_n\dot{x}(t) + \omega_n^2 x(t) = 0 \tag{3.22}$$

where

$$\zeta = \frac{c}{2m\omega_n} \tag{3.23}$$

is known as the *viscous damping factor*. The solution $x(t)$ is subject to the initial conditions given by Eqs. (3.13). Laplace transforming Eq. (3.22) and using Eqs. (3.4), we have

$$s^2X(s) - sx(0) - \dot{x}(0) + 2\zeta\omega_n\left[sX(s) - x(0)\right] + \omega_n^2 X(s) = 0 \tag{3.24}$$

Solving for $X(s)$ and using Eqs. (3.13), we obtain

$$X(s) = \frac{s + 2\zeta\omega_n}{s^2 + 2\zeta\omega_n s + \omega_n^2}x_0 + \frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}v_0 \tag{3.25}$$

The response of the damped single-degree-of-freedom system is given by the inverse Laplace transform of Eq. (3.25). To this end, we must distinguish between the case in which $\zeta < 1$, corresponding to an *underdamped system*, and that in which $\zeta > 1$, corresponding to an *overdamped system*.

Using the table of Laplace transforms in Appendix A and inverse Laplace transforming Eq. (3.25), the response of an *underdamped* single-degree-of-freedom system to initial excitations can be shown to be

$$x(t) = e^{-\zeta\omega_n t}\left[x_0\left(\cos\omega_d t + \frac{\zeta\omega_n}{\omega_d}\sin\omega_d t\right) + \frac{v_0}{\omega_d}\sin\omega_d t\right] \tag{3.26}$$

The response can be rewritten in the convenient form

$$x(t) = Ae^{-\zeta^{\omega_n t}}\cos\left(\omega_d t - \phi\right) \tag{3.27}$$

where now the constant A is given by

$$A = \frac{1}{\omega_d}\left[(\omega_d x_0)^2 + (\zeta\omega_n x_0 + v_0)^2\right]^{1/2} \tag{3.28}$$

and the phase angle by

$$\phi = \tan^{-1}\frac{\zeta\omega_n x_0 + v_0}{\omega_d x_0} \tag{3.29}$$

Moreover,

$$\omega_d = \left(1 - \zeta^2\right)^{1/2}\omega_n \tag{3.30}$$

is known as the *frequency of damped free vibration*. The multiplying factor

$$A\exp\left(-\zeta\omega_n t\right)$$

in Eq. (3.27) can be regarded as a time-dependent amplitude modulating the harmonic oscillation at the frequency ω_d, so that the motion for $0 < \zeta < 1$ represents *decaying oscillation*. A typical plot of $x(t)$ versus t is displayed in Fig. 3.5, with the modulating envelope $\pm A\exp\left(-\zeta\omega_n t\right)$ versus t shown in dashed lines. Whereas the constant A can be identified as the magnitude of the modulating envelope, the phase angle ϕ does not lend itself to easy interpretation.

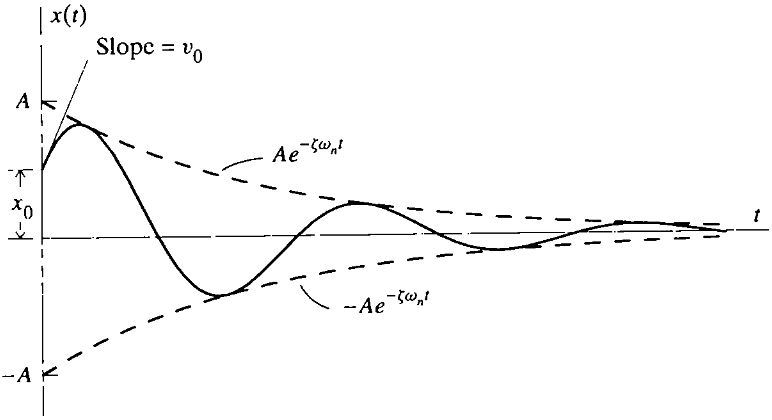


Figure 3.5 Free response of an underdamped single-degree-of-freedom system

The response of an *overdamped* single-degree-of-freedom system to initial excitations can be obtained from Eq. (3.26) by simply replacing ω_d by $i\left(\zeta^2-1\right)^{1/2}\omega_n$. Hence, recognizing that $\cos i\alpha = \cosh\alpha$, $\sin i\alpha = i\sinh\alpha$, Eq. (3.26) yields

$$x(t) = e^{-\zeta\omega_n t}\left\{x_0\left[\cosh\left(\zeta^2-1\right)^{1/2}\omega_n t + \frac{\zeta}{\left(\zeta^2-1\right)^{1/2}}\sinh\left(\zeta^2-1\right)^{1/2}\omega_n t\right] + \frac{v_0}{\left(\zeta^2-1\right)^{1/2}\omega_n}\sinh\left(\zeta^2-1\right)^{1/2}\omega_n t\right\} \tag{3.31}$$

which represents *aperiodically decaying motion.* A typical plot of $x(t)$ versus t is shown in Fig. 3.6.

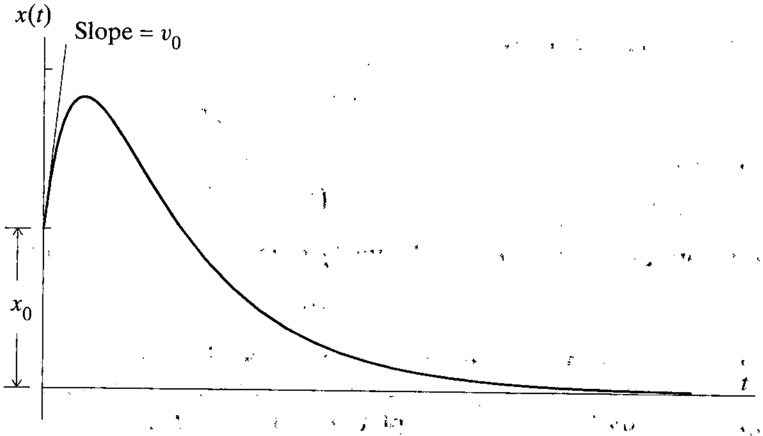


Figure 3.6 Free response of an overdamped single-degree-of-freedom system

The case in which $\zeta = 1$ is commonly known as *critical damping.* Letting $\zeta = 1$ in Eq. (3.31), the response of a critically damped single-degree-of-freedom system can be shown to have the form (see Problem 3.5)

$$x(t) = e^{-\omega_n t}\left[x_0\left(1+\omega_n t\right) + v_0 t\right] \tag{3.32}$$

As in the overdamped case, critically damped motion also *decays aperiodically.* In fact, it is the fastest decaying aperiodic motion. Clearly, there is nothing critical about $\zeta = 1$. It is merely a borderline case separating oscillatory decay from aperiodic decay.

It may prove of interest at this point to contrast the approach to the response to initial excitations presented in Sec. 1.4 to that used here. In Sec. 1.4, we assumed the exponential solution

$$x(t) = Ae^{st} \tag{3.33}$$

where, in the case of the first-order system, Eq. (3.2), s is the solution of the characteristic equation

$$s + a = 0 \tag{3.34}$$

and A is obtained by satisfying the initial condition, Eq. (3.3). On the other hand, in the case of the second-order system, Eq. (3.22), the characteristic equation is

$$s^2 + 2\zeta\omega_n s + \omega_n^2 = 0 \tag{3.35}$$

and the constants A_1 and A_2 corresponding to the roots s_1 and s_2 of the characteristic equation are determined by invoking the initial conditions, Eqs. (3.13). All these steps are carried out automatically in the solution by the Laplace transformation. Indeed, Eqs. (3.4) for the Laplace transform of derivatives take into account the initial conditions automatically. Moreover, recognizing that the characteristic polynomial appears at the denominator of the Laplace transform $X(s)$ of the response, as attested by Eqs. (3.6), (3.15) and (3.25), we conclude that the solution of the characteristic equation and the combination of the various exponential terms are implicit in the Laplace inversion process. Of course, a great deal of work is eliminated from the inversion process by the use of Laplace transform tables.

3.2 RESPONSE TO HARMONIC EXCITATIONS

In Sec. 1.5, we discussed the response of linear time-invariant systems to harmonic excitations in a general way. In this section, we propose to apply the theory developed there to first-order and second-order mechanical systems.

We shall find it convenient to express the excitation in the complex form

$$f(t) = Ake^{i\omega t} \tag{3.36}$$

and we recall from Sec. 1.5 that the complex notation has certain advantages over the real notation, i.e., the notation in terms of trigonometric functions. Also we note that A has units of displacement. Inserting Eq. (3.36) into Eq. (3.1), the differential equation of motion of a damper-spring system subjected to harmonic excitations can be written as

$$c\dot{x}(t) + kx(t) = f(t) = Ake^{i\omega t} \tag{3.37}$$

Then, dividing through by c, we obtain

$$\dot{x}(t) + ax(t) = Aae^{i\omega t}, \qquad a = k/c \tag{3.38}$$

But Eq. (3.38) is virtually identical to Eq. (a) of Example 1.2, except that here $e^{i\omega t}$ replaces $\sin\omega t$. Hence, using the analogy with Eq. (b) of Example 1.2, we can write the steady-state harmonic response in the form

$$x(t) = A|G(i\omega)|e^{i(\omega t - \phi)} \tag{3.39}$$

where

$$|G(i\omega)| = \frac{1}{\left[1 + (\omega\tau)^2\right]^{1/2}} \tag{3.40}$$

is the *magnitude* and

$$\phi(\omega) = \tan^{-1}\omega\tau \tag{3.41}$$

is the *phase angle of the frequency response*

$$G(i\omega) = \frac{1}{1 + i\omega\tau} \tag{3.42}$$

in which

$$\tau = \frac{1}{a} = \frac{c}{k} \tag{3.43}$$

is the time constant. Of course, if the excitation is $f(t) = Ak \cos \omega t$, we retain the real part of Eq. (3.39) as the response, and if the excitation is $f(t) = Ak \sin \omega t$ we retain the imaginary part, which was actually done in Example 1.2.

Equation (3.39) states that the response to harmonic excitation is harmonic and has the same frequency as the excitation, where the amplitude and phase angle of the response depend on the excitation frequency ω. This information is not readily accessible from plots of the response versus time. However, as pointed out in Sec. 1.5, a great deal of information concerning the amplitude and phase angle of the response can be gained from *frequency-response plots*, namely, $|G(i\omega)|$ versus ω and $\phi(\omega)$ versus ω.

It must be pointed out here that the nature of the amplitude and phase angle is distinctly different in the case of response to harmonic excitation from that in the case of response to initial excitations. Indeed, in the case at hand, we consider the amplitude and phase angle of the response in relation to the magnitude and phase angle of the excitation, respectively. Here the constant A plays no particular role, as it drops out when the ratio of the response amplitude to the excitation amplitude is considered, leaving the nondimensional magnitude $|G(i\omega)|$ of this ratio as the quantity of interest. Moreover, here the phase angle ϕ is clearly defined. It represents a measure of the time the peak response lags behind the peak excitation.

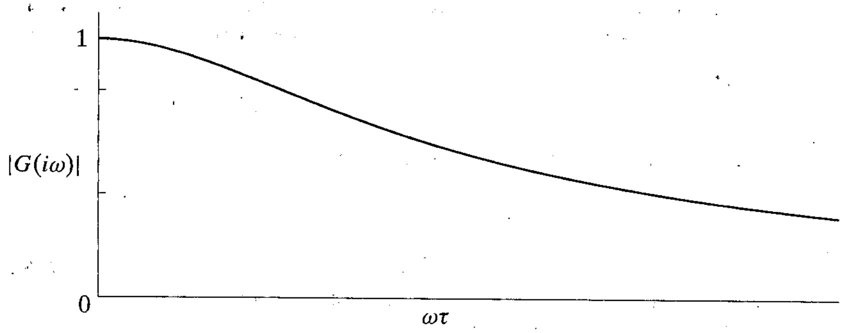


Figure 3.7 Magnitude of the frequency response for a damper-spring system versus nondimensional excitation frequency

Figure 3.7 displays the plot $|G(i\omega)|$ versus $\omega\tau$ for the first-order system given by Eq. (3.38); the plot is based on Eq. (3.40). We observe from Fig. 3.7 that the magnitude is attenuated greatly for large values of $\omega\tau$, and it remains largely unaffected for small values of $\omega\tau$. Hence, for a given τ, the first-order system acts like a filter, and in particular a *low-pass filter*. The plot $\phi(\omega)$ versus $\omega\tau$ is based on Eq. (3.41) and is shown in Fig. 3.8. We observe that the phase angle tends to 90° as $\omega\tau$ increases,

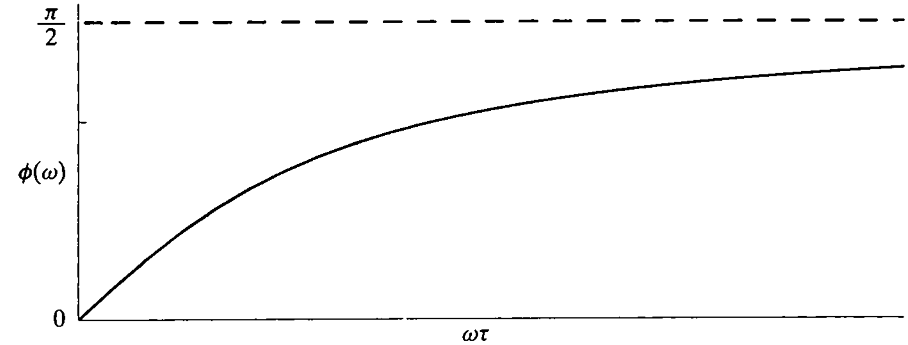


Figure 3.8 Phase angle of the frequency response for a damper-spring system versus nondimensional excitation frequency

so that the response tends to be 90° out of phase with the excitation for large values of $\omega\tau$.

Next we turn our attention to second-order systems subjected to harmonic excitations. Inserting Eq. (3.36) into Eq. (3.10), we obtain the differential equation of motion

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) = Ake^{i\omega t} \tag{3.44}$$

Dividing through by m, Eq. (3.44) can be rewritten as

$$\ddot{x}(t) + 2\zeta\omega_n\dot{x}(t) + \omega_n^2 x(t) = A\omega_n^2 e^{i\omega t} \tag{3.45}$$

where ζ is the viscous damping factor and ω_n is the natural frequency of undamped oscillation. Following the procedure of Sec. 1.5, the solution of Eq. (3.45) has the same general form as for first-order systems, namely,

$$x(t) = A\,|G(i\omega)|\,e^{i(\omega t-\phi)} \tag{3.46}$$

For the second-order system at hand, however, Eq. (1.45) yields the frequency response

$$G(i\omega) = \frac{\omega_n^2}{(i\omega)^2 + 2\zeta\omega_n(i\omega) + \omega_n^2} = \frac{1}{1 - (\omega/\omega_n)^2 + i2\zeta\omega/\omega_n} \tag{3.47}$$

so that, using Eq. (1.47), the magnitude is

$$|G(i\omega)| = \left[G(i\omega)\overline{G}(i\omega)\right]^{1/2} = \frac{1}{\left\{[1 - (\omega/\omega_n)^2]^2 + (2\zeta\omega/\omega_n)^2\right\}^{1/2}} \tag{3.48}$$

and, using Eq. (1.48), the phase angle is

$$\phi(\omega) = \tan^{-1}\frac{-\text{Im}\,G(i\omega)}{\text{Re}\,G(i\omega)} = \tan^{-1}\frac{2\zeta\omega/\omega_n}{1 - (\omega/\omega_n)^2} \tag{3.49}$$

To gain some insight into the behavior of single-degree-of-freedom systems as the excitation frequency varies, we make use once again of frequency response plots. Based on Eq. (3.48), Fig. 3.9 shows plots of $|G(i\omega)|$ versus ω/ω_n, with the viscous damping factor ζ acting as a parameter. In the first place, we note that all curves begin at $|G(i0)| = 1$. In the special case of an undamped system, $\zeta = 0$, the plot experiences a discontinuity at $\omega/\omega_n = 1$, at which point the displacement amplitude becomes infinite, a phenomenon referred to as *resonance*. Of course, infinite displacements are not possible for real physical systems, for which displacements must remain finite. In fact, the solution on which the plot is based rests on the assumption that displacements are sufficiently small so as to remain within the linear range. Hence, a separate solution must be produced for an undamped system at resonance. Nevertheless, the frequency response plot corresponding to $\zeta = 0$ serves as a warning that the system is likely to experience violent vibration when the excitation frequency passes through resonance. For $0 < \zeta < 1/\sqrt{2}$, the frequency response curves experience peaks at $\omega = (1 - 2\zeta^2)^{1/2}\omega_n$ and then approach zero asymptotically as ω increases. For $\zeta \geq 1/\sqrt{2}$, the curves experience no peaks but approach zero asymptotically as ω increases, albeit at a lower rate than for $\zeta < 1/\sqrt{2}$.

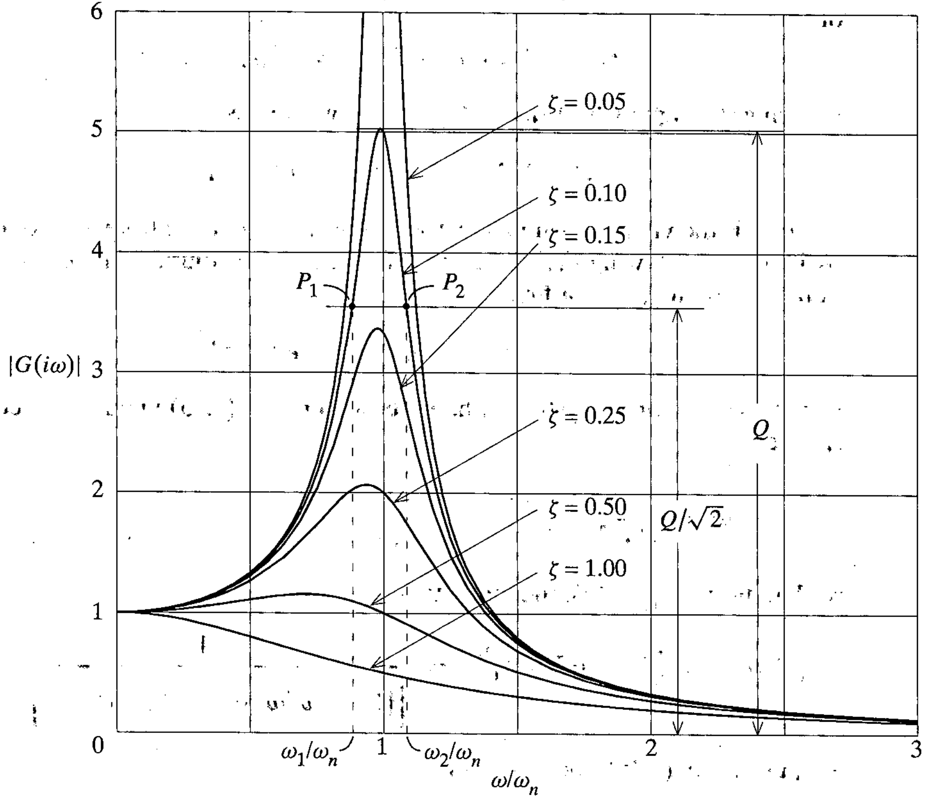


Figure 3.9 Magnitude of the frequency response for a damped single-degree-of-freedom system versus normalized excitation frequency with the damping factor as a parameter

For most systems of interest, damping tends to be light, e.g., $\zeta < 0.05$. In such cases, the curves $|G(i\omega)|$ versus ω/ω_n experience a maximum in the neighborhood of $\omega/\omega_n = 1$ and they are nearly symmetric with respect to a vertical through that point. Denoting the peak values by Q, $|G(i\omega)|_{\max} = Q$, it can be shown that for small ζ

$$Q \cong \frac{1}{2\zeta} \tag{3.50}$$

Note that Q is sometimes referred to as the *quality factor*. Moreover, the points P_1 and P_2 on each side of the peak corresponding to an amplitude of $|G(i\omega)|$ equal to $Q/\sqrt{2}$ are called *half-power points*. The excitation frequencies corresponding to P_1 and P_2 are denoted by ω_1 and ω_2, respectively, and the frequency band $\Delta\omega = \omega_2 - \omega_1$ is known as the *bandwidth* of the system. It can be shown that for small values of ζ

$$Q \cong \frac{1}{2\zeta} \cong \frac{\omega_n}{\omega_2 - \omega_1} \tag{3.51}$$

The second of the frequency response plots are phase angle plots. Figure 3.10 presents plots of $\phi(\omega)$ versus ω/ω_n for selected values of ζ. All curves pass through

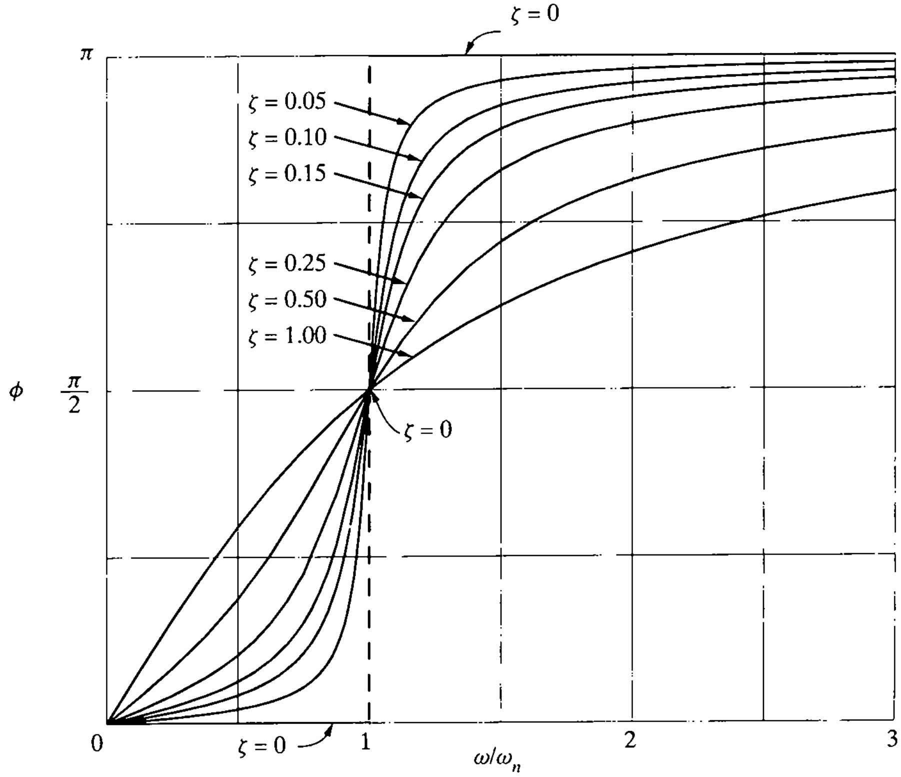


Figure 3.10 Phase angle of the frequency response for a damped single-degree-of-freedom system versus normalized excitation frequency with the damping factor as a parameter

the point $\phi = \pi/2$, $\omega/\omega_n = 1$, and for $\omega/\omega_n < 1$ they approach zero, whereas for $\omega/\omega_n > 1$ they approach π. For $\zeta = 0$, $\phi = 0$ for $\omega/\omega_n < 1$, and $\phi = \pi$ for $\omega/\omega_n > 1$. Note that for $\phi = 0$ the displacement $x(t)$ is in the same direction as the force $f(t)$. On the other hand, for $\phi = \pi$ the displacement is opposite in direction to the excitation. At $\omega/\omega_n = 1$, ϕ experiences a discontinuity, with the actual value there being equal to $\pi/2$.

Now a word about an earlier statement that, according to the superposition principle, the response to initial excitations and the response to external excitations can be obtained separately and combined linearly. Whereas this is true in general, such a statement must be questioned in the case in which the external excitation is harmonic. The reason for this lies in the different nature of the responses. Indeed, the response to initial excitations represents a transient response, which depends strongly on time, whereas the response to harmonic excitations is a steady-state response, which can be defined as the response that obtains after the transients have died down, for which time plays no particular role. Hence, combining the response to initial excitations and the response to harmonic excitations is not very meaningful.

Many systems in real life can be approximated by second-order systems subjected to harmonic excitations. Some of these systems have one characteristic in common, namely, they involve rotating eccentric masses, such as the system discussed in the example that follows.

Finally, we wish to explain the statement concerning the real part and imaginary part of the solution made earlier in this section. To this end, we consider a geometric representation of Eq. (3.45) in the complex plane. From Fig. 1.4, we conclude that the excitation $A\omega_n^2 e^{i\omega t}$ can be represented in the complex plane as a vector of magnitude $A\omega_n^2$ and making an angle ωt with respect to the real axis. Moreover, from Eq. (3.46), the response can be represented as a vector of magnitude $A\,|G\,(i\omega)|$ and making an angle $\omega t - \phi$ relative to the real axis. The two vectors are shown in Fig. 3.11. Differentiating Eq. (3.46) with respect to time, and recognizing that $i = \cos\dfrac{\pi}{2} + i\sin\dfrac{\pi}{2} = e^{i\pi/2}$, we can write

$$\dot{x}(t) = i\omega A\,|G\,(i\omega)|\,e^{i(\omega t-\phi)} = i\omega x(t) = \omega x(t)e^{i\pi/2} \tag{3.52}$$

so that the velocity can be interpreted as a vector of magnitude ω times the magnitude of the displacement vector and preceding the displacement vector by the phase angle $\pi/2$. Similarly, differentiating Eq. (3.46) a second time and considering the relation $-1 = \cos\pi + i\sin\pi = e^{i\pi}$, we obtain

$$\ddot{x}(t) = (i\omega)^2 A\,|G(i\omega)|\,e^{i(\omega t-\phi)} = -\omega^2 x(t) = \omega^2 x(t)e^{i\pi} \tag{3.53}$$

so that the acceleration can be interpreted as a vector of magnitude ω^2 times the magnitude of the displacement vector and leading the displacement vector by the phase angle π. In view of this, Eq. (3.45) can be represented geometrically by the trapezoidal diagram shown in Fig. 3.11. The angle ωt increases proportionally with time, so that the entire diagram rotates counterclockwise in the complex plane with the constant angular velocity ω. Retaining the real part of the excitation and of the response is tantamount to projecting the diagram on the real axis, and we observe that these projections vary harmonically with time, as they should. A similar statement

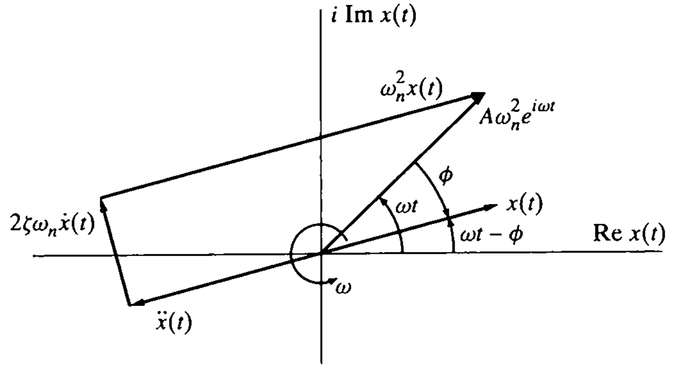


Figure 3.11 Geometric representation of Eq. (3.45) in the complex plane

can be made concerning the imaginary part of the excitation and response. We note that it was implicitly assumed that A is a real quantity. There is no loss of generality in doing so. Indeed, if A were to be complex, we could express it in the form $|A|\, e^{i\psi}$, thus advancing all four sides of the trapezoid by the same phase angle ψ, without affecting their relative positions and magnitudes. Also note that the diagram of Fig. 3.11 is equally valid for the first-order system described by Eq. (3.38), except that the trapezoid reduces to a right triangle.

Example 3.1

The system shown in Fig. 3.12a consists of a main mass $M - m$ and two eccentric masses $m/2$ rotating in opposite directions with the constant angular velocity ω. The system is mounted on two linear springs of stiffness $k/2$ each and a damper with the viscous damping coefficient c. Derive the system response and plot the corresponding amplitude and phase angle as functions of the driving frequency with the damping factor as a parameter.

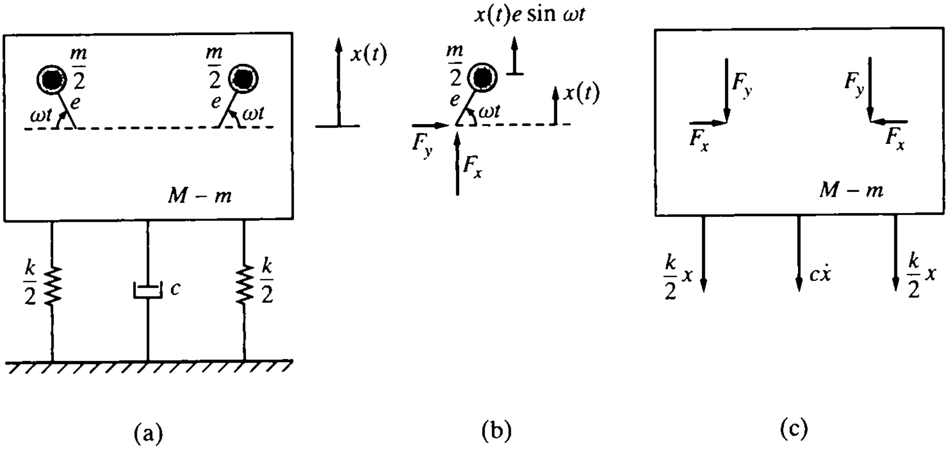


Figure 3.12 **(a)** System with unbalanced rotating masses **(b)** Free-body diagram for the right eccentric mass **(c)** Free-body diagram for the main mass

Although there are three masses, the motion of the eccentric masses relative to the main mass is known, so that this is only a single-degree-of-freedom system. In view of this, we define the response as the displacement $x(t)$ of the main mass. For convenience, we measure $x(t)$ from the static equilibrium position. In that position the springs are compressed by the amount $\delta_{st} = Mg/k$, where δ_{st} represents the static deflection, so that there is a combined constant force $k\delta_{st}$ in the springs balancing the total weight Mg of the system at all times. As a result, in deriving the equation of motion, the weight Mg can be ignored. We propose to derive the equation of motion by means of Newton's second law. To this end, we consider two free-body diagrams, one for the main mass and one for the right eccentric mass, as shown in Figs. 3.12b and 3.12c, respectively. Due to symmetry, a free-body diagram for the left eccentric mass is not necessary. Figure 3.12b shows two pairs of forces F_x and F_y exerted by the rotating masses on the main mass. The two horizontal forces F_y cancel each other, so that the system undergoes no horizontal motion. On the other hand, the two vertical forces F_x reinforce each other. Note that, consistent with ignoring the force $Mg/2$ in each of the springs, we ignore the force $mg/2$ in F_x.

Using Newton's second law in conjunction with Fig. 3.12b, the equation for the vertical motion of the main mass is

$$-2F_x - c\dot{x}(t) - 2\frac{k}{2}x(t) = (M - m)\ddot{x}(t) \tag{a}$$

Moreover, from Fig. 3.12c we observe that the vertical displacement of the eccentric mass is $x(t) + e\sin\omega t$, so that the equation for the vertical motion of the eccentric mass is

$$F_x = \frac{m}{2}\frac{d^2}{dt^2}\left[x(t) + e\sin\omega t\right] = \frac{m}{2}\left[\ddot{x}(t) - e\omega^2\sin\omega t\right] \tag{b}$$

Equations (a) and (b) can be combined into the system differential equation of motion

$$M\ddot{x}(t) + c\dot{x}(t) + kx(t) = me\omega^2\sin\omega t = \mathrm{Im}\left(me\omega^2 e^{i\omega t}\right) \tag{c}$$

Hence, the rotating eccentric masses exert a harmonic excitation on the system.

The solution of Eq. (c) can be written down directly by using results obtained earlier in this section. To this end, we divide both sides of Eq. (c) by M and write

$$\ddot{x}(t) + 2\zeta\omega_n\dot{x}(t) + \omega_n^2 x(t) = \mathrm{Im}\left(\frac{m}{M}e\omega^2 e^{i\omega t}\right) \tag{d}$$

where

$$2\zeta\omega_n = c/M, \quad \omega_n^2 = k/M \tag{e}$$

Then, comparing Eq. (d) to Eq. (3.45) and recalling that the solution of Eq. (3.45) is given by Eq. (3.46), we obtain the response

$$\begin{aligned} x(t) &= \mathrm{Im}\left[\frac{m}{M}e\left(\frac{\omega}{\omega_n}\right)^2 |G(i\omega)|\, e^{i(\omega t-\phi)}\right] \\ &= \frac{m}{M}e\left(\frac{\omega}{\omega_n}\right)^2 |G(i\omega)|\sin(\omega t - \phi) \end{aligned} \tag{f}$$

where $|G(i\omega)|$ and ϕ are given by Eqs. (3.48) and (3.49), respectively. Hence , in this particular case, the magnitude plot is $(\omega/\omega_n)^2\,|G(i\omega)|$ versus ω/ω_n. It is displayed in Fig. 3.13 with ζ playing the role of a parameter. The phase angle plot remains as in Fig. 3.10.

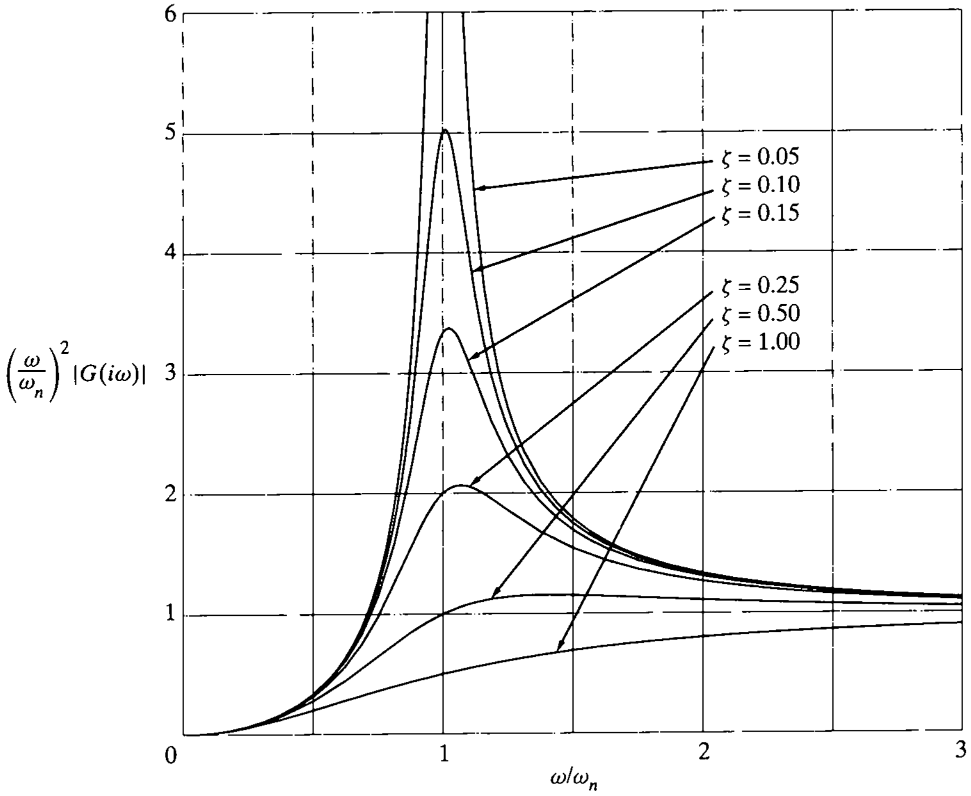


Figure 3.13 Magnitude of the frequency response for a system with unbalanced rotating masses versus normalized excitation frequency with the damping factor as a parameter

3.3 SYSTEMS WITH STRUCTURAL DAMPING

Structural damping is generally attributed to energy loss due to the hysteresis of elastic materials experiencing cyclic stress. Mathematically, structural damping is commonly treated as an equivalent viscous damping. To establish the analogy between structural and viscous damping, we consider the energy dissipated by the single-degree-of-freedom system under harmonic excitation given by Eq. (3.44) during one cycle of motion in the general form

$$\Delta E_{\text{cyc}} = \int_{\text{cyc}} f dx = \int_0^{2\pi/\omega} f\dot{x}dt \tag{3.54}$$

where f is the harmonic force, ω the frequency of the harmonic force and $\dot{x}$ is the velocity. But, f is given by the right side of Eq. (3.44) and $\dot{x}$ can be obtained from Eq. (3.46). Hence, inserting the real part of both f and $\dot{x}$ into Eq. (3.54), carrying out the integration and recalling Eq. (3.49), we obtain

$$\Delta E_{\text{cyc}} = \int_0^{2\pi/\omega} (\text{Re } f)(\text{Re } \dot{x})dt = -k\omega A^2 |G(i\omega)| \int_0^{2\pi/\omega} \cos\omega t \sin(\omega t - \phi)dt$$

$$= k\pi A^2 \, |G(i\omega)| \sin\phi = c\pi\omega X(i\omega)^2 \tag{3.55}$$

where $c = 2m\zeta\omega_n$ is the coefficient of viscous damping and

$$X(i\omega) = A\,|G(i\omega)| \tag{3.56}$$

is the displacement amplitude.

At this point, we turn our attention to the concept of structural damping. Experience shows that energy is dissipated in all systems, including systems regarded as conservative. Indeed, conservative vibrating systems represent more of a mathematical convenience than a physical reality. In support of this statement, we note the fact that, in the absence of persistent excitations , the vibration of a mass-spring system does not last forever but dies out eventually. This can be attributed to internal friction in the spring. Unlike viscous damping, this type of damping does not depend on the time rate of strain. Experiments performed by Kimball and Lovell (Ref. 1) indicate that for a large variety of materials, such as metals, glass, rubber and maple wood, subjected to cyclic stress in a way that the strains remain below the elastic limit, the internal friction is entirely independent of the time rate of strain. Their experiments indicate that, over a considerable frequency range, the internal friction depends on the amplitude of oscillation. In particular, the energy loss per cycle of stress was found to be proportional to the amplitude squared, or

$$\Delta E_{\text{cyc}} = \alpha X^2 \tag{3.57}$$

where α is a constant independent of the frequency of the harmonic oscillation. The type of damping causing this energy loss is referred to as *structural damping* and is generally attributed to the *hysteresis* in the elastic materials. During loading, a piece of material in cyclic stress follows a stress-strain path that differs from the path during unloading, as shown in Fig. 3.14, even when the strain amplitude is well below the elastic limit of the material. The stress-strain curve forms a *hysteresis loop*, and the energy dissipated during one cycle of stress is proportional to the area inside the hysteresis loop, shown as the shaded area in Fig. 3.14. For this reason, structural damping is also known as *hysteretic damping*.

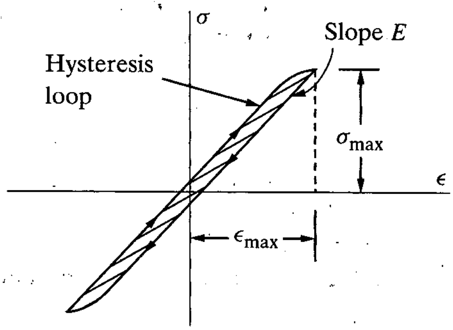


Figure 3.14 Stress-strain diagram showing a hysteresis loop

The fact that both for viscous damping and for structural damping the energy loss is proportional to the displacement amplitude squared, as given by Eq. (3.55) and Eq. (3.57), respectively, suggests an analogy whereby structurally damped systems subjected to harmonic excitation can be treated as viscously damped systems with the equivalent coefficient of viscous damping

$$c_{\text{eq}} = \frac{\alpha}{\pi\omega} \tag{3.58}$$

Under these circumstances, if we replace the parameter c in Eq. (3.44) by c_{eq}, as given by Eq. (3.58), we obtain the equation of motion of a structurally damped single-degree-of-freedom system in the form

$$m\ddot{x}(t) + \frac{\alpha}{\pi\omega}\dot{x}(t) + kx(t) = Ake^{i\omega t} \tag{3.59}$$

Then, recalling Eq. (3.52), Eq. (3.59) can be rewritten as

$$m\ddot{x}(t) + k(1 + i\gamma)x(t) = Ake^{i\omega t} \tag{3.60}$$

where

$$\gamma = \frac{\alpha}{\pi k} \tag{3.61}$$

is known as the *structural damping factor*. The quantity $k(1 + i\gamma)$ is referred to at times as *complex stiffness* and at other times as *complex damping*. From Eq. (3.60), it is clear that structural damping is proportional to the displacement and opposite in direction to the velocity.

The solution of Eq. (3.60) is

$$x(t) = \frac{Ae^{i\omega t}}{1 - (\omega/\omega_n)^2 + i\gamma} \tag{3.62}$$

and we note that the maximum response of structurally damped systems occurs exactly at $\omega = \omega_n$, in contrast to viscously damped systems (see Sec. 3.2). It must be stressed again that *the analogy between structural damping and viscous damping is valid only in the case of harmonic excitation.*

3.4 RESPONSE TO PERIODIC EXCITATIONS

In Sec. 3.2, we derived the response of linear time-invariant systems to harmonic excitations. It is well known, however, that periodic functions can be expanded into Fourier series, namely, series of harmonic functions. Hence, by invoking the superposition principle, it is possible to express the response to a periodic excitation as a series of harmonic responses. In this section, we propose to derive such a series.

There are essentially two types of Fourier series, real and complex. The real form is in terms of trigonometric functions and the complex form is in terms of exponential functions with imaginary exponents. Although the two forms are equivalent, and indeed one can be deduced from the other, our interest lies in the complex form, as in Sec. 3.2 we derived the response to harmonic excitations in complex form. Indeed, in Sec. 3.2 we demonstrated that the response to an excitation in the form

$$f(t) = Ake^{i\omega t} \tag{3.63}$$

can be expressed as

$$x(t) = A|G(i\omega)|e^{i(\omega t - \phi)} \tag{3.64}$$

with the understanding that if the actual excitation is $f(t) = Ak \cos \omega t$, we retain $\operatorname{Re} x(t)$ as the response, and if the excitation is $f(t) = Ak \sin \omega t$, we retain $\operatorname{Im} x(t)$.

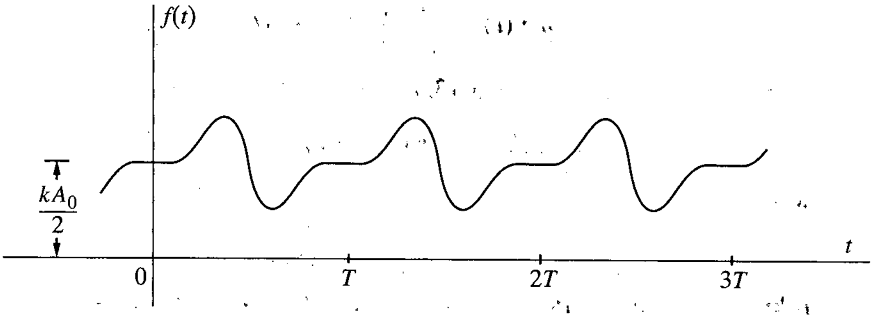


Figure 3.15 Periodic function

Next we consider a periodic function $f(t)$ as that depicted in Fig. 3.15. For the complex form of Fourier series to be equivalent to the real form, negative frequencies must be included in the series. Negative frequencies can be avoided by considering a Fourier series for the excitation in the form

$$f(t) = k\left(\frac{1}{2}A_0 + \operatorname{Re}\sum_{p=1}^{\infty} A_p e^{ip\omega_0 t}\right), \qquad \omega_0 = 2\pi/T \tag{3.65}$$

where ω_0 is the *fundamental frequency* and T is the *period* of $f(t)$. For $f(t)$ to represent a periodic function, p must be an integer, so that $p\omega_0$ $(p = 1, 2, \ldots)$ are *higher harmonics* with frequencies equal to integer multiples of the fundamental frequency. The coefficient A_0 is real and the coefficients A_p $(p = 1, 2, \ldots)$ are in general complex. They can all be obtained by means of the formula

$$A_p = \frac{2}{T}\int_0^T e^{-ip\omega_0 t} f(t)dt, \qquad p = 0, 1, 2, \ldots \tag{3.66}$$

and it should be noted that the integration limits can be changed without affecting the results, as long as the integration is carried out over a complete period. For example, $T/2$ and $-T/2$ are equally suitable as upper and lower limit, respectively, and in some cases they may be even more convenient.

According to the superposition principle, the response of a linear time-invariant system to the periodic excitation given by Eq. (3.65) can be obtained in the form of a linear combination of harmonic responses. Hence, using the analogy with Eq. (3.64), we can write the response in the general form

$$x(t) = \frac{1}{2}A_0 + \text{Re}\sum_{p=1}^{\infty} A_p|G_p|e^{i(p\omega_0 t-\phi_p)} \tag{3.67}$$

where $|G_p|$ and ϕ_p are the magnitude and phase angle of the frequency response corresponding to the excitation frequency $p\omega_0$. In the case of a first-order system, using the analogy with Eq. (3.42), the frequency response is

$$G_p = G(ip\omega_0) = \frac{1}{1 + ip\omega_0\tau} \tag{3.68}$$

which has the magnitude

$$|G_p| = \frac{1}{\left[1 + (p\omega_0\tau)^2\right]^{1/2}} \tag{3.69}$$

and the phase angle

$$\phi_p = \tan^{-1} p\omega_0\tau \tag{3.70}$$

On the other hand, considering Eq. (3.47), the frequency response for a second-order system is

$$G_p = \frac{1}{1 - (p\omega_0/\omega_n)^2 + i2\zeta p\omega_0/\omega_n} \tag{3.71}$$

which has the magnitude

$$|G_p| = \frac{1}{\left\{\left[1 - (p\omega_0/\omega_n)^2\right]^2 + (2\zeta p\omega_0/\omega_n)^2\right\}^{1/2}} \tag{3.72}$$

and the phase angle

$$\phi_p = \tan^{-1}\frac{2\zeta p\omega_0/\omega_n}{1 - (p\omega_0/\omega_n)^2} \tag{3.73}$$

Once again the question arises as to how to present the information concerning the response so as to gain as much physical insight into the system behavior as possible. Of course, one can always plot $x(t)$ versus t, but this is likely to be very tedious and not very informative. An efficient way of presenting the information is to plot the magnitude of the response versus the frequency, a plot known as a *frequency spectrum*. Because the response consists of a linear combination of harmonic components at the discrete frequencies $\omega = p\omega_0$ $(p = 0, 1, 2, \ldots)$, this is a discrete frequency spectrum. Of course, such a discrete frequency spectrum can be plotted for the excitation as well.

Example 3.2

The first-order system described by the differential equation (3.1) is subjected to the periodic force shown in Fig. 3.16. Determine the system response for the time constant $\tau = T/10$ and plot the excitation and response frequency spectra.

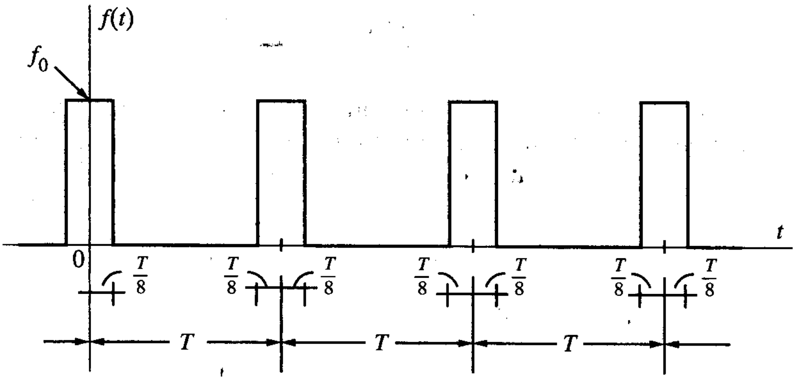


Figure 3.16 Periodic force

In this particular case, it is more convenient to work with the domain of integration $-T/2 < t < T/2$. Hence, the Fourier series for the excitation is given by Eq. (3.65), where in the case at hand

$$f(t) = \begin{cases} f_0, & -T/8 < t < T/8 \\ 0, & -T/2 < t < -T/8, \quad T/8 < t < T/2 \end{cases} \tag{a}$$

Inserting Eq. (a) into Eq. (3.66), we obtain the coefficients

$$\begin{aligned} A_0 &= \frac{2}{kT}\int_{-T/2}^{T/2} f(t)\,dt = \frac{f_0}{2k} \\ A_p &= \frac{2}{kT}\int_{-T/2}^{T/2} e^{-ip\omega_0 t} f(t)dt = \frac{2f_0}{kT}\frac{e^{-ip\omega_0 t}}{-ip\omega_0}\bigg|_{-T/8}^{T/8} \\ &= \frac{if_0}{kp\pi}\left(e^{-ip\pi/4} - e^{ip\pi/4}\right) = \frac{2f_0}{kp\pi}\sin p\pi/4, \qquad p = 1, 2, \ldots \end{aligned} \tag{b}$$

Figure 3.17a shows a normalized plot of the excitation frequency spectrum.

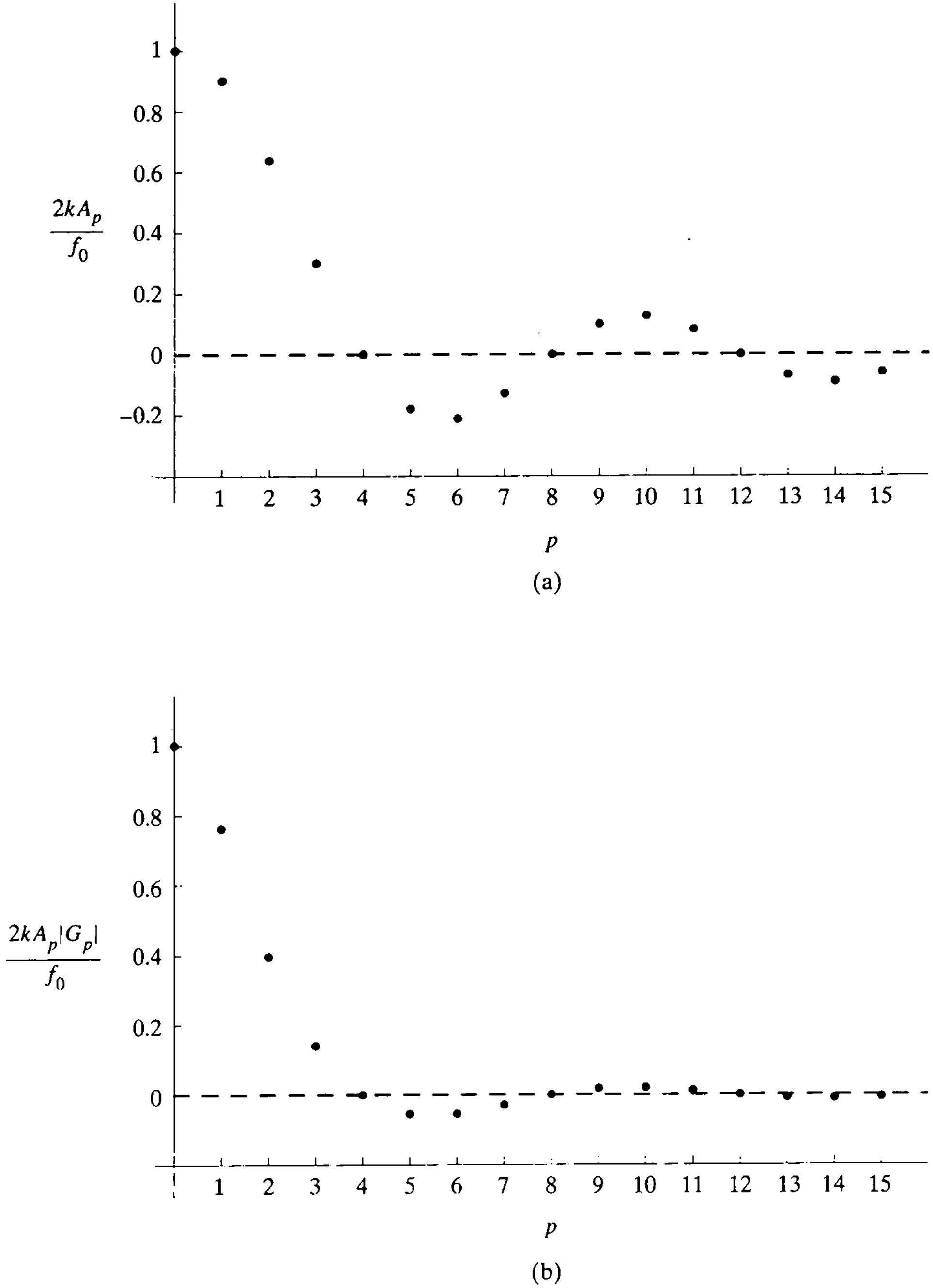
Introducing $\tau = T/10$ into Eq. (3.69), we obtain the frequency response magnitude for the first-order system at hand

$$|G_p| = \frac{1}{\left[1 + (p\omega_0 T/10)^2\right]^{1/2}} = \frac{1}{\left[1 + (p\pi/5)^2\right]^{1/2}}, \qquad p = 1, 2, \ldots \tag{d}$$

so that

$$A_p\,|G_p| = \frac{2f_0}{k\pi}\frac{\sin p\pi/4}{p\left[1 + (p\pi/5)^2\right]^{1/2}}, \qquad p = 1, 2, \ldots \tag{e}$$

A normalized plot of the response frequency spectrum is shown in Fig. 3.17b.



(a)

(b)

Figure 3.17 **(a)** Normalized frequency spectrum for the force of Fig. 3.16 **(b)** Normalized response frequency spectrum for a damper-spring system subjected to the force of Fig. 3.16.

3.5 RESPONSE TO ARBITRARY EXCITATIONS

It was demonstrated in Sec. 1.8 that the response of a linear time-invariant system to arbitrary excitations can be expressed in the form of a convolution integral, or

$$x(t) = \int_0^t f(\tau)g(t-\tau)d\tau = \int_0^t f(t-\tau)g(\tau)d\tau \tag{3.74}$$

where $f(t)$ is the excitation and $g(t)$ is the impulse response. In this section, we wish to expand on this subject in the context of first-order and second-order mechanical systems.

From Sec. 3.1, the differential equation of motion of a first-order system subjected to an arbitrary excitation $f(t)$ has the form

$$c\dot{x}(t) + kx(t) = f(t) \tag{3.75}$$

But, in Sec. 1.7, we have shown that the impulse response is the inverse Laplace transformation of the transfer function. Using Eq. (1.52), the transfer function for our first-order system can be written as

$$G(s) = \frac{1}{cs+k} = \frac{1}{c}\frac{1}{s+a}, \qquad a = k/c \tag{3.76}$$

Essentially the same first-order system was discussed in Example 1.5, from which we can write the impulse response

$$g(t) = \frac{1}{c}e^{-at}u(t) \tag{3.77}$$

Hence, inserting Eq. (3.77) into Eq. (3.74), we obtain the response of the first-order system in the form of the convolution integral

$$\begin{aligned} x(t) &= \int_0^t f(t-\tau)g(\tau)d\tau = \frac{1}{c}\int_0^t f(t-\tau)e^{-a\tau}u(\tau)d\tau \\ &= \frac{1}{c}\int_0^t f(t-\tau)e^{-a\tau}d\tau \end{aligned} \tag{3.78}$$

where we ignored $u(\tau)$ because it is equal to 1 over the interval of integration.

As an application, we propose to use Eq. (3.78) to derive the step response. Letting the excitation be the unit step function, $f(t) = u(t)$, the step response takes the form

$$s(t) = \frac{1}{c}\int_0^t u(t-\tau)e^{-a\tau}d\tau = \frac{1}{c}\frac{e^{-a\tau}}{-a}\bigg|_0^t = \frac{1}{k}\left(1 - e^{-at}\right)u(t) \tag{3.79}$$

where we multiplied the result by $u(t)$ to reflect the fact that the step response is zero for $t < 0$. The response is plotted in Fig. 3.18. Note also that the ramp response was derived in Example 1.5. However, because of a slight difference in the differential equation, the ramp response obtained in Example 1.5 must be divided by c.

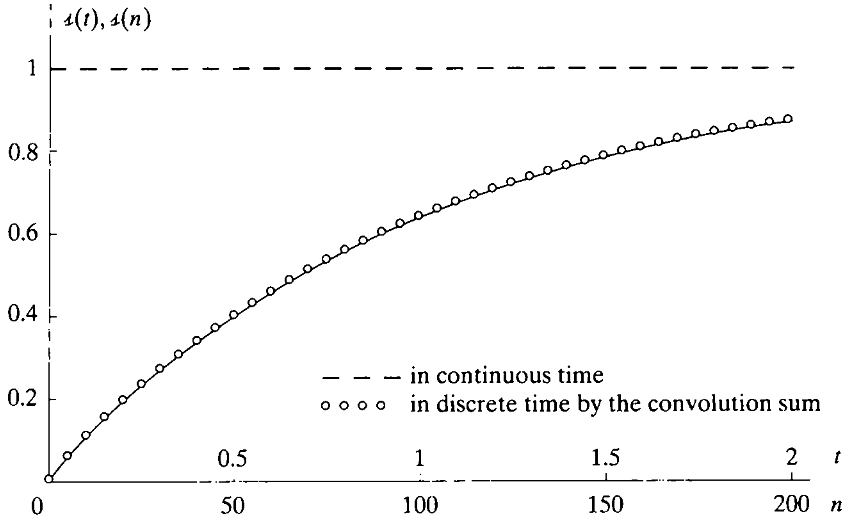


Figure 3.18 Step response for a damper-spring system.

Also from Sec. 3.1, we obtain the differential equation of motion of a second-order system in the form

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = f(t) \tag{3.80}$$

so that, using Eq. (1.58), we obtain the transfer function

$$G(s) = \frac{1}{ms^2 + cs + k} = \frac{1}{m}\frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{3.81}$$

Assuming that the system is underdamped, $\zeta < 1$, and using the table of Laplace transforms in Appendix A, we can write the impulse response as the inverse Laplace transform of $G(s)$ in the form

$$g(t) = \frac{1}{m\omega_d}e^{-\zeta\omega_n t}\sin\omega_d t\,\mathscr{u}(t), \qquad \omega_d = (1 - \zeta^2)^{1/2}\omega_n \tag{3.82}$$

so that, inserting Eq. (3.82) into Eq. (3.74), we obtain the response of a damped single-degree-of-freedom system to arbitrary excitations as the convolution integral

$$x(t) = \frac{1}{m\omega_d}\int_0^t f(t - \tau)e^{-\zeta\omega_n\tau}\sin\omega_d\tau\, d\tau \tag{3.83}$$

In Sec. 1.3, we made the comment that the distinction between initial excitations and external excitations is somewhat artificial, as some initial conditions are generated by external excitations. It appears that the time has arrived to explain this comment. Equation (3.8) represents the response of a first-order system to the initial excitation x_0, and Eq. (3.77) represents the impulse response of the same system. Yet the two responses are essentially the same, except for a multiplying factor. Hence, comparing Eqs. (3.8) and (3.77), we conclude that the effect of the unit impulse $\delta(t)$ acting on the first-order system described by Eq. (3.75) is to produce the equivalent initial displacement

$$x(0+) = \frac{1}{c} \tag{3.84}$$

Similarly, for $x_0 = 0$, Eq. (3.26) represents the response of a second-order system to the initial velocity v_0 and Eq. (3.82) represents the impulse response of the same system, and the two response differ only by a multiplying factor. Comparing Eqs. (3.26) and (3.82), we reach the conclusion that the effect of the unit impulse $\delta(t)$ acting on the second-order system described by Eq. (3.80) is to generate the equivalent initial velocity

$$v(0+) = \frac{1}{m} \tag{3.85}$$

Hence, some initial conditions are indeed generated by initial impulses.

In Sec. 1.9, we discussed a method for deriving the response of a system in the state space. Whereas the approach is better suited for multi-input, multi-output systems, we will find that the state-space formulation, which is based on first-order equations, has some merit even for single-degree-of-freedom systems, such as that described by Eq. (3.80). Before we cast Eq. (3.80) in state form, we rewrite it as

$$\ddot{x}(t) + 2\zeta\omega_n\dot{x}(t) + \omega_n^2 x(t) = m^{-1}f(t) \tag{3.86}$$

where ω_n is the natural frequency and ζ is the viscous damping factor, defined by Eqs. (3.12) and (3.23), respectively. Next, we define the two-dimensional state vector as $\mathbf{x}(t) = [x_1(t)\ x_2(t)]^T$, where $x_1(t) = x(t)$ and $x_2(t) = \dot{x}(t)$. Then, adding the identity $\dot{x}_1(t) = \dot{x}_1(t)$, Eq. (3.86) can be rewritten according to Eq. (1.119) in the state form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}f(t) \tag{3.87}$$

in which

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ m^{-1} \end{bmatrix} \tag{3.88a, b}$$

Then, from Eq. (1.121), the state response is given by

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t-\tau)\,\mathbf{b}f(\tau)d\tau \tag{3.89}$$

where $\Phi(t)$ is the transition matrix, which can be obtained conveniently by means of Eq. (1.114) as follows:

$$\Phi(t) = \mathcal{L}^{-1}(sI - A)^{-1} = \mathcal{L}^{-1}\begin{bmatrix} s & -1 \\ \omega_n^2 & s + 2\zeta\omega_n \end{bmatrix}^{-1}$$

$$= \mathcal{L}^{-1}\frac{1}{s^2 + 2\zeta\omega_n s + \omega_n^2}\begin{bmatrix} s + 2\zeta\omega_n & 1 \\ -\omega_n^2 & s \end{bmatrix}$$

$$= \frac{1}{\omega_d} e^{-\zeta\omega_n t} \begin{bmatrix} \omega_d \cos\omega_d t + \zeta\omega_n \sin\omega_d t & \sin\omega_d t \\ -\omega_n^2 \sin\omega_d t & \omega_d \cos\omega_d t - \zeta\omega_n \sin\omega_d t \end{bmatrix} \tag{3.90}$$

in which use has been made of the table of Laplace transform pairs in Appendix A. Hence, inserting Eqs. (3.88b) and (3.90) into Eq. (3.89), we obtain the state response in the general form

$$\begin{aligned} \mathbf{x}(t) = {} & \frac{1}{\omega_d} e^{-\zeta\omega_n t} \begin{bmatrix} \omega_d \cos\omega_d t + \zeta\omega_n \sin\omega_d t & \sin\omega_d t \\ -\omega_n^2 \sin\omega_d t & \omega_d \cos\omega_d t - \zeta\omega_n \sin\omega_d t \end{bmatrix} \mathbf{x}(0) \\ & + \frac{1}{m\omega_d} \int_0^t e^{-\zeta\omega_n (t-\tau)} \begin{bmatrix} \sin\omega_d (t - \tau) \\ \omega_d \cos\omega_d (t - \tau) - \zeta\omega_n \sin\omega_d (t - \tau) \end{bmatrix} f(\tau) d\tau \end{aligned} \tag{3.91}$$

Comparing Eqs. (3.83) and (3.91), we conclude that the response based on the transition matrix provides not only the displacement but also the velocity caused by external forces, as well as the displacement and velocity due to the initial displacement and velocity. The transition matrix, Eq. (3.90), will prove useful in our study of the discrete-time response of damped single-degree-of-freedom systems.

Example 3.3

Derive the step response of the single-degree-of-freedom system given by Eq. (3.80) by means of the convolution integral.

The general response of the a single-degree-of-freedom system to arbitrary excitations has the form of the convolution integral given by Eq. (3.83). To obtain the step response, we let

$$f(t - \tau) = \mathscr{u}(t - \tau) \tag{a}$$

in Eq. (3.83), recognize that $\mathscr{u}(t - \tau) = 1$ for $0 < \tau < t$ and obtain, after some algebraic operations,

$$\begin{aligned} \mathscr{s}(t) &= \frac{1}{m\omega_d} \int_0^t e^{-\zeta\omega_n \tau} \sin\omega_d \tau \, d\tau \\ &= \frac{1}{2im\omega_d} \int_0^t e^{-\zeta\omega_n \tau} \left(e^{i\omega_d \tau} - e^{-i\omega_d \tau}\right) d\tau \\ &= \frac{1}{2im\omega_d} \left[\frac{e^{-(\zeta\omega_n - i\omega_d)\tau}}{-(\zeta\omega_n - i\omega_d)} - \frac{e^{-(\zeta\omega_n + i\omega_d)\tau}}{-(\zeta\omega_n + i\omega_d)} \right] \Bigg|_0^t \\ &= \frac{1}{k} \left[1 - e^{-\zeta\omega_n t} \left(\cos\omega_d t + \frac{\zeta\omega_n}{\omega_d} \sin\omega_d t \right) \right] \mathscr{u}(t) \end{aligned} \tag{b}$$

3.6 DISCRETE-TIME SYSTEMS

In Sec. 3.5, we discussed the problem of deriving the response of first-order and second-order systems to arbitrary excitations by means of the convolution integral. The response can be evaluated in closed form only when the excitations represent relatively simple functions, or can be expressed as a superposition of simple functions. In the more general case, closed-form evaluation of the convolution integral is not possible, and the response must be obtained numerically, most likely by means of a digital computer. But time is a continuous independent variable, and digital computers accept only digital information, which is discrete in time. This problem was discussed in Sec. 1.10, where continuous-time systems are approximated by discrete-time systems. This amounts to regarding the input as a sequence of numbers, representing the excitation at given sampling times, and generating the output in the form of another sequence of numbers, representing the response at given discrete times. The understanding is that the sampling period, i.e., the time interval between two consecutive samplings, is sufficiently small that the error incurred because of the discretization-in-time process is insignificant. In this section, we apply the developments of Sec. 1.10 to first-order and second-order systems.

There are two techniques for computing the response in discrete time (see Sec. 1.10). The first uses a convolution sum and is suitable for single-input, single-output systems. One disadvantage of the convolution sum is that it is not a recursive process, so that the computation of the response at a given sampling time does not use results from the preceding computation. Moreover, the number of terms in the sum increases with each sampling. The second technique is based on the transition matrix and has none of these disadvantages. Indeed, it is a recursive process, it is suitable for multi-input, multi-output systems and the amount of computation is the same for each sampling.

As in Sec. 1.10, we denote the sampling times by t_j $(j = 0, 1, 2, \ldots)$ and let for convenience the sampling times be equally spaced, $t_j = jT$, where T is the sampling period. Then, from Eq. (1.131), the discrete-time response sequence is given by the convolution sum

$$x(n) = \sum_{j=0}^{n} f(j)g(n-j), \qquad n = 1, 2, \ldots \tag{3.92}$$

where $f(j)$ is the discrete-time excitation sequence and $g(j)$ is the discrete-time impulse response at $t_j = jT$, and we note that T was omitted from the arguments for brevity.

Next, we use Eq. (3.92) to obtain the discrete-time step response of the first-order system considered in Sec. 3.5. The excitation sequence for a unit step function is simply

$$f(j) = 1, \qquad j = 0, 1, 2 \ldots \tag{3.93}$$

Moreover, the discrete-time impulse response has the expression

$$g(j) = \frac{T}{c} e^{-jaT}, \qquad j = 0, 1, 2, \ldots \tag{3.94}$$

where $a = k/c$, in which k is the spring constant and c the coefficient of viscous damping. It should be noted that the discrete-time impulse response can be obtained by letting $t = jT$ in the continuous-time impulse response and multiplying the result by T. The multiplication by T can be explained by the difference in the definition of the discrete-time and continuous-time unit impulse. Equation (3.94) can be proved more formally by means of the scalar version of Eq. (1.140) (see Problem 3.30). Introducing Eqs. (3.93) and (3.94) into Eq. (3.92), we obtain the discrete-time step response sequence

$$\begin{aligned}
\mathscr{s}(0) &= f(0)g(0) = \frac{T}{c} \\
\mathscr{s}(1) &= \sum_{j=0}^{1} f(j)g(1-j) = \frac{T}{c}\left(1 + e^{-aT}\right) \\
\mathscr{s}(2) &= \sum_{j=0}^{2} f(j)g(2-j) = \frac{T}{c}\left(1 + e^{-aT} + e^{-2aT}\right) \\
&\vdots \\
\mathscr{s}(n) &= \sum_{j=0}^{n} f(j)g(n-j) = \frac{T}{c}\left(1 + e^{-aT} + \ldots + e^{-naT}\right)
\end{aligned} \tag{3.95}$$

The response for $c = 1\ \mathrm{N \cdot s/m}$, $a = 1\ \mathrm{s}^{-1}$ and $T = 0.01$ s is plotted in Fig. 3.18. As can be concluded from Fig. 3.18, the discrete-time step response matches quite well the continuous-time step response. Even better agreement can be obtained by decreasing the sampling period T. It should be stressed that, although the response sequence, Eqs. (3.95), has the appearance of a closed-form solution, the objective of the discrete-time approach is to carry out numerical computations. To emphasize this point, in Example 3.4 we process the response sequence entirely numerically.

Also from Sec. 1.10, the response of a damped single-degree-of-freedom system can be obtained by means of the recursive process

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \boldsymbol{\gamma} f(k), \qquad k = 0, 1, 2, \ldots \tag{3.96}$$

where $\mathbf{x}(k) = [x_1(k)\ x_2(k)]^T$ is the state vector, $f(k)$ the excitation and

$$\Phi = e^{AT}, \qquad \boldsymbol{\gamma} = A^{-1}\left(e^{AT} - I\right)\mathbf{b} \tag{3.97a, b}$$

are coefficient matrices, the first one being recognized as the discrete-time transition matrix and the second as a vector, in which

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 0 \\ m^{-1} \end{bmatrix} \tag{3.98a, b}$$

are the coefficient matrices in the continuous-time state equations. The use of the recursive process, Eq. (3.96), is illustrated in Example 3.5, in which the same problem as in Example 3.4 is solved.

Example 3.4

Use the convolution sum to derive the response of a damped single-degree-of-freedom system to the triangular pulse shown in Fig. 3.19. The system parameters are $m = 1$ kg, $\zeta = 0.1$, $\omega_n = 2$ rad s^{-1}.

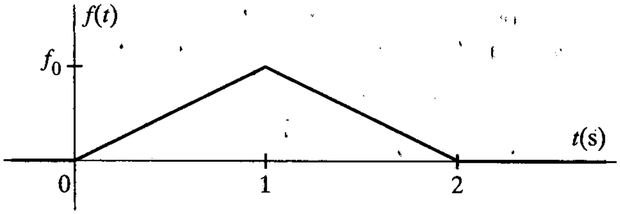


Figure 3.19 Triangular pulse

We choose the sampling period $T = 0.05$ s, so that the excitation sequence has the form

$$f(n) = \begin{cases} 0.05nf_0, & 0 < n \le 20 \\ [1 - 0.05\,(n - 20)]\,f_0, & 20 < n \le 40 \\ 0, & n > 40 \end{cases} \tag{a}$$

Moreover, from Eq. (3.82), the discrete-time impulse response can be written in the form (see Problem 3.31)

$$g(n) \doteq \frac{T}{m\omega_d} e^{-n\zeta\omega_n T} \sin n\omega_d T = 0.0251e^{-0.01n} \sin 0.0995n, \qquad n = 0, 1, 2, \ldots \tag{b}$$

Introducing Eqs. (a) and (b) into Eq. (3.92) and using the numerical values of the given parameters, we obtain the response sequence

$$\begin{aligned}
x(0) &= f(0)g(0) = 0 \\
x(1) &= \sum_{j=0}^{1} f(j)g(1-j) = 0 \\
x(2) &= \sum_{j=0}^{2} f(j)g(2-j) = 0.05f_0 \times 0.0025 = 0.0001f_0 \\
x(3) &= \sum_{j=0}^{3} f(j)g(3-j) = 0.05f_0 \times 0.0049 + 0.1f_0 \times 0.0025 = 0.0005f_0 \\
x(4) &= \sum_{j=0}^{4} f(j)g(4-j) \\
&= 0.05f_0 \times 0.0072 + 0.1f_0 \times 0.0049 + 0.15f_0 \times 0.0025 = 0.0012f_0 \\
x(5) &= \sum_{j=0}^{5} f(j)g(5-j) \\
&= 0.05f_0 \times 0.0093 + 0.1f_0 \times 0.0072 + 0.15f_0 \times 0.0049 + 0.2f_0 \times 0.0025 \\
&= 0.0024f_0
\end{aligned} \tag{c}$$

$$x(6) = \sum_{j=0}^{6} f(j)g(6-j)$$

$$=0.05f_0 \times 0.0114 + 0.1f_0 \times 0.0093 + 0.15f_0 \times 0.0072 + 0.20f_0 \times 0.0049$$
$$+ 0.25f_0 \times 0.0025 = 0.0042f_0$$

. . . .

It is clear that this is not a recursive process and that the number of operations increases with each sampling. The discrete-time response is shown in Fig. 3.20, in which the individual points are marked by white circles. The discrete-time response agrees quite well with the continuous-time response given by the solid line in Fig. 3.20 (see Problem 3.21).

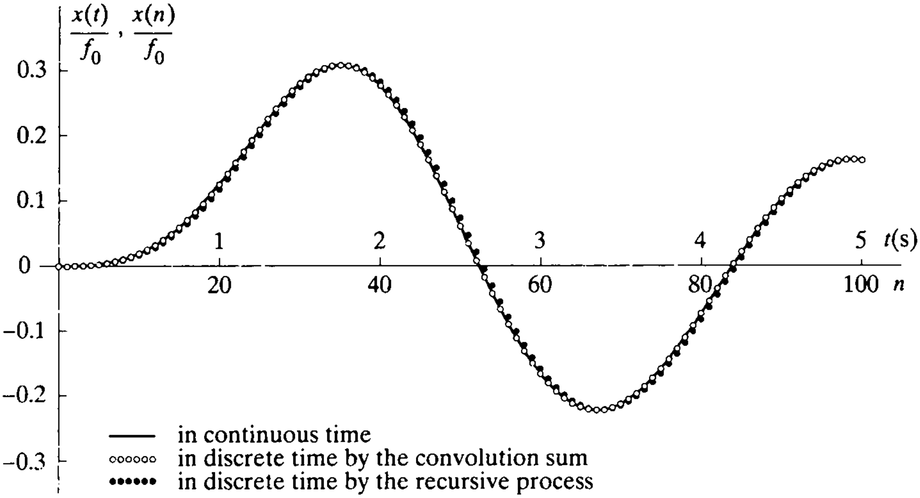


Figure 3.20 Reponse of a damped single-degree-of-freedom system

Example 3.5

Solve the problem of Example 3.4 by means of the recursive process given by Eq. (3.96), compare the results with those obtained in Example 3.4, as well as with the continuous-time response, and draw conclusions.

Introducing the values of the system parameters used in Example 3.4 into Eqs. (3.98), we obtain the coefficient matrices

$$A = \begin{bmatrix} 0 & 1 \\ -\omega_n^2 & -2\zeta\omega_n \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -4 & -0.4 \end{bmatrix}, \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{a}$$

so that, using Eqs. (3.97) with $T = 0.05$ s, the discrete-time coefficient matrices can be shown to have the form

$$\Phi = e^{AT} = \begin{bmatrix} 0.9950 & 0.0494 \\ -0.1977 & 0.9753 \end{bmatrix}, \qquad \gamma = A^{-1}\left(e^{AT} - I\right)\mathbf{b} = \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} \tag{b}$$

Then, the recursive process can be written as

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \gamma f(k), \qquad k = 0, 1, 2, \ldots \tag{c}$$

where $\mathbf{x}(0) = \mathbf{0}$ and $f(k)$ is the excitation sequence given by Eq. (a) of Example 3.4. Hence, inserting Eq. (a) of Example 3.4 and Eqs. (b) of this example into Eq. (c), we obtain the response sequence

$$\begin{aligned}
\mathbf{x}(1) &= \Phi\mathbf{x}(0) + \boldsymbol{\gamma} f(0) = \mathbf{0} \\
\mathbf{x}(2) &= \Phi\mathbf{x}(1) + \boldsymbol{\gamma} f(1) = \mathbf{0} + \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} 0.05 f_0 = \begin{bmatrix} 0.0001 \\ 0.0025 \end{bmatrix} f_0 \\
\mathbf{x}(3) &= \Phi\mathbf{x}(2) + \boldsymbol{\gamma} f(2) \\
&= \begin{bmatrix} 0.9950 & 0.0494 \\ -0.1977 & 0.9753 \end{bmatrix} \begin{bmatrix} 0.0001 \\ 0.0025 \end{bmatrix} f_0 + \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} 0.1 f_0 = \begin{bmatrix} 0.0003 \\ 0.0073 \end{bmatrix} f_0 \\
\mathbf{x}(4) &= \Phi\mathbf{x}(3) + \boldsymbol{\gamma} f(3) \\
&= \begin{bmatrix} 0.9950 & 0.0494 \\ -0.1977 & 0.9753 \end{bmatrix} \begin{bmatrix} 0.0003 \\ 0.0073 \end{bmatrix} f_0 + \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} 0.15 f_0 = \begin{bmatrix} 0.0009 \\ 0.0145 \end{bmatrix} f_0 \\
\mathbf{x}(5) &= \Phi\mathbf{x}(4) + \boldsymbol{\gamma} f(4) \\
&= \begin{bmatrix} 0.9950 & 0.0494 \\ -0.1977 & 0.9753 \end{bmatrix} \begin{bmatrix} 0.0009 \\ 0.0145 \end{bmatrix} f_0 + \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} 0.2 f_0 = \begin{bmatrix} 0.0018 \\ 0.0239 \end{bmatrix} f_0 \\
\mathbf{x}(6) &= \Phi\mathbf{x}(5) + \boldsymbol{\gamma} f(5) \\
&= \begin{bmatrix} 0.9950 & 0.0494 \\ -0.1977 & 0.9753 \end{bmatrix} \begin{bmatrix} 0.0018 \\ 0.0239 \end{bmatrix} f_0 + \begin{bmatrix} 0.0012 \\ 0.0494 \end{bmatrix} 0.25 f_0 = \begin{bmatrix} 0.0033 \\ 0.0353 \end{bmatrix} f_0 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{d}$$

For comparison purposes, the discrete-time displacement, given by the top component of the vectors $\mathbf{x}(n)$, is also shown in Fig. 3.20, in which the individual points are marked by black circles. As can be concluded from Fig. 3.20, there is a difference between the two responses resembling one caused by a phase shift, with the response obtained by the convolution sum leading the response obtained by the discrete-time transition matrix. The response by the discrete-time transition matrix is not as close to the continuous-time response as the response by the convolution sum, which can be attributed to the error involved in the approximation of the discrete-time excitation vector. Indeed, the effect of the approximation is to shift the staircase representing the discrete-time pulse by the amount $T/2$ relative to the continuous-time pulse, as well as relative to the sampled pulse, which explains the phase shift mentioned above. Of course, this error tends to decrease with a decrease in the sampling period.

3.7 SYNOPSIS

This chapter represents the beginning of our study of vibrations. It may appear as a timid beginning, as the material is generally covered in a first course on vibrations, but a good case can be made for the inclusion of the material in this chapter. Indeed, first-order and second-order systems, particularly the latter, are fundamental to a study of vibrations. The inclusion of this material is not only for completeness but also for convenience, as during the course of our study we will refer to it repeatedly.

There are many mechanical systems of practical interest whose behavior can be simulated by single-degree-of-freedom models. Whereas input-output relations for such systems were discussed in Chapter 1, in this chapter we examine these relations in greater depth and for a larger variety of inputs. One important conclusion reached is that the nature of the excitations dictates the manner in which the input-output relations are described. In particular, for transient inputs, such as initial and arbitrary excitations, time-domain descriptions of the response are indicated, and for steady-state inputs, such as harmonic and periodic excitations, frequency-domain descriptions are capable of yielding more useful information. Single-degree-of-freedom systems are important for other reasons as well. Indeed, using a technique known as modal analysis, it is shown in Chapters 4 and 7, that the equations of motion of linear, time-invariant multi-degree-of-freedom discrete systems and distributed-parameter systems, respectively, can be transformed into sets of independent second-order systems.

Finally, the subject of discrete-time systems, not ordinarily included in books on vibrations, must be noted. In processing solutions on a digital computer, it is necessary to list the excitation and response for discrete values of time. Discrete-time systems is a generic term which merely refers to a formal framework for programming solutions on a digital computer.

PROBLEMS

3.1 Obtain the response of the first-order system defined by Eq. (3.2) to the initial excitation given by Eq. (3.3) by means of the classical approach, i.e., by assuming an exponential solution, solving a characteristic equation for the exponent and making the solution satisfy the initial condition.

3.2 Repeat Problem 3.1 for the harmonic oscillator defined by Eq. (3.11) and subject to the initial conditions (3.13).

3.3 Plot the response of a damped single-degree-of-freedom system to the initial conditions $x_0 = 2$ cm, $v_0 = 4$ cm s^{-1}. The system parameters are $\zeta = 0.2$, $\omega_n = 2.5$ rad s^{-1}.

3.4 Repeat Problem 3.3, but with the viscous damping factor $\zeta = \sqrt{2}$.

3.5 Derive the response of a critically damped second-order system to the initial excitations $x(0) = x_0$, $\dot{x}(0) = v_0$, thus verifying Eq. (3.32).

3.6 Derive the response of a first-order system, Eq. (3.37), to the harmonic excitation $f(t) = Ak \sin \omega t$ by assuming a solution in terms of trigonometric functions and draw conclusions.

3.7 Derive the response of a second-order system, Eq. (3.44), to the harmonic excitation $f(t) = Ak \cos \omega t$ by assuming a solution in terms of trigonometric functions and draw conclusions.

3.8 The support of the mass-damper-spring system shown in Fig. 3.21 undergoes the harmonic motion $y(t) = A \sin \omega t$. Derive the system response and plot the magnitude and phase angle versus ω/ω_n diagrams for $\zeta = 0$, 0.1 and 2.

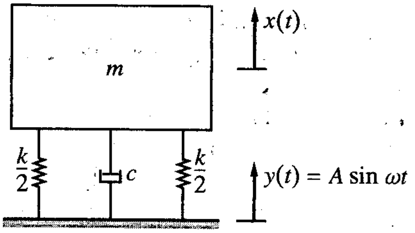


Figure 3.21 Single-degree-of-freedom system with support undergoing harmonic motion

3.9 A vehicle modeled as a mass-damper-spring system is traveling on a wavy road at the constant velocity v, as shown in Fig. 3.22. Derive the system response and plot the magnitude and phase angle of the force transmitted to the vehicle versus $\omega L/v$ for $\zeta = 0$, 0.2 and 1.5.

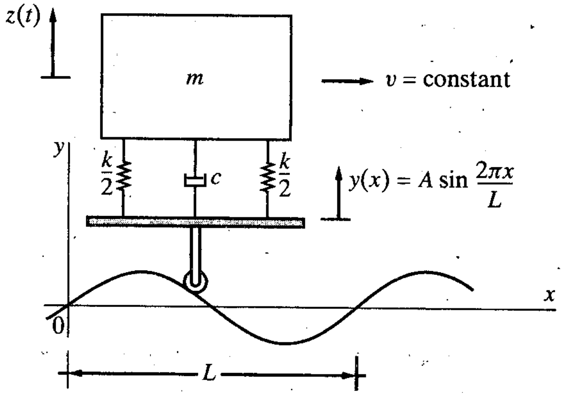


Figure 3.22 Vehicle traveling on a wavy road

3.10 A rigid disk of mass M with a mass m attached at a distance ℓ from the center of the disk is mounted on a viscously damped, simply supported massless shaft of flexural rigidity EI, as shown in Fig. 3.23. The shaft is whirling with the angular velocity ω. Determine the minimum viscous damping factor so that the peak bending displacement of the disk center will not exceed $3Me/m$, where e is the eccentricity.

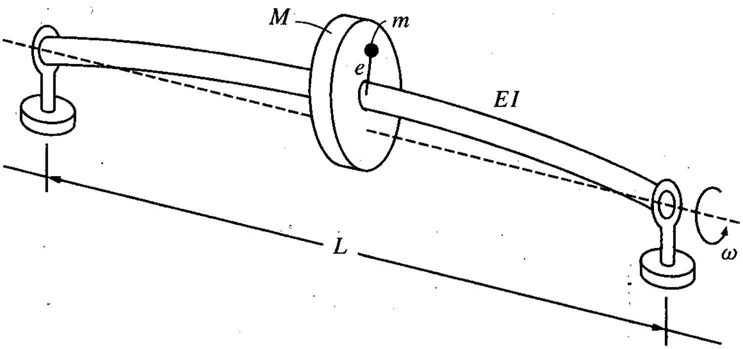


Figure 3.23 Whirling elastic shaft with an eccentric mass

3.11 Derive the response of a first-order system to the periodic excitation shown in Fig. 3.24. Plot the excitation and response frequency spectra.

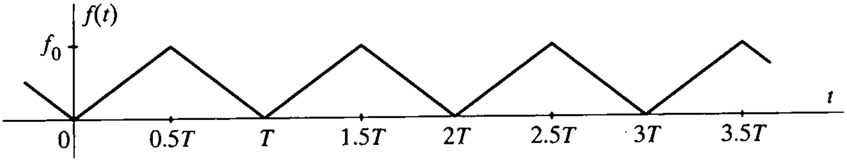


Figure 3.24 Periodic excitation

3.12 Derive the response of a first-order system to the periodic excitation shown in Fig. 3.25. Plot the excitation and response frequency spectra.

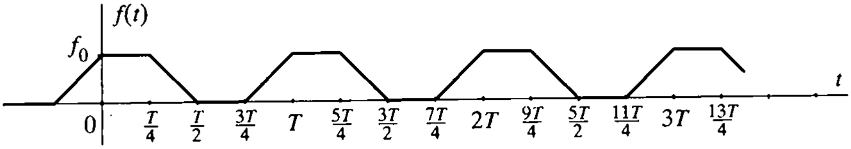


Figure 3.25 Periodic excitation

3.13 A harmonic oscillator is acted upon by the periodic force shown in Fig. 3.24. Derive the system response and plot the excitation and response frequency spectra and explain the response spectrum. Use $\omega_0/\omega_n = 0.35$.

3.14 Repeat Problem 3.13 for the periodic force shown in Fig. 3.25.

3.15 The damped single-degree-of-freedom system described by Eq. (3.44) is acted upon by the periodic force shown in Fig. 3.26. Derive the system response and plot the excitation and response frequency spectra and explain the response spectrum. Use the parameters $\zeta = 0.25$, $\omega_0/\omega_n = 0.30$.

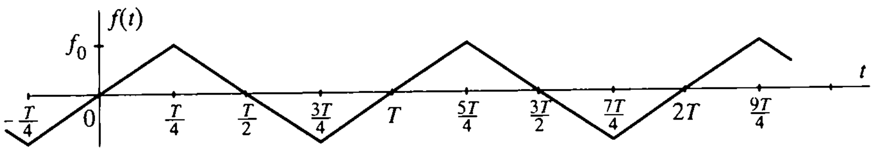


Figure 3.26 Periodic force

3.16 Repeat Problem 3.13 for the periodic excitation force of Fig. 3.27.

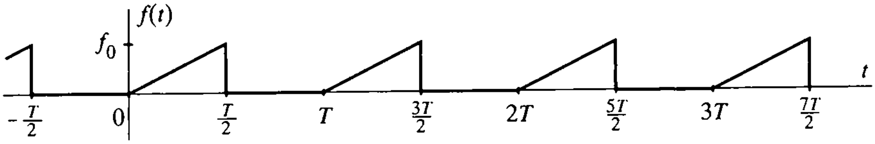


Figure 3.27 Periodic force

3.17 The support of the mass-damper-spring system shown in Fig. 3.21 undergoes the periodic motion depicted in Fig. 3.28. Derive the response for $\zeta = 0.2$ and $\omega_0/\omega_n = 0.4$ and plot the excitation and response frequency spectra. Explain the response frequency spectrum.

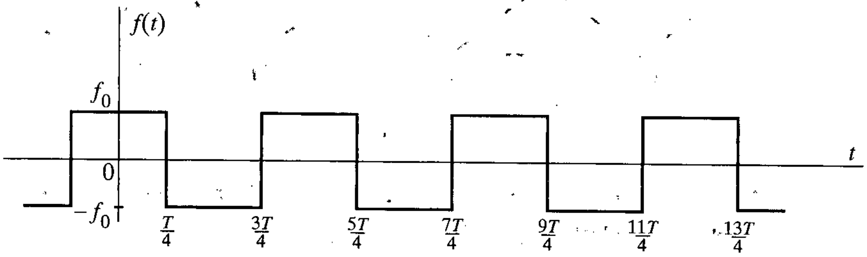


Figure 3.28 Periodic motion of support

3.18 Derive the ramp response of the first-order system given by Eq. (3.75). Plot the response for $a = 1\ \text{s}^{-1}$.

3.19 Derive the ramp response of the damped second-order system described by Eq. (3.44). Then, verify that the ramp response is equal to the integral of the step response, Eq. (b) of Example 3.3.

3.20 Use the convolution process of Fig. 1.12 to derive the response of the first-order system described by Eq. (3.75) to the triangular pulse shown in Fig. 3.19. Plot the response for $a = 1\ \text{s}^{-1}$.

3.21 Use the convolution process of Fig. 1.12 to derive the response of the system of Example 3.4 to the triangular pulse shown in Fig. 3.19. Plot the response over the time interval $0 < t < 3$ s.

3.22 Use the convolution process of Fig. 1.12 to derive the response of the damped second-order system described by Eq. (3.44) to the rectangular pulse shown in Fig. 3.29. Plot the response for $m = 1$ kg, $\zeta = 0.1$, $\omega_n = 4$ rad s^{-1} and $t_1 = 2$ s.

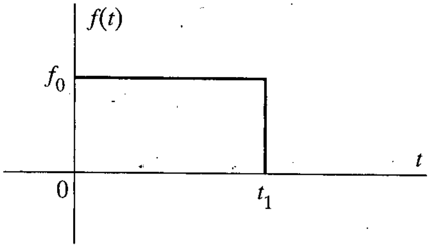


Figure 3.29 Rectangular pulse

3.23 Repeat Problem 3.22 for the trapezoidal pulse shown in Fig. 3.30. Use $t_1 = 1$ s and $t_2 = 2$ s.

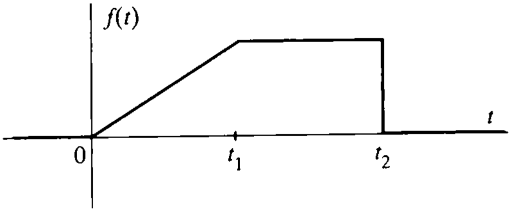


Figure 3.30 Trapezoidal pulse

3.24 Solve Problem 3.20 by the approach based on the transition matrix. Plot the displacement only.

3.25 Compute the state for the system of Problem 3.21 by the approach based on the transition matrix. Plot the displacement only.

3.26 Compute the state for the system of Problem 3.23 by the approach based on the transition matrix. Plot the displacement only.

3.27 Show that Problem 3.21 can be solved by regarding the response as a combination of ramp responses.

3.28 Show that Problem 3.22 can be solved by regarding the response as a combination of step responses.

3.29 Show that Problem 3.23 can be solved by regarding the response as a combination of ramp and step responses.

3.30 Prove Eq. (3.94) by means of the scalar version of Eq. (1.140).

3.31 Use Eq. (1.140) to show that the discrete-time impulse response of a damped single-degree-of-freedom system has the expression

$$g(n) = \frac{T}{m\omega_d} e^{-n\zeta\omega_n T} \sin n\omega_d T$$

3.32 Solve Problem 3.22 in discrete time using the convolution sum with $T = 0.05$ s.

3.33 Solve Problem 3.23 in discrete time using the convolution sum with $T = 0.05$ s.

3.34 Solve Problem 3.20 by the approach based on the discrete-time transition matrix. Use $T = 0.05$ s and compare results with those obtained in Problem 3.20.

3.35 Solve the problem of Example 3.5 using $T = 0.01$ s and assess the improvement in the computed response.

3.36 Solve Problem 3.22 by the approach based on the discrete-time transition matrix for $T = 0.05$ s and $T = 0.01$ s. Plot the displacement versus time for the two cases, compare results with the continuous-time solution and assess the accuracy of the discrete-time solutions.

3.37 Repeat Problem 3.36 for the case solved in Problem 3.23.

BIBLIOGRAPHY

1. Kimball, A. L. and Lovell, D. E., "Internal Friction in Solids," *Physical Reviews*, Vol. 30 (2nd ser.), 1927, pp. 948–959.
2. Meirovitch, L., *Introduction to Dynamics and Control*, Wiley, New York, 1985.
3. Meirovitch, L., *Elements of Vibration Analysis*, 2nd ed., McGraw-Hill, New York, 1986.
4. Thomson, W. T., *Theory of Vibration With Applications*, 4th ed., Prentice Hall, Englewood Cliffs, NJ, 1993.

4

MULTI-DEGREE-OF-FREEDOM SYSTEMS

Many vibrating systems can be represented by simple mathematical models, such as single-degree-of-freedom systems. Although these are mere idealizations of more complex physical systems, they are frequently capable of capturing the essential dynamic characteristics of the system. Quite often, however, such idealizations are not possible, and more refined mathematical models are advised. There are two types of mathematical models in common use, *discrete*, or *lumped models* and *distributed*, or *continuous models*. The choice of model depends on the nature of the system parameters, namely, mass, damping and stiffness. The dynamic behavior of discrete systems is described by a finite number of time-dependent variables. We recall from Sec. 2.7 that the minimum number of variables required to describe the dynamic behavior fully is referred to as the number of degrees of freedom of the system. On the other hand, the dynamic behavior of distributed systems is described by one or several variables depending on both time and space. Distributed systems are said to possess an infinite number of degrees of freedom. This chapter is concerned with the vibration of discrete systems. Distributed systems are discussed later in this text.

As shown in Chapter 2, the equations of motion of multi-degree-of-freedom systems can be derived conveniently by the Lagrangian approach. They consist of a set of simultaneous ordinary differential equations relating the system response to the excitations. The problem of solving the equations of motion for the response is of vital importance in vibrations. The equations of motion are frequently linear, but they can be nonlinear. Corresponding to given initial conditions, the solution of the equations of motion can be envisioned geometrically as tracing a trajectory in the state space. General solutions for multi-degree-of-freedom nonlinear systems are not possible. Under certain circumstances, however, the equations of motion admit special solutions in the state space. These special solutions are characterized by constant displacements and zero velocities, for which reason they are called *equilibrium points*. In vibrations, there is considerable interest in motions in the small neighborhood of equilibrium points. Such motions are governed by linearized equations about a given equilibrium point. The equations can be expressed conveniently in matrix form.

Linear conservative natural systems occupy a central position in vibrations. Such systems are capable of so-called *natural motions*, in which all the system coordinates execute harmonic oscillation at a given frequency and form a certain displacement pattern, where the oscillation frequencies and displacement patterns are called *natural frequencies* and *natural modes*, respectively. The natural frequencies and modes represent an inherent characteristic of the system and can be obtained by solving the so-called *algebraic eigenvalue problem* for the system, namely, a set of homogeneous algebraic equations. The eigenvalue problem for conservative natural systems can be defined in terms of a single real symmetric matrix. Its solution consists of *real eigenvalues*, which are related to the natural frequencies, and *real orthogonal eigenvectors*, which represent the natural modes. The orthogonality property is very powerful, as it permits the transformation of a set of simultaneous ordinary differential equations of motion to a set of independent equations. Each of the independent equations is of second order and resembles entirely the equation of motion of a single-degree-of-freedom system, so that the equations can be solved by the methods of Chapter 3. This procedure for solving the differential equations of motion is known as *modal analysis*. Another class of systems of interest in vibrations is that of *conservative gyroscopic systems*. It is demonstrated that gyroscopic systems possess many of the desirable properties of natural systems. Conservative systems represent a mathematical idealization, and in reality all vibrating systems dissipate energy. This idealization can be justified when the energy dissipation is sufficiently small that it can be ignored. Systems with perceptible energy dissipation belong to the class of *nonconservative systems*; the class includes systems with *viscous damping forces* and *circulatory forces*. The eigenvalue problem for nonconservative systems is characterized by *complex eigenvalues* and *complex biorthogonal eigenvectors*. The equations of motion for nonconservative systems can be solved by means of a method based on the transition matrix, as discussed in Sec. 1.9, or by means of a modal analysis based on the biorthogonality property. In many cases, particularly for systems subjected to arbitrary external forces, the evaluation of the response can cause difficulties. In such cases, it is advisable to determine the response on a digital computer. The formalism for this approach is referred to as *discrete-time systems*, and it involves treating the time as if it were a discrete variable. Finally, there is the question of *nonlinear systems* for which linearization under the small-motions assumption cannot be justified. In such cases, one must be content with a numerical solution for the response. The algorithms for numerical integration of nonlinear differential equations of motion tend to be more involved than for linear ones, but the idea of treating the time as a discrete variable remains the same. All the topics mentioned are discussed in this chapter.

4.1 EQUATIONS OF MOTION

In Sec. 2.7, we have indicated that the motion of an n-degree-of-freedom system is fully described by n generalized coordinates $q_k(t)(k = 1, 2, \ldots, n)$. Then, in Sec. 2.11, we have demonstrated that the generalized coordinates satisfy the Lagrange's equations of motion

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} = Q_{knc}, \qquad k = 1, 2, \ldots, n \tag{4.1}$$

where $\dot{q}_k$ are generalized velocities, $L = T - V$ is the Lagrangian, in which T is the kinetic energy and V is the potential energy, and Q_{knc} are the nonconservative generalized forces.

From Sec. 2.12, we recall that in the case of a *nonnatural system*, the kinetic energy can be written in the form

$$T = T_2 + T_1 + T_0 \tag{4.2}$$

in which

$$T_2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\dot{q}_i\dot{q}_j \tag{4.3}$$

is quadratic in the generalized velocities, where $m_{ij} = m_{ji}$ are symmetric coefficients depending in general on the generalized coordinates, $m_{ij} = m_{ij}(q_1, q_2, \ldots, q_n)$,

$$T_1 = \sum_{j=1}^{n} f_j\dot{q}_j \tag{4.4}$$

is linear in the generalized velocities, where f_j are coefficients depending on the generalized coordinates, $f_j = f_j(q_1, q_2, \ldots, q_n)$, and

$$T_0 = T_0(q_1, q_2, \ldots, q_n) \tag{4.5}$$

is a function of the generalized coordinates alone and contains no generalized velocities. It should also be noted here that, although gyroscopic forces, such as those arising from T_1, are most commonly associated with spinning bodies, they can also arise in elastic pipes containing flowing liquid (Ref. 16).

The nonconservative generalized forces Q_{knc} include a class of forces derivable from a dissipation function. This is the class of *viscous damping forces*, which depend on generalized velocities and can be expressed in the form

$$Q_{k\text{visc}} = -\frac{\partial \mathcal{F}}{\partial \dot{q}_k}, \qquad k = 1, 2, \ldots, n \tag{4.6}$$

where

$$\mathcal{F} = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}\dot{q}_i\dot{q}_j \tag{4.7}$$

is known as *Rayleigh's dissipation function*, in which c_{ij} are *damping coefficients*; they are symmetric, $c_{ij} = c_{ji}$, and in general constant. Another class of nonconservative forces is that of *circulatory forces* (Refs. 1, 6 and 17), which arise in power-transmitting components such as cranks, shafts, etc. Circulatory forces often occur in tandem with viscous damping forces and can be expressed in the form

$$Q_{k\text{circ}} = -\frac{\partial \mathcal{F}'}{\partial \dot{q}_k}, \qquad k = 1, 2, \ldots, n \tag{4.8}$$

where

$$\mathcal{F}' = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} c'_{ij}\dot{q}_i\dot{q}_j + \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\dot{q}_i q_j \tag{4.9}$$

will be referred to as the *circulatory function*, in which the coefficients c'_{ij} are symmetric, $c'_{ij} = c'_{ji}$, and the coefficients h_{ij} are skew symmetric, $h_{ij} = -h_{ji}$. Actually, $\mathcal{F}'$ can contain also terms in the generalized coordinates alone, but such terms do not contribute to $Q_{k\text{circ}}$, so that they can be ignored. The two types of forces can be treated together by introducing the *modified Rayleigh's dissipation function*

$$\mathcal{F}^* = \mathcal{F} + \mathcal{F}' = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} c^*_{ij}\dot{q}_i\dot{q}_j + \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\dot{q}_i q_j \tag{4.10}$$

where

$$c^*_{ij} = c_{ij} + c'_{ij} \tag{4.11}$$

Then, excluding viscous damping forces and circulatory forces from Q_{knc}, we can rewrite Lagrange's equations as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_k}\right) - \frac{\partial L}{\partial q_k} + \frac{\partial \mathcal{F}^*}{\partial \dot{q}_k} = Q_k, \qquad k = 1, 2, \ldots, n \tag{4.12}$$

where now Q_k represent the nonconservative applied forces alone, and note that we omitted "nc" from the subscript to simplify the notation. Similarly, Hamilton's equations, Eqs. (2.149), can be rewritten as

$$\dot{q}_k = \frac{\partial \mathcal{H}}{\partial p_k}, \qquad k = 1, 2, \ldots, n \tag{4.13a}$$

$$\dot{p}_k = -\frac{\partial \mathcal{H}}{\partial q_k} - \frac{\partial \mathcal{F}^*}{\partial \dot{q}_k} + Q_k, \qquad k = 1, 2, \ldots, n \tag{4.13b}$$

For future reference, we wish to recast some of the results obtained in this section in matrix form. We begin with Lagrange's equations, Eqs. (4.12), which can be expressed symbolically as

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} + \frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}} = \mathbf{Q} \tag{4.14}$$

where $\mathbf{q} = [q_1\, q_2\, \dots\, q_n]^T$ is the generalized coordinate vector, $\dot{\mathbf{q}} = [\dot{q}_1\, \dot{q}_2\, \dots \dot{q}_n]^T$ the generalized velocity vector and $\mathbf{Q} = [Q_1\, Q_2\, \dots Q_n]^T$ the generalized applied force vector. Moreover,

$$\frac{\partial L}{\partial \dot{\mathbf{q}}} = \begin{bmatrix} \partial L/\partial \dot{q}_1 \\ \partial L/\partial \dot{q}_2 \\ \vdots \\ \partial L/\partial \dot{q}_n \end{bmatrix} \tag{4.15}$$

represents a symbolic n-vector, and a similar statement can be made concerning $\partial L/\partial \mathbf{q}$ and $\partial \mathcal{F}^*/\partial \dot{\mathbf{q}}$. Using Eq. (2.155) in conjunction with Eqs. (4.3) and (4.4), the Lagrangian for a nonnatural system can be expressed as

$$L = T_2 + T_1 - U \tag{4.16}$$

where

$$T_2 = \frac{1}{2}\dot{\mathbf{q}}^T M \dot{\mathbf{q}}, \qquad T_1 = \dot{\mathbf{q}}^T \mathbf{f} \tag{4.17a, b}$$

in which $M = [m_{ij}] = M^T$ is a symmetric $n \times n$ matrix of coefficients and $\mathbf{f} = [f_1\, f_2\, \dots\, f_n]^T$ is an n-vector, and $U = V - T_0$ is the dynamic potential, a function of the generalized coordinates alone according to Eq. (2.156). Similarly, using Eq. (4.10), the modified Rayleigh's dissipation function can be written in the matrix form

$$\mathcal{F}^* = \mathcal{F} + \mathcal{F}' = \frac{1}{2}\dot{\mathbf{q}}^T C^* \dot{\mathbf{q}} + \dot{\mathbf{q}}^T H \mathbf{q} \tag{4.18}$$

where

$$\mathcal{F} = \frac{1}{2}\dot{\mathbf{q}}^T C \dot{\mathbf{q}} \tag{4.19}$$

is the matrix form of Rayleigh's dissipation function, Eq. (4.7), in which $C = [c_{ij}] = C^T$ is the symmetric $n \times n$ *damping matrix*, and

$$\mathcal{F}' = \frac{1}{2}\dot{\mathbf{q}}^T C' \dot{\mathbf{q}} + \dot{\mathbf{q}}^T H \mathbf{q} \tag{4.20}$$

is the matrix form of the circulatory function, in which $C' = [c'_{ij}] = C'^T$ is a symmetric $n \times n$ *equivalent damping matrix* and $H = [h_{ij}] = -H^T$ is the skew symmetric $n \times n$ *circulatory matrix*. As pointed out earlier, in general $\mathcal{F}'$ also contains terms depending on $\mathbf{q}$ alone, but such terms are irrelevant and thus ignored. From Eqs. (4.18) - (4.20), we conclude that

$$C^* = C + C' \tag{4.21}$$

which is the matrix counterpart of Eq. (4.11). Then, according to Eqs. (4.6) and (4.8), we can write

$$\frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}} = \frac{\partial \mathcal{F}}{\partial \dot{\mathbf{q}}} + \frac{\partial \mathcal{F}'}{\partial \dot{\mathbf{q}}} = -\left(\mathbf{Q}_{\text{visc}} + \mathbf{Q}_{\text{circ}}\right) = C^* \dot{\mathbf{q}} + H\mathbf{q} \tag{4.22}$$

Introducing Eq. (4.22) into Eq. (4.14), the matrix form of Lagrange's equations is simply

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} = \mathbf{Q} + \mathbf{Q}_{\text{visc}} + \mathbf{Q}_{\text{circ}} = \mathbf{Q} - C^*\dot{\mathbf{q}} - H\mathbf{q} \tag{4.23}$$

It is not difficult to verify that Hamilton's equations have the matrix form

$$\dot{\mathbf{q}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \tag{4.24a}$$

$$\dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \mathbf{Q} + \mathbf{Q}_{\text{visc}} + \mathbf{Q}_{\text{circ}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \mathbf{Q} - C^*\dot{\mathbf{q}} - H\mathbf{q} \tag{4.24b}$$

where $\mathbf{p} = [p_1\ p_2\ \dots\ p_n]^T$ is the *generalized momentum vector*, and we note that $\dot{\mathbf{q}}$ in Eq. (4.24b) must be replaced by an expression in terms of $\mathbf{q}$ and $\mathbf{p}$ obtained from

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} \tag{4.25}$$

Example 4.1

The system shown in Fig. 4.1 consists of a mass m connected to a rigid ring through a viscous damper and two nonlinear springs. The ring rotates with the angular velocity Ω relative to the inertial system X, Y, the damper has a coefficient of viscous damping c and the nonlinear springs are characterized by the force-displacement relations

$$F_x(x) = -k_x(x + \epsilon_x x^3),\ F_y(y) = -k_y(y + \epsilon_y y^3) \tag{a}$$

where x and y are displacements measured relative to the rotating reference system x, y embedded in the ring and k_x, k_y, ϵ_x and ϵ_y are constants. Note that the minus sign implies that the spring forces are opposed to the displacements. In addition, the mass m is subjected to external damping forces proportional to the absolute velocities $\dot{X}$ and $\dot{Y}$, where the proportionality constant is h. Derive the equations of motion by means of Lagrange's equations, Eqs. (4.12).

From Fig. 4.1, the position vector of m is given by

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} \tag{b}$$

where $\mathbf{i}$ and $\mathbf{j}$ are unit vectors along the rotating body axes x and y, respectively. Moreover, the angular velocity vector of the body axes has the form

$$\boldsymbol{\omega} = \Omega\mathbf{k} \tag{c}$$

where $\mathbf{k}$ is a unit vector normal to $\mathbf{i}$ and $\mathbf{j}$. Taking the time derivative of Eq. (b) and recognizing that the unit vectors $\mathbf{i}$ and $\mathbf{j}$ rotate with the angular velocity $\boldsymbol{\omega}$, we obtain the absolute velocity vector

$$\begin{aligned}\mathbf{v} = \dot{\mathbf{r}}_{\text{rel}} + \boldsymbol{\omega} \times \mathbf{r} &= \dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \Omega\mathbf{k}\times(x\mathbf{i} + y\mathbf{j}) \\ &= (\dot{x} - \Omega y)\mathbf{i} + (\dot{y} + \Omega x)\mathbf{j}\end{aligned} \tag{d}$$

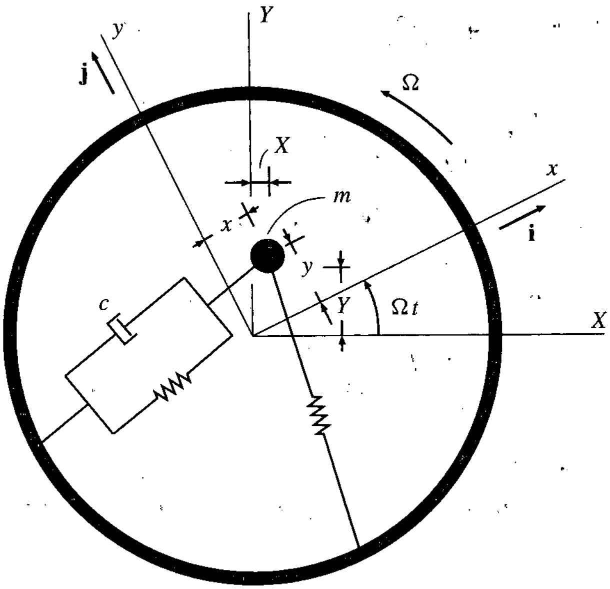


Figure 4.1 Mass connected to a rotating rigid ring through damper and springs

where $\dot{\mathbf{r}}_{\text{rel}}$ is the velocity of m relative to the rotating body axes. Hence, the kinetic energy is simply

$$
\begin{aligned}
T &= \frac{1}{2}m\mathbf{v}^T\mathbf{v} = \frac{1}{2}m\left[(\dot{x} - \Omega y)^2 + (\dot{y} + \Omega x)^2\right] \\
&= \frac{1}{2}m\left[\dot{x}^2 + \dot{y}^2 + 2\Omega\left(x\dot{y} - \dot{x}y\right) + \Omega^2\left(x^2 + y^2\right)\right] = T_2 + T_1 + T_0 \qquad \text{(e)}
\end{aligned}
$$

where

$$
T_2 = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right), \qquad T_1 = m\Omega\left(x\dot{y} - \dot{x}y\right), \qquad T_0 = \frac{1}{2}m\Omega^2\left(x^2 + y^2\right) \qquad \text{(f)}
$$

so that this is a nonnatural system. Inserting Eqs. (a) into Eq. (2.37) and choosing the origin O as the reference position, we obtain the potential energy

$$
\begin{aligned}
V &= \int_x^0 F_x(\xi)d\xi + \int_y^0 F_y(\eta)d\eta = -k_x\int_x^0(\xi + \epsilon_x\xi^3)d\xi - k_y\int_y^0(\eta + \epsilon_y\eta^3)d\eta \\
&= \frac{1}{2}\left[k_x\left(x^2 + \frac{1}{2}\epsilon_x x^4\right) + k_y\left(y^2 + \frac{1}{2}\epsilon_y y^4\right)\right] \qquad \text{(g)}
\end{aligned}
$$

Hence, using Eq. (2.155), the Lagrangian can be written in the form

$$
\begin{aligned}
L &= T - V = T_2 + T_1 - U = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) + m\Omega\left(x\dot{y} - \dot{x}y\right) \\
&\quad - \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x^2 + \left(k_y - m\Omega^2\right)y^2 + \frac{1}{2}k_x\epsilon_x x^4 + \frac{1}{2}k_y\epsilon_y y^4\right] \qquad \text{(h)}
\end{aligned}
$$

so that the dynamic potential can be identified as

$$U = V - T_0 = \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x^2 + \left(k_y - m\Omega^2\right)y^2 + \frac{1}{2}k_x\epsilon_x x^4 + \frac{1}{2}k_y\epsilon_y y^4\right] \tag{i}$$

The Rayleigh dissipation function is due to the damper alone and has the simple form

$$\mathcal{F} = \frac{1}{2}c\dot{x}^2 \tag{j}$$

The external damping forces can be derived from the dissipation function

$$\mathcal{F}' = \frac{1}{2}h\left(\dot{X}^2 + \dot{Y}^2\right) \tag{k}$$

To express $\dot{X}$ and $\dot{Y}$ in terms of x, y, $\dot{x}$ and $\dot{y}$, we refer to Fig. 4.1 and write

$$X = x\cos\Omega t - y\sin\Omega t, \qquad Y = x\sin\Omega t + y\cos\Omega t \tag{l}$$

so that, taking time derivatives, we have

$$\begin{aligned}\dot{X} &= \dot{x}\cos\Omega t - \dot{y}\sin\Omega t - \Omega\left(x\sin\Omega t + y\cos\Omega t\right)\\ &= \left(\dot{x} - \Omega y\right)\cos\Omega t - \left(\dot{y} + \Omega x\right)\sin\Omega t\\ \dot{Y} &= \dot{x}\sin\Omega t + \dot{y}\cos\Omega t + \Omega\left(x\cos\Omega t - y\sin\Omega t\right)\\ &= \left(\dot{x} - \Omega y\right)\sin\Omega t + \left(\dot{y} + \Omega x\right)\cos\Omega t\end{aligned} \tag{m}$$

and we observe that $\dot{X}$ and $\dot{Y}$ are merely the projections of $\mathbf{v}$ on the inertial axes X and Y, respectively. Inserting Eqs. (m) into Eq. (k), we obtain

$$\mathcal{F}' = \frac{1}{2}h\left[\left(\dot{x} - \Omega y\right)^2 + \left(\dot{y} + \Omega x\right)^2\right] \tag{n}$$

where we note that $\mathcal{F}'$ is really a circulatory function, which explains the notation. Finally, using Eq. (4.10), we obtain the modified Rayleigh's dissipation function

$$\mathcal{F}^* = \mathcal{F} + \mathcal{F}' = \frac{1}{2}c\dot{x}^2 + \frac{1}{2}h\left[\left(\dot{x} - \Omega y\right)^2 + \left(\dot{y} + \Omega x\right)^2\right] \tag{o}$$

At this point, we are in the position of deriving the desired equations of motion. Using $q_1 = x$ and $q_2 = y$ as generalized coordinates, Eqs. (4.12) become

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} + \frac{\partial \mathcal{F}^*}{\partial \dot{x}} &= 0\\ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{y}}\right) - \frac{\partial L}{\partial y} + \frac{\partial \mathcal{F}^*}{\partial \dot{y}} &= 0\end{aligned} \tag{p}$$

Introducing Eqs. (h) and (o) into Eqs. (p) and rearranging, we obtain Lagrange's equations of motion

$$\begin{aligned}m\ddot{x} + (c + h)\dot{x} - 2m\Omega\dot{y} + \left(k_x - m\Omega^2\right)x - h\Omega y + k_x\epsilon_x x^3 &= 0\\ m\ddot{y} + 2m\Omega\dot{x} + h\dot{y} + h\Omega x + \left(k_y - m\Omega^2\right)y + k_y\epsilon_y y^3 &= 0\end{aligned} \tag{q}$$

At this point, it may prove of interest to rewrite some of the expressions just derived in matrix notation. To this end, we introduce the configuration vector $\mathbf{q} = [x \; y]^T$, so that T_2, the first of Eqs. (f), can be reduced to the form given by Eq. (4.17a), where

$$M = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \qquad \text{(r)}$$

is a constant diagonal matrix. Moreover, T_1 as given by the second Eqs. (f) can be written in the form of Eq. (4.17b), in which

$$\mathbf{f} = F\mathbf{q} \qquad \text{(s)}$$

where

$$F = \begin{bmatrix} 0 & -m\Omega \\ m\Omega & 0 \end{bmatrix} \qquad \text{(t)}$$

Then, using Eqs. (4.17b), (s) and (t), we can write

$$\frac{d}{dt}\left(\frac{\partial T_1}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial T_1}{\partial \mathbf{q}} = G\dot{\mathbf{q}} \qquad \text{(u)}$$

in which

$$G = F - F^T = \begin{bmatrix} 0 & -2m\Omega \\ 2m\Omega & 0 \end{bmatrix} \qquad \text{(v)}$$

represents the gyroscopic matrix for the system at hand, and we note that gyroscopic effects, also known as Coriolis effects, arise when the relative velocity vector has a component normal to the angular velocity vector. We return to this subject in Sec. 4.4, in which the gyroscopic matrix is defined in a more general way. Finally, comparing Eqs. (j) and (4.19), we conclude that the damping matrix has the form

$$C = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{(w)}$$

and, comparing Eqs. (n) and (4.20), we obtain the equivalent damping matrix

$$C' = \begin{bmatrix} h & 0 \\ 0 & h \end{bmatrix} \qquad \text{(x)}$$

as well as the circulatory matrix

$$H = \begin{bmatrix} 0 & -\Omega h \\ \Omega h & 0 \end{bmatrix} \qquad \text{(y)}$$

4.2 GEOMETRIC REPRESENTATION OF MOTION. EQUILIBRIUM POINTS

Equations (4.12) represent a set of nonhomogeneous, generally nonlinear, ordinary differential equations of second order. Closed-form solutions of such sets of nonlinear equations do not exist in general. However, under certain circumstances Eqs. (4.12) admit special solutions. To discuss these special solutions, it is convenient to consider a geometric interpretation of the motion. It should be stated from the outset that a geometric description of the solution is essentially qualitative and its main purpose is to gain physical insights into the nature of the motion. The description can be used for quantitative results for systems defined by a single second-order differential equation only.

The solution of Eqs. (4.12) consists of the n generalized coordinates $q_k(t)(k = 1, 2, \ldots, n)$ and can be represented geometrically by conceiving of an n-dimensional

Euclidean space with axes q_k (Fig. 4.2), where the space is known as the *configuration space*. This is the same configuration space encountered in Sec. 2.10 in conjunction with Hamilton's principle, except that here the configuration space is in terms of generalized coordinates. Then, the solution of Eqs. (4.12) at any time t can be represented in the configuration space by the n-dimensional *configuration vector* $\mathbf{q}(t) = [q_1(t)\, q_2(t) \ldots q_n(t)]^T$. As time unfolds, the tip of the vector $\mathbf{q}(t)$ traces a curve in the configuration space known as the *dynamical path*, and we note that the time t does not appear explicitly in this representation but plays the role of a parameter only.

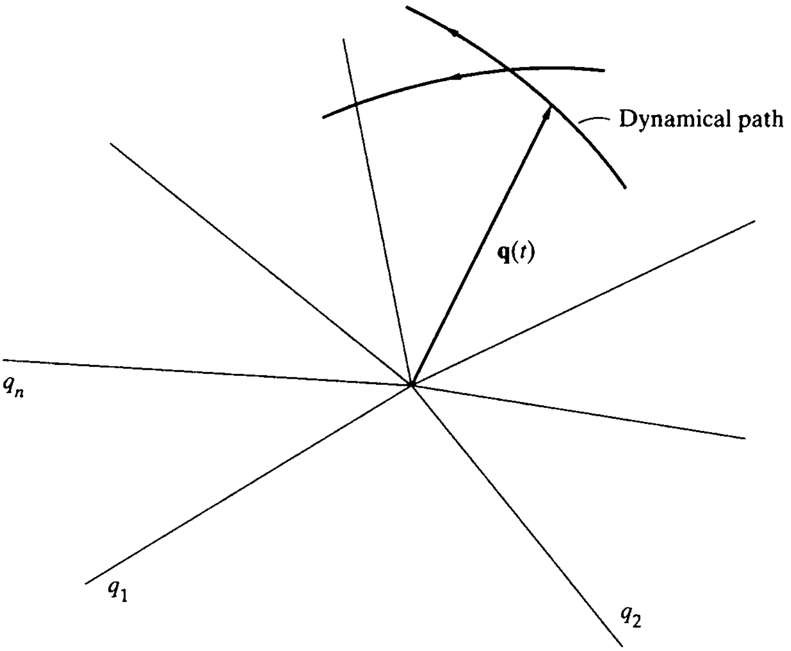


Figure 4.2 Dynamical paths in configuration space

The geometric representation in the configuration space has certain drawbacks. The main drawback is that it does not define the state of the system uniquely. Indeed, as shown in Fig. 4.2, two dynamical paths corresponding to different initial conditions can intersect. For a given point in the configuration space the motion can start in any direction with arbitrary initial velocity, so that the picture of the totality of paths is one of confusion. To render the geometric representation unique, it is necessary to provide additional information, which can be done by specifying the generalized velocities $\dot{q}_k(t)$. Then, regarding the generalized velocities as a set of auxiliary variables, the motion can be represented geometrically in a $2n$-dimensional Euclidean space defined by q_k and $\dot{q}_k$ and known as the *state space* (Fig. 4.3). Moreover, adjoining the identities $\dot{q}_k \equiv \dot{q}_k$ $k = 1, 2, \ldots, n)$ and transforming Lagrange's equations, Eqs. (4.12), into a set of n first-order differential equations, we obtain a set of $2n$

first-order equations known as *state equations*. Alternatively, we can specify the generalized momenta $p_k(t) = \partial L/\partial \dot{q}_k(t)$, where L is the Lagrangian, and represent the motion geometrically in a $2n$-dimensional Euclidean space defined by q_k and p_k and known as the *phase space*. Note that the generalized momenta were introduced as auxiliary variables in Sec. 2.12 in conjunction with Hamilton's equations. From a dynamical point of view, there is no material difference between the state space and phase space, as a point in either space defines the state of the system uniquely. In this text, we work primarily with the state space.

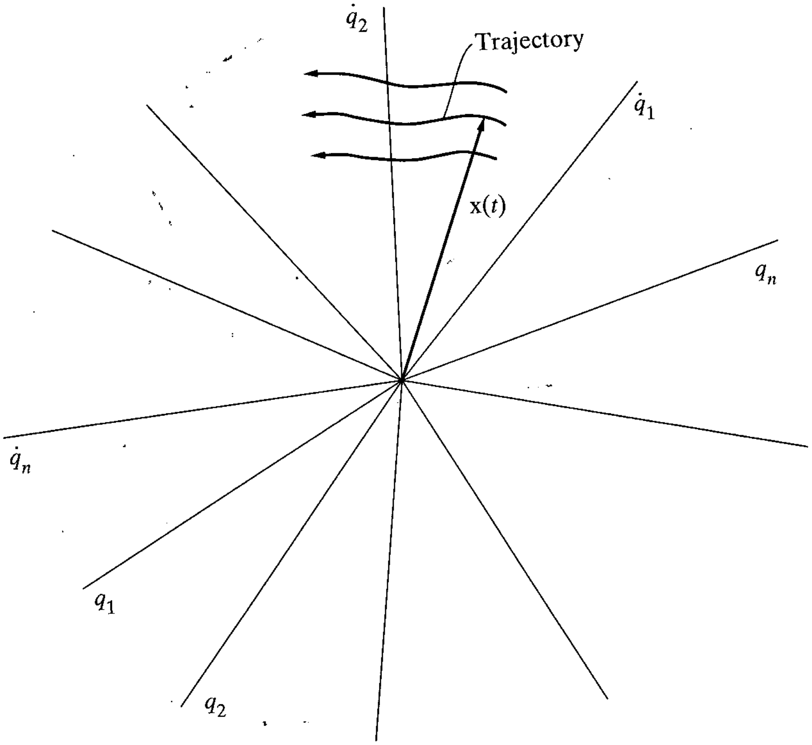


Figure 4.3 Trajectories in state space

The set of $2n$ variables q_k and $\dot{q}_k$ $(k = 1, 2, \ldots, n)$ defines the $2n$-dimensional vector $\mathbf{x} = \left[\mathbf{q}^T \; \dot{\mathbf{q}}^T\right]^T$, called the *state vector*. The tip of the state vector traces a curve in the state space known as a *trajectory*, which depicts geometrically the manner in which the solution of the state equations corresponding to a given initial state evolves with time. But, unlike dynamical paths in the configuration space, two trajectories in the state space corresponding to different initial conditions never intersect in finite time. In fact, as shown in Fig. 4.3, the motion of the dynamical system as represented by trajectories resembles the motion of a fluid, with each streamline representing the motion of the dynamical system corresponding to a given set of initial conditions. Hence, the geometric representation of the motion in the state space is unique, in the sense that through a given point in the state space passes a single trajectory. As a matter of interest, it should be pointed out that the configuration space can be regarded as the projection of the state space parallel to the $\dot{q}_i$ axes, which implies that a trajectory in the state space projects into a dynamical path in the configuration space.

There are certain special solutions in the state space of particular interest in vibrations. They represent constant solutions, $\mathbf{x} = \mathbf{x}_0 =$ constant, and are known as *equilibrium points*, or *singular points*. All other points are said to be *ordinary*, or *regular*. Trajectories never intersect at regular points, although they can intersect at equilibrium points. However, this cannot happen in finite time (Ref. 9). Indeed, if at all, equilibrium points are reached as $t \to \pm\infty$. The fact that the state is constant at an equilibrium point implies that $\mathbf{q} = \mathbf{q}_0 =$ constant and $\dot{\mathbf{q}} = \dot{\mathbf{q}}_0 = \mathbf{0}$. It further implies that $\ddot{\mathbf{q}} = \ddot{\mathbf{q}}_0 = \mathbf{0}$, which explains why such points are called equilibrium points. Because $\dot{\mathbf{q}}_0 = \mathbf{0}$, it follows that *all equilibrium points lie in the configuration space*. In the special case in which $\mathbf{q}_0$ is also zero, in which case the equilibrium point lies at the origin of the state space, the equilibrium point is said to be *trivial*. All other equilibrium points are *nontrivial*. In the case of second-order systems, the state space reduces to the state plane. In this case it is possible to analyze the motion quantitatively by plotting trajectories in the state plane corresponding to various initial conditions. For systems of order larger than two it is no longer feasible to plot trajectories, placing a quantitative analysis beyond reach. Note that constant solutions carry the implication that there are no time-dependent generalized applied forces at equilibrium points, $\mathbf{Q}(t) = \mathbf{0}$, although there can be constant forces and forces depending on the state.

The question remains as to how to determine the equilibrium points for a given dynamical system. Clearly, they must represent constant solutions and they must satisfy the homogeneous Lagrange's equations, Eq. (4.14) with $\mathbf{Q} = \mathbf{0}$. For a general nonnatural system, the kinetic energy consists of three parts, T_2, T_1 and T_0, as can be seen from Eq. (4.2). It is clear that the term T_2 does not contribute to constant solutions, as the pertinent derivatives in Eq. (4.14) will always contain $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$, which are zero at an equilibrium point. Similarly, the term T_1 does not contribute to a constant solution, because the derivatives indicated in Eq. (4.14) contain $\dot{\mathbf{q}}$. On the other hand, the term T_0 does contribute to a constant solution, as it depends on generalized coordinates alone. The same can be said about the potential energy V. The modified Rayleigh's dissipation function $\mathcal{F}^*$ is the sum of $\mathcal{F}$ and $\mathcal{F}'$, given by Eqs. (4.19) and (4.20), respectively. Using a similar argument to that used above, we conclude that $\mathcal{F}$ does not affect the equilibrium positions. On the other hand, $\mathcal{F}'$ does contribute to constant solutions, because the coefficients h_{ij} are generally constant. Hence, from Eq. (4.14) we conclude that the constant vector $\mathbf{q}_0$ must be a solution of the algebraic vector equation

$$\left(\frac{\partial U}{\partial \mathbf{q}} + \frac{\partial \mathcal{F}'}{\partial \dot{\mathbf{q}}}\right)\Bigg|_{\mathbf{q}=\mathbf{q}_0,\dot{\mathbf{q}}=\mathbf{0}} = \mathbf{0} \tag{4.26}$$

where $U = V - T_0$ is the dynamic potential, a function of the generalized coordinates alone. Equation (4.26) represents the *equilibrium equation*. But, $\partial\mathcal{F}'/\partial\dot{\mathbf{q}}$ is linear in the generalized coordinates. Hence, if the dynamic potential U is quadratic, then $\partial U/\partial \mathbf{q}$ is linear in the generalized coordinates. In this case, Eq. (4.26) represents a set of linear algebraic equations, and there is only one equilibrium point. If the equations are homogeneous, then the equilibrium is the trivial one and if the equations are nonhomogeneous, the equilibrium is nontrivial. If U contains terms of degree higher than two in the generalized coordinates, then there can be more than one equilibrium point.

In the absence of circulatory forces, $\mathcal{F}' = 0$, the equilibrium equation, Eq. (4.26), reduces to

$$\left.\frac{\partial U}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} = \mathbf{0} \tag{4.27}$$

which implies that *the dynamic potential has a stationary value at an equilibrium point.*

In the case of a natural system $T_0 = 0$, so that Eq. (4.27) reduces to the familiar equilibrium equation

$$\left.\frac{\partial V}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} = \mathbf{0} \tag{4.28}$$

which states that *for a natural system the potential energy has a stationary value at an equilibrium point.*

Example 4.2

Derive the equilibrium equations for the system of Example 4.1 by means of Eq. (4.26) and discuss the existence of equilibrium points.

Equation (4.26) calls for the dynamic potential U and the circulatory function $\mathcal{F}'$. Both functions are readily available from Example 4.1. Indeed, Eqs. (i) and (n) of Example 4.1 have the respective form

$$\begin{aligned} U &= V - T_0 \\ &= \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x^2 + \left(k_y - m\Omega^2\right)y^2 + \frac{1}{2}k_x\epsilon_x x^4 + \frac{1}{2}k_y\epsilon_y y^4\right] \end{aligned} \tag{a}$$

and

$$\mathcal{F}' = \frac{1}{2}h\left[(\dot{x} - \Omega y)^2 + (\dot{y} + \Omega x)^2\right] \tag{b}$$

so that, letting $q_1 = x$ and $q_2 = y$ and setting $\dot{x} = \dot{y} = 0$ and $x = x_0$, $y = y_0$, Eq. (4.26) yields the equilibrium equations

$$\begin{aligned} \left(k_x - m\Omega^2\right)x_0 + k_x\epsilon_x x_0^3 - h\Omega y_0 &= 0 \\ \left(k_y - m\Omega^2\right)y_0 + k_y\epsilon_y y_0^3 + h\Omega x_0 &= 0 \end{aligned} \tag{c}$$

In the absence of circulatory forces, $h = 0$, Eqs. (c) reduce to the two independent equilibrium equations

$$\begin{aligned} x_0(k_x - m\Omega^2 + k_x\epsilon_x x_0^2) &= 0 \\ y_0(k_y - m\Omega^2 + k_y\epsilon_y y_0^2) &= 0 \end{aligned} \tag{d}$$

From Eqs. (d), it is immediately obvious that the trivial solution $x_0 = y_0 = 0$ represents an equilibrium point, and this equilibrium point exists regardless of the values of the system parameters. On the other hand, the existence of nontrivial equilibrium points does depend on the values of the system parameters. Indeed, equating the expressions inside parentheses in Eqs. (d) to zero, we can write

$$x_0 = \pm\sqrt{\left(m\Omega^2 - k_x\right)/k_x\epsilon_x}, \qquad y_0 = \pm\sqrt{\left(m\Omega^2 - k_y\right)/k_y\epsilon_y} \tag{e}$$

For nontrivial equilibrium points to exist, the system parameters must be such that x_0 and/or y_0 are real. If both x_0 and y_0 are real, then there are nine equilibrium points, including the trivial one, as shown in Fig. 4.4a. This can happen in the four different ways: (i) $m\Omega^2 > k_x$, $\epsilon_x > 0$ and $m\Omega^2 > k_y$, $\epsilon_y = 0$, (ii) $m\Omega^2 > k_x$, $\epsilon_x > 0$ and $m\Omega^2 < k_y$, $\epsilon_y < 0$, (iii) $m\Omega^2 < k_x$, $\epsilon_x < 0$ and $m\Omega^2 > k_y$, $\epsilon_y > 0$ and (iv) $m\Omega^2 < k_x$, $\epsilon_x < 0$ and $m\Omega^2 < k_y$, $\epsilon_y < 0$. The nine equilibrium points of Fig. 4.4a represent the maximum number possible. Indeed, if the system parameters are such that x_0 is real and y_0 is imaginary, or x_0 is imaginary and y_0 is real, there are only three equilibrium points, as depicted in Fig. 4.4b, or Fig. 4.4c, respectively. Of course, when the parameters are such that both x_0 and y_0 are imaginary, then the only equilibrium point is the trivial one.

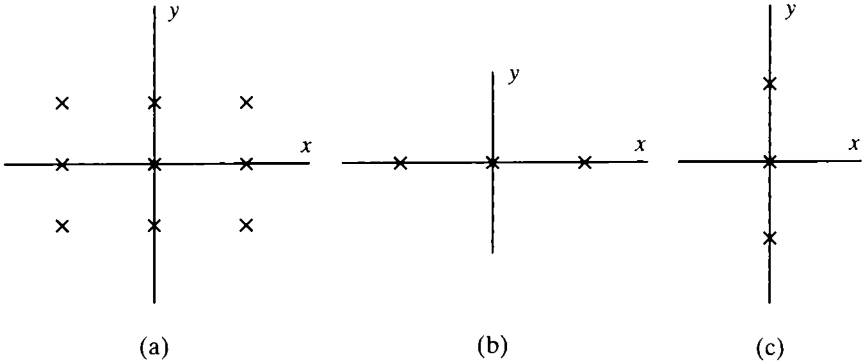


Figure 4.4 Equilibrium points **(a)** Both x_0 and y_0 real **(b)** x_0 real and y_0 imaginary **(c)** x_0 imaginary and y_0 real

Next, we consider the case in which $h \neq 0$. To determine the equilibrium points, we first solve the first of Eqs. (c) for y_0 and write

$$y_0 = \frac{x_0}{h\Omega}\left(k_x - m\Omega^2 + k_x\epsilon_x x_0^2\right) \tag{f}$$

Then, inserting Eq. (f) into the second of Eqs. (c) and rearranging, we obtain an equation for x_0 alone, or

$$\begin{aligned} x_0\Big\{ & k_y\epsilon_y\left(k_x\epsilon_x\right)^3 x_0^8 + 3k_y\epsilon_y\left(k_x\epsilon_x\right)^2\left(k_x - m\Omega^2\right)x_0^6 \\ & + 3k_y\epsilon_y k_x\epsilon_x\left(k_x - m\Omega^2\right)^2 x_0^4 + \Big[k_x\epsilon_x\left(k_y - m\Omega^2\right)(h\Omega)^2 \\ & + k_y\epsilon_y\left(k_x - m\Omega^2\right)^3\Big]x_0^2 + \left(k_x - m\Omega^2\right)\left(k_y - m\Omega^2\right)(h\Omega)^2 + (h\Omega)^4\Big\} = 0 \end{aligned} \tag{g}$$

It is clear from Eqs. (f) and (g) that, as in the case $h = 0$, the trivial solution $x_0 = y_0 = 0$ is an equilibrium point, independently of the system parameters. To examine the existence of nontrivial equilibrium points, we observe that the expression inside braces in Eq. (g) represents an eighth-degree polynomial in x_0. For nontrivial equilibrium points to exist, it is necessary that the polynomial admit some real roots x_0. Then, inserting these real values of x_0 into Eq. (f), we obtain corresponding real values for y_0, thus defining the equilibrium points. We note that, because the eighth-degree polynomial contains even powers of x_0 only, if x_0 is a root, then $-x_0$ is also a root, and the same can be said about y_0, as can be concluded from Eq. (f). It can be verified that, including the trivial one, the number of equilibrium points takes one of the values $1, 3, 9, 15, \ldots, 81$, depending on the system parameters. The nontrivial equilibrium points can be determined only numerically.

It is clear from Eqs. (c) that in the linear case, in which $\epsilon_x = \epsilon_y = 0$, the origin is the only equilibrium point. This confirms the statement made earlier that linear systems admit only one equilibrium point. In the case at hand, Eqs. (c) with $\epsilon_x = \epsilon_y = 0$ are homogeneous, so that the equilibrium point is the trivial one.

4.3 STABILITY OF EQUILIBRIUM POINTS. THE LIAPUNOV DIRECT METHOD

A question of particular interest in vibrations is whether a slight perturbation of a dynamical system from an equilibrium state will produce a motion remaining in the neighborhood of the equilibrium point or a motion tending to leave that neighborhood. This question arises mainly in the case of homogeneous systems, namely, systems for which $\mathbf{Q} = \mathbf{0}$ in Eq. (4.14), and can be rendered more precise by invoking one of the *Liapunov stability definitions*. To this end, we consider the case in which a given trajectory $\mathbf{x}(t)$ in the state space has the value $\mathbf{x}(t_0) = \mathbf{x}_\delta$ at time $t_0 > 0$, where $\mathbf{x}_\delta$ lies inside the spherical region $\|\mathbf{x} - \mathbf{x}_0\| < \delta$, in which $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T\mathbf{v}}$ denotes the *Euclidean norm* of the vector $\mathbf{v}$, i.e., the magnitude of the vector. Then, the Liapunov stability definitions can be stated as follows:

1. The equilibrium point $\mathbf{x}_0$ is *stable in the sense of Liapunov* if for any arbitrary positive ϵ and time t_0 there exists a $\delta = \delta(\epsilon, t_0) > 0$ such that if the inequality
$$\|\mathbf{x}_\delta - \mathbf{x}_0\| < \delta \tag{4.29}$$
is satisfied, then the inequality
$$\|\mathbf{x}(t) - \mathbf{x}_0\| < \epsilon, \qquad t_0 \le t < \infty \tag{4.30}$$
is implied. If δ is independent of t_0, the stability is said to be *uniform*.
2. The equilibrium point $\mathbf{x_0}$ is *asymptotically stable* if it is Liapunov stable and in addition
$$\lim_{t\to\infty} \|\mathbf{x}(t) - \mathbf{x}_0\| = 0 \tag{4.31}$$
3. The equilibrium point $\mathbf{x}_0$ is *unstable* if for any arbitrarily small δ and any time t_0 such that
$$\|\mathbf{x}_\delta - \mathbf{x}_0\| < \delta \tag{4.32}$$

the motion at some finite time t_1 satisfies

$$\|\mathbf{x}(t_1) - \mathbf{x}_0\| = \epsilon, \qquad t_1 > t_0 \tag{4.33}$$

This is equivalent to the statement that a motion initiated inside the open sphere of radius δ and with the center at $\mathbf{x}_0$ reaches the boundary of the sphere of radius ϵ in finite time. The implication is that the trajectory reaches the boundary on its way out of the spherical region $\|\mathbf{x} - \mathbf{x}_0\| < \epsilon$.

The three possible cases are displayed in Fig. 4.5.

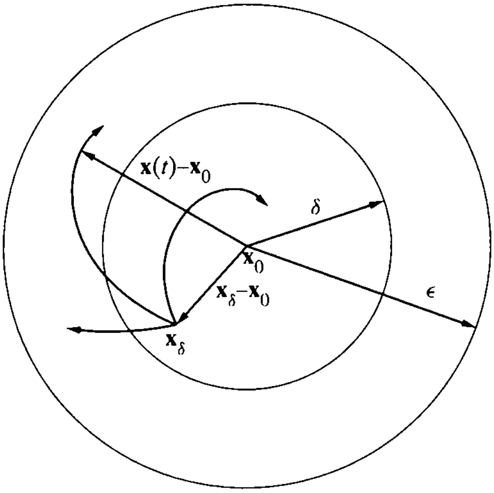

$\mathbf{x}(t)-\mathbf{x}_0$
δ
$\mathbf{x}_0$
$\mathbf{x}_\delta-\mathbf{x}_0$
ϵ
$\mathbf{x}_\delta$

Figure 4.5 Stable, asymptotically stable and unstable trajectories in the state space

A stability analysis based on the solution of the equations of motion is not always feasible, as such solutions are not generally available, particularly for nonlinear systems. The *Liapunov direct method*, also known as the *Liapunov second method*, represents an approach to the problem of stability of dynamical system not requiring the solution of the differential equations of motion. The method consists of devising for the dynamical system a suitable scalar testing function defined in the state space, which can be used in conjunction with its total time derivative in an attempt to determine the stability characteristics of equilibrium points. Such testing functions, if they can be found, are referred to as *Liapunov functions* for the system. When a Liapunov function can be devised for the system, stability conclusions can be reached on the basis of the sign properties of the function and its total time derivative, where the latter is evaluated along a trajectory of the system. The evaluation along a trajectory does not imply that a solution of the differential equations must be produced but only that the differential equations are used in calculating the time derivative of the Liapunov function. The Liapunov direct method can be used for both linear and nonlinear systems, and for problems from a variety of fields. There is no unique Liapunov function for a given system, and there is a large degree of flexibility in the selection of a Liapunov function. In the problems of interest to our study, the choice of a Liapunov function can often be made on a rational basis.

We consider a dynamical system described by a set of $2n$ *state equations*, i.e., $2n$ first-order differential equations having the vector form

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{4.34}$$

where $\mathbf{x}$ is the state vector and $\mathbf{f}$ is a continuous generally nonlinear vector function of $\mathbf{x}$ in the spherical region $D_\epsilon : \|\mathbf{x}\| < \epsilon$ with the center at the origin of the state space and of radius ϵ, where ϵ is a positive constant. It should be noted, in passing, that the upper half of $\mathbf{f}$ is the n-dimensional null vector. We assume that the origin of the state space is an equilibrium point, so that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$, and propose to investigate the stability of a trajectory of Eq. (4.34) initiated in the region $D_\delta : \|\mathbf{x}\| < \delta$, where $\delta < \epsilon$. To this end, we consider a real continuous scalar function $\mathcal{V}(\mathbf{x})$ possessing continuous first partial derivatives with respect to the state variables x_i $(i = 1, 2, \ldots, 2n)$ in D_ϵ and vanishing at the origin of the state space, $\mathcal{V}(\mathbf{0}) = 0$. For such a function, we introduce the following sign definitions:

1. The function $\mathcal{V}(\mathbf{x})$ is said to be *positive (negative) definite* in D_ϵ if $\mathcal{V}(\mathbf{x}) > 0$ (< 0) for all $\mathbf{x} \neq \mathbf{0}$.
2. The function $\mathcal{V}(\mathbf{x})$ is said to be *positive (negative) semidefinite* in D_ϵ if $\mathcal{V}(\mathbf{x}) \geq 0$ (≤ 0) and it can vanish also for some $\mathbf{x} \neq \mathbf{0}$.
3. The function $\mathcal{V}(\mathbf{x})$ is said to be *sign-variable* if it can assume both positive and negative values in D_ϵ regardless of how small ϵ is.

Figures 4.6a,b and c provide a geometric interpretation for positive definite, positive semidefinite and sign-variable functions, respectively. The figures resemble a bowl, a cylinder and a saddle, respectively.

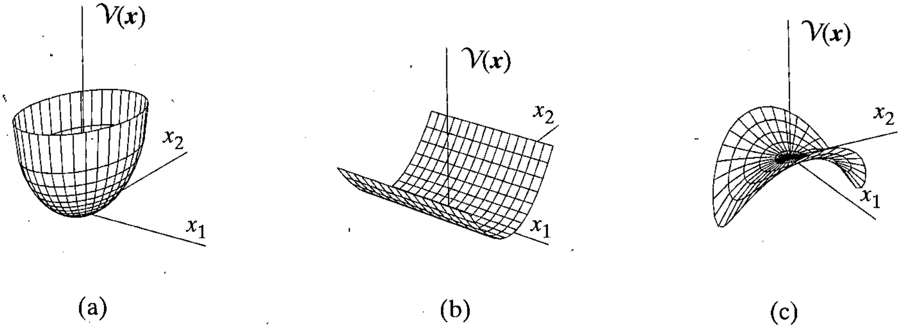


Figure 4.6 **(a)** Positive definite function **(b)** Positive semidefinite function **(c)** Sign-variable function

A case of special interest is that in which $\mathcal{V}$ has the quadratic form

$$\mathcal{V} = \sum_{i=1}^{2n} \sum_{j=1}^{2n} b_{ij} x_i x_j = \mathbf{x}^T B \mathbf{x} \tag{4.35}$$

where $B = [b_{ij}]$ is a real symmetric matrix. In this case, the sign properties of $\mathcal{V}$ defined above carry over to the matrix B. Conversely, the sign properties of B carry over to the function $\mathcal{V}$. In fact, it is often more convenient to ascertain the sign properties of $\mathcal{V}$ by means of the matrix B. Indeed, assuming that the linear transformation

$$\mathbf{x} = T\mathbf{z} \tag{4.36}$$

diagonalizes B, so that

$$\mathcal{V} = \mathbf{z}^T T^T BT\mathbf{z} = \mathbf{z}^T D\mathbf{z} \tag{4.37}$$

where

$$D = T^T BT \tag{4.38}$$

is a diagonal real matrix, the following statements are true: B is positive (negative) definite if all the diagonal elements of D are positive (negative), B is positive (negative) semidefinite if the diagonal elements of D are nonnegative (nonpositive), i.e., some are positive (negative) and the remaining ones are zero, and B is sign-variable if the diagonal elements of D are of both signs. Perhaps a more ready way of checking whether B is positive definite or not is by means of *Sylvester's criterion* (Ref. 9), which can be stated as follows: *The necessary and sufficient conditions for the matrix B to be positive definite is for all the principal minor determinants of B to be positive.* Mathematically, the conditions can be written in the form

$$\det[b_{rs}] > 0, \qquad r, s = 1, 2, \dots, k;\ k = 1, 2, \dots, 2n \tag{4.39}$$

The total time derivative of $\mathcal{V}$ with respect to time evaluated along a trajectory of the system is obtained by using Eq. (4.34) and writing

$$\dot{\mathcal{V}} = \frac{d\mathcal{V}}{dt} = \sum_{i=1}^{2n} \frac{\partial \mathcal{V}}{\partial x_i}\frac{dx_i}{dt} = \dot{\mathbf{x}}^T \nabla \mathcal{V} = \mathbf{f}^T \nabla \mathcal{V} \tag{4.40}$$

where $\nabla \mathcal{V}$ is the gradient of $\mathcal{V}$.

There are several theorems due to Liapunov, as well as Chetayev and Krasovskii, concerned with stability, asymptotic stability and instability of equilibrium points of systems of the type given by Eq. (4.34). The idea is that if a Liapunov function $\mathcal{V}$ can be found so that $\mathcal{V}$ and $\dot{\mathcal{V}}$ satisfy the conditions prescribed by one of the theorems, then stability, asymptotic stability or instability can be established. On the other hand, the fact that a Liapunov function cannot be found simply means that stability, asymptotic stability or instability cannot be demonstrated, and it does not mean that none of the three motion characteristics exists. The most important Liapunov stability theorems can be stated as follows:

Theorem 1. If there exists for system (4.34) a positive (negative) definite function $\mathcal{V}(\mathbf{x})$ whose total time derivative $\dot{\mathcal{V}}(\mathbf{x})$ is negative (positive) semidefinite along every trajectory of (4.34), then the trivial solution is stable.

Theorem 2. If there exists for system (4.34) a positive (negative) definite function $\mathcal{V}(\mathbf{x})$ whose total time derivative $\dot{\mathcal{V}}(\mathbf{x})$ is negative (positive) definite along every trajectory of (4.34), then the trivial solution is asymptotically stable.

Theorem 3. If there exists for system (4.34) a function $\mathcal{V}(\mathbf{x})$ whose total time derivative $\dot{\mathcal{V}}(\mathbf{x})$ is positive (negative) definite along every trajectory of (4.34) and the function itself can assume positive (negative) values for arbitrarily small values of $\mathbf{x}$, then the trivial solution is unstable.

A geometric interpretation in the form of trajectories illustrating Liapunov's stability theorems, Theorems 1 and 2, and Liapunov's instability theorem, Theorem 3, is provided in Figs. 4.7a and 4.7b, respectively. A more extensive discussion of the Liapunov direct method, including proofs of the theorems, can be found in Ref. 9.

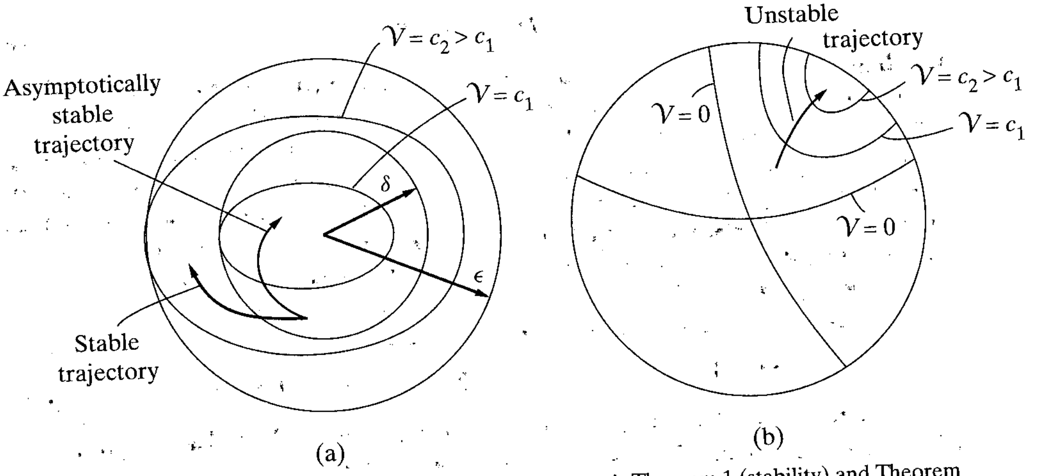


Figure 4.7 **(a)** Geometric interpretation of Liapunov's Theorem 1 (stability) and Theorem 2 (asymptotic stability) **(b)** Geometric interpretation of Liapunov's Theorem 3 (instability)

Liapunov's theorems are unduly restrictive and some of the conditions can be relaxed in a meaningful way. A generalization of Liapunov's instability theorem, known as Chetayev's instability theorem, essentially states that $\dot{\mathcal{V}}$ need not be positive (negative) in the entire neighborhood of the origin but only in the subdomain in which $\mathcal{V}$ is positive (negative). A generalization of Liapunov's asymptotic stability theorem proposed by Barbasin and Krasovskii states that $\dot{\mathcal{V}}$ need be only negative (positive) semidefinite, provided $\dot{\mathcal{V}}$ does not reduce to zero and stays zero for all subsequent times for $\mathbf{x} \neq \mathbf{0}$. A similar generalization due to Krasovskii exists for Liapunov's instability theorem.

At this point, we return to the dynamical system of Sec. 4.1. In Sec. 2.12, we derived an expression for the time derivative of the Hamiltonian, Eq. (2.160), for the case in which the Hamiltonian does not depend explicitly on time, and established the circumstances under which the Hamiltonian is conserved. Considering Eq. (4.23) and letting $\mathbf{Q} = \mathbf{0}$, Eq. (2.160) can be rewritten in the matrix form

$$\dot{\mathcal{H}} = \frac{d\mathcal{H}}{dt} = \dot{\mathbf{q}}^T \left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} \right] = \dot{\mathbf{q}}^T \left(\mathbf{Q} - C^*\dot{\mathbf{q}} - H\mathbf{q}\right) = -\dot{\mathbf{q}}^T \left(C^*\dot{\mathbf{q}} + H\mathbf{q}\right) \tag{4.41}$$

Moreover, we recall Eq. (2.157) stating that the Hamiltonian for a nonnatural system has the expression

$$\mathcal{H} = T_2 + U \tag{4.42}$$

and note that the various quantities entering into Eqs. (4.41) and (4.42) were defined in Sec. 4.1.

The fact that $\mathcal{H}$ and $\dot{\mathcal{H}}$ contain functions with known characteristics suggests the possibility that the Hamiltonian $\mathcal{H}$ can serve as a Liapunov function. The function T_2 is positive definite on physical grounds, so that the nature of $\mathcal{H}$ depends on the dynamic potential U. It is clear that if U is negative definite, negative semidefinite or sign-variable, then $\mathcal{H}$ is sign-variable. If in the subdomain of D_ϵ in which $\mathcal{H}$ is negative $\dot{\mathcal{H}}$ is also negative, then according to Chetayev's instability theorem the equilibrium point is *unstable*. If U is nonnegative, then $\mathcal{H}$ is positive definite. Because of the presence of circulatory forces, $\dot{\mathcal{H}}$ can assume positive values. Hence, according to Liapunov's instability theorem, the equilibrium point is *unstable*.

The situation is quite different in the absence of circulatory forces, $C' = H = 0$, in which case Eq. (4.41) reduces to

$$\dot{\mathcal{H}} = -\dot{\mathbf{q}}^T C \dot{\mathbf{q}} \tag{4.43}$$

In this case, sharper conclusions can be drawn. Indeed, $\dot{\mathcal{H}}$ is nonpositive. Moreover, we consider the case in which $\dot{\mathcal{H}}$ does not reduce to zero and stays zero for $\mathbf{x} \neq \mathbf{0}$. This is the case of *pervasive damping* (Ref. 13), defined as the case in which the damping matrix C couples all the equations of motion. Then, if U is nonnegative, $\mathcal{H}$ is positive definite and according to the Barbasin and Krasovskii theorem, the equilibrium point is *asymptotically stable*. On the other hand, if U is nonpositive, so that $\mathcal{H}$ can take negative values, the equilibrium point is *unstable*, according to Krasovskii's theorem.

Finally, we consider the case in which both the viscous damping and circulatory forces are absent, $C = C' = H = 0$. In this case, Eq. (4.41) reduces to

$$\dot{\mathcal{H}} = 0 \tag{4.44}$$

Then, if U is positive definite, $\mathcal{H}$ is positive definite and, according to Liapunov's first theorem, the equilibrium point is *stable*. On the other hand, if U is nonpositive, $\mathcal{H}$ is sign-variable and the equilibrium point is *unstable*, according to Krasovskii's theorem. Note that in this case, Eq. (4.44) yields

$$\mathcal{H} = \text{constant} \tag{4.45}$$

Equation (4.45) represents a *conservation principle* stating that *in the absence of applied forces, viscous damping forces and circulatory forces the Hamiltonian is conserved.*

It should be pointed out that gyroscopic forces have no effect on the above results.

In the case of *natural systems*, $T_2 = T$, $T_1 = 0$, the Hamiltonian reduces to the total energy E, or

$$\mathcal{H} = E = T + V \tag{4.46}$$

It follows that all the above results remain valid, provided $\mathcal{H}$ is replaced by E and U by V. In particular, the conservation of the Hamiltonian principle defined by Eq. (4.45) reduces to the *conservation of energy principle*

$$E = T + V = \text{constant} \tag{4.47}$$

Hence, the conservation of the Hamiltonian can be regarded as a generalization to nonnatural systems of the conservation of energy for natural systems. Note that the two conservation principles, Eqs. (4.45) and (4.47), were encountered in Sec. 2.13.

The stability statement following Eq. (4.44) can be stated as the following theorem: *If the dynamic potential U has a minimum at an equilibrium point, then the equilibrium point is stable.* The theorem represents a generalization to nonnatural systems of *Lagrange's theorem* for natural systems: *If the potential energy V has a minimum at an equilibrium point, then the equilibrium point is stable.*

It was pointed out earlier that the Liapunov function is not necessarily unique for a given system. In fact, it does not even need to have physical meaning, in contrast to the present case.

Example 4.3

Consider the system of Examples 4.1 and 4.2, assume that the circulatory forces are absent, $h = 0$, identify the equilibrium points and determine their stability by means of the Liapunov direct method for the case in which $m\Omega^2 > k_x$, $m\Omega^2 > k_y$, $\epsilon_x > 0$ and $\epsilon_y > 0$.

From Eqs. (e) of Example 4.2, we conclude that for the given parameters, x_0 and y_0 are real, so that the system admits the nine equilibrium points shown in Fig. 4.4a. We denote the equilibrium points as follows:

$$\begin{aligned}
&E_1 : x_0 = 0, y_0 = 0 \\
&E_2 : x_0 = 0, \qquad y_0 = \sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y} \\
&E_3 : x_0 = 0, \qquad y_0 = -\sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y} \\
&E_4 : x_0 = \sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = 0 \\
&E_5 : x_0 = -\sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = 0 \\
&E_6 : x_0 = \sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = \sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y} \\
&E_7 : x_0 = \sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = -\sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y} \\
&E_8 : x_0 = -\sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = \sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y} \\
&E_9 : x_0 = -\sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}, \qquad y_0 = -\sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y}
\end{aligned} \tag{a}$$

Although geometrically there are nine equilibrium points, the points can be divided into groups with the same stability characteristics.

Next, we wish to check the stability of the equilibrium points. To this end, we consider the Hamiltonian $\mathcal{H}$ as a candidate for a Liapunov function. For the case at hand, insertion of the first of Eqs. (f) and Eq. (i) of Example 4.1 into Eq. (4.42) yields

$$\begin{aligned}\mathcal{H} &= T_2 + U \\ &= \frac{1}{2}m\left(\dot{x}^2+\dot{y}^2\right) + \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x^2 + \left(k_y - m\Omega^2\right)y^2 + \frac{1}{2}k_x\epsilon_x x^4 + \frac{1}{2}k_y\epsilon_y y^4\right]\end{aligned} \tag{b}$$

Moreover, introducing Eq. (j) of Example 4.1 into Eq. (4.43) and recalling Eq. (4.19), we obtain

$$\dot{\mathcal{H}} = -2\mathcal{F} = -c\dot{x}^2 \tag{c}$$

But, from the equations of motion, Eqs. (q) of Example 4.1 with $h = 0$, we conclude that damping is pervasive, because the coupling is such that $\dot{x}$ cannot be zero without $\dot{y}$ being also zero at the same time. Hence, the stability depends on the sign properties of $\mathcal{H}$. If $\mathcal{H}$ is positive definite, then the equilibrium point is asymptotically stable. On the other hand, if $\mathcal{H}$ can take negative values in the neighborhood of the equilibrium point, then the equilibrium point is unstable. Because $\dot{\mathcal{H}} \leq 0$ and damping is pervasive, mere stability is not possible.

Before we check the stability of the individual equilibrium points, we shall find it convenient to introduce a transformation of coordinates translating the origin of the state space to an equilibrium point. This transformation has the simple form

$$x(t) = x_0 + x_1(t), \qquad y(t) = y_0 + y_1(t) \tag{d}$$

so that the new state space is defined by $x_1, y_1, \dot{x}_1$ and $\dot{y}_1$ and its origin is at x_0, y_0. We can expand the Hamiltonian in the neighborhood of a given equilibrium point E_i by inserting Eqs. (d) into the Hamiltonian, Eq. (b), and using the values of x_0 and y_0 corresponding to $E_i (i = 1, 2, \ldots, 9)$. Any constant in $\mathcal{H}$ can be ignored as irrelevant.

In the case of the equilibrium point E_1, x_0 and y_0 are zero and the Hamiltonian, Eq. (b), reduces to

$$\begin{aligned}\mathcal{H} &= \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) \\ &+ \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x_1^2 + \left(k_y - m\Omega^2\right)y_1^2 + \frac{1}{2}k_x\epsilon_x x_1^4 + \frac{1}{2}k_y\epsilon_y y_1^4\right]\end{aligned} \tag{e}$$

Because $k_x - m\Omega^2 < 0$ and $k_y - m\Omega^2 < 0$, $\mathcal{H}$ can take negative values in the neighborhood of the origin, so that *the equilibrium point E_1 is unstable.*

For the equilibrium points E_2 and E_3, $x_0 = 0$ and $y_0 = \pm\sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y}$ and the Hamiltonian becomes

$$\mathcal{H} = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \frac{1}{2}\left[\left(k_x - m\Omega^2\right)x_1^2 + \frac{1}{2}k_x\epsilon_x x_1^4\right] + \frac{1}{4}k_y\epsilon_y\left(2y_0 + y_1\right)^2 y_1^2 \quad \text{(f)}$$

But, $k_x - m\Omega^2 < 0$, from which we conclude that $\mathcal{H}$ can take negative values in the neighborhood of E_2 and E_3, so that *E_2 and E_3 are unstable*. Using similar arguments, it is easy to verify that *E_4 and E_5 are also unstable*.

For the equilibrium points E_6, E_7, E_8 and E_9, $x_0 = \pm\sqrt{(m\Omega^2 - k_x)/k_x\epsilon_x}$ and $y_0 = \pm\sqrt{(m\Omega^2 - k_y)/k_y\epsilon_y}$ and the Hamiltonian takes the form

$$\mathcal{H} = \frac{1}{4}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \frac{1}{4}\left[k_x\epsilon_x\left(2x_0 + x_1\right)^2 x_1^2 + k_y\epsilon_y\left(2y_0 + y_1\right)^2 y_1^2\right] \quad \text{(g)}$$

Clearly, the Hamiltonian is positive definite, so that *the equilibrium points E_6, E_7, E_8 and E_9 are all asymptotically stable*.

4.4 LINEARIZATION ABOUT EQUILIBRIUM POINTS

As indicated in Sec. 4.2, general closed-form solutions of sets of nonlinear differential equations do not exist. On the other hand, special solutions do exist in the form of equilibrium points, defined as constant solutions in the state space, $\mathbf{x} = \mathbf{x}_0$. A problem of considerable interest is concerned with the motion in a *small neighborhood* of equilibrium points, where the neighborhood is defined as

$$\|\mathbf{x} - \mathbf{x}_0\| = \left[(\mathbf{x} - \mathbf{x}_0)^T (\mathbf{x} - \mathbf{x}_0)\right]^{1/2} < \epsilon \quad (4.48)$$

in which ϵ is a small positive constant and $\|\mathbf{x} - \mathbf{x}_0\|$ is the *Euclidean norm* of $\mathbf{x} - \mathbf{x}_0$. Inequality (4.48) defines a $2n$-dimensional *spherical region* in the state space with the *origin* at $\mathbf{x} = \mathbf{x}_0$ and of *radius* ϵ.

The stipulation that the motions remain in the neighborhood $\|\mathbf{x} - \mathbf{x}_0\| < \epsilon$ of $\mathbf{x}_0$ is commonly known as the *small motions assumption*, and it carries the implication that in that neighborhood the system behaves as if it were linear. This implies further that the behavior of the system in the neighborhood of $\mathbf{x} = \mathbf{x}_0$ is governed by a special version of the equations of motion, obtained by expanding Eq. (4.14) about $\mathbf{x} = \mathbf{x}_0$ and retaining the linear terms in $\mathbf{x} - \mathbf{x}_0$ alone. The resulting equations, known as the *linearized Lagrange's equations of motion*, play a very important role in vibrations.

To derive the linearized equations, it is convenient to introduce the notation

$$\mathbf{x}(t) = \mathbf{x}_0 + \mathbf{x}_1(t) \quad (4.49)$$

where $\mathbf{x}_0$ is the equilibrium state vector and $\mathbf{x}_1(t)$ is a vector of *small perturbations* in the state variables from equilibrium. Before we proceed with the linearization of the equations of motion, we recognize that retention of linear terms in $\mathbf{x}_1$ alone in the equations of motion requires that terms higher than quadratic in $\mathbf{x}_1$ in T, V and $\mathcal{F}^*$ be ignored. To this end, we will find it expedient to separate the state vector

$\mathbf{x} = \left[\mathbf{q}^T \ \dot{\mathbf{q}}^T\right]^T$ into the generalized displacement vector $\mathbf{q}$ and generalized velocity vector $\dot{\mathbf{q}}$, so that Eq. (4.49) can be rewritten as

$$\mathbf{q}(t) = \mathbf{q}_0 + \mathbf{q}_1(t), \qquad \dot{\mathbf{q}}(t) = \dot{\mathbf{q}}_1(t) \tag{4.50}$$

where $\mathbf{q}_0$ is the constant equilibrium configuration vector and $\mathbf{q}_1$ and $\dot{\mathbf{q}}_1$ are perturbation vectors.

We propose to derive the linearized equations for a nonnatural system, so that the kinetic energy is given by Eq. (4.2). From Eq. (4.3), the quadratic part in the generalized velocities has the form

$$T_2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\dot{q}_i\dot{q}_j = \frac{1}{2}\dot{\mathbf{q}}^T M\dot{\mathbf{q}} \cong \frac{1}{2}\dot{\mathbf{q}}_1^T M\dot{\mathbf{q}}_1 \tag{4.51}$$

in which

$$M = \left[m_{ij}\right] = M(\mathbf{q}_0) = M^T = \text{constant} \tag{4.52}$$

is the symmetric *mass matrix*, or *inertia matrix*. The entries $m_{ij}(\mathbf{q}_0) = m_{ji}(\mathbf{q}_0)$ are known as the *mass coefficients*, or *inertia coefficients*. On the other hand, from Eq. (4.4), the linear part in the generalized velocities can be expressed as

$$T_1 = \sum_{j=1}^{n} f_j\dot{q}_j = \dot{\mathbf{q}}^T\mathbf{f} \cong \dot{\mathbf{q}}_1^T\mathbf{f}_0 + \dot{\mathbf{q}}_1^T\mathbf{f}_1 \tag{4.53}$$

where $\mathbf{f}_0$ is a constant vector and $\mathbf{f}_1$ is given by

$$\mathbf{f}_1 = F\mathbf{q}_1 \tag{4.54}$$

in which

$$F = \left[f_{ij}\right] = \left[\partial f_i/\partial q_j\right]\Big|_{\mathbf{q}=\mathbf{q}_0} = \text{constant} \tag{4.55}$$

is a constant matrix. Hence, inserting Eq. (4.54) into Eq. (4.53), we have

$$T_1 \cong \dot{\mathbf{q}}_1^T\mathbf{f}_0 + \dot{\mathbf{q}}_1^T F\mathbf{q}_1 \tag{4.56}$$

Next, we wish to remove the higher-order terms in T_0, where T_0 depends on generalized coordinates alone. However, the potential energy also depends on generalized coordinates alone, so that it is more efficient to treat T_0 and V together. To this end, we recall Eq. (2.156) defining the dynamic potential U as $V - T_0$ and expand the following Taylor's series

$$\begin{aligned} U(q_1, q_2, \ldots, q_n) &\cong U(q_{01}, q_{02}, \ldots, q_{0n}) + \sum_{i=1}^{n}\frac{\partial U}{\partial q_i}\Bigg|_{\mathbf{q}=\mathbf{q}_0} q_{1i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 U}{\partial q_i\partial q_j}\Bigg|_{\mathbf{q}=\mathbf{q}_0} q_{1i}q_{1j} \\ &= U(\mathbf{q}_0) + \mathbf{q}_1^T\frac{\partial U}{\partial \mathbf{q}}\Bigg|_{\mathbf{q}=\mathbf{q}_0} + \frac{1}{2}\mathbf{q}_1^T K\mathbf{q}_1 \end{aligned} \tag{4.57}$$

where $U(\mathbf{q}_0)$ is a constant scalar and

$$K = [k_{ij}] = \left[\frac{\partial^2 U}{\partial q_i \partial q_j}\right]_{\mathbf{q}=\mathbf{q}_0} = K^T = \text{constant} \tag{4.58}$$

is the symmetric *stiffness matrix*. The entries k_{ij}, known as *stiffness coefficients*, can be divided into two types, the first arising from the potential energy V and called *elastic stiffness coefficients* and the second arising from the centrifugal term T_0 in the kinetic energy and referred to as *geometric stiffness coefficients*. Finally, assuming that the coefficients c^*_{ij} and h_{ij} are constant and using Eqs. (4.50), the modified Rayleigh's dissipation function, Eq. (4.10), can be rewritten in the form

$$\begin{aligned}\mathcal{F}^* &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} c^*_{ij}\dot{q}_i\dot{q}_j + \sum_{i=1}^{n}\sum_{j=1}^{n} h_{ij}\dot{q}_i q_j \\ &= \frac{1}{2}\dot{\mathbf{q}}^T C^*\dot{\mathbf{q}} + \dot{\mathbf{q}}^T H\mathbf{q} = \frac{1}{2}\dot{\mathbf{q}}_1^T C^*\dot{\mathbf{q}}_1 + \dot{\mathbf{q}}_1^T H\mathbf{q}_0 + \dot{\mathbf{q}}_1^T H\mathbf{q}_1\end{aligned} \tag{4.59}$$

where

$$C^* = [c^*_{ij}] = C^{*T} = \text{constant}, \qquad H = [h_{ij}] = -H^T = \text{constant} \tag{4.60}$$

in which C^* is the symmetric *damping matrix* and H is the skew symmetric *circulatory matrix*. The entries c^*_{ij} and h_{ij} are called *damping coefficients* and *circulatory coefficients*, respectively.

At this point, we have all the material required for the derivation of the equations of motion. In examining Eq. (4.14), however, we observe that the various derivatives are with respect to $\mathbf{q}$ and $\dot{\mathbf{q}}$ and all our expressions are in terms of $\mathbf{q}_1$ and $\dot{\mathbf{q}}_1$. This presents no problem, as Eqs. (4.50) can be used to write

$$\frac{\partial L}{\partial \dot{\mathbf{q}}} = \frac{\partial L}{\partial \dot{\mathbf{q}}_1}, \qquad \frac{\partial L}{\partial \mathbf{q}} = \frac{\partial L}{\partial \mathbf{q}_1}, \qquad \frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}} = \frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}_1} \tag{4.61}$$

Hence, Eq. (4.14) can be replaced by

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}_1}\right) - \frac{\partial L}{\partial \mathbf{q}_1} + \frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}_1} = \mathbf{Q} \tag{4.62}$$

Inserting Eqs. (4.51), (4.56) and (4.57) into Eq. (4.16), we obtain the Lagrangian

$$L = T_2 + T_1 - U = \frac{1}{2}\dot{\mathbf{q}}_1^T M\dot{\mathbf{q}}_1 + \dot{\mathbf{q}}_1^T \mathbf{f}_0 + \dot{\mathbf{q}}_1^T F\mathbf{q}_1 - \mathbf{q}_1^T \left.\frac{\partial U}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} - \frac{1}{2}\mathbf{q}_1^T K\mathbf{q}_1 \tag{4.63}$$

where the constant term $U(\mathbf{q}_0)$ was ignored as inconsequential. The first two terms in Eq. (4.62) take the form

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}_1}\right) = M\ddot{\mathbf{q}}_1 + F\dot{\mathbf{q}}_1 \tag{4.64}$$

and

$$\frac{\partial L}{\partial \mathbf{q}_1} = F^T \dot{\mathbf{q}}_1 - \left.\frac{\partial U}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} - K\mathbf{q}_1 \tag{4.65}$$

Moreover, using Eq. (4.59), the third term in Eq. (4.62) is

$$\frac{\partial \mathcal{F}^*}{\partial \dot{\mathbf{q}}_1} = C^* \dot{\mathbf{q}}_1 + H\mathbf{q}_0 + H\mathbf{q}_1 \tag{4.66}$$

Hence, inserting Eqs. (4.64)–(4.66) into Eq. (4.62), we obtain

$$M\ddot{\mathbf{q}}_1 + F\dot{\mathbf{q}}_1 - F^T\dot{\mathbf{q}}_1 + \left.\frac{\partial U}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} + K\mathbf{q}_1 + C^*\dot{\mathbf{q}}_1 + H\mathbf{q}_0 + H\mathbf{q}_1 = \mathbf{Q} \tag{4.67}$$

Equation (4.67) contains terms of different orders of magnitude. Consistent with a perturbation approach, we separate these terms into zero-order terms, denoted by $O\left(\mathbf{q}_1^0\right)$, and first-order terms, denoted by $O\left(\mathbf{q}_1\right)$. The zero-order terms yield the *equilibrium equation*

$$\left.\frac{\partial U}{\partial \mathbf{q}}\right|_{\mathbf{q}=\mathbf{q}_0} + H\mathbf{q}_0 = \mathbf{0} \tag{4.68}$$

an algebraic equation generally nonlinear, and we note that Eq. (4.68) is equivalent to Eq. (4.26) obtained earlier. On the other hand, the first-order terms yield the *linearized equations of motion*

$$M\ddot{\mathbf{q}} + \left(C^* + G\right)\dot{\mathbf{q}} + (K + H)\mathbf{q} = \mathbf{Q} \tag{4.69}$$

where

$$G = F - F^T = -G^T \tag{4.70}$$

is the skew symmetric *gyroscopic matrix*. Its entries, $g_{ij} = f_{ij} - f_{ji}$, are known as *gyroscopic coefficients*. Note that we dropped the subscript from $\mathbf{q}$ in Eq. (4.69) with the understanding that the components of the vector represent displacements from equilibrium.

In deriving the linearized Lagrange's equations, Eq. (4.69), it is not really necessary to take derivatives. Indeed, the equations are fully defined by the coefficient matrices M, C^*, G, K and H and, of course, the virtual work $\overline{\delta W} = \mathbf{Q}^T \delta\mathbf{q}$ performed by the applied forces.

Under certain circumstances, the small-motions assumption is violated, so that the linearized equations are not valid. This occurs, of course, when the equilibrium point is unstable. It can also occur when the equilibrium point is stable but initial excitations or external forces cause large displacements. In such cases, the response can be obtained by integrating the nonlinear differential equations of motion numerically. Numerical methods for determining the system response are discussed in Sec. 4.12.

Example 4.4

Linearize the equations of motion of the system of Examples 4.1 and 4.2 and cast the equations in matrix form.

The nonlinear equations were derived in Example 4.1 in the form of Eqs. (q). To linearize the equations, we consider the coordinate transformation

$$x(t) = x_0 + x_1(t), \qquad y(t) = y_0 + y_1(t) \tag{a}$$

where x_0 and y_0 define the equilibrium positions, and hence they represent constants satisfying Eqs. (c) of Example 4.2, and $x_1(t)$ and $y_1(t)$ are small perturbations. Introducing Eqs. (a) from this example into Eqs. (q) of Example 4.1 and ignoring terms of order higher than one in x_1 and y_1, we obtain the general linearized equtions

$$\begin{aligned} &m\ddot{x}_1 + (c+h)\dot{x}_1 - 2m\Omega\dot{y}_1 + \left(k_x - m\Omega^2 + 3k_x\epsilon_x x_0^2\right)x_1 - h\Omega y_1 = 0 \\ &m\ddot{y}_1 + 2m\Omega\dot{x}_1 + h\dot{y}_1 + h\Omega x_1 + \left(k_y - m\Omega^2 + 3k_y\epsilon_y y_0^2\right)y_1 = 0 \end{aligned} \tag{b}$$

which are valid for all equilibrium points. The matrix form of Eqs. (b) is simply

$$M\ddot{\mathbf{q}} + \left(C^* + G\right)\dot{\mathbf{q}} + (K + H)\mathbf{q} = \mathbf{0} \tag{c}$$

where

$$M = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \tag{d}$$

is the mass matrix,

$$C^* = \begin{bmatrix} c+h & 0 \\ 0 & h \end{bmatrix} \tag{e}$$

is the modified damping matrix, in the sense that it includes contributions from viscous damping forces and circulatory forces,

$$G = \begin{bmatrix} 0 & -2m\Omega \\ 2m\Omega & 0 \end{bmatrix} \tag{f}$$

is the gyroscopic matrix,

$$K = \begin{bmatrix} k_x - m\Omega^2 + 3k_x\epsilon_x x_0^2 & 0 \\ 0 & k_y - m\Omega^2 + 3k_y\epsilon_y y_0^2 \end{bmatrix} \tag{g}$$

is the stiffness matrix and

$$H = \begin{bmatrix} 0 & -h\Omega \\ h\Omega & 0 \end{bmatrix} \tag{h}$$

is the circulatory matrix.

It is clear from the above that the matrix equation, Eq. (c), has the same form for all equilibrium points. As we conclude from Eqs. (d)–(h), the only difference lies in the explicit form of the stiffness matrix, which is as follows:

$$E_1 : x_0 = y_0 = 0$$

$$K = \begin{bmatrix} k_x - m\Omega^2 & 0 \\ 0 & k_y - m\Omega^2 \end{bmatrix} \tag{i}$$

$$E_2, E_3 : x_0 = 0, \qquad y_0 = \pm\sqrt{\left(m\Omega^2 - k_y\right)/k_y\epsilon_y}$$

$$K = \begin{bmatrix} k_x - m\Omega^2 & 0 \\ 0 & 2\left(m\Omega^2 - k_y\right) \end{bmatrix} \tag{j}$$

$$E_4, E_5 : x_0 = \pm\sqrt{\left(m\Omega^2 - k_x\right)/k_x\epsilon_x}, \qquad y_0 = 0$$

$$K = \begin{bmatrix} 2\left(m\Omega^2 - k_x\right) & 0 \\ 0 & k_y - m\Omega^2 \end{bmatrix} \tag{k}$$

$$E_6, E_7, E_8, E_9 : x_0 = \pm\sqrt{\left(m\Omega^2 - k_x\right)/k_x\epsilon_x}, \qquad y_0 = \pm\sqrt{\left(m\Omega^2 - k_y\right)/k_y\epsilon_y}$$

$$K = \begin{bmatrix} 2\left(m\Omega^2 - k_x\right) & 0 \\ 0 & 2\left(m\Omega^2 - k_y\right) \end{bmatrix} \tag{l}$$

4.5 STABILITY OF LINEARIZED SYSTEMS

In Sec. 4.3, we introduced various stability definitions and presented a method for testing system stability, namely, the Liapunov direct method. Although it is generally referred to as a method, the Liapunov direct method represents a philosophy of approach more than a method and its success depends on the ability to generate a Liapunov function, which is by no means guaranteed. The Liapunov direct method can be used both for linear and nonlinear systems.

For strictly linear systems, it is possible to test the system stability by solving the associated algebraic eigenvalue problem and examining the eigenvalues, an approach guaranteed to yield results. But, the algebraic eigenvalue problem is basically a numerical problem, which can be carried out only for given values of the system parameters. By contrast, the Liapunov direct method is a qualitative procedure that, provided a Liapunov function can be found, is capable of yielding stability criteria, as can be concluded from Example 4.3. Of course, we can always generate stability criteria by solving the eigenvalue problem repeatedly for varying values of the system parameters, but such an approach is likely to prove unwieldy. Fortunately, the approach is not really necessary in vibrations, where for the most part the interest lies in the system response and not in stability criteria. Still, the connection between eigenvalues and stability represents an important subject, and we propose to pursue it here.

The stability definitions presented in Sec. 4.3 carry the implication that the externally applied forces are absent. Hence, letting $\mathbf{Q} = \mathbf{0}$ in Eq. (4.69), we obtain the homogeneous part of the linearized equation in the form

$$M\ddot{\mathbf{q}} + \left(C^* + G\right)\dot{\mathbf{q}} + (K + H)\mathbf{q} = \mathbf{0} \tag{4.71}$$

where we recognize that the vector $\mathbf{q}$ in Eq. (4.71) represents a vector of perturbations from the equilibrium position $\mathbf{q}_0$. To simplify the discussion, and without loss of generality, we assume that the equilibrium position is trivial, $\mathbf{q}_0 = \mathbf{0}$. In view of this, we define the perturbation state vector as

$$\mathbf{x}(t) = \left[\mathbf{q}^T(t)\ \dot{\mathbf{q}}^T(t)\right]^T \tag{4.72}$$

Then, adding the identity $\dot{\mathbf{q}} = \dot{\mathbf{q}}$, Eq. (4.71) can be cast in the state form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) \tag{4.73}$$

where

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}(K+H) & -M^{-1}(C^*+G) \end{array}\right] \tag{4.74}$$

is the coefficient matrix.

The nature of the motion in the neighborhood of the trivial equilibrium can be ascertained by simply solving Eq. (4.73). Because the matrix A is constant, Eq. (4.73) represents a homogeneous linear time-invariant system. Its solution has the exponential form

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{x} \tag{4.75}$$

where λ is a constant scalar and $\mathbf{x}$ a constant vector. To determine λ and $\mathbf{x}$, we insert Eq. (4.75) into Eq. (4.73), divide through by $e^{\lambda t}$ and obtain

$$A\mathbf{x} = \lambda\mathbf{x} \tag{4.76}$$

Equation (4.76) represents the *algebraic eigenvalue problem*, a fundamental problem in linear system theory in general and vibrations in particular. The algebraic eigenvalue problem can be stated as the problem of determining the values of the parameter λ, so that Eq. (4.76) has nontrivial solutions. Because the dimension of the eigenvalue problem is $2n$, its solution consists of $2n$ pairs λ_j and $\mathbf{x}_j$, where λ_j are known as the *eigenvalues of* A and $\mathbf{x}_j$ as the *eigenvectors belonging to* λ_j $(j = 1, 2, \ldots, 2n)$. Both λ_j and $\mathbf{x}_j$ are in general complex quantities, although real eigenvalues and eigenvectors are possible. The general solution of Eq. (4.73) can be expressed as the linear combination

$$\mathbf{x}(t) = \sum_{j=1}^{2n} c_j e^{\lambda_j t}\mathbf{x}_j \tag{4.77}$$

where c_j are generally complex coefficients. They represent constants of integration depending on the initial conditions. We should note at this point that, because $\mathbf{x}(t)$ is real, if $c_j e^{\lambda_j t}\mathbf{x}_j$ is a complex solution, then $\bar{c}_j e^{\bar{\lambda}_j t}\bar{\mathbf{x}}_j$ must also be a solution, where the overbar denotes a complex conjugate.

It is clear from Eq. (4.77) that the behavior of the system is governed by the exponential terms, which in turn implies the eigenvalues λ_j. These eigenvalues can be expressed in the general form

$$\lambda_j = \alpha_j + i\beta_j, \qquad j = 1, 2, \ldots, 2n \tag{4.78}$$

The real part α_j of the eigenvalue determines the amplitude of the jth term in the series in Eq. (4.77). Indeed, $c_j e^{\alpha_j t}$ plays the role of a time-varying amplitude. On the other hand, the complex part β_j of the eigenvalue represents the frequency of the jth term, as $e^{i\beta_j t}$ is a unit vector rotating in the complex plane with the angular velocity β_j (see Sec. 1.5). Clearly, only the real part of the eigenvalues controls the stability characteristics of the system.

If all the eigenvalues are complex with zero real part, $\alpha_j = 0$ $(j = 1, 2, \ldots, 2n)$, so that they are all pure imaginary complex conjugates, then the response neither reduces to zero nor increases without bounds as time unfolds. In this case, according to the first stability definition, the equilibrium point is merely *stable*. If all the eigenvalues possess negative real part, $\alpha_j < 0$ $(j = 1, 2, \ldots, 2n)$, or if there are real eigenvalues they are all negative, then the response approaches zero as $t \to \infty$. In this case, according to the second stability definition, the equilibrium point is *asymptotically stable*. Note that, if all the eigenvalues are complex, it is sufficient that a single pair of complex conjugate eigenvalues possess zero real part, while the remaining eigenvalues have negative real part, for the equilibrium to be merely stable instead of being asymptotically stable. Finally, if the real part of at least one pair of complex conjugate eigenvalues is positive, or at least one eigenvalue is real and positive, then the response approaches infinity as $t \to \infty$. In this case, according to the third stability definition, the equilibrium point is *unstable*. Of course, infinity is only a mathematical concept and is to be interpreted to mean that the motions tend to become very large. Clearly, in this case the small motions assumption is violated, so that the linearized equations cease to be valid.

The above stability conclusions were reached on the basis of linearized equations of motion, so that the question arises to what extent the conclusions apply to the original nonlinear system. Before answering this question, it helps if we introduce two definitions. In particular, if the equilibrium point is either *asymptotically stable* or *unstable*, the system is said to possess *significant behavior*. On the other hand, if the equilibrium point is merely *stable*, the system exhibits *critical behavior*. Then, *if the linearized system displays significant behavior, the stability characteristics of the nonlinear system are the same as those of the linearized system.* On the other hand, *if the linearized system exhibits critical behavior, the stability conclusions cannot be extended to the nonlinear system.* In this case, the stability analysis must be based on the full nonlinear equations.

Example 4.5

Consider the linearized system of Example 4.4, test the stability of the equilibrium points by the Liapunov direct method and verify the stability conclusions reached in Example 4.3. Then, solve the algebraic eigenvalue problem for several combinations of the system parameters chosen so as to demonstrate the connection between the eigenvalues and the system stability.

A stability analysis by the Liapunov direct method for the linearized system follows the same pattern as for the nonlinear system. The only difference lies in the Hamiltonian, with the Hamiltonian for the linearized system reducing to a quadratic form, obtained by ignoring terms in x_1 and y_1 of degree three and higher in the Hamiltonian for the nonlinear system. Hence, from Eqs. (e)–(g) of Example 4.3, we can write the Hamiltonian for each of the equilibrium points as follows:

$$E_1: \quad x_0 = y_0 = 0$$

$$\mathcal{H} = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \frac{1}{2}\left(k_x - m\Omega^2\right)x_1^2 + \frac{1}{2}\left(k_y - m\Omega^2\right)y_1^2 \tag{a}$$

$$E_2, E_3: \quad x_0 = 0, \qquad y_0 = \pm\sqrt{\left(m\Omega^2 - k_y\right)/k_y\epsilon_y}$$

$$\mathcal{H} = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \frac{1}{2}\left(k_x - m\Omega^2\right)x_1^2 + \left(m\Omega^2 - k_y\right)y_1^2 \tag{b}$$

$$E_4, E_5: \quad x_0 = \pm\sqrt{\left(m\Omega^2 - k_x\right)/k_x\epsilon_x}, \qquad y_0 = 0$$

$$\mathcal{H} = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \left(m\Omega^2 - k_x\right)x_1^2 + \frac{1}{2}\left(k_y - m\Omega^2\right)y_1^2 \tag{c}$$

$$E_6, E_7, E_8, E_9: \quad x_0 = \pm\sqrt{\left(m\Omega^2 - k_x\right)/k_x\epsilon_x}, \qquad y_0 = \pm\sqrt{\left(m\Omega^2 - k_y\right)/k_y\epsilon_y}$$

$$\mathcal{H} = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \left(m\Omega^2 - k_x\right)x_1^2 + \left(m\Omega^2 - k_y\right)y_1^2 \tag{d}$$

Moreover, in all nine cases

$$\dot{\mathcal{H}} = -c\dot{x}_1^2 \tag{e}$$

From the linearized equations of motion, Eqs. (b) of Example 4.4, we conclude that damping is pervasive, so that an equilibrium point is asymptotically stable if $\mathcal{H}$ is positive definite and unstable if $\mathcal{H}$ can take negative values in the neighborhood of the equilibrium point. Mere stability is not possible.

Recalling from Example 4.3 that the system parameters are such that $k_x - m\Omega^2 < 0$ and $k_y - m\Omega^2 < 0$, we conclude from Eqs. (a)–(c) that $\mathcal{H}$ is sign-variable for the equilibrium points $E_1, E_2, \dots, E_5$, so that these equilibrium points are *unstable*. On the other hand, from Eq. (d), $\mathcal{H}$ is positive definite for the equilibrium points E_6, E_7, E_8, E_9, so that these equilibrium points are *asymptotically stable*.

The stability conclusions reached here on the basis of the linearized system are the same as those for the nonlinear system. This is to be expected, because asymptotic stability and instability imply significant behavior.

The eigenvalue problem for the linearized system is defined by

$$(A - \lambda I)\mathbf{x} = \mathbf{0} \tag{f}$$

where $\mathbf{x} = [x_1 \;\; y_1 \;\; \dot{x}_1 \;\; \dot{y}_1]^T$ is the state vector and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}K & -M^{-1}(C + G) \end{array}\right] \tag{g}$$

is the coefficient matrix, in which

$$M = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix}, \qquad C = \begin{bmatrix} c & 0 \\ 0 & 0 \end{bmatrix}, \qquad G = \begin{bmatrix} 0 & -2m\Omega \\ 2m\Omega & 0 \end{bmatrix} \tag{h}$$

Moreover, the stiffness matrix depends on the equilibrium points, as follows:

$$E_1 : K = \begin{bmatrix} k_x - m\Omega^2 & 0 \\ 0 & k_y - m\Omega^2 \end{bmatrix} \tag{i}$$

$$E_2, E_3 : K = \begin{bmatrix} k_x - m\Omega^2 & \\ 0 & 2\left(m\Omega^2 - k_y\right) \end{bmatrix} \tag{j}$$

$$E_4, E_5 : K = \begin{bmatrix} 2\left(m\Omega^2 - k_x\right) & 0 \\ 0 & k_y - m\Omega^2 \end{bmatrix} \tag{k}$$

$$E_6, E_7, E_8, E_9 : K = \begin{bmatrix} 2\left(m\Omega^2 - k_x\right) & 0 \\ 0 & 2\left(m\Omega^2 - k_y\right) \end{bmatrix} \tag{l}$$

The eigenvalue problem has been solved for the parameters $m = 1$ kg, $\Omega = 2$rad/s, $k_x = k_y = 3$ N/m, $c = 0.2$ N · s/m The eigenvalues corresponding to the nine equilibrium points are as follows:

$$\begin{aligned}
&E_1: &\lambda_1 &= 0.0077 + 0.2678i & \lambda_3 &= -0.1077 + 3.7309i\\
& &\lambda_2 &= 0.0077 - 0.2678i & \lambda_4 &= -0.1077 - 3.7309i\\
&E_2, E_3: &\lambda_1 &= 0.3298 & \lambda_3 &= -0.0877 + 4.1358i\\
& &\lambda_2 &= -0.3544 & \lambda_4 &= -0.0877 - 4.1358i\\
&E_4, E_5: &\lambda_1 &= -0.3367 & \lambda_3 &= -0.1051 + 4.1362i\\
& &\lambda_2 &= 0.3470 & \lambda_4 &= -0.1051 - 4.1362i\\
&E_6, E_7, E_8, E_9: &\lambda_1 &= -0.0092 + 0.4494i & \lambda_3 &= -0.0908 + 4.4482i\\
& &\lambda_2 &= -0.0092 - 0.4494i & \lambda_4 &= -0.0908 - 4.4482i
\end{aligned} \tag{m}$$

An examination of the eigenvalues permits us to conclude that for the equilibrium point E_1 there is one pair of complex conjugate eigenvalues with positive real part and for each of the equilibrium points $E_2, \ldots, E_5$ there is one real positive eigenvalue, so that all these points are *unstable*. On the other hand, for the equilibrium points E_6, E_7, E_8 and E_9, all four eigenvalues are complex with negative real part, so that these points are *asymptotically stable*.

We observe that the stability conclusions based on the eigenvalues are consistent with the conclusions reached on the basis of the Liapunov direct method.

4.6 LINEAR CONSERVATIVE NATURAL SYSTEMS. THE SYMMETRIC EIGENVALUE PROBLEM

As can be concluded from Sec. 4.3, conservative natural systems imply the absence of gyroscopic, viscous damping, circulatory and externally impressed forces. In the case of linear systems, this implies further that the gyroscopic matrix G, damping matrix C^*, circulatory matrix H and force vector $\mathbf{Q}$ are all zero. Under these circumstances, Eq. (4.69) reduces to the *linear conservative natural system*

$$M\ddot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{0} \tag{4.79}$$

where M and K are real symmetric $n \times n$ mass and stiffness matrices, respectively. Moreover, M is positive definite.

A very important case in the study of vibrations is that in which all the coordinates, i.e., all the components $q_i(t)$ of $\mathbf{q}(t)$, execute the same motion in time $(i = 1, 2, \ldots, n)$. In this case, the system is said to execute *synchronous motion*. To examine the possibility that such motions exist, we consider a solution of Eq. (4.79) in the exponential form

$$\mathbf{q}(t) = e^{st}\mathbf{u} \tag{4.80}$$

where s is a constant scalar and $\mathbf{u}$ a constant n-vector. Introducing Eq. (4.80) into Eq. (4.79) and dividing through by e^{st}, we can write

$$K\mathbf{u} = \lambda M\mathbf{u}, \quad \lambda = -s^2 \tag{4.81}$$

Equation (4.81) represents a set of n simultaneous homogeneous algebraic equations in the unknowns u_i $(i = 1, 2, \ldots, n)$, with λ playing the role of a parameter. The problem of determining the values of the parameter λ for which Eq. (4.81) admits nontrivial solutions $\mathbf{u}$ is known as the *algebraic eigenvalue problem*, or simply the *eigenvalue problem*. It is also known as the *characteristic-value problem*. We note that Eq. (4.81) represents a special case of the more general eigenvalue problem given by Eq. (4.76).

The eigenvalue problem, Eq. (4.81), is in terms of two real symmetric matrices. This creates a slight inconvenience, as the eigenvalue problem and the properties of its solution can be best discussed when the problem is defined by a single matrix alone, such as the eigenvalue problem described by Eq. (4.76). Of course, multiplication of Eq. (4.81) on the left by M^{-1} yields such an eigenvalue problem, but the matrix $M^{-1}K$ is generally not symmetric, which tends to obscure the properties of the solution. Because K and M are real and symmetric and, moreover, because M is positive definite, the eigenvalue problem can be transformed into one in terms of a single real symmetric matrix, a highly desirable class of problems. Indeed, from linear algebra (Ref. 14), it can be shown that the matrix M can be decomposed into

$$M = Q^T Q \tag{4.82}$$

where Q is a real nonsingular matrix. Inserting Eq. (4.82) into Eq. (4.81), we obtain

$$K\mathbf{u} = \lambda Q^T Q\mathbf{u} \tag{4.83}$$

Next, we introduce the notation

$$Q\mathbf{u} = \mathbf{v} \tag{4.84}$$

Equation (4.84) represents a linear transformation and the relation

$$\mathbf{u} = Q^{-1}\mathbf{v} \tag{4.85}$$

represents the inverse transformation, where Q^{-1} is the inverse of the matrix Q. The inverse is guaranteed to exist because Q is nonsingular. Introducing Eqs. (4.84) and (4.85) into Eq. (4.83) and multiplying on the left by $\left(Q^T\right)^{-1}$, we obtain the eigenvalue problem

$$A\mathbf{v} = \lambda\mathbf{v} \tag{4.86}$$

where, considering the relation $(Q^T)^{-1} = (Q^{-1})^T$, we conclude that

$$A = \left(Q^T\right)^{-1} K Q^{-1} = \left(Q^{-1}\right)^T K Q^{-1} = A^T \tag{4.87}$$

is a real symmetric matrix. An eigenvalue problem in terms of a single matrix, such as that given by Eq. (4.86), is said to be in *standard form*. Note that Eq. (4.86) is still a special case of Eq. (4.76), because here the matrix A is symmetric.

Next, we propose to discuss the nature of the eigenvalue problem in general terms and examine the implications of its solution on the motion of the system. To this end, we rewrite Eq. (4.86) in the form

$$(A - \lambda I)\mathbf{v} = \mathbf{0} \tag{4.88}$$

where I is the $n \times n$ identity matrix. It is well known from linear algebra that a set of n linear homogeneous algebraic equations in n unknowns possesses a nontrivial solution if and only if the determinant of the coefficients is zero, or

$$\det(A - \lambda I) = 0 \tag{4.89}$$

Equation (4.89) is known as the *characteristic equation* and det $(A - \lambda I)$ represents the *characteristic determinant*, a polynomial of degree n in λ. The characteristic equation possesses n solutions λ_r $(r = 1, 2, \dots, n)$, which represent the roots of the *characteristic polynomial.* The numbers $\lambda = \lambda_r$ $(r = 1, 2, \dots, n)$ are known as the *eigenvalues* of A. Hence, the eigenvalues λ_r are the values of λ that render the matrix $A - \lambda I$ *singular.* To every eigenvalue λ_r corresponds a vector $\mathbf{v}_r$, where $\mathbf{v}_r$ is referred to as the *eigenvector belonging to* λ_r and can be obtained by solving the matrix equation

$$(A - \lambda_r I)\,\mathbf{v}_r = \mathbf{0}, \qquad r = 1, 2, \dots, n \tag{4.90}$$

The eigenvectors $\mathbf{v}_r$ are unique, except for the magnitude. Indeed, because Eq. (4.90) is homogeneous, if $\mathbf{v}_r$ is a solution, then $c_r\mathbf{v}_r$ is also a solution, where c_r is a constant scalar. The implication is that *only the direction of a given eigenvector is unique, not its magnitude.* The magnitude of the eigenvectors can be rendered unique by a process known as *normalization.*

At this point, we wish to explore how the eigenvalues of a matrix change if a given number μ is subtracted from the main diagonal elements of the matrix. To this end, we subtract μI from both sides of Eq. (4.86) and write

$$(A - \mu I)\mathbf{v} = (\lambda - \mu)\mathbf{v} \tag{4.91}$$

Equation (4.91) states that, if the matrix A has the eigenvalue λ, then the matrix $A - \mu I$ has the eigenvalue $\lambda - \mu$. Hence, *subtraction of the constant μ from the main diagonal elements of A results in a shift in the eigenvalues of A by the same constant μ.* This fact can be used to accelerate the convergence of certain iteration processes for computing the eigenvalues of a matrix (Sec. 6.9). We observe that the eigenvectors of A are not affected by the subtraction process.

Eigenvalues and eigenvectors associated with real symmetric matrices have very important properties. To demonstrate these properties, we consider the eigenvalue, eigenvector pair λ_r, $\mathbf{v}_r$ and assume that they are complex. Because A is real, it follows that the complex conjugate pair $\bar{\lambda}_r$, $\bar{\mathbf{v}}_r$ must also constitute an eigenvalue, eigenvector pair. The two pairs satisfy the equations

$$A\mathbf{v}_r = \lambda_r\mathbf{v}_r \tag{4.92a}$$

$$A\bar{\mathbf{v}}_r = \bar{\lambda}_r\bar{\mathbf{v}}_r \tag{4.92b}$$

Next, we premultiply Eq. (4.92a) by $\bar{\mathbf{v}}_r^T$ and Eq. (4.92b) by $\mathbf{v}_r^T$, subtract the transpose of the second from the first, recall that $A^T = A$ and obtain

$$\bar{\mathbf{v}}_r^T A\mathbf{v}_r - \left(\mathbf{v}_r^T A\bar{\mathbf{v}}_r\right)^T = 0 = \lambda_r\bar{\mathbf{v}}_r^T\mathbf{v}_r - \bar{\lambda}_r\left(\mathbf{v}_r^T\bar{\mathbf{v}}_r\right) = \left(\lambda_r - \bar{\lambda}_r\right)\|\mathbf{v}_r\|^2 \tag{4.93}$$

But, for any complex vector $\mathbf{x} = [x_1\ x_2\ \dots\ x_n]^T$,

$$\|\mathbf{x}\|^2 = \overline{\mathbf{x}}^T\mathbf{x} = \sum_{i=1}^{n} \overline{x}_i x_i = \sum_{i=1}^{n} |x_i|^2 > 0 \tag{4.94}$$

is defined as the square of the norm of $\mathbf{x}$, which is a positive number and cannot be zero by definition. It follows that

$$\lambda_r - \overline{\lambda}_r = 0 \tag{4.95}$$

which can be satisfied if and only if λ_r is real. This result can be stated in the form of the theorem: *The eigenvalues of a real symmetric matrix are real.* As a corollary, *the eigenvectors of a real symmetric matrix are real.* Indeed, because the eigenvalues are real, complex numbers need never appear.

Now we consider two distinct eigenvalues, λ_r and λ_s, and the respective eigenvectors $\mathbf{v}_r$ and $\mathbf{v}_s$. Clearly, they must satisfy the relations

$$A\mathbf{v}_r = \lambda_r \mathbf{v}_r \tag{4.96a}$$

$$A\mathbf{v}_s = \lambda_s \mathbf{v}_s \tag{4.96b}$$

Following the same pattern as that just used, we premultiply Eq. (4.96a) by $\mathbf{v}_s^T$ and Eq. (4.96b) by $\mathbf{v}_r^T$, subtract the transpose of the second from the first, consider the symmetry of A and write

$$\mathbf{v}_s^T A\mathbf{v}_r - \left(\mathbf{v}_r^T A\mathbf{v}_s\right)^T = 0 = \lambda_r \mathbf{v}_s^T \mathbf{v}_r - \lambda_s \left(\mathbf{v}_r^T \mathbf{v}_s\right)^T = (\lambda_r - \lambda_s)\mathbf{v}_s^T \mathbf{v}_r \tag{4.97}$$

Because the eigenvalues are distinct, Eq. (4.97) can be satisfied if and only if

$$\mathbf{v}_s^T \mathbf{v}_r = 0, \qquad \lambda_r \neq \lambda_s \tag{4.98a}$$

so that the eigenvectors $\mathbf{v}_r$ and $\mathbf{v}_s$ are orthogonal. This permits us to state the following theorem: *Two eigenvectors of a real symmetric matrix belonging to two distinct eigenvalues are mutually orthogonal.* Equation (4.98a) is not the only orthogonality relation satisfied by the system eigenvectors. Indeed, premultiplying Eq. (4.96a) by $\mathbf{v}_s^T$ and considering Eq. (4.98a), we conclude that

$$\mathbf{v}_s^T A\mathbf{v}_r = 0, \qquad \lambda_r \neq \lambda_s \tag{4.98b}$$

Equation (4.98b) represents the accompanying theorem: *Two eigenvectors belonging to two distinct eigenvalues of a real symmetric matrix A are orthogonal with respect to A.* We note that, whereas Eq. (4.98a) represents mutual orthogonality, Eq. (4.98b) represents orthogonality with respect to A, where A plays the role of a weighting matrix.

The above proof of orthogonality was based on the assumption that the eigenvalues are distinct. The question arises as to what happens when there are repeated eigenvalues, i.e., when two or more eigenvalues have the same value, and we note that when an eigenvalue λ_i is repeated m_i times, where m_i is an integer, λ_i is said to have *multiplicity* m_i. The answer to the question lies in the following theorem (Ref. 14): *If an eigenvalue λ_i of a real symmetric matrix A has multiplicity m_i, then*

A has exactly m_i mutually orthogonal eigenvectors belonging to m_i. These eigenvectors are not unique, as *any linear combination of the eigenvectors belonging to a repeated eigenvalue is also an eigenvector.* Of course, the eigenvectors belonging to the repeated eigenvalue are orthogonal to the remaining eigenvectors. Hence, *all the eigenvectors of a real symmetric matrix A are orthogonal regardless of whether there are repeted eigenvalues or not.*

As pointed out earlier, only the direction of a given eigenvector is unique, not its magnitude. The magnitude can be rendered unique through a normalization process, whereby the magnitude is assigned a certain value. The normalization process itself is not unique and the assigned value is arbitrary. One of the most convenient normalization schemes is that in which the magnitude is equal to one, in which case all the eigenvectors become unit vectors. The normalization process can be expressed in the form

$$\mathbf{v}_r^T \mathbf{v}_r = 1, \qquad \mathbf{v}_r^T A \mathbf{v}_r = \lambda_r, \qquad r = 1, 2, \ldots, n \tag{4.99a, b}$$

Then, combining Eqs. (4.98a) and (4.99a), we can write

$$\mathbf{v}_s^T \mathbf{v}_r = \delta_{rs}, r, s = 1, 2, \ldots, n \tag{4.100}$$

where δ_{rs} is the Kronecker delta. In this case, the eigenvectors are said to be *orthonormal.* Similarly, combining Eqs. (4.98b) and (4.99b), we obtain

$$\mathbf{v}_s^T A \mathbf{v}_r = \lambda_r \delta_{rs}, \qquad r, s = 1, 2, \ldots, n \tag{4.101}$$

Clearly, *the normalization is a mere convenience, and is entirely devoid of any physical content.*

The preceding developments can be cast in matrix form. To this end, we introduce the $n \times n$ matrices of eigenvalues and eigenvectors

$$\Lambda = \operatorname{diag}[\lambda_r], \qquad V = [\mathbf{v}_1 \; \mathbf{v}_2 \; \ldots \; \mathbf{v}_n] \tag{4.102a, b}$$

respectively. Then, the n solutions of the eigenvalue problem, Eq. (4.86), can be written in the compact form

$$AV = V\Lambda \tag{4.103}$$

Similarly, Eq. (4.100) can be expressed as

$$V^T V = I \tag{4.104}$$

Equation (4.104) states that the columns of V are *mutually orthonormal,* in which case V is known as an *orthonormal matrix.* From Eq. (4.104), we conclude that

$$V^{-1} = V^T \tag{4.105}$$

or *the inverse of an orthonormal matrix is equal to its transpose.* Multiplying Eq. (4.105) on the left by V, we obtain

$$VV^T = I \tag{4.106}$$

which indicates that the rows of V are also mutually orthogonal. Finally, in view of Eqs. (4.102), the matrix form of Eqs. (4.101) is

$$V^T A V = \Lambda \tag{4.107}$$

The left side of Eq. (4.107) represents an *orthogonal transformation*, which is a special case of the larger class of *similarity transformations* to be discussed later in this chapter. Equation (4.107) states that *a real symmetric matrix A is diagonizable by means of an orthogonal transformation whereby the transformation matrix is the orthonormal matrix V of the eigenvectors and the diagonal matrix* Λ *is the matrix of the eigenvalues*. This implies that *the eigenvalues do not change in orthogonal transformations*, which further implies that *the characteristic polynomial is invariant in orthogonal transformations*. This suggests the possibility of solving the eigenvalue problem for real symmetric matrices by means of orthogonal transformations. Indeed, many computational algorithms are based on this idea.

A question of particular interest yet to be answered concerns the sign properties of the eigenvalues of real symmetric matrices. To answer this question, we consider the quadratic form

$$f = \mathbf{x}^T A\mathbf{x} \tag{4.108}$$

where A is an $n \times n$ real symmetric matrix and $\mathbf{x}$ a real n-vector. Moreover, we consider the linear transformation

$$\mathbf{x} = V\mathbf{y} \tag{4.109}$$

in which V is the $n \times n$ matrix of eigenvectors of A and $\mathbf{y}$ is a real n-vector. Inserting Eq. (4.109) into Eq. (4.108) and recalling Eq. (4.107), we obtain

$$f = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T V^T AV\mathbf{y} = \mathbf{y}^T \Lambda\mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2 \tag{4.110}$$

The extreme right expression in Eq. (4.110) represents the canonical form of f, with the coefficients of the canonical form being equal to the eigenvalues of A. Hence, Eq. (4.110) relates directly the sign properties of the quadratic form f, and hence the sign properties of the matrix A, to the sign properties of the eigenvalues λ_i of A. In particular, the quadratic form f is positive (negative) for all real nontrivial vectors $\mathbf{y}$ if and only if all the coefficients λ_i are positive (negative). This permits us to state the theorem: *The real symmetric matrix A is positive (negative)definite if all its eigenvalues* λ_i $(i = 1, 2, \ldots, n)$ *are positive (negative)*. Conversely, *if the real symmetric matrix A is positive (negative) definite, then all the eigenvalues* λ_i $(i = 1, 2, \ldots, n)$ *are real and positive (negative)*. Examining the canonical form of f, we conclude that, if one or more of the eigenvalues is zero, then f can be zero for some nontrivial vector $\mathbf{y}$. This conclusion can be stated as the theorem: *If the real symmetric matrix A is positive (negative) semidefinite, then the eigenvalues* λ_i $(i = 1, 2, \ldots, n)$ *are real and nonnegative (nonpositive)*.

The orthogonality of the system eigenvectors is of fundamental importance in linear system theory in general and in vibrations in particular. Indeed, orthogonality plays an indispensable role in the solution of the differential equations of motion associated with the vibration of linear systems. To introduce the ideas, we recall from Chapter 2 (Fig. 2.1) that the position vector of a mass particle can be expressed in terms of rectangular in the form

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{4.111}$$

where **i**, **j** and **k** are unit vectors along the orthogonal axes x, y and z, respectively. Clearly, the unit vectors are orthonormal, as they satisfy the relations

$$\begin{aligned} &\mathbf{i}\cdot\mathbf{i} = 1, \quad \mathbf{j}\cdot\mathbf{j} = 1, \quad \mathbf{k}\cdot\mathbf{k} = 1 \\ &\mathbf{i}\cdot\mathbf{j} = \mathbf{j}\cdot\mathbf{i} = 0, \quad \mathbf{i}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{i} = 0, \quad \mathbf{j}\cdot\mathbf{k} = \mathbf{k}\cdot\mathbf{j} = 0 \end{aligned} \tag{4.112}$$

Then, using Eqs. (4.112), the rectangular components of **r** can be obtained by writing

$$x = \mathbf{i}\cdot\mathbf{r}, \qquad y = \mathbf{j}\cdot\mathbf{r}, \qquad z = \mathbf{k}\cdot\mathbf{r} \tag{4.113}$$

The idea can be generalized somewhat, and cast in a form more consistent with our objectives, by using matrix notation. To this end, we introduce the *standard unit vectors*

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \tag{4.114}$$

which can be regarded as the matrix counterpart of the unit vectors **i**, **j**, **k**. It is evident that the standard unit vectors are orthonormal, as they satisfy

$$\mathbf{e}_i^T\mathbf{e}_j = \delta_{ij}, \qquad i, j = 1, 2, 3 \tag{4.115}$$

Hence, any three-dimensional vector **x** can be expressed in terms of the standard unit vectors as follows:

$$\mathbf{x} = [x_1\, x_2\, x_3]^T = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3 = \sum_{i=1}^{3} x_i\mathbf{e}_i \tag{4.116}$$

The same idea can be further generalized by conceiving of an n-dimensional space defined by the axes $x_1, x_2, \ldots, x_n$ with directions coinciding with the directions of the n orthonormal standard unit vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, respectively. Then, the standard unit vectors $\mathbf{e}_1, \mathbf{e}_2 \ldots, \mathbf{e}_n$ can be used as a *basis* for an n-dimensional *vector space*, which implies that any n-vector **x** can be expressed in the form

$$\mathbf{x} = [x_1\, x_2 \ldots x_n]^T = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n = \sum_{i=1}^{n} x_i\mathbf{e}_i \tag{4.117}$$

Note that x_i $(i = 1, 2, \ldots, n)$ are referred to as the *coordinates of the vector* **x** *with respect to the standard basis* $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$.

At this point, we propose to argue the case for using the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, normalized according to Eqs. (4.99a), as a basis for an n-dimensional vector space. Indeed, the eigenvectors are mutually orthogonal, as they satisfy Eqs. (4.100), so that any n-vector **v** can be expressed in terms of components along $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. This implies that any arbitrary nonzero n-vector **v** can be written as the linear combination

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \sum_{i=1}^{n} c_i\mathbf{v}_i = V\mathbf{c} \tag{4.118}$$

where V is the orthonormal matrix of eigenvectors of the real symmetric matrix A and $\mathbf{c} = [c_1\, c_2\, \ldots\, c_n]^T$ is the n-vector of coordinates of **v** with respect to the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. To obtain the vector $\mathbf{c}$, we premultiply Eq. (4.118) by V^T and consider the orthonormality relations, Eq. (4.104), with the result

$$\mathbf{c} = V^T\mathbf{v} \tag{4.119}$$

Of course, the same operations could be carried out trivially with the standard basis, or not so trivially with any other basis for an n-dimensional vector space. Hence, the question arises as to why should we choose to work with the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. The answer to this question becomes obvious when we consider that

$$A\mathbf{v} = \sum_{i=1}^{n} c_i A\mathbf{v}_i = \sum_{i=1}^{n} c_i \lambda_i \mathbf{v}_i = V\Lambda\mathbf{c} \tag{4.120}$$

where use has been made of Eq. (4.96a), so that the n-vector $A\mathbf{v}$ can also be expressed as a linear combination of the eigenvectors of A, except that every coordinate c_i is multiplied by λ_i. Then, premultiplying Eq. (4.120) by V^T and considering once again the orthonormality relation, Eq. (4.104), we can write

$$\Lambda\mathbf{c} = V^T A\mathbf{v} \tag{4.121}$$

Clearly, what is unique about the basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is that no other basis can be used to represent an arbitrary n-vector $\mathbf{v}$ as well as the companion n-vector $A\mathbf{v}$. We refer to Eqs. (4.118)–(4.121) as the *expansion theorem*.

The above developments are valid for eigenvalue problems in terms of a single real symmetric matrix, Eq. (4.86), so that the question arises as to how they relate to eigenvalue problems in terms of two real symmetric matrices, such as that described by Eq. (4.81). The answer is that *the developments are pertinent provided that one of the matrices is positive definite*, so that a decomposition of the type (4.82) is possible. In this case, *the two eigenvalue problems, Eqs.* (4.81) *and* (4.86), *yield the same eigenvalues.* On the other hand, the eigenvector s are different. Still, they are related, as the eigenvectors $\mathbf{u}_r$ and $\mathbf{v}_r$ $(r = 1, 2, \ldots, n)$ can be obtained from one another by means of the linear transformation and its inverse

$$\mathbf{v}_r = Q\mathbf{u}_r, \qquad \mathbf{u}_r = Q^{-1}\mathbf{v}_r, \qquad r = 1, 2, \ldots, n \tag{4.122a, b}$$

as can be concluded from Eqs. (4.84) and (4.85), respectively.

The eigenvectors $\mathbf{v}_i$ $(i = 1, 2, \ldots, n)$ of the real symmetric matrix A were shown to be mutually orthogonal and orthogonal with respect to A, as indicated by Eqs. (4.98a) and (4.98b), respectively. Moreover, the eigenvectors can be normalized so as to render them orthonormal, with the orthonormality relations given by Eqs. (4.100). We now wish to show that the orthogonality property extends to the eigenvectors $\mathbf{u}_i (i = 1, 2, \ldots, n)$, albeit in a somewhat different form. To this end, we introduce Eq. (4.121a) into Eqs. (4.100), recall Eq. (4.82) and write

$$\mathbf{v}_s^T\mathbf{v}_r = \mathbf{u}_s^T Q^T Q\mathbf{u}_r = \mathbf{u}_s^T M\mathbf{u}_r = \delta_{rs}, \; r, s = 1, 2, \ldots, n \tag{4.123}$$

so that *the eigenvectors* $\mathbf{u}_r$ $(r = 1, 2, \ldots, n)$ *are orthogonal with respect to the mass matrix* M, rather than being mutually orthogonal as the eigenvectors $\mathbf{v}_r$. We note that Eqs. (4.123) represent not only orthogonality relations but a normalization scheme as well. Similarly, inserting Eqs. (4.122a) into Eqs. (4.101) and considering Eq. (4.87), we obtain

$$\mathbf{u}_s^T K \mathbf{u}_r = \lambda_r \delta_{rs}, \qquad r, s = 1, 2, \ldots, n \tag{4.124}$$

so that *the eigenvectors* $\mathbf{u}_r$ $(r = 1, 2, \ldots, n)$ *are orthogonal with respect to the stiffness matrix* K *as well.*

The original eigenvalue problem, Eq. (4.81), and Eqs. (4.123) and (4.124) can be cast in matrix form. Indeed, introducing the matrix of eigenvectors

$$U = [\mathbf{u}_1 \; \mathbf{u}_2 \; \ldots \; \mathbf{u}_n] \tag{4.125}$$

we conclude that the matrix counterpart of Eq. (4.81) is

$$KU = MU\Lambda \tag{4.126}$$

where Λ is the diagonal matrix of eigenvalues, Eq. (4.102a). Moreover, Eqs. (4.123) can be written in the compact matrix form

$$U^T MU = I \tag{4.127}$$

and Eqs. (4.124) can be condensed into

$$U^T KU = \Lambda \tag{4.128}$$

The expansion theorem can also be formulated in terms of the eigenvectors $\mathbf{u}_r$ $(r = 1, 2, \ldots, n)$. Indeed, it is not difficult to verify that in this case the expansion theorem has the form

$$\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_n\mathbf{u}_n = \sum_{i=1}^{n} c_i\mathbf{u}_i = U\mathbf{c} \tag{4.129}$$

where, using Eqs. (4.127) and (4.128),

$$\mathbf{c} = U^T M\mathbf{u} \tag{4.130a}$$

$$\Lambda\mathbf{c} = U^T K\mathbf{u} \tag{4.130b}$$

The expansion theorem, Eqs. (4.129) and (4.130), plays a pivotal role in the solution of the equations of motion for linear systems (see Sec. 4.10).

At this point, we return to the question of existence of synchronous motions, raised in the beginning of this section. It should be obvious by now that this question is intimately related to the existence of solutions of the eigenvalue problem. Because in general the eigenvalue problem in the case at hand admits n distinct solutions, we conclude that the system can execute synchronous motions in n different ways. The nature of these motions depends on the system eigenvalues. We established already that the eigenvalues are real, so that the only remaining question is that of sign. *The mass matrix* M *is positive definite* by definition, *and has no effect on the sign of the eigenvalues.* Hence, *the sign of the eigenvalues depends on the sign properties of the stiffness matrix* K *alone.* In the following, we examine the various possibilities.

In the most frequently encountered case, *the stiffness matrix* K *is positive definite,* in which case *the system is positive definite and all the eigenvalues are positive,*

$\lambda_j > 0 \; (j = 1, 2, \ldots, n)$. In view of the fact that all $\lambda_j \; (j = 1, 2, \ldots, n)$ are positive, it is convenient to introduce the notation

$$\lambda_j = \omega_j^2, \qquad j = 1, 2, \ldots, n \tag{4.131}$$

where $\omega_j \; (j = 1, 2, \ldots, n)$ are real numbers. Inserting Eqs. (4.131) into the second of Eqs. (4.81), we conclude that to each eigenvalue λ_j there corresponds the pair of pure imaginary complex conjugate exponents

$$\begin{matrix} s_j \\ \bar{s}_j \end{matrix} = \pm i\omega_j, \qquad j = 1, 2, \ldots, n \tag{4.132}$$

Introducing these exponents into Eq. (4.80), we conclude that Eq. (4.79) admits synchronous solutions of the form

$$\mathbf{q}_j(t) = \left(a_j e^{i\omega_j t} + \bar{a}_j e^{-i\omega_j t}\right)\mathbf{u}_j = A_j \cos\left(\omega_j t - \phi_j\right)\mathbf{u}_j, \qquad j = 1, 2, \ldots, n \tag{4.133}$$

where A_j and ϕ_j are known as *amplitude* and *phase angle*, respectively, and we note that the coefficient $\bar{a}_j$ of $e^{-i\omega_j t}$ was taken as the complex conjugate of a_j, because $\mathbf{q}_j(t)$ must be real. Equations (4.133) indicate that a positive definite system admits synchronous motions varying harmonically with time, where $\omega_j \; (j = 1, 2, \ldots, n)$ are known as *natural frequencies* of vibration. Consistent with this, the eigenvectors $\mathbf{u}_j \; (j = 1, 2, \ldots, n)$ are called *natural modes*. They are also referred to as *modal vectors*. The synchronous solutions $\mathbf{q}_j(t)$ represent *natural motions*, and they are an inherent characteristic of the system. In general, the solution of Eq. (4.79) consists of a linear combination of the natural motions, or

$$\mathbf{q}(t) = \sum_{j=1}^{n} \mathbf{q}_j(t) = \sum_{j=1}^{n} A_j \cos\left(\omega_j t - \phi_j\right)\mathbf{u}_j \tag{4.134}$$

in which A_j and ϕ_j play the role of constants of integration. Their value depends on the initial conditions, i.e., initial displacements and velocities. Note that, by adjusting the initial conditions, each of the natural motions can be excited independently of the other, in which case the system will vibrate in the particular mode excited.

Equation (4.134) represents the homogeneous solution, i.e., the response of the system in the absence of impressed forces. For this reason, it is known as the *free response* of the system. Because Eq. (4.134) is a combination of harmonic terms, which oscillate between given limits, the response neither approaches zero nor increases secularly with time. Hence, the free response of a positive definite conservative system is *stable*.

When *the stiffness matrix K is only positive semidefinite, the system is positive semidefinite, and the eigenvalues are nonnegative*, $\lambda_j \geq 0 \; (j = 1, 2, \ldots, n)$. This implies that the system admits some zero eigenvalues, with the rest of the eigenvalues being positive. Corresponding to a zero eigenvalue, say $\lambda_s = 0$, we have the solution

$$\mathbf{q}_s(t) = (a_s + tb_s)\,\mathbf{u}_s \tag{4.135}$$

which is divergent, and hence *unstable*. Zero eigenvalues occur when the system is *unrestrained*, in which case the associated eigenvectors can be identified as *rigid-body modes*. They satisfy the relations

$$K\mathbf{u}_s = \mathbf{0}, \qquad s = 1, 2, \ldots, r \tag{4.136}$$

where r is the number of rigid-body modes. The free response of the system in this case is simply

$$\mathbf{q}(t) = \sum_{s=1}^{r} (a_s + tb_s)\,\mathbf{u}_s + \sum_{j=r+1}^{n} A_j \cos\left(\omega_j t - \phi_j\right)\mathbf{u}_j \tag{4.137}$$

In the case in which K *is sign-variable, the system admits negative eigenvalues*. Corresponding to a negative eigenvalue, say $\lambda_j < 0$, we obtain the exponents

$$\begin{matrix} s_j \\ s_{j+1} \end{matrix} = \pm\sqrt{|-\lambda_j|} \tag{4.138}$$

Whereas the solution corresponding to $s_{j+1} = -\sqrt{|-\lambda_j|}$ decays exponentially with time, the solution corresponding to $s_j = \sqrt{|-\lambda_j|}$ diverges, so that *the system is unstable*. Note that it is sufficient that a single eigenvalue be negative for the system to be unstable. When K is *negative definite* or *negative semidefinite*, the eigenvalues are negative or nonpositive, respectively, so that the system is *unstable*.

It should be pointed out that unstable solutions associated with negative eigenvalues are still consistent with conservative systems, except that at some point the small motions assumption is violated and the system may no longer be linear.

Example 4.6

Derive the eigenvalue problem for the vibrating system of Fig. 4.8, reduce it to one in terms of a single real symmetric matrix and verify the properties of the eigenvalues and eigenvectors. The parameters are as follows: $m_1 = 2m$, $m_2 = 3m$, $m_3 = m$, $k_1 = 2k$, $k_2 = 3k$, $k_3 = 2k$, $k_4 = k$.

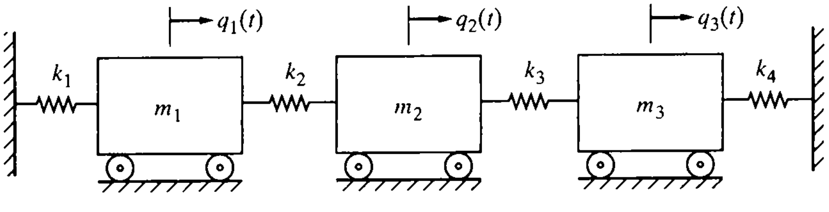


Figure 4.8 Undamped three-degree-of-freedom system

The eigenvalue problem, Eq. (4.81), is defined by the mass and stiffness matrices, which can be obtained from the kinetic energy and potential energy, respectively. The kinetic energy is simply

$$T = \frac{1}{2}\left[m_1\dot{q}_1^2(t) + m_2\dot{q}_2^2(t) + m_3\dot{q}_3^2(t)\right] = \frac{1}{2}\dot{\mathbf{q}}^T(t)M\dot{\mathbf{q}}(t) \tag{a}$$

where $\mathbf{q}(t) = [q_1(t)\ q_2(t)\ q_3(t)]^T$ is the configuration vector and

$$M = \begin{bmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{bmatrix} = \begin{bmatrix} 2m & 0 & 0 \\ 0 & 3m & 0 \\ 0 & 0 & m \end{bmatrix} \tag{b}$$

is the mass matrix. Moreover, the potential energy has the expression

$$\begin{aligned} V &= \frac{1}{2}\left\{k_1 q_1^2(t) + k_2[q_2(t) - q_1(t)]^2 + k_3[q_3(t) - q_2(t)]^2 + k_4 q_3^2(t)\right\} \\ &= \frac{1}{2}\left[(k_1 + k_2)q_1^2(t) + (k_2 + k_3)q_2^2(t) + (k_3 + k_4)q_3^2(t) - 2k_2 q_1(t)q_2(t)\right. \\ &\qquad \left. - 2k_3 q_2(t)q_3(t)\right] \\ &= \frac{1}{2}\mathbf{q}^T(t)K\mathbf{q}(t) \end{aligned} \tag{c}$$

in which

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 + k_4 \end{bmatrix} = \begin{bmatrix} 5k & -3k & 0 \\ -3k & 5k & -2k \\ 0 & -2k & 3k \end{bmatrix} \tag{d}$$

is the stiffness matrix. Inserting Eqs. (b) and (d) into Eq. (4.81) and considering Eq. (4.131), we can write the eigenvalue problem in the form

$$\begin{bmatrix} 5 & -3 & 0 \\ -3 & 5 & -2 \\ 0 & -2 & 3 \end{bmatrix}\mathbf{u} = \lambda \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{u}, \qquad \lambda = \frac{m\omega^2}{k} \tag{e}$$

To reduce the eigenvalue problem (e) to one in terms of a single matrix, we must use the linear transformation (4.84) with the matrix Q obtained from the decomposition of M, Eq. (4.82). In the case at hand M is a diagonal matrix, so that Q simply reduces to

$$Q = M^{1/2} = m^{1/2}\begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{f}$$

Hence, following the established process, the eigenvalue problem in terms of a single real symmetric matrix, Eq. (4.86), is defined by the matrix

$$\begin{aligned} A = M^{-1/2}KM^{-1/2} &= \frac{k}{m}\begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 5 & -3 & 0 \\ -3 & 5 & -2 \\ 0 & -2 & 3 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \frac{k}{m}\begin{bmatrix} 5/2 & -\sqrt{3/2} & 0 \\ -\sqrt{3/2} & 5/3 & -2/\sqrt{3} \\ 0 & -2/\sqrt{3} & 3 \end{bmatrix} \end{aligned} \tag{g}$$

Moreover from Eq. (4.85) the eigenvectors of the two problems are related by

$$\mathbf{u} = M^{-1/2}\mathbf{v} = m^{-1/2}\begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}\mathbf{v} = m^{-1/2}\begin{bmatrix} v_1/\sqrt{2} \\ v_2/\sqrt{3} \\ v_3 \end{bmatrix} \tag{h}$$

The solution of the eigenvalue problem (4.86) with A given by Eq. (g) consists of the eigenvalues

$$\lambda_1 = 0.4256, \qquad \lambda_2 = 2.7410, \qquad \lambda_3 = 4.0000 \tag{i}$$

and eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} 0.4743 \\ 0.8033 \\ 0.3603 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -0.7435 \\ 0.1463 \\ 0.6525 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 0.4714 \\ -0.5774 \\ 0.6667 \end{bmatrix} \tag{j}$$

Clearly, the eigenvalues are real, because A is real and symmetric. Moreover, they are positive, because A is positive definite. It is easy to verify that the eigenvectors satisfy the orthogonality relations (4.98). In fact, they have been normalized so as to satisfy the orthonormality relations (4.100) and (4.101).

The eigenvalues given by Eqs. (i) can be used to compute the natural frequencies

$$\omega_1 = 0.6524\sqrt{\frac{k}{m}} \text{ rad/s}, \quad \omega_2 = 1.6556\sqrt{\frac{k}{m}} \text{ rad/s}, \quad \omega_3 = 2.0000\sqrt{\frac{k}{m}} \text{ rad/s} \tag{k}$$

Moreover, introducing Eqs. (j) into Eq. (h), we obtain the modal vectors

$$\mathbf{u}_1 = m^{-1/2}\begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix}, \quad \mathbf{u}_2 = m^{-1/2}\begin{bmatrix} -0.5257 \\ 0.0845 \\ 0.6525 \end{bmatrix}, \quad \mathbf{u}_3 = m^{-1/2}\begin{bmatrix} 0.3333 \\ -0.3333 \\ 0.6667 \end{bmatrix} \tag{l}$$

It is not difficult to verify that the modal vectors satisfy the orthonormality relations, Eqs. (4.123) and (4.124).

4.7 GYROSCOPIC CONSERVATIVE SYSTEMS

In Sec. 2.13, we have shown that, in the case of nonnatural systems, the Hamiltonian is conserved if it does not depend explicitly on time and all the nonconservative forces are zero. The nonconservative forces include the damping, circulatory and externally applied forces, as demonstrated in Sec. 4.3. Gyroscopic forces, like viscous damping forces and part of the circulatory forces, are proportional to the velocities, so that it is tempting to conclude that they are nonconservative as well. But, as shown in Sec. 4.3, gyroscopic forces do not prevent the conservation of the Hamiltonian, so that they represent conservative forces. It should be pointed out here that the gyroscopic forces are the ones responsible for a system being nonnatural. In this section, we first verify the conservative nature of the gyroscopic forces and then present a method taking advantage of this fact to simplify the eigenvalue problem for gyroscopic conservative systems significantly.

Let us consider once again the system described by Eq. (4.69), assume that the damping matrix C^*, circulatory matrix H and force vector $\mathbf{Q}$ are all zero and obtain the *linear gyroscopic system*

$$M\ddot{\mathbf{q}}(t) + G\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{0} \tag{4.139}$$

where M is the real symmetric positive definite mass matrix, G is the real skew symmetric gyroscopic matrix and K is the real symmetric positive definite, or positive semidefinite stiffness matrix, all $n \times n$ matrices. Because Eq. (4.139) represents a linear homogeneous time-invariant system, it admits a solution of the exponential form

$$\mathbf{q}(t) = e^{st}\mathbf{u} \tag{4.140}$$

where s is a constant scalar and $\mathbf{u}$ is a constant vector, both in general complex. Introducing Eq. (4.140) into Eq. (4.139) and following the usual steps, we obtain the eigenvalue problem

$$s^2 M\mathbf{u} + sG\mathbf{u} + K\mathbf{u} = \mathbf{0} \tag{4.141}$$

Then, if we premultiply Eq. (4.141) by $\bar{\mathbf{u}}^T$, where $\bar{\mathbf{u}}$ is the complex conjugate of $\mathbf{u}$, we can write

$$ms^2 + igs + k = 0 \tag{4.142}$$

where

$$m = \bar{\mathbf{u}}^T M\mathbf{u} > 0, \qquad ig = \bar{\mathbf{u}}^T G\mathbf{u}, \qquad k = \bar{\mathbf{u}}^T K\mathbf{u} \geq 0 \tag{4.143}$$

are scalars. Equation (4.142) represents a quadratic equation in s having the roots

$$\begin{matrix} s_1 \\ s_2 \end{matrix} = i\left(-\frac{g}{2m} \pm \frac{1}{2m}\sqrt{g^2 + 4mk}\right) \tag{4.144}$$

The nature of the roots depends on k, and hence on K. We distinguish the following cases:

i. If $k > 0$, so that K is positive definite, then

$$g^2 + 4mk > 0 \tag{4.145}$$

In this case *the roots are pure imaginary*, which implies *pure oscillatory motion*, or *stable motion.*

ii. If $k \geq 0$, and $k = 0$ for some $\mathbf{u} \neq 0$, then K is positive semidefinite and *at least one root is zero*, which implies *divergent motion.*

iii. If $k < 0$, so that K is negative definite, there are two possibilities. *If inequality* (4.145) *still holds, the motion represents pure oscillation*, which is *stable*. Hence, gyroscopic forces can stabilize an unstable conservative system. On the other hand, if

$$g^2 + 4mk < 0 \tag{4.146}$$

then *at least one of the roots has positive real part*, which implies *unstable motion.*

In the case of structural vibration, the stiffness matrix is generally positive definite or positive semidefinite, so that our interest lies in Cases i and ii. Case ii tends to arise when rigid-body motions are possible, in which case the associated eigenvectors are in the nullspace of K, $K\mathbf{u} = \mathbf{0}$. But, rigid-body motions can be removed by constraining the system so as to undergo elastic motions alone. In view of this, we do not lose generality if *we confine ourselves to the case in which K is positive definite.*

Under the assumption that both M and K are positive definite, which corresponds to case i, all the eigenvalues are pure imaginary. Hence, if we substitute $s = i\omega$ in Eq. (4.141), we obtain the eigenvalue problem

$$-\omega^2 M\mathbf{u} + i\omega G\mathbf{u} + K\mathbf{u} = \mathbf{0} \tag{4.147}$$

where ω must satisfy the characteristic equation

$$\det\left[-\omega^2 M + i\omega G + K\right] = 0 \tag{4.148}$$

But the determinant of a matrix is equal to the determinant of the transposed matrix. Hence, recalling that M and K are symmetric and G is skew symmetric, the characteristic equation can also be written in the form

$$\det\left[-\omega^2 M + i\omega G + K\right]^T = \det\left[-\omega^2 M - i\omega G + K\right] = 0 \qquad (4.149)$$

from which we conclude that if $i\omega$ is a root of the characteristic equation, then $-i\omega$ is also a root. It follows that *the eigenvalues of a gyroscopic conservative system occur in pairs of pure imaginary complex conjugates, $s_r = i\omega_r, \bar{s}_r = -i\omega_r$ $(r = 1, 2, \ldots, n)$*, where ω_r are recognized as the *natural frequencies*. As a corollary, it follows that *the eigenvectors belonging to the eigenvalues $\pm i\omega_r$ are complex conjugates*, although they are *not necessarily pure imaginary*.

The eigenvalue problem given by Eq. (4.147) contains both ω and ω^2 and is complex, so that it does not permit a ready solution. The first difficulty can be overcome by recasting the problem in state form, but the problem remains complex. A method developed in Ref. 10 takes advantage of the fact that the eigenvalues are pure imaginary to remove the second objection. To describe the method, we begin by recasting the equations of motion, Eq. (4.139), in the state form

$$M^*\dot{\mathbf{x}}(t) = -G^*\mathbf{x}(t) \qquad (4.150)$$

where $\mathbf{x}(t) = \left[\mathbf{q}^T(t)\ \dot{\mathbf{q}}^T(t)\right]^T$ is the $2n$-dimensional state vector and

$$M^* = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} = M^{*T}, \qquad G^* = \begin{bmatrix} 0 & -K \\ K & G \end{bmatrix} = -G^{*T} \qquad (4.151\text{a, b})$$

are $2n \times 2n$ coefficient matrices, in which M^* is real symmetric and positive definite and G^* is real and skew symmetric. In view of Eq. (4.140), and considering the fact that the eigenvalues are pure imaginary, the solution of Eq. (4.150) has the form

$$\mathbf{x}(t) = e^{i\omega t}\mathbf{x} \qquad (4.152)$$

where $\mathbf{x}$ is a constant $2n$-vector. Introducing Eq. (4.152) into Eq. (4.150) and dividing both sides by $e^{i\omega t}$, we obtain the eigenvalue problem

$$i\omega M^*\mathbf{x} = -G^*\mathbf{x} \qquad (4.153)$$

Now the eigenvalue problem, albeit of order $2n$, contains ω to the first power only, but the problem is still complex. To reduce the eigenvalue problem to real form, we introduce $\mathbf{x} = \mathbf{y} + i\mathbf{z}$ into Eq. (4.153), equate both the real and imaginary part on both sides of the resulting equation and write

$$\omega M^*\mathbf{y} = -G^*\mathbf{z}, \qquad \omega M^*\mathbf{z} = G^*\mathbf{y} \qquad (4.154\text{a, b})$$

Solving Eq. (4.154b) for $\mathbf{z}$ and introducing into Eq. (4.154a), then solving Eq. (4.154a) for $\mathbf{y}$ and introducing into Eq. (4.154b), we obtain

$$K^*\mathbf{y} = \lambda M^*\mathbf{y}, \qquad K^*\mathbf{z} = \lambda M^*\mathbf{z}, \qquad \lambda = \omega^2 \qquad (4.155)$$

where

$$K^* = G^{*T}M^{*\,-1}G^* = \begin{bmatrix} KM^{-1}K & KM^{-1}G \\ G^TM^{-1}K & K + G^TM^{-1}G \end{bmatrix} = K^{*T} \qquad (4.156)$$

is a real symmetric positive definite matrix. Hence, we not only reduced the complex eigenvalue problem (4.153) to a real one, but to one in terms of two real symmetric positive definite matrices of the type given by Eq. (4.81). It follows that we can use the approach of Sec. 4.6 to reduce the eigenvalue problem (4.155) to standard form, i.e., one in terms of a single real symmetric matrix.

Because both eigenvalue problems, Eqs. (4.81) and (4.155), belong to the same class, their solutions share many of the characteristics. Moreover, both problems can be solved by the same efficient computational algorithms for real symmetric matrices. Still, some differences exist. In the first place, eigenvalue problem (4.155) is of order $2n$, whereas eigenvalue problem (4.81) is of order n only. Moreover, we observe that the real part $\mathbf{y}$ and the imaginary part $\mathbf{z}$ of $\mathbf{x}$ satisfy the same eigenvalue problem. It follows that the eigenvalue problem given by Eqs. (4.155) is characterized by the fact that *every eigenvalue has multiplicity two*, as to every eigenvalue $\lambda_r = \omega_r^2$ belong two eigenvectors, $\mathbf{y}_r$ and $\mathbf{z}_r$ $(r = 1, 2, \ldots, n)$. This fact presents no problem as far as solving the eigenvalue problem is concerned. Indeed, because the problem is positive definite, the eigenvectors $\mathbf{y}_r$ and $\mathbf{z}_r$ are independent and can be rendered orthogonal. Of course, they are orthogonal to the remaining $n-1$ pairs of eigenvectors. The solution to the real eigenvalue problem, Eqs. (4.155), can be used to construct the solution to the complex eigenvalue problem, Eq. (4.153). Indeed, the complex eigensolutions can be written in the form

$$\begin{matrix} s_r \\ \bar{s}_r \end{matrix} = \pm i\omega_r, \qquad \begin{matrix} \mathbf{x}_r \\ \bar{\mathbf{x}}_r \end{matrix} = \mathbf{y}_r \pm i\mathbf{z}_r, \qquad r = 1, 2, \ldots, n \tag{4.157}$$

As in the case of natural systems, it is advantageous to cast the eigenvalue problem, Eqs. (4.155), into a form defined by a single real symmetric matrix instead of two. This presents no problem, as both M^* and K^* are real symmetric positive definite matrices. Hence, by analogy with Eq. (4.82), we decompose the matrix M^* into

$$M^* = Q^{*T} Q^* \tag{4.158}$$

where Q^* is a $2n \times 2n$ nonsingular matrix. Then, using the linear transformation

$$Q^*\mathbf{y} = \mathbf{v}_y, \qquad Q^*\mathbf{z} = \mathbf{v}_z \tag{4.159}$$

and recognizing that it is not really necessary to distinguish between $\mathbf{v}_y$ and $\mathbf{v}_z$, we can reduce Eq. (4.155) to the standard form

$$A\mathbf{v} = \lambda\mathbf{v}, \qquad \lambda = \omega^2 \tag{4.160}$$

where $\mathbf{v}$ stands for both $\mathbf{v}_y$ and $\mathbf{v}_z$ and

$$A = \left(Q^{*T}\right)^{-1} K^* Q^{*-1} = \left(Q^{*-1}\right)^T K^* Q^{*-1} = A^T \tag{4.161}$$

is a $2n \times 2n$ real symmetric positive definite matrix. Clearly, the two eigenvalue problems, Eqs. (4.155) and (4.160), possess the same eigenvalues, so that every eigenvalue of A, as given by Eq. (4.161), retains the multiplicity two. We express this multiplicity in the form

$$\lambda_{2r-1} = \lambda_{2r}, \qquad r = 1, 2, \ldots, n \tag{4.162}$$

Then, by analogy with Eqs. (4.159), we express the real and imaginary part of $\mathbf{x}_r$ as

$$\mathbf{y}_r = Q^{*-1}\mathbf{v}_{2r-1}, \qquad \mathbf{z}_r = Q^{*-1}\mathbf{v}_{2r}, \qquad r = 1, 2, \ldots, n \tag{4.163}$$

where $\mathbf{v}_{2r-1}$ and $\mathbf{v}_{2r}$ are pairs of orthogonal eigenvectors belonging to $\lambda_{2r-1} = \lambda_{2r}$ $(r = 1, 2, \ldots, n)$. Of course, the eigenvectors are orthogonal to the remaining pairs of eigenvectors. The eigenvectors can be normalized so as to satisfy the orthonormality relations

$$\mathbf{v}_{2r-1}^T\mathbf{v}_{2r} = 0, \quad \mathbf{v}_{2s-1}^T\mathbf{v}_{2r-1} = \mathbf{v}_{2s}^T\mathbf{v}_{2r} = \delta_{rs}, \quad \mathbf{v}_{2s-1}^T A\mathbf{v}_{2r-1} = \mathbf{v}_{2s}^T A\mathbf{v}_{2r} = \lambda_r\delta_{rs},$$
$$r, s = 1, 2, \ldots, n \tag{4.164}$$

At this point, we return to the free vibration problem, Eq. (4.139). As amply demonstrated in this section, in the case of gyroscopic systems, we must work with the state form, namely, Eq. (4.150). Equation (4.150) admits solutions in the form of linear combinations of the eigensolutions. Because the solution must be real, by analogy with natural systems, we express the free response of a gyroscopic conservative system as the linear combination

$$\begin{aligned}\mathbf{x}(t) &= \sum_{r=1}^{n}\left(c_r e^{i\omega_r t}\mathbf{x}_r + \bar{c}_r e^{-i\omega_r t}\bar{\mathbf{x}}_r\right)\\ &= \sum_{r=1}^{n} A_r\left[\cos(\omega_r t - \phi_r)\,\mathbf{y}_r - \sin(\omega_r t - \phi_r)\,\mathbf{z}_r\right]\end{aligned} \tag{4.165}$$

where the amplitudes A_r and phase angles ϕ_r $(r = 1, 2, \ldots, n)$ depend on the initial condition $\mathbf{x}(0) = \left[\mathbf{q}^T(0)\ \dot{\mathbf{q}}^T(0)\right]^T$. Clearly, Eq. (4.165) gives not only the displacement vector $\mathbf{q}(t)$ but also the velocity vector $\dot{\mathbf{q}}(t)$.

Example 4.7

The system of Example 4.5 represents a damped gyroscopic system. Let $m = 1\,\text{kg}$, $\Omega = 2\,\text{rad/s}$, $k_x = 5\,\text{N/m}$, $k_y = 10\,\text{N/m}$, $c = 0$ and solve the eigenvalue problem associated with A, Eq. (g) of Example 4.5, for the equilibirium point E_1. Then, solve the same problem by the approach presented in this section and compare the results.

Using the given parameter values and Eq. (i) of Example 4.5, the matrix A has the form

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}K & -M^{-1}G \end{array}\right]$$
$$= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -m^{-1}\left(k_x - m\Omega^2\right) & 0 & 0 & 2\Omega \\ 0 & -m^{-1}\left(k_y - m\Omega^2\right) & -2\Omega & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 4 \\ 0 & -6 & -4 & 0 \end{bmatrix} \tag{a}$$

which has the eigenvalues

$$\begin{aligned} \lambda_1 &= i\omega_1 = 0.5137i \text{ rad/s}, \qquad \lambda_2 = \bar{\lambda}_1 = -i\omega_1 = -0.5137i \text{ rad/s} \\ \lambda_3 &= i\omega_2 = 4.7682i \text{ rad/s}, \qquad \lambda_4 = \bar{\lambda}_3 = -i\omega_2 = -4.7682i \text{ rad/s} \end{aligned} \tag{b}$$

where ω_1 and ω_2 are natural frequencies, and the right and left eigenvectors (see Sec. 4.8)

$$\mathbf{v}_1 = \begin{bmatrix} 0.8374 \\ -0.3000i \\ 0.4302i \\ 0.1541 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0.8374 \\ 0.3000i \\ -0.4302i \\ 0.1541 \end{bmatrix}$$

$$\mathbf{v}_3 = \begin{bmatrix} 0.1354 \\ 0.1543i \\ 0.6455i \\ -0.7356 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} 0.1354 \\ -0.1543i \\ -0.6455i \\ -0.7356 \end{bmatrix} \tag{c}$$

and

$$\mathbf{w}_1 = \begin{bmatrix} 0.5775 \\ -1.2413i \\ 0.2967i \\ 0.1063 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 0.5775 \\ 1.2413i \\ -0.2967i \\ 0.1063 \end{bmatrix}$$

$$\mathbf{w}_3 = \begin{bmatrix} 0.1210 \\ 0.8272i \\ 0.5769i \\ -0.6574 \end{bmatrix}, \quad \mathbf{w}_4 = \begin{bmatrix} 0.1210 \\ -0.8272i \\ -0.5769i \\ -0.6574 \end{bmatrix} \tag{d}$$

respectively. Note that the eigenvectors have been normalized so that $\mathbf{w}_i^T \mathbf{v}_i = 1$ $(i = 1, 2, 3, 4)$.

Next, we insert the same parameter values into Eqs. (4.151a) and (4.156) and obtain

$$M^* = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} = \begin{bmatrix} k_x - m\Omega^2 & 0 & 0 & 0 \\ 0 & k_y - m\Omega^2 & 0 & 0 \\ 0 & 0 & m & 0 \\ 0 & 0 & 0 & m \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 6 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{e}$$

and

$$\begin{aligned} K^* &= \begin{bmatrix} KM^{-1}K & KM^{-1}G \\ G^TM^{-1}K & K + G^TM^{-1}G \end{bmatrix} \\ &= \begin{bmatrix} m^{-1}\left(k_x - m\Omega^2\right)^2 & 0 & 0 & -2\Omega\left(k_x - m\Omega^2\right) \\ 0 & m^{-1}\left(k_y - m\Omega^2\right)^2 & 2\Omega\left(k_y - m\Omega^2\right) & 0 \\ 0 & 2\Omega\left(k_y - m\Omega^2\right) & k_x + 3m\Omega^2 & 0 \\ -2\Omega\left(k_x - m\Omega^2\right) & 0 & 0 & k_y + 3m\Omega^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 & -4 \\ 0 & 36 & 24 & 0 \\ 0 & 24 & 17 & 0 \\ -4 & 0 & 0 & 22 \end{bmatrix} \end{aligned} \tag{f}$$

respectively. Inserting Eqs. (e) and (f) into Eqs. (4.155) and solving the eigenvalue problem, we obtain the repeated eigenvalues

$$\lambda_1 = \lambda_2 = \omega_1^2 = 0.2639, \qquad \lambda_3 = \lambda_4 = \omega_2^2 = 22.736 \tag{g}$$

yielding the natural frequencies

$$\omega_1 = 0.5137 \text{ rad/s}, \qquad \omega_2 = 4.7682 \text{ rad/s} \tag{h}$$

which coincide with those obtained solving the eigenvalue problem for A. Moreover, the eigenvectors are

$$\mathbf{y}_1 = \begin{bmatrix} -0.1810 \\ 0 \\ 0 \\ 0.9835 \end{bmatrix}, \quad \mathbf{z}_1 = \begin{bmatrix} 0 \\ -0.2063 \\ -0.8630 \\ 0 \end{bmatrix}$$
$$\mathbf{y}_2 = \begin{bmatrix} -0.9835 \\ 0 \\ 0 \\ -0.1810 \end{bmatrix}, \quad \mathbf{z}_2 = \begin{bmatrix} 0 \\ -0.3523 \\ 0.5052 \\ 0 \end{bmatrix} \tag{i}$$

The eigenvectors have been normalized so that $\mathbf{y}_r^T M \mathbf{y}_r = \mathbf{z}_r^T M \mathbf{z}_r = 1 (r = 1, 2)$.

Clearly, the eigenvalue problem in symmetric form, Eqs. (4.155), is considerably easier to solve than the nonsymmetric eigenvalue problem, and the eigenvector structure is considerably simpler. In fact, we note from Eqs. (e) and (f) that the symmetric eigenvalue problem of order $2n$ can be separated into two symmetric eigenvalue problems of order n.

4.8 NONCONSERVATIVE SYSTEMS. THE NONSYMMETRIC EIGENVALUE PROBLEM

As demonstrated in Sec. 4.3, viscous damping forces and circulatory forces render a system nonconservative, in the sense that the Hamiltonian is not conserved. In this case, the free response is no longer pure oscillatory, so that the desirable characteristics of conservative systems no longer exist. It is possible to treat the case in which only viscous damping forces are present separately from the case in which both viscous damping and circulatory forces are present. In this text, we go directly to the second case and will make the distinction between the two cases when appropriate. Hence, we consider the general homogeneous linear system given by Eq. (4.71), or

$$M\ddot{\mathbf{q}}(t) + \left(C^* + G\right)\dot{\mathbf{q}}(t) + (K + H)\,\mathbf{q}(t) = \mathbf{0} \tag{4.166}$$

where the various matrices are as defined in Sec. 4.4.

To explore the system characteristics, it is necessary to cast Eq. (4.166) in state form. To this end, we premultiply Eq. (4.166) by M^{-1}, adjoin the identity $\dot{\mathbf{q}}(t) = \dot{\mathbf{q}}(t)$ and rewrite the free vibration equations in the standard state form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) \tag{4.167}$$

where $\mathbf{x}(t) = \left[\mathbf{q}^T(t)\;\dot{\mathbf{q}}^T(t)\right]^T$ is the $2n$-dimensional state vector and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}(K + H) & -M^{-1}(C^* + G) \end{array}\right] \tag{4.168}$$

is the $2n \times 2n$ real nonsymmetric coefficient matrix, and we recall that we first encountered Eqs. (4.167) and (4.168) in Sec. 4.5 in conjunction with system stability.

In this section, we concentrate on the eigenvalue problem and the nature of its solutions.

The solution of Eq. (4.167) has the exponential form

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{x} \tag{4.169}$$

where λ is a constant scalar and $\mathbf{x}$ a constant $2n$-vector. Inserting Eq. (4.169) into Eq. (4.167) and dividing through by $e^{\lambda t}$, we obtain the *general algebraic eigenvalue problem*

$$A\mathbf{x} = \lambda\mathbf{x} \tag{4.170}$$

Equation (4.170) admits solutions in the form of the *eigenvalues* λ_i and corresponding *eigenvectors* $\mathbf{x}_i$ $(i = 1, 2, \dots, 2n)$. They satisfy the equations

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i, \qquad i = 1, 2, \dots, 2n \tag{4.171}$$

The question arises naturally as to whether the eigenvectors are orthogonal and whether an expansion theorem exists. We confine ourselves to the case in which all the eigenvalues of A are distinct, from which it follows that all the eigenvectors are independent. Although independent eigenvectors can be rendered orthogonal, mutual orthogonality is not sufficient. Indeed, to serve as a basis for the problem at hand, the eigenvectors must be not only mutually orthogonal but also orthogonal with respect to the matrix A. The eigenvectors $\mathbf{x}_i$ cannot be orthogonal with respect to A, however, because A is not symmetric. But, whereas the eigenvectors are not orthogonal in the ordinary sense, they are orthogonal in some fashion. Before we explore the nature of the orthogonality, we recall that $\det A^T = \det A$, so that

$$\det (A - \lambda I)^T = \det(A^T - \lambda I) = \det(A - \lambda I) \tag{4.172}$$

from which we conclude that *A and A^T possess the same eigenvalues*. Hence, we can write the eigenvalue problem associated with A^T in the form

$$A^T\mathbf{y} = \lambda\mathbf{y} \tag{4.173}$$

The eigenvalue problem for A^T, Eq. (4.173), is referred to as the *adjoint eigenvalue problem* of the eigenvalue problem for A and it admits solutions in the form of the eigenvalues λ_j and the eigenvectors $\mathbf{y}_j$, where $\mathbf{y}_j$ are called *adjoint eigenvectors* $(j = 1, 2, \dots, 2n)$ of the eigenvectors $\mathbf{x}_i$ $(i = 1, 2, \dots, 2n)$. They satisfy the equations

$$A^T\mathbf{y}_j = \lambda_j\mathbf{y}_j, \qquad j = 1, 2, \dots, 2n \tag{4.174}$$

Equations (4.174) can be rewritten in the form

$$\mathbf{y}_j^T A = \lambda_j\mathbf{y}_j^T, \qquad j = 1, 2, \dots, 2n \tag{4.175}$$

Because of their position to the left of the matrix A, the adjoint eigenvectors $\mathbf{y}_j$ are known as *left eigenvectors* of A. Consistent with this, the eigenvectors $\mathbf{x}_j$ are called *right eigenvectors* of A. It is perhaps of interest to note that when A is a real symmetric matrix, $A = A^T$, the adjoint eigenvectors $\mathbf{y}_j$ coincide with the eigenvectors $\mathbf{x}_j$ $(j = 1, 2, \dots, 2n)$, in which case the eigenvalue problem is said to be *self-adjoint*.

As in the symmetric case, *if a constant μ is subtracted from the main diagonal elements of a nonsymmetric matrix A, then the eigenvalues of A are shifted by the same constant μ.* Indeed, Eq. (4.91) is valid for arbitrary matrices A and is in no way restricted to symmetric matrices.

Next, we multiply Eq. (4.171) on the left by $\mathbf{y}_j^T$ and Eq. (4.175) on the right by $\mathbf{x}_i$, subtract the second result from the first and obtain

$$\left(\lambda_i - \lambda_j\right)\mathbf{y}_j^T\mathbf{x}_i = 0 \tag{4.176}$$

But, according to our assumption, all eigenvalues are distinct, so that we must have

$$\mathbf{y}_j^T\mathbf{x}_i = 0, \qquad \lambda_i \neq \lambda_j, \qquad i, j = 1, 2, \ldots, 2n \tag{4.177}$$

Equations (4.177) state that *the right eigenvectors and left eigenvectors of the real nonsymmetric matrix A belonging to distinct eigenvalues are orthogonal.* It must be stressed here that the orthogonality embodied by Eqs. (4.177) is materially different from the mutual orthogonality exhibited by the eigenvectors of a real symmetric matrix. Indeed, the two sets of eigenvectors, *the right eigenvectors* $\mathbf{x}_i$ *and the left eigenvectors* $\mathbf{y}_j$, *are biorthogonal.* Next, we premultiply Eqs. (4.171) by $\mathbf{y}_j^T$, consider Eqs. (4.177) and obtain

$$\mathbf{y}_j^T A\mathbf{x}_i = 0, \qquad \lambda_i \neq \lambda_j \; i, j = 1, 2, \ldots, 2n \tag{4.178}$$

so that *the right and left eigenvectors are biorthogonal with respect to the matrix A as well.* The biorthogonality does not extend to eigenvectors belonging to the same eigenvalue. These pairs of eigenvectors can be normalized by letting $\mathbf{y}_i^T\mathbf{x}_i = 1$ $(i = 1, 2, \ldots, 2n)$, in which case the two sets of eigenvectors become *biorthonormal.* They satisfy the *biorthonormality relations*

$$\mathbf{y}_j^T\mathbf{x}_i = \delta_{ij}, \qquad i, j = 1, 2, \ldots, 2n \tag{4.179}$$

Moreover, premultiplying Eqs. (4.171) by $\mathbf{y}_j^T$ and considering Eqs. (4.179), we obtain

$$\mathbf{y}_j^T A\mathbf{x}_i = \lambda_i\delta_{ij}, \qquad i, j = 1, 2, \ldots, 2n \tag{4.180}$$

The biorthonormality property given by Eqs. (4.179) can be used to develop a more general expansion theorem. Because Eqs. (4.179) involve both the right and left eigenvectors, the new expansion theorem requires both sets of vectors. Moreover, we have the choice of expanding any arbitrary vector $\mathbf{v}$ in a $2n$-dimensional space in terms of the right eigenvectors or in terms of the left eigenvectors. An expansion in terms of the right eigenvectors has the form

$$\mathbf{v} = \sum_{i=1}^{2n} a_i\mathbf{x}_i \tag{4.181}$$

Premultiplying both sides of Eq. (4.181) by $\mathbf{y}_j^T$ and $\mathbf{y}_j^T A$, in sequence, and considering Eqs. (4.179) and (4.180), we obtain

$$a_i = \mathbf{y}_i^T\mathbf{v}, \qquad i = 1, 2, \ldots, 2n \tag{4.182a}$$

$$\lambda_i a_i = \mathbf{y}_i^T A\mathbf{v}, \qquad i = 1, 2, \ldots, 2n \tag{4.182b}$$

On the other hand, an expansion in terms of the left eigenvectors has the form

$$\mathbf{v} = \sum_{j=1}^{2n} b_j \mathbf{y}_j \tag{4.183}$$

where

$$b_j = \mathbf{x}_j^T \mathbf{v}, \qquad j = 1, 2, \ldots, 2n \tag{4.184a}$$

$$\lambda_j b_j = \mathbf{x}_j^T A\mathbf{v}, \qquad j = 1, 2, \ldots, 2n \tag{4.184b}$$

Equations (4.181)–(4.184) represent a *dual expansion theorem.* It should be pointed out that the $2n$-dimensional space under consideration is in general complex. Moreover, because the dual expansion theorem involves both the right and the left eigenvectors, it is necessary to solve the eigenvalue problem twice, once for A and once for A^T.

The preceding developments can be expressed in a compact matrix form. To this end, we introduce the matrix of eigenvalues

$$\Lambda = \mathrm{diag}\,(\lambda_i) \tag{4.185}$$

as well as the matrices of right and left eigenvectors

$$X = [\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_{2n}], \qquad Y = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots\ \mathbf{y}_{2n}] \tag{4.186a, b}$$

Then, the biorthonormality relations, Eqs. (4.179) and (4.180), can be written as

$$Y^T X = I, \qquad Y^T A X = \Lambda \tag{4.187a, b}$$

Equation (4.187a) implies that

$$Y^T = X^{-1} \tag{4.188}$$

so that, instead of solving the eigenvalue problem for A^T, it is possible to obtain the left eigenvectors by inverting the matrix of right eigenvectors. Inserting Eq. (4.188) into Eq. (4.187b), we obtain

$$X^{-1} A X = \Lambda \tag{4.189}$$

Equation (4.189) represents a *similarity transformation*, and the matrices A and Λ are said to be *similar*. Hence, assuming that all eigenvalues are distinct, the matrix A can be diagonalized by means of a similarity transformation. This implies that *the eigenvalues do not change in similarity transformations*, which further implies that *the characteristic polynomial is invariant in similarity transformations.*

The expansion theorem can also be expressed in more compact form. Indeed, introducing the vectors of coefficients $\mathbf{a} = [a_1\ a_2\ \ldots\ a_{2n}]^T$ and $\mathbf{b} = [b_1\ b_2\ \ldots b_{2n}]^T$, Eqs. (4.181) and (4.182) can be expressed as

$$\mathbf{v} = X\mathbf{a} \tag{4.190a}$$

$$\mathbf{a} = Y^T\mathbf{v}, \qquad \Lambda\mathbf{a} = Y^T A\mathbf{v} \tag{4.190b, c}$$

and Eqs. (4.183) and (4.184) as

$$\mathbf{v} = Y\mathbf{b} \tag{4.191a}$$

$$\mathbf{b} = X^T\mathbf{v}, \qquad \Lambda\mathbf{b} = X^T A\mathbf{v} \tag{4.191b, c}$$

At this time, we turn our attention to the solution of Eq. (4.167). This solution consists of the response to the initial excitation $\mathbf{x}(0)$. Recalling Eq. (4.169) and recognizing that there are $2n$ ways in which the exponential form can satisfy Eq. (4.167), we express the solution as the linear combination

$$\mathbf{x}(t) = \sum_{i=1}^{2n} \mathbf{x}_i e^{\lambda_i t} a_i \tag{4.192}$$

To determine the coefficients a_i, we let $t = 0$ in Eq. (4.192) and write

$$\mathbf{x}(0) = \sum_{i=1}^{2n} \mathbf{x}_i a_i \tag{4.193}$$

Premultiplying both sides of Eq. (4.193) by $\mathbf{y}_j^T$ and considering the orthonormality conditions, Eqs. (4.179), we obtain

$$a_i = \mathbf{y}_i^T\mathbf{x}(0), \qquad i = 1, 2, \ldots, 2n \tag{4.194}$$

so that, introducing Eqs. (4.194) into Eq. (4.192), the solution becomes

$$\mathbf{x}(t) = \sum_{i=1}^{2n} \mathbf{x}_i e^{\lambda_i t}\mathbf{y}_i^T\mathbf{x}(0) \tag{4.195}$$

Equation (4.195) can be expressed in a more compact form. Indeed, recalling Eqs. (4.185) and (4.186), the response of the system to initial excitations can be rewritten in the form

$$\mathbf{x}(t) = Xe^{\Lambda t}Y^T\mathbf{x}(0) \tag{4.196}$$

where $e^{\Lambda t}$ can be computed by means of the series

$$e^{\Lambda t} = I + t\Lambda + \frac{t^2}{2!}\Lambda^2 + \frac{t^3}{3!}\Lambda^3 + \ldots \tag{4.197}$$

Convergence of the series is guaranteed, but the rate of convergence depends on $t \max|\lambda_i|$, in which $\max|\lambda_i|$ denotes the magnitude of the eigenvalue of A of largest modulus.

Equation (4.196) requires the solution of the eigenvalue problems for A and A^T. It is possible, however, to determine the response without solving these eigenvalue problems. Indeed, as can be concluded from Eq. (1.109), the solution of Eq. (4.167) is simply

$$\mathbf{x}(t) = e^{At}\mathbf{x}(0) = \Phi(t)\mathbf{x}(0) \tag{4.198}$$

where, from Eq. (1.108),

$$\Phi(t) = e^{At} = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \ldots \tag{4.199}$$

is the transition matrix. Clearly, the two solutions, Eqs. (4.196) and (4.198), must be equivalent. To show this, we premultiply Eqs. (4.187) by X and postmultiply by X^{-1}, consider Eq. (4.188) and obtain

$$XY^T = I, \qquad A = X\Lambda Y^T \tag{4.200a, b}$$

Then, inserting Eq. (4.200b) into Eq. (4.199) and using Eq. (4.200a), we have

$$\begin{aligned} e^{At} &= XY^T + tX\Lambda Y^T + \frac{t^2}{2!}X\Lambda Y^T X\Lambda Y^T + \frac{t^3}{3!}X\Lambda Y^T X\Lambda Y^T X\Lambda^T + \dots \\ &= XY^T + tX\Lambda Y^T + \frac{t^2}{2!}X\Lambda^2 Y^T + \frac{t^3}{3!}X\Lambda^3 Y^T + \dots \\ &= X\left(I + t\Lambda + \frac{t^2}{2!}\Lambda^2 + \frac{t^3}{3!}\Lambda^3 + \dots\right)Y^T = Xe^{\Lambda t}Y^T \end{aligned} \tag{4.201}$$

which verifies the equivalence of the two solutions. It is obvious that the earlier convergence statement applies to solution (4.198) as well.

It should be pointed out here that, although our interest was confined to state equations of order $2n$, the discussion of Eq. (4.167) and its solution is not restricted in any way and is applicable not only to systems of even order but also of odd order.

Example 4.8

Solve the eigenvalue problem for the linearized system of Example 4.4 about the trivial equilibrium, Eqs. (b) with $x_0 = y_0 = 0$, for the parameter values

$$m = 1\,\text{kg}, \quad \Omega = 2\,\text{rad/s}, \quad k_x = 5\,\text{N/m}, \quad k_y = 10\,\text{N/m},$$
$$c = 0.1\,\text{N}\cdot\text{s/m}, \quad h = 0.2\,\text{N}\cdot\text{s/m}$$

Equation (c) of Example 4.4 represents the equations in matrix form. Using Eqs. (d)–(h) of Example 4.4 in conjunction with the parameter values listed above, we obtain the coefficient matrices

$$\begin{aligned} M &= \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C^* = \begin{bmatrix} c+h & 0 \\ 0 & h \end{bmatrix} = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.2 \end{bmatrix} \\ G &= \begin{bmatrix} 0 & -2m\Omega \\ 2m\Omega & 0 \end{bmatrix} = \begin{bmatrix} 0 & -4 \\ 4 & 0 \end{bmatrix} \\ K &= \begin{bmatrix} k_x - m\Omega^2 & 0 \\ 0 & k_y - m\Omega^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix} \\ H &= \begin{bmatrix} 0 & -h\Omega \\ h\Omega & 0 \end{bmatrix} = \begin{bmatrix} 0 & -0.4 \\ 0.4 & 0 \end{bmatrix} \end{aligned} \tag{a}$$

Hence, inserting Eqs. (a) into Eq. (4.168), we can write the system matrix as

$$\begin{aligned} A &= \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}(K+H) & -M^{-1}(C^*+G) \end{array}\right] \\ &= \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0.4 & -0.3 & 4 \\ -0.4 & -6 & -4 & -0.2 \end{bmatrix} \end{aligned} \tag{b}$$

The solution of the eigenvalue problem consists of the eigenvalues

$$\lambda_1 = -0.1128 + 0.5083i, \qquad \lambda_2 = -0.1128 - 0.5083i$$
$$\lambda_3 = -0.1372 + 4.7653i, \qquad \lambda_4 = -0.1372 - 4.7653i \tag{c}$$

the right eigenvectors

$$\mathbf{x}_1 = \begin{bmatrix} -0.8326 - 0.0750i \\ -0.0347 + 0.2946i \\ 0.1320 - 0.4147i \\ -0.1458 - 0.0509i \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} -0.8326 + 0.0750i \\ -0.0347 - 0.2946i \\ 0.1320 + 0.4147i \\ -0.1458 + 0.0509i \end{bmatrix}$$
$$\mathbf{x}_3 = \begin{bmatrix} -0.1279 + 0.0442i \\ -0.0524 - 0.1452i \\ -0.1933 - 0.6158i \\ 0.6990 - 0.2297i \end{bmatrix}, \qquad \mathbf{x}_4 = \begin{bmatrix} -0.1279 - 0.0442i \\ -0.0524 + 0.1452i \\ -0.1933 + 0.6158i \\ 0.6990 + 0.2297i \end{bmatrix} \tag{d}$$

and the left eigenvectors

$$\mathbf{y}_1 = \begin{bmatrix} -0.5723 + 0.0983i \\ -0.1248 - 1.2501i \\ 0.0280 + 0.2992i \\ -0.1064 + 0.0070i \end{bmatrix}, \qquad \mathbf{y}_2 = \begin{bmatrix} -0.5723 - 0.0983i \\ -0.1248 + 1.2501i \\ 0.0280 - 0.2992i \\ -0.1064 - 0.0070i \end{bmatrix}$$
$$\mathbf{y}_3 = \begin{bmatrix} -0.1315 + 0.0152i \\ -0.2000 + 0.8059i \\ -0.1946 + 0.5435i \\ -0.6225 + 0.2135i \end{bmatrix}, \qquad \mathbf{y}_4 = \begin{bmatrix} -0.1315 - 0.0152i \\ -0.2000 - 0.8059i \\ -0.1946 - 0.5435i \\ -0.6225 - 0.2135i \end{bmatrix} \tag{e}$$

where the eigenvectors have been normalized so that $\mathbf{y}_j^T\mathbf{x}_i = \delta_{ij}$ $(i, j = 1, 2, 3, 4)$.

4.9 RESPONSE OF DISCRETE SYSTEMS TO HARMONIC EXCITATIONS

In Sec. 1.5, we considered the response of linear time-invariant systems to harmonic excitations. We recall that the response to harmonic excitations represents a steady-state response, which is to be treated separately from the response to transient excitations, such as initial displacements and velocities. We found it convenient in Sec. 1.5 to express harmonic excitations in exponential form, and we propose to use the same approach for the multi-degree-of-freedom systems considered in this section. Hence, we rewrite Eq. (4.69) as

$$M\ddot{\mathbf{q}}(t) + \left(C^* + G\right)\dot{\mathbf{q}}(t) + (K + H)\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{4.202}$$

where ω is the driving frequency and $\mathbf{Q}_0$ is a constant vector. The various coefficient matrices are as defined in Sec. 4.4.

By analogy with the approach of Sec. 1.5, we express the steady-state solution of Eq. (4.202) in the exponential form

$$\mathbf{x}(t) = e^{i\omega t}\mathbf{X}(i\omega) \tag{4.203}$$

in which $\mathbf{X}(i\omega)$ is a vector of amplitudes depending on the excitation frequency ω. Introducing Eq. (4.203) into Eq. (4.202) and dividing through by $e^{i\omega t}$, we obtain

$$Z(i\omega)\mathbf{X}(i\omega) = \mathbf{Q}_0 \tag{4.204}$$

where

$$Z(i\omega) = -\omega^2 M + i\omega\left(C^* + G\right) + K + H \tag{4.205}$$

is known as the *impedance matrix*. The solution of Eq. (4.204) is simply

$$\mathbf{X}(i\omega) = Z^{-1}(i\omega)\mathbf{Q}_0 \tag{4.206}$$

But, from matrix theory (Appendix B), the inverse of the matrix Z can be determined by means of the formula

$$Z^{-1}(i\omega) = \frac{\text{adj } Z(i\omega)}{\det Z(i\omega)} \tag{4.207}$$

where

$$\text{adj } Z = \left[(-1)^{j+k} \det M_{jk}\right]^T \tag{4.208}$$

is the *adjugate* matrix, in which $(-1)^{j+k}\det M_{jk}$ is the *cofactor* corresponding to the entry Z_{jk} of Z, where M_{jk} is the submatrix obtained by striking out the jth row and kth column from Z. Inserting Eqs. (4.206) and (4.207) into Eq. (4.203), we obtain the response of a discrete system to harmonic excitations in the form

$$\mathbf{q}(t) = e^{i\omega t}\frac{\text{adj } Z(i\omega)\mathbf{Q}_0}{\det Z(i\omega)} \tag{4.209}$$

As in Sec. 1.5, if the excitation is $\cos\omega t\ \mathbf{Q}_0$, the response is Re $\mathbf{q}(t)$, and if the excitation is $\sin\omega t\ \mathbf{Q}_0$, the response is Im $\mathbf{q}(t)$.

Because of difficulties in evaluating the inverse of the impedance matrix, the approach is suitable only for relatively low-order systems. For higher-order systems, it is more efficient to use modal analysis, as described in the next section.

Example 4.9

Derive the response of the damped two-degree-of-freedom system shown in Fig. 4.9 to harmonic excitations.

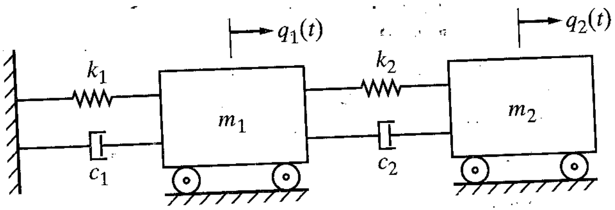


Figure 4.9 Damped two-degree-of-freedom system

In the absence of gyroscopic and circulatory forces, the matrix equation of motion, Eq. (4.202), reduces to

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{a}$$

where the mass matrix M, damping matrix C and stiffness matrix K can be obtained from the kinetic energy, Rayleigh's dissipation function and potential energy, respectively. The kinetic energy is simply

$$T = \frac{1}{2}\left[m_1\dot{q}_1^2(t) + m_2\dot{q}_2^2(t)\right] = \frac{1}{2}\dot{\mathbf{q}}^T(t)M\dot{\mathbf{q}}(t) \tag{b}$$

in which $\mathbf{q}(t) = [q_1(t)\ q_2(t)]^T$ is the two-dimensional configuration vector, $\mathbf{Q}_0 = [Q_{01}\ Q_{02}]^T$ is the force amplitude vector and

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \tag{c}$$

is the diagonal mass matrix. The Rayleigh's dissipation function can be shown to have the form

$$\mathcal{F} = \frac{1}{2}\left\{c_1\dot{q}_1^2(t) + c_2[\dot{q}_2(t) - \dot{q}_1(t)]^2\right\} = \frac{1}{2}\dot{\mathbf{q}}^T(t)C\dot{\mathbf{q}}(t) \tag{d}$$

where

$$C = \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix} \tag{e}$$

is the damping matrix. Moreover, the potential energy has the expression

$$V = \frac{1}{2}\left\{k_1 q_1^2(t) + k_2[q_2(t) - q_1(t)]^2\right\} = \frac{1}{2}\mathbf{q}^T(t)K\mathbf{q}(t) \tag{f}$$

in which

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} \tag{g}$$

is the stiffness matrix.

Equation (4.209) gives the system response in the form

$$\mathbf{q}(t) = e^{i\omega t}\frac{\text{adj}\, Z^T(i\omega)\mathbf{Q}_0}{\det Z(i\omega)} \tag{h}$$

where, using Eq. (4.205), the impedance matrix has the expression

$$Z(i\omega) = \begin{bmatrix} Z_{11}(i\omega) & Z_{12}(i\omega) \\ Z_{12}(i\omega) & Z_{22}(i\omega) \end{bmatrix} = -\omega^2 M + i\omega C + K \tag{i}$$

Using Eqs. (c), (e) and (g), the entries of the impedance matrix are

$$\begin{aligned} Z_{11}(i\omega) &= -\omega^2 m_1 + i\omega(c_1 + c_2) + k_1 + k_2 \\ Z_{12}(i\omega) &= -i\omega c_2 - k_2 \\ Z_{22}(i\omega) &= -\omega^2 m_2 + i\omega c_2 + k_2 \end{aligned} \tag{j}$$

The adjugate matrix is given by

$$\text{adj}\, Z(i\omega) = \begin{bmatrix} Z_{22}(i\omega) & -Z_{12}(i\omega) \\ -Z_{12}(i\omega) & Z_{11}(i\omega) \end{bmatrix} \tag{k}$$

and the determinant has the expression

$$\det Z(i\omega) = |Z(i\omega)| = Z_{11}(i\omega)Z_{22}(i\omega) - Z_{12}^2(i\omega) \tag{l}$$

Inserting Eqs. (k) and (l) into Eq. (h), we can write the response by components

$$\begin{aligned} q_1(t) &= \frac{Z_{22}Q_{01} - Z_{12}Q_{02}}{Z_{11}Z_{22} - Z_{12}^2}e^{i\omega t} \\ q_2(t) &= \frac{-Z_{12}Q_{01} + Z_{11}Q_{02}}{Z_{11}Z_{22} - Z_{12}^2}e^{i\omega t} \end{aligned} \tag{m}$$

Of course, one must retain the real part or the imaginary part of $q_i(t)$ $(i = 1, 2)$, depending on whether the excitation is $\cos\omega t\, \mathbf{Q}_0$ or $\sin\omega t\, \mathbf{Q}_0$, respectively.

4.10 RESPONSE OF MULTI-DEGREE-OF-FREEDOM SYSTEMS TO ARBITRARY EXCITATIONS

In Sec. 4.4, we derived the linearized equations of motion of multi-degree-of-freedom systems in the neighborhood of equilibrium positions in the form of a set of simultaneous ordinary differential equations with constant coefficients, Eq. (4.69). As shown in Chapter 1, the response of generic dynamical systems to arbitrary excitations can be obtained by a method based on the transition matrix. This approach is certainly applicable to the multi-degree-of-freedom systems discussed in this chapter. Moreover, in both Chapter 1 and Chapter 3, we presented methods for deriving the response of low-order systems to a variety of excitations, including arbitrary excitations. A prima facie judgment is likely to lead to the conclusion that the techniques for the response of first-order and second-order systems cannot help very much in the case of multi-degree-of-freedom. This would be an incorrect conclusion, however, because a linear transformation using the modal matrix is capable of reducing a set of simultaneous equations of motion to a set of independent equations, known as *modal equations*, which can indeed be treated by the techniques of Chapters 1 and 3. The procedure for producing the response of multi-degree-of-freedom systems by first solving the eigenvalue problem and then using the modal matrix to decouple the system into modal equations is called *modal analysis*.[1] The approach is extremely efficient for undamped systems and systems with special types of damping. In the case of general nonconservative systems, the response can be obtained by means of the approach based on the transition matrix (Sec. 1.9), or by means of a modal analysis in terms of state modes. In discussing the free vibration problem earlier in this chapter, we found it convenient to distinguish among various classes of systems. We propose to follow a similar pattern here.

i. Undamped natural systems

We are concerned with the case in which the gyroscopic, viscous damping and circulatory forces are zero, in which case Eq. (4.69) reduces to

$$M\ddot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{Q}(t) \tag{4.210}$$

where M and K are $n \times n$ real symmetric mass and stiffness matrices, respectively, and $\mathbf{q}(t)$ and $\mathbf{Q}(t)$ are n-dimensional generalized displacement and force vectors, respectively. The mass matrix is positive definite by definition, but the stiffness matrix is assumed to be only positive semidefinite. The solution of Eq. (4.210) is subject to the initial conditions

$$\mathbf{q}(0) = \mathbf{q}_0, \quad \dot{\mathbf{q}}(0) = \dot{\mathbf{q}}_0 \tag{4.211a, b}$$

Equation (4.210) represents a set of n simultaneous second-order ordinary differential equations describing the motion of a *natural system*. A solution can be produced by expressing the equations in state form, i.e, a set of $2n$ first-order equations obtained by introducing the velocities as auxiliary variables and adding to

[1] The term "modal analysis" is also being used to describe a procedure for identifying the dynamic characteristics of a system experimentally.

Eq. (4.210) the n identities $\dot{\mathbf{q}} = \dot{\mathbf{q}}$, as shown in Sec. 4.5. The state equations can be solved by the method based on the transition matrix discussed in Sec. 1.9. However, in the case of undamped natural systems there are more efficient methods of solution. Indeed, a linear transformation with the modal matrix as the transformation matrix permits the reduction of Eq. (4.210) to a set of independent equations lending themselves to ready solution. To this end, it is first necessary to solve the eigenvalue problem defined by

$$K\mathbf{u} = \lambda M\mathbf{u}, \qquad \lambda = \omega^2 \tag{4.212}$$

As discussed in Sec. 4.6, the solution of Eq. (4.212) consists of n eigenvalues $\lambda_i = \omega_i^2$, in which ω_i are the natural frequencies, and n eigenvectors, or modal vectors $\mathbf{u}_i$ $(i = 1, 2, \ldots, n)$. The eigenvectors are orthogonal and have been normalized so as to satisfy the orthonormality relations given by Eqs. (4.123) and (4.124), or Eqs. (4.127) and (4.128), where in the latter Λ is the diagonal matrix of eigenvalues and U is the orthonormal matrix of eigenvectors.

In view of the expansion theorem, Eq. (4.129), the solution of Eq. (4.212) can be expressed as the linear combination

$$\mathbf{q}(t) = \sum_{i=1}^{n} \eta_i(t)\mathbf{u}_i = U\boldsymbol{\eta}(t) \tag{4.213}$$

where $\boldsymbol{\eta}(t) = [\eta_1(t)\ \eta_2(t) \ldots \eta_n(t)]^T$ is an n-vector of *principal coordinates*, *natural coordinates*, or *modal coordinates*. Inserting Eq. (4.213) into Eq. (4.210), premultiplying the result by U^T and considering the orthonormality conditions, Eqs. (4.127) and (4.128), we obtain

$$\ddot{\boldsymbol{\eta}}(t) + \Lambda\boldsymbol{\eta}(t) = \mathbf{N}(t) \tag{4.214}$$

in which

$$\mathbf{N}(t) = U^T\mathbf{Q}(t) \tag{4.215}$$

is an n-vector of *generalized forces*, or *modal forces* with the components

$$N_i(t) = \mathbf{u}_i^T\mathbf{Q}(t), \qquad i = 1, 2, \ldots, n \tag{4.216}$$

Equation (4.214) represents a set of n independent *modal equations*. Assuming that the system possesses r zero eigenvalues, the modal equations can be written in the scalar form

$$\ddot{\eta}_i(t) = N_i(t), \qquad i = 1, 2, \ldots, r \tag{4.217a}$$

$$\ddot{\eta}_i(t) + \omega_i^2\eta_i(t) = N_i(t), \qquad i = r+1, r+2, \ldots, n \tag{4.217b}$$

We recall from Sec. 4.6 that zero eigenvalues are associated with rigid-body modes. Hence, Eqs. (4.217a) and (4.217b) represent the differential equations for the rigid-body modes and elastic modes, respectively. Because $\eta_i(t)$ represent coordinates with respect to an orthonormal basis, $\eta_i(t)$ are also known as *normal coordinates*.

Equations (4.217) constitute a set of independent second-order ordinary differential equations of the type discussed in more general form in Chapter 3. Indeed, in Sec. 3.1 we discussed the homogeneous solution of second-order systems, i.e., the solution to initial excitations, and in Secs. 3.2, 3.4 and 3.5 we discussed the particular solution. More specifically, we discussed the solution to harmonic, periodic and arbitrary external forces. Of course, for linear time-invariant systems the principle of superposition applies, so that the total solution is the sum of the homogeneous and particular solutions. This statement should really be limited to transient response only. Indeed, we recall from Chapter 3 that the response to harmonic excitations or periodic excitations is a steady-state response, and it is not meaningful to combine a steady-state response with the response to initial conditions, which is a transient response.

In this section, we concentrate on the response to transient excitations and propose to derive the solution to Eqs. (4.217) subject to initial and external excitations simultaneously, which requires expressions for the initial modal displacements and velocities. To this end, we let $t = 0$ in Eq. (4.213), use Eq. (4.211a) and write

$$\mathbf{q}(0) = \mathbf{q}_0 = \sum_{i=1}^{n} \eta_i(0)\mathbf{u}_i \tag{4.218}$$

Then, premultiplying both sides of Eq. (4.218) by $\mathbf{u}_j^T M$ and considering Eqs. (4.123), we obtain the initial modal displacements

$$\eta_i(0) = \mathbf{u}_i^T M \mathbf{q}_0, \qquad i = 1, 2, \ldots, n \tag{4.219a}$$

Following the same pattern in conjunction with Eq. (4.211b), it is easy to verify that the initial modal velocities are given by

$$\dot{\eta}_i(0) = \mathbf{u}_i^T M \dot{\mathbf{q}}_0, \qquad i = 1, 2, \ldots, n \tag{4.219b}$$

Then, using developments from Chapter 3, the complete solution of Eqs. (4.217) can be shown to be

$$\eta_i(t) = \int_0^t \left[\int_0^\tau N_i(\sigma) d\sigma \right] d\tau + \eta_i(0) + t\dot{\eta}_i(0), \quad i = 1, 2, \ldots, r \tag{4.220a}$$

$$\eta_i(t) = \frac{1}{\omega_i} \int_0^t N_i\,(t - \tau) \sin \omega_i \tau d\tau + \eta_i(0) \cos \omega_i t + \frac{\dot{\eta}_i(0)}{\omega_i} \sin \omega_i t, \qquad i = r + 1, r + 2, \ldots, n \tag{4.220b}$$

The formal solution of Eq. (4.210) is obtained by inserting Eqs. (4.216), (4.219) and (4.220) into Eq. (4.213), with the result

$$\begin{aligned}\mathbf{q}(t) = {} & \sum_{i=1}^{r} \left\{ \int_0^t \left[\mathbf{u}_i^T \int_0^\tau \mathbf{Q}(\sigma) d\sigma \right] d\tau + \mathbf{u}_i^T M \left(\mathbf{q}_0 + t\dot{\mathbf{q}}_0\right) \right\} \mathbf{u}_i \\ & + \sum_{i=r+1}^{n} \left[\frac{1}{\omega_i} \mathbf{u}_i^T \int_0^t \mathbf{Q}\,(t - \tau) \sin \omega_i \tau d\tau + \mathbf{u}_i^T M \mathbf{q}_0 \cos \omega_i t \right. \\ & \left. + \frac{\mathbf{u}_i^T M \dot{\mathbf{q}}_0}{\omega_i} \sin \omega_i t \right] \mathbf{u}_i \end{aligned} \tag{4.221}$$

The just described procedure for obtaining the response of multi-degree-of-freedom natural conservative systems by reducing the equations of motion to a set of independent equations for the modal coordinates is commonly known as the *classical modal analysis*.

The classical modal analysis has the significant advantage that it only requires the solution of a real symmetric eigenvalue problem, which is by far the most desirable type. In fact, this feature is so attractive that the question arises whether the classical modal analysis can be applied to systems other than natural conservative ones. The answer is affirmative in the case of systems with proportional viscous damping and a qualified yes in the case of systems with small viscous damping, or with structural damping.

ii. Systems with proportional viscous damping

The equations of motion for a viscously damped system can be obtained from Eq. (4.69) by ignoring the gyroscopic and circulatory forces. The result is

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{Q}(t) \tag{4.222}$$

where C is the $n \times n$ real symmetric damping matrix. Introducing the linear transformation given by Eq. (4.213) into Eq. (4.222), premultiplying the result by U^T and considering the orthonormality relations, Eqs. (4.127) and (4.128), we obtain

$$\ddot{\boldsymbol{\eta}}(t) + C'\dot{\boldsymbol{\eta}}(t) + \Lambda\boldsymbol{\eta}(t) = \mathbf{N}(t) \tag{4.223}$$

where $\mathbf{N}(t)$ is given by Eq. (4.215) and

$$C' = U^T C U \tag{4.224}$$

is a real symmetric matrix. In general, C' is not diagonal, in which case the classical modal analysis is not able to reduce the equations of motion to an independent set.

In the special case in which the damping matrix is a linear combination of the mass matrix and stiffness matrix of the form

$$C = \alpha M + \beta K \tag{4.225}$$

where α and β are real constant scalars, C' reduces to the diagonal matrix

$$C' = \alpha I + \beta\Lambda \tag{4.226}$$

Viscous damping characterized by a matrix of the form given by Eq. (4.225) is known as *proportional damping*. Clearly, the classical modal analysis is capable of decoupling systems with proportional damping. To obtain the response, it is convenient to introduce the notation

$$C' = \operatorname{diag}\left(\alpha_i + \beta_i\omega_i^2\right) = \operatorname{diag}\left(2\zeta_i\omega_i\right) \tag{4.227}$$

so that Eq. (4.223) can be written in the form of the independent equations

$$\ddot{\eta}_i(t) + 2\zeta_i\omega_i\dot{\eta}_i(t) + \omega_i^2\eta_i(t) = N_i(t), \qquad i = 1, 2, \ldots, n \tag{4.228}$$

where $N_i(t)$ are as given by Eq. (4.216). Note that Eqs. (4.228) imply that there are no rigid-body modes. If rigid-body modes are present, then Eqs. (4.228) must be modified by analogy with Eqs. (4.217). Equations of the type (4.228) were studied in Secs. 3.1 and 3.5 in conjunction with the response of single-degree-of-freedom systems to arbitrary excitations. Hence, using the analogy with Eqs. (3.26) and (3.83), the solution of Eqs. (4.228) is simply

$$\begin{aligned}\eta_i(t) &= \frac{1}{\omega_{di}}\int_0^t N_i(t-\tau)e^{-\zeta_i\omega_i t}\sin\omega_{di}\tau\,d\tau \\ &\quad + e^{-\zeta_i\omega_i t}\left[\eta_i(0)\left(\cos\omega_{di}t + \frac{\zeta_i\omega_i}{\omega_{di}}\sin\omega_{di}t\right) + \frac{\dot{\eta}_i(0)}{\omega_{di}}\sin\omega_{di}t\right], \\ &\qquad i = 1, 2, \ldots, n \end{aligned} \tag{4.229}$$

in which

$$\omega_{di} = \left(1-\zeta_i^2\right)^{1/2}\omega_i, \qquad i = 1, 2, \ldots, n \tag{4.230}$$

is the frequency of damped oscillation in the ith mode. The modal initial conditions are as given by Eqs. (4.219). The formal solution of Eq. (4.222) is obtained by introducing Eqs. (4.229) into Eq. (4.213).

The assumption of proportional viscous damping is made quite frequently, many times only implicitly.

There is another case in which the classical modal transformation decouples a viscously damped system. Indeed, it was shown in Ref. 3 that, if the matrices $M^{-1}C$ and $M^{-1}K$ commute, then a linear transformation involving the modal matrix U is once again capable of decoupling the equations of motion. This case is not very common.

iii. Systems with small damping

We are concerned with the case in which damping is small, although not necessarily of the proportional type. The implication is that the entries of the matrix C are one order of magnitude smaller than the entries of the matrices M and K. This permits a perturbation solution of Eq. (4.222). A first-order perturbation solution has the form

$$\mathbf{q}(t) = \mathbf{q}_0(t) + \mathbf{q}_1(t) \tag{4.231}$$

where the subscripts 0 and 1 denote zero-order and first-order quantities, respectively, with zero-order quantities being one order of magnitude larger than first-order quantities. Inserting Eq. (4.231) into Eq. (4.222), regarding C as a first-order quantity, separating terms of different order of magnitude and ignoring second-order terms, we obtain the zero-order equation

$$M\ddot{\mathbf{q}}_0(t) + K\mathbf{q}_0(t) = \mathbf{Q}(t) \tag{4.232}$$

and the first-order equation

$$M\ddot{\mathbf{q}}_1 + K\mathbf{q}_1(t) = -C\mathbf{q}_0(t) \tag{4.233}$$

Equations (4.232) and (4.233) can both be solved by the classical modal analysis for undamped systems described earlier in this section. As with any perturbation solution, Eq. (4.232) can be solved for $\mathbf{q}_0(t)$ first, independently of Eq. (4.233). Then, inserting the zero-order solution $\mathbf{q}_0(t)$ into Eq. (4.233), the first-order equation, with $-C\mathbf{q}_0(t)$ playing the role of an excitation force, can be solved for the first-order perturbation $\mathbf{q}_1(t)$.

It should be mentioned here that it is common practice to treat systems with small damping by assuming that the off-diagonal terms in C' are of second order in magnitude, and hence sufficiently small to be ignored. The perturbation approach defined by Eqs. (4.232) and (4.233) does not support such an assumption.

iv. Systems with structural damping

The concept of structural damping introduced in Sec. 3.3 for single-degree-of-freedom systems can be extended to multi-degree-of-freedom systems provided that *all the excitation forces are harmonic and of the same frequency*. Under these circumstances, Eq. (4.222) takes the special form

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{4.234}$$

where $\mathbf{Q}_0$ is a constant n-vector. Consistent with Eq. (3.58) for single-degree-of-freedom systems, we invoke the analogy between viscous damping, and structural damping and introduce the *hysteretic damping matrix* (Ref. 8)

$$C = \frac{1}{\pi\omega}\alpha \tag{4.235}$$

in which α is an $n \times n$ symmetric matrix of coefficients. Inserting Eq. (4.235) into Eq. (4.234), we obtain

$$M\ddot{\mathbf{q}}(t) + \frac{1}{\pi\omega}\alpha\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{4.236}$$

so that structurally damped multi-degree-of-freedom systems can be treated as if they were viscously damped. It is customary to assume that the hysteretic damping matrix is proportional to the stiffness matrix, or

$$\alpha = \pi\gamma K \tag{4.237}$$

where γ is a *structural damping factor*, so that Eq. (4.236) can be rewritten as

$$M\ddot{\mathbf{q}}(t) + \frac{\gamma}{\omega}K\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{4.238}$$

Moreover, for harmonic oscillation,

$$\dot{\mathbf{q}}(t) = i\omega\mathbf{q}(t) \tag{4.239}$$

so that Eq. (4.238) reduces to

$$M\ddot{\mathbf{q}}(t) + (1 + i\gamma)K\mathbf{q}(t) = e^{i\omega t}\mathbf{Q}_0 \tag{4.240}$$

in which $(1 + i\gamma)K$ is a *complex stiffness matrix*.

Equation (4.240) is in a form that lends itself to a solution by the classical modal analysis. Indeed, following the steps outlined earlier in this section, we obtain the independent modal equations

$$\ddot{\eta}_r(t) + (1 + i\gamma)\omega_r^2\eta_r(t) = e^{i\omega t}N_r, \qquad r = 1, 2, \ldots, n \tag{4.241}$$

where

$$N_r = \mathbf{u}_r^T \mathbf{Q}_0, \qquad r = 1, 2, \ldots, n \tag{4.242}$$

are amplitudes of the modal forces. The solution of Eq. (4.241) is simply

$$\eta_r(t) = \frac{e^{i\omega t} N_r}{(1 + i\gamma)\,\omega_r^2 - \omega^2}, \qquad r = 1, 2, \ldots, n \tag{4.243}$$

from which we obtain the solution of Eq. (4.240) in the form

$$\mathbf{q}(t) = \sum_{r=1}^{n} \frac{e^{i\omega t} \mathbf{u}_r^T \mathbf{Q}_0}{(1 + i\gamma)\,\omega_r^2 - \omega^2} \mathbf{u}_r \tag{4.244}$$

The extension of the concept of structural damping to multi-degree-of-freedom systems hinges on the assumption that the analogy with viscous damping is as given by Eqs. (4.235) and (4.237). This assumption has not received sufficient experimental substantiation, so that the results presented here must be used judiciously.

v. Undamped gyroscopic systems

The equations of motion for undamped gyroscopic systems can be obtained from Eq. (4.69) by excluding the viscous damping and circulatory forces, with the result

$$M\ddot{\mathbf{q}}(t) + G\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{Q}(t) \tag{4.245}$$

where G is the $n \times n$ skew symmetric gyroscopic matrix. It is easy to verify that in this case the classical modal analysis fails to produce a meaningful solution, as the matrix $U^T G U$ is an $n \times n$ null matrix, so that the corresponding modal equations do not contain gyroscopic terms. This erroneous result is due to the fact that the classical modal matrix U is real, and the eigenvectors associated with Eq. (4.245) are complex. Moreover, as shown in Sec. 4.7, no solution to the eigenvalue problem is possible in the configuration space, so that the problem must be cast in the state space.

Following the approach of Sec. 4.7, Eq. (4.245) can be expressed in the state form

$$M^*\dot{\mathbf{x}}(t) = -G^*\mathbf{x}(t) + \mathbf{X}(t) \tag{4.246}$$

where $\mathbf{x}(t) = \left[\mathbf{q}^T(t)\ \dot{\mathbf{q}}^T(t)\right]^T$ is the $2n$-dimensional state vector, $\mathbf{X}(t) = \left[\mathbf{0}^T\ \mathbf{Q}^T(t)\right]^T$ is the associated $2n$-dimensional excitation vector and

$$M^* = \begin{bmatrix} K & 0 \\ 0 & M \end{bmatrix} = M^{*T}, \qquad G^* = \begin{bmatrix} 0 & -K \\ K & G \end{bmatrix} = -G^{*T} \tag{4.247a, b}$$

are $2n \times 2n$ coefficient matrices, the first real symmetric and positive definite and the second real and skew symmetric. The eigenvalue problem can be expressed in the real form

$$\omega_r M^* \mathbf{y}_r = -G^* \mathbf{z}_r, \qquad \omega_r M^* \mathbf{z}_r = G^* \mathbf{y}_r, \qquad r = 1, 2, \ldots, n \tag{4.248a, b}$$

where $\mathbf{y}_r$ and $\mathbf{z}_r$ are the real and imaginary parts of the complex state eigenvector $\mathbf{x}_r$ $(r = 1, 2, \ldots, n)$. Of course, the complex conjugate $\bar{\mathbf{x}}_r$ is also an eigenvector, but this is inconsequential here because we work with real quantities alone. The solution to the real eigenvalue problem, Eqs. (4.248), was discussed in Sec. 4.7.

The general solution of the Eq. (4.246) can be obtained by means of a modal analysis for gyroscopic systems developed in Ref. 11, which is essentially the solution presented here. To this end, we introduce the $2n \times 2n$ real modal matrix

$$P = [\mathbf{y}_1\ \mathbf{z}_1\ \mathbf{y}_2\ \mathbf{z}_2\ \ldots\ \mathbf{y}_n\ \mathbf{z}_n] \tag{4.249}$$

Then, using results from Sec. 4.7, it is easy to verify that the modal matrix satisfies the orthonormality equations

$$P^T M^* P = I, \qquad P^T G^* P = A \tag{4.250a, b}$$

in which

$$A = \text{block} - \text{diag}\begin{bmatrix} 0 & -\omega_r \\ \omega_r & 0 \end{bmatrix} \tag{4.251}$$

Next, we consider the linear transformation

$$\mathbf{x}(t) = \sum_{r=1}^{n} [\xi_r(t)\mathbf{y}_r + \eta_r(t)\mathbf{z}_r] = P\mathbf{w}(t) \tag{4.252}$$

where

$$\mathbf{w}(t) = [\xi_1(t)\ \eta_1(t)\ \xi_2(t)\ \eta_2(t)\ \ldots\ \xi_n(t)\ \eta_n(t)]^T \tag{4.253}$$

is a $2n$-vector of modal coordinates. Introducing Eq. (4.252) into Eq. (4.246) and premultiplying both sides by P^T, we obtain

$$\dot{\mathbf{w}}(t) = -A\mathbf{w}(t) + \mathbf{W}(t) \tag{4.254}$$

in which

$$\mathbf{W}(t) = [Y_1(t)\ Z_1(t)\ Y_2(t)\ Z_2(t)\ \ldots\ Y_n(t)\ Z_n(t)]^T = P^T\mathbf{X}(t) \tag{4.255}$$

where

$$Y_r(t) = \mathbf{y}_r^T\mathbf{X}(t), \qquad Z_r(t) = \mathbf{z}_r^T\mathbf{X}(t) \tag{4.256}$$

Whereas the matrix A is not diagonal, the decoupling is just as effective, as A is block-diagonal and every block is 2×2. Indeed, Eq. (4.254) represents a set of n independent pairs of first-order equations, having the explicit form

$$\dot{\xi}_r(t) - \omega_r\eta_r(t) = Y_r(t), \qquad \dot{\eta}_r(t) + \omega_r\xi_r(t) = Z_r(t), \qquad r = 1, 2, \ldots, n \tag{4.257}$$

Clearly, each pair of equations can be solved for the conjugate modal coordinates $\xi_r(t)$, $\eta_r(t)$ independently of any other pair.

The solution of Eqs. (4.257) can be obtained conveniently by means of the Laplace transformation. Letting $\overline{\xi}_r(s) = \mathcal{L}\xi_r(t)$, $\overline{\eta}_r(s) = \mathcal{L}\eta_r(t)$, $\overline{Y}_r(s) = \mathcal{L}Y_r(t)$ and $\overline{Z}_r(s) = \mathcal{L}Z_r(t)$ and transforming both sides of Eqs. (4.257), we obtain

$$\begin{aligned} s\overline{\xi}_r(s) - \xi_r(0) - \omega_r\overline{\eta}_r(s) &= \overline{Y}_r(s), \\ s\overline{\eta}_r(s) - \eta_r(0) + \omega_r\overline{\xi}_r(s) &= \overline{Z}_r(s), \end{aligned} \qquad r = 1, 2, \ldots, n \tag{4.258}$$

where the initial modal coordinates $\xi_r(0)$ and $\eta_r(0)$ can be obtained from Eq. (4.252) with $t = 0$ and the orthonormality relations, Eq. (4.250a), as follows:

$$\xi_r(0) = \mathbf{y}_r^T M^* \mathbf{x}(0), \qquad \eta_r(0) = \mathbf{z}_r^T M^* \mathbf{x}(0) \tag{4.259}$$

in which $\mathbf{x}(0)$ is the initial state vector. Equations (4.258) can be solved for the pair $\overline{\xi}_r(s), \overline{\eta}_r(s)$ of transformed modal coordinates, with the result

$$\begin{aligned}\overline{\xi}_r(s) &= \frac{1}{s^2 + \omega_r^2}\left[s\overline{Y}_r(s) + \omega_r \overline{Z}_r(s) + s\xi_r(0) + \omega_r \eta_r(0)\right],\\ \overline{\eta}_r(s) &= \frac{1}{s^2 + \omega_r^2}\left[s\overline{Z}_r(s) - \omega_r \overline{Y}_r(s) + s\eta_r(0) - \omega_r \xi_r(0)\right],\\ &\qquad r = 1, 2, \dots, n\end{aligned} \tag{4.260}$$

Using the convolution theorem (Appendix A) and considering Eqs. (4.256) and (4.259), we can write the solution of Eqs. (4.260) in terms of convolution integrals in the form

$$\begin{aligned}\xi_r(t) &= \int_0^t \left[\mathbf{y}_r^T \mathbf{X}(\tau) \cos\omega_r(t-\tau) + \mathbf{z}_r^T \mathbf{X}(\tau) \sin\omega_r(t-\tau)\right] d\tau\\ &\quad + \mathbf{y}_r^T M^* \mathbf{x}(0) \cos\omega_r t + \mathbf{z}_r^T M^* \mathbf{x}(0) \sin\omega_r t,\\ \eta_r(t) &= \int_0^t \left[\mathbf{z}_r^T \mathbf{X}(\tau) \cos\omega_r(t-\tau) - \mathbf{y}_r^T \mathbf{X}(\tau) \sin\omega_r(t-\tau)\right] d\tau\\ &\quad + \mathbf{z}_r^T M^* \mathbf{x}(0) \cos\omega_r t - \mathbf{y}_r^T M^* \mathbf{x}(0) \sin\omega_r t,\\ &\qquad r = 1, 2, \dots, n\end{aligned} \tag{4.261}$$

Finally, inserting Eqs. (4.261) into series (4.252), we obtain the complete response of a conservative gyroscopic system in the form of the state vector

$$\begin{aligned}\mathbf{x}(t) = \sum_{r=1}^{n} \Bigg\{ \int_0^t &\Big[\left(\mathbf{y}_r\mathbf{y}_r^T + \mathbf{z}_r\mathbf{z}_r^T\right)\mathbf{X}(\tau)\cos\omega_r(t-\tau)\\ &+ \left(\mathbf{y}_r\mathbf{z}_r^T - \mathbf{z}_r\mathbf{y}_r^T\right)\mathbf{X}(\tau)\sin\omega_r(t-\tau)\Big]\, d\tau\\ + \left(\mathbf{y}_r\mathbf{y}_r^T + \mathbf{z}_r\mathbf{z}_r^T\right) M^*\mathbf{x}(0)\cos\omega_r t &+ \left(\mathbf{y}_r\mathbf{z}_r^T - \mathbf{z}_r\mathbf{y}_r^T\right) M^*\mathbf{x}(0)\sin\omega_r t\Bigg\}\end{aligned} \tag{4.262}$$

Equation (4.262) represents the transient response, which implies that the force vector $\mathbf{Q}(t)$, and hence $\mathbf{X}(t)$, is defined only for $t \geq 0$ and is zero for $t < 0$. In the case of steady-state excitations, one can derive the response starting with Eqs. (4.257) directly.

vi. General nonconservative systems

General nonconservative systems differ from conservative systems in two important ways, namely, they cannot be described by a single real symmetric matrix, which implies that the eigensolutions are complex for the most part, and solutions of the equations of motion require a state space description. Of course, there are special cases of nonconservative systems that can be treated in the configuration space by the classical modal analysis, but here the interest lies in systems that do not lend themselves to such a treatment. There is a modal analysis designed especially for arbitrary viscously damped systems (Refs. 8 and 12), but its advantages over other methods for treating general nonconservative systems are questionable, as the eigenvalues still tend to be complex and solutions still require a state space formulation. In view of this, we choose to discuss general nonconservative systems directly.

Equation (4.69) expresses the equations of motion for a general nonconservative system in the matrix notation

$$M\ddot{\mathbf{q}}(t) + \left(C^* + G\right)\dot{\mathbf{q}}(t) + (K + H)\mathbf{q} = \mathbf{Q} \tag{4.263}$$

in which the various coefficient matrices have been defined in Sec. 4.4. Adding the identity $\dot{\mathbf{q}}(t) = \dot{\mathbf{q}}(t)$, Eq. (4.263) can be cast in the state form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{Q}(t) \tag{4.264}$$

where $\mathbf{x}(t) = \left[\mathbf{q}^T(t)\, \dot{\mathbf{q}}^T(t)\right]^T$ is the $2n$-dimensional state vector and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}(K + H) & -M^{-1}(C^* + G) \end{array}\right], \qquad B = \left[\begin{array}{c} 0 \\ \hline M^{-1} \end{array}\right] \tag{4.265a,b}$$

are $2n \times 2n$ and $2n \times n$ coefficient matrices, respectively, both real.

Equations of the type given by Eq. (4.264) were discussed in Chapter 1. Indeed, from Sec. 1.9, we can write the solution of Eq. (4.264) directly in the form

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t - \tau)B\mathbf{Q}(\tau)d\tau \tag{4.266}$$

in which $\Phi(t - \tau)$ is the transition matrix, given by the infinite series

$$\begin{aligned}\Phi(t - \tau) &= \Phi(t, \tau) = e^{A(t-\tau)} \\ &= I + (t - \tau)A + \frac{(t - \tau)^2}{2!}A^2 + \frac{(t - \tau)^3}{3!}A^3 + \dots \end{aligned} \tag{4.267}$$

The transition matrix was encountered several times before in this text. Clearly, Eq. (4.266) contains both the homogeneous and particular solutions, and we recall that we obtained the homogeneous solution in Sec. 4.8 in the form of Eq. (4.198). The convergence of the series for the transition matrix is guaranteed, but the rate of convergence depends on the eigenvalue of A of largest modulus and the time interval $t - \tau$.

In a limited number of cases, evaluation of the response can be carried out in closed form. This is often the case when the system is of low order and the excitation forces are relatively simple. In particular, the order of the system must be sufficiently low that the transition matrix can be derived by means of the inverse Laplace transform formula (see Sec. 1.9)

$$\Phi(t) = \mathcal{L}^{-1}(sI - A)^{-1} \tag{4.268}$$

which implies that the order must be sufficiently low so as to permit closed-form inversion of the matrix $sI - A$. The fact that the transition matrix can be obtained in closed form does not guarantee that the response can be obtained in closed form. Indeed, in addition, the excitation forces must represent sufficiently simple functions that the integral in Eq. (4.266) can be evaluated in closed form.

We conclude from the above that in most cases evaluation of the response by means of Eq. (4.266) must be carried out numerically, which implies repeated computation of the transition matrix. For practical reasons, inclusion of an infinite number of terms in the series is not possible, so that the transition matrix must be approximated by truncating the series. An approximation including terms through nth power in A only has the form

$$\Phi_n = I + tA + \frac{t^2}{2!}A^2 + \frac{t^3}{3!}A^3 + \ldots + \frac{t^n}{n!}A^n \tag{4.269}$$

where we let $\tau = 0$ in Eq. (4.267) for simplicity. The computation of Φ_n can be carried out efficiently by rewriting Eq.(4.269) as

$$\Phi_n = I + tA\left(I + \frac{t}{2}A\left(I + \frac{t}{3}A\left(I + \ldots + \frac{t}{n-1}A\left(I + \frac{t}{n}A\right)\ldots\right)\right)\right) \tag{4.270}$$

and using the recursive process

$$\begin{aligned}
\psi_1 &= I + \frac{t}{n}A \\
\psi_2 &= I + \frac{t}{n-1}A\psi_1 \\
\psi_3 &= I + \frac{t}{n-2}A\psi_2 \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\Phi_n &= \psi_n = I + tA\psi_{n-1}
\end{aligned} \tag{4.271}$$

The computation of Φ_n by means of Eqs. (4.271) requires $n-1$ matrix multiplications.

The transition matrix possesses the properties

$$\Phi(t, t) = I \tag{4.272}$$

$$\Phi(t_1, t_3) = \Phi(t_1, t_2)\Phi(t_2, t_3) \tag{4.273}$$

Moreover, letting $t_3 = t_1$ in Eq. (4.273) and considering Eq. (4.272), we conclude that

$$\Phi^{-1}(t_1, t_2) = \Phi(t_2, t_1) \tag{4.274}$$

Equation (4.273), referred to as the *semigroup property*, can be used to expedite the convergence of the series (4.267) by dividing the interval $t - \tau$ into $k+1$ smaller subintervals and writing

$$\Phi(t, \tau) = \Phi(t, t_k)\Phi(t_k, t_{k-1}) \dots \Phi(t_2, t_1)\Phi(t_1, \tau) \tag{4.275}$$

Now the computation of the transition matrix requires the continuous product of $k+1$ transition matrices, but the convergence of each of these transition matrices is considerably faster than the convergence of the overall transition matrix.

The remarkable aspect of solution (4.266) is that it does not require the solution of the eigenvalue problem for the coefficient matrix A. On the other hand, it does require repeated evaluation of the transition matrix. Of course, if the solution of the eigenvalue problem for A is available, then the derivation of the response by means of a modal analysis for general dynamic systems is quite efficient.

We recall from Sec. 4.8 that the solution of the eigenvalue problem for an arbitrary real matrix A consists of the eigenvalues λ_i, right eigenvectors $\mathbf{x}_i$ and left eigenvectors $\mathbf{y}_i$ $(i = 1, 2, \dots, 2n)$. The corresponding matrices are Λ, X and Y, respectively. The two sets of eigenvectors are biorthogonal and can be normalized so as to satisfy Eqs. (4.187). In view of the expansion theorem, Eqs. (4.190), we consider a solution of Eq. (4.264) in the form

$$\mathbf{x}(t) = X\boldsymbol{\zeta}(t) \tag{4.276}$$

where $\boldsymbol{\zeta}(t) = [\zeta_1(t)\ \zeta_2(t)\ \dots\ \zeta_{2n}(t)]^T$ is a $2n$-vector of modal coordinates. Introducing Eq. (4.276) into Eq. (4.264), premultiplying both sides of the resulting equation by Y^T and considering the biorthonormality relations, Eqs. (4.187), we obtain

$$\dot{\boldsymbol{\zeta}}(t) = \Lambda\boldsymbol{\zeta}(t) + \mathbf{Z}(t) \tag{4.277}$$

where

$$\mathbf{Z}(t) = Y^T B\mathbf{Q}(t) \tag{4.278}$$

is a modal force vector. Equation (4.277) represents a set of $2n$ independent equations of the form

$$\dot{\zeta}_i(t) = \lambda_i\zeta_i(t) + Z_i(t), \qquad i = 1, 2, \dots, 2n \tag{4.279}$$

Using results obtained in Secs. 1.4 and 1.8, the solution of Eqs. (4.279) can be shown to be

$$\zeta_i(t) = e^{\lambda_i t}\zeta_i(0) + \int_0^t e^{\lambda_i(t-\tau)}Z_i(\tau)d\tau, \qquad i = 1, 2, \dots, 2n \tag{4.280}$$

where, premultiplying Eq. (4.276) by $\mathbf{y}_j^T$, letting $t = 0$ and using the orthonormality relations, the initial modal coordinates are

$$\zeta_i(0) = \mathbf{y}_i^T\mathbf{x}(0), \qquad i = 1, 2, \dots, 2n \tag{4.281}$$

The formal solution is obtained by inserting Eqs. (4.280) in conjunction with Eqs. (4.281) into Eq. (4.276).

At this point, it may prove of interest to show that the solution obtained by means of modal analysis is equivalent to that obtained by the transition matrix. To this end, we introduce Eq. (4.201) into Eq. (4.266), recall Eq. (4.267) and write

$$\mathbf{x}(t) = Xe^{\Lambda t}Y^T\mathbf{x}(0) + \int_0^t Xe^{\Lambda(t-\tau)}Y^T B\mathbf{Q}(\tau)d\tau \tag{4.282}$$

Next, we premultiply Eq. (4.282) by Y^T, use Eqs. (4.276) and (4.278), consider Eq. (4.187a) and obtain

$$\boldsymbol{\zeta}(t) = e^{\Lambda t}\boldsymbol{\zeta}(0) + \int_0^t e^{\Lambda(t-\tau)}\mathbf{Z}(\tau)d\tau \tag{4.283}$$

It is not difficult to see that Eq. (4.283) represents the matrix form of Eqs. (4.280).

Example 4.10

Derive the response of the three-degree-of-freedom system of Example 4.6 to the excitation

$$Q_1(t) = Q_2(t) = 0, \qquad Q_3(t) = Q_0 u(t) \tag{a}$$

where Q_0 is a constant and $u(t)$ is the unit step function.

Recognizing that the system has no rigid-body modes, the response can be obtained from Eq. (4.221) in the form

$$\mathbf{q}(t) = \sum_{i=1}^{3} \frac{\mathbf{u}_i}{\omega_i}\mathbf{u}_i^T \int_0^t \mathbf{Q}(t-\tau)\sin\omega_i\tau\, d\tau \tag{b}$$

where $\mathbf{Q}$ has the components given by Eqs. (a). Moreover, from Example 4.6, the natural frequencies are

$$\omega_1 = 0.6524\sqrt{\frac{k}{m}}\ \text{rad/s}, \quad \omega_2 = 1.6556\sqrt{\frac{k}{m}}\ \text{rad/s}, \quad \omega_3 = 2.0000\sqrt{\frac{k}{m}}\ \text{rad/s} \tag{c}$$

and the orthonormal modal vectors are

$$\mathbf{u}_1 = m^{-1/2}\begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix}, \quad \mathbf{u}_2 = m^{-1/2}\begin{bmatrix} -0.5257 \\ 0.0845 \\ 0.6525 \end{bmatrix}, \quad \mathbf{u}_3 = m^{-1/2}\begin{bmatrix} 0.3333 \\ -0.3333 \\ 0.6667 \end{bmatrix} \tag{d}$$

Equation (b) involves the integral

$$\int_0^t u(t-\tau)\sin\omega_i\tau d\tau = \int_0^t \sin\omega_i\tau d\tau = \frac{1}{\omega_i}(1-\cos\omega_i t) \tag{e}$$

Hence, inserting Eqs. (a), (c), (d) and (e) into Eq. (b), we obtain the response

$$\mathbf{q}(t) = \frac{1}{k}\left\{\frac{0.3603}{0.6524^2}\begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix}\left(1-\cos 0.6524\sqrt{\frac{k}{m}}t\right)\right.$$
$$+\frac{0.6525}{1.6556^2}\begin{bmatrix} -0.5257 \\ 0.0845 \\ 0.6525 \end{bmatrix}\left(1-\cos 1.6556\sqrt{\frac{k}{m}}t\right)$$
$$\left.+\frac{0.6667}{2.0000^2}\begin{bmatrix} 0.3333 \\ -0.3333 \\ 0.6667 \end{bmatrix}\left(1-\cos 2.0000\sqrt{\frac{k}{m}}t\right)\right\}$$

$$= \frac{1}{k}\left\{\begin{bmatrix}0.2839\\0.3926\\0.3050\end{bmatrix}\left(1-\cos 0.6524\sqrt{\frac{k}{m}}t\right)+\begin{bmatrix}-0.1251\\0.0201\\0.1553\end{bmatrix}\left(1-\cos 1.6556\sqrt{\frac{k}{m}}t\right)\right.$$

$$\left.+\begin{bmatrix}0.0556\\-0.0556\\0.1111\end{bmatrix}\left(1-\cos 2.0000\sqrt{\frac{k}{m}}t\right)\right\} \qquad \text{(f)}$$

We observe that the largest contribution to the response is from the first mode, with the contribution from the higher modes diminishing as the mode number increases.

Example 4.11

Investigate the convergence characteristics of the transition matrix for the system of Example 4.8.

From Eq. (4.201), the transition matrix can be expressed in the form of the infinite matrix series

$$\Phi = e^{At} = Xe^{\Lambda t}Y^T = X\left(I + t\Lambda + \frac{t^2}{2!}\Lambda^2 + \frac{t^3}{3!}\Lambda^3 + \dots\right)Y^T \qquad \text{(a)}$$

where $\Lambda = \text{diag}\ (\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4)$ is the diagonal matrix of the eigenvalues and X and Y are the matrices of right and left eigenvectors. Because X and Y are constant, the convergence rate of the transition matrix depends on the rate of convergence of $e^{\Lambda t}$, which in turn depends on the magnitude of the eigenvalue of largest modulus. In our particular case, both λ_3 and $\lambda_4 = \overline{\lambda}_3$ share this property. From Example 4.8, we can express λ_3 in the form

$$\lambda_3 = |\lambda_3|\,e^{-i\phi_3}, \qquad \phi_3 = \tan^{-1}\frac{-\text{Im}\,\lambda_3}{\text{Re}\,\lambda_3} \qquad \text{(b)}$$

where

$$|\lambda_3| = \left[(-0.1372)^2 + 4.7653^2\right]^{1/2} = 4.7673 \qquad \text{(c)}$$

is the magnitude. The phase angle is irrelevant. Because Λ is diagonal, the convergence of Φ depends on the convergence of the infinite scalar series

$$\begin{aligned} e^{\lambda_3 t} &= 1 + t\,|\lambda_3|\,e^{-i\phi_3} + t^2\frac{|\lambda_3|^2}{2!}e^{-i2\phi_3} + t^3\frac{|\lambda_3|^3}{3!}e^{-i3\phi_3} + t^4\frac{|\lambda_3|^4}{4!}e^{-i4\phi_3} \\ &\quad + t^5\frac{|\lambda_3|^5}{5!}e^{-i5\phi_3} + \dots \\ &= 1 + 4.7673te^{-i\phi_3} + 11.3635t^2e^{-i2\phi_3} + 18.0576t^3e^{-i3\phi_3} + 21.5214t^4e^{-i4\phi_3} \\ &\quad + 20.5196t^5e^{-i5\phi_3} + \dots \end{aligned} \qquad \text{(d)}$$

Clearly, the exponential terms have unit magnitude, so that they do not affect the convergence rate. The question of convergence is intimately related to the question of precision, as the number of terms required varies with the desired accuracy. For a given precision, the rate of convergence depends on the value of $|\lambda_3|$, as well as on t. For a 10^{-4} precision, and for $|\lambda_3| = 4.7673$ and $t = 0.1$ s, it is necessary to take terms through the fifth power. On the other hand, for $t = 0.5$ s, terms through eleventh power are required, and for $t = 1$ s terms through nineteenth power. Of course, as t increases, computational savings may be achieved by using the semigroup property, as reflected in Eq. (4.275).

It should be pointed out that, although we based the analysis on the behavior of $e^{\Lambda t}$, it is common practice to compute the transition matrix by means of e^{At}, i.e., without solving the eigenvalue problem for A first. Because in this case the eigenvalues are not available, no rational estimate of the number of terms required for convergence of the series for the transition matrix can be made.

4.11 DISCRETE-TIME SYSTEMS

In Sec. 4.10, we discussed the problem of deriving the response of multi-degree-of-freedom systems to arbitrary excitations by means of modal analysis, whereby sets of simultaneous ordinary differential equations of motion are reduced to low-order independent equations lending themselves to relatively easy solution. This makes modal analysis a very desirable method, particularly for conservative systems, which are characterized by real eigensolutions and potent algorithms for computing them. The situation is not nearly so good for nonconservative systems, characterized by complex eigensolutions and significantly less desirable computational algorithms.

In the case in which modal analysis is used to transform a set of simultaneous equations of motion to a set of independent first-order or second-order equations, the response can be obtained by the methods presented in Chapter 3. Of course, the use of modal analysis is not an absolute necessity and a solution can be produced by the method based on the transition matrix presented in Sec. 1.9, even for conservative systems.

The solution based on the transition matrix has the appearance of a closed-form solution, but must be evaluated numerically for the most part. The procedure is computationally intensive, as this involves repeated evaluation of the transition matrix, which requires an increasing amount of effort as the time t increases. As suggested in Example 4.11, some advantage can accrue by dividing the time into smaller increments and using Eq. (4.275). Perhaps a better approach, however, is to discretize the system in time. From Sec. 4.10, the state equations for a general dynamical system have the matrix form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{Q}(t) \tag{4.284}$$

where $\mathbf{x}(t) = \left[\mathbf{q}^T(t)\ \dot{\mathbf{q}}(t)\right]^T$ is the state vector and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}(K+H) & -M^{-1}(C^* + G) \end{array}\right], \qquad B = \left[\begin{array}{c} 0 \\ \hline M^{-1} \end{array}\right] \tag{4.285a,b}$$

are coefficient matrices. From Sec. 1.10, the discrete-time equivalent of Eq. (4.284) is given by the sequence

$$\mathbf{x}(k+1) = \Phi\mathbf{x}(k) + \Gamma\mathbf{Q}(k), \qquad k = 0, 1, 2, \ldots \tag{4.286}$$

where

$$\Phi = e^{AT} = I + TA + \frac{T^2}{2!}A^2 + \frac{T^3}{3!}A^3 + \ldots \tag{4.287a}$$

and

$$\Gamma = A^{-1}\left(e^{AT} - I\right)B = T\left(I + \frac{T}{2!}A + \frac{T^2}{3!}A^2 + \frac{T^3}{4}A^3 + \dots\right)B \quad (4.287b)$$

in which T is the sampling period. Moreover, Φ is recognized as the discrete-time transition matrix. By taking T sufficiently small, the transition matrix can be computed with a relatively small number of terms. In addition to T, the number of terms depends on the magnitude of the eigenvalue of largest modulus, namely, $|\lambda_n|$.

Example 4.12

Compute the discrete-time response sequence of the system of Example 4.8 to an initial impulse applied in the x-direction. List results through $\mathbf{x}(4)$.

From Example 4.8, the system matrix is

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0.4 & -0.3 & 4 \\ -0.4 & -6 & -4 & -0.2 \end{bmatrix} \quad \text{(a)}$$

Hence, if we use results from Example 4.8 to calculate the largest modulus $|\lambda_3| = |\lambda_4| = 4.7673$ and if we choose $T = 0.05$ s and an accuracy of 10^{-4}, then the transition matrix can be computed from Eq. (4.287a) with five terms, as follows:

$$\Phi = e^{AT} \cong I + TA + \frac{T^2}{2!}A^2 + \frac{T^3}{3!}A^3 + \frac{T^4}{4!}A^4$$
$$= \begin{bmatrix} 0.9987 & -0.0055 & 0.0492 & 0.0049 \\ -0.0049 & 0.9927 & -0.0049 & 0.0492 \\ -0.0512 & 0.0101 & 0.9641 & 0.1961 \\ -0.0148 & 0.2977 & -0.1961 & 0.9627 \end{bmatrix} \quad \text{(b)}$$

and, from Eq. (4.287b), we obtain

$$\Gamma = A^{-1}\left(e^{AT} - I\right)B \cong T\left(I + \frac{T}{2!}A + \frac{T^2}{3!}A^2\right)B = \begin{bmatrix} -0.0013 & 0.0001 \\ -0.0001 & 0.0012 \\ 0.0493 & 0.0050 \\ -0.0050 & 0.0493 \end{bmatrix} \quad \text{(c)}$$

Moreover, the discrete-time excitation sequence is

$$\mathbf{Q}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{Q}(k) = \mathbf{0}, \qquad k = 1, 2, \dots \quad \text{(d)}$$

Inserting Eqs. (b)–(d) into Eq. (4.286) and letting $\mathbf{x}(0) = \mathbf{0}$, we obtain the discrete-time response sequence

$$\mathbf{x}(1) = \Phi\mathbf{x}(0) + \Gamma\mathbf{Q}(0) = \begin{bmatrix} -0.0013 & 0.0001 \\ -0.0001 & 0.0012 \\ 0.0493 & 0.0050 \\ -0.0050 & 0.0493 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.0013 \\ -0.0001 \\ 0.0493 \\ -0.0050 \end{bmatrix}$$

$$\mathbf{x}(2) = \Phi\mathbf{x}(1) + \Gamma\mathbf{Q}(1)$$

$$= \begin{bmatrix} 0.9987 & -0.0055 & 0.0492 & 0.0049 \\ -0.0049 & 0.9927 & -0.0049 & 0.0492 \\ -0.0512 & 0.0101 & 0.9641 & 0.1961 \\ -0.0148 & 0.2977 & -0.1961 & 0.9627 \end{bmatrix} \begin{bmatrix} -0.0013 \\ -0.0001 \\ 0.0493 \\ -0.0050 \end{bmatrix} = \begin{bmatrix} 0.0012 \\ -0.0005 \\ 0.0466 \\ -0.0145 \end{bmatrix}$$

$$\mathbf{x}(3) = \Phi\mathbf{x}(2) + \Gamma\mathbf{Q}(2)$$

$$= \begin{bmatrix} 0.9987 & -0.0055 & 0.0492 & 0.0049 \\ -0.0049 & 0.9927 & -0.0049 & 0.0492 \\ -0.0512 & 0.0101 & 0.9641 & 0.1961 \\ -0.0148 & 0.2977 & -0.1961 & 0.9627 \end{bmatrix} \begin{bmatrix} 0.0012 \\ -0.0005 \\ 0.0466 \\ -0.0145 \end{bmatrix} = \begin{bmatrix} 0.0025 \\ -0.0014 \\ 0.0420 \\ -0.0231 \end{bmatrix}$$

$$\mathbf{x}(4) = \Phi\mathbf{x}(3) + \Gamma\mathbf{Q}(3)$$

$$= \begin{bmatrix} 0.9987 & -0.0055 & 0.0492 & 0.0049 \\ -0.0049 & 0.9927 & -0.0049 & 0.0492 \\ -0.0512 & 0.0101 & 0.9641 & 0.1961 \\ -0.0148 & 0.2977 & -0.1961 & 0.9627 \end{bmatrix} \begin{bmatrix} 0.0025 \\ -0.0014 \\ 0.0420 \\ -0.0231 \end{bmatrix} = \begin{bmatrix} 0.0044 \\ -0.0025 \\ 0.0359 \\ -0.0304 \end{bmatrix} \tag{e}$$

4.12 NUMERICAL SOLUTION OF NONLINEAR INITIAL-VALUE PROBLEMS

The various algorithms for determining the system response discussed in Secs. 4.6–4.11 were confined to linear systems and systems linearized about equilibrium positions. Linearization implies the assumption that the motions about equilibrium are small. In certain cases, this assumption is not valid, so that the determination of the response must be based on the original nonlinear differential equations. Except for some special cases, nonlinear differential equations do not lend themselves to closed-form solutions, so that one must be content with numerical solutions. As in Sec. 4.11, numerical solutions can be carried out on digital computers in discrete time, but solutions for nonlinear equations tend to be more involved than for linear equations. In this section, we concentrate on numerical solutions by the Runge-Kutta methods.

In general, the equations of motion for vibrating mechanical systems represent sets of second-order ordinary differential equations. However, numerical integration of differential equations is carried out most conveniently in terms of first-order equations, which requires that the equations of motion be cast in state form. State equations were introduced in Sec. 4.3 in the form of Eq. (4.34). In this section, we enlarge the scope by considering vectors **f** depending explicitly not only on the state vector **x** but also on the time t. This enables us to compute the response to initial excitations and external forces simultaneously.

For convenience, we begin our discussion by considering an initial-value problem described by the single first-order differential equation

$$\dot{x}(t) = f(x(t), t) \tag{4.288}$$

where f is a nonlinear function of $x(t)$ and t. The solution $x(t)$ is subject to the initial condition $x(0) = x_0$. In attempting a solution of an initial-value problem, it is of interest to know that a unique solution exists. Moreover, it is of interest to know that small changes in the problem statement, as reflected in the function $f(x(t), t)$

brings about small changes in the solution $x(t)$. This places certain restrictions on the function f. To discuss these restrictions, we consider a function $f = f(x(t), t)$ defined at every point of a region R of the plane (x, t). The function f is said to satisfy a *Lipschitz condition* with respect to x in R if there exists a positive constant L such that the inequality

$$|f(x_1, t) - f(x_2, t)| \leq L|x_1 - x_2| \tag{4.289}$$

is satisfied for every pair of points (x_1, t) and (x_2, t) in R. The constant L for which the above inequality holds true is called a *Lipschitz constant* and the function f is said to be a *Lipschitz function*, or *Lipschitzian* in R. It can be shown (Ref. 4) that $f(x, t)$ is a Lipschitz function if the region R is convex and the partial derivative $\partial f/\partial x$ exists and is bounded in R. Note that a region is convex provided that whenever two points belong to R, the entire straight line connecting the two points also belongs to R. We consider regions $R\{0 \leq t \leq t_f, -\infty < x < \infty\}$ of the plane (x, t), which can be verified to be convex. According to a theorem demonstrated in Ref. 4, the differential equation (4.288) has a unique solution if f is Lipschitzian in R. The theorem asserts not only that a solution exists but that the solution is determined uniquely by the initial condition $x(0) = x_0$. In the following discussions, we will be concerned only with Lipschitzian functions f. Note that analogous statements can be made in the case in which Eq. (4.288) represents a vector equation rather than a scalar equation (Ref. 4).

Numerical integration provides only an approximate solution whose accuracy depends on the order of the approximation, among other things. We consider here solutions by the *Runge-Kutta methods*, a family of algorithms characterized by different orders of approximation, where the order is related to the number of terms in Taylor series expansions. Derivation of the high-order Runge-Kutta algorithms is very tedious, and the details are not particularly useful. To develop a feel for the approach, however, in this text we derive the second-order method and simply list the equations for the higher-order algorithms in common use. To this end, we expand the solution of Eq. (4.288) in the Taylor series

$$x(t + T) = x(t) + Tx^{(1)}(t) + \frac{T^2}{2!}x^{(2)}(t) + \frac{T^3}{3!}x^{(3)}(t) + \ldots \tag{4.290}$$

where T is a small time increment and the superscript (i) denotes the ith derivative with respect to time $(i = 1, 2, 3, \ldots)$. For numerical purposes, we must limit the number of terms in the series, which amounts to using a finite series to approximate $x(t + T)$. Assuming that the solution $x(t)$ has $N + 1$ continuous derivatives, we can rewrite Eq. (4.290) in terms of an Nth-degree Taylor polynomial about t as follows:

$$x(t + T) = x(t) + Tx^{(1)}(t) + \frac{T^2}{2!}x^{(2)}(t) + \ldots + \frac{T^N}{N!}x^{(N)}(t) + \frac{T^{N+1}}{(N + 1)!}x^{(N+1)}(\xi) \tag{4.291}$$

for some ξ, $t < \xi < t + T$.

Successive differentiation of $x(t)$ with due consideration of Eq. (4.288) yields

$$x^{(i)}(t) = \frac{d^i x(t)}{dt^i} = \frac{d^{i-1} f(x(t), t)}{dt^{i-1}} = f^{(i-1)}(x(t), t), \qquad i = 1, 2, \ldots, N \tag{4.292}$$

Moreover, for our numerical algorithm, we are interested in the discrete-time version of Eq. (4.291). To this end, we consider the discrete times $t = t_k$, $t + T = t_{k+1} = t_k + T$ $(k = 0, 1, 2, \ldots)$, where T is known as the *step size*, and introduce the notation

$$\begin{aligned} x(t) &= x(t_k), \qquad x(t+T) = x(t_{k+1}), \qquad k = 0, 1, 2, \ldots \\ f^{(i)}(x(t), t) &= f^{(i)}(x(t_k), t_k), \qquad i = 0, 1, \ldots, N-1; \; k = 0, 1, 2, \ldots \end{aligned} \tag{4.293}$$

Introducing Eqs. (4.292) and (4.293) into Eq. (4.291), we have

$$x(t_{k+1}) = x(t_k) + \sum_{j=0}^{N} \frac{T^j}{j!} f^{(j-1)}(x(t_k), t_k) + \frac{T^{N+1}}{(N+1)!} f^{(N)}(x(\xi_k), \xi_k) \tag{4.294}$$

where $t_k < \xi_k < t_{k+1}$.

An approximation of the solution of Eq. (4.291) is obtained by ignoring the remainder term in Eq. (4.294), i.e., the term involving ξ_k. Hence, denoting the approximate values of $x(t_k)$, $x(t_{k+1})$ and $f^{(i)}(x(t_k), t_k)$ by w_k, w_{k+1} and $f_k^{(i)}$, respectively, we can write the discrete-time version of the Nth-order Taylor series in the form

$$w_{k+1} = w_k + Tf_k + \frac{T^2}{2!} f_k^{(1)} + \ldots + \frac{T^N}{N!} f_k^{(N-1)}, \qquad k = 0, 1, 2, \ldots; \; w_0 = x_0 \tag{4.295}$$

The method for computing the numerical solution of Eq. (4.288) by means of Eq. (4.295) is called the *Taylor method of order N*. It forms the basis for the Runge-Kutta methods.

The error in the approximation (4.295) is known as the *local truncation error* and is defined by

$$e_{k+1} = x(t_{k+1}) - x(t_k) - \sum_{j=0}^{N} \frac{T^j}{j!} f^{(j-1)}(x(t_k), t_k) = \frac{T^{N+1}}{(N+1)!} f^{(N)}(x(\xi_k), \xi_k) \tag{4.296}$$

It follows that the Taylor method of order N has the desirable feature that the local truncation error is of order $O(T^{N+1})$. The *global truncation error* is defined as

$$e_M = x(t_M) - x_M \tag{4.297}$$

where M is the number of steps. In the case of the Taylor method of order N, the global truncation error can be shown (Ref. 7) to be of order $O(T^N)$. Hence, by making the order N sufficiently large, or the step size T for a given N sufficiently small, the global truncation error can be made as small as desired. Note that, in addition to truncation errors, there may be roundoff errors.

The lowest-order approximation is obtained by retaining terms in Eq. (4.295) through the first order only and has the form

$$w_{k+1} = w_k + Tf_k \tag{4.298}$$

The method of computing the first-order approximation by means of Eq. (4.298) is known as *Euler's method.* It represents linearization of the nonlinear differential equation and tends to be very inaccurate, as the global truncation error is of the same order as the step size T. For this reason, Euler's method is not recommended.

The Taylor methods have a serious drawback in that they require derivatives of $f(x,t)$. Indeed, quite often this makes the process very tedious, so that the Taylor methods, as defined by Eq. (4.295), have limited appeal. The Runge-Kutta methods, which are based on the Taylor methods, retain the desirable characteristic of high-order truncation errors of the Taylor methods, but avoid the requirement for derivatives of f. To illustrate the ideas, we consider the second-order Runge-Kutta method, denoted by RK2. To this end, we write the second-order approximation in the form

$$x(t+T) = x(t) + c_1 g_1 + c_2 g_2 \tag{4.299}$$

where c_1 and c_2 are constants and

$$g_1 = Tf(x,t), \qquad g_2 = Tf(x + \alpha_1 g_1,\ t + \alpha_2 T) \tag{4.300}$$

in which α_1 and α_2 are constants. But, from Eq. (4.294) with $N = 2$, the second-order Taylor method is defined by

$$\begin{aligned} x(t+T) &= x(t) + Tf(x,t) + \frac{T^2}{2!}\frac{df(x,t)}{dt} \\ &= x(t) + Tf(x,t) + \frac{T^2}{2}\left(\frac{\partial f(x,t)}{\partial x}f(x,t) + \frac{\partial f(x,t)}{\partial t}\right) \end{aligned} \tag{4.301}$$

Comparing Eq. (4.299), in conjunction with Eqs. (4.300), with Eq. (4.301), we conclude that if the constants c_1, c_2, α_1 and α_2 satisfy the relations

$$c_1 + c_2 = 1, \qquad \alpha_1 c_2 = \frac{1}{2}, \qquad \alpha_2 c_2 = \frac{1}{2} \tag{4.302}$$

then the RK2 method will have truncation errors of the same order of magnitude as the second-order Taylor method.

We observe that Eqs. (4.302) represent three algebraic equations in four unknowns, so that they are underdetermined. This implies that one of the unknowns can be chosen arbitrarily, which determines the three other unknowns uniquely. One satisfactory choice, and one that gives Eq. (4.299) symmetric form, is $c_1 = 1/2$, which yields $c_1 = c_2 = 1/2$, $\alpha_1 = \alpha_2 = 1$. In this case, the RK2 method can be written in the form

$$w_{k+1} = w_k + \frac{T}{2}\{f(w_k, t_k) + f[w_k + Tf(w_k, t_k), t_k + T]\}, \qquad k = 0, 1, 2, \ldots;\ w_0 = x_0 \tag{4.303}$$

This version of the RK2 method is known as *Heun's method.* Note that Ref. 2 refers to another version of RK2 as Heun's method.

Another choice is $c_1 = 0$, which yields $c_2 = 1$, $\alpha_1 = \alpha_2 = 1/2$. In this case, the RK2 method has the form

$$w_{k+1} = w_k + Tf\left[w_k + \frac{T}{2}f(w_k, t_k),\ t_k + \frac{T}{2}\right], \qquad k = 0, 1, 2, \ldots;\ w_0 = x_0 \tag{4.304}$$

This version of the RK2 method is called the *modified Euler method.*

The most common Runge-Kutta method is of order four. It is denoted by RK4 and is defined by

$$w_{k+1} = w_k + \frac{1}{6}(g_1 + 2g_2 + 2g_3 + g_4) \tag{4.305}$$

where

$$\begin{aligned} g_1 &= Tf(w_k, t_k), \\ g_2 &= Tf\left(w_k + \frac{1}{2}g_1, t_k + \frac{T}{2}\right), \\ g_3 &= Tf\left(w_k + \frac{1}{2}g_2, t_k + \frac{T}{2}\right), \\ g_4 &= Tf(w_k + g_3,\ t_k + T), \end{aligned} \qquad k = 0, 1, 2, \ldots \tag{4.306}$$

It has a local truncation error of order $O(T^5)$ and global truncation error of order $O(T^4)$. For the most part, this represents very good accuracy. The RK4 method is also easy to implement, which explains why the method is used widely.

The question of the step size required to guarantee a given accuracy remains. One way of addressing this question is to solve the problem twice, once using the step size T and the other using the step size $T/2$, and compare the results. If there is no satisfactory agreement, the procedure must be repeated. This process is not very efficient, as it requires a significant amount of computation.

The *Runge-Kutta-Fehlberg method*, denoted by RKF45, resolves the above problem in an efficient manner. Indeed, the approach involves the determination of an optimal step size for a given accuracy. The RKF45 method requires two different approximations at each step. The first step is the RK4 approximation defined by

$$w_{k+1} = w_k + \frac{25}{216}g_1 + \frac{1,408}{2,565}g_3 + \frac{2,197}{4,104}g_4 - \frac{1}{5}g_5 \tag{4.307}$$

and the second is the RK5 approximation having the form

$$\tilde{w}_{k+1} = w_k + \frac{16}{135}g_1 + \frac{6,656}{12,825}g_3 + \frac{28,561}{56,430}g_4 - \frac{9}{50}g_5 + \frac{2}{55}g_6 \tag{4.308}$$

where

$$\begin{aligned}
g_1 &= Tf\,(w_k, t_k) \\
g_2 &= Tf\left(w_k + \frac{1}{4}g_1, t_k + \frac{T}{4}\right) \\
g_3 &= Tf\left(w_k + \frac{3}{32}g_1 + \frac{9}{32}g_2,\ \ t_k + \frac{3T}{8}\right) \\
g_4 &= Tf\left(w_k + \frac{1,932}{2,197}g_1 - \frac{7,200}{2,197}g_2 + \frac{7.296}{2,197}g_3,\ t_k + \frac{12T}{13}\right) \\
g_5 &= Tf\left(w_k + \frac{439}{216}g_1 - 8g_2 + \frac{3,680}{513}g_3 - \frac{845}{4,104}g_4,\ t_k + T\right) \\
g_6 &= Tf\left(w_k - \frac{8}{27}g_1 + 2g_2 - \frac{3,544}{2,565}g_3 + \frac{1,859}{4,104}g_4 - \frac{11}{40}g_5, t_k + \frac{T}{2}\right)
\end{aligned} \tag{4.309}$$

The RKF45 method requires the evaluation of six functions, $g_1, g_2, \ldots, g_6$, per step. Note that, although g_2 is not required explicitly to compute w_{k+1} and $\tilde{w}_{k+1}$, it is required to compute g_3, g_4, g_5 and g_6. By contrast, RK4 and RK5 require the evaluation of four functions and six functions, respectively, for a total of ten.

The RKF45 method contains a so-called "error-control" procedure permitting an optimal determination of the step size for a given accuracy. Denoting the specified error-control tolerance by ε and the corresponding optimal step size by sT, the scalar s can be determined by means of the formula (Ref. 7)

$$s = \left(\frac{\varepsilon T}{2|\tilde{w}_{k+1} - w_{k+1}|}\right)^{1/4} \tag{4.310}$$

Formula (4.310) provides a conservative choice for s so as to avoid extensive computations involved in the repetition of steps. Then, w_{k+1} is computed using the step size sT instead of T. Moreover, the next step, consisting of the computation of w_{k+1} and $\tilde{w}_{k+1}$ with k advanced by 1, is carried out initially with the step size sT. Note that, if w_{k+1} and $\tilde{w}_{k+1}$ agree to more significant digits than required, s can be larger than 1, which implies an increase in the step size.

The Runge-Kutta methods are *one-step methods*, as information from step k only is used to compute w_{k+1}. Such methods are said to be *self-starting*. The Runge-Kutta methods have the advantages that they are easy to code and are numerically stable for a large class of problems.

Because the information at previous points, namely, $w_{k-1}, w_{k-1}, \ldots$ and f_{k-1}, $f_{k-2}, \ldots$, is readily available and because the global truncation error tends to increase with each step, the question arises whether accuracy can be improved by using this information to compute w_{k+1}. Methods using information from more than one previous point are referred to as *multistep methods*. A commonly used multistep method is the *fourth-order Adams-Bashforth-Moulton predictor-corrector method*

(Refs. 2, 5 and 7). It uses the *Adams-Bashforth predictor* defined by

$$w_{k+1} = w_k + \frac{T}{24}\left[55f(w_k, t_k) - 59f(w_{k-1}, t_{k-1}) + 37f(w_{k-2}, t_{k-2}) - 9f(w_{k-3}, t_{k-3})\right], \quad k = 3, 4, \ldots \tag{4.311}$$

and the *Adams-Moulton corrector* given by

$$w_{k+1} = w_k + \frac{T}{24}\left[9f(w_{k+1}, t_{k+1}) + 19f(w_k, t_k) - 5f(w_{k-1}, t_{k-1}) + f(w_{k-2}, t_{k-2})\right], \quad k = 2, 3, \ldots \tag{4.312}$$

Multistep methods are not self-starting. Hence, before the predictor-corrector method can be applied, it is necessary to generate the starting values w_1, w_2 and w_3 for a given initial value $w_0 = x_0$ by a single-step method, such as the RK4 method. Then, the predictor is used to compute $w_4^{(0)}$, an initial approximation to w_4, as follows:

$$w_4^{(0)} = w_3 + \frac{T}{24}\left[55f(w_3, t_3) - 59f(w_2, t_2) + 37f(w_1, t_1) - 9f(x_0, 0)\right] \tag{4.313}$$

The approximation is improved by using the corrector and writing

$$w_4^{(1)} = w_3 + \frac{T}{24}\left[9f(w_4^{(0)}, t_4) + 19f(w_3, t_3) - 5f(w_2, t_2) + f(w_1, t_1)\right] \tag{4.314}$$

The corrector can be used again in conjunction with $w_4^{(1)}$ to obtain

$$w_4^{(2)} = w_3 + \frac{T}{24}\left[9f(w_4^{(1)}, t_4) + 19f(w_3, t_3) - 5f(w_2, t_2) + f(w_1, t_1)\right] \tag{4.315}$$

The process can be continued until convergence to w_4 is achieved. However, the process converges to an approximation given by the corrector, rather than to the solution $w(t_4)$. In practice, it is more efficient to reduce the step size, if improved accuracy is needed, and accept $w_4^{(1)}$ as the approximation to $w(t_4)$. Then, the process continues by using the predictor in conjunction with $w_4 = w_4^{(1)}$ to compute $w_5^{(0)}$ and the corrector in conjunction with $w_5^{(0)}$ to compute $w_5^{(1)}$, where the latter is accepted as the approximation to $x(t_5)$, etc. The local truncation error for both the predictor and the corrector is of the order $O(T^5)$.

Two other popular multistep methods are the *Milne-Simpson method* and the *Hamming method* (Ref. 7).

Difficulties with numerical methods can be expected when the exact solution of the differential equation contains terms of the exponential form $e^{\lambda t}$, where λ is a complex number with negative real part. Whereas this term tends to zero as t increases, the approximation does not necessarily exhibit this characteristic. Such a differential equation is said to be *stiff* and can arise in damped systems. Unless the step size is sufficiently small, the results can be meaningless. Even when the step size is reduced, the improvement tends to be temporary, and eventually truncation and roundoff errors lead to instability. This tends to occur when the solution contains a steady-state part, in addition to the transient part. For the steady-state part, a larger step size should be used, but the transient part, which may have decayed already, dictates a smaller step size. But, as indicated in Sec. 3.2, it is not advisable to combine steady-state solutions with transient solutions.

In the vibration of single- and multi-degree-of-freedom systems, the interest lies in sets of 2 and $2n$ first-order differential equations, respectively, rather than in a single first-order equation. The generalization from a single equation to a set of equations is relatively simple. Indeed, by analogy with Eq. (4.34), we extend Eq. (4.288) to state form by writing

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) \tag{4.316}$$

where $\mathbf{x}(t)$ is the $2n$-dimensional state vector and $\mathbf{f}$ is a $2n$-dimensional nonlinear vector function of $\mathbf{x}(t)$ and t. Then, by analogy with Eqs. (4.305) and (4.306), the fourth-order Runge-Kutta method for state equations can be defined by

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{1}{6}(\mathbf{g}_1 + 2\mathbf{g}_2 + 2\mathbf{g}_3 + \mathbf{g}_4), \qquad k = 0, 1, 2, \ldots \tag{4.317}$$

where

$$\left.\begin{aligned} \mathbf{g}_1 &= T\mathbf{f}(\mathbf{w}_k, t_k), \\ \mathbf{g}_2 &= T\mathbf{f}\left(\mathbf{w}_k + \frac{1}{2}\mathbf{g}_1,\ t_k + \frac{T}{2}\right), \\ \mathbf{g}_3 &= T\mathbf{f}\left(\mathbf{w}_k + \frac{1}{2}\mathbf{g}_2,\ t_k + \frac{T}{2}\right), \\ \mathbf{g}_4 &= T\mathbf{f}(\mathbf{w}_k + \mathbf{g}_3,\ t_k + T), \end{aligned}\right. \qquad k = 0, 1, 2, \ldots \tag{4.318}$$

are $2n$-dimensional vectors. The fourth-order Runge-Kutta method requires four evaluations of the vector $\mathbf{f}$ for each integration step, which implies a significant amount of computations.

The Runge-Kutta-Fehlberg method (RKF45) and the Adams-Bashforth-Moulton predictor-corrector method can be extended to state form in the same simple manner.

The Runge-Kutta methods are quite accurate and easy to code. They are used widely for numerical integration of nonlinear differential equations associated with vibration problems. Predictor-corrector methods tend to be more accurate than Runge-Kutta methods, but they are more difficult to code. Moreover, they must rely on one-step methods, such as the Runge-Kutta methods, to generate the necessary starting values.

Example 4.13

The oscillation of a simple pendulum is described by the differential equation

$$\ddot{\theta} + 4\sin\theta = 0 \tag{a}$$

The pendulum is subject to the initial conditions

$$\theta(0) = 0, \qquad \dot{\theta}(0) = 3 \text{ rad/s} \tag{b}$$

Compute the response by the fourth-order Runge-Kutta method using the step size $T = 0.01$ s and plot $\theta(t)$ versus t for $0 < t < 5$ s.

Introducing the notation

$$\theta(t) = x_1(t), \qquad \dot{\theta}(t) = x_2(t) \tag{c}$$

Eq. (a) can be replaced by the state equations

$$\dot{x}_1(t) = x_2(t), \qquad \dot{x}_2(t) = -4\sin\, x_1(t) \tag{d}$$

Hence, with reference to Eq. (4.316), the state vector and the vector **f** have the form

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}, \qquad \mathbf{f}(\mathbf{x}(t), t) = \mathbf{f}(\mathbf{x}(t)) = \begin{bmatrix} x_2(t) \\ -4\sin x_1(t) \end{bmatrix} \tag{e}$$

so that **f** does not depend on t explicitly.

The computational algorithm is defined by Eqs. (4.317) and (4.318). From the latter, the vectors $\mathbf{g}_1$, $\mathbf{g}_2$, $\mathbf{g}_3$ and $\mathbf{g}_4$ have the explicit form

$$\begin{aligned}
\mathbf{g}_1 &= \begin{bmatrix} g_{11k} \\ g_{12k} \end{bmatrix} = \begin{bmatrix} Tf_1(\mathbf{w}_k) \\ Tf_2(\mathbf{w}_k) \end{bmatrix} = \begin{bmatrix} Tw_{2k} \\ -4T\sin w_{1k} \end{bmatrix}, \\
\mathbf{g}_2 &= \begin{bmatrix} g_{21k} \\ g_{22k} \end{bmatrix} = \begin{bmatrix} Tf_1(\mathbf{w}_k + 0.5\mathbf{g}_1) \\ Tf_2(\mathbf{w}_k + 0.5\mathbf{g}_1) \end{bmatrix} = \begin{bmatrix} T(w_{2k} + 0.5g_{12k}) \\ -4T\sin(w_{1k} + 0.5g_{11k}) \end{bmatrix}, \\
\mathbf{g}_3 &= \begin{bmatrix} g_{31k} \\ g_{32k} \end{bmatrix} = \begin{bmatrix} Tf_1(\mathbf{w}_k + 0.5\mathbf{g}_2) \\ Tf_2(\mathbf{w}_k + 0.5\mathbf{g}_2) \end{bmatrix} = \begin{bmatrix} T(w_{2k} + 0.5g_{22k}) \\ -4T\sin(w_{1k} + 0.5g_{21k}) \end{bmatrix}, \\
\mathbf{g}_4 &= \begin{bmatrix} g_{41k} \\ g_{42k} \end{bmatrix} = \begin{bmatrix} Tf_1(\mathbf{w}_k + \mathbf{g}_3) \\ Tf_2(\mathbf{w}_k + \mathbf{g}_3) \end{bmatrix} = \begin{bmatrix} T(w_{2k} + g_{32k}) \\ -4T\sin(w_{1k} + g_{31k}) \end{bmatrix}, \\
& \qquad k = 0, 1, 2, \ldots
\end{aligned} \tag{f}$$

in which, from Eqs. (b) and (c), the initial state vector is given by

$$\mathbf{w}_0 = \mathbf{x}_0 = [x_{10}\ x_{20}]^T = [0\ 2]^T \tag{g}$$

The plot $\theta(t)$ versus t is shown in Fig. 4.10. Whereas the plot has the appearance of a sine curve, the response is periodic, rather than harmonic as in the linear case. As the initial velocity $\dot{\theta}(0)$ increases, the response differs more and more from a sine curve. In fact, for $\dot{\theta}(0) > 4$ rad/s, there are no equilibrium positions and the nature of the motion changes from oscillation about the equilibrium position $\theta = 0$ to rotation with variable angular velocity.

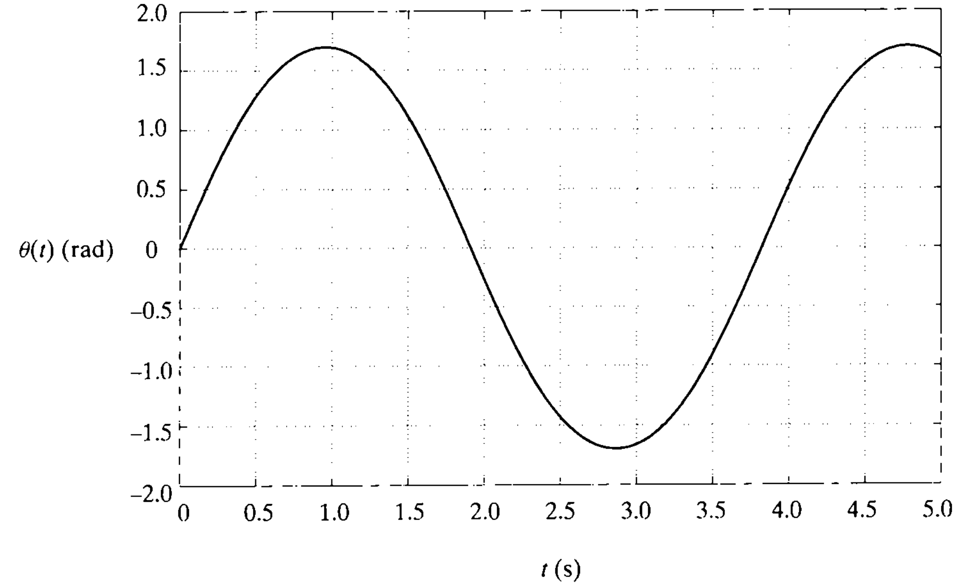


Figure 4.10 Response of a simple pendulum undergoing nonlinear oscillation

4.13 SYNOPSIS

Multi-degree-of-freedom systems are described by sets of simultaneous ordinary differential equations, derived conveniently by the methods of analytical dynamics discussed in Chapter 2. A great deal of insight into the system behavior can be gained from a geometric representation of the solution in the state space. Such a representation is essential to a good understanding of the concepts of equilibrium points and small motions in the neighborhood of equilibrium points. A very important issue is whether displacements from equilibrium are small or not, as this factor decides at times whether a system can be treated as linear or nonlinear. This question is related to system stability, as the motion of unstable systems increases without bounds, thus becoming nonlinear. The converse is not true, however, as there are many stable nonlinear systems, such as a simple pendulum undergoing large displacements (see Example 4.13). In many cases, the issue of linearity versus nonlinearity can be settled on the basis of physical grounds.

A very large number of vibrating systems falls in the class of *linear time-invariant systems*, more commonly known as *linear systems with constant coefficients*. The most important and desirable characteristic of linear time-invariant systems is that they are subject to the *principle of superposition*. For such systems, we can avail ourselves to a wealth of solution techniques. Among these, *modal analysis*, whereby a set of simultaneous ordinary differential equations of motion is transformed into a set of independent equations, plays a central role in vibrations, as it permits a relatively easy determination of the response. This implies the use of a linear transformation involving the *modal matrix*, which can be obtained by solving the *algebraic eigenvalue problem*. Due to its pivotal role in vibrations, the following two chapters are devoted to the algebraic eigenvalue problem, Chapter 5 to qualitative aspects and Chapter 6 to computational algorithms. For undamped natural systems, the eigenvalue problem is symmetric and the eigensolutions are real, with the modal matrix being orthogonal. The response of such systems is defined in the configuration space, and the procedure for obtaining the response is referred to at times as the *classical modal analysis*. Undamped gyroscopic systems share many of the properties of undamped natural systems, except that the eigenvalue problem and the response are defined in the state space, rather than in the configuration space. Nonconservative systems, of which damped systems are a special case, differ from conservative systems in that the eigenvalue problem is not symmetric and the eigensolutions are either complex, or real, or some complex and some real. The eigenvalue problem is defined in the state space and is significantly more involved than in the symmetric case. Moreover, the eigenvalue problem must be solved twice, once for the system matrix and once for the transposed system matrix. The eigenvalues are the same, but the two sets of eigenvectors are different; they possess the biorthogonality property. A modal analysis can be formulated for the response of nonconservative systems, but it is appreciably more complicated than for conservative systems. The modal analysis is in the state space and must use both sets of eigenvectors. In view of the fact that the modal analysis for nonconservative systems is quite laborious, an approach avoiding the need for solving nonsymmetric eigenvalue problems has certain appeal. Such an approach is based on the transition matrix, which requires the evaluation of a matrix series in the system matrix. The approach based on the transition matrix tends to be computationally intensive, which dictates the use of a digital computer. This, in turn requires transformation of the equations of motion from continuous time to discrete time, solution of the discretized-in-time problem on a digital computer and transformation of the discrete-time solution back to continuous time. More often than not, the last step is carried out automatically by plotting the discrete-time solution sequence as a function of time; for fine resolution, such plots tend to appear as continuous in time. The approach just described is referred to as *discrete-time systems*, and provides the formalism for computer coding of vibration problems. Finally, there is the problem of *nonlinear equations* of motion not lending themselves to linearization. In such cases, it is necessary to evaluate the response in discrete time through numerical integration. Widely used numerical integration techniques are the Runge-Kutta methods, such as the *Runge-Kutta-Fehlberg* method, and predictor-corrector methods, such as the *Adams-Bashforth-Moulton method.*

It should be pointed out that the techniques presented in this chapter apply not only to lumped systems but also to distributed-parameter systems discretized in the spatial variables. Indeed, except for certain cases, almost all involving parameters distributed uniformly, vibration problems for distributed systems do not admit exact solutions, so that the interest lies in approximate solutions. In one form or other, such approximate solutions require spatial discretization, as we shall see in Chapters 8 and 9. The equations of motion for the resulting discretized systems have the same form as those for the lumped-parameter systems considered in this chapter and the next two chapters. Hence, the techniques presented in this chapter apply equally well to discrete models approximating distributed-parameter systems.

PROBLEMS

4.1 Derive Lagrange's equations of motion for the system of Problem 2.2 under the assumptions that the displacements y_1 and y_2 are relatively large and that the string tension T remains constant throughout the motion.

4.2 The system of Fig. 4.11 consists of a mass m suspended on a string fixed at both ends and rotating at a constant angular velocity Ω about a horizontal axis passing through the two ends. For convenience, introduce a reference frame y, z rotating with the constant angular velocity Ω with respect to the inertial axes X, Y, Z and express the displacement of m in terms of components measured relative to y, z. Derive Lagrange's equations of motion under the assumptions that the displacements y, z are relatively large and that the string tension T is constant at all times. Ignore gravity.

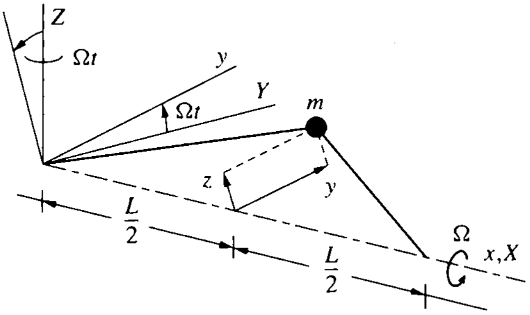


Figure 4.11 Mass on a rotating string

4.3 The system shown in Fig. 4.12 is similar to that in Problem 4.2, except that there are two masses m_1 and m_2 suspended on the rotating string, instead of one. The displacement components of m_1 and m_2 relative to the rotating frame y, z are y_1, z_1 and y_2, z_2, respectively. Derive Lagrange's equations of motion under the same assumptions as in Problem 4.2. Ignore gravity.

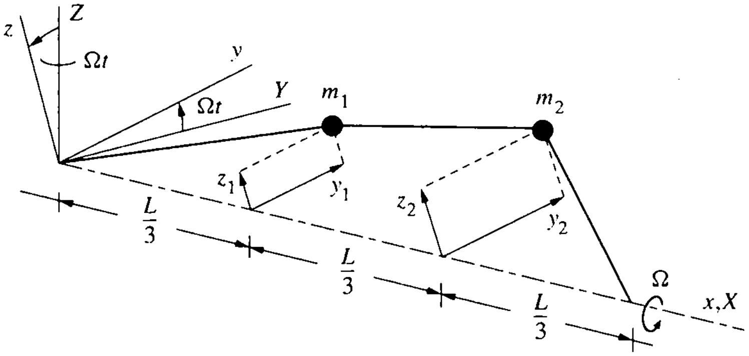


Figure 4.12 Two masses on a rotating string

4.4 The system of Fig. 4.13 consists of two rigid links of total mass m_i and length L_i $(i = 1, 2,)$ hinged to a shaft rotating with the constant angular velocity Ω about a vertical axis. The links are hinged so as to permit motion of the links in the rotating vertical plane and their angular displacements θ and ϕ are restrained by torsional springs of stiffness k_1 and k_2, respectively. Derive Lagrange's equations of motion for arbitrarily large angles θ and ϕ.

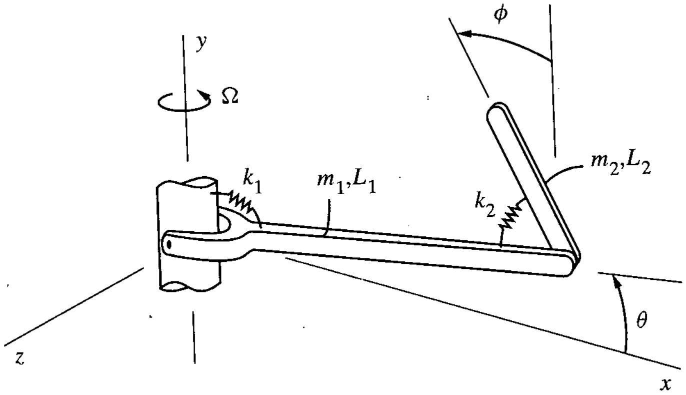


Figure 4.13 Two links hinged to a rotating shaft

4.5 The system shown in Fig. 4.14 is similar to that in Problem 4.4, except that the second link is hinged to the first so as to permit angular displacements ϕ in a plane normal to the first link. Derive Lagrange's equations of motion for arbitrarily large angles θ and ϕ.

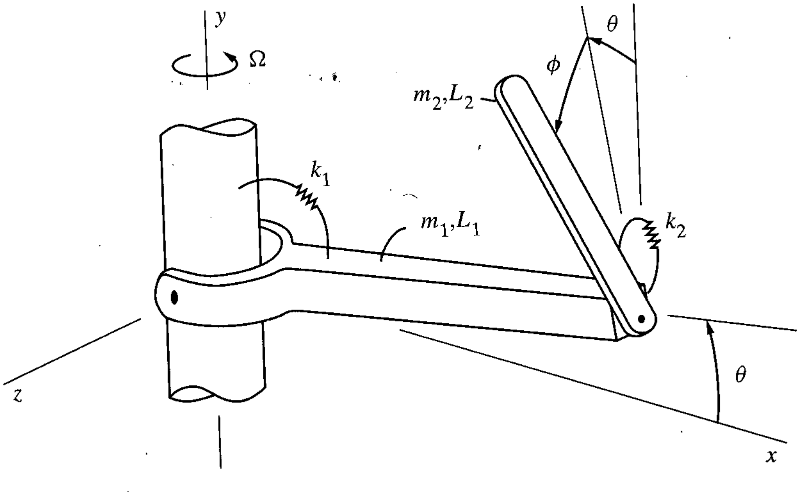


Figure 4.14 Two links hinged to a rotating shaft

4.6 A massless elastic beam with a rigid disk of mass M attached at midspan is rotating with the constant angular velocity Ω, as shown in Fig. 4.15. Let Y, Z be a set of inertial axes and y, z a set of body axes rotating with the angular velocity Ω with respect to Y, Z and derive Lagrange's equations of motion of M in terms of displacement components y and z along the body axes. Assume that the beam has equivalent spring constants k_y and k_z for bending in the y and z directions, respectively, and that the system is subject to internal damping forces proportional to $\dot{y}$ and $\dot{z}$, where the constant of proportionality is c, and external damping forces proportional to $\dot{Y}$ and $\dot{Z}$, where the proportionality constant is h.

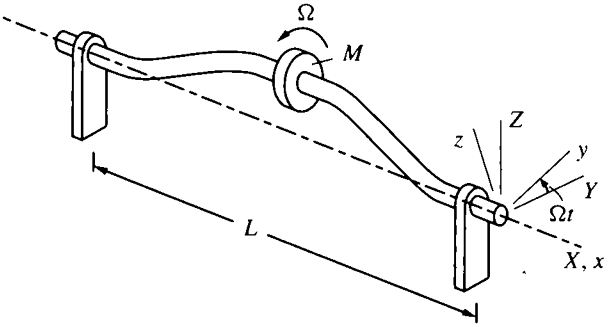


Figure 4.15 Rotating elastic beam with a rigid disk

4.7 Determine the equilibrium points for the system of Problem 2.19 by means of the approach of Sec. 4.2. Note that $T = T_2 + T_0$.

4.8 Derive the equilibrium equations for the system of Problem 4.1 by means of the approach of Sec. 4.2. Then, let the displacements y_1 and y_2 be small and determine the equilibrium position.

4.9 Determine the equilibrium positions for the system of Problem 4.2 by means of the approach of Sec. 4.2.

4.10 Derive the equilibrium equations for the system of Problem 4.3 by means of the approach of Sec. 4.2.

4.11 Derive the equilibrium equations for the system of Problem 4.4.

4.12 Derive the equilibrium equations for the system of Problem 4.5.

4.13 Test the stability of each of the equilibrium points for the system of Problem 4.7 by means of the Liapunov direct method.

4.14 Test the stability of the equilibrium position of the system of Problem 4.8 by means of the Liapunov direct method.

4.15 Assume that the angles defining the equilibrium equations in Problem 4.11 are small and test the stability of the equilibrium position by means of the Liapunov direct method.

4.16 Assume that the angles defining the equilibrium equations in Problem 4.12 are small and test the stability of the equilibrium position by means of the Liapunov direct method.

4.17 Test the stability of the equilibrium position of the system of Problem 4.6 by means of the Liapunov direct method.

4.18 Derive the linearized equation of motion of the system of Problem 4.7 about each of the equilibrium points.

4.19 Derive the linearized equations of motion about the equilibrium position for the system of Problem 4.15.

4.20 Derive the linearized equation of motion about the equilibrium position for the system of Problem 4.16.

4.21 Derive the linearized equations of motion about the trivial equilibrium position for the system of Problem 4.17.

4.22 Four discrete masses m_i $(i = 1, 2, 3, 4)$ suspended on a string are vibrating in a vertical plane, as shown in Fig. 4.16. Assume that the tension T in the string is constant and that the displacements y_i $(i = 1, 2, 3, 4)$, measured from the equilibrium position, are small, derive the eigenvalue problem and set it in symmetric standard form.

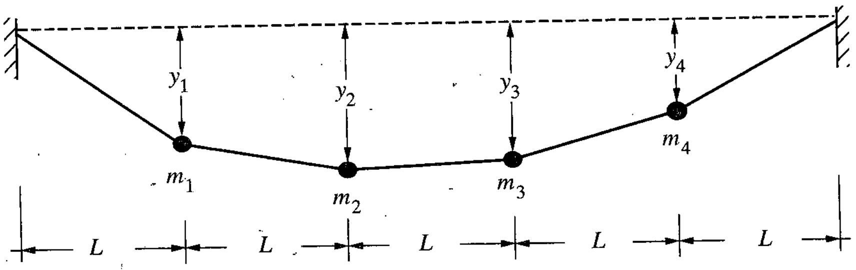


Figure 4.16 Four masses on a string in transverse vibration

4.23 Derive the eigenvalue problem for the system of Fig. 4.17 and set it in symmetric standard form.

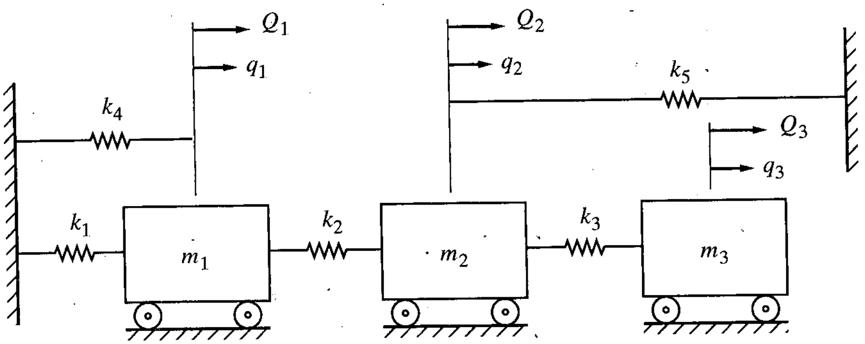


Figure 4.17 Three masses in axial vibration

4.24 The system of Fig. 4.18 consists of four rigid disks of mass moments of inertia I_i $(i = 1, 2, 3, 4)$ mounted on two shafts in torsion. The first shaft consists of three segments of torsional stiffnesses GJ_i $(i = 1, 2, 3)$, where G is the shear modulus and J_i are area polar moments of inertia, and the second shaft consists of one segment of torsional stiffness GJ_4. Disks 3 and 4 roll on each other without slip. Derive the eigenvalue problem and set it in symmetric standard form.

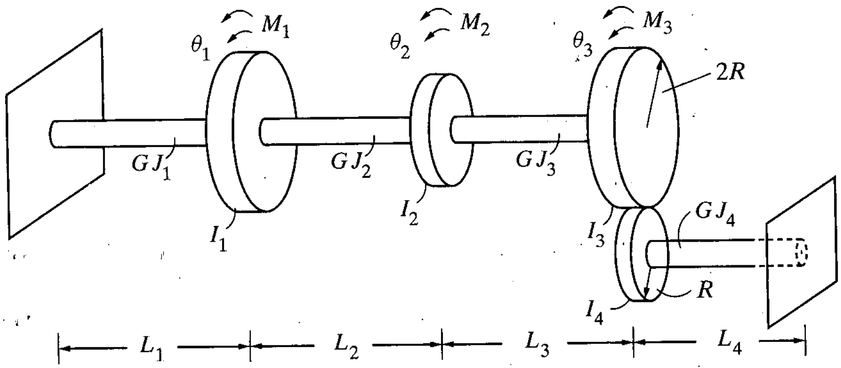


Figure 4.18 Four constrained disks in torsional vibration

4.25 The n-story building depicted in Fig. 4.19 consists of rigid floors of mass m_i each supported by massless columns of bending stiffness EI_i, where E is the modulus of elasticity and I_i are the area moments of inertia $(i = 1, 2, \ldots, n)$. Use the concept of equivalent springs for the columns, under the assumption that the right angles between floors and columns remain so during motion, and derive the eigenvalue problem.

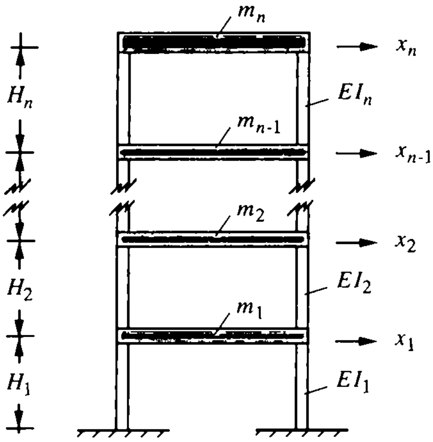


Figure 4.19 An n-story building in horizontal vibration

4.26 Derive the linearized equations of motion for the system of Problem 4.9 about the trivial equilibrium position and set up the eigenvalue problem in the most convenient standard form.

4.27 Derive the linearized equations of motion for the system of Problem 4.10 about the trivial equilibrium position and set up the eigenvalue problem in the most convenient standard form.

4.28 Derive the eigenvalue problem for the system of Fig. 4.20 and set it in standard form.

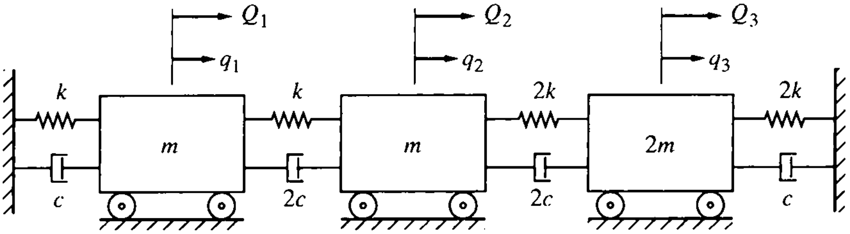


Figure 4.20 Damped three-degree-of-freedom system

4.29 Derive the eigenvalue problem for the system of Problem 4.6 and set it in standard form.

4.30 Determine the response of the system of Problem 4.24 to the torque $M_3(t) = M_3 \sin \omega t$, where M_3 is a constant amplitude; the other two torques are zero. Let $L_1 = L_2 = L_3 = L_4 = L$, $I_1 = I_2 = I_3 = I$, $I_4 = 2I$, $J_1 = J_2 = J_4 = J$, $J_3 = 2J$.

4.31 The n-story building of Problem 4.25 is subjected to the horizontal ground motion $u_g(t) = A \cos \omega t$, where A is a constant amplitude having units of displacement. Determine the general response of the building.

4.32 Determine the response of the system of Fig. 4.1 and Example 4.7 to the forces $F_x(t) = F_x \sin \omega t$, $F_y(t) = F_y \sin \omega t$ applied to mass m.

4.33 Determine the response of the system of Fig. 4.20 to the force $Q_2(t) = Q_2 \cos(\omega t - \psi)$, the other two forces are zero. Let $m = 1$, $c = 0.1$, $k = 1$.

4.34 Consider the string of Problem 4.22, let the masses have the values $m_1 = m, m_2 = m_3 = 2m, m_4 = m$ and determine the response to

(a) the initial displacements $\mathbf{q} = [0.5\ \ 1\ \ 1\ \ 0.5]^T$

(b) the initial velocities $\dot{\mathbf{q}}_0 = [1\ \ 1\ \ -1\ \ -1]^T$

Discuss the mode participation in the response in each of the two cases.

4.35 Determine the response of the torsional system of Problem 4.30 to

(a) the initial displacements $\boldsymbol{\theta}_0 = [0.5\ \ 0.8\ \ 1]^T$

(b) the initial velocities $\dot{\boldsymbol{\theta}}_0 = [0\ \ 0\ \ 1]^T$

4.36 Determine the response of the system of Example 4.7 to

(a) the initial displacements $\mathbf{q}_0 = [1\ \ 0]^T$

(b) the initial velocities $\dot{\mathbf{q}}_0 = [1\ \ 1]^T$

4.37 Determine the response of the system of Problem 4.33 to

(a) the initial displacements $\mathbf{q}_0 = [0.7\ \ 1\ \ 0.7]^T$

(b) the initial velocities $\dot{\mathbf{q}}_0 = [1\ \ 0\ \ -1]^T$

Discuss the mode participation in the response in each of the two cases.

4.38 Determine the response of the system of Example 4.6 to the impulsive excitation $\mathbf{Q}(t) = \hat{Q}_0 \delta(t)[0\ \ 1\ \ 0]^T$

4.39 Determine the response of the system of Problem 4.34 to an external excitation in the form of the pulse $\mathbf{Q}(t) = Q_0(u(t) - u(t-T))[0.5\ \ 1\ \ 1\ \ 0.5]^T$. Discuss the mode participation in the response.

4.40 The system of Problem 4.35 is subjected to the excitation $\mathbf{M}(t) = \hat{\mathbf{M}}_0 \delta(t)[0\ \ 0\ \ 1]^T$. Determine the response, compare the results with those obtained in Problem 4.35 and draw conclusions.

4.41 Determine the response of the system of Problem 4.35 to the excitation $\mathbf{M}(t) = M_0(r(t) - r(t-T))[0\ \ 0\ \ 1]^T$, where $r(t)$ denotes the unit ramp function (Sec. 1.7).

4.42 Derive the response of the system of Example 4.7 to the impulsive excitation $\mathbf{Q}(t) = \hat{Q}_0 \delta(t)[1\ \ 1]^T$. Compare the results with those obtained in Problem 4.36(b) and draw conclusions.

4.43 Determine the response of the system of Problem 4.33 to the excitation $\mathbf{Q}(t) = Q_0 \times u(t)[0\ \ 1\ \ 0]^T$.

4.44 Derive the response of the system of Problem 4.6 to the impulsive excitation $\mathbf{F}(t) = [F_x\ \ F_y]^T = \hat{F}_0 \delta(t)[1\ \ 0]^T$. The system parameters are $M = 1$, $c = 0.1$, $h = 0.1$, $\Omega = 2$, $k_x = 8$ and $k_y = 16$.

4.45 Solve Problem 4.41 in discrete time.

4.46 Solve Problem 4.42 in discrete time.

4.47 Solve Problem 4.43 in discrete time.

4.48 Solve Problem 4.44 in discrete time.

4.49 Let $L_1 = L_2 = L_3 = L$, $m_1 = 2m$, $m_2 = m$, $T/mL = 1$ and use the fourth-order Runge-Kutta method (RK4) to integrate numerically the equations of motion derived in Problem 4.1 for the initial displacements $y_1(0) = y_2(0) = 0.2L$. Work with the nondimensional displacements y_1/L and y_2/L.

4.50 Let $T/mL = 1, \Omega = 1$ and use the fourth-order Runge-Kutta method (RK4) to integrate numerically the equations of motion derived in Problem 4.2 for the initial velocities $\dot{x}(0) = 0.3L$, $\dot{y}(0) = 0$. Work with the nondimensional displacements x/L and y/L.

4.51 Solve Problem 4.49 by the Runge-Kutta-Fehlberg (RKF45) method.

4.52 Solve Problem 4.50 by the Runge-Kutta-Fehlberg (RKF45) method.

4.53 Solve Problem 4.49 by the Adams-Bashforth-Moulton predictor-corrector method

4.54 Solve Problem 4.50 by the Adams-Bashforth-Moulton predictor-corrector method.

4.55 The system of Problem 4.49 is acted upon by the forces $F_1(t) = F_2(t) = F_0 u(t)$, where $F_0/T = 0.25$. The initial conditions are zero. Obtain the response by the fourth-order Runge-Kutta method (RK4).

4.56 Solve Problem 4.55 by the Runge-Kutta-Fehlberg method (RKF45).

4.57 Solve Problem 4.55 by the Adams-Bashforth-Moulton predictor-corrector method.

BIBLIOGRAPHY

1. Bolotin, V. V., *The Dynamic Stability of Elastic Systems*, Holden-Day, San Francisco, 1964.
2. Burden, R. L. and Faires, J. D., *Numerical Analysis*, 5th ed., Prindle, Weber and Schmidt, Boston, 1993.
3. Caughey, T. K. and O'Kelley, M. E. J., "Classical Normal Modes in Damped Linear Dynamic Systems," *Journal of Applied Mechanics*, Vol. 32, 1965, pp. 583–588.
4. Coddington, E. A. and Levinson, N., *Theory of Ordinary Differential Equations*, R. E. Krieger, Melbourne, FL, 1984.
5. Forsythe, G. E., Malcolm, M. A. and Moler, C. B., *Computer Methods for Mathematical Computations*, Prentice Hall, Englewood Cliffs, NJ, 1977.
6. Huseyn, K., *Vibrations and Stability of Multiple Parameter Systems*, Sijthoff & Noordhoff, Alphen aan den Rijn, The Netherlands, 1978.
7. Mathews, J. H., *Numerical Methods for Computer Science, Engineering, and Mathematics*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1992.
8. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan, New York, 1967.
9. Meirovitch, L., *Methods of Analytical Dynamics*, McGraw-Hill, New York, 1970.
10. Meirovitch, L., "A New Method of Solution of the Eigenvalue Problem for Gyroscopic Systems," *AIAA Journal*, Vol. 12, No. 10, 1974, pp. 1337–1342.
11. Meirovitch, L., "A Modal Analysis for the Response of Linear Gyroscopic Systems," *Journal of Applied Mechanics*, Vol. 42, No. 2, 1975, pp. 446–450.
12. Meirovitch, L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.
13. Mingori, D. L., "A Stability Theorem for Mechanical Systems With Constraint Damping," *Journal of Applied Mechanics*, Vol. 37, 1970, pp. 253–258.
14. Murdoch, D. C., *Linear Algebra*, Wiley, New York, 1970.
15. Noble, B. and Daniel, J. W., *Applied Linear Algebra*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1977.
16. Paidoussis, M. P. and Issid, N. T., "Dynamic Stability of Pipes Conveying Fluid," *Journal of Sound and Vibration*, Vol. 33, No. 3, 1974, pp. 267–294.
17. Ziegler, H., *Principles of Structural Stability*, Blaisdell, Waltham, MA, 1968.

5

QUALITATIVE ASPECTS OF THE ALGEBRAIC EIGENVALUE PROBLEM

In Chapter 4, we established the fact that the algebraic eigenvalue problem plays a crucial role in vibrations. Indeed, its solution contains a great deal of information concerning the dynamic characteristics of the system and is instrumental in producing the system response.

Conservative systems represent a very important class in vibrations. They are characterized by a real symmetric eigenvalue problem, which is by far the most desirable type. Indeed, the solution of the eigenvalue problem for real symmetric matrices consists of real eigenvalues and real orthogonal eigenvectors, as demonstrated in Sec. 4.6. These properties of the eigensolutions can be used to prove the stationarity of Rayleigh's quotient, which provides a great deal of insight into the qualitative behavior of the eigensolution. Moreover, it permits the development of the maximin theorem, which in turn permits the development of the separation theorem. Note that the separation theorem can be used to demonstrate the convergence of discrete models approximating distributed ones. From a computational point of view, algorithms for solving real symmetric eigenvalue problems are by far the most stable and efficient.

This chapter is concerned with qualitative aspects of the algebraic eigenvalue problem for both symmetric and nonsymmetric matrices, with the emphasis on the symmetric case.

5.1 GEOMETRIC INTERPRETATION OF THE SYMMETRIC EIGENVALUE PROBLEM

The eigenvalue problem for real symmetric matrices lends itself to a geometric interpretation that is not only very interesting but also suggests a method of solution.

To present this interpretation, we recall the quadratic form given by Eq. (4.108) and consider the equation

$$f = \mathbf{x}^T A\mathbf{x} = 1 \tag{5.1}$$

where A is a real symmetric $n \times n$ matrix and $\mathbf{x}$ is a real nonzero n-vector. Equation (5.1) represents a surface in an n-dimensional Euclidean space. In the case in which A is positive definite the surface represents an n-dimensional ellipsoid with the center at the origin of the Euclidean space. Figure 5.1 depicts only a three-dimensional ellipsoid, $n = 3$, but we will treat it as if it were n-dimensional.

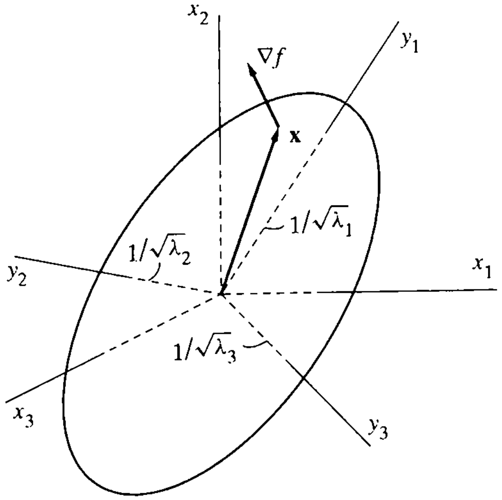


Figure 5.1 Three-dimensional ellipsoid

Next, we consider the gradient of f, namely, a vector ∇f normal to the surface of the ellipsoid and located at a point on the ellipsoid defined by the tip of the vector $\mathbf{x}$. In an n-dimensional space, the gradient of f can be expressed symbolically in the form of the n-vector

$$\nabla f = \left[\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \cdots \; \frac{\partial f}{\partial x_n}\right]^T = \frac{\partial f}{\partial \mathbf{x}} \tag{5.2}$$

so that, inserting Eq. (5.1) into Eq. (5.2), we obtain simply

$$\nabla f = 2A\mathbf{x} \tag{5.3}$$

But, we recall from geometry that the principal axes of an ellipsoid are normal to the surface of the ellipsoid. It follows that, if the vector $\mathbf{x}$ shown in Fig. 5.1 is to be aligned with a principal axis, then it must coincide with ∇f, the difference between the two being a constant of proportionality. Denoting the constant of proportionality by 2λ, we conclude that the condition for the vector $\mathbf{x}$ to be aligned with a principal axis is

$$\nabla f = 2\lambda\mathbf{x} \tag{5.4}$$

Comparing Eqs. (5.3) and (5.4), we conclude that the above condition can be written in the form

$$A\mathbf{x} = \lambda\mathbf{x} \tag{5.5}$$

which is recognized as the eigenvalue problem for the matrix A. Hence, *the eigenvalue problem for real symmetric positive definite matrices A can be interpreted geometrically as the problem of finding the principal axes of the ellipsoid* $f = \mathbf{x}^T A\mathbf{x} = 1$.

In view of the preceding discussion, we consider the problem of solving the eigenvalue problem by finding the principal axes of an ellipsoid. We concentrate first on the planar case, $n = 2$, in which case the ellipsoid described by Eq. (5.1) reduces to an ellipse. Then, we recall from analytic geometry that the problem of determining the principal axes of an ellipse can be solved by means of a coordinate transformation representing a rotation of axes x_1, x_2 through an angle θ, as shown in Fig. 5.2. The angle θ is chosen such that the equation of the ellipse assumes the canonical form, which amounts to annihilating the cross-product entry in the quadratic form resulting from the transformation. From Fig. 5.2, the relation between axes x_1, x_2 and y_1, y_2 is simply

$$\begin{aligned} x_1 &= y_1 \cos\theta - y_2 \sin\theta \\ x_2 &= y_1 \sin\theta + y_2 \cos\theta \end{aligned} \tag{5.6}$$

which can be written in the matrix form

$$\mathbf{x} = R\mathbf{y} \tag{5.7}$$

where $\mathbf{x} = [x_1\ x_2]^T$ and $\mathbf{y} = [y_1\ y_2]^T$ are two-dimensional vectors and

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{5.8}$$

is a 2 × 2 transformation matrix, known as a *rotation matrix*. Using Eq. (5.7), the transformation from the original quadratic form of the ellipse to the canonical form can be written as

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T R^T A R\mathbf{y} = \mathbf{y}^T D\mathbf{y} = 1 \tag{5.9}$$

in which

$$D = R^T A R \tag{5.10}$$

is a diagonal matrix. But, according to Eq. (4.110), if D is a diagonal matrix, then it must by necessity be the matrix Λ of eigenvalues. Moreover, the rotation matrix R must be the orthonormal matrix V of eigenvectors. Clearly, R is an orthonormal matrix, as it satisfies

$$R^T R = RR^T = I \tag{5.11}$$

which is typical of all rotation matrices. Letting $D = \Lambda$ and considering Eq. (5.8), we can write Eq. (5.10) in the explicit form

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{5.12}$$

so that, equating corresponding entries on both sides, we obtain

$$\lambda_1 = a_{11}\cos^2\theta + 2a_{12}\sin\theta\cos\theta + a_{22}\sin^2\theta \tag{5.13a}$$

$$\lambda_2 = a_{11}\sin^2\theta - 2a_{12}\sin\theta\cos\theta + a_{22}\cos^2\theta \tag{5.13b}$$

$$0 = -(a_{11} - a_{22})\sin\theta\cos\theta + a_{12}\left(\cos^2\theta - \sin^2\theta\right) \tag{5.13c}$$

Equation (5.13c) can be rewritten as

$$\tan 2\theta = \frac{2a_{12}}{a_{11} - a_{22}} \tag{5.14}$$

which represents a formula for computing the angle θ required for axes y_1 and y_2 to be principal axes. Then, inserting the angle θ thus obtained into Eqs. (5.13a) and (5.13b), we can compute the eigenvalues of the real symmetric positive definite matrix A. To complete the solution of the eigenvalue problem, we must produce the eigenvectors. To this end, we introduce the same angle θ into the columns of the rotation matrix R, Eq. (5.8), and obtain the orthonormal eigenvectors

$$\mathbf{v}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \tag{5.15}$$

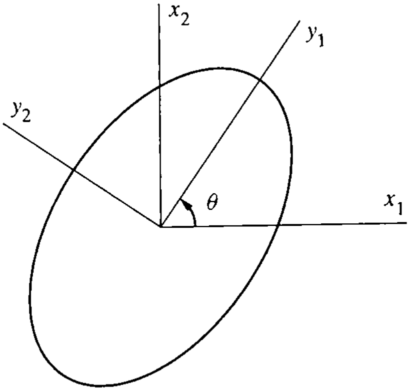


Figure 5.2 Rotation to the principal axes of an ellipse

For $n \geq 3$, the transformation to canonical form cannot be carried out in a single step but in a series of steps. Each of these steps involves a planar rotation designed to annihilate an off-diagonal entry of A (and the symmetric counterpart). Because the annihilated entry does not stay zero following the next transformation step, the determination of the principal axes of the ellipsoid represents an iteration process. The process is convergent. The Jacobi method for solving the eigenvalue problem for real symmetric matrices is based on this idea, for which reason the Jacobi method is referred to as *diagonalization by successive rotations*.

Now, let us premultiply Eq. (5.5) by $\mathbf{x}^T$, consider Eq. (5.1) and write the ellipsoid equation in the form

$$\lambda \mathbf{x}^T\mathbf{x} = \lambda \ \|\mathbf{x}\|^2 = 1 \tag{5.16}$$

where $\|\mathbf{x}\|$ is the Euclidean length of $\mathbf{x}$. Equation (5.16) can be used to relate any eigenvalue to the magnitude of the associated eigenvector. In this regard, it should be pointed out that Eq. (5.16) precludes the normalization process given by Eqs. (4.100), whereby the eigenvectors are rendered unit vectors. Letting $\lambda = \lambda_i$, $\mathbf{x} = \mathbf{v}_i$ in Eq. (5.16), we obtain

$$\lambda_i = \frac{1}{\|\mathbf{v}_i\|^2}, \qquad i = 1, 2, \ldots, n \tag{5.17}$$

so that the Euclidean length of the eigenvector $\mathbf{v}_i$ is inversely proportional to the square root of the associated eigenvalue λ_i (Fig. 5.1). When two eigenvalues are equal, the associated eigenvectors are equal in length. The two eigenvectors are linearly independent and can be rendered orthogonal. The fact that the eigenvectors are equal in length and orthogonal can be interpreted geometrically as the statement that the surface represented by Eq. (5.1) is an ellipsoid of revolution. Hence, any two orthogonal axes in the plane normal to the axis of revolution can be taken as principal axes.

Although we based the preceding discussion on the assumption that the matrix A is positive definite, the assumption was used only to determine the shape of the n-dimensional surface. In fact, the geometric interpretation can be extended to the case in which A is only positive semidefinite. In this case, we conclude from Eqs. (5.17) that the length of the eigenvector belonging to a zero eigenvalue is infinite, so that the ellipsoid degenerates into an infinitely long cylinder with the infinite axis corresponding to the zero eigenvalue.

Example 5.1

Solve the eigenvalue problem for the matrix

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \tag{a}$$

by finding the principal axes of the ellipse $\mathbf{x}^T A\mathbf{x} = 1$.

To obtain the angle θ to the principal axes, we use Eq. (5.14) and write

$$\tan 2\theta = \frac{2a_{12}}{a_{11} - a_{22}} - \frac{2(-1)}{2-1} = -2 \tag{b}$$

It is convenient to restrict the magnitude of the angle so as to satisfy $|\theta| \leq \pi/4$, so that Eq. (b) yields

$$\theta = -31.717475^\circ \tag{c}$$

Hence,

$$\sin\theta = -0.525731, \qquad \cos\theta = 0.850651 \tag{d}$$

Inserting these values in Eqs. (5.13a) and (5.13b), we obtain the eigenvalues

$$\begin{aligned} \lambda_1 &= a_{11}\cos^2\theta + 2a_{12}\sin\theta\cos\theta + a_{22}\sin^2\theta = 2.618034 \\ \lambda_2 &= a_{11}\sin^2\theta - 2a_{12}\sin\theta\cos\theta + a_{22}\cos^2\theta = 0.381966 \end{aligned} \tag{e}$$

Moreover, from Eqs. (5.15), the associated eigenvectors are

$$\mathbf{v}_1 = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} 0.850651 \\ -0.525731 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} = \begin{bmatrix} 0.525731 \\ 0.850651 \end{bmatrix} \tag{f}$$

The eigenvalue problem corresponds to the vibration of a two-degree-of-freedom system like the one shown in Fig. 4.9 with the parameters $m_1 = m_2 = 1$, $c_1 = c_2 = 0$ and $k_1 = k_2 = 1$. In view of this, we conclude on physical grounds that λ_2 and $\mathbf{v}_2$ actually correspond to the first mode and λ_1 and $\mathbf{v}_1$ to the second. The explanation for this is that the angle θ specified by Eq. (b) is to one of the principal axes, not necessarily to the axis corresponding to the lowest mode. Hence, relabeling the modes, we obtain the matrices of eigenvalues and eigenvectors

$$\Lambda = \begin{bmatrix} 0.381966 & 0 \\ 0 & 2.618034 \end{bmatrix}, \qquad V = \begin{bmatrix} 0.525731 & 0.850651 \\ 0.850651 & -0.525731 \end{bmatrix} \tag{g}$$

As we shall see in Sec. 6.4, it is typical of the Jacobi method that the modes do not necessarily appear in ascending order in the computed matrices of eigenvalues and eigenvectors, so that a rearrangement of these matrices may be required.

5.2 THE STATIONARITY OF RAYLEIGH'S QUOTIENT

In Sec. 4.6, we have shown that a real symmetric positive definite $n \times n$ matrix A possesses n real and positive eigenvalues λ_r and n mutually orthogonal real eigenvectors $\mathbf{v}_r$ $(r = 1, 2, \ldots, n)$ satisfying the eigenvalue problem

$$A\mathbf{v}_r = \lambda_r \mathbf{v}_r, \qquad r = 1, 2, \ldots, n \tag{5.18}$$

In this section, and the several following ones, we examine some qualitative properties of the solutions to the eigenvalue problem. To this end, we arrange the eigenvalues in ascending order of magnitude, so that they satisfy the inequalities $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Premultiplying both sides of Eqs. (5.18) by $\mathbf{v}_r^T$ and dividing by $\mathbf{v}_r^T \mathbf{v}_r$, we conclude that every eigenvalue λ_r can be expressed as the ratio

$$\lambda_r = \frac{\mathbf{v}_r^T A \mathbf{v}_r}{\mathbf{v}_r^T \mathbf{v}_r}, \qquad r = 1, 2, \ldots, n \tag{5.19}$$

Equations (5.19) imply that, provided the eigenvector $\mathbf{v}_r$ belonging to λ_r is known, the eigenvalue λ_r can be produced by simply computing the indicated ratio.

Next, we replace λ_r and $\mathbf{v}_r$ in Eqs. (5.19) by λ and $\mathbf{v}$, respectively, and obtain

$$\lambda(\mathbf{v}) = R(\mathbf{v}) = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \tag{5.20}$$

which is known as *Rayleigh's quotient* and is clearly a function of $\mathbf{v}$. A question of particular interest concerns the behavior of Rayleigh's quotient as $\mathbf{v}$ ranges over the entire n-dimensional Euclidean space. To this end, we recall from Sec. 4.6 that, according to the expansion theorem, Eqs. (4.118)–(4.121), any arbitrary n-vector $\mathbf{v}$ can be expressed as a linear combination of the system eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, and we note that these are unit eigenvectors satisfying the orthonormality relations (4.100) and (4.101). Hence, inserting Eq. (4.118) into Eq. (5.20) and considering Eqs. (4.104) and (4.107), which are the matrix counterpart of Eqs. (4.100) and (4.101), we obtain

$$\lambda = \frac{\mathbf{c}^T V^T A V \mathbf{c}}{\mathbf{c}^T V^T V \mathbf{c}} = \frac{\mathbf{c}^T \Lambda \mathbf{c}}{\mathbf{c}^T \mathbf{c}} = \frac{\sum_{i=1}^n \lambda_i c_i^2}{\sum_{i=1}^n c_i^2} \tag{5.21}$$

As the arbitrary vector $\mathbf{v}$ wanders over the n-dimensional Euclidean space, it will eventually enter a small neighborhood of a given eigenvector, say $\mathbf{v}_r$, as shown in Fig. 5.3. But, as discussed in Sec. 4.6, the coefficients c_i $(i = 1, 2, \ldots, n)$ in Eq. (4.118), as well as in Eq. (5.21), represent the coordinates of $\mathbf{v}$ with respect to the basis, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$, i.e., the projections of the vector $\mathbf{v}$ onto the axes defined by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. But, because $\mathbf{v}$ is inside a small neighborhood of $\mathbf{v}_r$, it follows that the magnitude of the coefficients must satisfy

$$|c_r| \gg |c_i|, \qquad i = 1, 2, \ldots, n; \qquad i \neq r \tag{5.22}$$

which is the same as saying that

$$\frac{|c_i|}{|c_r|} = \epsilon_i, \qquad i = 1, 2, \ldots, n; \qquad i \neq r \tag{5.23}$$

where ϵ_i are small numbers. Inserting Eqs. (5.23) into Eq. (5.21), using the binomial approximation $(1 + \delta)^{-1} \cong 1 - \delta$, where δ is a small quantity, and ignoring higher-order terms in ϵ_i^2, we obtain

$$\lambda = \frac{\lambda_r c_r^2 + \sum_{\substack{i=1\\i\neq r}}^n \lambda_i c_i^2}{c_r^2 + \sum_{\substack{i=1\\i\neq r}}^n c_i^2} = \frac{\lambda_r + \sum_{\substack{i=1\\i\neq r}}^n \lambda_i \epsilon_i^2}{1 + \sum_{\substack{i=1\\i\neq r}}^n \epsilon_i^2}$$
$$\cong \left(\lambda_r + \sum_{\substack{i=1\\i\neq r}}^n \lambda_i \epsilon_i^2\right)\left(1 - \sum_{\substack{i=1\\i\neq r}}^n \epsilon_i^2\right) \cong \lambda_r + \sum_{i=1}^n (\lambda_i - \lambda_r)\,\epsilon_i^2 \tag{5.24}$$

But, Eqs. (5.23) imply that $\mathbf{v}$ differs from $\mathbf{v}_r$ by a small quantity of first order in ϵ_i, $\mathbf{v} = \mathbf{v}_r + \mathbf{O}(\epsilon)$. On the other hand, Eq. (5.24) states that the corresponding Rayleigh's quotient λ differs from λ_r by a small quantity of second order in ϵ_i, $\lambda = \lambda_r + O\left(\epsilon^2\right)$. This result can be stated in the form of the theorem: *Rayleigh's quotient corresponding to a real symmetric positive definite matrix has stationary values in the neighborhood of the eigenvectors, where the stationary values are equal to the associated eigenvalues.* Although demonstrated and worded by Rayleigh differently, this is the essence of *Rayleigh's principle* (Ref. 10, Sec. 88).

Assuming that λ is proportional to ω^2, the special case in which $r = 1$ is by far the most important one, not only because it corresponds to the lowest natural frequency ω_1, which tends to be the most important one, but also because in this case Rayleigh's quotient has a minimum. To show this, we let $r = 1$ in Eq. (5.24), so that

$$\lambda = \lambda_1 + \sum_{i=2}^n (\lambda_i - \lambda_1)\,\epsilon_i^2 \geq \lambda_1 \tag{5.25}$$

where we recognize that the series is always positive, because it represents a quadratic form with positive coefficients. Inequality (5.25) states that *Rayleigh's quotient is never lower than the lowest eigenvalue* λ_1. It is generally higher than λ_1, except

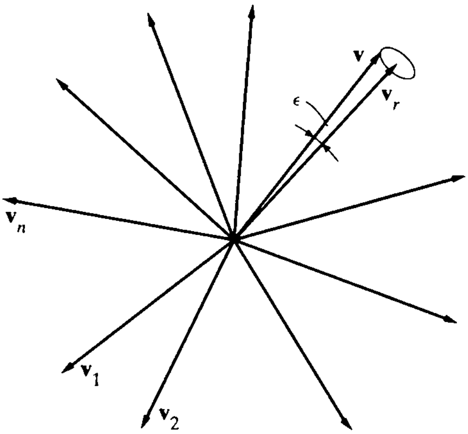


Figure 5.3 Arbitrary vector in the neighborhood of an eigenvector

when $\mathbf{v}$ is identically equal to $\mathbf{v}_1$, in which case *Rayleigh's quotient has a minimum value at* $\mathbf{v} = \mathbf{v}_1$ *equal to* λ_1, or

$$\lambda_1 = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \tag{5.26}$$

and this is the only minimum value. Because this case is of such importance relative to the other cases, the statement that λ_1 *is the minimum value of Rayleigh's quotient* is alone referred to as *Rayleigh's principle* (see, for example, Ref. 6, p. 31). This version of Rayleigh's principle has significant implications in deriving approximate solutions to the eigenvalue problem for distributed-parameter systems. Another interpretation of inequality (5.25) is that *Rayleigh's quotient provides an upper bound for the lowest eigenvalue* λ_1.

Following a similar argument, it is easy to verify that for $r = n$ Eq. (5.24) yields

$$\lambda = \lambda_n - \sum_{i=1}^{n-1} (\lambda_n - \lambda_i)\epsilon_i^2 \le \lambda_n \tag{5.27}$$

or, *Rayleigh's quotient is never higher than the highest eigenvalue* λ_n. It is generally lower than λ_n, except when $\mathbf{v}$ is identically equal to $\mathbf{v}_n$, in which case *Rayleigh's quotient has a maximum value at* $\mathbf{v} = \mathbf{v}_n$ *equal to* λ_n, or

$$\lambda_n = \max \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \tag{5.28}$$

Inequality (5.27) can also be interpreted as saying that *Rayleigh's quotient provides a lower bound for the highest eigenvalue* λ_n.

The behavior of Rayleigh's quotient can be interpreted geometrically in a manner similar to that in Sec. 5.1. To this end, we recall that in the case in which A is a real symmetric positive definite matrix the equation

$$\mathbf{v}^T A \mathbf{v} = \lambda \mathbf{v}^T \mathbf{v} = \lambda \, \|\mathbf{v}\|^2 = 1 \tag{5.29}$$

represents an n-dimensional ellipsoid with the length of the semiaxes equal to $\|\mathbf{v}_i\| = 1/\sqrt{\lambda_i}$, consistent with Eqs. (5.17), and raises the question as to how Rayleigh's quotient changes as the tip of the vector **v** slides across the surface of the ellipsoid. We recognize that not much happens at arbitrary points, so that we confine our discussions to motions of **v** in planes defined by pairs of eigenvectors. As it moves in such a plane, **v** aligns itself eventually with an eigenvector. But, because eigenvectors coincide with principal axes, they are normal to the surface of the ellipsoid, so that the rate of change of $\|\mathbf{v}\|$, and hence the rate of change of λ is zero at an eigenvector, which indicates that λ has a stationary value at an eigenvector. To examine the nature of the stationary value, and provide a clearer picture at the same time, it is convenient to refer once again to a three-dimensional ellipsoid, such as that shown in Fig. 5.4. With **v** temporarily at $\mathbf{v}_1$, a move toward $\mathbf{v}_2$, or toward $\mathbf{v}_3$, tends to produce a decrease in $\|\mathbf{v}\|$, and hence an increase in λ, according to Eq. (5.29). This indicates that λ *has a minimum at* $\mathbf{v}_1$, so that in the neighborhood of $\mathbf{v}_1$ Rayleigh's quotient resembles the bowl shown in Fig. 5.5a. Using the same argument, with **v** at $\mathbf{v}_2$, a move toward $\mathbf{v}_1$ tends to produce a decrease in λ and a move toward $\mathbf{v}_3$ tends to produce an increase in λ, which is equivalent to the statement that λ *has a mere stationary value at* $\mathbf{v}_2$. Hence, in the neighborhood of $\mathbf{v}_2$, Rayleigh's quotient resembles the saddle shown in Fig. 5.5b. Finally, at $\mathbf{v} = \mathbf{v}_3$, λ decreases as **v** moves toward $\mathbf{v}_1$ or toward $\mathbf{v}_2$, which indicates that λ *has a maximum at* $\mathbf{v}_3$. Consistent with this, Rayleigh's quotient resembles an inverted bowl with the apex at the tip of $\mathbf{v}_3$, as shown in Fig. 5.5c.

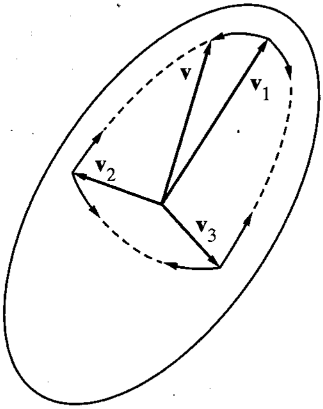


Figure 5.4 Three-dimensional ellipsoid with eigenvectors as semiaxes

The question arises as to how the preceding developments apply to our vibration problems. To answer this question, we insert Eqs. (4.84) and (4.87) into Eq. (5.20), consider Eq. (4.82) and obtain the Rayleigh's quotient in the form

$$\lambda = \frac{\mathbf{u}^T Q^T \left(Q^T\right)^{-1} K Q^{-1} Q \mathbf{u}}{\mathbf{u}^T Q^T Q \mathbf{u}} = \frac{\mathbf{u}^T K \mathbf{u}}{\mathbf{u}^T M \mathbf{u}} \tag{5.30}$$

where **u** is a vector of actual displacement amplitudes, as opposed to **v** which is not, K is the stiffness matrix and M is the mass matrix. It follows that all the properties of Rayleigh's quotient demonstrated for the form (5.20) hold true when the quotient is in the form (5.30). Although the form (5.20) is more useful in interpreting the eigenvalue problem geometrically, the form (5.30) has its own advantages, the most important one being that it returns us to the physical world. Indeed, because λ is proportional to ω^2, Eq. (5.30) permits us to conclude that the natural frequencies can be increased by increasing the stiffness, or decreasing the mass, or both, and vice versa.

Rayleigh's quotient provides an expedient way of estimating eigenvalues, and in particular the lowest eigenvalue. The procedure amounts to guessing the shape of a certain mode of vibration, inserting the guessed mode into Rayleigh's quotient, Eq. (5.30), and computing an estimate of the corresponding eigenvalue. It should be noted here that, because of the stationarity of Rayleigh's quotient, estimates of eigenvalues tend to be one order of magnitude more accurate than the guessed eigenvectors. The usefulness of the procedure is limited primarily to the lowest eigenvalue, because the shape of the lowest mode is the easiest to guess. Indeed, quite often a reasonably accurate guess consists of the static displacement vector of the system subjected to forces proportional to the masses. No such guessing aids exist for the higher modes. The fact that Rayleigh's quotient is able to provide reasonably accurate estimates of the lowest natural frequency is very fortunate, because the lowest natural frequency is more often than not the most important.

Rayleigh's quotient is a concept of towering importance to the eigenvalue problem associated with vibrating discrete systems. The usefulness of the concept is pervasive, extending not only to analytical developments but also to computational algorithms. Moreover, the concept is as vital to differential eigenvalue problems associated with distributed-parameter systems, to be discussed in later chapters, as it is to the algebraic eigenvalue problem discussed in this chapter.

Example 5.2

Verify that Rayleigh's quotient associated with the real symmetric positive definite matrix

$$A = \begin{bmatrix} 2.5 & -1 & 0 \\ -1 & 5 & -1.414214 \\ 0 & -1.414214 & 10 \end{bmatrix} \tag{a}$$

has a minimum at the lowest eigenvector and a stationary value at the second eigenvector. The matrices of eigenvalues and eigenvectors of A are

$$\Lambda = \begin{bmatrix} 2.119322 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10.380678 \end{bmatrix}$$

$$V = \begin{bmatrix} 0.932674 & -0.359211 & 0.032965 \\ 0.355049 & 0.898027 & -0.259786 \\ 0.063715 & 0.254000 & 0.965103 \end{bmatrix} \tag{b}$$

To verify that Rayleigh's quotient has a minimum at $\mathbf{v}_1$, we consider the trial vector $\mathbf{v} = [0.8968\ 0.4449\ 0.0891]^T$, which corresponds roughly to $\mathbf{v}_1 + 0.1\mathbf{v}_2$, and we note that $\mathbf{v}$ has not been normalized. Hence, the value of Rayleigh's quotient is

$$\lambda = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\begin{bmatrix} 0.8968 \\ 0.4449 \\ 0.0891 \end{bmatrix}^T \begin{bmatrix} 2.5 & 0 & 0 \\ -1 & 5 & -1.414214 \\ 0 & -1.414214 & 10 \end{bmatrix} \begin{bmatrix} 0.8968 \\ 0.4449 \\ 0.0891 \end{bmatrix}}{\begin{bmatrix} 0.8968 \\ 0.4449 \\ 0.0891 \end{bmatrix}^T \begin{bmatrix} 0.8968 \\ 0.4449 \\ 0.0891 \end{bmatrix}}$$

$$= 2.1477 \tag{c}$$

so that the estimate is higher than $\lambda_1 \cong 2.1193$. The error in the estimate is

$$\epsilon_1 = \frac{\lambda - \lambda_1}{\lambda_1} \cong \frac{2.1477 - 2.1153}{2.1193} = 0.0134 \tag{d}$$

which is one order of magnitude smaller than the difference between $\mathbf{v}$ and $\mathbf{v}_1$. It can be verified that, regardless of the choice of $\mathbf{v}$, the value of λ will always be larger than λ_1, so that Rayleigh's quotient has a minimum at $\mathbf{v}_1$.

To verify the stationarity of Rayleigh's quotient, we first consider the trial vector $\mathbf{v} = [-0.2659 \ \ 0.9335 \ \ 0.2604]^T$, which is roughly equal to $\mathbf{v}_2 + 0.1\mathbf{v}_1$. Hence, we write

$$\lambda = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\begin{bmatrix} -0.2659 \\ 0.9335 \\ 0.2604 \end{bmatrix}^T \begin{bmatrix} 2.5 & 0 & 0 \\ -1 & 5 & -1.414214 \\ 0 & -1.414214 & 10 \end{bmatrix} \begin{bmatrix} -0.2659 \\ 0.9335 \\ 0.2604 \end{bmatrix}}{\begin{bmatrix} -0.2659 \\ 0.9335 \\ 0.2604 \end{bmatrix}^T \begin{bmatrix} -0.2659 \\ 0.9335 \\ 0.2604 \end{bmatrix}}$$

$$= 4.9716 \tag{e}$$

so that the estimate is lower than $\lambda_2 = 5$. The error is

$$\epsilon_2 = \frac{\lambda - \lambda_2}{\lambda_2} = \frac{4.9716 - 5}{5} = -0.0057 \tag{f}$$

which is clearly one order of magnitude smaller than the difference between the trial vector $\mathbf{v}$ and $\mathbf{v}_2$. Next, we consider the trial vector $\mathbf{v} = [-0.3559 \ \ 0.8720 \ \ 0.3505]^T$, which is roughly equal to $\mathbf{v}_2 + 0.1\mathbf{v}_3$. Inserting this trial vector into Rayleigh's quotient, we obtain

$$\lambda = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\begin{bmatrix} -0.3559 \\ 0.8720 \\ 0.3505 \end{bmatrix}^T \begin{bmatrix} 2.5 & 0 & 0 \\ -1 & 5 & -1.414214 \\ 0 & -1.414214 & 10 \end{bmatrix} \begin{bmatrix} -0.3559 \\ 0.8720 \\ 0.3505 \end{bmatrix}}{\begin{bmatrix} -0.3559 \\ 0.8720 \\ 0.3505 \end{bmatrix}^T \begin{bmatrix} -0.3559 \\ 0.8720 \\ 0.3505 \end{bmatrix}}$$

$$= 5.0528 \tag{g}$$

so that this estimate is higher than λ_2. The corresponding error is

$$\epsilon_2 = \frac{\lambda - \lambda_2}{\lambda_2} = \frac{5.0528 - 5}{5} = 0.0106 \tag{h}$$

which is once again one order of magnitude smaller than the difference between $\mathbf{v}$ and $\mathbf{v}_2$. Clearly, Rayleigh's quotient has a mere stationary value at $\mathbf{v} = \mathbf{v}_2$, as the estimate can be either larger or smaller than λ_2.

At this point, we wish to provide a geometric interpretation of the results. The trial vector $\mathbf{v} = \mathbf{v}_1 + 0.1\mathbf{v}_2$ represents a vector with the tip lying on the curve marked "Toward $\mathbf{v}_2$" in Fig. 5.5a. The value of λ corresponding to this choice is larger than the value λ_1 at $\mathbf{v}_1$. It is obvious that any other choice would not change the general picture, as the surface depicting λ is a bowl with the lowest point at $\mathbf{v}_1$. On the other hand, the surface depicting λ in the neighborhood of $\mathbf{v}_2$ is a saddle, as shown in Fig. 5.5b. The first choice of trial vector corresponds to a point on the curve marked "Toward $\mathbf{v}_1$," for which the surface dips below λ_2, and the second choice corresponds to a point on the curve "Toward $\mathbf{v}_3$," for which the surface rises above λ_2.

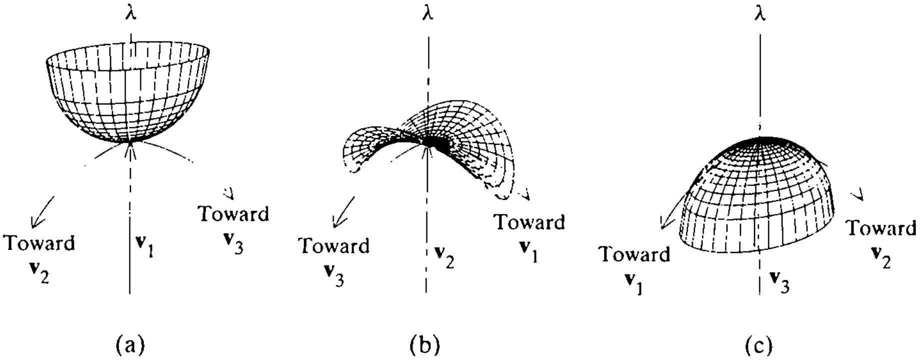


Figure 5.5 **(a)** Minimum value of Rayleigh's quotient at lowest eigenvector **(b)** Stationary value of Rayleigh's quotient at intermediate eigenvector **(c)** Maximum value of Rayleigh's quotient at highest eigenvector

5.3 MAXIMUM-MINIMUM CHARACTERIZATION OF THE EIGENVALUES

In Sec. 5.2, we have shown that Rayleigh's quotient associated with a real symmetric positive definite matrix has a stationary value in the neighborhood of an eigenvector, where the stationary value is equal to the associated eigenvalue. The lowest eigenvalue λ_1 plays a special role for various reasons. First among these is the fact that it corresponds to the lowest natural frequency ω_1, which is the most important one in vibrations. Moreover, in the case of the lowest eigenvalue, the stationary value is actually a minimum. Indeed, as shown in Sec. 5.2, the lowest eigenvalue λ_1 of a vibrating system is the minimum value Rayleigh's quotient $\lambda(\mathbf{v})$, Eq. (5.20), can take as the arbitrary n-vector $\mathbf{v}$ ranges over the n-dimensional Euclidean space, or

$$\lambda_1 = \min \lambda(\mathbf{v}) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \tag{5.31}$$

Another desirable feature of the lowest eigenvalue is that it is the easiest to estimate, because it is much easier to generate a vector resembling the lowest eigenvector than any other eigenvector. Finally, the fact that Rayleigh's quotient cannot fall below λ_1 makes it possible to improve the estimate by devising a sequence of vectors $\mathbf{v}$ designed to minimize the value of Rayleigh's quotient, as such a minimization process is certain to cause $\lambda(\mathbf{v})$ to approach λ_1. Clearly, this desirable feature is not possessed by the intermediate eigenvalues, as the intermediate eigenvectors are only saddle points.

In view of the above, the question arises as to whether there are any circumstances under which statements similar to those made for λ_1 can be made for the intermediate eigenvalues. In addressing this question, we concentrate first on λ_2. To this end, we propose to modify the series expansion for $\mathbf{v}$, Eq. (4.118), by omitting the first eigenvector $\mathbf{v}_1$, so that

$$\mathbf{v} = \sum_{i=2}^{n} c_i \mathbf{v}_i = V\mathbf{c} \tag{5.32}$$

where this time the vector $\mathbf{c}$ has the form

$$\mathbf{c} = [0\; c_2\; c_3\; \dots\; c_n]^T \tag{5.33}$$

This implies that the vector $\mathbf{v}$ is not entirely arbitrary but orthogonal to the first eigenvector $\mathbf{v}_1$. Indeed, premultiplying $\mathbf{v}_1$ by $\mathbf{v}^T$ where $\mathbf{v}$ is given by Eq. (4.118), and recalling the orthonormality relations, Eqs. (4.100), we can write the orthogonality condition as follows:

$$\mathbf{v}^T\mathbf{v}_1 = \mathbf{c}^T V^T \mathbf{v}_1 = \mathbf{c}^T [\mathbf{v}_1\; \mathbf{v}_2\; \dots\; \mathbf{v}_n]^T \mathbf{v}_1 = \mathbf{c}^T [1\; 0\; \dots\; 0]^T = c_1 = 0 \tag{5.34}$$

Equation (5.34) represents a constraint equation implying that the vector $\mathbf{v}$ can only range over an $(n-1)$-dimensional Euclidean *space of constraint orthogonal to the vector* $\mathbf{v}_1$. This concept can be best visualized geometrically by means of Fig. 5.6, in which the space of constraint is simply the plane orthogonal to $\mathbf{v}_1$, i.e., the plane defined by the eigenvectors $\mathbf{v}_2$ and $\mathbf{v}_3$. Returning to the n-dimensional case and following the same approach as in Sec. 5.2, it is not difficult to show that

$$\lambda(\mathbf{v}) = \lambda_2 + \sum_{i=3}^{n} (\lambda_i - \lambda_2)\epsilon_i^2 \geq \lambda_2, \qquad \mathbf{v}^T\mathbf{v}_1 = 0 \tag{5.35}$$

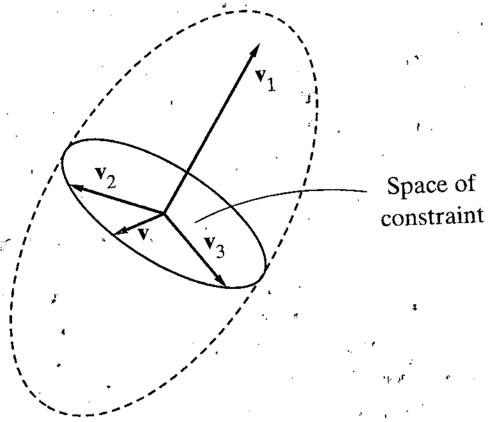


Figure 5.6. Arbitrary vector constrained to a plane orthogonal to $\mathbf{v}_1$

from which it follows that

$$\lambda_2 = \min \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}, \qquad \mathbf{v}^T\mathbf{v}_1 = 0 \tag{5.36}$$

or *Rayleigh's quotient has the minimum value of* λ_2 *for all trial vectors* $\mathbf{v}$ *orthogonal to the first eigenvector* $\mathbf{v}_1$, where *the minimum is reached at* $\mathbf{v} = \mathbf{v}_2$. Hence, at least in theory, it is possible to produce an estimate of λ_2 by guessing a vector $\mathbf{v}$ approximating the eigenvector $\mathbf{v}_2$ within a small quantity of first order. Clearly, this vector $\mathbf{v}$ must be rendered orthogonal to $\mathbf{v}_1$, which can be done by a process such as the Gram-Schmidt orthogonalization process (see Appendix B).

The approach can be extended to higher eigenvalues by constraining the trial vector $\mathbf{v}$ to be orthogonal to a suitable number of lower eigenvectors. For example, λ_{r+1} can be characterized by requiring that $\mathbf{v}$ be from the $(n - r)$-dimensional Euclidean space of constraint orthogonal to the first r eigenvectors, or

$$\lambda_{r+1} = \min \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}, \qquad \mathbf{v}^T\mathbf{v}_i = 0, \qquad i = 1, 2, \ldots, r \tag{5.37}$$

Causing a given eigenvalue, say λ_j, to acquire the extremum characteristics of λ_1 by requiring that the trial vector $\mathbf{v}$ be from the space of constraint orthogonal to the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1}$ is not very practical, because these lower eigenvectors are generally not available. Hence, the question arises naturally as to whether it is possible to achieve the same objective without relying on the corresponding lower eigenvectors. In the following, we do indeed develop such a method. However, the usefulness of this development does not lie so much in estimating eigenvalues as in providing a rigorous analytical foundation to a characterization of the eigenvalues with significant implications in approximate solutions to differential eigenvalue problems, as discussed later in this text.

We consider a given n-vector $\mathbf{w}$ and constrain the trial vector $\mathbf{v}$ to be from the $(n - 1)$-dimensional Euclidean space of constraint orthogonal to $\mathbf{w}$, so that $\mathbf{v}$ is not entirely arbitrary but must satisfy the constraint equation

$$\mathbf{v}^T\mathbf{w} = 0 \tag{5.38}$$

Geometrically, Eq. (5.38) confines the vector $\mathbf{v}$ to an $(n - 1)$-dimensional ellipsoid of constraint defined by the intersection of the n-dimensional ellipsoid associated with the real symmetric positive definite matrix A and the $(n - 1)$-dimensional Euclidean space of constraint orthogonal to $\mathbf{w}$. To clarify the idea, we refer to the three-dimensional ellipsoid of Fig. 5.7, from which we conclude that the space of constraint orthogonal to $\mathbf{w}$ is once again a plane, but this time the plane is a general one, not necessarily containing two eigenvectors, or even one eigenvector. Hence, $\mathbf{v}$ is confined to the ellipse resulting from the intersection of the ellipsoid and the plane normal to $\mathbf{w}$, as shown in Fig. 5.7.

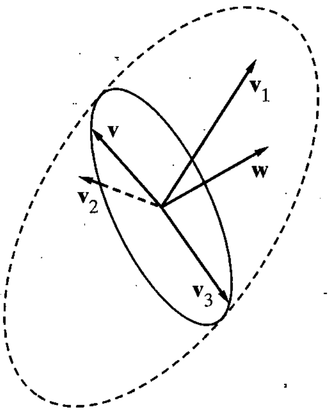


Figure 5.7 Arbitrary vector constrained to a space orthogonal to $\mathbf{w}$

The $n \times n$ real symmetric positive definite matrix A has the eigenvalues λ_1, $\lambda_2, \ldots, \lambda_n$. We can envision an $(n-1) \times (n-1)$ real symmetric positive definite matrix $\tilde{A}$ corresponding to the $(n-1)$-dimensional ellipsoid of constraint resulting from imposing on $\mathbf{v}$ the constraint given by Eq. (5.38) and denote the eigenvalues of $\tilde{A}$ by $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_{n-1}$. The question of interest here is how the eigenvalues $\tilde{\lambda}_1$, $\tilde{\lambda}_2, \ldots, \tilde{\lambda}_{n-1}$ of the constrained system relate to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the original unconstrained system. To answer this question, we concentrate first on $\tilde{\lambda}_1$ and introduce the definition

$$\tilde{\lambda}_1(\mathbf{w}) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \qquad \mathbf{v}^T \mathbf{w} = 0 \tag{5.39}$$

where $\tilde{\lambda}_1$ clearly depends on $\mathbf{w}$. But, the longest principal axis of the $(n-1)$-dimensional ellipsoid of constraint associated with $\tilde{A}$ is generally shorter than the eigenvector $\mathbf{v}_1$ belonging to λ_1, so that by analogy with Eq. (5.17) we conclude that $\tilde{\lambda}_1 > \lambda_1$. The only exception is when $\mathbf{w}$ coincides with one of the higher eigenvectors $\mathbf{v}_r$ $(r = 2, 3, \ldots, n)$, in which case $\mathbf{w} = \mathbf{v}_1$ is in the $(n-1)$-dimensional space of constraint and $\tilde{\lambda}_1 = \lambda_1$. Hence, $\tilde{\lambda}_1$ satisfies the inequality

$$\lambda_1 \leq \tilde{\lambda}_1 \tag{5.40}$$

This is consistent with the fact that constraints tend to increase the system stiffness. The question remains as to the highest value $\tilde{\lambda}_1$ can reach. To answer this question, we consider the trial vector

$$\mathbf{v} = c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 \tag{5.41}$$

Clearly, this choice is possible, as the fact that $\mathbf{v}$ must be orthogonal to $\mathbf{w}$ only means that c_1 and c_2 depend on one another. Next, we introduce Eq. (5.41) into Rayleigh's quotient, invoke the orthonormality relations, Eqs. (4.100) and (4.101), and write

$$\lambda(\mathbf{v}) = \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}} = \frac{(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)^T A (c_1\mathbf{v}_1 + c_2\mathbf{v}_2)}{(c_1\mathbf{v}_1 + c_2\mathbf{v}_2)^T (c_1\mathbf{v}_1 + c_2\mathbf{v}_2)} = \frac{c_1^2\lambda_1 + c_2^2\lambda_2}{c_1^2 + c_2^2}$$
$$\leq \frac{\left(c_1^2 + c_2^2\right)\lambda_2}{c_1^2 + c_2^2} = \lambda_2 \tag{5.42}$$

But Eq. (5.39) implies that

$$\tilde{\lambda}_1(\mathbf{w}) \leq \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}, \qquad \mathbf{v}^T\mathbf{w} = 0 \tag{5.43}$$

Hence, combining inequalities (5.40), (5.42) and (5.43), we conclude that

$$\lambda_1 \leq \tilde{\lambda}_1 \leq \lambda_2 \tag{5.44}$$

or *the first eigenvalue of a system with one constraint lies between the first and the second eigenvalue of the original unconstrained system.* This statement is known as *Rayleigh's theorem for systems with one constraint.* It should be noted here that the choice of $\mathbf{v}$ in the form given by Eq. (5.41) was motivated by our desire to define the range of $\tilde{\lambda}_1$ as sharply as possible. Indeed, any other choice of $\mathbf{v}$ would replace λ_2 in the right inequality in (5.44) by a higher value.

Rayleigh's theorem for systems with one constraint can be given a geometric interpretation by referring to the three-dimensional ellipsoid $\mathbf{v}^T A\mathbf{v} = 1$ shown in Fig. 5.7. As can be concluded from Fig. 5.7, the space of constraint is an ellipse resulting from the intersection of the plane normal to $\mathbf{w}$ and the ellipsoid. When $\mathbf{w}$ coincides with $\mathbf{v}_2$, the ellipse has principal axes $\mathbf{v}_1$ and $\mathbf{v}_3$. Because $\|\mathbf{v}_1\| \geq \|\mathbf{v}_3\|$, if we recall that at any point on the ellipsoid $\lambda = 1/\|\mathbf{v}\|$, it follows that the minimum value $\tilde{\lambda}_1$ can take on the ellipse defined by $\mathbf{v}_1$ and $\mathbf{v}_3$ is $\tilde{\lambda}_1 = \lambda_1$. If $\mathbf{w} \neq \mathbf{v}_2$, the longest axis of the intersecting ellipse is shorter than $\|\mathbf{v}_1\|$, so that $\tilde{\lambda}_1 \geq \lambda_1$. When $\mathbf{w}$ coincides with $\mathbf{v}_1$, the ellipse of constraint has principal axes $\mathbf{v}_2$ and $\mathbf{v}_3$. Hence, following the same line of thought, we conclude that the minimum value $\tilde{\lambda}_1$ can take on the ellipse defined by $\mathbf{v}_2$ and $\mathbf{v}_3$ is $\tilde{\lambda}_1 = \lambda_2$. If $\mathbf{w} \neq \mathbf{v}_1$, the longest axis of the intersecting ellipse normal to $\mathbf{w}$ is longer than $\mathbf{v}_2$, so that $\tilde{\lambda}_1 \leq \lambda_2$. This completes the geometric proof of inequalities (5.44), and hence of the theorem.

Inequality (5.44) characterizes $\tilde{\lambda}_1$ in relation to λ_1 and λ_2. The right side of the inequality, however, can be regarded as characterizing λ_2. Indeed, using Eq. (5.39), it can be reinterpreted as stating that

$$\lambda_2 = \max\tilde{\lambda}_1(\mathbf{w}) = \max_{\mathbf{w}}\left(\min_{\mathbf{v}}\frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\right), \qquad \mathbf{v}^T\mathbf{w} = 0 \tag{5.45}$$

This new interpretation of Rayleigh's theorem for systems with one constraint can be stated in the form of the theorem: *The second eigenvalue λ_2 of a real symmetric positive definite matrix A is the maximum value that can be given to* min $\left(\mathbf{v}^T A\mathbf{v}/\mathbf{v}^T\mathbf{v}\right)$ *by the imposition of the single constraint* $\mathbf{v}^T\mathbf{w} = 0$, where the maximum is with respect to $\mathbf{w}$ and the minimum is with respect to all vectors $\mathbf{v}$ satisfying the imposed constraint.

The preceding theorem can be extended to any number r of constraints, $r < n$, thus providing a characterization of the eigenvalue λ_{r+1} of the real symmetric positive definite matrix A that is independent of the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ of A. To this end, we consider r independent n-vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ and introduce the definition

$$\tilde{\lambda}_r(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \qquad \mathbf{v}^T \mathbf{w}_i = 0, \qquad i = 1, 2, \ldots, r \tag{5.46}$$

where $\tilde{\lambda}_r$ is a continuous function of $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$. In the special case in which the arbitrary constraint vectors $\mathbf{w}_i$ coincide with the eigenvectors $\mathbf{v}_i$ of $A(i = 1, 2, \ldots, r)$, we conclude from Eq. (5.37) that

$$\tilde{\lambda}_r(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r) = \lambda_{r+1}, \qquad \mathbf{w}_i = \mathbf{v}_i, \qquad i = 1, 2, \ldots, r \tag{5.47}$$

Next, we assume that the constraint vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ are given. Then, following the same pattern as that followed earlier for one constraint, we assume that the trial vector $\mathbf{v}$ has the form

$$\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_{r+1}\mathbf{v}_{r+1} = \sum_{i=1}^{r+1} c_i \mathbf{v}_i \tag{5.48}$$

so that, using the same argument as for one constraint, it can be shown that this choice is consistent with our minimization problem, Eq. (5.46). Hence, inserting Eq. (5.48) into Rayleigh's quotient and considering the orthonormality relations, Eqs. (4.100) and (4.101), we obtain

$$\lambda(\mathbf{v}) = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} = \frac{\left(\sum_{i=1}^{r+1} c_i \mathbf{v}_i^T\right) A \left(\sum_{j=1}^{r+1} c_j \mathbf{v}_j\right)}{\left(\sum_{i=1}^{r+1} c_i \mathbf{v}_i^T\right)\left(\sum_{j=1}^{r+1} c_j \mathbf{v}_j\right)} = \frac{\sum_{i=1}^{r+1} c_i^2 \lambda_i}{\sum_{i=1}^{r+1} c_i^2} \leq \frac{\left(\sum_{i=1}^{r+1} c_i^2\right)\lambda_{r+1}}{\sum_{i=1}^{r+1} c_i^2} = \lambda_{r+1} \tag{5.49}$$

But Eq. (5.46) implies that

$$\tilde{\lambda}_r(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r) \leq \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \qquad \mathbf{v}^T \mathbf{w}_i = 0, \qquad i = 1, 2, \ldots, r \tag{5.50}$$

so that, comparing inequalities (5.49) and (5.50), we can write

$$\tilde{\lambda}_r(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r) \leq \lambda_{r+1} \tag{5.51}$$

Inequality (5.51) can be interpreted as stating that

$$\lambda_{r+1} = \max \tilde{\lambda}_r(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r) \tag{5.52}$$

which, according to Eq. (5.47), occurs when $\mathbf{w}_i = \mathbf{v}_i$ $(i = 1, 2, \ldots, r)$. Hence, introducing Eq. (5.46) into Eq. (5.52), we conclude that

$$\lambda_{r+1} = \max_{\mathbf{w}} \left(\min_{\mathbf{v}} \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \right), \qquad \mathbf{v}^T \mathbf{w}_i = 0, \qquad i = 1, 2, \ldots, r \tag{5.53}$$

which can be stated as the theorem: *The eigenvalue λ_{r+1} of a real symmetric positive definite matrix A is the maximum value that can be given to* min $\left(\mathbf{v}^T A\mathbf{v}/\mathbf{v}^T\mathbf{v}\right)$ *by the imposition of the r constraints* $\mathbf{v}^T\mathbf{w}_i = 0$ $(i = 1, 2, \ldots, r)$, where the maximum is with respect to all sets containing $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ and the minimum is with respect to all vectors $\mathbf{v}$ satisfying the imposed constraints. According to Courant and Hilbert (Ref. 2), the maximum-minimum character of the eigenvalues was first mentioned by Fischer (Ref. 3) in connection with quadratic forms with real coefficients. Weyl (Ref. 13) and Courant (Ref. 1) applied the theorem to the theory of vibrations. The theorem is known as the *Courant-Fischer maximin theorem.* In the case in which the eigenvalues are arranged in descending order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, "max" and "min" reverse order in Eq. (5.53), in which case the theorem is called the *Courant-Fischer minimax theorem.*

5.4 SEPARATION THEOREM FOR NATURAL SYSTEMS

The Courant-Fischer maximin theorem characterizes the eigenvalues of a real symmetric positive definite matrix subjected to given constraints. Although it can be used to estimate higher eigenvalues, the maximin theorem is not really a computational tool. Our interest in the theorem lies in the fact that it facilitates the development of a theorem defining the manner in which the eigenvalues of the mathematical model of a given system behave as the number of degrees of freedom of the model is increased or decreased.

We consider an n-degree-of-freedom natural system whose eigenvalue problem is defined by the $n \times n$ real symmetric positive definite matrix A and confine ourselves to the c'se in which a reduction in the number of degrees of freedom by one is equivalent to removing one row and the corresponding column from A, thus obtaining an $(n-1) \times (n-1)$ real symmetric positive definite matrix A'. It is immaterial which row and column are removed, but for the sake of this discussion we assume that they are the last row and column. Hence, the matrices A and A' are such that

$$A = \begin{bmatrix} \begin{matrix} & A' & \\ & & \end{matrix} & \begin{matrix} \times \\ \times \end{matrix} \\ \times \quad \times & \times \end{bmatrix} \tag{5.54}$$

where the removed entries are indicated by one row and one column of $\times$'s. Matrices having the structure described by Eq. (5.54) are said to possess the *embedding property.* We denote the eigenvalues of A by $\lambda_1, \lambda_2 \ldots, \lambda_n$ and the eigenvalues of A' by $\lambda'_1, \lambda'_2, \ldots, \lambda'_{n-1}$ and pose the question as to how the two sets of eigenvalues relate to one another. To answer this question, we consider an arbitrary n-vector $\mathbf{v} = [v_1\ v_2\ \ldots\ v_n]^T$ and a reduced $(n-1)$-vector $\mathbf{v}' = [v_1\ v_2\ \ldots\ v_{n-1}]^T$, obtained from $\mathbf{v}$ by removing the last component. Regarding the system associated with A' as unconstrained, we can use Rayleigh's principle, Eq. (5.26), and write

$$\lambda_1' = \min \frac{\mathbf{v}'^T A' \mathbf{v}'}{\mathbf{v}'^T \mathbf{v}'} \tag{5.55}$$

But, Rayleigh's quotient for the system defined by A' is equal to Rayleigh's quotient for the system defined by A, provided the trial vector $\mathbf{v}$ satisfies the equation

$$\mathbf{v}^T \mathbf{e}_n = 0 \tag{5.56}$$

where $\mathbf{e}_n = [0\ 0\ \ldots 0\ 1]^T$ is the nth standard unit vector. Hence, we can write

$$\frac{\mathbf{v}'^T A' \mathbf{v}'}{\mathbf{v}'^T \mathbf{v}'} = \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \qquad \mathbf{v}^T \mathbf{e}_n = 0 \tag{5.57}$$

Equation (5.56) can be regarded as a constraint equation imposed on the original system defined by A, so that Eqs. (5.55) and (5.57) can be combined into

$$\lambda_1' = \tilde{\lambda}_1(\mathbf{e}_n) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \qquad \mathbf{v}^T \mathbf{e}_n = 0 \tag{5.58}$$

Then, using Rayleigh's theorem for systems with one constraint, inequalities (5.44), we conclude that

$$\lambda_1 \leq \lambda_1' \leq \lambda_2 \tag{5.59}$$

Next, we assume that the trial vector $\mathbf{v}$ is subjected to $r-1$ constraints defined by $\mathbf{v}^T \mathbf{w}_i = 0$ $(i = 1, 2, \ldots, r-1;\ r < n)$, where $\mathbf{w}_i$ are linearly independent n-vectors, and introduce the notation

$$\tilde{\lambda}_{r-1}(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{r-1}) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \quad \mathbf{v}^T \mathbf{w}_i = 0, \quad i = 1, 2, \ldots, r-1;\ r < n \tag{5.60}$$

Moreover, we assume that, in addition to the constraints $\mathbf{v}^T \mathbf{w}_i = 0$, the vector $\mathbf{v}$ is subjected to the constraint $\mathbf{v}^T \mathbf{e}_n = 0$ and define the eigenvalues of the system thus constrained by

$$\tilde{\tilde{\lambda}}_{r-1}(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{r-1}) = \min \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}}, \quad \mathbf{v}^T \mathbf{w}_i = 0,$$

$$i = 1, 2, \ldots, r-1;\ r < n, \qquad \mathbf{v}^T \mathbf{e}_n = 0 \tag{5.61}$$

Because in the latter case the system is subjected to one additional constraint than in the former, we can write

$$\tilde{\lambda}_{r-1} \leq \tilde{\tilde{\lambda}}_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.62}$$

At this point, we introduce the truncated $(n-1)$-dimensional constraint vectors $\mathbf{w}_i' = [w_{1i}\ w_{2i}\ \ldots w_{n-1,i}]^T$, so that the constraints $\mathbf{v}^T \mathbf{w}_i = 0$ $(i = 1, 2, \ldots, r-1; r <$

n), $\mathbf{v}^T\mathbf{e}_n = 0$ are equivalent to the constraints $\mathbf{v}'^T\mathbf{w}'_i = 0\ (i = 1, 2, \ldots, r-1;\ r < n)$. Hence, the definition given by Eq. (5.61) is equivalent to the definition

$$\tilde{\lambda}'_{r-1}\left(\mathbf{w}'_1, \mathbf{w}'_2, \ldots, \mathbf{w}'_{r-1}\right) = \min \frac{\mathbf{v}'^T A'\mathbf{v}'}{\mathbf{v}'^T\mathbf{v}'},\ \mathbf{v}'^T\mathbf{w}'_i = 0,$$

$$i = 1, 2, \ldots, r-1;\ r < n \tag{5.63}$$

from which it follows that

$$\tilde{\lambda}'_{r-1} = \tilde{\tilde{\lambda}}_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.64}$$

so that inequalities (5.62) can be replaced by

$$\tilde{\lambda}_{r-1} \leq \tilde{\lambda}'_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.65}$$

But, considering Eq. (5.52), we can write

$$\lambda_r = \max \tilde{\lambda}_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.66}$$

as well as

$$\lambda'_r = \max \tilde{\lambda}'_{r-1} = \max \tilde{\tilde{\lambda}}_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.67}$$

Moreover, inequalities (5.65) can be regarded as implying that

$$\max \tilde{\lambda}_{r-1} \leq \max \tilde{\lambda}'_{r-1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.68}$$

so that, considering Eqs. (5.66) and (5.67), we conclude that

$$\lambda_r \leq \lambda'_r, \qquad r = 2, 3, \ldots, n-1 \tag{5.69}$$

Inequalities (5.69) state that the addition of the constraint $\mathbf{v}^T\mathbf{e}_n = 0$ tends to raise the eigenvalues, a result that agrees with the intuition.

Inequalities (5.69) present a one-sided picture, however, as they only relate λ'_r to λ_r ($r = 2, 3, \ldots, n-1$). To complete the picture, we must have a relation between λ'_r and λ_{r+1} ($r = 2, 3, \ldots, n-1$). To this end, we combine Eqs. (5.61) and (5.67) and write

$$\lambda'_r = \max_{\mathbf{w}}\left(\min_{\mathbf{v}} \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\right), \qquad \mathbf{v}^T\mathbf{w}_i = 0,$$

$$i = 1, 2, \ldots, r-1;\ r < n, \qquad \mathbf{v}^T\mathbf{e}_n = 0 \tag{5.70}$$

On the other hand, Eq. (5.53) states that

$$\lambda_{r+1} = \max_{\mathbf{w}}\left(\min_{\mathbf{v}} \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^T\mathbf{v}}\right), \quad \mathbf{v}^T\mathbf{w}_i = 0, \quad i = 1, 2, \ldots, r;\ r < n \tag{5.71}$$

We observe from Eqs. (5.70) and (5.71) that in both cases the system is subjected to r constraints and that the first $r-1$ of these constraints are the same. But, whereas in Eq. (5.70) the rth constraint involves $\mathbf{e}_n$, a given vector, in Eq. (5.71) it involves $\mathbf{w}_r$, an arbitrary vector that can be chosen so as to maximize λ_{r+1}. It follows that there are many more choices in maximizing min $\left(\mathbf{v}^T A\mathbf{v}/\mathbf{v}^T\mathbf{v}\right)$ in Eq. (5.71) than in Eq. (5.70), from which we conclude that

$$\lambda_r' \leq \lambda_{r+1}, \qquad r = 2, 3, \ldots, n-1 \tag{5.72}$$

Hence, combining inequalities (5.59), (5.69) and (5.72), we obtain

$$\lambda_1 \leq \lambda_1' \leq \lambda_2 \leq \lambda_2' \leq \ldots \leq \lambda_{n-1} \leq \lambda_{n-1}' \leq \lambda_n \tag{5.73}$$

We refer to inequalities (5.73) as the *separation theorem*, which can be stated as follows: *The eigenvalues $\lambda_1', \lambda_2', \ldots, \lambda_{n-1}'$ of the $(n-1) \times (n-1)$ real symmetric positive definite matrix A', obtained by striking out one row and the corresponding column from the $n \times n$ real symmetric positive definite matrix A, separate the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of A.* The separation theorem is known under a variety of names. Indeed, whereas Wilkinson (Ref. 14) refers to it as the separation theorem, Golub and Van Loan (Ref. 5) call it the *interlacing property* and Strang (Ref. 12) uses the term *intertwining of eigenvalues.* Using a slightly different interpretation of the theorem, Franklin (Ref. 4) and Meirovitch (Ref. 7) refer to it as the *inclusion principle.*

The separation theorem was proved for the first time by Rayleigh himself (Ref. 10, Vol. 1, Sec. 92a) for the special case of a two-degree-of-freedom system. The result was presented in Sec. 5.3 of this text as Rayleigh's theorem for systems with one constraint in the form of inequalities (5.44). Ironically, Rayleigh carried out his proof by an approach based on Lagrange's equations, without using Rayleigh's quotient. Using an approach similar to Rayleigh's, Routh (Ref. 11, Sec. 78) proved the theorem for an n-degree-of-freedom system.

The proof of the separation theorem, inequalities (5.73), was carried out for a Rayleigh's quotient in terms of a single real symmetric positive definite $n \times n$ matrix A and the corresponding $(n-1) \times (n-1)$ matrix A', where the matrices possess the embedding property given by Eq. (5.54). However, the separation theorem is equally valid for a Rayleigh's quotient in terms of two real symmetric positive definite $n \times n$ mass and stiffness matrices M and K, respectively, and the associated $(n-1) \times (n-1)$ matrices M' and K', provided M and M' on the one hand and K and K' on the other hand possess the embedding property (Ref. 8). To show this, we consider the eigenvalue problems

$$K'\mathbf{u} = \lambda' M'\mathbf{u}, \qquad K\mathbf{u} = \lambda M\mathbf{u} \tag{5.74a, b}$$

where the mass and stiffness matrices are related by

$$M = \begin{bmatrix} M' & \mathbf{m} \\ \mathbf{m}^T & m \end{bmatrix}, \qquad K = \begin{bmatrix} K' & \mathbf{k} \\ \mathbf{k}^T & k \end{bmatrix} \tag{5.75a, b}$$

where $\mathbf{m} = [m_{n1} \; m_{n2} \; \ldots \; m_{n,n-1}]^T$ and $\mathbf{k} = [k_{n1} \; k_{n2} \; \ldots \; k_{n,n-1}]^T$ are $(n-1)$-vectors and $m = m_{nn}$ and $k = k_{nn}$ are scalars. Because M' is real symmetric and positive definite, it has the Cholesky decomposition (see Sec. 6.2)

$$M' = L'L'^T \tag{5.76}$$

where L' is a nonsingular lower triangular matrix. Then, using the linear transformation

$$L'^T\mathbf{u} = \mathbf{v}, \qquad \mathbf{u} = (L'^T)^{-1}\mathbf{v} \tag{5.77a, b}$$

Eq. (5.74a) can be reduced to the standard form

$$A'\mathbf{v} = \lambda'\mathbf{v} \tag{5.78}$$

in which

$$A' = (L')^{-1}K'(L'^T)^{-1} \tag{5.79}$$

is a real symmetric matrix of order $n - 1$. Using the same process, we can first carry out the decomposition $M = LL^T$ and then reduce the eigenvalue problem (5.74b) to the standard form

$$A\mathbf{v} = \lambda\mathbf{v} \tag{5.80}$$

where

$$A = L^{-1}K(L^T)^{-1} \tag{5.81}$$

is a real symmetric matrix of order n. But, the matrix L can be expressed in the form

$$L = \begin{bmatrix} L' & \mathbf{0} \\ \mathbf{l}^T & l \end{bmatrix} \tag{5.82}$$

in which $\mathbf{l}$ is an $(n - 1)$-vector and l is a scalar given by

$$\mathbf{l} = (L')^{-1}\mathbf{m}, \qquad l = m - \mathbf{m}^T(L'^T)^{-1}\mathbf{m} \tag{5.83a, b}$$

It is not difficult to show that

$$L^{-1} = \begin{bmatrix} (L')^{-1} & \mathbf{0} \\ -\mathbf{l}^T(L')^{-1}/l & \mathbf{1}/l \end{bmatrix} \tag{5.84}$$

so that, inserting Eq. (5:84) into Eq. (5.81), we obtain (Ref. 8)

$$A = \begin{bmatrix} A' & (L')^{-1}(\mathbf{k} - (L')^{-1}K'(L'^T)^{-1}\mathbf{l})/l \\ \text{symm} & (\mathbf{l}^T(L')^{-1}K'(L'^T)^{-1}\mathbf{l} - 2\mathbf{k}^T(L'^T)^{-1}\mathbf{l} + k)/l^2 \end{bmatrix} \tag{5.85}$$

from which we conclude that the matrices A' and A, Eqs. (5.79) and (5.81), possess the embedding property. It follows that the eigenvalues $\lambda'_1, \lambda'_2, \ldots, \lambda'_{n-1}$ and $\lambda_1, \lambda_2, \ldots, \lambda_n$ of the eigenvalue problems (5.74a) and (5.74b), respectively, satisfy the separation theorem, inequalities (5.73).

The separation theorem has significant implications in approximate solutions to the eigenvalue problem for self-adjoint distributed-parameter systems. Indeed, the convergence of the Rayleigh-Ritz method for producing such approximate solutions can be demonstrated on the basis of the separation theorem. The Rayleigh-Ritz method is discussed later in this text.

Example 5.3

Verify that the eigenvalues of the 3×3 matrix A of Example 5.2 and the eigenvalues of the 2×2 matrix A' obtained by removing one row and the corresponding column from the matrix A satisfy the separation theorem, and that this is true regardless of the row and column removed.

From Example 5.2, the matrix A is

$$A = \begin{bmatrix} 2.5 & -1 & 0 \\ -1 & 5 & -\sqrt{2} \\ 0 & -\sqrt{2} & 10 \end{bmatrix} \tag{a}$$

and has the eigenvalues

$$\lambda_1 = 2.119322, \qquad \lambda_2 = 5, \qquad \lambda_3 = 10.380678 \tag{b}$$

Removing the third row and column from A, we obtain

$$A' = \begin{bmatrix} 2.5 & -1 \\ -1 & 5 \end{bmatrix} \tag{c}$$

which has the eigenvalues

$$\lambda_1' = 2.149219, \qquad \lambda_2' = 5.350781 \tag{d}$$

The two sets of eigenvalues given by Eqs. (b) and (d) clearly satisfy the separation theorem, inequalities (5.73) with $n = 3$. If the second row and column are removed from A, the matrix A' is the trivial one

$$A' = \begin{bmatrix} 2.5 & 0 \\ 0 & 10 \end{bmatrix} \tag{e}$$

with the obvious eigenvalues

$$\lambda_1' = 2.5, \qquad \lambda_2' = 10 \tag{f}$$

The eigenvalues given by Eqs. (b) and (f) satisfy the separation theorem as well. Finally, striking out the first row and column from A, we have

$$A' = \begin{bmatrix} 5 & -\sqrt{2} \\ -\sqrt{2} & 10 \end{bmatrix} \tag{g}$$

which has the eigenvalues

$$\lambda_1' = 4.627719, \qquad \lambda_2' = 10.372281 \tag{h}$$

Examining the two sets of eigenvalues, Eqs. (b) and (h), we conclude that once again the separation theorem holds true. This verifies that the separation theorem is satisfied independently of the row and column removed.

5.5 SEPARATION THEOREM FOR GYROSCOPIC SYSTEMS

The fact that the eigenvalue problem for gyroscopic systems can be reduced to the standard form given by Eq. (4.160), in which the coefficient matrix A is real symmetric and positive definite, permits a characterization of the eigenvalues similar to that for natural systems. Indeed, it is shown in Ref. 9 that

$$\lambda_{2r+1} = \lambda_{2r+2} = \max_{\mathbf{w}} \left(\min_{\mathbf{v}} \frac{\mathbf{v}^T A \mathbf{v}}{\mathbf{v}^T \mathbf{v}} \right), \quad \mathbf{v}^T \mathbf{w}_{2i-1} = 0, \quad \mathbf{v}^T \mathbf{w}_{2i} = 0, \quad i = 1, 2, \ldots, r \tag{5.86}$$

where $\mathbf{w}_{2i-1}$ and $\mathbf{w}_{2i}$ $(i = 1, 2, \ldots, r)$ are r independent pairs of mutually orthogonal $2n$-vectors. Equation (5.86) represents a *maximin theorem for gyroscopic systems* and can be stated as follows: *The pair of repeated eigenvalues* λ_{2r+1}, $\lambda_{2r+2} = \lambda_{2r+1}$ *of a gyroscopic system is the maximum value that can be given to* $\min\left(\mathbf{v}^T A\mathbf{v}/\mathbf{v}^T\mathbf{v}\right)$ *by the imposition of the constraints* $\mathbf{v}^T\mathbf{w}_{2i-1} = 0$, $\mathbf{v}^T\mathbf{w}_{2i} = 0$ $(i = 1, 2, \ldots, r)$, where the maximum is with respect to all sets containing $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{2r}$ and the minimum is with respect to all vectors $\mathbf{v}$ satisfying the imposed constraints.

Next, we consider the case in which the number of degrees of freedom of the gyroscopic system described by Eq. (4.139) is reduced from n to $n-1$. Consistent with this, the eigenvalue problem can be expressed as

$$A'\mathbf{v}' = \lambda'\mathbf{v}' \tag{5.87}$$

where, by analogy with Eq. (5.54), A' is a $2(n-1) \times 2(n-1)$ real symmetric positive definite matrix obtained by striking out two rows and the corresponding two columns from A, so that A and A' are related by

$$A = \overset{A'}{\left[\begin{array}{cccccc} & \downarrow & \times & & \mid & \times \\ & & \times & & & \times \\ \times & \times & \times & & \downarrow & \times \\ & & & & & \times \\ & & & & & \times \\ \times & \times & \times & \times & \times & \times \end{array}\right]} \tag{5.88}$$

Moreover, $\mathbf{v}'$ is a $2(n-1)$-vector related to the $2n$-vector $\mathbf{v}$ by

$$\mathbf{v}' = \begin{bmatrix} v_1 & v_2 & \ldots & v_{n-1} & v_{n+1} & \ldots & v_{2n-1} \end{bmatrix}^T .$$

We assume that the eigenvalues of A are such that $\lambda_1 = \lambda_2 \le \lambda_3 = \lambda_4 \le \ldots \le \lambda_{2n-1} = \lambda_{2n}$ and those of A' satisfy $\lambda_1' = \lambda_2' \le \lambda_3' = \lambda_4' \le \ldots \le \lambda_{2n-3}' = \lambda_{2n-2}'$. Then, based on the maximin theorem for gyroscopic systems stated above, it is proved in Ref. 9 that the two sets of eigenvalues are related by

$$\begin{aligned} \lambda_1 &= \lambda_2 \le \lambda_1' = \lambda_2' \le \lambda_3 = \lambda_4 \le \lambda_3' = \lambda_4' \le \ldots \\ &\le \lambda_{2n-3} = \lambda_{2n-2} \le \lambda_{2n-3}' = \lambda_{2n-2}' \le \lambda_{2n-1} = \lambda_{2n} \end{aligned} \tag{5.89}$$

Inequalities (5.89) represent the *separation theorem for gyroscopic systems* and can be stated as follows: *The eigenvalues* $\lambda_1', \lambda_2' = \lambda_1', \lambda_3', \lambda_4' = \lambda_3', \ldots, \lambda_{2n-3}', \lambda_{2n-2}' = \lambda_{2n-3}'$ *of the* $2(n-1) \times 2(n-1)$ *real symmetric positive definite matrix* A' *defining an* $(n-1)$*-degree-of-freedom model of a gyroscopic system, obtained by striking out the* n*th and* $2n$*th rows and columns from the* $2n \times 2n$ *real symmetric matrix* A *defining an* n*-degree-of-freedom model of the same gyroscopic system, separate the eigenvalues* $\lambda_1, \lambda_2 = \lambda_1, \lambda_3, \lambda_4 = \lambda_3, \ldots, \lambda_{2n-1}, \lambda_{2n} = \lambda_{2n-1}$ *of* A.

As for natural systems, the separation theorem for gyroscopic systems not only characterizes approximate eigenvalues of distributed-parameter systems computed by means of the Rayleigh-Ritz method but can also be used to prove convergence of these approximate eigenvalues to the actual ones. An illustration of the separation theorem for gyroscopic systems is provided in Ref. 9.

5.6 GERSCHGORIN'S THEOREMS

Under certain circumstances, Gerschgorin's theorems permit the estimation of the location in the complex plane of the eigenvalues of an arbitrary square matrix A. In the case of a real symmetric matrix A, the complex plane projects onto the real axis.

Let us consider an $m \times m$ matrix A, where m is odd or even, and write the eigenvalue problem in the index notation

$$\sum_{j=1}^{m} a_{ij} x_j = \lambda x_i, \qquad i = 1, 2, \ldots, m \tag{5.90}$$

Then, assuming that x_k is the component of the vector $\mathbf{x}$ with the largest modulus, $|x_k| = \max |x_j|$ $(j = 1, 2, \ldots, m)$, we let $i = k$ in Eqs. (5.90) and write

$$(\lambda - a_{kk})\, x_k = \sum_{\substack{j=1 \\ j \neq k}}^{m} a_{kj} x_j \tag{5.91}$$

But

$$|\lambda - a_{kk}| \cdot |x_k| \leq \sum_{\substack{j=1 \\ j \neq k}}^{m} |a_{kj}| \cdot |x_j| \leq |x_k| \sum_{\substack{j=1 \\ j \neq k}}^{m} |a_{kj}| \tag{5.92}$$

so that, dividing through by $|x_k|$, we obtain

$$|\lambda - a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq k}}^{m} |a_{kj}| \tag{5.93}$$

Next, we introduce the notation

$$r_k = \sum_{\substack{j=1 \\ j \neq k}}^{m} |a_{kj}| \tag{5.94}$$

and rewrite inequality (5.93) as

$$|\lambda - a_{kk}| \leq r_k \tag{5.95}$$

First, we observe that $|\lambda - a_{kk}|$ represents the distance from the point a_{kk} in the complex plane to the eigenvalue λ, so that inequality (5.95) can be interpreted geometrically as defining a circular region with the center at a_{kk} and with the radius equal to r_k, as shown in Fig. 5.8. Then, recognizing that Eqs. (5.90) admits m solutions, we let $k = 1, 2, \ldots, m$ and express inequality (5.95) in the form of the theorem: *Every eigenvalue of the matrix A lies in at least one of the circular disks with centers at a_{kk} and radii r_k*. The theorem is known as *Gerschgorin's first theorem* and the disks are sometimes referred to as *Gerschgorin's disks.*

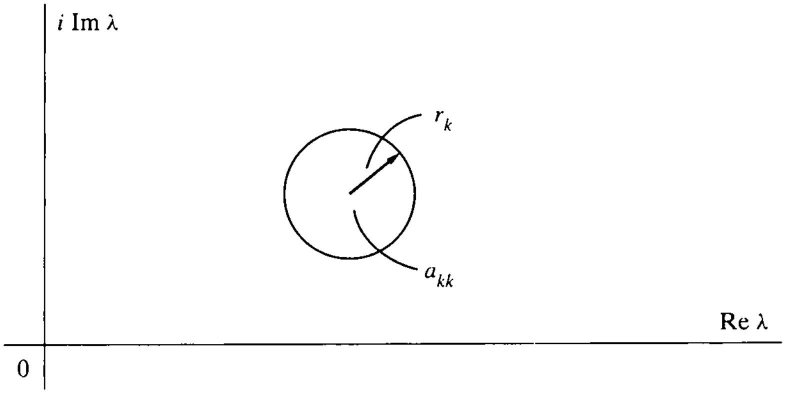


Figure 5.8 Gerschgorin disk in λ-plane

From Gerschgorin's first theorem, it is possible to conclude that a given eigenvalue can lie in more than one disk. A second theorem by Gerschgorin is concerned with the distribution of the eigenvalues among the disks. Gerschgorin's second theorem is based on the theorem on continuity (Ref. 4), which states: *The eigenvalues of a matrix A are continuous functions of the elements of A.* Any matrix A can always be written as the sum of two matrices of the form

$$A = D + O \tag{5.96}$$

where $D = \text{diag}\,(a_{kk})$ is the diagonal matrix obtained from A by omitting its off-diagonal entries and O is the matrix of the off-diagonal entries of A. Consistent with this, we can introduce the matrix

$$A\,(\epsilon) = D + \epsilon O \tag{5.97}$$

where ϵ is a parameter satisfying $0 \leq \epsilon \leq 1$. For $\epsilon = 0$, we have $A(0) = D = \text{diag}\,(a_{ii})$, and for $\epsilon = 1$, we have $A\,(1) = A$. The coefficients of the characteristic polynomial of $A\,(\epsilon)$ are polynomials in ϵ and, by continuity, they are continuous functions of ϵ. In view of Gerschgorin's first theorem, we conclude that the eigenvalues of $A\,(\epsilon)$ lie in the circular disks with centers at a_{kk} and with radii

$$\epsilon r_k = \sum_{\substack{j=1 \\ j\neq k}}^{m} \epsilon \left|a_{kj}\right| .$$

As ϵ varies from 1 to 0, the m eigenvalues of $A\,(\epsilon)$ move continuously from the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$ of $A\,(1)$ to $a_{11}, a_{22}, \ldots, a_{mm}$, respectively.

Next, we consider the case in which ℓ of the disks corresponding to A are disjoint and the remaining $m - \ell$ disks are connected. For the sake of this discussion, and without loss of generality, we assume that the first ℓ disks of A are disjoint. But, because the ℓ disks of A of radii $r_1, r_2, \ldots, r_\ell$ are disjoint, it follows that the first ℓ disks of $A(\epsilon)$ of radii $\epsilon r_1, \epsilon r_2, \ldots, \epsilon r_\ell$ are disjoint. Indeed, as ϵ decreases from 1 to 0, the first ℓ Gerschgorin's disks shrink to the points $\lambda = a_{kk}$ $(k = 1, 2, \ldots, \ell)$. In the process, the eigenvalues contained in these disks remain inside them. Assuming that the first ℓ eigenvalues of A are distinct, we conclude that each disjoint disk contains exactly one eigenvalue. We then conclude that the remaining $m - \ell$ eigenvalues must lie in the connected domain representing the union of the corresponding $m - \ell$ disks. These conclusions form the basis of *Gerschgorin's second theorem*: *If ℓ of the Gerschgorin's disks corresponding to the matrix A are disjoint and the remaining $m - \ell$ disks form a connected domain, which is isolated from the first ℓ disks, then there are exactly $m - \ell$ eigenvalues of A contained in the connected domain.*

For Gerschgorin's theorem to be useful, the radii of the disks should be relatively small. This, in turn, implies that the off-diagonal entries of A should be relatively small. Unfortunately, this rules out the use of Gerschgorin's theorems for nonconservative systems, in which case the matrix A has the form given by Eq. (4.168). Clearly, this form does not lend itself to meaningful application of Gerschgorin's theorems. Indeed, one half of the state equations represent mere kinematical identities, having no bearing on the eigenvalues, and the other half are not diagonally dominant. The situation is considerably better in the case of conservative systems, natural or nonnatural, as in this case the matrix A is real and symmetric, a form more conducive to meaningful results. Indeed, owing to the fact that the eigenvalues of a real symmetric matrix are real, the disks collapse into segments on the real axis with centers at a_{kk} and of length $2r_k$ $(k = 1, 2, \ldots, m)$. Note that Gerschgorin's first theorem is used in Sec. 6.8 to locate approximately an eigenvalue in conjunction with Givens' method for computing the eigenvalues of a symmetric tridiagonal matrix.

Example 5.4

Verify the applicability of Gerschgorin's theorems to the matrix A of Example 5.2.

From Example 5.2, the matrix A is

$$A = \begin{bmatrix} 2.5 & -1 & 0 \\ -1 & 5 & -\sqrt{2} \\ 0 & -\sqrt{2} & 10 \end{bmatrix} \tag{a}$$

Hence, because the matrix is real and symmetric, Gerschgorin's disks collapse into segments on the real axis. Nevertheless, we will continue to refer to them as disks.

There are three Gerschgorin disks with the centers given by

$$a_{11} = 2.5, \qquad a_{22} = 5, \qquad a_{33} = 10 \tag{b}$$

and with the radii, defined by Eq. (5.94), or

$$\begin{aligned} r_1 &= |a_{12}| + |a_{13}| = |-1| + 0 = 1 \\ r_2 &= |a_{21}| + |a_{23}| = |-1| + |-\sqrt{2}| = 1 + \sqrt{2} \\ r_3 &= |a_{31}| + |a_{32}| = 0 + |-\sqrt{2}| = \sqrt{2} \end{aligned} \tag{c}$$

The collapsed disks, i.e., the segments, are shown on the λ-axis in Fig. 5.9. The actual eigenvalues

$$\lambda_1 = 2.119322, \qquad \lambda_2 = 5, \qquad \lambda_3 = 10.380678 \tag{d}$$

marked by circles in Fig. 5.9, clearly fall within these segments, thus verifying Gerschgorin's first theorem. Moreover, the first two disks intersect and there are two eigenvalues inside the connected domain, thus satisfying Gerschgorin's second theorem.

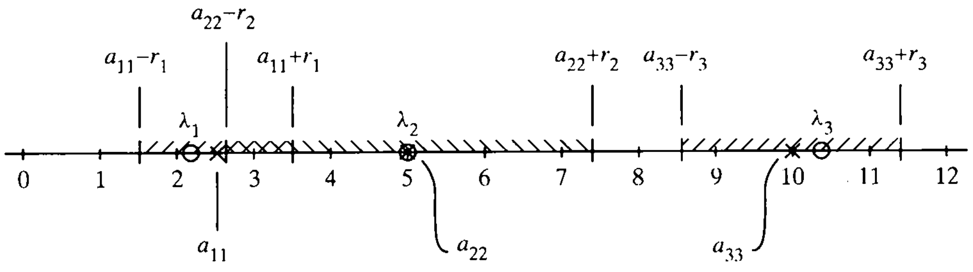


Figure 5.9 Gerschgorin disks collapsed into segments

5.7 PERTURBATION OF THE EIGENVALUE PROBLEM

A problem of great interest in vibrations, as well as in related areas such as controls, is how the dynamic characteristics change as a result of changes in the system parameters. The parameter changes may be due to design modifications or to improved knowledge of the system. Regardless of the reasons, the net effect is that the matrix A is different from the original one. Then, the question arises as to how these changes affect the eigenvalues and eigenvectors. The changes in the eigenvalues and eigenvectors can always be determined by solving the eigenvalue problem associated with the new matrix A. This may not be necessary when the new matrix A does not differ very much from the original one. Indeed, based on the assumption that the change in the matrix A represents a small quantity of first order, in this section we develop a perturbation method permitting the computation of the changes in the eigensolution in terms of the original eigensolution and the change in the matrix A.

We consider the case in which the original system is defined by the $n \times n$ real arbitrary matrix A_0 and denote by λ_{0i}, $\mathbf{x}_{0i}$ and $\mathbf{y}_{0i}$ $(i = 1, 2, \dots, n)$ the associated eigenvalues, right eigenvectors and left eigenvectors, respectively. They satisfy the two eigenvalue problems

$$A_0\mathbf{x}_{0i} = \lambda_{0i}\mathbf{x}_{0i}, \qquad i = 1, 2, \dots, n \tag{5.98a}$$

$$\mathbf{y}_{0i}^T A_0 = \lambda_{0i}\mathbf{y}_{0i}^T, \qquad i = 1, 2, \dots, n \tag{5.98b}$$

We confine ourselves to the case in which *all the eigenvalues are distinct*. As shown in Sec. 4.8, the right and left eigenvectors are biorthogonal and can be normalized so as to satisfy

$$\mathbf{y}_{0j}^T\mathbf{x}_{0i} = \delta_{ij}, \qquad \mathbf{y}_{0j}^T A_0\mathbf{x}_{0i} = \lambda_{0i}\delta_{ij}, \qquad i, j = 1, 2, \dots, n \tag{5.99a, b}$$

where δ_{ij} is the Kronecker delta. Next, we define the new system by the $n \times n$ real arbitrary matrix A and assume that the two matrices are related by

$$A = A_0 + A_1 \tag{5.100}$$

where A_1 is an $n \times n$ real arbitrary matrix. We consider the case in which A_1 is "small" relative to A_0, in the sense that the entries of A_1 are small quantities of first order compared to the entries of A_0. Consistent with this, we refer to A_1 as a first-order *perturbation matrix*, to A_0 as the *unperturbed matrix* and to A as the *perturbed matrix*. By analogy with Eqs. (5.98), the perturbed eigenvalue problems have the form

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \qquad i = 1, 2, \ldots, n \tag{5.101a}$$

$$\mathbf{y}_i^T A = \lambda_i \mathbf{y}_i^T, \qquad i = 1, 2, \ldots, n \tag{5.101b}$$

where λ_i, $\mathbf{x}_i$ and $\mathbf{y}_i$ $(i = 1, 2, \ldots, n)$ are the perturbed eigenvalues, right eigenvectors and left eigenvectors, respectively. We assume that the perturbed eigenvalues are distinct and that the perturbed eigenvectors satisfy the biorthonormality relations

$$\mathbf{y}_j^T \mathbf{x}_i = \delta_{ij}, \qquad \mathbf{y}_j^T A \mathbf{x}_i = \lambda_i \delta_{ij}, \qquad i, j = 1, 2, \ldots, n \tag{5.102a, b}$$

At this point, we consider a formal *first-order perturbation solution* of the eigenvalue problems (5.101), with A being given by Eq. (5.100), in the form

$$\lambda_i = \lambda_{0i} + \lambda_{1i}, \qquad \mathbf{x}_i = \mathbf{x}_{0i} + \mathbf{x}_{1i}, \qquad \mathbf{y}_i = \mathbf{y}_{0i} + \mathbf{y}_{1i}, \qquad i = 1, 2, \ldots, n \tag{5.103a, b, c}$$

where λ_{1i}, $\mathbf{x}_{1i}$ and $\mathbf{y}_{1i}$ $(i = 1, 2, \ldots, n)$ are first-order perturbations in the eigenvalues, right eigenvectors and left eigenvectors, respectively. Inserting Eqs. (5.103) into Eqs. (5.101), considering Eqs. (5.98) and ignoring second-order terms in the perturbations, we obtain

$$A_0 \mathbf{x}_{1i} + A_1 \mathbf{x}_{0i} = \lambda_{0i} \mathbf{x}_{1i} + \lambda_{1i} \mathbf{x}_{0i}, \qquad i = 1, 2, \ldots, n \tag{5.104a}$$

$$\mathbf{y}_{0i}^T A_1 + \mathbf{y}_{1i}^T A_0 = \lambda_{0i} \mathbf{y}_{1i}^T + \lambda_{1i} \mathbf{y}_{0i}^T, \qquad i = 1, 2, \ldots, n \tag{5.104b}$$

Hence, the problem has been reduced to the determination of the perturbations $\lambda_{1i}, \mathbf{x}_{1i}, \mathbf{y}_{1i}$ on the assumption that A_0, A_1, λ_{0i}, $\mathbf{x}_{0i}$ and $\mathbf{y}_{0i}$ are all known. Clearly, the objective is to carry out this process without solving any new eigenvalue problems. To this end, we concentrate first on Eq. (5.104a), recognize that the unperturbed right eigenvectors $\mathbf{x}_{0i}$ $(i = 1, 2, \ldots, n)$ can be used as a basis for an n-dimensional vector space and expand the perturbations in the right eigenvectors as follows:

$$\mathbf{x}_{1i} = \sum_{k=1}^{n} \epsilon_{ik} \mathbf{x}_{0k}, \qquad i = 1, 2, \ldots, n \tag{5.105}$$

where ϵ_{ik} are small quantities of first order. But, considering Eqs. (5.99a) and (5.102a) and ignoring second-order quantities, we obtain

$$\begin{aligned} \mathbf{y}_i^T \mathbf{x}_i &= \left(\mathbf{y}_{0i}^T + \mathbf{y}_{1i}^T\right)(\mathbf{x}_{0i} + \mathbf{x}_{1i}) \\ &\cong \mathbf{y}_{0i}^T \mathbf{x}_{0i} + \mathbf{y}_{0i}^T \mathbf{x}_{1i} + \mathbf{y}_{1i}^T \mathbf{x}_{0i} = 1 + \mathbf{y}_{0i}^T \mathbf{x}_{1i} + \mathbf{y}_{1i}^T \mathbf{x}_{0i} = 1 \end{aligned} \tag{5.106}$$

which permits us to write

$$\mathbf{y}_{0i}^T \mathbf{x}_{1i} = 0, \qquad \mathbf{x}_{0i}^T \mathbf{y}_{1i} = 0, \qquad i = 1, 2, \ldots, n \tag{5.107a, b}$$

Introducing Eq. (5.105) into Eq. (5.107a) and considering once again Eq. (5.99a), we have

$$\mathbf{y}_{0i}^T \sum_{k=1}^{n} \epsilon_{ik} \mathbf{x}_{0k} = \sum_{k=1}^{n} \epsilon_{ik} \mathbf{y}_{0i}^T \mathbf{x}_{0k} = \sum_{k=1}^{n} \epsilon_{ik} \delta_{ik} = \epsilon_{ii} = 0 \tag{5.108}$$

so that

$$\mathbf{x}_{1i} = \sum_{\substack{k=1\\k\neq i}}^{n} \epsilon_{ik}\mathbf{x}_{0k}, \qquad i = 1, 2, \ldots, n \tag{5.109}$$

Inserting Eq. (5.109) into Eq. (5.104a) and using Eq. (5.98a), we obtain

$$A_0 \sum_{\substack{k=1\\k\neq i}}^{n} \epsilon_{ik}\mathbf{x}_{0k} + A_1\mathbf{x}_{0i} = \sum_{\substack{k=1\\k\neq i}}^{n} \epsilon_{ik}\lambda_{0k}\mathbf{x}_{0k} + A_1\mathbf{x}_{0i} = \lambda_{0i}\sum_{\substack{k=1\\k\neq i}}^{n} \epsilon_{ik}\mathbf{x}_{0k} + \lambda_{1i}\mathbf{x}_{0i}, \qquad i = 1, 2, \ldots, n \tag{5.110}$$

Premultiplying Eq. (5.110) through by $\mathbf{y}_{0j}^T$ and considering Eqs. (5.99), we have

$$\epsilon_{ij}\left(\lambda_{0j} - \lambda_{0i}\right) + \mathbf{y}_{0j}^T A_1\mathbf{x}_{0i} = \lambda_{1i}\delta_{ij} \tag{5.111}$$

Letting $i = j$ and using Eq. (5.108), we obtain the eigenvalue perturbations

$$\lambda_{1i} = \mathbf{y}_{0i}^T A_1\mathbf{x}_{0i}, \qquad i = 1, 2, \ldots, n \tag{5.112}$$

On the other hand, for $i \neq j = k$, we obtain

$$\epsilon_{ik} = \frac{\mathbf{y}_{0k}^T A_1\mathbf{x}_{0i}}{\lambda_{0i} - \lambda_{0k}}, \qquad i, k = 1, 2, \ldots, n;\; i \neq k \tag{5.113}$$

The formal determination of the perturbations in the right eigenvectors is completed by inserting Eq. (5.113) into Eq. (5.109).

Using the same pattern and working with Eq. (5.104b), we conclude that the perturbations in the left eigenvectors have the form

$$\mathbf{y}_{1i} = \sum_{\substack{k=1\\k\neq i}}^{n} \gamma_{ik}\mathbf{y}_{0k}, \qquad i = 1, 2, \ldots, n \tag{5.114}$$

where

$$\gamma_{ik} = \frac{\mathbf{x}_{0k}^T A_1\mathbf{y}_{0i}}{\lambda_{0i} - \lambda_{0k}}, \qquad i, k = 1, 2, \ldots, n;\; i \neq k \tag{5.115}$$

Conservative systems play an important role in vibrations. Such systems can be described by real symmetric matrices A. The implication is that, irrespective of how the system parameters are altered, both the original matrix A_0 and the perturbation matrix A_1 are real and symmetric. Of course, in this case the left eigenvectors, also called the *adjoint eigenvectors*, coincide with the right eigenvectors and *the system is self-adjoint*. Of course, the eigenvalues and eigenvectors are all real. Letting $\mathbf{y}_{0i} = \mathbf{x}_{0i}$ in Eq. (5.112), we obtain the perturbations in the eigenvalues

$$\lambda_{1i} = \mathbf{x}_{0i}^T A_1\mathbf{x}_{0i}, \qquad i = 1, 2, \ldots, n \tag{5.116}$$

which are clearly real. Moreover, letting $\mathbf{y}_{0k} = \mathbf{x}_{0k}$, Eq. (5.113) yields the coefficients of the series for the perturbations in the eigenvectors

$$\epsilon_{ik} = \frac{\mathbf{x}_{0k}^T A_1 \mathbf{x}_{0i}}{\lambda_{0i} - \lambda_{0k}}, \qquad i, k = 1, 2, \ldots, n;\; i \neq k \tag{5.117}$$

which are real as well, so that perturbations in the eigenvectors themselves are real, as expected.

Higher-order perturbation solutions have the form

$$\lambda_i = \lambda_{0i} + \lambda_{1i} + \lambda_{2i}, \ldots, \mathbf{x}_i = \mathbf{x}_{0i} + \mathbf{x}_{1i} + \mathbf{x}_{2i}, \ldots, \mathbf{y}_i = \mathbf{y}_{0i} + \mathbf{y}_{1i} + \mathbf{y}_{2i} + \ldots \tag{5.118}$$

They can be obtained by introducing Eqs. (5.118) in conjunction with Eq. (5.100) into Eqs. (5.101), separate terms of different order of magnitude and ignore terms of order higher than the desired one. For example, for a second-order perturbation solution, third-order terms are ignored. Then, the second-order perturbation solution is obtained from the second-order equation in which the zero-order and first-order terms are assumed to be known.

Example 5.5

Obtain a first-order perturbation eigensolution for the matrix

$$A = \begin{bmatrix} 2.6 & -1.1 & 0 \\ -1.1 & 5.2 & -\sqrt{2} \\ 0 & -\sqrt{2} & 10 \end{bmatrix} \tag{a}$$

based on the eigensolution of

$$A_0 = \begin{bmatrix} 2.5 & -1 & 0 \\ -1 & 5 & -\sqrt{2} \\ 0 & -\sqrt{2} & 10 \end{bmatrix} \tag{b}$$

where the latter consists of the eigenvalues

$$\lambda_{01} = 2.119322, \qquad \lambda_{02} = 5, \qquad \lambda_{03} = 10.380678 \tag{c}$$

and eigenvectors

$$\mathbf{x}_{01} = \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix}, \qquad \mathbf{x}_{02} = \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix}, \qquad \mathbf{x}_{03} = \begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix} \tag{d}$$

We note that the eigenvectors have been normalized so as to satisfy $\mathbf{x}_{0j}^T \mathbf{x}_{0i} = \delta_{ij}$ $(i, j = 1, 2, 3)$.

In the first place, we observe that

$$A_1 = A - A_0 = \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{e}$$

which must be considered as "small" relative to A_0. Both matrices A_0 and A_1 are symmetric so that the perturbations are based on Eqs. (5.116) and (5.117).

The perturbations in the eigenvalues are obtained from Eqs. (5.116) in the form

$$\begin{aligned}
\lambda_{11} &= \mathbf{x}_{01}^T A_1 \mathbf{x}_{01} = \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & -0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix} \\
&= 0.045971 \\
\lambda_{12} &= \mathbf{x}_{02}^T A_1 \mathbf{x}_{02} = \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix} \\
&= 0.238710 \\
\lambda_{13} &= \mathbf{x}_{03}^T A_1 \mathbf{x}_{03} = \begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix} \\
&= 0.015319
\end{aligned} \tag{f}$$

The eigenvector perturbations are given by Eqs. (5.105) in conjunction with Eqs. (5.117). Hence,

$$\begin{aligned}
\epsilon_{12} &= -\epsilon_{21} = \frac{\mathbf{x}_{02}^T A_1 \mathbf{x}_{01}}{\lambda_{01} - \lambda_{02}} \\
&= \frac{\begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix}}{2.119322 - 5} = -0.014141 \\
\epsilon_{13} &= -\epsilon_{31} = \frac{\mathbf{x}_{03}^T A_1 \mathbf{x}_{01}}{\lambda_{01} - \lambda_{03}} \\
&= \frac{\begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix}}{2.119322 - 10.380678} = -0.000930 \\
\epsilon_{23} &= -\epsilon_{32} = \frac{\mathbf{x}_{03}^T A_1 \mathbf{x}_{02}}{\lambda_{02} - \lambda_{03}} \\
&= \frac{\begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix}^T \begin{bmatrix} 0.1 & -0.1 & 0 \\ -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix}}{5 - 10.380678} = -0.011176
\end{aligned} \tag{g}$$

Using Eqs. (5.105), we obtain

$$\begin{aligned}
\mathbf{x}_{11} &= \epsilon_{12}\mathbf{x}_{02} + \epsilon_{13}\mathbf{x}_{03} \\
&= -0.014141 \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix} - 0.000930 \begin{bmatrix} 0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix} = \begin{bmatrix} -0.005110 \\ 0.012941 \\ 0.002694 \end{bmatrix} \\
\mathbf{x}_{12} &= \epsilon_{21}\mathbf{x}_{01} + \epsilon_{23}\mathbf{x}_{03}
\end{aligned} \tag{h}$$

$$= 0.014141 \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix} - 0.011176 \begin{bmatrix} l0.032965 \\ -0.259786 \\ 0.965103 \end{bmatrix} = \begin{bmatrix} 0.012821 \\ 0.007924 \\ -0.009885 \end{bmatrix}$$

$$\mathbf{x}_{13} = \epsilon_{31}\mathbf{x}_{01} + \epsilon_{32}\mathbf{x}_{02}$$

$$= 0.000930 \begin{bmatrix} 0.932674 \\ 0.355049 \\ 0.063715 \end{bmatrix} + 0.011176 \begin{bmatrix} 0.359211 \\ -0.898027 \\ -0.254000 \end{bmatrix} = \begin{bmatrix} 0.004882 \\ -0.009706 \\ -0.002779 \end{bmatrix}$$

The first-approximation eigenvalues are given by Eq. (5.103a) in the form

$$\lambda_i \cong \lambda_{0i} + \lambda_{1i}, \qquad i = 1, 2, 3 \qquad \text{(i)}$$

Inserting Eqs. (c) and (f) into Eq. (i), we have

$$\lambda_1 = 2.165293, \qquad \lambda_2 = 5.238710, \qquad \lambda_3 = 10.395997 \qquad \text{(j)}$$

The first-approximation eigenvectors are given by Eq. (5.103b) as follows:

$$\mathbf{x}_i \cong \mathbf{x}_{0i} + \mathbf{x}_{1i}, \qquad i = 1, 2, 3 \qquad \text{(k)}$$

Introducing Eqs. (d) and (h) in Eq. (k) and normalizing so as to satisfy $\mathbf{x}_j^T \mathbf{x}_i = \delta_{ij}$ $(i, j = 1, 2, 3)$, we obtain

$$\mathbf{x}_1 = \begin{bmatrix} 0.927470 \\ 0.367953 \\ 0.066402 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0.371971 \\ -0.889958 \\ -0.263842 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0.037845 \\ -0.269475 \\ 0.962263 \end{bmatrix} \qquad \text{(l)}$$

For comparison purposes, the exact solution of the eigenvalue problem for the matrix A, i.e., obtained directly without any perturbation scheme, is

$$\lambda_1 = 2.164748, \qquad \lambda_2 = 5.238546, \qquad \lambda_3 = 10.396796 \qquad \text{(m)}$$

$$\mathbf{x}_1 = \begin{bmatrix} 0.927810 \\ 0.367120 \\ 0.066263 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0.371103 \\ -0.890158 \\ -0.264388 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 0.038078 \\ -0.269893 \\ 0.962137 \end{bmatrix} \qquad \text{(n)}$$

Comparing Eqs. (j) and (m) on the one hand and Eqs. (l) and (n) on the other hand, we conclude that the first-order perturbation solution produced reasonable results in this particular case.

5.8 SYNOPSIS

Although the algebraic eigenvalue problem is basically a numerical problem associated with lumped systems, there are certain qualitative aspects that make the computational task easier and more purposeful; they can also shed a great deal of light on the problem of approximating distributed-parameter systems by discrete ones. This is particularly true for conservative vibrating systems, characterized by real symmetric eigenvalue problems. The concepts presented in this chapter not only enrich our vibrations experience but also permit us to acquire a deeper understanding of the subject.

The idea of linear transformations, and in particular orthogonal transformations, is fundamental to the vibration of linear systems. Among orthogonal transformations, the coordinate transformation representing rotations has interesting implications, in view of the fact that finding the principal axes of an n-dimensional ellipsoid is equivalent to solving the eigenvalue problem for an $n \times n$ positive definite real symmetric matrix. This problem transcends the field of vibrations, as in three dimensions the same problem arises in finding principal stresses in stress analysis and principal moments of inertia in rigid-body dynamics. Many computational algorithms for solving the algebraic eigenvalue problem for real symmetric matrices use orthogonal transformations. The Jacobi method, in particular, uses transformations representing rotations (Sec. 6.4). Another concept of fundamental importance in vibrations is Rayleigh's quotient. The concept is important not only in the eigenvalue problem for conservative discrete systems but also in approximate techniques for distributed-parameter systems. Rayleigh's quotient permits a qualitative study of the eigensolution properties. For discrete systems, Rayleigh's quotient has a stationary value at an eigenvector, where the stationary value is the associated eigenvalue. Of particular significance is the fact that the minimum value Rayleigh's quotient can take is the lowest eigenvalue, which is sometimes referred to as Rayleigh's principle. The same idea can be used to characterize the higher eigenvalues by the imposition of constraints on the trial vectors. In particular, the separation theorem demonstrates how the eigenvalues of an $n \times n$ matrix A relate to the eigenvalues of an $(n-1) \times (n-1)$ matrix A' obtained from A by removing one row and the associated column. The theorem can be used to demonstrate convergence of approximate techniques for distributed systems, as shown in Chapter 8. Gerschgorin's theorems can be used to locate eigenvalues approximately. The theorems are not restricted to real symmetric matrices, but they are useful only for diagonally dominant matrices. For practical purposes, this rules out their use for certain nonsymmetric matrices, as discussed in Sec. 6.8. Finally, it is shown that, when the system parameters change slightly, a perturbation technique can be used to compute the eigensolutions of the new system by simply correcting the eigensolutions of the original system.

PROBLEMS

5.1 Draw the ellipse $\mathbf{x}^T A\mathbf{x} = 1$ corresponding to the matrix A of Example 5.1 and obtain graphically the eigenvalues and eigenvectors. Explain the inconsistency between the eigenvectors obtained here and those given by Eqs. (f) of Example 5.1.

5.2 Use Rayleigh's quotient to estimate the lowest natural frequency for the system of Example 4.6. Use as trial vector the vector of static displacements obtained by loading the system with forces proportional to the masses.

5.3 The masses in the system of Problem 4.22 have the values $m_1 = m$, $m_2 = m_3 = 2m$, $m_4 = m$. Use Rayleigh's quotient to estimate the two lowest natural frequencies.

5.4 The masses in the system of Problem 4.22 have the values $m_1 = m_2 = 2m$, $m_3 = m_4 = m$. Use Rayleigh's quotient to estimate the two lowest natural frequency.

5.5 The parameters in the system of Problem 4.24 have the values $L_1 = L_2 = L_3 = L_4 = L$, $I_1 = I_2 = I_3 = I$, $I_4 = 2I$, $J_1 = J_2 = J_4 = J$, $J_3 = 2J$. Estimate the lowest natural frequency by means of Rayleigh's quotient.

5.6 Verify that the eigenvalues of the 3×3 matrix A of Example 4.6 and the eigenvalues of the 2×2 matrix A' obtained by removing one row and the associated column from A satisfy the separation theorem, no matter which row and column are removed.

5.7 Verify that the eigenvalues of the 3 × 3 matrix

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 3 & -\sqrt{2} \\ 0 & -\sqrt{2} & 1 \end{bmatrix}$$

and the eigenvalues of the 2 × 2 matrix A' obtained by removing one row and the associated column from A satisfy the separation theorem, regardless of the row and column removed.

5.8 Verify that the eigenvalues of the 4 × 4 matrix A for the system of Problem 5.4 and the eigenvalues of the 3 × 3 matrix A' obtained by removing one row and the associated column from A satisfy the separation theorem, independently of the row and column removed.

5.9 Show that the eigenvalues of the matrix

$$A = \begin{bmatrix} 5 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 5 \end{bmatrix}$$

are consistent with both Gerschgorin's theorems.

5.10 Show that the eigenvalues of the matrix

$$A = \begin{bmatrix} 3.8867 & 0.5672 & 0 \\ 0.5672 & 0.7897 & 0.7481 \\ 0 & 0.7481 & 2.4903 \end{bmatrix}$$

are consistent with both Gerschgorin's theorems.

5.11 Use the approach of Sec. 5.7 to produce a first-order perturbation theory for the eigenvalue problem

$$K\mathbf{u} = \lambda M\mathbf{u}$$

where

$$K = K_0 + K_1, \qquad M = M_0 + M_1$$

in which K_0, M_0, K and M are positive definite real symmetric matrices and K_1 and M_1 are "small" relative to K_0 and M_0, respectively.

5.12 Use the approach of Sec. 5.7 to produce a second-order perturbation theory for the eigenvalue problem

$$A\mathbf{x} = \lambda\mathbf{x}$$

where

$$A = A_0 + A_1$$

in which A_0 and A are real symmetric matrices and A_1 is "small" relative to A_0.

5.13 Use the developments of Problem 5.12 to compute a second-order perturbation solution to the problem of Example 5.5. Compare results with those of Example 5.5 and draw conclusions.

5.14 A two-degree-of-freedom system is defined by the mass, damping and stiffness matrices

$$M = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \qquad C = \begin{bmatrix} 0.2 & -0.1 \\ -0.1 & 0.1 \end{bmatrix}, \qquad K = \begin{bmatrix} 5 & -4 \\ -4 & 4 \end{bmatrix}$$

Develop a procedure for solving the eigenvalue problem whereby the effect of damping is treated as a perturbation on the undamped system.

BIBLIOGRAPHY

1. Courant, R., "Über die Eigenwerte bei den Differentialgleichungen des mathematischen Physik," *Mathematische Zeitschrift*, Vol. 7, 1920, pp. 1–57.
2. Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, Vol. 1, Wiley, New York, 1989.
3. Fischer, E., "Über quadratische Formen mit reellen Koeffizienten," *Monatshefte für Mathematik und Physik*, Vol. 16, 1905, pp. 234–249.
4. Franklin, J. N., *Matrix Theory*, Prentice Hall, Englewood Cliffs, NJ, 1968.
5. Golub, G. H. and Van Loan, C. F., *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
6. Gould, S. H., *Variational Methods for Eigenvalue Problems: An Introduction to the Methods of Rayleigh, Ritz, Weinstein and Aronszajn*, Dover, New York, 1995
7. Meirovitch, L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.
8. Meirovitch, L. and Baruh, H., "On the Inclusion Principle for the Hierarchical Finite Element Method," *International Journal for Numerical Methods in Engineering*, Vol. 19, 1983, pp. 281–291.
9. Meirovitch, L., "A Separation Principle for Gyroscopic Conservative Systems," *Journal of Vibration and Acoustics* (forthcoming).
10. Rayleigh (Lord), *Theory of Sound*, Vol. 1, Dover, New York, 1945 (first American edition of the 1894 edition).
11. Routh, E. J., *Advanced Dynamics of a System of Rigid Bodies*, 6th ed., Dover, New York, 1905.
12. Strang, G., *Linear Algebra and Its Applications*, 3rd ed., Harcourt Brace Jovanovich, San Diego, 1988.
13. Weyl, H., "Das asymptotishe Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Auwendung auf die Theorie der Hohlraumstrahlung)," *Mathematische Annalen*, Vol. 71, 1912, pp. 441–479.
14. Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1988.

6

COMPUTATIONAL TECHNIQUES FOR THE ALGEBRAIC EIGENVALUE PROBLEM

As amply demonstrated in Chapter 4, the algebraic eigenvalue problem plays a pivotal role in the study of vibrations of multi-degree-of-freedom systems. Indeed, generally the equations for the small motions of vibrating multi-degree-of-freedom systems consist of a set of simultaneous ordinary differential equations. The solution of these equations, linear as they are, causes great difficulties when the equations are in simultaneous form. These difficulties can be obviated by carrying out a linear transformation rendering the equations independent, where the transformation matrix is the modal matrix, a square matrix with its columns representing the vibration modes. Because mathematically the vibration modes represent the system eigenvectors, it becomes necessary to solve the algebraic eigenvalue problem. Of course, the reason the linear transformation using the modal matrix is capable of decoupling the simultaneous differential equations lies in the orthogonality property of the eigenvectors, a remarkable property indeed.

The algebraic eigenvalue problem is essentially a numerical problem. The rapid rise in the ability of digital computers to process numerical solutions for systems of large order has stimulated an ever increasing interest in the development of computational algorithms for the algebraic eigenvalue problem. In this chapter, computational techniques most appropriate to the study of vibrations are presented. The choice of algorithms is based on both pedagogical and computational efficiency considerations.

Many computational algorithms for the algebraic eigenvalue problem call for the solution of nonhomogeneous linear algebraic equations. This chapter begins with a very efficient approach to this problem, namely, Gaussian elimination with back-substitution. The standard algebraic eigenvalue problem is described by means of a single matrix. Problems defined by a single real symmetric matrix, or problems that can be reduced to a single real symmetric matrix, have very desirable properties. In particular, the eigenvalues are real and the eigenvectors are real and orthogonal. In the vibration of conservative systems, the problem is generally defined in terms of two real symmetric matrices, the mass matrix and the stiffness matrix. If one of the matrices is positive definite, as the mass matrix is almost by definition, then the problem can be reduced to one in terms of a single symmetric matrix by means of the Cholesky decomposition.

In one form or another, virtually all algorithms for the algebraic eigenvalue problem are iterative in nature. One of the oldest and best known algorithms is matrix iteration by the power method. Although the method has many drawbacks, its inclusion can be justified on the basis of academic considerations. The method yields one eigensolution at a time and has some merits if the interest lies in only a few dominant eigensolutions. A method lending itself to a nice geometric interpretation is the Jacobi method. Indeed, for a single $n \times n$ real symmetric positive definite matrix A, the method can be shown to be equivalent to finding the principal axes of an n-dimensional ellipsoid. The method yields all the eigensolutions simultaneously. Many of the algorithms are efficient only if the matrix A is in tridiagonal form, a form encountered only occasionally in vibrations. Hence, for the most part, it is necessary to reduce a symmetric matrix to a symmetric tridiagonal one. To this end, three algorithms are presented, Givens' method, Householder's method and Lanczos' method. One of the most efficient methods for the computation of the eigenvalues of symmetric tridiagonal matrices is due to Givens, and is based on the separation principle discussed in Chapter 5. Another method is the QR algorithm, which becomes competitive only when the matrix is tridiagonal and shifts are used. Both the Givens method and the QR method yield eigenvalues alone. The eigenvectors belonging to these eigenvalues can be computed efficiently by inverse iteration. Other methods of interest are Rayleigh's quotient iteration, which targets individual eigensolutions, and simultaneous iteration, which permits computation of a limited number of dominant eigensolutions at the same time.

The eigenvalue problem for nonsymmetric matrices is considerably more involved than for symmetric matrices, particularly if some of the eigenvalues are complex. The simplest algorithm for such problems is matrix iteration by the power method, a significantly different version from the one for symmetric matrices. A much more powerful algorithm is the QR method, which yields eigenvalues alone. Here too, efficiency considerations dictate that the nonsymmetric matrix be reduced to a special form, in this case a Hessenberg form. Finally, the eigenvectors can be obtained by means of inverse iteration modified so as to accommodate complex conjugate eigenvectors.

6.1 SOLUTION OF LINEAR ALGEBRAIC EQUATIONS. GAUSSIAN ELIMINATION

In the process of solving algebraic eigenvalue problems, it becomes necessary at times to solve sets of linear algebraic equations, a most fundamental problem in linear algebra. The interest here lies in the case in which the number of equations is equal to the number of unknowns, so that we consider the system of equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n &= b_n \end{aligned} \tag{6.1}$$

Equations (6.1) can be written in the compact matrix form

$$A\mathbf{x} = \mathbf{b} \tag{6.2}$$

where $A = [a_{ij}]$ is the $n \times n$ coefficient matrix, $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]^T$ the n-vector of unknowns and $\mathbf{b}$ the n-vector of nonhomogeneous terms. To discuss the conditions under which Eq. (6.2) possesses a unique solution, it is convenient to write the matrix A in terms of its columns, as well as to introduce an augmented matrix A_b, as follows:

$$A = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n], \qquad A_b = [\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_n\ \mathbf{b}] \tag{6.3a, b}$$

Equation (6.3a) permits us to rewrite Eq. (6.2) in the form

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n = \mathbf{b} \tag{6.4}$$

which expresses the vector $\mathbf{b}$ as a linear combination of the column vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ of A. But, the rank of a matrix can be defined as the number of linearly independent columns of the matrix. Hence, Eq. (6.2) has a solution if (see Appendix B)

$$\text{rank}\, A = \text{rank}\, A_b = n \tag{6.5}$$

and the solution is unique. Equation (6.5) implies that A is nonsingular, which further implies that $\det A \neq 0$. Throughout this discussion, we assume that this is indeed the case.

It is well known that an analytical solution of Eqs. (6.1), or Eq. (6.2), can be obtained by Cramer's rule in the form of the ratio of two determinants, or

$$x_j = \frac{\det[\mathbf{a}_1\ \mathbf{a}_2\ \ldots\ \mathbf{a}_{j-1}\ \mathbf{b}\ \mathbf{a}_{j+1}\ \ldots\ \mathbf{a}_n]}{\det A}, \quad j = 1, 2, \ldots, n \tag{6.6}$$

Yet, Cramer's rule is not the method of choice for solving linear algebraic equations, particularly for $n > 3$. The reason can be traced to the fact that, in producing a numerical solution of Eqs. (6.1), or Eq. (6.2), an analytical solution may not necessarily be the most efficient computationally, and at times it may not even be feasible. In this regard, we define the most efficient computational algorithm as the one requiring the smallest number of multiplications. Judging by this criterion, Cramer's rule is a computational nightmare, as the evaluation of determinants requires an excessive number of multiplications. Indeed, the evaluation of an $n \times n$ determinant requires $n!$ multiplications, a number that increases very rapidly with n. As an example, the evaluation of a relatively moderate 10×10 determinant requires the stunning number of 3,628,800 multiplications. Clearly, Cramer's rule is not a computational tool, and a reasonable alternative is imperative.

The solution of Eq. (6.2) can be written in the simple form

$$\mathbf{x} = A^{-1}\mathbf{b} \tag{6.7}$$

where the inverse of A can be obtained by means of the formula (Appendix B)

$$A^{-1} = \frac{\operatorname{adj} A}{\det A} \tag{6.8}$$

in which

$$\operatorname{adj} A = \left[(-1)^{j+k} \det M_{jk}\right]^T \tag{6.9}$$

is a square matrix known as the *adjugate* of A, where $(-1)^{j+k} \det M_{jk}$ is the *cofactor* corresponding to the entry a_{jk} of A and M_{jk} is the submatrix obtained from A by striking out the jth row and kth column. But, once again we are facing the problem of evaluating determinants, which seems to suggest the nonsensical idea that the solution of linear algebraic equations represents a titanic task. Nothing could be farther from the truth, however, as the above conclusion was based on Eq. (6.8) for the inverse of a matrix. Indeed, a very efficient method for solving linear algebraic equations, which is implicitly equivalent to Eq. (6.7), does exist and is known as *Gaussian elimination with back-substitution*. Although the process of solving linear algebraic equations by eliminating variables is well known, in this section we present the Gaussian elimination in a matrix form suitable for computer programming. In the process, we also develop an efficient method for carrying out matrix inversions.

The Gaussian elimination is basically a procedure for solving sets of linear algebraic equations through elementary operations. The net effect of these elementary operations is to carry out a linear transformation on Eq. (6.2), which amounts to premultiplying Eq. (6.2) by the $n \times n$ matrix P, so that

$$PA\mathbf{x} = P\mathbf{b} \tag{6.10}$$

The transformation matrix P is such that PA is an upper triangular matrix U. Hence, introducing the notation

$$PA = U, \qquad P\mathbf{b} = \mathbf{c} \tag{6.11a, b}$$

Eq. (6.10) can be rewritten as

$$U\mathbf{x} = \mathbf{c} \tag{6.12}$$

The question remains as to how to generate the transformation matrix P required for the computation of U and $\mathbf{c}$. The process involves $n - 1$ steps, where the steps are perhaps best explained by beginning with the equations in index notation. To this end, we introduce the notation

$$A = A_0, \qquad \mathbf{b} = \mathbf{a}_{n+1}^{(0)} \tag{6.13a, b}$$

and rewrite Eqs. (6.1) as follows:

$$
\begin{aligned}
a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + \dots + a_{1n}^{(0)}x_n &= a_{1,n+1}^{(0)} \\
a_{21}^{(0)}x_1 + a_{22}^{(0)}x_2 + \dots + a_{2n}^{(0)}x_n &= a_{2,n+1}^{(0)} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
a_{n1}^{(0)}x_1 + a_{n2}^{(0)}x_2 + \dots + a_{nn}^{(0)}x_n &= a_{n,n+1}^{(0)}
\end{aligned}
\tag{6.14}
$$

The first step consists of subtracting $a_{i1}^{(0)}/a_{11}^{(0)}$ times the first of Eqs. (6.14) from the ith equation ($i = 2, 3, \dots, n$), provided $a_{11}^{(0)} \neq 0$. The result is

$$
\begin{aligned}
a_{11}^{(0)}x_1+a_{12}^{(0)}x_2 + \dots + a_{1n}^{(0)}x_n &= a_{1,n+1}^{(0)} \\
a_{22}^{(1)}x_2 + \dots + a_{2n}^{(1)}x_n &= a_{2,n+1}^{(1)} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
a_{n2}^{(1)}x_2 + \dots + a_{nn}^{(1)}x_n &= a_{n,n+1}^{(1)}
\end{aligned}
\tag{6.15}
$$

where the modified coefficients have the expressions

$$
a_{ij}^{(1)} = a_{ij}^{(0)} - \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}a_{1j}^{(0)}, \qquad i = 2, 3, \dots, n;\ j = 2, 3, \dots, n+1 \tag{6.16}
$$

Next, we assume that $a_{22}^{(1)} \neq 0$ in Eqs. (6.15) and subtract $a_{i2}^{(1)}/a_{22}^{(1)}$ times the second equation from the ith equation ($i = 3, 4, \dots, n$) and obtain

$$
\begin{aligned}
a_{11}x_1^{(0)} + a_{12}^{(0)}x_2+a_{13}^{(0)}x_3 + \dots + a_{1n}^{(0)}x_n &= a_{1,n+1}^{(0)} \\
a_{22}^{(1)}x_2+a_{23}^{(1)}x_3 + \dots + a_{2n}^{(1)}x_n &= a_{2,n+1}^{(1)} \\
a_{33}^{(2)}x_3 + \dots + a_{3n}^{(2)}x_n &= a_{3,n+1}^{(2)} \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
a_{n3}^{(2)}x_3 + \dots + a_{nn}^{(2)}x_n &= a_{n,n+1}^{(2)}
\end{aligned}
\tag{6.17}
$$

where

$$
a_{ij}^{(2)} = a_{ij}^{(1)} - \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}a_{2j}^{(1)}, \qquad i = 3, 4, \dots, n;\ j = 3, 4, \dots, n+1 \tag{6.18}
$$

After $n-1$ steps, the procedure yields

$$\begin{aligned} a_{11}^{(0)}x_1 + a_{12}^{(0)}x_2 + a_{13}^{(0)}x_3 + \dots + a_{1n}^{(0)}x_n &= a_{1,n+1}^{(0)} \\ a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \dots + a_{2n}^{(1)}x_n &= a_{2,n+1}^{(1)} \\ a_{33}^{(2)}x_3 + \dots + a_{3n}^{(2)}x_n &= a_{3,n+1}^{(2)} \\ \dots\dots\dots\dots\dots\dots \\ a_{nn}^{(n-1)}x_n &= a_{n,n+1}^{(n-1)} \end{aligned} \tag{6.19}$$

Comparing Eqs. (6.12) and (6.19), we conclude that

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ 0 & u_{22} & u_{23} & \dots & u_{2n} \\ 0 & 0 & u_{33} & \dots & u_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & u_{nn} \end{bmatrix} = \begin{bmatrix} a_{11}^{(0)} & a_{12}^{(0)} & a_{13}^{(0)} & \dots & a_{1n}^{(0)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} \\ 0 & 0 & a_{33}^{(2)} & \dots & a_{3n}^{(2)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn}^{(n-1)} \end{bmatrix} \tag{6.20a}$$

$$\mathbf{c} = [c_1 \ c_2 \ c_3 \ \dots \ c_n]^T = \left[a_{1,n+1}^{(0)} \ a_{2,n+1}^{(1)} \ a_{3,n+1}^{(2)} \ \dots \ a_{n,n+1}^{(n-1)}\right]^T \tag{6.20b}$$

The preceding operations can be cast in matrix form. To this end, Eqs. (6.15) can be rewritten as

$$A_1\mathbf{x} = \mathbf{a}_{n+1}^{(1)} \tag{6.21}$$

where the coefficient matrix A_1 and the vector $\mathbf{a}_{n+1}^{(1)}$ are obtained from A_0 and $\mathbf{a}_{n+1}^{(0)}$, respectively, by writing

$$A_1 = P_1A_0, \qquad \mathbf{a}_{n+1}^{(1)} = P_1\mathbf{a}_{n+1}^{(0)} \tag{6.22a, b}$$

in which the transformation matrix P_1 has the form

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -p_{21} & 1 & 0 & \dots & 0 \\ -p_{31} & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -p_{n1} & 0 & 0 & \dots & 1 \end{bmatrix} \tag{6.23}$$

where, according to Eqs. (6.16),

$$p_{i1} = \frac{a_{i1}^{(0)}}{a_{11}^{(0)}}, \qquad i = 2, 3, \dots, n \tag{6.24}$$

In a similar fashion, the result of the second step is

$$A_2\mathbf{x} = \mathbf{a}_{n+1}^{(2)} \tag{6.25}$$

in which

$$A_2 = P_2 A_1, \qquad \mathbf{a}_{n+1}^{(2)} = P_2 \mathbf{a}_{n+1}^{(1)} \tag{6.26a, b}$$

where

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & -p_{32} & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & -p_{n2} & 0 & \dots & 1 \end{bmatrix} \tag{6.27}$$

in which, according to Eqs. (6.18), the entries p_{i2} have the expressions

$$p_{i2} = \frac{a_{i2}^{(1)}}{a_{22}^{(1)}}, \qquad i = 3, 4, \dots, n \tag{6.28}$$

The process continues in the same fashion and ends after $n-1$ steps with the results

$$U = A_{n-1} = P_{n-1} A_{n-2} \tag{6.29a}$$

$$\mathbf{c} = \mathbf{a}_{n+1}^{(n-1)} = P_{n-1} \mathbf{a}_{n+1}^{(n-2)} \tag{6.29b}$$

By induction, Eqs. (6.29) yield

$$U = P_{n-1} A_{n-2} = P_{n-1} P_{n-2} A_{n-3} = \dots = P_{n-1} P_{n-2} \dots P_2 P_1 A \tag{6.30a}$$

$$\mathbf{c} = P_{n-1} \mathbf{a}_{n+1}^{(n-2)} = P_{n-1} P_{n-2} \mathbf{a}_{n+1}^{(n-3)} = \dots = P_{n-1} P_{n-2} \dots P_2 P_1 \mathbf{b} \tag{6.30b}$$

in which we replaced A_0 by A and $\mathbf{a}_{n+1}^{(0)}$ by $\mathbf{b}$, according to Eqs. (6.13). Then, comparing Eqs. (6.11) and (6.30), we conclude that the transformation matrix has the form of the continuous matrix product

$$P = P_{n-1} P_{n-2} \dots P_2 P_1 \tag{6.31}$$

which indicates that the transformation matrix P can be generated in $n-1$ steps. This is merely of academic interest, however, as P is never computed explicitly, because U and $\mathbf{c}$ are determined by means of Eqs. (6.22), (6.26)...(6.29) and not through Eqs. (6.11).

With U and $\mathbf{c}$ obtained from Eqs. (6.29), Eq. (6.12) can be solved with ease by *back-substitution*. Indeed, the bottom equation involves x_n alone, and can be solved with the result

$$x_n = c_n / u_{nn} \tag{6.32}$$

Then, having x_n, the $(n-1)$th equation can be solved to obtain

$$x_{n-1} = \frac{1}{u_{n-1,n-1}} \left(c_{n-1} - u_{n-1,n} x_n \right) \tag{6.33}$$

Next, upon substitution of x_{n-1} and x_n into the $(n-2)$th equation, we are able to solve for x_{n-2}. The procedure continues by solving in sequence for $x_{n-3}, \dots, x_2, x_1$.

The question can be raised as to why not reduce the matrix A to diagonal form, instead of upper triangular form, thus obviating the need for back-substitution. Indeed, if we subtract $a_{12}^{(0)}/a_{22}^{(1)}$ times the second of Eqs. (6.19) from the first, we obtain

$$\begin{aligned}
a_{11}^{(0)}x_1 + \qquad\qquad\quad a_{13}^{(2)}x_3 + \ldots + a_{1n}^{(2)}x_n &= a_{1,n+1}^{(2)} \\
a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \ldots + a_{2n}^{(1)}x_n &= a_{2,n+1}^{(1)} \\
a_{33}^{(2)}x_3 + \ldots + a_{3n}^{(2)}x_n &= a_{3,n+1}^{(2)} \qquad (6.34) \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\
a_{n3}^{(2)}x_3 + \ldots + a_{nn}^{(2)}x_n &= a_{n,n+1}^{(2)}
\end{aligned}$$

so that now both the first and second column have one element alone. The procedure continues in the same fashion, subtracting the equation with the entry to be isolated both from the equations below it and the equations above it, ultimately obtaining a diagonal matrix of coefficients. Such a variation on the Gaussian elimination does indeed exist, and is known as the *Gauss-Jordan reduction*. But, whereas the Gauss-Jordan reduction is intuitively appealing, it is not as efficient as Gaussian elimination. Indeed, the Gaussian elimination requires approximately $n^3/3$ operations and back-substitution about $n^2/2$, as opposed to Gauss-Jordan reduction, which requires approximately $n^3/2$ operations, where operations are defined as multiplications or divisions. Hence, for large n, Gaussian elimination is the method of choice.

The fact that the matrix A is not singular guarantees that Eqs. (6.1) admit a solution, but it does not guarantee that the computational process cannot break down. To explain this statement, we recall that the rth step in the Gaussian elimination involves division by the element $a_{rr}^{(r-1)}$, where $a_{rr}^{(r-1)}$ is known as the rth *pivot*. In the process presented above, the pivots are taken in order down the main diagonal. Clearly, the method breaks down when one of the pivots is zero. In fact, the pivot does not have to be exactly zero for difficulties to arise. Indeed, if the pivot is a very small number, large computational errors can occur. This should not be construed that the equations do not have a solution; it merely implies that, to permit a continuation of the process, it is necessary to interchange rows and/or columns. For example, if the rth pivot is unduly small, then we can choose as the new rth pivot the element in the rth column of largest magnitude, or

$$a_{rr}^{(r-1)} = \max\left|a_{sr}^{(r-1)}\right|, \qquad s = r, r+1, \ldots, n \qquad (6.35)$$

which amounts to interchanging the r and s rows in the equations in which x_r is in line for elimination below the main diagonal. This can be carried out by premultiplying the matrix A_{r-1} and the vector $\mathbf{a}_{n+1}^{(r-1)}$ by the *permutation matrix*

$$I_{rs} = \begin{array}{cc} \begin{array}{ccccccccc} & & & r & & s & & & \end{array} & \\ \left[\begin{array}{ccccccccc} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & \cdots & 1 & & & \\ & & & \vdots & \ddots & \vdots & & & \\ & & & 1 & \cdots & 0 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{array}\right] & \begin{array}{c} \\ \\ \\ r \\ \\ s \\ \\ \\ \\ \end{array} \end{array} \tag{6.36}$$

The process just described is known as *partial pivoting*. The search for a large pivot need not be confined to the rth column and can be extended to all columns $r \leq t \leq n$, so that the new pivot is chosen according to

$$a_{rr}^{(r-1)} = \max \left| a_{st}^{(r-1)} \right|, \qquad s, t = r, r+1, \dots, n \tag{6.37}$$

This requires interchanging the rows r and s and the columns r and t, which can be achieved by premultiplication of the matrix A_{r-1} and vector $\mathbf{a}_{n+1}^{(r-1)}$ by the permutation matrix I_{rs} and postmultiplication of the matrix $I_{rs}A_{r-1}$ by the permutation matrix I_{rt}. Note that an interchange of columns requires an interchange of the unknowns x_r and x_t (see Example 6.2). The process in which both rows and columns are interchanged is referred to as *complete pivoting*. The algorithm thus modified is called *Gaussian elimination with interchanges*, or *with pivoting for size*. The modified algorithm does guarantee a solution, provided the set of equations is consistent, which is implied by Eq. (6.5). Pivoting does complicate the solution process and degrades the performance of the algorithm, and should not be used unless the solution is in jeopardy. Unfortunately, such a judgment cannot be made a priori.

Equation (6.11a) can be given an interesting interpretation by expressing a typical intermediate transformation matrix in the form

$$P_r = I - \mathbf{p}_r \mathbf{e}_r^T, \qquad r = 1, 2, \dots, n-1 \tag{6.38}$$

where $\mathbf{e}_r$ is the rth standard unit vector and

$$\mathbf{p}_r = [0\, 0 \dots 0\; p_{r+1,r}\; p_{r+2,r}\; \cdots\; p_{nr}]^T, \qquad r = 1, 2, \dots, n-1 \tag{6.39}$$

Then, observing that $\mathbf{e}_r^T \mathbf{p}_r = 0$, it is easy to verify that

$$P_r^{-1} = I + \mathbf{p}_r \mathbf{e}_r^T, \qquad r = 1, 2, \dots, n-1 \tag{6.40}$$

so that, using Eqs. (6.31) and (6.40) and recognizing that $\mathbf{e}_i^T \mathbf{p}_j = 0,\ i < j$, we can write

$$\begin{aligned} P^{-1} &= P_1^{-1} P_2^{-1} \dots P_{n-2}^{-1} P_{n-1}^{-1} \\ &= \left(I + \mathbf{p}_1 \mathbf{e}_1^T\right)\left(I + \mathbf{p}_2 \mathbf{e}_2^T\right) \dots \left(I + \mathbf{p}_{n-2} \mathbf{e}_{n-2}^T\right)\left(I + \mathbf{p}_{n-1} \mathbf{e}_{n-1}^T\right) \end{aligned}$$

$$=I + \sum_{r=1}^{n-1} \mathbf{p}_r \mathbf{e}_r^T = L \tag{6.41}$$

where L is a unit lower triangular matrix, i.e., a lower triangular matrix with 1's on the main diagonal. Then, premultiplying Eq. (6.11a) by $P^{-1} = L$, we obtain

$$A = LU \tag{6.42}$$

which states that Gaussian elimination is equivalent to factorization of the coefficient matrix A into a product of a unit lower triangular matrix L and an upper triangular matrix U.

Example 6.1

Solve the linear algebraic equations

$$\begin{aligned} 4x_1 + 2x_2 - 2.4x_3 &= 1.6 \\ 2x_1 + 1.05x_2 + 0.2x_3 &= 5.8 \\ x_1 + 2x_2 - 3.6x_3 &= -7.1 \end{aligned} \tag{a}$$

by Gaussian elimination with back-substitution.

We propose to solve Eqs. (a) using the matrix formulation. To this end, we use the notation of Eqs. (6.13) and write

$$A_0 = \begin{bmatrix} 4 & 2 & -2.4 \\ 2 & 1.05 & 0.2 \\ 1 & 2 & -3.6 \end{bmatrix}, \qquad \mathbf{a}_4^{(0)} = \begin{bmatrix} 1.6 \\ 5.8 \\ -7.1 \end{bmatrix} \tag{b}$$

To determine the first transformation matrix, P_1, we use Eqs. (6.24) with $n = 3$ and write

$$p_{21} = \frac{a_{21}^{(0)}}{a_{11}^{(0)}} = \frac{2}{4} = 0.5, \qquad p_{31} = \frac{a_{31}^{(0)}}{a_{11}^{(0)}} = \frac{1}{4} = 0.25 \tag{c}$$

so that, using Eq. (6.23), the first transformation matrix is

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ -0.25 & 0 & 1 \end{bmatrix} \tag{d}$$

Inserting Eqs. (b) and (d) into Eqs. (6.22), we obtain

$$A_1 = P_1 A_0 = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ -0.25 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2.4 \\ 2 & 1.05 & 0.2 \\ 1 & 2 & -3.6 \end{bmatrix} = \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 1.5 & -3 \end{bmatrix} \tag{e}$$

and

$$\mathbf{a}_4^{(1)} = P_1 \mathbf{a}_4^{(0)} = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ -0.25 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1.6 \\ 5.8 \\ -7.1 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 5 \\ -7.5 \end{bmatrix} \tag{f}$$

Next, we use Eqs. (6.28) with $n = 3$ and write

$$p_{32} = \frac{a_{32}^{(1)}}{a_{22}^{(1)}} = \frac{1.5}{0.05} = 30 \tag{g}$$

so that, inserting Eq. (g) into Eq. (6.27), we obtain

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -30 & 1 \end{bmatrix} \tag{h}$$

Finally, using Eqs. (6.29) with $n = 3$, we have

$$U = A_2 = P_2A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -30 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 1.5 & -3 \end{bmatrix} = \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 0 & -45 \end{bmatrix} \tag{i}$$

and

$$\mathbf{c} = \mathbf{a}_4^{(2)} = P_2\mathbf{a}_4^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -30 & 1 \end{bmatrix} \begin{bmatrix} 1.6 \\ 5 \\ -7.5 \end{bmatrix} = \begin{bmatrix} 1.6 \\ 5 \\ -157.5 \end{bmatrix} \tag{j}$$

At this point, we are ready for the back-substitution. Using Eq. (6.32) with $n = 3$, we can write

$$x_3 = \frac{c_3}{u_{33}} = \frac{-157.5}{-45} = 3.5 \tag{k}$$

and from Eq. (6.33), we have

$$x_2 = \frac{1}{u_{22}}(c_2 - u_{23}x_3) = \frac{1}{0.05}(5 - 1.4 \times 3.5) = 2 \tag{l}$$

It is easy to verify that x_1 can be obtained by writing

$$x_1 = \frac{1}{u_{11}}(c_1 - u_{12}x_2 - u_{13}x_3) = \frac{1}{4}[1.6 - 2 \times 2 - (-2.4) \times 3.5] = 1.5 \tag{m}$$

which completes the solution.

Example 6.2

Solve Eqs. (a) of Example 6.1 by Gaussian elimination in two ways, with partial pivoting and with complete pivoting.

The first step remains as in Example 6.1. Hence, from Eqs. (e) and (f) of Example 6.1, we have

$$A_1 = \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 1.5 & -3 \end{bmatrix}, \quad \mathbf{a}_4^{(1)} = \begin{bmatrix} 1.6 \\ 5 \\ -7.5 \end{bmatrix} \tag{a}$$

The pivot $a_{22}^{(1)}$ is small, so that we propose to interchange rows 2 and 3. To this end, we use the permutation matrix

$$I_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{b}$$

and introduce the modified matrix

$$A_1^* = I_{23}A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 1.5 & -3 \end{bmatrix} = \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 1.5 & -3 \\ 0 & 0.05 & 1.4 \end{bmatrix} \tag{c}$$

and modified vector

$$\mathbf{a}_4^{*(1)} = I_{23}\mathbf{a}_4^{(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1.5 \\ 5 \\ -7.5 \end{bmatrix} = \begin{bmatrix} 1.6 \\ -7.5 \\ 5 \end{bmatrix} \tag{d}$$

From here on, the procedure follows the same pattern as in Example 6.1. To determine the second transformation matrix, P_2, we use Eqs. (6.28) with $n = 3$ and write

$$p_{32} = \frac{a_{32}^{*(1)}}{a_{22}^{*(1)}} = \frac{0.05}{1.5} = \frac{1}{30} \tag{e}$$

so that, from Eq. (6.27), we have

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/30 & 1 \end{bmatrix} \tag{f}$$

Hence, using Eqs. (6.29) with $n = 3$, we obtain

$$U = A_2 = P_2 A_1^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/30 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 1.5 & -3 \\ 0 & 0.05 & 1.4 \end{bmatrix} = \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 1.5 & -3 \\ 0 & 0 & 1.5 \end{bmatrix} \tag{g}$$

and

$$\mathbf{c} = a_4^{(2)} = P_2 \mathbf{a}_4^{*(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/30 & 1 \end{bmatrix} \begin{bmatrix} 1.6 \\ -7.5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1.6 \\ -7.5 \\ 5.25 \end{bmatrix} \tag{h}$$

The back-substitution is also as in Example 6.1, namely,

$$\begin{aligned} x_3 &= \frac{c_3}{u_{33}} = \frac{5.25}{1.5} = 3.5 \\ x_2 &= \frac{1}{u_{22}}(c_2 - u_{23}x_3) = \frac{1}{1.5}[-7.5 - (-3) \times 3.5] = 2 \\ x_1 &= \frac{1}{u_{11}}(c_1 - u_{12}x_2 - u_{13}x_3) = \frac{1}{4}[1.6 - 2 \times 2 - (-2.4) \times 3.5] = 1.5 \end{aligned} \tag{i}$$

The solution agrees with that obtained in Example 6.1. This is not surprising, as the pivot was not sufficiently small to cause loss of accuracy and thus to warrant interchanges of equations. Indeed, the partial pivoting carried out here was mainly to illustrate the process.

Next, we observe from the matrix A_1 of Example 6.1 that the entry with the largest magnitude in the 2×2 lower right corner submatrix is $a_{33}^{(1)} = -3$. Hence, we designate this entry as the second pivot, which requires complete pivoting. Because this involves an interchange of columns, we must redefine the vector of unknowns. To this end, we recognize that $I_{23}I_{23} = I$, where I is the identity matrix, and consider

$$I_{23}A_1I_{23}I_{23}\mathbf{x} = A_1^*\mathbf{y} \tag{j}$$

where this time

$$\begin{aligned} A_1^* = I_{23}A_1I_{23} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2.4 \\ 0 & 0.05 & 1.4 \\ 0 & 1.5 & -3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 4 & -2.4 & 2 \\ 0 & -3 & 1.5 \\ 0 & 1.4 & 0.05 \end{bmatrix} \end{aligned} \tag{k}$$

In addition,

$$\mathbf{y} = I_{23}\mathbf{x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_3 \\ x_2 \end{bmatrix} \tag{l}$$

The vector $\mathbf{a}_4^{*(1)}$ remains as that given by Eq. (d). Then, following the pattern established earlier in this example, we write

$$p_{32} = \frac{a_{32}^{*(1)}}{a_{22}^{*(1)}} = -\frac{1.4}{3} \tag{m}$$

so that

$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1.4/3 & 1 \end{bmatrix} \tag{n}$$

Moreover,

$$\begin{aligned} U = A_2 = P_2 A_1^* &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1.4/3 & 1 \end{bmatrix} \begin{bmatrix} 4 & -2.4 & 2 \\ 0 & -3 & 1.5 \\ 0 & 1.4 & 0.05 \end{bmatrix} \\ &= \begin{bmatrix} 4 & -2.4 & 2 \\ 0 & -3 & 1.5 \\ 0 & 0 & 0.75 \end{bmatrix} \end{aligned} \tag{o}$$

and

$$\mathbf{c} = \mathbf{a}_4^{(2)} = P_2 \mathbf{a}_4^{*(1)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1.4/3 & 1 \end{bmatrix} \begin{bmatrix} 1.6 \\ -7.5 \\ 5 \end{bmatrix} = \begin{bmatrix} 1.6 \\ -7.5 \\ 1.5 \end{bmatrix} \tag{p}$$

Finally, the back-substitution yields the solution

$$\begin{aligned} y_3 = x_2 &= \frac{c_3}{u_{33}} = \frac{1.5}{0.75} = 2 \\ y_2 = x_3 &= \frac{1}{u_{22}}(c_2 - u_{23}y_3) = \frac{1}{u_{22}}(c_2 - u_{23}x_2) = -\frac{1}{3}(-7.5 - 1.5 \times 2) = 3.5 \\ y_1 = x_1 &= \frac{1}{u_{11}}(c_1 - u_{12}y_2 - u_{13}y_3) = \frac{1}{u_{11}}(c_1 - u_{12}x_3 - u_{13}x_2) \\ &= \frac{1}{4}[1.6 - (-2.4) \times 3.5 - 2 \times 2] = 1.5 \end{aligned} \tag{q}$$

which agrees with the solutions obtained earlier without pivoting and with partial pivoting.

6.2 CHOLESKY DECOMPOSITION

As shown in Sec. 4.6, in attempting to reduce the eigenvalue problem for natural conservative systems from one in terms of two real symmetric matrices, with one of the two matrices being positive definite, to one in terms of a single real symmetric matrix, it is necessary to decompose the positive definite matrix into the product of a real nonsingular matrix and its transpose. Then, in Sec. 6.1, we demonstrated that a real arbitrary matrix A can be decomposed by means of Gaussian elimination into the product of a lower and an upper triangular matrix. In this section, we present a technique more efficient than Gaussian elimination, but one restricted to real symmetric positive definite matrices. To this end, we rewrite Eq. (6.42) as

$$A = L'U' \tag{6.43}$$

where L' is a unit lower triangular matrix and U' an upper triangular matrix. In the case in which A is symmetric, Eq. (6.43) can be rewritten in the form

$$A = L'DL'^T \tag{6.44}$$

in which D is a diagonal matrix. In general, the decomposition need not exist. When *A is real symmetric and positive definite*, however, the decomposition is guaranteed to exist and, in addition, the elements of D are all positive. In this case, letting

$$L'D^{1/2} = L \tag{6.45}$$

in which $D^{1/2}$ has the elements $\sqrt{d_{ii}}$, Eq. (6.44) reduces to

$$A = LL^T \tag{6.46}$$

where *L is a unique nonsingular lower triangular matrix with positive diagonal elements.* Equation (6.46) is known as the *Cholesky decomposition*, or *Cholesky factorization.*

The computational algorithm for producing L from the real symmetric positive definite matrix A using the Cholesky decomposition is extremely simple. Indeed, if L is given explicitly by

$$L = \begin{bmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{bmatrix} \tag{6.47}$$

then, by induction, the elements of L can be computed by means of the recursive formulae

$$l_{ii} = \left(a_{ii} - \sum_{j=1}^{i-1} l_{ij}^2\right)^{1/2}, \qquad i = 1, 2, \dots, n \tag{6.48a}$$

$$l_{ki} = \frac{1}{l_{ii}}\left(a_{ik} - \sum_{j=1}^{i-1} l_{ij}l_{kj}\right),$$

$$i = 1, 2, \dots, n; \; k = i+1, i+2, \dots, n \tag{6.48b}$$

Equations (6.48a) and (6.48b) are to be used alternately. For example, the process begins by using Eq. (6.48a) with $i = 1$ to compute l_{11}. Then, having l_{11}, we go to Eq. (6.48b) with $i = 1$ and $k = 2, 3, \dots, n$ to compute $l_{21}, l_{31}, \dots, l_{n1}$. Next, we return to Eq. (6.48a) with $i = 2$ to compute l_{22} and then to Eq. (6.48b) for $l_{32}, l_{42}, \dots l_{n2}$. The process is repeated in the same manner, and it terminates with the use of Eq. (6.48a) with $i = n$ to compute l_{nn}.

As far as efficiency is concerned, the Cholesky decomposition requires $n^3/6$ multiplications, and we recall from Sec. 6.1 that Gaussian elimination requires about $n^3/3$ multiplications. Hence, the Cholesky decomposition is about twice as efficient as the Gaussian elimination. This is to be expected, as the Cholesky decomposition takes full advantage of the symmetry of the matrix A.

The Cholesky decomposition is ideal for reducing an eigenvalue problem in terms of two real symmetric matrices to one defined by a single real symmetric matrix. Of course, this requires that one of the two matrices be positive definite, but this presents no problem, as the mass matrix is in general positive definite. We recall that in Sec. 4.6 we introduced a decomposition of the mass matrix M in the form of Eq. (4.82). Clearly, the developments of Sec. 4.6 were carried out with the Cholesky decomposition in mind, so that the matrix Q in Eqs. (4.82), (4.84), (4.85) and (4.87) should be replaced by L^T.

It should be reiterated that the Cholesky decomposition of a real symmetric matrix is possible only if the matrix is positive definite. In vibrations, positive definiteness can often be ascertained on physical grounds, as in the case of the mass matrix. In other cases, it can be ascertained by means of *Sylvester's criterion*, which states that *a real symmetric matrix is positive definite if and only if all its principal minor determinants are positive* (Ref. 10, p. 94). However, because the evaluation of an $n \times n$ determinant requires $n!$ multiplications, application of the criterion is feasible only for matrices of relatively small order.

Example 6.3

Verify that the real symmetric matrix

$$A = \begin{bmatrix} 6.25 & 3.25 & 2.875 \\ 3.25 & 15.7525 & 10.495 \\ 2.875 & 10.495 & 9.3325 \end{bmatrix} \tag{a}$$

is positive definite. Then, compute the matrix L for the Cholesky decomposition.

The principal minor determinants of A are

$$\begin{aligned} \Delta_1 &= a_{11} = 6.25 \\ \Delta_2 &= a_{11}a_{22} - a_{12}^2 = 6.25 \times 15.7525 - 3.25^2 = 87.8906 \\ \Delta_3 &= a_{11}(a_{22}a_{33} - a_{23}^2) - a_{12}(a_{12}a_{33} - a_{13}a_{23}) + a_{13}(a_{12}a_{23} - a_{13}a_{22}) \\ &= 6.25(15.7525 \times 9.3325 - 10.495^2) - 3.25(3.25 \times 9.3325 - 2.875 \times 10.495) \\ &\quad + 2.875(3.25 \times 10.495 - 2.875 \times 15.7525) = 197.7540 \end{aligned} \tag{b}$$

All principal minor determinants are positive, so that the matrix A is indeed positive definite.

We begin the decomposition process by letting $i = 1$ in Eq. (6.48a) and writing

$$l_{11} = \sqrt{a_{11}} = \sqrt{6.25} = 2.5 \tag{c}$$

Then, from Eq. (6.48b) with $i = 1$ and $k = 2$ and $k = 3$, we obtain

$$\begin{aligned} l_{21} &= \frac{1}{l_{11}}a_{12} = \frac{3.25}{2.5} = 1.3 \\ l_{31} &= \frac{1}{l_{11}}a_{13} = \frac{2.875}{2.5} = 1.15 \end{aligned} \tag{d}$$

At this point, we return to Eq. (6.48a) with $i = 2$ and write

$$l_{22} = \sqrt{a_{22} - l_{21}^2} = \sqrt{15.7525 - 1.3^2} = 3.75 \tag{e}$$

so that, from Eq. (6.48b) with $i = 2$ and $k = 3$, we have

$$l_{32} = \frac{1}{l_{22}}(a_{23} - l_{21}l_{31}) = \frac{1}{3.75}(10.495 - 1.3 \times 1.15) = 2.4 \tag{f}$$

Finally, from Eq. (6.48a) with $i = 3$, we obtain

$$l_{33} = \sqrt{a_{33} - l_{31}^2 - l_{32}^2} = \sqrt{9.3325 - 1.15^2 - 2.4^2} = 1.5 \tag{g}$$

Hence, the desired matrix is

$$L = \begin{bmatrix} 2.5 & 0 & 0 \\ 1.3 & 3.75 & 0 \\ 1.15 & 2.4 & 1.5 \end{bmatrix} \tag{h}$$

6.3 THE POWER METHOD FOR SYMMETRIC EIGENVALUE PROBLEMS

As pointed out in Chapter 4, for $n > 2$, the solution of the algebraic eigenvalue problem is a numerical problem. There are many algorithms for solving the eigenvalue problem and they all have one thing in common, they are all essentially *iterative* in nature. To introduce the idea, we consider a given problem with the solution x. In seeking a solution by an iterative method, we begin with a guess x_0 and compute a sequence of improved guesses $x_1, x_2, \ldots$. The iteration *converges* if, for every initial guess x_0, the sequence $x_1, x_2, \ldots$ tends to the true solution x, although the true solution itself is never reached. The fact that a method is known to converge is reassuring, but convergence alone is not enough. Indeed, one of the deciding factors in choosing a given method is the *rate of convergence*. There are certain definitions characterizing the rate of convergence. In *linear convergence* every step multiplies the error by a fixed factor $r < 1$, and the number of accurate significant figures increases by a constant amount at each iteration step. *Quadratic convergence* is characterized by the fact that the error is squared at every step, going from 10^{-1} to 10^{-2} to $10^{-4} \ldots$, or the number of accurate significant figures doubles at each step. In *cubic convergence* the error is cubed at every step, going from 10^{-1} to 10^{-3} to $10^{-9} \ldots$, or the number of accurate significant figures triples at each step. Convergence is a much more complex concept than the preceding definitions seem to imply, and on many occasions it does not lend itself to easy analysis or classification. Moreover, even when one of the definitions does apply, it may not necessarily apply over the entire iteration process.

We begin our discussion of computational algorithms for the eigenvalue problem with the *power method*, perhaps the most widely known of the iteration procedures. The power method can be used both for symmetric and nonsymmetric eigenvalue problems. Whereas the general ideas are the same in both cases, the details differ. Hence, we discuss the two cases separately, the symmetric eigenvalue problem in this section and the nonsymmetric one in Sec. 6.13.

We consider a real symmetric matrix A of order n and write the eigenvalue problem in the standard form

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \qquad i = 1, 2, \ldots, n \tag{6.49}$$

where the eigenvalues are ordered so that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$, in which λ_1 is referred to as the *dominant eigenvalue*. We assume that the n eigenvectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, n)$ are linearly independent and that they span an n-dimensional vector space. Hence, by the expansion theorem (Sec. 4.6) any arbitrary vector $\mathbf{v}_0$ in that space can be expressed as the linear combination

$$\mathbf{v}_0 = \sum_{i=1}^{n} c_i \mathbf{x}_i \tag{6.50}$$

The power method iteration process is defined by

$$\mathbf{v}_p = A\mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.51}$$

where

$$\mathbf{v}_p = A\mathbf{v}_{p-1} = A^2\mathbf{v}_{p-2} = \ldots = A^p\mathbf{v}_0 \tag{6.52}$$

so that the iteration process amounts to raising the matrix A to the indicated powers, although A^p is never computed explicitly. Inserting Eq. (6.50) into Eq. (6.51), using Eq. (6.49) repeatedly and factoring out λ_1^p, we obtain

$$\mathbf{v}_p = \lambda_1^p \left[c_1\mathbf{x}_1 + \sum_{i=2}^{n} c_i \left(\frac{\lambda_i}{\lambda_1} \right)^p \mathbf{x}_i \right] \tag{6.53}$$

On the assumption that $c_1 \neq 0$ and recalling that λ_1 is the largest eigenvalue, for sufficiently large p, Eq. (6.53) can be written in the form

$$\mathbf{v}_p = \lambda_1^p \left(c_1\mathbf{x}_1 + \boldsymbol{\epsilon}_p \right) \tag{6.54}$$

where $\boldsymbol{\epsilon}_p$ is a vector with small components. In fact, as $p \to \infty$, the vector $\boldsymbol{\epsilon}_p$ approaches the null vector asymptotically, so that

$$\lim_{p\to\infty} \mathbf{v}_p = \lambda_1^p c_1 \mathbf{x}_1 \tag{6.55}$$

Hence, the iteration process converges to the eigenvector $\mathbf{x}_1$ belonging to the dominant eigenvalue λ_1. An additional premultiplication by A yields

$$\lim_{p\to\infty} \mathbf{v}_{p+1} = \lambda_1^{p+1} c_1 \mathbf{x}_1 \tag{6.56}$$

and we note that convergence is characterized by the fact that two consecutive iterates, $\mathbf{v}_p$ and $\mathbf{v}_{p+1}$, are proportional to each other, where the constant of proportionality is λ_1. Indeed, Eqs. (6.55) and (6.56) can be used to write

$$\lambda_1 = \lim_{p\to\infty} \frac{v_{p+1,j}}{v_{p,j}} \tag{6.57}$$

where the subscript j indicates the jth component of the iterate. Hence, after achieving convergence, the dominant eigenvalue can be obtained as the ratio of two homologous (having the same relative position) components corresponding to two consecutive iterates.

We observe from Eq. (6.54) that, as p increases, the components of the vector $\mathbf{v}_p$ tend to become progressively large for $\lambda_1 > 1$ and progressively small for $\lambda_1 < 1$, which can be a problem. To obviate this problem, it is advisable in practice to modify slightly the iteration process, Eq. (6.51), by scaling the iterates so as to moderate the changes in magnitude, which implies normalization. To this end, we replace Eq. (6.51) by the iteration process

$$\mathbf{v}_p^* = A\mathbf{v}_{p-1}, \qquad \mathbf{v}_p = \alpha_p \mathbf{v}_p^*, \qquad p = 1, 2, \ldots \tag{6.58a, b}$$

where α_p is a normalization factor. If the iterates $\mathbf{v}_p^*$ are normalized so as to render the component of $\mathbf{v}_p$ of largest magnitude equal to 1, then

$$\alpha_p = \frac{1}{\max\limits_i |v_{pi}^*|}, \qquad p = 1, 2, \ldots \tag{6.59}$$

On the other hand, if the vectors $\mathbf{v}_p^*$ are normalized so that the magnitude of $\mathbf{v}_p$ is equal to 1, then

$$\alpha_p = 1/\|\mathbf{v}_p^*\|, \qquad p = 1, 2, \ldots \tag{6.60}$$

where $\|\mathbf{v}_p^*\|$ is the Euclidean norm of $\mathbf{v}_p^*$. In both cases, the convergence can be expressed in the simple form

$$\lim_{p\to\infty} \alpha_p = \lambda_1 \tag{6.61a}$$

and

$$\lim_{p\to\infty} \mathbf{v}_p = \mathbf{x}_1 \tag{6.61b}$$

where the eigenvector $\mathbf{x}_1$ is normalized according to one of the two schemes just described. In fact, when unit eigenvectors are required, as we shall see shortly, a good strategy is to normalize according to Eq. (6.59) during the iteration process and switch to Eq. (6.60) after convergence has been reached.

In the above process, we tacitly assumed that $\lambda_2 \neq \lambda_1$. In the case in which the dominant eigenvalue λ_1 has multiplicity m, Eq. (6.53) must be replaced by

$$\mathbf{v}_p = \lambda_1^p \left[\sum_{i=1}^{m} c_i \mathbf{x}_i + \sum_{i=m+1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^p \mathbf{x}_i \right] \tag{6.62}$$

and the iterates tend to some vector lying in the subspace spanned by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$. The problem of determining the remaining $m - 1$ eigenvectors is discussed shortly.

The question can be raised as to whether the iteration process fails to converge to the dominant eigenvalue, eigenvector pair $\lambda_1, \mathbf{x}_1$ if $c_1 = 0$ in Eq. (6.50), in which case the eigenvector is not represented in $\mathbf{v}_0$, and hence in the iterates. Whereas it may be possible in theory to conjure up an example in which $\mathbf{x}_1$ is absent from $\mathbf{v}_0$, in carrying out the iteration process on a digital computer this possibility does not exist, as digital computers do not perform arithmetical operations exactly. Indeed, a number stored in a digital computer has only a given number of significant figures, which tends to introduce rounding errors. As a result, the initial guess $\mathbf{v}_0$ and the iterates $\mathbf{v}_p$ are likely to acquire a component of $\mathbf{x}_1$. This component, even if extremely small in the beginning, tends to grow larger and larger progressively and finally assert itself essentially as the only component in the iterate. This is a reassuring thought as far as convergence to the dominant eigensolution is concerned, but the thought raises questions as to how to obtain the subdominant eigensolutions. In fact, it is not possible to converge to subdominant eigensolutions with the iteration process specified by Eq. (6.51), so that the process must be modified.

We assume that the dominant eigensolution $\lambda_1, \mathbf{x}_1$ satisfying Eq. (6.49) has been determined, and that the eigenvectors are to be normalized so as to satisfy the orthonormality conditions

$$\mathbf{x}_i^T \mathbf{x}_j = \delta_{ij}, \qquad i, j = 1, 2, \ldots, n \tag{6.63}$$

Then, we consider the matrix

$$A_2 = A - \lambda_1 \mathbf{x}_1 \mathbf{x}_1^T \tag{6.64}$$

Multiplying Eq. (6.64) on the right by $\mathbf{x}_i$ and considering Eqs. (6.63), we obtain

$$A_2\mathbf{x}_i = A\mathbf{x}_i - \lambda_1 \mathbf{x}_1 \mathbf{x}_1^T \mathbf{x}_i = \lambda_i \mathbf{x}_i - \delta_{i1}\lambda_1 \mathbf{x}_1 = \begin{cases} \mathbf{0} & \text{if } i = 1 \\ \lambda_i \mathbf{x}_i & \text{if } i \neq 1 \end{cases} \tag{6.64}$$

Equation (6.64) permits us to conclude that the matrix A_2 possesses the eigenvalues $0, \lambda_2, \ldots, \lambda_n$ and the eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. Hence, using the initial vector in the form given by Eq. (6.50) in conjunction with the matrix A_2, we obtain the first iterate

$$\mathbf{v}_1 = A_2\mathbf{v}_0 = \sum_{i=1}^{n} c_i A_2 \mathbf{x}_i = \sum_{i=2}^{n} c_i \lambda_i \mathbf{x}_i \tag{6.65}$$

which is entirely free of $\mathbf{x}_1$. It follows that, if we use the iteration process

$$\mathbf{v}_p = A_2 \mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.66}$$

then the pth iterate has the form

$$\mathbf{v}_p = \lambda_2^p \left[c_2\mathbf{x}_2 + \sum_{i=3}^{n} c_i \left(\frac{\lambda_i}{\lambda_2}\right)^p \mathbf{x}_i \right] \tag{6.67}$$

so that the iteration process involving A_2 converges to $\lambda_2, \mathbf{x}_2$ in the same way as the iteration process involving A converges to $\lambda_1, \mathbf{x}_1$. The matrix A_2 is known as a *deflated matrix*. As in the case of the dominant pair $\lambda_1, \mathbf{x}_1$, we actually use the iteration process

$$\mathbf{v}_p^* = A_2\mathbf{v}_{p-1}, \qquad \mathbf{v}_p = \alpha_p \mathbf{v}_p^*, \qquad p = 1, 2, \ldots \tag{6.68}$$

where, in view of the fact that the eigenvectors must be normalized so as to satisfy Eqs. (6.63), the normalization factor α_p is determined according to Eq. (6.60).

At this point, we return to the question of repeated eigenvalues. In the case examined earlier in which λ_1 has multiplicity m, the iteration process involving the deflated matrix A_2, Eq. (6.66), remains the same, but the expression for the pth iterate, Eq. (6.67), must be replaced by

$$\mathbf{v}_p = \lambda_1^p \left[\sum_{i=1}^{m} c_i \mathbf{x}_i + \sum_{i=m+1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^p \mathbf{x}_i \right] \tag{6.69}$$

so that the process converges to λ_1, $\mathbf{x}_2$. Here, once again, we use the iteration process given by Eqs. (6.68), instead of that given by Eqs. (6.66), so that $\mathbf{x}_2$ is automatically orthonormal to $\mathbf{x}_1$.

Matrix deflation can be used to compute the remaining subdominant eigensolutions. Indeed, it is easy to verify that the deflated matrix

$$A_k = A_{k-1} - \lambda_{k-1}\mathbf{x}_{k-1}\mathbf{x}_{k-1}^T, \qquad k = 2, 3, \ldots, n \tag{6.70}$$

has the eigenvalues $0, 0, \ldots, 0, \lambda_k, \lambda_{k+1}, \ldots, \lambda_n$ and the eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots,, \mathbf{x}_{k-1}, \mathbf{x}_k,, \mathbf{x}_{k+1}, \ldots,, \mathbf{x}_n$, respectively, so that the deflated matrix A_k can be used to iterate to the eigensolution $\lambda_k, \mathbf{x}_k$. In the case in which λ_1 has multiplicity m and $k < m$, the eigenvalues are $0, 0, \ldots, 0, \lambda_1, \lambda_1, \ldots, \lambda_1, \lambda_{m+1}, \ldots, \lambda_n$ and the eigenvectors are $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_k, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_m, \mathbf{x}_{m+1}, \ldots, \mathbf{x}_n$, respectively. Of course, this does not affect the algorithm itself, which remains the same, regardless of eigenvalue multiplicity. It should be pointed out that no iteration is really required for $k = n$, as $n - 1$ eigenvectors are sufficient to determine $\mathbf{x}_n$ by means of Eqs. (6.63). Then, λ_n can be computed using Eq. (6.49) with $i = n$. Still, whereas this information has educational value, it is simpler to complete the solution with $k = n$. The procedure for determining the subdominant eigensolutions using matrix deflation is due to Hotelling (Ref. 6). Various other deflation procedures are presented by Wilkinson (Ref. 13).

Next, we propose to make the connection between the developments of this section for the eigenvalue problem in terms of a single real symmetric matrix and the problem in terms of two real symmetric matrices of interest in vibrations. Using Eqs. (4.81) and (4.131), the eigenvalue problem can be written in the form

$$K\mathbf{u} = \omega^2 M\mathbf{u} \tag{6.71}$$

where K and M are the symmetric stiffness and mass matrices, respectively, $\mathbf{u}$ is a vector of displacement amplitudes and ω is the frequency of vibration. In reducing the eigenvalue problem (6.71) to standard form, it is important to keep in mind that the most accurate eigenvalue, eigenvector pair computed by means of the power method is the dominant pair $\lambda_1, \mathbf{x}_1$. But, in vibrations, it is generally the lowest natural frequency ω_1 holding the most interest. Hence, Eq. (6.71) must be reduced to standard form in such a way that λ is inversely proportional to ω^2. Another thing to consider is that the symmetry of the coefficient matrix A is not really required for the power method. Indeed, only the orthonormality of the eigenvectors is required, and only for the computation of the subdominant eigensolutions. In view of this, we premultiply both sides of Eq. (6.71) by K^{-1} and rewrite the result in the form

$$A\mathbf{u} = \lambda\mathbf{u}, \qquad \lambda = 1/\omega^2 \tag{6.72}$$

in which

$$A = K^{-1}M \tag{6.73}$$

and it is clear that this reduction is possible only if K is nonsingular. Note that the matrix given by Eq. (6.73) is commonly referred to as the *dynamical matrix* (Refs. 9 and 11).

The computation of the dominant mode follows the established procedure, with the initial trial vector $\mathbf{v}_0$ in the form of the linear combination

$$\mathbf{v}_0 = \sum_{i=1}^{n} c_i \mathbf{u}_i \tag{6.74}$$

where $\mathbf{u}_i$ $(i = 1, 2, \ldots, n)$ are the modal vectors. Then, assuming that the modal vectors are to be normalized so as to satisfy the orthonormality relations

$$\mathbf{u}_i^T M \mathbf{u}_j = \delta_{ij}, \qquad i, j = 1, 2, \ldots, n \tag{6.75}$$

we can construct the deflated matrices for iteration to the subdominant modes in the form

$$A_k = A_{k-1} - \lambda_{k-1}\mathbf{u}_{k-1}\mathbf{u}_{k-1}^T M, \qquad k = 2, 3, \ldots, n \tag{6.76}$$

When the stiffness matrix is positive semidefinite only, the matrix is singular and certain vectors $\mathbf{u}_i$ $(i = 1, 2, \ldots, r)$ satisfy the equation

$$K\mathbf{u}_i = \mathbf{0}, \qquad i = 1, 2, \ldots r \tag{6.77}$$

These vectors can be identified as *rigid-body modes* and, because $M_i\mathbf{u} \neq \mathbf{0}$, they are characterized by zero frequencies

$$\omega_i = 0, \qquad i = 1, 2, \ldots, r \tag{6.78}$$

The eigenvalue problem can be modified in this case by eliminating the rigid-body modes from the formulation, resulting in an eigenvalue problem defined by positive definite, real symmetric mass and stiffness matrices (Ref. 11).

The main advantage of the power method is simplicity. Two other advantages are that the algorithm yields eigenvalues and eigenvectors simultaneously and that it iterates to one eigensolution at a time, thus providing a partial solution if desired. A fourth advantage is that the method is able to accommodate very large sparse matrices A by storing only the nonzero elements, instead of the full array of n^2 elements. This advantage may not be as significant as it may seem, because it is the stiffness matrix K that tends to be banded, and Eq. (6.73) calls for the flexibility matrix K^{-1}, which tends to be fully populated. On the other side of the ledger, there is the question of convergence. There are two factors affecting convergence. The first is the choice of the initial vector $\mathbf{v}_0$. If $\mathbf{v}_0$ is relatively close to the first eigenvector $\mathbf{x}_1$, then the coefficient c_1 in Eq. (6.50) is much larger than the remaining coefficients, which tends to reduce the number of iteration steps. This factor is not as significant as it would seem. Indeed, far more important is the second factor, namely, the ratio λ_2/λ_1, which represents a characteristic of the system. Clearly, the smaller is the ratio, the faster is the convergence. Convergence problems can be expected if λ_1 is not strongly dominant. In particular, if λ_2/λ_1 is close to 1, then convergence will be painfully slow. In this case, convergence can be accelerated through a shift in the eigenvalues, which can be accomplished by replacing the matrix A in Eq. (6.58a) by $A - \mu I$. Now the process converges to $\lambda_1 - \mu$ and $\mathbf{x}_1$, and the rate of convergence depends on the ratio $(\lambda_2 - \mu) / (\lambda_1 - \mu)$. Note that the value of μ can be changed with each iteration step. A judicious choice of μ can accelerate convergence dramatically. Some choices are discussed by Wilkinson (Ref. 13, p. 572), but not all choices are desirable, as they tend to complicate the iteration process and cause programming difficulties. The procedure can be extended to the subdominant eigensolutions.

The power method has pedagogical value, but is not a serious contender as a computational algorithm for solving the algebraic eigenvalue problem, except when only a few dominant eigensolutions are required.

Example 6.4

Solve the eigenvalue problem for the system of Example 4.6 by the power method.

From Example 4.6, we obtain the mass and stiffness matrices

$$M = m\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad K = k\begin{bmatrix} 5 & -3 & 0 \\ -3 & 5 & -2 \\ 0 & -2 & 3 \end{bmatrix} \tag{a}$$

so that, using Eq. (6.73), the dynamical matrix can be verified to be

$$A = K^{-1}M = \begin{bmatrix} 0.7857 & 0.9643 & 0.2143 \\ 0.6429 & 1.6071 & 0.3571 \\ 0.4286 & 1.0714 & 0.5714 \end{bmatrix} \tag{b}$$

where the parameters m and k have been assigned to λ, so that

$$\lambda = k/m\omega^2 \tag{c}$$

We begin the iteration process with $\mathbf{v}_0 = [1\ 1\ 1]^T$. Introducing $\mathbf{v}_0$ into Eq. (6.58a) and using Eqs. (6.58b) and (6.59), all with $p = 1$, we obtain

$$\begin{bmatrix} 0.7857 & 0.9643 & 0.2143 \\ 0.6429 & 1.6071 & 0.3571 \\ 0.4286 & 1.0714 & 0.5714 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1.9643 \\ 2.6071 \\ 2.0714 \end{bmatrix} = 2.6071\begin{bmatrix} 0.7534 \\ 1.0000 \\ 0.7945 \end{bmatrix} \tag{d}$$

so that $\alpha_1 = 2.6071$ and $\mathbf{v}_1 = [0.7534\ 1.0000\ 0.7945]^T$. Repeating the process with $p = 2$, we write

$$\begin{bmatrix} 0.7857 & 0.9643 & 0.2143 \\ 0.6429 & 1.6071 & 0.3571 \\ 0.4286 & 1.0714 & 0.5714 \end{bmatrix}\begin{bmatrix} 0.7534 \\ 1.0000 \\ 0.7945 \end{bmatrix} = 2.3752\begin{bmatrix} 0.7269 \\ 1.0000 \\ 0.7782 \end{bmatrix} \tag{e}$$

The seventh iteration yields

$$\begin{bmatrix} 0.7857 & 0.9643 & 0.2143 \\ 0.6429 & 1.6071 & 0.3571 \\ 0.4286 & 1.0714 & 0.5714 \end{bmatrix}\begin{bmatrix} 0.7231 \\ 1.0000 \\ 0.7769 \end{bmatrix} = 2.3495\begin{bmatrix} 0.7231 \\ 1.0000 \\ 0.7769 \end{bmatrix} \tag{f}$$

so that

$$\lambda_1 = 2.3495, \qquad \mathbf{u}_1 = \begin{bmatrix} 0.7231 \\ 1.0000 \\ 0.7769 \end{bmatrix} \tag{g}$$

Using Eq. (c) and normalizing according to Eq. (6.75), we obtain the lowest natural frequency and modal vector

$$\omega_1 = 0.6524\sqrt{\frac{k}{m}}, \qquad \mathbf{u}_1 = m^{-1/2}\begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix} \tag{h}$$

To compute the second eigenvalue and eigenvector, we use Eq. (6.76) with $k = 2$ and construct the first deflated matrix

$$\begin{aligned} A_2 &= A - \lambda_1 \mathbf{u}_1 \mathbf{u}_1^T M \\ &= \begin{bmatrix} 0.7857 & 0.9643 & 0.2143 \\ 0.6429 & 1.6071 & 0.3571 \\ 0.4286 & 1.0714 & 0.5714 \end{bmatrix} - 2.3495 \begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix} \begin{bmatrix} 0.3354 \\ 0.4638 \\ 0.3603 \end{bmatrix}^T \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0.2572 & -0.1319 & -0.0696 \\ -0.0880 & 0.0911 & -0.0354 \\ -0.1392 & -0.1063 & 0.2664 \end{bmatrix} \end{aligned} \tag{i}$$

Then, using as a first trial vector $\mathbf{v}_0 = [1\ 0\ -1]^T$, we obtain

$$\begin{bmatrix} 0.2572 & -0.1319 & -0.0696 \\ -0.0880 & 0.0911 & -0.0354 \\ -0.1392 & -0.1063 & 0.2664 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 0.3268 \\ -0.0526 \\ -0.4056 \end{bmatrix} = 0.4056 \begin{bmatrix} 0.8057 \\ -0.1297 \\ -1.0000 \end{bmatrix} \tag{j}$$

so that $\alpha_1 = 0.4056$ and $\mathbf{v}_1 = [0.8057\ -0.1297\ -1.0000]^T$. Repeating the process with $p = 2$, we have

$$\begin{bmatrix} 0.2572 & -0.1319 & -0.0696 \\ -0.0880 & 0.0911 & -0.0354 \\ -0.1392 & -0.1063 & 0.2644 \end{bmatrix} \begin{bmatrix} 0.8057 \\ -0.1297 \\ -1.0000 \end{bmatrix} = 0.3648 \begin{bmatrix} 0.8057 \\ -0.1297 \\ -1.0000 \end{bmatrix} \tag{k}$$

and we conclude that convergence has been achieved already, which can be attributed to an extremely good guess for $\mathbf{v}_0$. Hence,

$$\lambda_2 = 0.3648, \qquad \mathbf{u}_2 = \begin{bmatrix} 0.8057 \\ -0.1297 \\ -1.0000 \end{bmatrix} \tag{l}$$

Using Eq. (c) and normalizing according to Eq. (6.75), we obtain the second natural frequency and modal vector

$$\omega_2 = 1.6556\sqrt{\frac{k}{m}}, \qquad \mathbf{u}_2 = m^{-1/2} \begin{bmatrix} 0.5258 \\ -0.0845 \\ -0.6525 \end{bmatrix} \tag{m}$$

The third and last eigenvector does not really require an iterative process, as $\mathbf{u}_3$ can be computed from the requirement that it be orthogonal with respect to M to $\mathbf{u}_1$ and $\mathbf{u}_2$. Nevertheless, it is simpler to continue with the iteration process. To this end, we use Eq. (6.76) with $k = 3$ and compute the second deflated matrix

$$\begin{aligned} A_3 = A_2 - \lambda_2 \mathbf{u}_2 \mathbf{u}_2^T M &= \begin{bmatrix} 0.2572 & -0.1319 & -0.0696 \\ -0.0880 & 0.0911 & -0.0354 \\ -0.1392 & -0.1063 & 0.2664 \end{bmatrix} \\ &\quad - 0.3648 \begin{bmatrix} 0.5258 \\ -0.0845 \\ -0.6525 \end{bmatrix} \begin{bmatrix} 0.5258 \\ -0.0845 \\ -0.6525 \end{bmatrix}^T \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 0.0556 & -0.0833 & 0.0556 \\ -0.0556 & 0.0833 & -0.0556 \\ 0.1111 & -0.1667 & 0.1111 \end{bmatrix} \end{aligned} \tag{n}$$

As a first trial vector, we use $\mathbf{v}_0 = [1 \ -1 \ 1]^T$ and write

$$\begin{bmatrix} 0.0556 & -0.0833 & 0.0556 \\ -0.0556 & 0.0833 & -0.0556 \\ 0.1111 & -0.1667 & 0.1111 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.1944 \\ -0.1944 \\ 0.3888 \end{bmatrix} = 0.3888 \begin{bmatrix} 0.5000 \\ -0.5000 \\ 1.0000 \end{bmatrix} \tag{o}$$

A second step yields

$$\begin{bmatrix} 0.0556 & -0.0833 & 0.0556 \\ -0.0556 & 0.0833 & -0.0556 \\ 0.1111 & -0.1667 & 0.1111 \end{bmatrix} \begin{bmatrix} 0.5000 \\ -0.5000 \\ 1.0000 \end{bmatrix} = 0.2500 \begin{bmatrix} 0.5000 \\ -0.5000 \\ 1.0000 \end{bmatrix} \tag{p}$$

so that

$$\lambda_3 = 0.2500, \qquad \mathbf{u}_3 = \begin{bmatrix} 0.5000 \\ -0.5000 \\ 1.0000 \end{bmatrix} \tag{q}$$

from which we obtain the third natural frequency and modal vector

$$\omega_3 = 2.0000\sqrt{\frac{k}{m}}, \qquad \mathbf{u}_3 = m^{-1/2} \begin{bmatrix} 0.3333 \\ -0.3333 \\ 0.6667 \end{bmatrix} \tag{r}$$

thus completing the solution.

6.4 THE JACOBI METHOD

In Sec. 5.1, we presented a geometric interpretation of the eigenvalue problem for real symmetric positive definite $n \times n$ matrices A whereby the solution of the eigenvalue problem was demonstrated to be equivalent to the problem of determining the principal axes of an n-dimensional ellipsoid defined by the equation

$$f = \mathbf{x}^T A \mathbf{x} = 1 \tag{6.79}$$

where $\mathbf{x}$ is an n-dimensional vector.

In the two-dimensional case, the ellipsoid reduces to an ellipse and the problem of determining its principal axes reduces to a coordinate transformation representing a rotation of axes. Analytically, the process amounts to reducing the equation of the ellipse to canonical form, which is equivalent to diagonalizing the matrix A, or

$$R^T A R = D \tag{6.80}$$

where R is the rotation matrix and D is the diagonal matrix. It is demonstrated in Sec. 5.1 that D is simply the diagonal matrix Λ of eigenvalues and R is the orthonormal matrix V of eigenvectors. In this section, we extend these ideas to the n-dimensional case.

In the two-dimensional case, the principal axes can be determined by means of a single rotation. In the n-dimensional case, it is not possible to determine the principal axes in a single step. Indeed, one planar rotation can be used to annihilate the off-diagonal element of A in the plane of rotation and its symmetric counterpart. But, a second rotation designed to annihilate another off-diagonal element and its symmetric counterpart in a different plane will cause the just annihilated element and its symmetric counterpart to acquire some nonzero value, albeit smaller than the original one, i.e., smaller than the value before annihilation. It follows that the diagonalization of A cannot be carried out in a finite number of steps and can only be performed iteratively, whereby each iteration represents one planar rotation. This is the essence of the *Jacobi method*, for which reason the Jacobi method is referred to as *diagonalization by successive rotations*. The method converges but in theory the number of rotations is infinite, as in any iterative process. In practice, it is finite and depends on the accuracy desired.

The iteration process is given by

$$A_k = R_k^T A_{k-1} R_k, \qquad k = 1, 2, \dots \tag{6.81}$$

where $A = A_0$ and R_k is the kth rotation matrix. Assuming that our objective is to annihilate the off-diagonal elements p, q and q, p of the matrix A_k, by analogy with Eq. (5.8), the matrix R_k is taken to represent a rotation in the p, q-plane and has the form

$$R_k = \begin{array}{cc}
\begin{array}{cccccccc} & & & p & & q & & \end{array} & \\
\left[\begin{array}{cccccccc}
1 & 0 & \dots & 0 & \dots & 0 & \dots & 0 \\
0 & 1 & \dots & 0 & \dots & 0 & \dots & 0 \\
\multicolumn{8}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
0 & 0 & \dots & \cos\theta_k & \dots & -\sin\theta_k & \dots & 0 \\
\multicolumn{8}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
0 & 0 & \dots & \sin\theta_k & \dots & \cos\theta_k & \dots & 0 \\
\multicolumn{8}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
0 & 0 & \dots & 0 & \dots & 0 & \dots & 1
\end{array}\right] &
\begin{array}{c} \\ \\ \\ p \\ \\ q \\ \\ \\ \end{array}
\end{array} \tag{6.82}$$

Denoting the i, j element of A_k by $a_{ij}^{(k)}$ and inserting Eq. (6.82) into Eq. (6.81), we can express the elements of A_k in terms of the elements of A_{k-1} and the rotation angle θ_k as follows:

$$a_{pp}^{(k)} = a_{pp}^{(k-1)} \cos^2\theta_k + 2a_{pq}^{(k-1)} \sin\theta_k \cos\theta_k + a_{qq}^{(k-1)} \sin^2\theta_k \tag{6.83a}$$

$$a_{qq}^{(k)} = a_{pp}^{(k-1)} \sin^2\theta_k - 2a_{pq}^{(k-1)} \sin\theta_k \cos\theta_k + a_{qq}^{(k-1)} \cos^2\theta_k \tag{6.83b}$$

$$\begin{aligned} a_{pq}^{(k)} = a_{qp}^{(k)} &= -\left(a_{pp}^{(k-1)} - a_{qq}^{(k-1)}\right) \sin\theta_k \cos\theta_k \\ &\quad + a_{pq}^{(k-1)} \left(\cos^2\theta_k - \sin^2\theta_k\right) \end{aligned} \tag{6.83c}$$

$$a_{ip}^{(k)} = a_{pi}^{(k)} = a_{ip}^{(k-1)} \cos\theta_k + a_{iq}^{(k-1)} \sin\theta_k, \qquad i \neq p, q \tag{6.83d}$$

$$a_{iq}^{(k)} = a_{qi}^{(k)} = -a_{ip}^{(k-1)} \sin\theta_k + a_{iq}^{(k-1)} \cos\theta_k, \qquad i \neq p, q \tag{6.83e}$$

$$a_{ij}^{(k)} = a_{ij}^{(k-1)}, \qquad i, j \neq p, q \tag{6.83f}$$

so that the only elements affected by the orthonormal transformation (6.80) are those in row and column p and in row and column q. From Eq. (6.83), we conclude that, to render the element $a_{pq}^{(k)}$ zero, the rotation angle θ_k must be chosen so as to satisfy

$$\tan 2\theta_k = \frac{2a_{pq}^{(k-1)}}{a_{pp}^{(k-1)} - a_{qq}^{(k-1)}} \tag{6.84}$$

But, from Eqs. (6.83), we observe that only $\sin\theta_k$ and $\cos\theta_k$ are required explicitly, and not θ_k. Introducing the notation

$$a_{pq}^{(k-1)} = b_{k-1}, \qquad \frac{1}{2}\left(a_{pp}^{(k-1)} - a_{qq}^{(k-1)}\right) = c_{k-1} \tag{6.85}$$

it is not difficult to verify that

$$\cos\theta_k = \left[\frac{1}{2} + \frac{c_{k-1}}{2\left(b_{k-1}^2 + c_{k-1}^2\right)^{1/2}}\right]^{1/2}, \qquad \sin\theta_k = \frac{b_{k-1}}{2\left(b_{k-1}^2 + c_{k-1}^2\right)^{1/2}\cos\theta_k} \tag{6.86}$$

where $\cos\theta_k$ is to be taken to be positive and $\sin\theta_k$ takes the sign of $\tan 2\theta_k$. When $a_{pp}^{(k-1)} = a_{qq}^{(k-1)}$, we take θ_k to be $\pm\pi/4$, according to the sign of $a_{pq}^{(k-1)}$. Because in general $a_{pq}^{(k+1)} \neq 0$, the process is iterative.

From Eqs. (6.81), we can write

$$\begin{aligned} A_k &= R_k^T A_{k-1} R_k = R_k^T R_{k-1}^T A_{k-2} R_{k-1} R_k = \dots \\ &= R_k^T R_{k-1}^T \dots R_2^T R_1^T A R_1 R_2 \dots R_{k-1} R_k \end{aligned} \tag{6.87}$$

The process is convergent, in the sense that A_k tends to a diagonal matrix as $k \to \infty$. By necessity, this diagonal matrix must be the matrix Λ of the eigenvalues, so that

$$\lim_{k\to\infty} A_k = \Lambda \tag{6.88}$$

Moreover, comparing Eq. (6.87) to Eq. (4.107) and considering Eq. (6.88), we conclude that

$$\lim_{k\to\infty} R_1 R_2 \dots R_{k-1} R_k = V \tag{6.89}$$

where V is the matrix of the eigenvectors, and we observe that, because every one of the rotation matrices is orthonormal, *the Jacobi method produces automatically an orthonormal matrix of eigenvectors.* It should be pointed out that the Jacobi method yields a complete solution of the eigenvalue problem, in contrast to the power method, which is capable of producing a partial solution, as well as a complete solution.

Next, we wish to prove convergence of the Jacobi method. To this end, we first express the matrix A_k as the sum

$$A_k = D_k + U_k + U_k^T \tag{6.90}$$

where D_k is the matrix of the diagonal elements of A_k and U_k is the upper triangular matrix of the off-diagonal elements of A_k. Then, we define the Euclidean norm of a matrix A as (see Appendix B)

$$\|A\|_E = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2\right)^{1/2} \tag{6.91}$$

In view of this definition, the Euclidean norm squared of A_k is simply

$$\|A_k\|_E^2 = \|D_k\|_E^2 + 2\|U_k\|_E^2 \tag{6.92}$$

But, Eqs. (6.83a)–(6.83c) can be used to show that

$$\left(a_{pp}^{(k)}\right)^2 + \left(a_{qq}^{(k)}\right)^2 + 2\left(a_{pq}^{(k)}\right)^2 = \left(a_{pp}^{(k-1)}\right)^2 + \left(a_{qq}^{(k-1)}\right)^2 + 2\left(a_{pq}^{(k-1)}\right)^2 \tag{6.93}$$

Moreover, using Eqs. (6.83d) and (6.83e), we can write

$$\left(a_{ip}^{(k)}\right)^2 + \left(a_{iq}^{(k)}\right)^2 = \left(a_{ip}^{(k-1)}\right)^2 + \left(a_{iq}^{(k-1)}\right)^2, \qquad i \neq p, q \tag{6.94}$$

Hence, using Eqs. (6.93) and (6.94), as well as Eq. (6.83f), and considering the symmetry of A_{k-1} and A_k, we obtain

$$\|A_k\|_E^2 = \|A_{k-1}\|_E^2 \tag{6.95}$$

which indicates that *the Euclidean norm of a real symmetric matrix is invariant under an orthonormal transformation.* Next we write

$$\|D_k\|_E^2 = \left(a_{pp}^{(k)}\right)^2 + \left(a_{qq}^{(k)}\right)^2 + \sum_{\substack{i=1\\ i\neq p,q}}^{n} \left(a_{ii}^{(k)}\right)^2 \tag{6.96}$$

so that, considering Eqs. (6.92), (6.93) and (6.96), as well as Eq. (6.83f), Eq. (6.95) can be shown to yield

$$\|U_k\|_E^2 - \left(a_{pq}^{(k)}\right)^2 = \|U_{k-1}\|_E^2 - \left(a_{pq}^{(k-1)}\right)^2 \tag{6.97}$$

But, θ_k is chosen so as to render $a_{pq}^{(k)}$ zero, so that

$$\|U_k\|_E^2 = \|U_{k-1}\|_E^2 - \left(a_{pq}^{(k-1)}\right)^2 \tag{6.98}$$

Finally, inserting Eqs. (6.92) and (6.98) into Eq. (6.95), we conclude that

$$\|D_k\|_E^2 = \|D_{k-1}\|_E^2 + 2\left(a_{pq}^{(k-1)}\right)^2 \tag{6.99}$$

Equation (6.99) demonstrates that one iteration step causes the sum of the diagonal elements squared of A_k to increase by the amount $2\left(a_{pq}^{(k-1)}\right)^2$ relative to the sum of the diagonal elements squared of A_{k-1}. In view of Eqs. (6.92) and (6.95), we further conclude that the increase in the sum of the diagonal elements squared is at the expense of the sum of the off-diagonal elements squared. Pictorially, the iteration process can be envisioned as a steady migration of numerical strength from the off-diagonal elements to the diagonal ones, until all off-diagonal elements lose any significance and A_k becomes diagonal. This proves that *the iteration process is convergent*; it leads in the limit to the solution of the eigenvalue problem according to Eqs. (6.88) and (6.89).

From Eq. (6.99), we conclude that convergence can be accelerated by choosing the element of U_{k-1} of largest modulus as the element $a_{pq}^{(k-1)}$ to be annihilated. The drawback of such a choice is that it makes it necessary to search through all the elements of U_{k-1}. The simplest approach is to perform the rotations sequentially in the planes $(1, 2), (1, 3), \ldots (1, n), (2, 3), (2, 4), \ldots, (2, n), \ldots$ and $(n-1, n)$, where the sequence of $n(n-1)/2$ rotations is referred to as a *sweep*, and we note that one complete sweep requires approximately $2n^3$ multiplications. This procedure is known as the *serial Jacobi method*. If the element $a_{pq}^{(k-1)}$ is much smaller than the general level of the elements of U_{k-1}, then the effort made in annihilating $a_{pq}^{(k-1)}$ is almost totally wasted. To render the annihilation of $a_{pq}^{(k-1)}$ meaningful, we can establish a threshold value for each sweep and omit any rotation involving an off-diagonal element whose magnitude lies below the threshold value. The process is terminated when $n(n-1)/2$ consecutive rotations are omitted. This version of the Jacobi method is known as the *threshold serial Jacobi method.*

The accuracy of the Jacobi method depends on how accurately $\sin\theta_k$ and $\cos\theta_k$ are computed. If $\sin\theta_k$ and $\cos\theta_k$ are computed with reasonable accuracy, then no significant growth of rounding error occurs. The accuracy of the eigenvectors depends on the separation of the eigenvalues. One of the most significant features of the Jacobi method is that, even if some eigenvalues are very close, the associated eigenvectors are almost exactly orthonormal. Hence, if the interest lies not only in the eigenvalues but also in the full set of orthonormal eigenvectors, then the Jacobi method may prove more desirable than faster methods that do not produce eigenvectors, particularly for diagonally dominant matrices A.

In vibrations, our interest lies in an eigenvalue problem in terms of the mass and stiffness matrices, as given by Eq. (6.71), so that once again we wish to transform the eigenvalue problem into one defined by a single matrix A. For the Jacobi method, however, A must be real symmetric and positive definite. On the other hand, because the iteration process converges to all eigenvalues and eigenvectors simultaneously, it is no longer necessary to define the eigenvalues as inversely proportional to the natural frequencies squared. To reduce the eigenvalue problem (6.71) to one in terms of a single symmetric matrix, we first carry out the Cholesky decomposition (Sec. 6.2)

$$M = LL^T \tag{6.100}$$

where L is a nonsingular lower triangular matrix. Then, introducing the linear transformation

$$L^T\mathbf{u} = \mathbf{v} \tag{6.101}$$

so that

$$\mathbf{u} = \left(L^T\right)^{-1}\mathbf{v} = \left(L^{-1}\right)^T\mathbf{v} \tag{6.102}$$

and premultiplying both sides of Eq. (6.71) by L^{-1}, we obtain the desired eigenvalue problem in the form

$$A\mathbf{v} = \lambda\mathbf{v}, \qquad \lambda = \omega^2 \tag{6.103}$$

where

$$A = L^{-1}K\left(L^{-1}\right)^T = A^T \tag{6.104}$$

The solution of the eigenvalue problem, Eq. (6.103), yields eigenvalues equal to the natural frequencies squared, $\lambda_i = \omega_i^2$ $(i = 1, 2, \ldots, n)$. On the other hand, the eigenvectors must be inserted into Eq. (6.102) to produce the modal vectors, $\mathbf{u}_i = \left(L^{-1}\right)^T \mathbf{v}_i$ $(i = 1, 2, \ldots, n)$.

The Jacobi method exhibits ultimate quadratic convergence (Ref. 5, p. 448; Ref. 13, p. 270).

Example 6.5

Solve the eigenvalue problem of Example 6.4 by the serial Jacobi method.

To reduce the eigenvalue problem of Example 6.4 to one in terms of a single real symmetric matrix, we wish to carry out the Cholesky decomposition indicated by Eq. (6.100). In the case at hand, the mass matrix is diagonal, so that

$$L = M^{1/2} \tag{a}$$

Hence, inserting Eq. (a) into Eq. (6.104) and using Eqs. (a) of Example 6.4, we obtain

$$A = M^{-1/2}KM^{-1/2} = \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & -3 & 0 \\ -3 & 5 & -2 \\ 0 & -2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 2.5000 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3.0000 \end{bmatrix} \tag{b}$$

where the parameter ratio k/m was included in λ.

We begin the iteration with a rotation in the $(1, 2)$-plane, $p = 1$, $q = 2$. Using Eqs. (6.85) with $k = 1$, we have

$$b_0 = a_{12}^{(0)} = -1.2247, \qquad c_0 = \frac{1}{2}\left(a_{11}^{(0)} - a_{22}^{(0)}\right) = \frac{1}{2}(2.5000 - 1.6667) = 0.4167 \tag{c}$$

so that, using Eqs. (6.86) with $k = 1$, we can write

$$\cos\theta_1 = \left[\frac{1}{2} + \frac{c_0}{2\left(b_0^2 + c_0^2\right)^{1/2}}\right]^{1/2}$$
$$= \left[0.5000 + \frac{0.4167}{2\left(1.2247^2 + 0.4167^2\right)^{1/2}}\right]^{1/2} = 0.8130 \tag{d}$$
$$\sin\theta_1 = \frac{b_0}{2\left(b_0^2 + c_0^2\right)^{1/2}\cos\theta_1} = \frac{-1.2247}{2\left(1.2247^2 + 0.4167^2\right)^{1/2} 0.8130} = -0.5822$$

Equations (d) define the first rotation matrix, Eq. (6.82) with $k = 1$: Then, using Eq. (6.81) with $k = 1$, we obtain

$$A_1 = R_1^T A R_1$$

$$= \begin{bmatrix} 0.8130 & -0.5822 & 0 \\ 0.5822 & 0.8130 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2.5000 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3.0000 \end{bmatrix} \begin{bmatrix} 0.8130 & 0.5822 & 0 \\ -0.5822 & 0.8130 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3.3771 & 0 & 0.6723 \\ 0 & 0.7897 & 0.9388 \\ 0.6723 & 0.9388 & 3.0000 \end{bmatrix} \tag{e}$$

Next, we carry out a rotation in the (1, 3)-plane, $p = 1, q = 3$. Hence, Eqs. (6.85) with $k = 2$ yield

$$b_1 = a_{13}^{(1)} = 0.6723, \quad c_1 = \frac{1}{2}\left(a_{11}^{(1)} - a_{33}^{(1)}\right) = \frac{1}{2}(3.3771 - 3.0000) = 0.1885 \tag{f}$$

so that, using Eqs. (6.86) with $k = 2$, we have

$$\cos\theta_2 = \left[\frac{1}{2} + \frac{c_1}{2\left(b_1^2 + c_1^2\right)^{1/2}}\right]^{1/2} = \left[0.5000 + \frac{0.1885}{2\left(0.6723^2 + 0.1885^2\right)^{1/2}}\right]^{1/2}$$

$$= 0.7969 \tag{g}$$

$$\sin\theta_2 = \frac{b_1}{2\left(b_1^2 + c_1^2\right)^{1/2}\cos\theta_2} = \frac{0.6723}{2\left(0.6723^2 + 0.1885^2\right)^{1/2} 0.7969} = 0.6041$$

Then, using Eq. (6.81) with $k = 2$, we obtain

$$A_2 = R_2^T A_1 R_2$$

$$= \begin{bmatrix} 0.7969 & 0 & 0.6041 \\ 0 & 1 & 0 \\ -0.6041 & 0 & 0.7969 \end{bmatrix} \begin{bmatrix} 3.3771 & 0 & 0.6723 \\ 0 & 0.7897 & 0.9388 \\ 0.6723 & 0.9388 & 3.0000 \end{bmatrix} \begin{bmatrix} 0.7969 & 0 & -0.6041 \\ 0 & 1 & 0 \\ 0.6041 & 0 & 0.7969 \end{bmatrix}$$

$$= \begin{bmatrix} 3.8867 & 0.5672 & 0 \\ 0.5672 & 0.7897 & 0.7481 \\ 0 & 0.7481 & 2.4903 \end{bmatrix} \tag{h}$$

and note that the 1, 2 element, annihilated in the first rotation, is no longer zero. For the third and final step in the first sweep, we carry out a rotation in the (2, 3)-plane, $p = 2, q = 3$. From Eqs. (6.85) with $k = 3$, we have

$$b_2 = a_{23}^{(2)} = 0.7481$$

$$c_2 = \frac{1}{2}\left(a_{22}^{(2)} - a_{33}^{(2)}\right) = \frac{1}{2}(0.7897 - 2.4903) = -0.8503 \tag{i}$$

Then, using Eqs. (6.86) with $k = 3$, we can write

$$\cos\theta_3 = \left[\frac{1}{2} + \frac{c_2}{2\left(b_2^2 + c_2^2\right)^{1/2}}\right]^{1/2} = \left[0.5000 + \frac{-0.8503}{2\left(0.7481^2 + 0.8503^2\right)^{1/2}}\right]^{1/2}$$

$$= 0.3530 \tag{j}$$

$$\sin\theta_3 = \frac{b_2}{2\left(b_2^2 + c_2^2\right)^{1/2}\cos\theta_3} = \frac{0.7481}{2\left(0.7481^2 + 0.8503^2\right)^{1/2} 0.3530} = 0.9356$$

Finally, using Eq. (6.81) with $k = 3$, we obtain

$$\begin{aligned} A_3 &= R_3^T A_2 R_3 \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.3530 & 0.9356 \\ 0 & -0.9356 & 0.3530 \end{bmatrix} \begin{bmatrix} 3.8867 & 0.5672 & 0 \\ 0.5672 & 0.7897 & 0.7481 \\ 0 & 0.7481 & 2.4903 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.3530 & -0.9356 \\ 0 & 0.9356 & 0.3530 \end{bmatrix} \\ &= \begin{bmatrix} 3.8867 & 0.2002 & -0.5307 \\ 0.2002 & 2.7408 & 0 \\ -0.5307 & 0 & 0.5392 \end{bmatrix} \end{aligned} \tag{k}$$

which completes the first sweep.

The process continues with a new sweep beginning with a rotation in the (1, 2)-plane. Omitting the details, we list the final results

$$\Lambda = \text{diag}\ (4.0000\ 2.7412\ 0.4256), \qquad V = \begin{bmatrix} 0.4712 & 0.7436 & 0.4743 \\ -0.5774 & -0.1464 & 0.8033 \\ 0.6667 & -0.6525 & 0.3603 \end{bmatrix} \tag{l}$$

and we note that we were close to the two highest eigenvalues after one sweep only. We also note that the eigenvalues and eigenvectors are in descending order. Using the first of Eqs. (l), we obtain the natural frequencies

$$\begin{aligned} \omega_1 &= \sqrt{\frac{\lambda_1 k}{m}} = \sqrt{\frac{0.4256k}{m}} = 0.6524\sqrt{\frac{k}{m}} \\ \omega_2 &= \sqrt{\frac{\lambda_2 k}{m}} = \sqrt{\frac{2.7412k}{m}} = 1.6556\sqrt{\frac{k}{m}} \\ \omega_3 &= \sqrt{\frac{\lambda_3 k}{m}} = \sqrt{\frac{4.0000k}{m}} = 2.0000\sqrt{\frac{k}{m}} \end{aligned} \tag{m}$$

Moreover, considering Eq. (6.102) and rearranging the columns of V, the modal matrix is

$$\begin{aligned} U = M^{-1/2}V &= m^{-1/2} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1/\sqrt{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0.4743 & 0.7436 & 0.4712 \\ 0.8033 & -0.1464 & -0.5774 \\ 0.3603 & -0.6525 & 0.6667 \end{bmatrix} \\ &= m^{-1/2} \begin{bmatrix} 0.3354 & 0.5258 & 0.3333 \\ 0.4638 & -0.0845 & -0.3333 \\ 0.3603 & -0.6525 & 0.6667 \end{bmatrix} \end{aligned} \tag{n}$$

6.5 GIVENS' TRIDIAGONALIZATION METHOD

Some efficient computational algorithms for the solution of the symmetric eigenvalue problem require that the matrix A be tridiagonal. Other algorithms, although capable of solving the eigenvalue problem for a fully populated real symmetric matrix, are not competitive unless the matrix is tridiagonal. Some algorithms working with tridiagonal matrices are so powerful that they remain competitive even when the matrix A must be tridiagonalized first and the effort to tridiagonalize A is taken into account.

As demonstrated in Sec. 6.4, the Jacobi method reduces a real symmetric matrix A to diagonal form by means of a series of orthogonal transformations representing planar rotations. The method is iterative in nature, which implies extensive computations. The *Givens method* uses the same concept of orthogonal transformations representing rotations toward a more modest objective, namely, tridiagonalization instead of diagonalization. In contrast with the Jacobi method, Givens' method is not iterative and it involves $(n-1)(n-2)/2$ steps, a smaller number than the number of steps in one sweep alone in the Jacobi method.

The Givens method borrows some of the features from the Jacobi method, but the two methods differ on two major respects, the scheduling of the element annihilation and the computation of the rotation angles θ_k. Indeed, Eqs. (6.81)–(6.83) remain the same for the Givens method. But, whereas in the Jacobi method one seeks to annihilate $a_{pq}^{(k)}$ and $a_{qp}^{(k)}$, in the Givens method the objective is to annihilate the elements $a_{iq}^{(k)}$ and $a_{qi}^{(k)}$, $i \neq p, q$. Hence, from Eq. (6.83e), we write

$$a_{iq}^{(k)} = a_{qi}^{(k)} = -a_{ip}^{(k-1)} \sin\theta_k + a_{iq}^{(k-1)} \cos\theta_k = 0, \qquad i \neq p, q \tag{6.105}$$

Equation (6.105) can be satisfied by taking simply

$$\sin\theta_k = g_k a_{iq}^{(k-1)}, \qquad \cos\theta_k = g_k a_{ip}^{(k-1)}, \qquad k = 1, 2, \dots \tag{6.106}$$

where

$$g_k = \left[\left(a_{ip}^{(k-1)}\right)^2 + \left(a_{iq}^{(k-1)}\right)^2\right]^{1/2}, \qquad k = 1, 2, \dots \tag{6.107}$$

Contrasting Eqs. (6.106) and (6.107) with Eqs. (6.85) and (6.86), we conclude that the computation of $\sin\theta_k$ and $\cos\theta_k$ is significantly simpler in the Givens method than in the Jacobi method.

The tridiagonalization of A can be carried out in a series of steps designed to annihilate all the elements in the upper (lower) triangular matrix excluding the main diagonal and the upper (lower) subdiagonal. This implies that elements reduced to zero in a previous step must remain zero throughout, which is guaranteed if we annihilate in sequence the elements $a_{13}^{(1)}, a_{14}^{(2)}, \dots, a_{1n}^{(n-2)}, a_{24}^{(n-1)}, a_{25}^{(n)}, \dots, a_{2n}^{(2n-5)}, \dots, a_{n-2,n}^{((n-1)(n-2)/2)}$ through rotations in the planes $(2,3), (2,4), \dots, (2,n), (3,4), (3,5), \dots, (3,n), \dots, (n-1,n)$, respectively. The process requires $(n-1)(n-2)/2$ rotations and results in the tridiagonal matrix

$$T = A_k = \begin{bmatrix} \alpha_1 & \beta_2 & 0 & \dots & 0 & 0 \\ \beta_2 & \alpha_2 & \beta_3 & \dots & 0 & 0 \\ 0 & \beta_3 & \alpha_3 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \alpha_{n-1} & \beta_n \\ 0 & 0 & 0 & \dots & \beta_n & \alpha_n \end{bmatrix}; \qquad k = \frac{(n-1)(n-2)}{2} \tag{6.108}$$

The Givens reduction to tridiagonal form requires approximately $\frac{4}{3}n^3$ multiplication, and we recall from Sec. 6.4 that the Jacobi method requires $2n^3$ multiplications for just one sweep. The figures are not really comparable, as the Jacobi method aims at the diagonalization of a real symmetric matrix, whereas the Givens method aims only at its tridiagonalization.

At this point, we wish to establish how the eigenvalue problems for A and T relate to one another. To this end, we recall Eq. (6.79) and write the equation of the ellipsoid in terms of the two matrices as follows:

$$f = \mathbf{x}^T A\mathbf{x} = \mathbf{y}^T T\mathbf{y} = 1 \tag{6.109}$$

But, the matrices A and T are related by the transformation

$$T = R^T AR \tag{6.110}$$

where

$$R = R_1 R_2 \dots R_k = \prod_{i=1}^{k} R_i, \qquad k = (n-1)(n-2)/2 \tag{6.111}$$

is an orthonormal matrix, in which R_i are individual rotation matrices. Because Eq. (6.110) represents an orthonormal transformation, the matrices A and T possess the same eigenvalues, so that the question is how the eigenvectors of A relate to the eigenvectors of T. The two eigenvalue problems have the form

$$A\mathbf{x} = \lambda\mathbf{x} \tag{6.112}$$

and

$$T\mathbf{y} = \lambda\mathbf{y} \tag{6.113}$$

Inserting Eq. (6.110) into Eq. (6.113), we obtain

$$R^T AR\mathbf{y} = \lambda\mathbf{y} \tag{6.114}$$

Premultiplying both sides of Eq. (6.114) by R and recognizing that for an orthonormal matrix $RR^T = I$, Eq. (6.114) can be rewritten as

$$AR\mathbf{y} = \lambda R\mathbf{y} \tag{6.115}$$

so that, comparing Eqs. (6.112) and (6.115), we conclude that the eigenvectors of A are related to the eigenvectors of T by the linear transformation

$$\mathbf{x} = R\mathbf{y} \tag{6.116}$$

It should be pointed out that, unlike in the Jacobi method, here R does not represent a matrix of eigenvectors of A, because T is merely a tridiagonal matrix and not the diagonal matrix of eigenvalues. Of course, the problem of solving Eq. (6.113) for the eigenvalues and eigenvectors of $T = A_k$ remains, but the tridiagonal form of A_k opens new and attractive possibilities.

Example 6.6

The eigenvalue problem for the torsional system shown in Fig. 6.1 can be written in the form

$$A\mathbf{x} = \lambda\mathbf{x}, \qquad \lambda = \omega^2 \frac{IL}{GJ} \tag{a}$$

where $\mathbf{x} = [\psi_1\ \psi_2\ \psi_3\ \psi_4]^T$ is the vector of twist angles and

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \tag{b}$$

is the coefficient matrix. Note that all parameters were grouped in λ, in which I is the mass moment of inertia of each disk, L is the distance between disks and GJ is the torsional stiffness of the connecting shafts. Use Givens' method to tridiagonalize the matrix A.

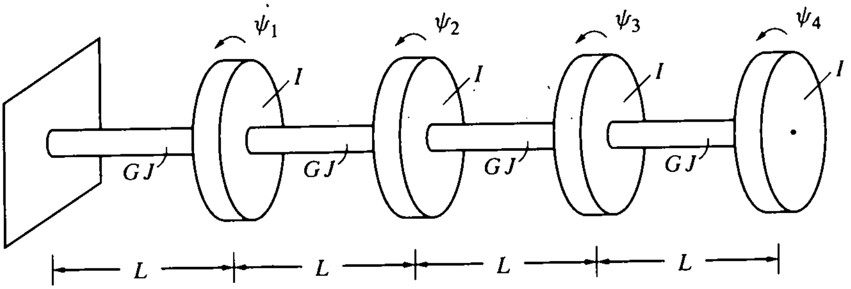


Figure 6.1 Four-degree-of-freedom torsional system

We begin with the annihilation of $a_{13}^{(0)}$. To this end, we use the rotation θ_1 in the (2,3) plane. Hence, letting $i = 1$, $p = 2$, $q = 3$ in Eqs. (6.106) and (6.107), we can write

$$\begin{aligned} \sin\theta_1 &= \frac{a_{13}^{(0)}}{\left[\left(a_{12}^{(0)}\right)^2 + \left(a_{13}^{(0)}\right)^2\right]^{1/2}} = \frac{1}{\left(1^2 + 1^2\right)^{1/2}} = \frac{1}{\sqrt{2}} = 0.7071 \\ \cos\theta_1 &= \frac{a_{12}^{(0)}}{\left[\left(a_{12}^{(0)}\right)^2 + \left(a_{13}^{(0)}\right)^2\right]^{1/2}} = \frac{1}{\left(1^2 + 1^2\right)^{1/2}} = \frac{1}{\sqrt{2}} = 0.7071 \end{aligned} \tag{c}$$

Then, inserting Eqs. (c) into Eqs. (6.81) and (6.82) with $k = 1$, we obtain

$$\begin{aligned} A_1 &= R_1^T A_0 R_1 = R_1^T A R_1 \\ &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.7071 & 0.7071 & 0 \\ 0 & -0.7071 & 0.7071 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.7071 & -0.7071 & 0 \\ 0 & 0.7071 & 0.7071 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1.4142 & 0 & 1 \\ 1.4142 & 4.5 & 0.5 & 3.5355 \\ 0 & 0.5 & 0.5 & 0.7071 \\ 1 & 3.5355 & 0.7071 & 4 \end{bmatrix} \end{aligned} \tag{d}$$

Next, we use the rotation θ_2 in the plane (2,4) to annihilate $a_{14}^{(1)}$, so that $i = 1$, $p = 2$, $q = 4$. From Eqs. (6.106) and (6.107) with $k = 2$, we have

$$\begin{aligned}
\sin\theta_2 &= \frac{a_{14}^{(1)}}{\left[\left(a_{12}^{(1)}\right)^2 + \left(a_{14}^{(1)}\right)^2\right]^{1/2}} = \frac{1}{\left(1.4142^2 + 1^2\right)^{1/2}} = 0.5774 \\
\cos\theta_2 &= \frac{a_{12}^{(1)}}{\left[\left(a_{12}^{(1)}\right)^2 + \left(a_{14}^{(1)}\right)^2\right]^{1/2}} = \frac{1.4142}{\left(1.4142^2 + 1^2\right)^{1/2}} = 0.8165
\end{aligned} \tag{e}$$

Inserting Eqs. (e) into Eqs. (6.81) and (6.82) with $k = 2$, we obtain

$$\begin{aligned}
A_2 &= R_2^T A_1 R_2 \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8165 & 0 & 0.5774 \\ 0 & 0 & 1 & 0 \\ 0 & -0.5774 & 0 & 0.8165 \end{bmatrix} \begin{bmatrix} 1 & 1.4142 & 0 & 1 \\ 1.4142 & 4.5 & 0.5 & 3.5355 \\ 0 & 0.5 & 0.5 & 0.7071 \\ 1 & 3.5355 & 0.7071 & 4 \end{bmatrix} \\
&\qquad \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8165 & 0 & -0.5774 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5744 & 0 & 0.8165 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1.7321 & 0 & 0 \\ 1.7321 & 7.6667 & 0.8165 & 0.9428 \\ 0 & 0.8165 & 0.5 & 0.2887 \\ 0 & 0.9428 & 0.2887 & 0.8333 \end{bmatrix}
\end{aligned} \tag{f}$$

Finally, we use the rotation θ_3 in the plane (3,4) to annihilate $a_{24}^{(2)}$, so that $i = 2$, $p = 3$, $q = 4$. To this end, we use Eqs. (6.106) and (6.107) with $k = 3$ and write

$$\begin{aligned}
\sin\theta_3 &= \frac{a_{24}^{(2)}}{\left[\left(a_{23}^{(2)}\right)^2 + \left(a_{24}^{(2)}\right)^2\right]^{1/2}} = \frac{0.9428}{\left(0.8165^2 + 0.9428^2\right)^{1/2}} = 0.7559 \\
\cos\theta_3 &= \frac{a_{23}^{(2)}}{\left[\left(a_{23}^{(2)}\right)^2 + \left(a_{24}^{(2)}\right)^2\right]^{1/2}} = \frac{0.8165}{\left(0.8165^2 + 0.9428^2\right)^{1/2}} = 0.6547
\end{aligned} \tag{g}$$

so that, introducing Eqs. (g) into Eqs. (6.81) and (6.82), we obtain

$$\begin{aligned}
T = A_3 &= R_3^T A_2 R_3 \\
&= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6547 & 0.7559 \\ 0 & 0 & -0.7559 & 0.6547 \end{bmatrix} \begin{bmatrix} 1 & 1.7321 & 0 & 0 \\ 1.7321 & 7.6667 & 0.8165 & 0.9428 \\ 0 & 0.8165 & 0.5 & 0.2887 \\ 0 & 0.9428 & 0.2887 & 0.8333 \end{bmatrix} \\
&\qquad \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6547 & -0.7559 \\ 0 & 0 & 0.7559 & 0.6547 \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} 1 & 1.7321 & 0 & 0 \\ 1.7321 & 7.6667 & 1.2472 & 0 \\ 0 & 1.2472 & 0.9762 & 0.1237 \\ 0 & 0 & 0.1237 & 0.3571 \end{bmatrix} \tag{h}$$

According to Eq. (6.116), the eigenvectors of A are related to the eigenvectors of T by the overall rotation matrix R. Hence, using Eq. (6.111), we can write

$$R = R_1R_2R_3$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.7071 & -0.7071 & 0 \\ 0 & 0.7071 & 0.7071 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.8165 & 0 & -0.5774 \\ 0 & 0 & 1 & 0 \\ 0 & 0.5774 & 0 & 0.8165 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6547 & -0.7559 \\ 0 & 0 & 0.7559 & 0.6547 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.5774 & -0.7715 & 0.2673 \\ 0 & 0.5774 & 0.1543 & -0.8018 \\ 0 & 0.5774 & 0.6172 & 0.5345 \end{bmatrix} \tag{i}$$

6.6 HOUSEHOLDER'S TRIDIAGONALIZATION METHOD

Another tridiagonalization technique, due to Householder, is intimately related to Givens' method. *Householder's method* also uses orthonormal transformations to reduce a real symmetric matrix to tridiagonal form, but it does it more efficiently than Givens' method. In contrast with Givens' method, in which two symmetric elements are annihilated at a time, in Householder's method a whole row and column (the symmetric counterpart of the row) are annihilated at a time, with the exception, of course, of the tridiagonal elements in that row and column. Like in Givens' method, subsequent transformations do not affect previously annihilated rows and columns, so that the tridiagonalization process requires $n-2$ transformations. But, in contrast with Givens' method, the transformations do not represent rotations.

Householder's tridiagonalization algorithm is defined by the transformation

$$A_k = P_kA_{k-1}P_k, \; A_0 = A, \; k = 1, 2, \ldots, n-2 \tag{6.117}$$

where

$$P_k = I - 2\mathbf{v}_k\mathbf{v}_k^T, \qquad \mathbf{v}_k^T\mathbf{v}_k = 1 \tag{6.118}$$

is a symmetric orthonormal matrix. Indeed, the symmetry of P_k can be verified with ease and the orthonormality follows from

$$\begin{aligned} P_k^TP_k &= \left(I - 2\mathbf{v}_k\mathbf{v}_k^T\right)\left(I - 2\mathbf{v}_k\mathbf{v}_k^T\right) = I - 4\mathbf{v}_k\mathbf{v}_k^T + 4\left(\mathbf{v}_k\mathbf{v}_k^T\right)\left(\mathbf{v}_k\mathbf{v}_k^T\right) \\ &= I - 4\mathbf{v}_k\mathbf{v}_k^T + 4\mathbf{v}_k\left(\mathbf{v}_k^T\mathbf{v}_k\right)\mathbf{v}_k^T = I \end{aligned} \tag{6.119}$$

The matrix P_k defined by Eqs. (6.118) represents a linear transformation in the real Euclidean space transforming one n-vector into another n-vector. The transformation can be interpreted geometrically as the extension to the n-dimensional case of a *reflection* through a given plane. For this reason, the matrix is referred to as an *elementary reflector*. However, P_k is better known as a *Householder transformation.*

The first transformation, defined by Eq. (6.117) with $k = 1$, must result in a matrix with the first row and column equal to zero, except for the two tridiagonal elements in the row and column. Hence, the matrix A_1 must have the form

$$A_1 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & 0 & \dots & 0 \\ a_{12}^{(1)} & a_{22}^{(1)} & a_{23}^{(1)} & \dots & a_{2n}^{(1)} \\ 0 & a_{23}^{(1)} & a_{33}^{(1)} & \dots & a_{3n}^{(1)} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & a_{2n}^{(1)} & a_{3n}^{(1)} & \dots & a_{nn}^{(1)} \end{bmatrix} \tag{6.120}$$

Equations (6.118) and (6.120) imply that the matrix P_1, and hence the vector $\mathbf{v}_1$, is subject to the $n - 1$ constraints

$$a_{13}^{(1)} = a_{14}^{(1)} = \dots = a_{1n}^{(1)} = 0, \qquad \mathbf{v}_1^T \mathbf{v}_1 = 1 \tag{6.121}$$

But, the fact that $n - 1$ constraints are imposed on the n components $v_{1,j}$ of the vector $\mathbf{v}_1$ implies that only one component is arbitrary. We designate this component to be $v_{1,1}$ and choose this value as zero, $v_{1,1} = 0$. Hence, the first transformation matrix has the form

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & & 0 \\ 0 & 1 - 2v_{1,2}^2 & -2v_{1,2}v_{1,3} & \dots & -2v_{1,2}v_{1,n} \\ 0 & -2v_{1,2}v_{1,3} & 1 - 2v_{1,3}^2 & \dots & -2v_{1,3}v_{1,n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & -2v_{1,2}v_{1,n} & -2v_{1,3}v_{1,n} & \dots & 1 - 2v_{1,n}^2 \end{bmatrix} \tag{6.122}$$

Next, we let $k = 1$ in Eq. (6.117) and write

$$A_1 = P_1 A_0 P_1, \qquad A_0 = A \tag{6.123}$$

where $A = A_0$ is a given matrix. Hence, the problem reduces to determining the components of the vector

$$\mathbf{v} = [0 \; v_{1,2} \; v_{1,3} \; \dots \; v_{1,n}]^T \tag{6.124}$$

in terms of the elements $a_{ij}^{(0)} = a_{ij}$ of the matrix $A_0 = A$. The algebraic operations are relatively lengthy and are omitted here. Details are given in Ref. 10, from which we obtain the results

$$v_{1,2} = \frac{1}{\sqrt{2}} \left(1 \mp \frac{a_{12}^{(0)}}{h_1} \right)^{1/2} \tag{6.125a}$$

$$v_{1,j} = \mp \frac{a_{1j}^{(0)}}{2h_1 v_{1,2}}, \qquad j = 3, 4, \dots, n \tag{6.125b}$$

in which

$$h_1 = \left[\sum_{j=2}^{n} \left(a_{ij}^{(0)}\right)^2\right]^{1/2} \tag{6.126}$$

and we note that the choice of sign in Eq. (6.125b) must be the same as in Eq. (6.125a).

In general, the vector $\mathbf{v}_k$ defining the transformation matrix P_k has the form

$$\mathbf{v}_k = [0\ 0\ \dots\ 0\ v_{k,k+1}\ v_{k,k+2}\ \dots\ v_{k,n}]^T, \qquad k = 1, 2, \dots, n-2 \tag{6.127}$$

where

$$v_{k,k+1} = \frac{1}{\sqrt{2}}\left(1 \mp \frac{a_{k,k+1}^{(k-1)}}{h_k}\right)^{1/2} \tag{6.128a}$$

$$v_{k,j} = \mp\frac{a_{kj}^{(k-1)}}{2h_k v_{k,k+1}}, \qquad j = k+2, k+3, \dots, n \tag{6.128b}$$

in which

$$h_k = \left[\sum_{j=k+1}^{n} \left(a_{kj}^{(k-1)}\right)^2\right]^{1/2}, \qquad k = 1, 2, \dots, n-2 \tag{6.129}$$

The Householder tridiagonalization method requires $\frac{2}{3}n^3$ multiplications, so that the method is twice as efficient as Givens' method.

Completion of $k = n-2$ transformations results in a tridiagonal matrix $T = A_k$ as given by Eq. (6.108). In fact, the matrix is essentially the same as that obtained by the Givens method, with the possible exception of inconsequential differences in the sign of the subdiagonal elements. Of course, the eigenvalues of A are the same as the eigenvalues of T, whereas the eigenvectors of A are related to the eigenvectors of T by

$$\mathbf{x} = P\mathbf{y} \tag{6.130}$$

where

$$P = P_1 P_2 \dots P_k = \prod_{i=1}^{k} P_i, \qquad k = n-2 \tag{6.131}$$

and note that A and T are related by

$$T = P^T A P \tag{6.132}$$

in which, by virtue of the symmetry of the matrices P_i $(i = 1, 2, \dots, k)$,

$$P^T = P_k \dots P_2 P_1 = \prod_{i=k}^{1} P_i, \qquad k = n-2 \tag{6.133}$$

Example 6.7

Carry out the tridiagonalization of the matrix A of Example 6.6 by means of Householder's method. Compare the results with those obtained in Example 6.6 and draw conclusions.

The tridiagonalization requires the transformation matrices P_i, which involve the vectors $\mathbf{v}_i$ $(i = 1, 2, \ldots, n-2)$. Using Eq. (6.126) in conjunction with Eq. (b) of Example 6.6, we can write

$$h_1 = \left[\sum_{j=2}^{4}\left(a_{ij}^{(0)}\right)^2\right]^{1/2} = \sqrt{1^2+1^2+1^2} = \sqrt{3} = 1.7321 \tag{a}$$

so that, from Eqs. (6.125), we obtain the nonzero components of $\mathbf{v}_1$

$$\begin{aligned} v_{1,2} &= \frac{1}{\sqrt{2}}\left(1 \mp \frac{a_{12}^{(0)}}{h_1}\right)^{1/2} = \frac{1}{\sqrt{2}}\left(1+\frac{1}{\sqrt{3}}\right)^{1/2} = 0.8881 \\ v_{1,3} &= \mp\frac{a_{13}^{(0)}}{2h_1 v_{1,2}} = \frac{1}{2 \times 1.7321 \times 0.8881} = 0.3251 \\ v_{1,4} &= \mp\frac{a_{14}^{(0)}}{2h_1 v_{1,2}} = \frac{1}{2 \times 1.7321 \times 0.8881} = 0.3251 \end{aligned} \tag{b}$$

Hence,

$$\mathbf{v}_1 = [0 \;\; 0.8881 \;\; 0.3251 \;\; 0.3251]^T \tag{c}$$

Inserting Eq. (c) into Eq. (6.118) with $k = 1$, we obtain

$$P_1 = I - 2\mathbf{v}_1\mathbf{v}_1^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.5774 & -0.5774 & -0.5774 \\ 0 & -0.5774 & 0.7887 & -0.2113 \\ 0 & -0.5774 & -0.2113 & 0.7887 \end{bmatrix} \tag{d}$$

Then, introducing Eq. (b) of Example 6.6 and the above Eq. (d) into Eq. (6.123), we have

$$A_1 = P_1 A P_1 = \begin{bmatrix} 1 & -1.7321 & 0 & 0 \\ -1.7321 & 7.6667 & -0.5447 & -1.1221 \\ 0 & -0.5447 & 0.3780 & 0.1667 \\ 0 & -1.1221 & 0.1667 & 0.9553 \end{bmatrix} \tag{e}$$

Next, we let $k = 2$ in Eq. (6.129) and use Eq. (e) to write

$$h_2 = \left[\sum_{j=3}^{4}\left(a_{2j}^{(1)}\right)^2\right]^{1/2} = \left[(-0.5447)^2 + (-1.1221)^2\right]^{1/2} = 1.2472 \tag{f}$$

so that, using Eqs. (6.128) with $k = 2$ and Eqs. (e) and (f), we have

$$\begin{aligned} v_{2,3} &= \frac{1}{\sqrt{2}}\left(1 \mp \frac{a_{23}^{(1)}}{h_2}\right)^{1/2} = \frac{1}{\sqrt{2}}\left(1+\frac{0.5447}{1.2472}\right)^{1/2} = 0.8476 \\ v_{2,4} &= \mp\frac{a_{24}^{(1)}}{2h_2 v_{2,3}} = \frac{1.1221}{2 \times 1.2472 \times 0.8476} = 0.5307 \end{aligned} \tag{g}$$

Hence, the vector $\mathbf{v}_2$ has the form

$$\mathbf{v}_2 = [0\ 0\ 0.8476\ 0.5307]^T \tag{h}$$

Introducing Eq. (h) into Eq. (6.118) with $k = 2$, we can write

$$P_2 = I - 2\mathbf{v}_2\mathbf{v}_2^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.4367 & -0.8996 \\ 0 & 0 & -0.8996 & 0.4367 \end{bmatrix} \tag{i}$$

so that, using Eq. (6.117) with $k = 2$, we obtain

$$T = A_2 = P_2A_1P_2 = \begin{bmatrix} 1 & -1.7321 & 0 & 0 \\ -1.7321 & 7.6667 & 1.2472 & 0 \\ 0 & 1.2472 & 0.9762 & -0.1237 \\ 0 & 0 & -0.1237 & 0.3571 \end{bmatrix} \tag{j}$$

Moreover, using Eq. (6.131) with $k = 2$, we have

$$P = P_1P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.5774 & 0.7715 & 0.2673 \\ 0 & -0.5774 & -0.1543 & -0.8018 \\ 0 & -0.5774 & -0.6172 & 0.5345 \end{bmatrix} \tag{k}$$

Comparing the above Eq. (j) with Eq. (h) of Example 6.6, we conclude that the tridiagonal matrix computed by means of Householder's method is essentially the same as that computed by means of Givens' method. The only difference is in the sign of some subdiagonal elements, which does not affect the eigenvalues. In addition, if we compare the above Eq. (k) with Eq. (i) of Example 6.6, we conclude that the Householder transformation matrix P is essentially the same as Givens' rotation matrix R, with the exception of some signs. We should note, however, that the sign differences are consistent. This implies that, whereas there may be a difference in the sign of various components of the eigenvectors of T, Eq. (j), the eigenvectors of $A = PTP$ will be the same as the eigenvectors of $A = R^TTR$

6.7 LANCZOS' TRIDIAGONALIZATION METHOD

The Lanczos method represents a direct method for the tridiagonalization of a symmetric matrix by means of an orthonormal transformation, whereby the tridiagonal form and the transformation matrix are obtained through a simple recursive process. We recall that Givens' method and Householder's method also use orthonormal transformations, but in Givens' method they represent rotations and in Householder's method they represent reflections. Although there seems to be general agreement that Householder's method is the most effective tridiagonalization procedure, a discussion of Lanczos' method is likely to prove rewarding.

Let us assume that the tridiagonal matrix T is related to the original real symmetric matrix A by

$$T = P^TAP \tag{6.134}$$

where $P = [\mathbf{p}_1\ \mathbf{p}_2\ \dots\ \mathbf{p}_n]$ is an orthonormal transformation matrix, in which $\mathbf{p}_j$ $(j = 1, 2, \dots, n)$ are unit vectors. Premultiplying both sides of Eq. (6.134) by P and recognizing that $PP^T = I$ for an orthonormal matrix, we obtain

$$AP = PT \tag{6.135}$$

Then, inserting Eq. (6.108) into Eq. (6.135), we can write

$$A\mathbf{p}_j = \beta_j \mathbf{p}_{j-1} + \alpha_j \mathbf{p}_j + \beta_{j+1}\mathbf{p}_{j+1}, \quad j = 1, 2, \ldots, n; \ \mathbf{p}_0 = \mathbf{0}, \mathbf{p}_{n+1} = \mathbf{0} \tag{6.136}$$

Premultiplying both sides of Eq. (6.136) by $\mathbf{p}_j^T$ and considering the orthonormality of the vectors $\mathbf{p}_j$, we obtain

$$\alpha_j = \mathbf{p}_j^T A \mathbf{p}_j, \qquad j = 1, 2, \ldots, n \tag{6.137}$$

Moreover, Eqs. (6.136) can be rewritten as

$$\mathbf{r}_{j+1} = \left(A - \alpha_j I\right)\mathbf{p}_j - \beta_j \mathbf{p}_{j-1}, \qquad j = 1, 2, \ldots, n-1; \ \mathbf{p}_0 = \mathbf{0} \tag{6.138}$$

in which we introduced the definitions

$$\beta_j = \|\mathbf{r}_j\|, \qquad \mathbf{p}_j = \mathbf{r}_j/\beta_j, \qquad j = 2, 3, \ldots, n \tag{6.139}$$

Equations (6.137)–(6.139) represent a set of recursive formulae that can be used to determine the elements of T and the columns of P beginning with a given unit vector $\mathbf{p}_1$. For simplicity, the vector $\mathbf{p}_1$ can be taken as the standard unit vector $\mathbf{e}_n$.

Example 6.8

Tridiagonalize the matrix of Example 6.6 by means of the Lanczos method.

Beginning with Eq. (6.137) and using $\mathbf{p}_1 = \mathbf{e}_4$ in conjunction with Eq. (b) of Example 6.6, we have for $j = 1$

$$\alpha_1 = \mathbf{p}_1^T A \mathbf{p}_1 = \mathbf{e}_4^T A \mathbf{e}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = 4 \tag{a}$$

Then, using Eq. (6.138) with $j = 1$, we obtain

$$\mathbf{r}_2 = (A - \alpha_1 I)\mathbf{p}_1 = \begin{bmatrix} -3 & 1 & 1 & 1 \\ 1 & -2 & 2 & 2 \\ 1 & 2 & -1 & 3 \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \tag{b}$$

so that, Eqs. (6.139) with $j = 2$ yield

$$\beta_2 = \|\mathbf{r}_2\| = \sqrt{1^2 + 2^2 + 3^2} = \sqrt{14}$$

$$\mathbf{p}_2 = \mathbf{r}_2/\beta_2 = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} \tag{c}$$

Next, we use Eq. (6.137) with $j = 2$ and write

$$\alpha_2 = \mathbf{p}_2^T A \mathbf{p}_2 = \frac{1}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = 5 \tag{d}$$

Then, using Eq. (6.138) with $j = 2$, we obtain

$$\begin{aligned} \mathbf{r}_3 &= (A - \alpha_2 I)\mathbf{p}_2 - \beta_2 \mathbf{p}_1 \\ &= \frac{1}{\sqrt{14}} \begin{bmatrix} -4 & 1 & 1 & 1 \\ 1 & -3 & 2 & 2 \\ 1 & 2 & -2 & 3 \\ 1 & 2 & 3 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} - \sqrt{14} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{14}} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \end{aligned} \tag{e}$$

so that, from Eqs. (6.139) with $j = 3$, we can write

$$\beta_3 = \|\mathbf{r}_3\| \frac{1}{\sqrt{14}} \sqrt{1^2 + 1^2 + (-1)^2} = \sqrt{3/14}$$

$$\mathbf{p}_3 = \mathbf{r}_3/\beta_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} \tag{f}$$

Now, we use Eq. (6.137) with $j = 3$ and write

$$\alpha_3 = \mathbf{p}_3^T A \mathbf{p}_3 = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} = \frac{2}{3} \tag{g}$$

so that Eq. (6.138) with $j = 3$ can be used to obtain

$$\begin{aligned} \mathbf{r}_4 &= (A - \alpha_3 I)\mathbf{p}_3 - \beta_3 \mathbf{p}_2 \\ &= \frac{1}{\sqrt{3}} \begin{bmatrix} 1/3 & 1 & 1 & 1 \\ 1 & 4/3 & 2 & 2 \\ 1 & 2 & 7/3 & 3 \\ 1 & 2 & 3 & 10/3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 0 \end{bmatrix} - \frac{\sqrt{3}}{14} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 0 \end{bmatrix} = \frac{1}{42\sqrt{3}} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 0 \end{bmatrix} \end{aligned} \tag{h}$$

Moreover, Eqs. (6.139) with $j = 4$ yield

$$\beta_4 = \|\mathbf{r}_4\| = \frac{1}{42\sqrt{3}} \sqrt{5^2 + (-4)^2 + 1^2} = \frac{1}{\sqrt{126}}$$

$$\mathbf{p}_4 = \mathbf{r}_4/\beta_4 = \frac{1}{\sqrt{42}} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 0 \end{bmatrix} \tag{i}$$

Finally, using Eq. (6.137) with $j = 4$, we obtain

$$\alpha_4 = \mathbf{p}_4^T A \mathbf{p}_4 = \frac{1}{42} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 0 \end{bmatrix}^T \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ -4 \\ 1 \\ 0 \end{bmatrix} = \frac{1}{3} \tag{j}$$

Using the above results, the tridiagonal matrix is

$$T = \begin{bmatrix} \alpha_1 & \beta_2 & 0 & 0 \\ \beta_2 & \alpha_2 & \beta_3 & 0 \\ 0 & \beta_3 & \alpha_3 & \beta_4 \\ 0 & 0 & \beta_4 & \alpha_4 \end{bmatrix} = \begin{bmatrix} 4 & \sqrt{14} & 0 & 0 \\ \sqrt{14} & 5 & \sqrt{3/14} & 0 \\ 0 & \sqrt{3/14} & 2/3 & 1/\sqrt{126} \\ 0 & 0 & 1/\sqrt{126} & 1/3 \end{bmatrix}$$

$$= \begin{bmatrix} 4 & 3.7417 & 0 & 0 \\ 3.7417 & 5 & 0.4629 & 0 \\ 0 & 0.4629 & 0.6667 & 0.0891 \\ 0 & 0 & 0.0891 & 0.3333 \end{bmatrix} \tag{k}$$

and the transformation matrix is

$$P = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \mathbf{p}_4] = \begin{bmatrix} 0 & 1/\sqrt{14} & 1/\sqrt{3} & 5/\sqrt{42} \\ 0 & 2/\sqrt{14} & 1/\sqrt{3} & -4/\sqrt{42} \\ 0 & 3/\sqrt{14} & -1/\sqrt{3} & 1/\sqrt{42} \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0.2673 & 0.5774 & 0.7715 \\ 0 & 0.5345 & 0.5774 & -0.6172 \\ 0 & 0.8018 & -0.5774 & 0.1543 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{l}$$

where the latter can be verified to be orthonormal.

In looking over Examples 6.6–6.8, the simplicity of Lanczos' method compared to Givens' method and Householder's method is immediately apparent. Moreover, whereas the tridiagonal matrix obtained by Lanczos' method differs to some extent from that obtained by Givens' method and Householder's method, the transformation matrix is virtually the same. Indeed, the only differences are in the sign and in the position of the elements within the individual columns of the transformation matrix. The implication of the difference in the tridiagonal matrix T is that convergence to the eigenvalues may be somewhat slower, depending on the algorithm used. On the other hand, the differences in sign and position in the transformation matrix P are consistent with the differences in sign and position in T in the sense that they are consistent with the relation $A = PTP^T$. This implies that, although the eigenvectors of T may differ from one tridiagonalization method to another, the eigenvectors of A do not. Hence, all differences are immaterial, as they do not affect the final outcome, i.e., the eigenvalues and eigenvectors of A.

6.8 GIVENS' METHOD FOR THE EIGENVALUES OF TRIDIAGONAL MATRICES

Givens' method (Ref. 4) is one of the most effective techniques for computing the eigenvalues of a real symmetric tridiagonal matrix. It is also one of the most versatile, as it permits the targeting of individual eigenvalues for computation, provided the general location of a given eigenvalue is known. In this regard, we recall from Sec. 5.6 that, according to Gerschgorin's first theorem, the eigenvalues lie inside Gerschgorin's disks, i.e., certain circular regions in the complex λ-plane. In the case of real symmetric matrices, the Gerschgorin disks collapse into segments of the real λ-axis. The determination of the approximate location of the eigenvalues is particularly effective when the matrix is diagonally dominant, which can be the case following tridiagonalization of a real symmetric matrix.

As discussed in Sec. 4.6, the eigenvalues of a real symmetric matrix can be obtained by finding the roots of the characteristic polynomial, which implies the expansion of the associated characteristic determinant, an $n \times n$ determinant. But, as demonstrated in Sec. 6.1, the taste for evaluating determinants disappears rapidly as n increases. The beauty of Givens' method lies in the fact that *it finds the roots of the characteristic polynomial without actually requiring the polynomial explicitly.*

The characteristic determinant associated with the tridiagonal matrix T given by Eq. (6.108) has the form

$$\det(T - \lambda I) = \begin{vmatrix} \alpha_1 - \lambda & \beta_2 & 0 & \dots & 0 & 0 \\ \beta_2 & \alpha_2 - \lambda & \beta_3 & \dots & 0 & 0 \\ 0 & \beta_3 & \alpha_3 - \lambda & \dots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \dots & \alpha_{n-1} - \lambda & \beta_n \\ 0 & 0 & 0 & \dots & \beta_n & \alpha_n - \lambda \end{vmatrix} \tag{6.140}$$

Denoting by $p_i(\lambda)$ the principal minor determinant of order i of the matrix $T - \lambda I$, it can be shown by induction that

$$\begin{aligned} p_1(\lambda) &= \alpha_1 - \lambda \\ p_i(\lambda) &= (\alpha_i - \lambda)\, p_{i-1}(\lambda) - \beta_i^2 p_{i-2}(\lambda), \qquad i = 2, 3, \dots, n \end{aligned} \tag{6.141}$$

in which $p_0(\lambda)$ is taken to be equal to 1 identically. We propose to establish some important properties of the polynomials $p_i(\lambda)$ $(i = 1, 2, \dots, n)$, without actually having explicit expressions for the polynomials. Of course, $p_n(\lambda)$ is the characteristic polynomial and

$$p_n(\lambda) = 0 \tag{6.142}$$

represents the characteristic equation. We observe from Eq. (6.140) that, if β_i is zero for some value of i, then the characteristic polynomial reduces to the product of two polynomials of correspondingly lower order and the roots of $p_n(\lambda) = 0$ are simply the roots of the two polynomials. Hence, without loss of generality, we can assume that no β_i is equal to zero.

Next, we consider the interval $a < \lambda < b$ on the real λ-axis and denote it by (a, b), where a and b are given numbers such that neither one of them is a root of the polynomials defined by Eqs. (6.141). The first step in the Givens method for the computation of the eigenvalues of T is the determination of the number of roots of the characteristic polynomial lying in the interval (a, b). As a preliminary to the exposition of Givens' method, we must show that the sequence of polynomials $p_1(\lambda)$, $p_2(\lambda)$, $\dots$, $p_n(\lambda)$, associated with the principal minor determinants of the matrix $T - \lambda I$ and defined by Eqs. (6.141), possesses the following properties:

i. $p_0(\lambda) \neq 0$.

ii. If $p_{i-1}(\mu) = 0$ for some $\lambda = \mu$, then $p_i(\mu)$ and $p_{i-2}(\mu)$ are nonzero and of opposite signs.

iii. As λ passes through a zero of $p_n(\lambda)$, the quotient $p_n(\lambda)/p_{n-1}(\lambda)$ changes sign from positive to negative.

The first property is true because $p_0(\lambda) \equiv 1$ by definition. To demonstrate the second property, we let $\lambda = \mu$ in Eqs. (6.141), assume that $p_{i-1}(\mu) = 0$ and obtain

$$p_i(\mu) = -\beta_i^2 p_{i-2}(\mu) \tag{6.143}$$

If we further assume that $p_i(\mu)$ is equal to zero, then according to Eq. (6.143) $p_{i-2}(\mu)$ must also be zero, so that three consecutive polynomials in the sequence are zero. Under these circumstances, we conclude from Eqs. (6.141) with $\lambda = \mu$

that $p_{i-3}(\mu) = p_{i-4}(\mu) = \dots = p_0(\mu) = 0$, which contradicts the fact that $p_0(\mu) \equiv 1$. Hence, if $p_{i-1}(\mu) = 0$, then $p_i(\mu) \neq 0$ and $p_{i-2}(\mu) \neq 0$, so that we conclude immediately from Eq. (6.143) that $p_i(\mu)$ and $p_{i-2}(\mu)$ must have opposite signs.

To prove the third property, we call upon the separation theorem (Sec. 5.4). First, we denote the roots of the characteristic polynomial $p_n(\lambda)$ by $\lambda_1, \lambda_2, \dots, \lambda_n$ and assume that they are ordered so as to satisfy $\lambda_1 < \lambda_2 < \dots < \lambda_n$. Moreover, the polynomial $p_{n-1}(\lambda)$ represents the determinant of the matrix obtained by striking out the last row and column from the matrix $T - \lambda I$. Consistent with the developments in Sec. 5.4, we denote the roots of $p_{n-1}(\lambda)$ by $\lambda_1', \lambda_2', \dots, \lambda_{n-1}'$ and assume that they are ordered so that $\lambda_1' < \lambda_2' < \dots < \lambda_{n-1}'$. Then, according to the separation theorem, the two sets of eigenvalues satisfy the inequalities

$$\lambda_1 < \lambda_1' < \lambda_2 < \lambda_2' < \lambda_3 < \dots < \lambda_{n-1} < \lambda_{n-1}' < \lambda_n \tag{6.144}$$

Typical plots of $p_n(\lambda)$ and $p_{n-1}(\lambda)$ are shown in Fig. 6.2, in which vertical dashed lines through $\lambda_1, \lambda_1', \lambda_2, \dots, \lambda_{n-1}, \lambda_{n-1}', \lambda_n$ separate regions in which the ratio $p_n(\lambda)/p_{n-1}(\lambda)$ possesses opposite signs. Note that, because the matrix T is positive definite, $p_{n-1}(0) > 0$ and $p_n(0) > 0$. It is clear from Fig. 6.2 that, as λ passes through the roots $\lambda_1, \lambda_2, \dots, \lambda_n$, the sign of $p_n(\lambda)/p_{n-1}(\lambda)$ changes from positive to negative. It follows that the sequence of polynomials $p_1(\lambda), p_2(\lambda), \dots, p_n(\lambda)$ possesses all three indicated properties. A sequence of polynomials possessing these three properties is known as a *Sturm sequence.*

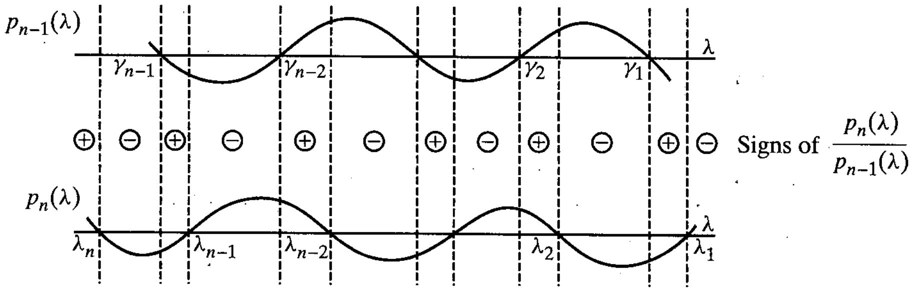


Figure 6.2 Signs of the ratio $p_n(\lambda)/p_{n-1}(\lambda)$

At this point, we are in the position to determine the number of eigenvalues lying in the interval $a < \lambda < b$ by simply invoking *Sturm's theorem*, which reads as follows: *If the polynomials $p_0(\lambda), p_1(\lambda), \dots, p_n(\lambda)$ represent a Sturm sequence on the interval (a, b) and if $s(\mu)$ denotes the number of sign changes in the consecutive sequence of numbers $p_0(\mu), p_1(\mu), \dots, p_n(\mu)$, then the number of roots of the polynomial $p_n(\lambda)$ in the interval (a, b) is equal to $s(b) - s(a)$.* If $p_i(\mu) = 0$ for some μ, then the sign of $p_i(\mu)$ is taken as the opposite of the sign of $p_{i-1}(\mu)$. It is clear from Eqs. (6.141) that this does not affect the number of sign changes, as from $p_{i-1}(\mu)$ to $p_{i+1}(\mu)$ there is only one sign change, independently of the sign assigned to $p_i(\mu)$. Sturm's theorem can be proved by induction. To this end, we assume that the number of sign changes $s(\mu)$ in the sequence of numbers

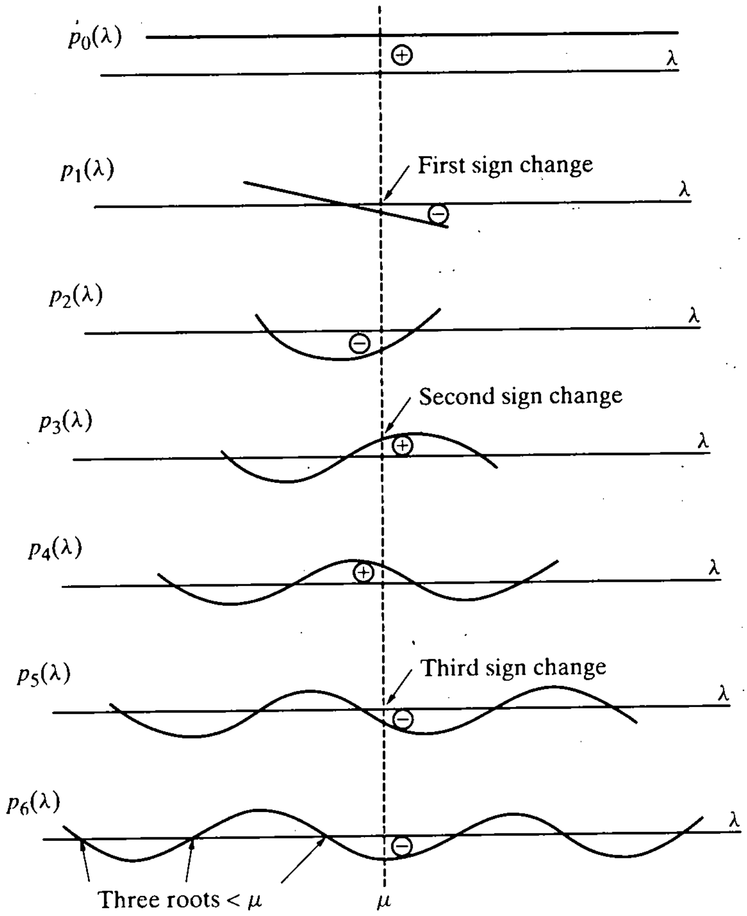


Figure 6.3 Plots of the polynomials $p_0(\lambda)$, $p_1(\lambda)$, ..., $p_6(\lambda)$ showing the sign changes for $\lambda = \mu$

$p_0(\mu)$, $p_1(\mu)$, ..., $p_n(\mu)$ is equal to the number of roots of $p_n(\lambda)$ corresponding to $\lambda < \mu$. As an example, we consider the sequence of seven polynomials $p_0(\lambda)$, $p_1(\lambda)$, ..., $p_6(\lambda)$ depicted in Fig. 6.3. For the particular value of μ shown, there are three sign changes in the sequence of numbers $p_0(\mu)$, $p_1(\mu)$, ..., $p_n(\mu)$ and there are exactly three roots, λ_1, λ_2 and λ_3, of the characteristic polynomial $p_6(\lambda)$ for $\lambda < \mu$. As μ increases, the number $s(\mu)$ remains the same until μ crosses the root λ_4, at which point $s(\mu)$ increases by one. This can be explained by the fact that, according to the second property of the Sturm sequence, the number of sign changes remains the same as μ crosses a root of $p_{i-1}(\lambda)$ $(i = 1, 2, \ldots, n)$. At the same time, according to the third property, there is one additional sign change as μ crosses a root of $p_n(\lambda)$. Hence, the number $s(\mu)$ increases by one every time μ crosses a root of $p_n(\lambda)$, which proves Sturm's theorem.

It should be emphasized here that, the polynomials $p_2(\lambda)$, $p_3(\lambda)$, ..., $p_n(\lambda)$ are never available in explicit form, nor is their explicit form necessary. Indeed, to determine the integers $s(a)$ and $s(b)$, it is only necessary to compute the values $p_1(a), p_2(a), \ldots, p_n(a)$ and $p_1(b), p_2(b), \ldots, p_n(b)$, which can be done recursively by using Eqs. (6.141). As far as the selection of the interval (a, b) designed to locate a given eigenvalue, Gerschgorin's first theorem can often provide some reasonable guidance, as discussed in the beginning of this section. If $s(b) \neq s(a)$, we know that there is at least one root in the interval (a, b). Then, the search for the desired eigenvalue can be narrowed by using the *bisection method*, which amounts to computing $s((a+b)/2)$ and checking the numbers $s((a+b)/2) - s(a)$ and $s(b) - s((a+b)/2)$. Then, if one of these two numbers is zero, the search is limited to the other interval, which is bisected again.

Convergence of Givens' method is linear, as the error is approximately halved at each step (Ref. 5, p. 437). However, the method yields eigenvalues with small relative error, regardless of their magnitude (Ref. 5, p. 439). Givens' method yields only the eigenvalues of a real symmetric tridiagonal matrix. Having the eigenvalues, the eigenvectors can be computed very efficiently by means of inverse iteration, as discussed in Sec. 6.10.

Example 6.9

Compute the lowest eigenvalue of the matrix of Example 6.6 by means of Givens' method based on Sturm's theorem.

Givens' method requires a tridiagonal matrix. From Example 6.6, the tridiagonal matrix is

$$T = \begin{bmatrix} \alpha_1 & \beta_2 & 0 & \\ \beta_2 & \alpha_2 & \beta_3 & 0 \\ 0 & \beta_3 & \alpha_3 & \beta_4 \\ 0 & 0 & \beta_4 & \alpha_4 \end{bmatrix} = \begin{bmatrix} 1 & 1.7321 & 0 & 0 \\ 1.7321 & 7.6667 & 1.2472 & 0 \\ 0 & 1.2472 & 0.9762 & 0.1237 \\ 0 & 0 & 0.1237 & 0.3571 \end{bmatrix} \tag{a}$$

so that, from Eqs. (6.141), we obtain the sequence of polynomials

$$\begin{aligned} p_1(\lambda) &= \alpha_1 - \lambda = 1 - \lambda \\ p_2(\lambda) &= (\alpha_2 - \lambda)\, p_1(\lambda) - \beta_2^2 p_0(\lambda) = (7.6667 - \lambda)\, p_1(\lambda) - 1.7321^2 \\ p_3(\lambda) &= (\alpha_3 - \lambda)\, p_2(\lambda) - \beta_3^2 p_1(\lambda) = (0.9762 - \lambda)\, p_2(\lambda) - 1.2472^2 p_1(\lambda) \\ p_4(\lambda) &= (\alpha_4 - \lambda)\, p_3(\lambda) - \beta_4^2 p_2(\lambda) = (0.3571 - \lambda)\, p_3(\lambda) - 0.1237^2 p_2(\lambda) \end{aligned} \tag{b}$$

Before we can apply Sturm's theorem, we must choose the interval (a, b). To this end, we consider using Gerschgorin's first theorem. Unfortunately, the subdiagonal elements β_2 and β_3 are too large for the theorem to yield useful results. Still, the element β_4 is sufficiently small to cause us to suspect that the lowest eigenvalue is not very far from α_4. Hence, from Sec. 5.6, we consider the possibility that the lowest eigenvalue lies in the segment with the center at α_4 and having the length $2r = 2\beta_4$, so that the end points of the segment are $\alpha_4 - \beta_4$ and $\alpha_4 + \beta_4$. Using Eq. (a), the end points have the value $\alpha_4 - \beta_4 = 0.2334$ and 0.4808. It turns out that a sharper estimate can be obtained using the tridiagonal matrix computed by means of Lanczos' method. Indeed, using Eq. (k) of Example 6.8, we can write $\alpha_4 - \beta_4 = 0.2442$ and $\alpha_4 + \beta_4 = 0.4224$. In consideration of this, we begin the iteration process with $a = 0.2$ and $b = 0.4$. The computation results are displayed in Table 6.1. Clearly, the choice $a = 0.2$, $b = 0.4$ was a good one, as 0.2 and 0.4 bracket the lowest eigenvalue, as witnessed by the fact that $s(0.2) = 0$ and $s(0.4) = 1$. Note that after the fifth iteration convergence is quite rapid, as can be concluded from the behavior of $p_4(\lambda)$. Convergence is achieved when $p_4(\lambda)$ reduces to zero, which is close at hand, as the actual lowest eigenvalue is $\lambda_1 = 0.283119$. Note that at this point we are sufficiently close to convergence that it is more expedient to abandon the iteration process and complete the computation of the eigenvalue through interpolation. Indeed, a simple linear interpolation yields

$$\begin{aligned}\lambda_1 &= \lambda^{(13)} - \frac{p_4\left(\lambda^{(13)}\right)}{p_4\left(\lambda^{(13)}\right) - p_4\left(\lambda^{(11)}\right)}\left(\lambda^{(13)} - \lambda^{(11)}\right)\\ &= 0.283106 - \frac{0.000010}{0.000010 - (-0.000070)}(0.283106 - 0.283204)\\ &= 0.283106 + 0.000012 = 0.283118 \qquad \text{(c)}\end{aligned}$$

where the superscript in parentheses indicates the iteration number.

TABLE 6.1

	λ	$p_1(\lambda)$	$p_2(\lambda)$	$p_3(\lambda)$	$p_4(\lambda)$	$s(\lambda)$
1	0.2	0.8	2.973333	1.063427	0.121600	0
2	0.4	0.6	1.360000	−0.149715	−0.027232	1
3	0.3	0.7	2.156667	0.369428	−0.011900	1
4	0.25	0.75	2.562500	0.694196	0.035156	0
5	0.275	0.725	2.358958	0.526301	0.007125	0
6	0.2875	0.7125	2.257656	0.446492	−0.003461	1
7	0.28125	0.71875	2.308268	0.486052	0.001557	0
8	0.284375	0.715625	2.282952	0.466186	−0.001020	1
9	0.282813	0.717187	2.295604	0.476095	0.000251	0
10	0.283594	0.716406	2.289278	0.471136	−0.000388	1
11	0.283204	0.716796	2.292437	0.473611	−0.000070	1
12	0.283009	0.716991	2.294016	0.474849	0.000089	0
13	0.283106	0.716894	2.293230	0.474233	0.000010	0

6.9 THE QR METHOD FOR SYMMETRIC EIGENVALUE PROBLEMS

The QR *method* is an iteration technique for computing the eigenvalues of a general matrix A by reducing the matrix to triangular form through orthonormal similarity transformations. The algorithm was developed independently by Francis (Ref. 2) and Kublanovskaya (Ref. 8). In this section, our interest lies in the eigenvalues of a real symmetric matrix A, in which case the matrix is reduced to diagonal form.

The iteration process consists of the matrix decomposition

$$A_s = Q_s R_s, \qquad s = 1, 2, \ldots;\ A_1 = A \qquad (6.145)$$

where Q_s is an orthonormal matrix and R_s is an upper triangular[1] matrix, followed by the computation of the matrix product in reverse order, or

$$A_{s+1} = R_s Q_s, \qquad s = 1, 2, \dots \tag{6.146}$$

Multiplying Eq. (6.145) by Q_s^T, considering the fact that Q_s is orthonormal and introducing the result into Eq. (6.146), we obtain

$$A_{s+1} = Q_s^T A_s Q_s, \qquad s = 1, 2, \dots \tag{6.147}$$

so that Eqs. (6.145) and (6.146) do indeed represent an orthonormal similarity transformation. For fully populated matrices, although the method converges, convergence can be very slow. However, if the matrix A is first reduced to tridiagonal form, then convergence of the QR algorithm is much faster, but still not competitive. Before the QR algorithm becomes truly effective, one additional refinement is necessary, namely, the incorporation of eigenvalue shifts, referred to as *shifts in origin*. The QR algorithm with shifts is defined by

$$A_s - \mu_s I = Q_s R_s, \qquad s = 1, 2, \dots \tag{6.148}$$

and

$$A_{s+1} = R_s Q_s + \mu_s I, \qquad s = 1, 2, \dots \tag{6.149}$$

and note that in general the value of the shift μ_s varies from step to step. The question remains as to the strategy to be employed for choosing the timing and value of the shifts. In this regard, it should be mentioned that, although the iteration process converges to all eigenvalues, convergence is not simultaneous but to one eigenvalue at a time. Indeed, the bottom right corner element $a_{nn}^{(s)}$ is the first to approach an eigenvalue, namely, the lowest eigenvalue. Consistent with this, the strategy recommended by Wilkinson (Ref. 13, Sec. 8.24) consists of solving the eigenvalue problem associated with the 2×2 lower right corner matrix

$$\begin{bmatrix} a_{n-1,n-1}^{(s)} & a_{n-1,n}^{(s)} \\ a_{n,n-1}^{(s)} & a_{nn}^{(s)} \end{bmatrix}$$

and take μ_s as the eigenvalue of this matrix closest to $a_{nn}^{(s)}$. The shift can be carried out at every stage or it can be delayed until the shift gives some indication of convergence. It is suggested in Ref. 13 that μ_s be accepted as a shift as soon as the criterion

$$|(\mu_s/\mu_{s-1}) - 1| < \frac{1}{2} \tag{6.150}$$

is satisfied. Convergence to the lowest eigenvalue is recognized by the fact that the last row and column have been reduced to a single nonzero element, the bottom right corner element, which is equal to λ_1. This amounts to an automatic deflation, as the iteration process to the remaining $n - 1$ eigenvalues continues with only an $(n-1) \times (n-1)$ matrix. Convergence to the eigenvalues $\lambda_2, \lambda_3, \dots, \lambda_n$ takes place at an accelerating rate as the element $a_{nn}^{(s)}$ approaches λ_1, because at the same time the

[1] The symbol R_s derives from the term "right triangular" used by Francis.

elements $a^{(s)}_{n-1,n-1}, \ldots, a^{(s)}_{22}, a^{(s)}_{11}$ make significant strides toward $\lambda_2, \ldots, \lambda_{n-1}, \lambda_n$, respectively. This is reflected in the fact that the matrix A_s resembles more and more a diagonal matrix. Convergence of $a^{(s)}_{n-1,n-1}$ to λ_2 follows soon after convergence to λ_1 has been achieved, at which time the last row and column of the deflated $(n-1) \times (n-1)$ matrix consist of a single nonzero element, the bottom right corner element, which is now equal to λ_2. Clearly, the iteration process to λ_3 continues with a deflated $(n-2) \times (n-2)$ matrix. This establishes the pattern for iteration to the remaining eigenvalues. The accelerating convergence rate is due to the progress made by all the eigenvalues during the iteration process, as well as to the automatic matrix deflation.

At this point, we turn our attention to the actual computational process. In view of the fact that the iteration is carried out using a tridiagonal matrix, we rewrite the algorithm in the form

$$T_s = Q_s R_s, \qquad s = 1, 2, \ldots \tag{6.151a}$$

$$T_{s+1} = R_s Q_s, \qquad s = 1, 2, \ldots \tag{6.151b}$$

The factorization indicated by Eq. (6.151a) amounts to determining the orthonormal matrix Q_s reducing T_s to upper triangular form the R_s. This matrix Q_s can be constructed in the form of a product of $n-1$ rotation matrices or a product of $n-1$ reflection matrices, such as in Givens' method and in Householder's method, respectively. Because now we must annihilate a single element per column, Householder's method loses the competitive edge, so that Givens' method is quite adequate. To annihilate the lower subdiagonal elements of T_s, we carry out rotations in the planes $(1, 2), (2, 3), \ldots, (n-1, n)$ using rotation matrices of the form

$$\Theta_k = \begin{array}{cccccccc}
 & & & k & k+1 & & & \\
1 & 0 & \cdots & 0 & 0 & \cdots & 0 & \\
0 & 1 & \cdots & 0 & 0 & \cdots & 0 & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
0 & 0 & \cdots & \cos\theta_k & \sin\theta_k & \cdots & 0 & k \\
0 & 0 & \cdots & -\sin\theta_k & \cos\theta_k & \cdots & 0 & k+1 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
0 & 0 & \cdots & 0 & 0 & \cdots & 1 &
\end{array}, \qquad k = 1, 2, \ldots, n-1 \tag{6.152}$$

where

$$\left.\begin{aligned}
\sin\theta_k &= \frac{t^{(k-1)}_{k+1,k}}{\left[\left(t^{(k-1)}_{k,k}\right)^2 + \left(t^{(k-1)}_{k+1,k}\right)^2\right]^{1/2}}, \\
\cos\theta_k &= \frac{t^{(k-1)}_{k,k}}{\left[\left(t^{(k-1)}_{k,k}\right)^2 + \left(t^{(k-1)}_{k+1,k}\right)^2\right]^{1/2}},
\end{aligned}\right. \qquad k = 1, 2, \ldots, n-1 \tag{6.153}$$

in which θ_k is the rotation angle and $t_{k+1,k}^{(k-1)}$ denotes the element in the $k+1$ row and k column of the matrix $T_s^{(k-1)}$, where the matrix $T_s^{(k-1)}$ is obtained from the tridiagonal matrix T_s through the recursive formula

$$T_s^{(k)} = \Theta_k T_s^{(k-1)}, k = 1, 2, \ldots, n-1; \quad T_s^{(0)} = T_s, \quad T_s^{(n-1)} = R_s \quad (6.154)$$

Moreover, the matrix Q_s, needed for the computation of T_{s+1}, can be written in the form

$$Q_s = \Theta_1^T \Theta_1^T \ldots \Theta_{n-1}^T \quad (6.155)$$

The process including shifts and the shifting strategy remain the same as that given by Eqs. (6.148)-(6.150), except that A_s must be replaced by T_s.

The QR method with the shifting strategy described in this section exhibits better than cubic convergence (Ref. 13, p. 562), which is quite remarkable.

Example 6.10

Compute the eigenvalues of the tridiagonal matrix of Example 6.6 by means of the QR method. The tridiagonal matrix of Example 6.6, with six decimal places accuracy, is

$$T = T_1 = \begin{bmatrix} 1 & 1.732051 & 0 & 0 \\ 1.732051 & 7.666667 & 1.247219 & 0 \\ 0 & 1.247219 & 0.976190 & 0.123718 \\ 0 & 0 & 0.123718 & 0.357143 \end{bmatrix} \quad \text{(a)}$$

The QR method calls for the reduction of the matrix T to an upper triangular form through premultiplication by an orthogonal matrix representing the product of Givens rotations in the planes (1,2), (2,3) and (3,4), where the rotation matrices are defined by Eqs. (6.152) and (6.153). Hence, letting $k = 1$ in Eqs. (6.153), in conjunction with the notation $T_1 = T_1^{(0)}$, we can write

$$\begin{aligned} \sin\theta_1 &= \frac{t_{21}^{(0)}}{\sqrt{\left(t_{11}^{(0)}\right)^2 + \left(t_{21}^{(0)}\right)^2}} = \frac{1.732051}{\sqrt{1^2 + 1.732051^2}} = 0.866025 \\ \cos\theta_1 &= \frac{t_{11}^{(0)}}{\sqrt{\left(t_{11}^{(0)}\right)^2 + \left(t_{21}^{(0)}\right)^2}} = \frac{1}{\sqrt{1^2 + 1.732051^2}} = 0.5 \end{aligned} \quad \text{(b)}$$

so that, from Eq. (6.152), the first rotation matrix is

$$\Theta_1 = \begin{bmatrix} 0.5 & 0.866025 & 0 & 0 \\ -0.866025 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{(c)}$$

Inserting Eqs. (a) and (c) into Eqs. (6.154) with $k = 1$ and $s = 1$, we obtain

$$T_1^{(1)} = \Theta_1 T_1^{(0)} = \begin{bmatrix} 2.000000 & 7.505553 & 1.080123 & 0 \\ 0 & 2.333333 & 0.623610 & 0 \\ 0 & 1.247219 & 0.976190 & 0.123718 \\ 0 & 0 & 0.123718 & 0.357143 \end{bmatrix} \quad \text{(d)}$$

Next, letting $k = 2$ in Eqs. (6.153) and using the indicated elements of $T_1^{(1)}$, Eq. (d), we have

$$\sin\theta_2 = \frac{t_{32}^{(1)}}{\sqrt{\left(t_{22}^{(1)}\right)^2 + \left(t_{32}^{(1)}\right)^2}} = \frac{1.247219}{\sqrt{2.333333^2 + 1.247219^2}} = 0.471405$$

$$\cos\theta_2 = \frac{t_{22}^{(1)}}{\sqrt{\left(t_{22}^{(1)}\right)^2 + \left(t_{32}^{(1)}\right)^2}} = \frac{2.333333}{\sqrt{2.333333^2 + 1.247219^2}} = 0.881917 \qquad \text{(e)}$$

so that the second rotation matrix has the form

$$\Theta_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.881917 & 0.471405 & 0 \\ 0 & -0.471405 & 0.881917 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \text{(f)}$$

Introducing Eqs. (d) and (f) into Eq. (6.154) with $k = 2$ and $s = 1$, we compute

$$T_1^{(2)} = \Theta_2 T_1^{(1)} = \begin{bmatrix} 2.000000 & 7.505553 & 1.080123 & 0 \\ 0 & 2.645751 & 1.010153 & 0.058321 \\ 0 & 0 & 0.566947 & 0.109109 \\ 0 & 0 & 0.123718 & 0.357143 \end{bmatrix} \qquad \text{(g)}$$

Letting $k = 3$ in Eqs. (6.153) in conjunction with the indicated elements of $T_1^{(2)}$, Eq. (g), we can write

$$\sin\theta_3 = \frac{t_{43}^{(2)}}{\sqrt{\left(t_{33}^{(2)}\right)^2 + \left(t_{43}^{(2)}\right)^2}} = \frac{0.123718}{\sqrt{0.566947^2 + 0.123718^2}} = 0.213201$$

$$\cos\theta_3 = \frac{t_{33}^{(2)}}{\sqrt{\left(t_{33}^{(2)}\right)^2 + \left(t_{43}^{(2)}\right)^2}} = \frac{0.566947}{\sqrt{0.566947^2 + 0.123718^2}} = 0.977008 \qquad \text{(h)}$$

so that the third rotation matrix is

$$\Theta_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.977008 & 0.213201 \\ 0 & 0 & -0.213201 & 0.977008 \end{bmatrix} \qquad \text{(i)}$$

Finally, premultiplying Eq. (g) by Eq. (i), we obtain

$$R_1 = T_1^{(3)} = \Theta_3 T_1^{(2)} = \begin{bmatrix} 2.000000 & 7.505553 & 1.080123 & 0 \\ 0 & 2.645751 & 1.010153 & 0.058321 \\ 0 & 0 & 0.580288 & 0.182743 \\ 0 & 0 & 0 & 0.325669 \end{bmatrix} \qquad \text{(j)}$$

Moreover, using Eq. (6.155) in conjunction with Eqs. (c), (f) and (i), we compute

$$Q_1 = \Theta_1^T \Theta_2^T \Theta_3^T = \begin{bmatrix} 0.500000 & -0.763763 & 0.398862 & -0.087039 \\ 0.866025 & 0.440959 & -0.230283 & 0.050252 \\ 0 & 0.471405 & 0.861640 & -0.188025 \\ 0 & 0 & 0.213201 & 0.977008 \end{bmatrix} \qquad \text{(k)}$$

Then, introducing Eqs. (j) and (k) into Eq. (6.151b) with $s = 1$, we have

$$T_2 = R_1 Q_1 = \begin{bmatrix} 7.500000 & 2.291288 & 0 & 0 \\ 2.291288 & 1.642857 & 0.273551 & 0 \\ 0 & 0.273551 & 0.538961 & 0.069433 \\ 0 & 0 & 0.069433 & 0.318182 \end{bmatrix} \tag{l}$$

At this point, before we begin the second iteration stage, we must decide whether to shift or not. To this end, we solve the eigenvalue problem for the 2 × 2 matrix in the lower right corner of T_1 and T_2. The eigenvalue closest to entry t_{44} is $\mu_1 = 0.333333$ in the case of T_1 and $\mu_2 = 0.298161$ in the case of T_2. Inserting these values into inequality (6.150), we can write

$$\left|(\mu_2/\mu_1) - 1\right| = \left|(0.298161/0.333333) - 1\right| = 0.105517 < \frac{1}{2} \tag{m}$$

so that a shift is in order. Hence, the new iteration stage involves the QR decomposition of the matrix

$$T_2 - \mu_2 I = \begin{bmatrix} 7.201839 & 2.291288 & 0 & 0 \\ 2.291288 & 1.344696 & 0.273551 & 0 \\ 0 & 0.273551 & 0.240800 & 0.069433 \\ 0 & 0 & 0.069433 & 0.020021 \end{bmatrix} \tag{n}$$

Omitting the details, we list the resulting upper triangular matrix

$$R_2 = \begin{bmatrix} 7.557547 & 2.591130 & 0.082935 & 0 \\ 0 & 0.647370 & 0.338011 & 0.029339 \\ 0 & 0 & 0.128474 & 0.063768 \\ 0 & 0 & 0 & -0.017165 \end{bmatrix} \tag{o}$$

and the orthogonal matrix

$$Q_2 = \begin{bmatrix} 0.952934 & -0.274782 & 0.107789 & -0.069236 \\ 0.303179 & 0.863677 & -0.338796 & 0.217618 \\ 0 & 0.422555 & 0.762572 & -0.489821 \\ 0 & 0 & 0.540441 & 0.841380 \end{bmatrix} \tag{p}$$

Hence, inserting Eqs. (o) and (p) into Eq. (6.149) with $s = 2$ and with A replaced by T, we obtain

$$T_3 = R_2 Q_2 + \mu_2 I = \begin{bmatrix} 8.285580 & 0.196269 & 0 & 0 \\ 0.196269 & 1.000108 & 0.054287 & 0 \\ 0 & 0.054287 & 0.430595 & -0.009277 \\ 0 & 0 & -0.009277 & 0.283719 \end{bmatrix} \tag{q}$$

Comparing the off-diagonal elements of T_3 with those of T_2, we conclude that the iteration converges rapidly.

The iteration steps are clear by now, so that we merely list the results of the next iteration stage, as follows:

$$\mu_3 = 0.283136$$

$$T_4 = \begin{bmatrix} 8.290819 & 0.017507 & 0 & 0 \\ 0.017507 & 0.999833 & 0.010888 & 0 \\ 0 & 0.010888 & 0.426229 & 0 \\ 0 & 0 & 0 & 0.283119 \end{bmatrix} \tag{r}$$

It is clear from T_4 that convergence has been achieved, so that

$$\lambda_1 = 0.283119 \tag{s}$$

which is essentially the same result as that obtained in Example 6.9. It is also clear that significant progress has been made toward convergence to the remaining eigenvalues.

To iterate to the next eigenvalue, we use the deflated matrix consisting of the 3 × 3 upper left corner of T_4. The results of the next iteration stage are

$$\mu_4 = 0.426022$$
$$T_5 = \begin{bmatrix} 8.290860 & 0.001277 & 0 \\ 0.001277 & 1.000000 & 0 \\ 0 & 0 & 0.426022 \end{bmatrix} \tag{t}$$

from which it is clear that

$$\lambda_2 = 0.426022 \tag{u}$$

The iteration to the next eigenvalue is to be carried out with the upper left corner 2 × 2 matrix. In view of the fact that the sole of f-diagonal term is quite small, we can dispense with further computations and accept as the remaining two eigenvalues

$$\lambda_3 = 1.000000, \qquad \lambda_4 = 8.290860 \tag{v}$$

6.10 INVERSE ITERATION

As shown in Sec. 6.3, the power method for solving the eigenvalue problem is defined by the process

$$\mathbf{v}_p = A\mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.156}$$

which iterates to the largest eigenvalue and associated eigenvector first. On the other hand, the process

$$\mathbf{v}_p = A^{-1}\mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.157}$$

iterates to the smallest eigenvalue first. The process defined by Eq. (6.157) is sometimes referred to as *inverse iteration*. To avoid the need for inverting the matrix A, we premultiply both sides of Eq. (6.157) by A and obtain

$$A\mathbf{v}_p = \mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.158}$$

which implies that, given the vector $\mathbf{v}_{p-1}$, the iterate $\mathbf{v}_p$ can be computed by solving a set of n nonhomogeneous algebraic equations in n unknowns. In this regard, we recall from Sec. 6.1 that Gaussian elimination with back-substitution is ideally suited for this task.

The approach suggested by Eq. (6.158) is not very attractive, unless appropriate modifications are made. One such modification is a shift in origin, so that the new iteration process is now defined by

$$(A - \lambda I)\mathbf{v}_p = \mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.159}$$

where λ is a given scalar. But, from the expansion theorem (Sec. 4.6), we can express the initial trial vector $\mathbf{v}_0$ as the linear combination

$$\mathbf{v}_0 = \sum_{i=1}^{n} c_i\mathbf{x}_i \tag{6.160}$$

where $\mathbf{x}_i$ are mutually orthonormal eigenvectors satisfying Eqs. (6.63). Inserting Eq. (6.160) into Eq. (6.159), it is not difficult to verify that

$$\mathbf{v}_p = \sum_{i=1}^{n} \frac{c_i}{(\lambda_i - \lambda)^p} \mathbf{x}_i \tag{6.161}$$

Next, we choose the shift λ to be very close to the eigenvalue λ_r. Then, Eq. (6.161) can be rewritten in the form

$$\mathbf{v}_p = c_r \mathbf{x}_r + \sum_{\substack{i=1 \\ i \neq r}}^{n} c_i \left(\frac{\lambda_r - \lambda}{\lambda_i - \lambda} \right)^p \mathbf{x}_i \tag{6.162}$$

where the scaling factor $(\lambda_r - \lambda)^{-p}$ has been ignored as inconsequential. Because λ is very close to λ_r, the summation terms decrease very fast as p increases. This is true even when $\mathbf{v}_0$ is highly deficient in $\mathbf{x}_r$. Hence, we can write

$$\lim_{p \to \infty} \mathbf{v}_p = c_r \mathbf{x}_r \tag{6.163}$$

We refer to the iteration process given by Eq. (6.159) as *inverse iteration*. In contrast to the ordinary power method, inverse iteration is capable of producing the eigenvectors in no particular order, and the speed of convergence does not depend on the ratio between two eigenvalues but on how close the choice λ is to λ_r. Clearly, inverse iteration is highly suitable for computing the eigenvectors corresponding to known eigenvalues.

In using the iteration algorithm described by Eq. (6.159), it is desirable to keep the magnitude of the iterates from becoming too large. To this end, we must normalize the newly computed iterate, which can be done by means of a scaling factor, in a manner similar to that used in Sec. 6.3 for the power method. Hence, by analogy with Eqs. (6.58), we rewrite the iteration process, Eq. (6.159), in the form

$$(A - \lambda I)\mathbf{v}_p^* = \mathbf{v}_{p-1}, \quad \mathbf{v}_p = \alpha_p \mathbf{v}_p^*, \qquad p = 1, 2, \dots \tag{6.164}$$

where α_p is the scaling factor. If the iterates are normalized so that the component of $\mathbf{v}_p$ largest in magnitude is equal to one, then α_p is given by Eq. (6.59). On the other hand, if the iterates are to have unit magnitude, then α_p is given by Eq. (6.60).

When A is a fully populated matrix, the computational effort can be excessive. In view of this, it is advisable to use inverse iteration in conjunction with tridiagonal matrices, in which case Eqs. (6.164) are replaced by

$$(T - \lambda I)\mathbf{v}_p^* = \mathbf{v}_{p-1}, \qquad \mathbf{v}_p = \alpha_p \mathbf{v}_p^*, \qquad p = 1, 2, \dots \tag{6.165}$$

Assuming that matrices T and A are related by

$$T = R^T A R \tag{6.166}$$

where R is obtained by means of Givens' method (Sec. 6.5), Householder's method (Sec. 6.6), or Lanczos' method (Sec. 6.7), and denoting the eigenvectors of T by $\mathbf{y}_i$, the eigenvectors of A can be recovered from the eigenvectors of T by writing

$$\mathbf{x}_i = R\mathbf{y}_i, \qquad i = 1, 2, \dots, n \tag{6.167}$$

To solve the first of Eqs. (6.165), we recall from Sec. 6.1 that Gaussian elimination amounts to reducing the coefficient matrix to triangular form. Hence, following the procedure described in Sec. 6.1, we rewrite the iteration process in the form

$$U\mathbf{v}_p^* = \mathbf{w}_{p-1}, \qquad \mathbf{w}_p = \alpha_p P\mathbf{v}_p^*, \qquad p = 1, 2, \ldots \tag{6.168}$$

where

$$U = P(T - \lambda I) \tag{6.169}$$

is an upper triangular matrix and P is a transformation matrix obtained in $n - 1$ steps (see Sec. 6.1).

Finally, we wish to discuss some numerical aspects of the inverse iteration process embodied in Eqs. (6.168). In the first place, we observe that the matrix $T - \lambda I$ is nearly singular, because the value of λ is chosen to be close to an eigenvalue. This implies that one of the diagonal elements of U will be close to zero. This should not be interpreted as an indication of impending significant numerical problems. Assuming that the small number is in the bottom right corner, a simple scaling of the vector $\mathbf{w}_{p-1}$ can dispose of the problem. Indeed, this small number only affects the magnitude of the eigenvector, which is inconsequential, as for an eigenvector only the direction is unique and the magnitude is arbitrary. Note that this magnitude is generally adjusted later, during the normalization process. It should be pointed out that, if λ is extremely close to an eigenvalue, it may be necessary to increase the numerical accuracy of the triangularization. Convergence of inverse iteration is extremely fast.

Example 6.11

Compute the eigenvector belonging to the eigenvalue $\lambda_1 = 0.283119$ of the matrix considered in Example 6.10 by means of inverse iteration.

Using Eq. (a) of Example 6.10 in conjunction with $\lambda = 0.283119$, we can write

$$T - \lambda I = \begin{bmatrix} 0.716881 & 1.732051 & 0 & 0 \\ 1.732051 & 7.383548 & 1.247219 & 0 \\ 0 & 1.247219 & 0.693071 & 0.123718 \\ 0 & 0 & 0.123718 & 0.074024 \end{bmatrix} \tag{a}$$

The first task to be carried out is the triangularization of the matrix $T - \lambda I$. To this end, we use the matrix formulation of the Gaussian elimination described in Sec. 6.1 and construct the first transformation matrix

$$P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -\dfrac{1.732051}{0.716881} & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2.416093 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{b}$$

Then, premultiplying Eq. (a) by Eq. (b), we obtain

$$P_1(T - \lambda I) = \begin{bmatrix} 0.716881 & 1.732051 & 0 & 0 \\ 0 & 3.198753 & 1.247219 & 0 \\ 0 & 1.247219 & 0.693071 & 0.123718 \\ 0 & 0 & 0.123718 & 0.074024 \end{bmatrix} \tag{c}$$

The next transformation matrix is simply

$$P_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -\dfrac{1.247219}{3.198753} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -0.389908 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{d}$$

so that, premultiplying Eq. (c) by Eq. (d), we have

$$P_2P_1(T - \lambda I) = \begin{bmatrix} 0.716881 & 1.732051 & 0 & 0 \\ 0 & 3.198753 & 1.247219 & 0 \\ 0 & 0 & 0.206743 & 0.123718 \\ 0 & 0 & 0.123718 & 0.074024 \end{bmatrix} \tag{e}$$

The third and final transformation matrix is

$$P_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -\dfrac{0.123718}{0.206743} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -0.598414 & 1 \end{bmatrix} \tag{f}$$

so that the desired upper triangular matrix is

$$U = P_3P_2P_1(T - \lambda I) = \begin{bmatrix} 0.716881 & 1.732051 & 0 & \\ 0 & 3.198753 & 1.247219 & \\ 0 & 0 & 0.206743 & 0.123718 \\ 0 & 0 & 0 & -0.000011 \end{bmatrix} \tag{g}$$

and the overall transformation matrix is

$$P = P_3P_2P_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -2.416093 & 1 & 0 & 0 \\ 0.942054 & -0.389908 & 1 & 0 \\ -0.563738 & 0.233326 & -0.598414 & 1 \end{bmatrix} \tag{h}$$

In view of the small value of u_{44}, we begin the iteration process with

$$\mathbf{w}_0 = 10^{-5}[1 \;\; 1 \;\; 1 \;\; 1]^T \tag{i}$$

Inserting Eq. (i) into the first of Eqs. (6.168) and considering Eq. (g), we write

$$\begin{aligned} 0.716881v_1^* + 1.732051v_2^* &= 10^{-5} \\ 3.198753v_2^* + 1.247219v_3^* &= 10^{-5} \\ 0.206743v_3^* + 0.123718v_4^* &= 10^{-5} \\ -0.000011v_4^* &= 10^{-5} \end{aligned} \tag{j}$$

Using back-substitution, we obtain the solution of Eqs. (j), which can be displayed in the vector form

$$\mathbf{v}_1^* = [v_1^* \;\; v_2^* \;\; v_3^* \;\; v_4^*]^T = [0.512528 \;\; -0.212137 \;\; 0.544062 \;\; -0.909091]^T \tag{k}$$

Then, using the second of Eqs. (6.168) and factoring out $\alpha_1 = -1.573094 \times 10^5$ so as to keep the bottom element of the vector $\mathbf{w}_1$ equal to 10^{-5}, we have

$$\mathbf{w}_1 = 10^{-5}[-0.325809 \;\; 0.922038 \;\; - 0.705365 \;\; 1]^T \tag{l}$$

so that we can replace Eqs. (j) by

$$\begin{aligned} 0.716881v_1^* + 1.732051v_2^* &= -0.325809 \times 10^{-5} \\ 3.198753v_2^* + 1.247219v_3^* &= 0.922038 \times 10^{-5} \\ 0.206743v_3^* + 0.123718v_4^* &= -0.705365 \times 10^{-5} \\ -0.000011v_4^* &= 10^{-5} \end{aligned} \tag{m}$$

which have the solution

$$\mathbf{v}_2^* = 10^{-5}[0.512447 \ \ -0.212099 \ \ 0.543979 \ \ -0.909091]^T \tag{n}$$

Using the second of Eqs. (6.168) and factoring out $\alpha_2 = -1.572990 \times 10^5$, we obtain

$$\mathbf{w}_2 = 10^{-5}[-0.325779 \ \ 0.921950 \ \ -0.705300 \ \ 1]^T \tag{o}$$

One more back-substitution yields the same result, $\mathbf{v}_3^* = \mathbf{v}_2^*$, so that we accept $\mathbf{v}_2^*$ as the eigenvector of T belonging to $\lambda_1 = 0.283119$, or

$$\mathbf{y}_1 = \mathbf{v}_2^* = [0.512447 \ \ -0.212099 \ \ 0.543979 \ \ -0.909091]^T \tag{p}$$

To produce the eigenvector of A, we recall Eq. (6.167), use the matrix R given by Eq. (i) of Example 6.6 (with six decimal places accuracy) and write

$$\mathbf{x}_1 = R\mathbf{y}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0.577350 & -0.771517 & 0.267261 \\ 0 & 0.577350 & 0.154303 & -0.801784 \\ 0 & 0.577350 & 0.617213 & 0.534522 \end{bmatrix} \begin{bmatrix} 0.512447 \\ -0.212099 \\ 0.543979 \\ -0.909091 \end{bmatrix}$$

$$= \begin{bmatrix} 0.512447 \\ -0.785109 \\ 0.690378 \\ -0.272633 \end{bmatrix} \tag{q}$$

which, upon normalizing so that $\|\mathbf{x}_1\| = 1$, has the form

$$\mathbf{x}_1 = [0.428537 \ \ -0.656553 \ \ 0.577333 \ \ -0.227991]^T \tag{r}$$

and it should be noted that λ_1 and $\mathbf{x}_1$ really correspond to the fourth vibration mode.

6.11 RAYLEIGH'S QUOTIENT ITERATION

The inverse iteration algorithm discussed in Sec. 6.10 permits a very efficient computation of an eigenvector $\mathbf{x}_i$ corresponding to a given eigenvalue λ_i $(i = 1, 2, \ldots, n)$. There are occasions in which the eigenvalue λ_i is not known, but a good approximation of the eigenvector $\mathbf{x}_i$ is available. Under these circumstances, the inverse iteration can be modified so as to produce both the eigenvalue λ_i and the eigenvector $\mathbf{x}_i$. The modification consists of inserting the initial estimate of $\mathbf{x}_i$ into Rayleigh's quotient to generate an even better estimate of λ_i than the initial estimate of $\mathbf{x}_i$. This estimate of λ_i and the initial estimate of $\mathbf{x}_i$ can be used in conjunction with inverse iteration to compute an improved estimate of $\mathbf{x}_i$. Then, the process is repeated as many times as necessary to achieve convergence to λ_i and $\mathbf{x}_i$. This is the essence of the *Rayleigh's quotient iteration* algorithm. The implication is that the algorithm permits the computation of any eigenvalue, eigenvector pair, provided there exists a good guess of the associated vector.

To describe Rayleigh's quotient iteration, we consider an initial vector $\mathbf{v}_0^*$ known to be closer to the eigenvector $\mathbf{x}_i$ than to any other eigenvector. Then, assuming that the initial vector has been normalized according to $\mathbf{v}_0 = \mathbf{v}_0^*/\|\mathbf{v}_0^*\|$, we can begin the iteration process, which is defined by

$$\mu_{p-1} = \mathbf{v}_{p-1}^T A \mathbf{v}_{p-1}, \qquad p = 1, 2, \ldots \tag{6.170}$$

and

$$\left(A - \mu_{p-1} I\right) \mathbf{v}_p^* = \mathbf{v}_{p-1}, \qquad \mathbf{v}_p = \mathbf{v}_p^*/\|\mathbf{v}_p\|, \qquad p = 1, 2, \ldots \tag{6.171a, b}$$

and we note that the shift changes from iteration to iteration. Equation (6.171a) is solved by Gaussian elimination with back-substitution. The process converges cubically in the neighborhood of each eigenvector (Ref. 13, p. 636).

Example 6.12

Solve the eigenvalue problem of Example 6.4 by means of Rayleigh's quotient iteration.

Ignoring the parameters k and m and using the mass and stiffness matrices given by Eqs. (a) of Example 6.4, we can write

$$A = M^{-1/2} K M^{-1/2} = \begin{bmatrix} 2.5 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3 \end{bmatrix} \tag{a}$$

and we note that the natural frequencies and modal vectors computed in Example 6.4 are related to the eigenvectors computed here by

$$\omega_i = \sqrt{\lambda_i}, \qquad \mathbf{u}_i = M^{-1/2}\mathbf{x}_i = \begin{bmatrix} 0.7071 & 0 & 0 \\ 0 & 0.5774 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_i, \qquad i = 1, 2, 3 \tag{b}$$

To begin the iteration process, we must choose an initial unit vector. To this end, we recognize that the system of Example 6.4 is the same as that of Example 4.6, and it represents the vibrating system of Fig. 4.8. The first eigenvector is characterized by no sign changes. Hence, we choose as initial unit vector

$$\mathbf{v}_0 = \frac{1}{\sqrt{3}} [1 \;\; 1 \;\; 1]^T \tag{c}$$

and note that $\mathbf{v}_0$ is a relatively crude guess. In fact, the only resemblance to the first eigenvector is that it has no sign changes. Inserting Eqs. (a) and (c) into Eq. (6.170) with $p = 1$ we obtain the first shift in the form

$$\mu_0 = \mathbf{v}_0^T A \mathbf{v}_0 = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2.5 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0.8026 \tag{d}$$

Then, introducing Eqs. (a), (c) and (d) into the Eq. (6.171a) with $p = 1$, we can write

$$(A - \mu_0 I)\mathbf{v}_1^* = \begin{bmatrix} 1.6974 & -1.2247 & 0 \\ -1.2247 & 0.8641 & -1.1547 \\ 0 & -1.1547 & 2.1974 \end{bmatrix} \mathbf{v}_1^* = \mathbf{v}_0 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{e}$$

The solution of Eq. (e), obtained by Gaussian elimination with back-substitution, is

$$\mathbf{v}_1^* = [-1.1545 \;\; -2.0716 \;\; -0.8258]^T \tag{f}$$

Upon normalization according to Eq. (6.171b), the first iterate is

$$\mathbf{v}_1 = [0.4597 \ 0.8249 \ 0.3288]^T \tag{g}$$

Next, we insert Eq. (g) into Eq. (6.170) with $p = 2$ and obtain the second shift

$$\mu_1 = \mathbf{v}_1^T A \mathbf{v}_1 = \begin{bmatrix} 0.4597 \\ 0.8249 \\ 0.3288 \end{bmatrix}^T \begin{bmatrix} 2.5 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3 \end{bmatrix} \begin{bmatrix} 0.4597 \\ 0.8249 \\ 0.3288 \end{bmatrix} = 0.4315 \tag{h}$$

so that, following the established pattern, we can write

$$(A - \mu_1 I)\mathbf{v}_2^* = \begin{bmatrix} 2.0685 & -1.2247 & 0 \\ -1.2247 & 1.2352 & -1.1547 \\ 0 & -1.1547 & 2.5685 \end{bmatrix} \mathbf{v}_2^* = \mathbf{v}_1 = \begin{bmatrix} 0.4597 \\ 0.8249 \\ 0.3288 \end{bmatrix} \tag{i}$$

Solving Eq. (i) for $\mathbf{v}_2^*$ and normalizing, we have

$$\mathbf{v}_2 = [0.4743 \ 0.8032 \ 0.3604]^T \tag{j}$$

Introducing Eq. (j) into Eq. (6.170) with $p = 3$, we obtain

$$\mu_2 = \mathbf{v}_2^T A \mathbf{v}_2 = 0.4257 \tag{k}$$

One more iteration corresponding to $p = 3$ yields

$$\mathbf{v}_3 = [0.4742 \ 0.8033 \ 0.3603]^T \tag{l}$$

Comparing Eqs. (j) and (l), we conclude that no additional iterations are necessary, so that we accept $\lambda_1 = \mu_2$ and $\mathbf{x}_1 = \mathbf{v}_3$ as the first eigenvalue and eigenvector of A. Hence, using Eqs. (b), we can write

$$\omega_1 = \sqrt{\lambda_1} = 0.6524, \qquad \mathbf{u}_1 = M^{-1/2}\mathbf{x}_1 = [0.3353 \ 0.4638 \ 0.3603]^T \tag{m}$$

Next, we propose to compute the second eigenvalue and eigenvector. To this end, we recognize that the second eigenvector is characterized by one sign change, so that we choose as the initial unit vector for the second mode

$$\mathbf{v}_0 = \frac{1}{\sqrt{3}}[-1 \ -1 \ 1]^T \tag{n}$$

Then, the first shift is

$$\mu_0 = \mathbf{v}_0^T A \mathbf{v}_0 = \frac{1}{3}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2.5 & -1.2247 & 0 \\ -1.2247 & 1.6667 & -1.1547 \\ 0 & -1.1547 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} = 2.3422 \tag{o}$$

Hence, the first iteration for the second mode is defined by

$$(A - \mu_0 I)\mathbf{v}_1^* = \begin{bmatrix} 0.1578 & -1.2247 & 0 \\ -1.2247 & -0.6755 & -1.1547 \\ 0 & -1.1547 & 0.6578 \end{bmatrix} \mathbf{v}_1^* = \mathbf{v}_0 = \frac{1}{\sqrt{3}}\begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \tag{p}$$

Solving Eq. (p) and normalizing, we obtain

$$\mathbf{v}_1 = [-0.5879 \ 0.1791 \ 0.7889]^T \tag{q}$$

This permits us to compute the second shift

$$\mu_1 = \mathbf{v}_1^T A \mathbf{v}_1 = 2.6926 \tag{r}$$

so that the next iteration is defined by

$$(A - \mu_1 I)\,\mathbf{v}_2^* = \begin{bmatrix} -0.1926 & -1.2247 & 0 \\ -1.2247 & -1.0259 & -1.1547 \\ 0 & -1.1547 & 0.3074 \end{bmatrix} \mathbf{v}_2^* = \mathbf{v}_1 = \begin{bmatrix} -0.5879 \\ 0.1791 \\ 0.7889 \end{bmatrix} \tag{s}$$

Upon solving Eq. (s) and normalizing, we can write

$$\mathbf{v}_2 = [-0.7425 \;\; 0.1405 \;\; 0.6550]^T \tag{t}$$

The next two iteration stages yield

$$\mu_2 = 2.7413, \qquad \mathbf{v}_3 = [-0.7448 \;\; 0.1466 \;\; 0.6510]^T \tag{u}$$

and

$$\mu_3 = 2.7410, \qquad \mathbf{v}_4 = [-0.7436 \;\; 0.1464 \;\; 0.6525]^T \tag{v}$$

It is obvious that convergence has been reached, so that we accept as eigenvalue and eigenvector $\lambda_2 = \mu_3$, $\mathbf{x}_2 = \mathbf{v}_4$. Hence, inserting λ_2 and $\mathbf{x}_2$ into Eqs. (b), we obtain

$$\omega_2 = \sqrt{\lambda_2} = 1.6556, \qquad \mathbf{u}_2 = M^{-1/2}\mathbf{x}_2 = [-0.5258 \;\; 0.0845 \;\; 0.6525]^T \tag{w}$$

Comparing Eqs. (m) and (w) to Eqs. (h) and (m) of Example 6.4, respectively, we conclude that the results are nearly identical. The only difference is in the top component of $\mathbf{u}_1$, which can be attributed to rounding errors. The computation of the third natural frequency and eigenvector is left as an exercise to the reader.

6.12 SIMULTANEOUS ITERATION

Many problems in structural dynamics involve mathematical models of high order. A complete solution of the eigenvalue problem for high-order systems is time-consuming, and may not even be necessary. In the first place, higher modes are characterized by high frequencies and can seldom be excited. Moreover, it is typical of discrete models of distributed systems that the higher modes tend to be inaccurate. In view of this, our interest is in a partial solution to the eigenvalue problem. This brings immediately to mind the power method, which iterates to one mode at a time. *Simultaneous iteration*, due to Jennings (Ref. 7) and Clint and Jennings (Ref. 1), can be regarded as an extension of the power method whereby iteration is carried out to a given number of modes simultaneously.

We are concerned with the eigenvalue problem

$$A\mathbf{x}_i = \lambda_i \mathbf{x}_i, \qquad \lambda_i = 1/\omega_i^2, \qquad i = 1, 2, \ldots, n \tag{6.172}$$

where A is a real symmetric matrix. The mutually orthogonal eigenvectors are assumed to be normalized so as to satisfy $\mathbf{x}_j^T \mathbf{x}_i = \delta_{ij}$. Simultaneous iteration is defined by the relation

$$V_p^* = AV_{p-1}, \qquad p = 1, 2, \ldots \tag{6.173}$$

where V_{p-1} is an $n \times m$ matrix of mutually orthonormal vectors $\mathbf{v}_i$ related to the matrix V_{p-1}^* of independent vectors $\mathbf{v}_i^*$ by

$$V_{p-1} = V_{p-1}^* U_{p-1}, \qquad p = 1, 2, \ldots \tag{6.174}$$

where U_{p-1} is an $m \times m$ upper triangular matrix. Equation (6.174) expresses the orthonormalization of m independent vectors. The orthonormalization can be carried out by means of the Gram-Schmidt method (Appendix B) or through solving an $m \times m$ eigenvalue problem, and we note that this must be done at every iteration step. Of course, the purpose of the orthogonalization process is to prevent all the vectors $\mathbf{v}_i$ from converging to $\mathbf{x}_1$, as they would in the absence of orthogonalization. The iteration process defined by Eqs. (6.173) and (6.174) converges with the result

$$\lim_{p\to\infty} V_p = X^{(m)}, \qquad \lim_{p\to\infty} U_p = \Lambda^{(m)} \tag{6.175a, b}$$

where $X^{(m)} = [\mathbf{x}_1\ \mathbf{x}_2\ \dots\ \mathbf{x}_m]$ is the matrix of the m lowest orthonormal eigenvectors and $\Lambda^{(m)} = \operatorname{diag}[\lambda_1\ \lambda_2\ \dots\ \lambda_m]$ is the diagonal matrix of the m lowest eigenvalues.

The preceding formulation can be modified to accommodate eigenvalue problems in terms of two real symmetric matrices of the type

$$K\mathbf{x}_i = \omega_i^2 M_i \mathbf{x}_i, \qquad i = 1, 2, \dots, n \tag{6.176}$$

In this case, the iteration process is defined by

$$K V_p^* = M V_{p-1}, \qquad p = 1, 2, \dots \tag{6.177}$$

But, unlike Eq. (6.173), in which V_p^* is obtained by simple matrix multiplication, the solution of Eq. (6.177) for V_p^* requires the solution of n nonhomogeneous algebraic equations, which can be obtained by Gaussian elimination with back-substitution. Note that, although the orthonormalization process can still be written in the form (6.174), in this case the matrix V_p must be orthonormal with respect to M.

The Gram-Schmidt orthogonalization method often gives inaccurate results in the sense that the vectors are not quite orthogonal. A method proposed by Clint and Jennings (Ref. 1), whereby orthogonalization is achieved by solving an eigenvalue problem of reduced order, is quite efficient computationally. The iteration process is based on the same Eq. (6.177). On the other hand, the Gram-Schmidt orthonormalization given by Eq. (6.174) is replaced by one requiring the solution of the eigenvalue problem

$$K_p P_p = M_p P_p \Lambda_p, \qquad p = 1, 2, \dots \tag{6.178}$$

where

$$K_p = \left(V_p^*\right)^T K V_p^*, \qquad M_p = \left(V_p^*\right)^T M V_p^*, \qquad p = 1, 2, \dots \tag{6.179a, b}$$

are $m \times m$ real symmetric matrices. The solution of the eigenvalue problem (6.178) consists of the matrix of eigenvectors P_p and the matrix of eigenvalues Λ_p, where P_p is assumed to be normalized with respect to M_p so that $P_p^T M_p P_p = I$. Then, the next iteration step is carried out with the matrix

$$V_p = V_p^* P_p, \qquad p = 1, 2, \dots \tag{6.180}$$

which is orthonormal with respect to M. Indeed, using Eq. (6.179b), we can write

$$P_p^T M_p P_p = P_p^T \left(V_p^*\right)^T M V_p^* P_p = V_p^T M V_p = I \tag{6.181}$$

The iteration process converges, but now the convergence expressions are

$$\lim_{p\to\infty} V_p = X^{(m)}, \qquad \lim_{p\to\infty} \Lambda_p = \Lambda^{(m)} \tag{6.182}$$

This version of simultaneous iteration is sometimes referred to as *subspace iteration.*

It should be pointed out that, although we must solve an eigenvalue problem at each iteration step, these eigenvalue problems are of significantly lower order than the order of the eigenvalue problem for the original system, $m \ll n$. We should also note that M_p and K_p tend to become diagonal as p increases, which tends to expedite the solution of the associated eigenvalue problem.

Example 6.13

Use simultaneous iteration in conjunction with the Gram-Schmidt orthogonalization to obtain the two lowest eigensolutions of the matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix} \tag{a}$$

As the initial trial matrix, we use

$$V_0 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{b}$$

Inserting Eqs. (a) and (b) into Eq. (6.173) with $p = 1$, we obtain

$$V_1^* = AV_0 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 3.5 & 0.5 \\ 4.5 & 1.5 \\ 5 & 2 \end{bmatrix} \tag{c}$$

Introducing the notation $V_1 = [\mathbf{v}_{11}\ \ \mathbf{v}_{12}]$, $V_1^* = \left[\mathbf{v}_{11}^*\ \ \mathbf{v}_{12}^*\right]$, the Gram-Schmidt orthonormalization given by Eq. (6.174) with $p = 1$ can be expressed in the form

$$\begin{gathered} \mathbf{v}_{11} = u_{11}\mathbf{v}_{11}^*, \qquad \mathbf{v}_{12} = u_{12}\mathbf{v}_{11}^* + u_{22}\mathbf{v}_{12}^*, \qquad \mathbf{v}_{11}^T\mathbf{v}_{12} = 0 \\ \|\mathbf{v}_{11}\| = 1, \qquad \|\mathbf{v}_{12}\| = 1 \end{gathered} \tag{d}$$

where u_{11}, u_{12} and u_{22} are the elements of the upper triangular matrix U_1. Solving Eqs. (d), we obtain

$$V_1 = \begin{bmatrix} 0.255031 & -0.622200 \\ 0.446304 & -0.571751 \\ 0.573819 & 0.151347 \\ 0.637577 & 0.512895 \end{bmatrix}, \qquad U_1 = \begin{bmatrix} 0.127515 & -0.311101 \\ 0 & 1.034198 \end{bmatrix} \tag{e}$$

Next, we introduce Eq. (a) and the first of Eqs. (e) into Eq. (6.173) with $p = 2$ and obtain

$$V_2^* = AV_1 = \begin{bmatrix} 1.912731 & -0.529709 \\ 3.570431 & -0.437218 \\ 4.781827 & 0.227024 \\ 5.419404 & 0.739919 \end{bmatrix} \tag{f}$$

so that, following the Gram-Schmidt orthonormalization established in Eqs. (d), we have

$$V_2 = \begin{bmatrix} 0.230865 & -0.606625 \\ 0.430948 & -0.574671 \\ 0.577162 & 0.051955 \\ 0.654117 & 0.546865 \end{bmatrix}, \qquad U_2 = \begin{bmatrix} 0.120699 & -0.037139 \\ 0 & 1.011100 \end{bmatrix} \tag{g}$$

Repeating the procedure for $p = 3, 4, 5$ and 6, we obtain the results

$$V_3 = \begin{bmatrix} 0.228337 & -0.591880 \\ 0.428828 & -0.574997 \\ 0.577340 & 0.019516 \\ 0.656237 & 0.564515 \end{bmatrix}, \qquad U_3 = \begin{bmatrix} 0.120616 & -0.004398 \\ 0 & 1.001850 \end{bmatrix} \tag{h}$$

$$V_4 = \begin{bmatrix} 0.228051 & -0.584031 \\ 0.428562 & -0.575867 \\ 0.577350 & 0.007691 \\ 0.656502 & 0.572037 \end{bmatrix}, \qquad U_4 = \begin{bmatrix} 0.120615 & -0.000537 \\ 0 & 1.000325 \end{bmatrix} \tag{i}$$

$$V_5 = \begin{bmatrix} 0.228018 & -0.580322 \\ 0.428530 & -0.576565 \\ 0.577350 & 0.003115 \\ 0.656534 & 0.575141 \end{bmatrix}, \qquad U_5 = \begin{bmatrix} 0.120615 & -0.000063 \\ 0 & 1.000053 \end{bmatrix} \tag{j}$$

and

$$V_6 = \begin{bmatrix} 0.228014 & -0.578650 \\ 0.428526 & -0.576970 \\ 0.577350 & 0.001284 \\ 0.656538 & 0.576426 \end{bmatrix}, \qquad U_6 = \begin{bmatrix} 0.120615 & -0.000007 \\ 0 & 1.000011 \end{bmatrix} \tag{k}$$

At this point, we observe that we are near convergence. In fact λ_1 and $\mathbf{v}_1$ have already converged, but λ_2 and $\mathbf{v}_2$ have not, although they are very close. Indeed, the two lowest eigenvalues are $\lambda_1 = 0.120615$ and $\lambda_2 = 1$.

Example 6.14

Solve the problem of Example 6.13 by means of subspace iteration.

The subspace iteration is defined by Eq. (6.177), or

$$KV_p^* = MV_{p-1}, \qquad p = 1, 2, \ldots \tag{a}$$

where, for the system of Example 6.13,

$$K = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \qquad M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{b}$$

As in Example 6.13, we use the initial trial matrix

$$V_0 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{c}$$

Inserting Eqs. (b) and (c) into Eq. (a) and using Gaussian elimination with back-substitution, we obtain

$$V_1^* = \begin{bmatrix} 2 & 0 \\ 3.5 & 0.5 \\ 4.5 & 1.5 \\ 5 & 2 \end{bmatrix} \tag{d}$$

Then, using Eqs. (6.179) with $p = 1$, we compute

$$K_1 = \left(V_1^*\right)^T K V_1^* = \begin{bmatrix} 7.5 & 2 \\ 2 & 1.5 \end{bmatrix}, \qquad M_1 = \left(V_1^*\right)^T M V_1^* = \begin{bmatrix} 61.5 & 18.5 \\ 18.5 & 6.5 \end{bmatrix} \tag{e}$$

Inserting Eqs. (e) into Eq. (6.178) with $p = 1$ and solving the 2×2 eigenvalue problem, we obtain

$$\Lambda_1 = \begin{bmatrix} 0.120715 & 0 \\ 0 & 1.044503 \end{bmatrix}, \qquad P_1 = \begin{bmatrix} 0.116049 & -0.315557 \\ 0.037835 & 1.033506 \end{bmatrix} \tag{f}$$

so that the next iteration stage begins with the matrix

$$V_1 = V_1^* P_1 = \begin{bmatrix} 0.232097 & -0.631114 \\ 0.425088 & -0.587696 \\ 0.578972 & 0.130252 \\ 0.655914 & 0.489226 \end{bmatrix} \tag{g}$$

Inserting Eq. (g) into Eq. (a) with $p = 2$ and using Gaussian elimination in conjunction with back-substitution, we obtain

$$V_2^* = \begin{bmatrix} 1.892071 & -0.599332 \\ 3.552044 & -0.567549 \\ 4.786930 & 0.051929 \\ 5.442843 & 0.541156 \end{bmatrix} \tag{h}$$

so that, using Eqs. (6.179) with $p = 2$, we have

$$\begin{aligned} K_2 &= \left(V_2^*\right)^T K V_2^* = \begin{bmatrix} 8.290608 & 0.004655 \\ 0.004655 & 0.983305 \end{bmatrix} \\ M_2 &= \left(V_2^*\right)^T M V_2^* = \begin{bmatrix} 68.736187 & 0.044068 \\ 0.044068 & 0.976857 \end{bmatrix} \end{aligned} \tag{i}$$

Introducing Eqs. (i) into Eq. (6.178) with $p = 2$ and solving the eigenvalue problem, we obtain

$$\Lambda_2 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.006627 \end{bmatrix}, \qquad P_2 = \begin{bmatrix} 0.120617 & -0.000660 \\ 0.000092 & 1.011791 \end{bmatrix} \tag{j}$$

so that, using Eq. (6.180) with $p = 2$, we compute the 4×2 orthonormal matrix

$$V_2 = V_2^* P_2 = \begin{bmatrix} 0.228160 & -0.607646 \\ 0.428383 & -0.576585 \\ 0.577388 & 0.049384 \\ 0.656547 & 0.543946 \end{bmatrix} \tag{k}$$

Comparing Eqs. (f) and (j) on the one hand and Eqs. (g) and (k) on the other, we observe that the first mode has almost reached convergence.

The third iteration stage begins by inserting Eq. (k) into Eq. (a) with $p = 3$ and solving for V_3^* by Gaussian elimination and back-substitution. For brevity, we omit details and list the numerical results for the third through the sixth iteration stages.

$$\Lambda_3 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.001075 \end{bmatrix}, \qquad V_3 = \begin{bmatrix} 0.228020 & -0.592002 \\ 0.428520 & -0.575226 \\ 0.577350 & 0.019206 \\ 0.656540 & 0.564163 \end{bmatrix} \tag{l}$$

$$\Lambda_4 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.000185 \end{bmatrix}, \qquad V_4 = \begin{bmatrix} 0.228014 & -0.584042 \\ 0.428525 & -0.575897 \\ 0.577350 & 0.007655 \\ 0.656539 & 0.571995 \end{bmatrix} \tag{m}$$

$$\Lambda_5 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.000033 \end{bmatrix}, \qquad V_5 = \begin{bmatrix} 0.228013 & -0.580322 \\ 0.428525 & -0.576570 \\ 0.577350 & 0.003111 \\ 0.656539 & 0.575137 \end{bmatrix} \tag{n}$$

$$\Lambda_6 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.000006 \end{bmatrix}, \qquad V_6 = \begin{bmatrix} 0.228013 & -0.578650 \\ 0.428525 & -0.576971 \\ 0.577350 & 0.001283 \\ 0.656539 & 0.576426 \end{bmatrix} \tag{o}$$

Finally, we begin the seventh iteration stage with V_6, use Eq. (a) with $p = 7$ and compute

$$V_7^* = \begin{bmatrix} 1.890427 & -0.577912 \\ 3.552841 & -0.577174 \\ 4.786730 & 0.000535 \\ 5.443268 & 0.576961 \end{bmatrix} \tag{p}$$

so that, inserting Eq. (p) into Eqs. (6.179) with $p = 7$, we obtain

$$\begin{aligned} K_7 &= \left(V_7^*\right)^T K V_7^* = \begin{bmatrix} 8.290859 & 0 \\ 0 & 0.999998 \end{bmatrix} \\ M_7 &= \left(V_7^*\right)^T M V_7^* = \begin{bmatrix} 68.738349 & 0 \\ 0 & 0.999996 \end{bmatrix} \end{aligned} \tag{q}$$

which indicates that convergence has been achieved. Indeed, solving the 2×2 eigenvalue problem, Eq. (6.178) with $p = 7$, we have

$$\Lambda_7 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.000001 \end{bmatrix}, \qquad P_7 = \begin{bmatrix} 0.120615 & 0 \\ 0 & 1.000002 \end{bmatrix} \tag{r}$$

so that, using Eq. (6.180) with $p = 7$, we obtain the matrix of orthonormal eigenvectors

$$V_7 = V_7^* P_7 = \begin{bmatrix} 0.228013 & -0.577913 \\ 0.428525 & -0.577175 \\ 0.577350 & 0.000535 \\ 0.656539 & 0.576962 \end{bmatrix} \tag{s}$$

At this point, we regard Λ_7 and V_7 as the matrices of eigenvalues and eigenvectors, respectively. It should be pointed out here that the actual eigenvalues are $\lambda_1 = 0.120615$ and $\lambda_2 = 1$.

Contrasting the results obtained here with those obtained in Example 6.13, we conclude that subspace iteration and simultaneous iteration have similar convergence characteristics.

6.13 THE POWER METHOD FOR NONSYMMETRIC EIGENVALUE PROBLEMS

As shown in Sec. 4.8, the eigenvalue problem for nonconservative systems can be written in the form

$$A\mathbf{x} = \lambda\mathbf{x} \tag{6.183}$$

where A is a real nonsymmetric $2n \times 2n$ matrix. The eigenvalue problem admits $2n$ solutions in the form of the eigenvalue, eigenvector pairs $\lambda_r, \mathbf{x}_r$ $(r = 1, 2, \ldots, 2n)$, in the sense that a given eigenvector $\mathbf{x}_r$ belongs to the eigenvalue λ_r, and not to any other eigenvalue. The eigenvector $\mathbf{x}_r$ is a $2n$-dimensional state vector, which implies that it has the form $\mathbf{x}_r = \begin{bmatrix}\mathbf{q}_r^T & \lambda_r\mathbf{q}_r^T\end{bmatrix}^T$, where $\mathbf{q}_r$ is an n-dimensional configuration vector. In general, the eigenvalues and eigenvectors tend to be complex quantities, although real eigenvalues and eigenvectors are possible. If an eigenvalue is complex (real), then the eigenvector belonging to it is complex (real). Because A is real, if $\lambda_r, \mathbf{x}_r$ is an eigensolution, then the complex conjugates $\bar{\lambda}_r, \bar{\mathbf{x}}_r$ also constitute an eigensolution. Also from Sec. 4.8, we recall that associated with the eigenvalue problem (6.183) there is the adjoint eigenvalue problem

$$A^T\mathbf{y} = \lambda\mathbf{y} \tag{6.184}$$

which admits eigensolutions $\lambda_s, \mathbf{y}_s$ $(s = 1, 2, \ldots, 2n)$, where $\mathbf{y}_s$ are referred to as adjoint eigenvectors. Hence, whereas the eigenvalues of A and A^T are the same, the eigenvectors are different. If Eq. (6.184) is transposed, then the eigenvector, albeit transposed, appears to the left of A, in contrast to Eq. (6.183) in which the eigenvector appears to the right of A. Because of this juxtaposition of eigenvectors relative to A, $\mathbf{x}_r$ are generally referred to as right eigenvectors and $\mathbf{y}_s$ as left eigenvectors. The two sets of eigenvectors are biorthogonal and can be conveniently normalized so as to satisfy the biorthonormality relations

$$\mathbf{y}_s^T\mathbf{x}_r = \delta_{rs}, \qquad r, s = 1, 2, \ldots, 2n \tag{6.185}$$

Algorithms for the solution of the nonsymmetric eigenvalue problem do not come close to possessing the desirable characteristics of algorithms for the symmetric eigenvalue problem. In this section, we consider solving the nonsymmetric eigenvalue problem by matrix iteration using the power method. There are significant differences between the case in which the eigensolutions are real and the case in which they are complex. We consider first the case in which the eigensolutions are real and assume that the eigenvalues are arranged in descending order of magnitude, or $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_{2n}|$, where we note that the eigenvalues can be of both signs. In this case, the iteration process for the dominant eigensolution $\lambda_1, \mathbf{x}_1$ is exactly the same as that for real symmetric matrices described in Sec. 6.3. Differences begin to surface with the computation of the first subdominant eigensolution. Indeed, before λ_2 and $\mathbf{x}_2$ can be computed, it is necessary to compute $\mathbf{y}_1$, because $\mathbf{x}_2$ is orthogonal to $\mathbf{y}_1$ and not to $\mathbf{x}_1$. To this end, we observe that the same iteration process using the transposed matrix A^T yields the dominant adjoint eigensolution $\lambda_1, \mathbf{y}_1$. Then, normalizing $\mathbf{x}_1$ and $\mathbf{y}_1$ so as to satisfy $\mathbf{y}_1^T\mathbf{x}_1 = 1$ and using the analogy with the deflation process discussed in Sec. 6.3, we iterate to the first subdominant eigensolution λ_2, $\mathbf{x}_2$ by using the deflated matrix

$$A_2 = A - \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T \tag{6.186a}$$

Similarly, the transposed matrix

$$A_2^T = A^T - \lambda_1 \mathbf{y}_1 \mathbf{x}_1^T \tag{6.186b}$$

is used to iterate to $\lambda_2, \mathbf{y}_2$. The iteration process can be generalized by using the deflated matrix

$$A_k = A_{k-1} - \lambda_{k-1}\mathbf{x}_{k-1}\mathbf{y}_{k-1}^T, \qquad A_1 = A, \qquad k = 2, 3, \dots, 2n \tag{6.187a}$$

to iterate to the eigensolution λ_k, $\mathbf{x}_k$ and the transposed matrix

$$A_k^T = A_{k-1}^T - \lambda_{k-1}\mathbf{y}_{k-1}\mathbf{x}_{k-1}^T, \qquad k = 2, 3, \dots, 2n \tag{6.187b}$$

to iterate to $\lambda_k, \mathbf{y}_k$. It should be stressed here that, before constructing a new deflated matrix, the right and left eigenvectors must be normalized so as to satisfy $\mathbf{y}_{k-1}^T\mathbf{x}_{k-1} = 1$.

When the eigensolutions are complex, matrix iteration by the power method for nonsymmetric matrices A is appreciably more involved than when the eigensolutions are real. Indeed, in iterating with complex vectors, one must bear in mind that complex quantities are characterized not only by magnitude but also by phase angle. Figure 6.4 shows the components of two successive iterated vectors, $\mathbf{v}_\ell$ and $\mathbf{v}_{\ell+1}$, in the complex plane. At first sight, there appears to be no relation between the two vectors. This is true even when the two vectors represent the same eigenvector, which happens at convergence. Indeed, because they are complex, for two vectors to represent the same eigenvector, the magnitude ratios of homologous components must be the same and the phase angle difference between any pair of components from one vector must be the same as the phase angle difference between homologous components from the other vector. The equivalence of two complex vectors, or lack of it, can be verified by bringing homologous components of $\mathbf{v}_\ell$ and $\mathbf{v}_{\ell+1}$, say $v_{i,\ell}$ and $v_{i,\ell+1}$, into coincidence through a rotation of one of the vectors. This difficulty in interpreting complex iteration results makes convergence difficult to recognize. Some of these difficulties can be mitigated by insisting that the iteration process be carried out with real iterated vectors. Fortunately, this is possible because complex eigensolutions for real matrices occur in pairs of complex conjugates. Even in working with real iterated vectors, however, recognizing convergence is not as simple as in the case of real eigensolutions.

Our next task is to derive a convergence criterion. To this end, we first assume that the eigenvalues satisfy $|\lambda_1| = |\lambda_2| = |\overline{\lambda}_1| \geq |\lambda_3| = |\lambda_4| = |\overline{\lambda}_3| \geq \dots \geq |\lambda_{2n-1}| = |\lambda_{2n}| = |\overline{\lambda}_{2n-1}|$. Then, we express the dominant pair of complex conjugate eigenvalues in the form

$$\lambda_1 = |\lambda_1|e^{i\theta_1}, \qquad \lambda_2 = \overline{\lambda}_1 = |\lambda_1|e^{-i\theta_1} \tag{6.188}$$

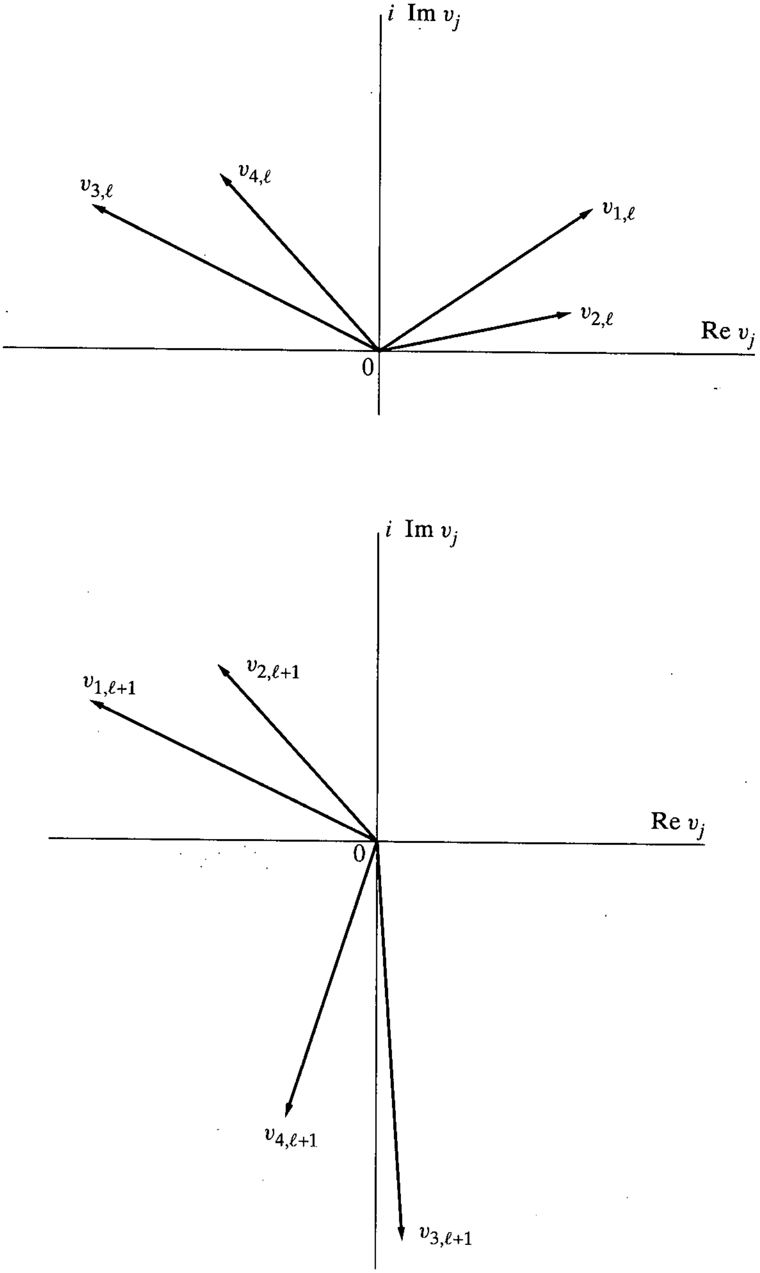


Figure 6.4 Two successive iterated vectors in the complex plane

where the magnitude $|\lambda_1|$ is such that $|\lambda_1| > |\lambda_r|$, $r = 3, 4, \ldots, 2n$. But, according to the expansion theorem of Sec. 4.8, Eq. (4.181), we can assume an initial trial vector as a linear combination of the right eigenvectors of the form

$$\mathbf{v}_0 = \sum_{r=1}^{2n} c_r \mathbf{x}_r = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_{2n}\mathbf{x}_{2n} = c_1\mathbf{x}_1 + \bar{c}_1\bar{\mathbf{x}}_1 + \ldots + \bar{c}_n\bar{\mathbf{x}}_n \quad (6.189)$$

and we observe that $\mathbf{v}_0$ is a real vector, because every sum $c_r\mathbf{x}_r + \bar{c}_r\bar{\mathbf{x}}_r$ of complex conjugates is a real vector. Then, following the analogy with the matrix iteration of Sec. 6.3 and using Eqs. (6.188), the pth iterate can be written as

$$\begin{aligned} \mathbf{v}_p &= A\mathbf{v}_{p-1} = c_1\lambda_1^p\mathbf{x}_1 + \bar{c}_1\bar{\lambda}_1^p\bar{\mathbf{x}}_1 + \boldsymbol{\epsilon}_p \\ &= |\lambda_1|^p\left(c_1e^{ip\theta_1}\mathbf{x}_1 + \bar{c}_1e^{-ip\theta_1}\bar{\mathbf{x}}_1\right) + \boldsymbol{\epsilon}_p \end{aligned} \quad (6.190)$$

As $p \to \infty$, $\boldsymbol{\epsilon}_p \to \mathbf{0}$, from which it follows that

$$\lim_{p\to\infty} \mathbf{v}_p = c_1\lambda_1^p\mathbf{x}_1 + \bar{c}_1\bar{\lambda}_1^p\bar{\mathbf{x}}_1 \quad (6.191)$$

Equation (6.191) demonstrates that the iteration process converges to the pair of dominant complex conjugate eigensolutions, but it provides no clues as to when convergence has been achieved and what the numerical value of the eigensolutions is. To address these issues, we consider three consecutive iterates $\mathbf{v}_p$, $\mathbf{v}_{p+1}$ and $\mathbf{v}_{p+2}$, use Eq. (6.190) and form the expression

$$\begin{aligned} \mathbf{v}_{p+2} + \xi\mathbf{v}_{p+1} + \eta\mathbf{v}_p &= \left(\lambda_1^2 + \xi\lambda_1 + \eta\right)c_1\lambda_1^p\mathbf{x}_1 + \left(\bar{\lambda}_1^2 + \xi\bar{\lambda}_1 + \eta\right)\bar{c}_1\bar{\lambda}_1^p\bar{\mathbf{x}}_1 \\ &\quad + \boldsymbol{\epsilon}_{p+2} + \xi\boldsymbol{\epsilon}_{p+1} + \eta\boldsymbol{\epsilon}_p \end{aligned} \quad (6.192)$$

But, if λ_1 and $\bar{\lambda}_1$ are roots of the quadratic equation

$$\lambda^2 + \xi\lambda + \eta = 0 \quad (6.193)$$

i.e., if they are such that

$$\begin{matrix}\lambda_1 \\ \bar{\lambda}_1\end{matrix} = -\frac{1}{2}\xi \pm \frac{i}{2}\sqrt{4\eta - \xi^2} \quad (6.194)$$

and if p is sufficiently large that $\boldsymbol{\epsilon}_p$, $\boldsymbol{\epsilon}_{p+1}$ and $\boldsymbol{\epsilon}_{p+2}$ are negligibly small, then Eq. (6.192) reduces to

$$\mathbf{v}_{p+2} + \xi\mathbf{v}_{p+1} + \eta\mathbf{v}_p = \mathbf{0} \quad (6.195)$$

Equation (6.195) states that *at convergence three successive iterates are linearly dependent.*

Next, we propose to develop a convergence criterion. To this end, we consider three successive iterates $\mathbf{v}_\ell$, $\mathbf{v}_{\ell+1}$ and $\mathbf{v}_{\ell+2}$. Before reaching convergence a linear combination of the type given by Eq. (6.195) is not zero, so that we define an error vector in the form

$$\mathbf{r}_\ell = \mathbf{v}_{\ell+2} + \xi_\ell\mathbf{v}_{\ell+1} + \eta_\ell\mathbf{v}_\ell \quad (6.196)$$

Then, to minimize the error, we use the method of least squares and write

$$\begin{aligned}\frac{\partial\left(\mathbf{r}_\ell^T\mathbf{r}_\ell\right)}{\partial\xi_\ell} &= \mathbf{v}_{\ell+1}^T\left(\mathbf{v}_{\ell+2}+\xi_\ell\mathbf{v}_{\ell+1}+\eta_\ell\mathbf{v}_\ell\right)=0\\ \frac{\partial\left(\mathbf{r}_\ell^T\mathbf{r}_\ell\right)}{\partial\eta_\ell} &= \mathbf{v}_\ell^T\left(\mathbf{v}_{\ell+2}+\xi_\ell\mathbf{v}_{\ell+1}+\eta_\ell\mathbf{v}_\ell\right)=0\end{aligned} \tag{6.197}$$

Equations (6.197) represent two homogeneous algebraic equations in the unknowns ξ_ℓ and η_ℓ and have the solution

$$\begin{aligned}\xi_\ell &= \frac{\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_\ell\right)\left(\mathbf{v}_\ell^T\mathbf{v}_{\ell+2}\right)-\left(\mathbf{v}_\ell^T\mathbf{v}_\ell\right)\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_{\ell+2}\right)}{\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_{\ell+1}\right)\left(\mathbf{v}_\ell^T\mathbf{v}_\ell\right)-\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_\ell\right)^2},\\ \eta_\ell &= \frac{\left(\mathbf{v}_\ell^T\mathbf{v}_{\ell+1}\right)\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_{\ell+2}\right)-\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_{\ell+1}\right)\left(\mathbf{v}_\ell^T\mathbf{v}_{\ell+2}\right)}{\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_{\ell+1}\right)\left(\mathbf{v}_\ell^T\mathbf{v}_\ell\right)-\left(\mathbf{v}_{\ell+1}^T\mathbf{v}_\ell\right)^2}\end{aligned} \tag{6.198}$$

Equations (6.198) form the basis for the desired convergence criterion, which can be stated as follows: *The iteration process achieves convergence when three successive iterates* $\mathbf{v}_\ell$, $\mathbf{v}_{\ell+1}$ *and* $\mathbf{v}_{\ell+2}$ *are such that* ξ_ℓ *and* η_ℓ, *as calculated by means of Eqs.* (6.198), *reach constant values* ξ *and* η. Then, the dominant eigenvalues λ_1 and $\bar{\lambda}_1$ are calculated by inserting the constants ξ and η thus obtained into Eq. (6.194).

The question remains as to how to determine the eigenvector $\mathbf{x}_1$ belonging to λ_1. In view of the fact that eigenvectors can only be determined within a multiplying constant, upon considering Eq. (6.191), we can write without loss of generality

$$\operatorname{Re}\mathbf{x}_1 = \frac{1}{2}\left(\mathbf{x}_1+\bar{\mathbf{x}}_1\right) = \frac{1}{2}\mathbf{v}_p \tag{6.199}$$

Similarly, we can write

$$\mathbf{v}_{p+1} = A\mathbf{v}_p = A\left(\mathbf{x}_1+\bar{\mathbf{x}}_1\right) = \lambda_1\mathbf{x}_1+\bar{\lambda}_1\bar{\mathbf{x}}_1 = 2\left(\operatorname{Re}\lambda_1\operatorname{Re}\mathbf{x}_1-\operatorname{Im}\lambda_1\operatorname{Im}\mathbf{x}_1\right) \tag{6.200}$$

from which we conclude that

$$\operatorname{Im}\mathbf{x}_1 = \frac{1}{\operatorname{Im}\lambda_1}\left(\operatorname{Re}\lambda_1\operatorname{Re}\mathbf{x}_1-\frac{1}{2}\mathbf{v}_{p+1}\right) = -\frac{1}{\sqrt{4\eta-\xi^2}}\left(\frac{1}{2}\xi\mathbf{v}_p+\mathbf{v}_{p+1}\right) \tag{6.201}$$

The process just described iterates to the two dominant complex conjugate eigensolutions $\lambda_1, \mathbf{x}_1$ and $\lambda_2 = \bar{\lambda}_1$, $\mathbf{x}_2 = \bar{\mathbf{x}}_1$. An entirely similar process involving A^T instead of A can be used to iterate to the dominant adjoint eigensolutions $\lambda_1, \mathbf{y}_1$ and $\lambda_2 = \bar{\lambda}_1, \mathbf{y}_2 = \bar{\mathbf{y}}_1$. To compute subdominant eigensolutions by the power method, it is necessary to use matrix deflation. Although in presenting the deflation process earlier in this section it was assumed that the eigensolutions are all real, the same process is applicable to complex conjugate eigensolutions. The only difference is that we must consider pairs of complex conjugates at a time, which enables us to construct real deflated matrices. Hence, using Eq. (6.187a), we can express the deflated matrix for iteration to $\lambda_3, \mathbf{x}_3$ in the form of the real matrix

$$A_3 = A - \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T - \lambda_2 \mathbf{x}_2 \mathbf{y}_2^T = A - \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T - \bar{\lambda}_1 \bar{\mathbf{x}}_1 \bar{\mathbf{y}}_1^T \tag{6.202}$$

Then, postmultiplying Eq. (6.202) by $\mathbf{x}_i$ and considering Eqs. (6.185), we obtain

$$\begin{aligned} A_3 \mathbf{x}_i &= A\mathbf{x}_i - \lambda_1 \mathbf{x}_1 \mathbf{y}_1^T \mathbf{x}_i - \lambda_2 \mathbf{x}_2 \mathbf{y}_2^T \mathbf{x}_i \\ &= \lambda_i \mathbf{x}_i - \lambda_1 \mathbf{x}_1 \delta_{1i} - \lambda_2 \mathbf{x}_2 \delta_{2i} = \begin{cases} 0 & \text{if } i = 1, 2 \\ \lambda_i \mathbf{x}_i & \text{if } i \neq 1, 2 \end{cases} \end{aligned} \tag{6.203}$$

It follows that A_3 has the eigenvalues $0, 0, \lambda_3, \lambda_4, \ldots, \lambda_{2n}$ and the eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots, \mathbf{x}_{2n}$. Hence, an iteration process using A_3 in conjunction with an initial vector $\mathbf{v}_0$ in the form of Eq. (6.189) iterates to $\lambda_3, \mathbf{x}_3$ in the same way the process using A iterates to $\lambda_1, \mathbf{x}_1$. Of course, the next step is to use the transposed deflated matrix

$$A_3^T = A^T - \lambda_1 \mathbf{y}_1 \mathbf{x}_1^T - \lambda_2 \mathbf{y}_2 \mathbf{x}_2^T \tag{6.204}$$

to iterate to $\lambda_3, \mathbf{y}_3$.

The iteration process just described can be generalized by using the deflated matrix

$$A_{2k-1} = A_{2k-3} - \lambda_{2k-3} \mathbf{x}_{2k-3} \mathbf{y}_{2k-3}^T - \lambda_{2k-2} \mathbf{x}_{2k-2} \mathbf{y}_{2k-2}^T, \qquad k = 2, 3, \ldots, n \tag{6.205a}$$

to iterate to $\lambda_{2k-1}, \mathbf{x}_{2k-1}$ and the transposed matrix

$$A_{2k-1}^T = A_{2k-3}^T - \lambda_{2k-3} \mathbf{y}_{2k-3} \mathbf{x}_{2k-3}^T - \lambda_{2k-2} \mathbf{y}_{2k-2} \mathbf{x}_{2k-2}^T, \qquad k = 2, 3, \ldots, n \tag{6.205b}$$

to iterate to $\lambda_{2k-1}, \mathbf{y}_{2k-1}$, where $A_1 = A$.

In general, a real nonsymmetric matrix A can have both real and complex conjugate eigensolutions, and quite often it is not possible to know in advance whether the dominant eigenvalue is of one type or the other. Hence, the question arises as to the iteration process to be used, the one for real eigenvalues or the one for complex conjugate eigenvalues. We address this problem by proposing a test based on the assumption that the dominant eigenvalue is real. To this end, we write the iteration process in the form

$$A\mathbf{v}_{\ell-1} = \mathbf{v}_\ell = \lambda^{(\ell)} \mathbf{v}_{\ell-1}, \qquad \ell = 1, 2, \ldots \tag{6.206}$$

and introduce thc error vector

$$\mathbf{r}_\ell = \mathbf{v}_\ell - \lambda^{(\ell)} \mathbf{v}_{\ell-1}, \qquad \ell = 1, 2, \ldots \tag{6.207}$$

Then, using the method of least squares in a manner akin to the one used earlier in this section, we can write

$$\frac{\partial \left(\mathbf{r}_\ell^T \mathbf{r}_\ell\right)}{\partial \lambda^{(\ell)}} = 2\mathbf{v}_{\ell-1}^T \left(\mathbf{v}_\ell - \lambda^{(\ell)} \mathbf{v}_{\ell-1}\right) = 0, \qquad \ell = 1, 2, \ldots \tag{6.208}$$

which yields

$$\lambda^{(\ell)} = \frac{\mathbf{v}_{\ell-1}^T \mathbf{v}_\ell}{\|\mathbf{v}_{\ell-1}\|^2}, \qquad \ell = 1, 2, \ldots \tag{6.209}$$

The approach consists of using Eq. (6.206) to compute several successive estimates, say $\lambda^{(1)}$, $\lambda^{(2)}$ and $\lambda^{(3)}$. If the sequence $\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}$ reveals a stable pattern, i.e., a tendency to converge, then the dominant eigenvalue is real and the process converges to λ_1. On the other hand, if the sequence exhibits an erratic pattern with no sign of convergence, such as a change in sign with each iteration step, then the dominant eigenvalue is likely to be complex. In this case, we switch from Eq. (6.209) to the convergence criterion given by Eqs. (6.198). In this regard, it should be pointed out that the computations involving Eqs. (6.206) and (6.209) were not all wasted, as the same iterates $\mathbf{v}_1, \mathbf{v}_2, \ldots$ are used in Eqs. (6.198). Clearly, the same procedure must be followed in iterating to each subsequent eigensolution.

The deflation process for nonsymmetric matrices is also due to Hotelling (Ref. 6). It has the advantage of simplicity, but has poor stability characteristics. Other deflation methods, with better stability characteristics, are discussed by Wilkinson (Ref. 13).

Example 6.15

Solve the eigenvalue problem for the damped two-degree-of-freedom system shown in Fig. 6.5. Use the parameters $m_1 = 1,\ m_2 = 2,\ c_1 = c_2 = 0.2,\ k_1 = 1$ and $k_2 = 4$.

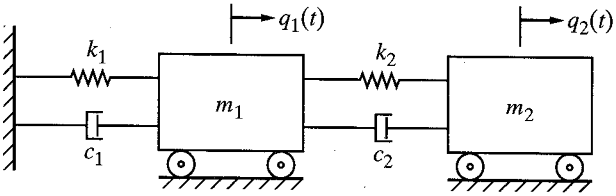


Figure 6.5 Damped two-degree-of-freedom system

The kinetic energy, Rayleigh's dissipation function and potential energy for the system are as follows:

$$T = \frac{1}{2}\dot{\mathbf{q}}^T M\dot{\mathbf{q}}, \qquad \mathcal{F} = \frac{1}{2}\dot{\mathbf{q}}^T C\dot{\mathbf{q}}, \qquad V = \frac{1}{2}\mathbf{q}^T K\mathbf{q} \tag{a}$$

where $\mathbf{q} = [q_1\ q_2]^T$ is the configuration vector and

$$M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$
$$C = \begin{bmatrix} c_1 + c_2 & -c_2 \\ -c_2 & c_2 \end{bmatrix} = \begin{bmatrix} 0.4 & -0.2 \\ -0.2 & 0.2 \end{bmatrix} \tag{b}$$
$$K = \begin{bmatrix} k_1 + k_2 & -k_2 \\ -k_2 & k_2 \end{bmatrix} = \begin{bmatrix} 5 & -4 \\ -4 & 4 \end{bmatrix}$$

are the mass, damping and stiffness matrix, respectively. Equations (b) permit us to calculate the coefficient matrix

$$A = \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5 & 4 & -0.4 & 0.2 \\ 2 & -2 & 0.1 & -0.1 \end{bmatrix} \tag{c}$$

Because the damping coefficients are relatively small, the eigensolutions are likely to consist of two pairs of complex conjugates. Nevertheless, to illustrate the testing procedure, we operate on the assumption that we do not know whether the dominant eigensolution is real or complex and proceed with the test based on Eq. (6.209). To this end, we choose the initial vector $\mathbf{v}_0 = [1\ \ 1\ \ 1\ \ 1]^T$, use Eq. (6.206) with $\ell = 1$ and obtain the first iterate

$$\mathbf{v}_1 = A\mathbf{v}_0 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5 & 4 & -0.4 & 0.2 \\ 2 & -2 & 0.1 & -0.1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1.2 \\ 0 \end{bmatrix} \tag{d}$$

Then, using Eq. (6.209) with $\ell = 1$, we can write

$$\lambda^{(1)} = \frac{\mathbf{v}_0^T \mathbf{v}_1}{\|\mathbf{v}_0\|^2} = 0.2 \tag{e}$$

Following the same pattern, we compute the second iterate

$$\mathbf{v}_2 = A\mathbf{v}_1 = [-1.2\quad 0\quad -0.52\quad -0.12]^T \tag{f}$$

so that

$$\lambda^{(2)} = \frac{\mathbf{v}_1^T \mathbf{v}_2}{\|\mathbf{v}_1\|^2} = -0.1674 \tag{g}$$

In view of the fact that $\lambda^{(1)}$ and $\lambda^{(2)}$ are of opposite signs, we conclude that the dominant eigensolution must be complex.

At this point, we turn our attention to the iteration process for complex eigensolutions based on Eqs. (6.198). However, before we can compute ξ_1 and η_1, we must have the third iterate $\mathbf{v}_3$, so that we write

$$\mathbf{v}_3 = A\mathbf{v}_2 = [-0.52\quad -0.12\quad 6.184\quad -2.44]^T \tag{h}$$

Hence, using Eqs. (6.198) with $\ell = 1$, we compute

$$\begin{aligned} \xi_1 &= \frac{(\mathbf{v}_2^T\mathbf{v}_1)(\mathbf{v}_1^T\mathbf{v}_3) - (\mathbf{v}_1^T\mathbf{v}_1)(\mathbf{v}_2^T\mathbf{v}_3)}{(\mathbf{v}_2^T\mathbf{v}_2)(\mathbf{v}_1^T\mathbf{v}_1) - (\mathbf{v}_2^T\mathbf{v}_1)^2} = 2.240665 \\ \eta_1 &= \frac{(\mathbf{v}_1^T\mathbf{v}_2)(\mathbf{v}_2^T\mathbf{v}_3) - (\mathbf{v}_2^T\mathbf{v}_2)(\mathbf{v}_1^T\mathbf{v}_3)}{(\mathbf{v}_2^T\mathbf{v}_2)(\mathbf{v}_1^T\mathbf{v}_1) - (\mathbf{v}_2^T\mathbf{v}_1)^2} = 2.718437 \end{aligned} \tag{i}$$

Next, we obtain the fourth iterate

$$\mathbf{v}_4 = A\mathbf{v}_3 = [6.184\quad -2.44\quad -0.8416\quad 0.0624]^T \tag{j}$$

which permits us to compute

$$\begin{aligned} \xi_2 &= \frac{(\mathbf{v}_3^T\mathbf{v}_2)(\mathbf{v}_2^T\mathbf{v}_4) - (\mathbf{v}_2^T\mathbf{v}_2)(\mathbf{v}_3^T\mathbf{v}_4)}{(\mathbf{v}_3^T\mathbf{v}_3)(\mathbf{v}_2^T\mathbf{v}_2) - (\mathbf{v}_3^T\mathbf{v}_2)^2} = 0.424882 \\ \eta_2 &= \frac{(\mathbf{v}_2^T\mathbf{v}_3)(\mathbf{v}_3^T\mathbf{v}_4) - (\mathbf{v}_3^T\mathbf{v}_3)(\mathbf{v}_2^T\mathbf{v}_4)}{(\mathbf{v}_3^T\mathbf{v}_3)(\mathbf{v}_2^T\mathbf{v}_2) - (\mathbf{v}_3^T\mathbf{v}_2)^2} = 4.619323 \end{aligned} \tag{k}$$

At this point the procedure is clear, so that we merely list the results leading to convergence

$$\begin{aligned}
\mathbf{v}_5 &= [-0.8416 \quad 0.0624 \quad -40.331 \quad 17.158]^T \\
\mathbf{v}_6 &= [-40.331 \quad 17.158 \quad 24.021 \quad -7.557]^T \\
\mathbf{v}_7 &= [24.021 \quad -7.557 \quad 259.165 \quad -111.819]^T \\
\mathbf{v}_8 &= [0.259165 \quad -0.111189 \quad -0.276365 \quad 0.100255]^T \times 10^3 \\
\mathbf{v}_9 &= [-0.276365 \quad 0.100255 \quad -1.612504 \quad 0.704306]^T \times 10^3 \\
\mathbf{v}_{10} &= [-1.612504 \quad 0.704306 \quad 2.568705 \quad -0.984920]^T \times 10^3 \\
\mathbf{v}_{11} &= [2.568705 \quad -0.984920 \quad 9.655277 \quad -4.278257]^T \times 10^3 \\
\mathbf{v}_{12} &= [9.655277 \quad -4.278257 \quad -21.500970 \quad 8.500605]^T \times 10^3 \\
\mathbf{v}_{13} &= [-21.500970 \quad 8.500605 \quad -55.088906 \quad 24.866912]^T \times 10^3
\end{aligned} \tag{l}$$

and

$$\begin{aligned}
\xi_3 &= 0.564512, & \eta_3 &= 6.643729 \\
\xi_4 &= 0.450787, & \eta_4 &= 6.648980 \\
\xi_5 &= 0.450232, & \eta_5 &= 6.697480 \\
\xi_6 &= 0.445574, & \eta_6 &= 6.694921 \\
\xi_7 &= 0.445399, & \eta_7 &= 6.697054 \\
\xi_8 &= 0.445216, & \eta_8 &= 6.696866 \\
\xi_9 &= 0.445200, & \eta_9 &= 6.696963 \\
\xi_{10} &= 0.445193, & \eta_{10} &= 6.696953 \\
\xi_{11} &= 0.445192, & \eta_{11} &= 6.696957
\end{aligned} \tag{m}$$

Hence, convergence has been achieved after eleven iterations, so that we have

$$\xi = \xi_{11} = 0.445192, \qquad \eta = \eta_{11} = 6.696957 \tag{n}$$

In addition, the iterate at convergence and the one immediately following convergence are

$$\begin{aligned}
\mathbf{v}_p &= \mathbf{v}_{11} = [2.568705 \quad -0.984920 \quad 9.655277 \quad -4.278257]^T \times 10^3 \\
\mathbf{v}_{p+1} &= \mathbf{v}_{12} = [9.655277 \quad -4.278257 \quad -21.500970 \quad 8.500605]^T \times 10^3
\end{aligned} \tag{o}$$

Inserting Eqs. (n) into Eq. (6.194), we obtain the dominant pair of complex conjugate eigenvalues

$$\begin{matrix}\lambda_1 \\ \overline{\lambda}_1\end{matrix} = -\frac{1}{2}\xi \pm \frac{i}{2}\sqrt{4\eta - \xi^2} = -0.222596 \pm 2.578257i \tag{p}$$

Moreover, introducing Eqs. (o) into Eqs. (6.199) and (6.201), we can write

$$\text{Re}\,\mathbf{x}_1 = \frac{1}{2}\mathbf{v}_p = [1.284353 \quad -0.492460 \quad 4.827639 \quad -2.139129]^T \times 10^3$$

$$\operatorname{Im}\mathbf{x}_1 = -\frac{1}{\sqrt{4\eta - \xi^2}}\left(\frac{1}{2}\xi\mathbf{v}_p + \mathbf{v}_{p+1}\right) \tag{q}$$

$$= [-1.983328 \quad 0.872197 \quad 3.752865 \quad -1.463834]^T \times 10^3$$

Following the same procedure with A^T as with A, we obtain the same values for ξ and η as the values given by Eqs. (n), as well as the iterates $\mathbf{w}_p$ and $\mathbf{w}_{p+1}$. Hence, the eigenvalues λ_1 and $\bar{\lambda}_1$ remain the same as those given by Eq. (p), as expected. On the other hand, using the analogy with Eqs. (q) in conjunction with the iterates $\mathbf{w}_p$ and $\mathbf{w}_{p+1}$, we compute Re $\mathbf{y}_1$ and Im $\mathbf{y}_1$. At this point, we are in a position to normalize the right and left eigenvectors belonging to λ_1 by writing $\mathbf{y}_1^T\mathbf{x}_1 = 1$, according to Eq. (6.185) with $r = s = 1$. The normalized right and left eigenvectors and the complex conjugates are

$$\begin{matrix}\mathbf{x}_1\\ \bar{\mathbf{x}}_1\end{matrix} = \begin{bmatrix} 0.138171 \mp 0.156517i \\ -0.054440 \pm 0.069777i \\ 0.372785 \pm 0.391081i \\ -0.167784 \mp 0.155893i \end{bmatrix}, \quad \begin{matrix}\mathbf{y}_1\\ \bar{\mathbf{y}}_1\end{matrix} = \begin{bmatrix} 1.259528 \pm 1.241648i \\ -1.074222 \mp 1.053273i \\ 0.526929 \mp 0.435837i \\ -0.423504 \pm 0.395909i \end{bmatrix} \tag{r}$$

To compute the eigensolutions $\lambda_3, \mathbf{x}_3, \bar{\lambda}_3, \bar{\mathbf{x}}_3$, we must produce the deflated matrix A_3. Introducing Eqs. (c), (p) and (r) into Eq. (6.202), we obtain

$$A_3 = A - \lambda_1\mathbf{x}_1\mathbf{y}_1^T - \bar{\lambda}_1\bar{\mathbf{x}}_1\bar{\mathbf{y}}_1^T = A - 2\operatorname{Re}\left(\lambda_1\mathbf{x}_1\mathbf{y}_1^T\right)$$

$$= \begin{bmatrix} 0.020244 & -0.010725 & -0.087772 & 0.926615 \\ 0.064174 & -0.057336 & 0.463307 & 0.606787 \\ 2.292088 & -2.204316 & 0.148014 & -0.120941 \\ -1.102962 & 0.639655 & -0.060470 & -0.025354 \end{bmatrix} \tag{s}$$

Using the same matrix iteration process with A_3 as with A_1 we compute the subdominant eigenvalues

$$\begin{matrix}\lambda_3\\ \bar{\lambda}_3\end{matrix} = -0.027404 \pm 0.545795i \tag{t}$$

and the normalized right and left eigenvectors

$$\begin{matrix}\mathbf{x}_3\\ \bar{\mathbf{x}}_3\end{matrix} = \begin{bmatrix} 0.433803 \mp 0.452506i \\ 0.516529 \mp 0.525161i \\ 0.235088 \pm 0.249168i \\ 0.272475 \pm 0.296311i \end{bmatrix}, \quad \begin{matrix}\mathbf{y}_3\\ \bar{\mathbf{y}}_3\end{matrix} = \begin{bmatrix} 0.186794 \pm 0.111816i \\ 0.339785 \pm 0.366585i \\ 0.275539 \mp 0.274295i \\ 0.655774 \mp 0.636294i \end{bmatrix} \tag{u}$$

6.14 REDUCTION OF NONSYMMETRIC MATRICES TO HESSENBERG FORM

As discussed in Sec. 6.8, Givens' method for the computation of the eigenvalues of real symmetric matrices requires that the matrix be in tridiagonal form. Then, in Sec. 6.9, we presented the QR method, designed for the same purpose. Convergence of the QR method can be very slow, unless the matrix is in tridiagonal form. However, in general matrices defining the eigenvalue problem for vibrating conservative systems, albeit symmetric, tend to be fully populated. Hence, if the interest lies in solving the eigenvalue problem by one of these two algorithms, it is necessary to transform the original matrix to tridiagonal form. But, because the tridiagonal matrix must have the same eigenvalues as the original matrix, the two matrices must be related by a similarity transformation. In the case of symmetric matrices, the transformation is actually orthogonal, i.e., a special case of the similarity transformation. Three methods capable of transforming a real symmetric matrix to tridiagonal form are Givens' method (Sec. 6.5), Householder's method (Sec. 6.6) and Lanczos' method (Sec. 6.7), among others. Denoting the original matrix by A and the tridiagonal matrix by T, the orthogonal transformation has the form

$$T = P^T A P \tag{6.210}$$

where the transformation matrix P is orthonormal, i.e., it satisfies $P^T P = I$. In Givens' method, P is obtained through a sequence of $(n-1)(n-2)/2$ rotations, where n is the order of the matrix A, and in Householder's method through a series of $n-2$ reflections. In contrast, Lanczos' method is a direct method whereby the matrix P is obtained by means of a recursive process.

In the case of a nonsymmetric matrix A, significant simplification of the eigenvalue problem accrues by reducing the matrix to *Hessenberg form*, denoted here by H. There are two such forms, an *upper Hessenberg* and a *lower Hessenberg form.* Of the two, the upper Hessenberg form is more commonly encountered, because various algorithms for the computation of eigenvalues are based on such forms. The upper Hessenberg matrix, defined by $h_{ij} = 0$, $i \geq j+2$, has the form

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} & \dots & h_{1,n-1} & h_{1n} \\ h_{21} & h_{22} & h_{23} & \dots & h_{2,n-1} & h_{2n} \\ 0 & h_{32} & h_{33} & \dots & h_{3,n-1} & h_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \dots & h_{n-1,n-1} & h_{n-1,n} \\ 0 & 0 & 0 & \dots & h_{n,n-1} & h_{nn} \end{bmatrix} \tag{6.211}$$

A real nonsymmetric matrix can be reduced to Hessenberg form by means of either Givens' method or Householder's method. In this regard, a slight inconvenience arises, because Givens' method and Householder's method are commonly formulated so as to reduce symmetric matrices to tridiagonal form through annihilation of the upper off-diagonal elements. Of course, as a by-product due to symmetry, the corresponding lower off-diagonal elements are also annihilated. In the case of nonsymmetric matrices, Givens' algorithm and Householder's algorithm, as presented in Secs. 6.5 and 6.6, respectively, result in lower Hessenberg matrices. Modification of the algorithms to produce upper Hessenberg matrices is relatively trivial and amounts to an interchange of the subscripts involved.

In this section, we consider a direct method for reducing a real nonsymmetric matrix to upper Hessenberg form. The approach reminds of the Gaussian elimination for the reduction of a general matrix to upper triangular form through elementary operations presented in Sec. 6.1. Using the same idea, and recognizing that the current objective is to solve an eigenvalue problem, instead of solving nonhomogeneous algebraic equations, we propose to reduce the nonsymmetric $n \times n$ matrix A to an upper Hessenberg form H by means of a transformation that preserves the eigenvalues. Hence, whereas in both cases use is made of elementary transformations, our interest lies in a transformation to Hessenberg form by means of a similarity transformation. In view of this, we consider the transformation

$$P^{-1}AP = H \tag{6.212}$$

where P is a unit lower triangular matrix of the type encountered in Sec. 6.1. But, because our objective is to produce only an upper Hessenberg matrix, instead of an upper triangular matrix, the transformation matrix P has the form

$$P = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & p_{32} & 1 & \dots & 0 & 0 \\ 0 & p_{42} & p_{43} & \dots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & p_{n-1,2} & p_{n-1,3} & \dots & 1 & 0 \\ 0 & p_{n2} & p_{n3} & \dots & p_{n,n-1} & 1 \end{bmatrix} \tag{6.213}$$

The problem consists of determining not only the matrix H but also the matrix P. To solve this problem, we premultiply Eq. (6.212) by P and obtain

$$AP = PH \tag{6.214}$$

which represents a set of n^2 equations in n^2 unknowns, $(n^2 - 3n + 2)/2$ nonzero p_{ij} and $(n^2 + 3n - 2)/2$ nonzero h_{ij}.

Because of the structure of P and H, we can determine the nonzero elements p_{ij} and h_{ij} by equating every column of AP to the corresponding column of PH. In particular, by equating the rth columns, we determine column r of $H (r = 1, 2, \dots, n)$ and column $r + 1$ of P $(r = 1, 2, \dots, n - 2)$. We note that it is only necessary to determine $n - 2$ columns of P, because the first and last columns are known. The solution can be carried out in a recursive manner, beginning by equating the first column on both sides of Eq. (6.214), or

$$A\mathbf{e}_1 = P\mathbf{h}_1 \tag{6.215}$$

where $\mathbf{e}_1 = [1 \;\; 0 \;\; 0 \;\; \dots \;\; 0]^T$ is recognized as the first standard unit vector and $\mathbf{h}_1 = [h_{11} \; h_{21} \; 0 \; \dots \; 0]^T$. Considering Eq. (6.213), Eq. (6.215) yields

$$h_{11} = a_{11}, \qquad h_{21} = a_{21} \tag{6.216a}$$

which determines the first column of H and

$$p_{i2} = \frac{a_{i1}}{h_{21}}, \qquad i = 3, 4, \dots, n \tag{6.216b}$$

which defines the second column of P. Note that, in choosing the first standard unit vector $\mathbf{e}_1$ as the first column of P, we obtain the simplest solution. Indeed, we could have used any other vector instead of $\mathbf{e}_1$ and still obtained a solution. Equating the second column on both sides of Eq. (6.214), we have

$$A\mathbf{p}_2 = P\mathbf{h}_2 \tag{6.217}$$

where $\mathbf{p}_2 = [0\ 1\ p_{32}\ p_{42}\ \dots\ p_{n2}]^T$, $\mathbf{h}_2 = [h_{12}\ h_{22}\ h_{32}\ 0\ \dots\ 0]^T$. Following the same pattern, the first three rows of Eq. (6.217) determine the second column of H, or by components

$$h_{i2} = a_{i2} + \sum_{k=3}^{n} a_{ik} p_{k2} - \sum_{k=2}^{i-1} p_{ik} h_{k2}, \qquad i = 1, 2, 3 \tag{6.218a}$$

and the remaining $n - 3$ rows can be used to determine the third column of P in the form

$$p_{i3} = \frac{1}{h_{32}} \left(a_{i2} + \sum_{k=3}^{n} a_{ik} p_{k2} \right) - p_{i2} h_{22}, \qquad i = 4, 5, \dots, n \tag{6.218b}$$

The procedure can be generalized by writing

$$h_{ir} = a_{ir} + \sum_{k=r+1}^{n} a_{ik} p_{kr} - \sum_{k=2}^{i-1} p_{ik} h_{kr},$$
$$r = 1, 2, \dots, n-1;\ i = 1, 2, \dots, r+1 \tag{6.219a}$$

$$p_{i,r+1} = \frac{1}{h_{r+1,r}} \left(a_{ir} + \sum_{k=r+1}^{n} a_{ik} p_{kr} - \sum_{k=2}^{r} p_{ik} h_{kr} \right),$$
$$r = 1, 2, \dots, n-2;\ i = r+2, r+3, \dots, n \tag{6.219b}$$

$$h_{in} = a_{in} - \sum_{k=2}^{i-1} p_{ik} h_{kn}, \qquad i = 1, 2, \dots, n \tag{6.219c}$$

where it must be remembered that $p_{k1} = 0\,(k = 2, 3, \dots, n)$. The computations alternate between Eqs. (6.219a) and (6.219b), until $r = n - 1$, at which point the computations skip from Eq. (6.219a) to Eq. (6.219c). This can be explained by the fact that the matrix P has only $n - 2$ unknown columns.

Equation (6.219b) contains a potential source of difficulties in that it requires division by $h_{r+1,r}$. Indeed, if $h_{r+1,r}$ is zero, the algorithm breaks down. Even when $h_{r+1,r}$ is only very small, and not necessarily zero, numerical instability can occur. In such cases, it becomes necessary to carry out suitable row and column interchanges, which can be done by means of permutation matrices, Eq. (6.36). However, unlike in Gaussian elimination, the transformation using the permutation matrix must be a similarity transformation so as to preserve the eigenvalues. But, because the inverse of a permutation matrix is equal to the permutation matrix itself, the similarity transformation implies premultiplication and postmultiplication by the same permutation matrix, which amounts to an interchange of both rows and the corresponding columns. As an example, if rows s and t are to be interchanged, so are columns s and t. In the case in which the matrix A represents the coefficient matrix in the state equations, such row and column interchanges are an absolute necessity. In fact, entire blocks must be interchanged, leading to the matrix

$$A' = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} 0 & I \\ -M^{-1}K & -M^{-1}C \end{bmatrix} \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} = \begin{bmatrix} -M^{-1}C & -M^{-1}K \\ I & 0 \end{bmatrix} \tag{6.220}$$

Example 6.16

Derive the upper Hessenberg matrix corresponding to the coefficient matrix of Example 6.15.

Inserting Eq. (c) of Example 6.15 into Eq. (6.220), we obtain the coefficient matrix

$$A' = \begin{bmatrix} -M^{-1}C & -M^{-1}K \\ I & 0 \end{bmatrix} = \begin{bmatrix} -0.4 & 0.2 & -5 & 4 \\ 0.1 & -0.1 & 2 & -2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{a}$$

In deriving the upper Hessenberg matrix corresponding to A' by means of Eqs. (6.219), we will treat A' as if it were A, in the sense that we will ignore the prime.

Using Eq. (6.219a) with $r = 1$, we obtain

$$h_{11} = a_{11} = -0.4, \qquad h_{21} = a_{21} = 0.1 \tag{b}$$

Then, from Eq. (6.219b) with $r = 1$, we have

$$p_{32} = \frac{a_{31}}{h_{21}} = 10, \qquad p_{42} = \frac{a_{41}}{h_{21}} = 0 \tag{c}$$

Next, we let $r = 2$ in Eqs. (6.219a) and (6.219b) and write

$$\begin{aligned} h_{12} &= a_{12} + a_{13}p_{32} + a_{14}p_{42} = -49.8 \\ h_{22} &= a_{22} + a_{23}p_{32} + a_{24}p_{42} = 19.9 \\ h_{32} &= a_{32} + a_{33}p_{32} + a_{34}p_{42} - p_{32}h_{22} = -199 \end{aligned} \tag{d}$$

and

$$p_{43} = \frac{1}{h_{32}}(a_{42} + a_{43}p_{32} + a_{44}p_{42} - p_{42}h_{22}) = -\frac{1}{199} = -0.005025 \tag{e}$$

respectively. At this point, the pattern changes somewhat. Indeed, Eq. (6.219a) with $r = 3$ yields

$$\begin{aligned} h_{13} &= a_{13} + a_{14}p_{43} = -\frac{999}{199} = -5.020101 \\ h_{23} &= a_{23} + a_{24}p_{43} = \frac{400}{199} = 2.010050 \\ h_{33} &= a_{33} + a_{34}p_{43} - h_{23}p_{32} = -\frac{4000}{199} = -20.100503 \\ h_{43} &= a_{43} + a_{44}p_{43} - h_{23}p_{42} - h_{33}p_{43} = -\frac{4000}{199^2} = -0.101008 \end{aligned} \tag{f}$$

Then, skipping to Eq. (6.219c), we obtain

$$\begin{aligned} h_{14} &= a_{14} = 4 \\ h_{24} &= a_{24} = -2 \\ h_{34} &= a_{34} - p_{32}h_{24} = 20 \\ h_{44} &= a_{44} - p_{42}h_{24} - p_{43}h_{34} = \frac{20}{199} = 0.100503 \end{aligned} \tag{g}$$

Hence, using Eqs. (b), (d), (f) and (g), the desired upper Hessenberg matrix is

$$H = \begin{bmatrix} -0.4 & -49.8 & -5.020101 & 4 \\ 0.1 & 19.9 & 2.010050 & -2 \\ 0 & -199 & -20.100503 & 20 \\ 0 & 0 & -0.101008 & 0.100503 \end{bmatrix} \tag{h}$$

Moreover, using Eqs. (c) and (e), the transformation matrix is

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 10 & 1 & 0 \\ 0 & 0 & -0.005025 & 1 \end{bmatrix} \tag{i}$$

6.15 THE QR METHOD FOR NONSYMMETRIC EIGENVALUE PROBLEMS

As demonstrated in Sec. 6.8, Givens' method for the computation of the eigenvalues of a real symmetric tridiagonal matrix is very efficient and easy to implement. Hence, the question arises as to whether the method can be extended to nonsymmetric matrices. Whereas general nonsymmetric matrices can be reduced to tridiagonal form, the process is potentially unstable (Ref. 13, p. 404). However, there is a much more serious obstacle preventing the extension of the method to nonsymmetric matrices. Indeed, in the case of symmetric tridiagonal matrices T the polynomials $p_i(\lambda)$ corresponding to the principal minor determinants of the matrix $T - \lambda I$ form a Sturm sequence. Givens' method is based on the Sturm sequence property, which in turn is based on the separation theorem. The Sturm sequence property cannot be demonstrated for nonsymmetric matrices, as the separation theorem holds only for real symmetric matrices.

In contrast with Givens' method for the computation of eigenvalues, the QR method does work for nonsymmetric matrices. In fact, the QR method is the most effective algorithm for computing the eigenvalues of general matrices. Still, before computational efficiencies can be realized, it is necessary to reduce the general matrix to Hessenberg form and to carry out shifts in origin. When all the eigenvalues are real, the QR method reduces the Hessenberg matrix to triangular form, with the eigenvalues lying on the main diagonal. Our interest lies in real matrices, so that if there are complex eigenvalues, then they occur in pairs of complex conjugates. In this case, the matrix differs from a triangular matrix in that there is a 2 × 2 matrix straddling the main diagonal for every pair of complex conjugate eigenvalues, where the complex pair represents the eigenvalues of the 2 × 2 matrix.

For real eigenvalues, the QR algorithm for upper Hessenberg matrices is essentially the same as for tridiagonal matrices. In fact, even when complex eigenvalues occur, the algorithm is for the most part the same as that described in Sec. 6.9. Differences begin to arise when the eigenvalues of the 2 × 2 matrix in the lower right corner turn complex. Indeed, the algorithm of Sec. 6.9 is based on the implicit assumption that the shifts are real, and here we are faced with the problem of complex shifts, an undesirable prospect. To avoid complex shifts, it is necessary to modify the algorithm by considering two complete iteration steps at a time. To this end, we assume that, upon completion of s iteration steps with real shifts, the 2 × 2 matrix in the lower right corner of A_s has the eigenvalues μ_s and $\overline{\mu}_s$. Then, using μ_s as a first shift, we can write the next iteration step in the form

$$A_s - \mu_s I = Q_s R_s, \qquad A_{s+1} = \mu_s I + R_s Q_s \tag{6.221a, b}$$

where Q_s is an orthogonal matrix, $Q_s^T Q_s = I$, and R_s is an upper triangular matrix. Moreover, using $\overline{\mu}_s$ as a second shift, we can define the following iteration step as

$$A_{s+1} - \overline{\mu}_s I = Q_{s+1} R_{s+1}, \qquad A_{s+2} = \overline{\mu}_s I + R_{s+1} Q_{s+1} \tag{6.222a, b}$$

We should note here that A_s, A_{s+1} and A_{s+2} are all Hessenberg matrices.

At this point, we begin to develop an algorithm obviating the problem of complex shifts. To this end, we premultiply Eq. (6.221a) by Q_s^T, postmultiply the result by Q_s and write

$$Q_s^T A_s Q_s - \mu_s Q_s^T Q_s = Q_s^T Q_s R_s Q_s \tag{6.223}$$

so that, recalling that $Q_s^T Q_s = I$ and using Eq. (6.221b), we have

$$A_{s+1} = Q_s^T A_s Q_s \tag{6.224}$$

Next, we premultiply Eq. (6.222a) by Q_s, postmultiply by R_s and use Eqs. (6.221a) and (6.224) to obtain

$$\begin{aligned} Q_s A_{s+1} R_s - \overline{\mu}_s Q_s R_s &= Q_s Q_s^T A_s Q_s R_s - \overline{\mu}_s Q_s R_s \\ &= (A_s - \overline{\mu}_s I)(A_s - \mu_s I) = Q_s Q_{s+1} R_{s+1} R_s \end{aligned} \tag{6.225}$$

Then, introducing the notation

$$Q_s Q_{s+1} = Q, \qquad R_{s+1} R_s = R \tag{6.226a, b}$$

we can rewrite Eq. (6.225) in the form

$$R = Q^T (A_s - \overline{\mu}_s I)(A_s - \mu_s I) \tag{6.227}$$

and we observe that R is an upper triangular matrix, Q is an orthogonal matrix and $(A_s - \overline{\mu}_s I)(A_s - \mu_s I) = (A_s - \mathrm{Re}\, \mu_s I)^2 + (\mathrm{Im}\, \mu_s I)^2$ is a real matrix. Hence, the algorithm, as described by Eq. (6.227), amounts to the upper triangularization of a real matrix. The actual objective of the algorithm, however, is the matrix Q, and not R, because it is Q that must be used to determine the matrix A_{s+2} required for the next iteration step. Indeed, using Eq. (6.224) with s replaced by $s+1$ in conjunction with Eq. (6.226a), as well as Eq. (6.224) itself, we can write

$$A_{s+2} = Q_{s+1}^T A_{s+1} Q_{s+1} = Q_{s+1}^T Q_s^T A_s Q_s Q_{s+1} = Q^T A_s Q \tag{6.228}$$

The computation of Q can be carried out by Givens' method. Ordinarily, the triangularization of a matrix would require $n(n-1)/2$ rotations Θ_k, $(k = 1, 2, \ldots n(n-1)/2)$ in the planes $(1, 2)$, $(1, 3)$, $\ldots$, $(1, n)$, $(2, 3)$, $\ldots$, $(2, n)$, $\ldots$, $(n-1, n)$, so that

$$Q^T = \Theta_{n(n-1)/2} \cdots \Theta_2 \Theta_1 \tag{6.229}$$

We observe, however, that the matrix $(A_s - \overline{\mu}_s I)(A_s - \mu_s I)$ represents the product of a Hessenberg matrix and its complex conjugate and its elements corresponding to the rows $i = 4, 5, \ldots, n$ and the columns $j = 1, 2, \ldots, i-3$ are zero. It follows that many of the rotation matrices are identity matrices. For example, $\Theta_3, \Theta_4, \ldots, \Theta_n$, corresponding to the planes $(1,4), (1,5), \ldots, (1,n)$, are all identity matrices. In fact, only $(n-1)(n-2)/2$ rotations are necessary for the computation of Q^T.

The algorithm consists of performing two QR iteration steps at a time, yielding upper Hessenberg matrices $A_{s+2}, A_{s+4}, \ldots$, as can be concluded from Eq. (6.228). Convergence is achieved when the sequence of upper Hessenberg matrices leads to an isolated, constant 2×2 matrix in the lower right corner, which implies that the element $(n-1, n-2)$ has been annihilated and the elements of the 2×2 matrix no longer change. At this point, the pair of complex conjugate eigenvalues is equal to the pair of complex conjugate shifts. Then, the last two rows and columns can be deleted from the matrix and the process continued with an $(n-2) \times (n-2)$ deflated matrix for the next eigenvalue or eigenvalue pair.

Example 6.17

Compute the eigenvalues of the Hessenberg matrix of Example 6.16 by means of the QR method with shifts.

Adopting the notation of this section, the Hessenberg matrix of Example 6.16 is

$$A_1 = A = \begin{bmatrix} -0.4 & -49.8 & -5.020101 & 4 \\ 0.1 & 19.9 & 2.010050 & -2 \\ 0 & -199 & -20.100503 & 20 \\ 0 & 0 & -0.101008 & 0.100503 \end{bmatrix} \tag{a}$$

We begin the process by computing the eigenvalues of the 2×2 matrix in the lower right corner of A_1. The eigenvalues are 0 and 20, with the eigenvalue closest to a_{44} being 0. Hence, the first iteration stage is carried without a shift, so that the first stage is given by

$$A_1 = Q_1 R_1, \qquad A_2 = R_1 Q_1 \tag{b}$$

where Q_1 is obtained by means of three successive Givens rotations in the planes (1,2), (2,3) and (3,4). The process was demonstrated in Example 6.10, so that we dispense with the intermediate steps and list the results of the decomposition of A_1 directly

$$Q_1 = \begin{bmatrix} 0.970143 & -0.008803 & -0.005858 & 0.242305 \\ -0.242536 & -0.035212 & -0.023433 & 0.969220 \\ 0 & 0.999341 & -0.000877 & 0.036285 \\ 0 & 0 & 0.999708 & 0.024170 \end{bmatrix} \tag{c}$$

$$R_1 = \begin{bmatrix} -0.412311 & -53.139555 & -5.357722 & 4.365641 \\ 0 & -199.131207 & -20.113844 & 20.022034 \\ 0 & 0 & -0.101037 & 0.106361 \\ 0 & 0 & 0 & -0.241093 \end{bmatrix} \tag{d}$$

Inserting Eqs. (c) and (d) into the second of Eqs. (b), we obtain

$$A_2 = \begin{bmatrix} 12.488235 & -3.479423 & 5.616691 & -51.692714 \\ 48.296412 & -13.088826 & 24.700036 & -193.247856 \\ 0 & -0.100970 & -0.106418 & -0.001095 \\ 0 & 0 & -0.241023 & -0.005827 \end{bmatrix} \tag{e}$$

The eigenvalues of the 2 × 2 matrix in the lower right corner of A_2 are 0.108723 and −0.008132, and they are both real. The eigenvalue closest to the lower right corner element is the negative one, so that we choose as our shift $\mu_2 = -0.001832$. The results of the next iteration stage are

$$Q_2 = \begin{bmatrix} -0.250494 & 0.651485 & 0.669048 & 0.255332 \\ -0.968118 & -0.168567 & -0.173112 & -0.066065 \\ 0 & 0.739697 & -0.628711 & -0.239938 \\ 0 & 0 & 0.356552 & -0.934276 \end{bmatrix} \tag{f}$$

$$R_2 = \begin{bmatrix} -49.886898 & 13.535232 & -25.319500 & 200.035472 \\ 0 & -0.136502 & -0.419696 & -1.102569 \\ 0 & 0 & -0.675983 & -1.129950 \\ 0 & 0 & 0 & -0.433695 \end{bmatrix}. \tag{g}$$

and

$$A_3 = R_2Q_2 = \begin{bmatrix} -0.615469 & -53.510937 & 51.521792 & -194.445077 \\ 0.132150 & -0.295570 & -0.105625 & 1.139822 \\ 0 & -0.500023 & 0.013981 & 1.217879 \\ 0 & 0 & -0.154635 & 0.397058 \end{bmatrix} \tag{i}$$

The eigenvalues of the 2 × 2 matrix in the lower right corner of A_3 are $0.205520 \pm 0.389409i$, so that they are complex conjugates.

Next, we begin with the algorithm using complex shifts described in this section. Letting $\mu_3 = 0.205520 + 0.389409i$ and $\overline{\mu}_3 = 0.205520 - 0.389409i$ and using Eq. (i), we form

$$\left(A_3 - \overline{\mu}_3 I\right)(A_3 - \mu_3 I) = \begin{bmatrix} -6.245835 & 44.983550 & -16.447185 & 124.147732 \\ -0.174713 & -6.615951 & 6.705533 & -26.177482 \\ -0.066078 & 0.346330 & 0.052815 & -0.569937 \\ 0 & 0.077321 & 0 & 0 \end{bmatrix} \tag{j}$$

Then, using the QR decomposition defined by Eq. (6.227), we obtain

$$Q = \begin{bmatrix} 0.999553 & -0.028128 & 0.008647 & -0.005260 \\ 0.027960 & 0.999425 & 0.008489 & -0.017175 \\ 0.010575 & 0.016162 & -0.839774 & 0.542592 \\ 0 & -0.009822 & -0.542800 & -0.839804 \end{bmatrix} \tag{k}$$

and

$$R = \begin{bmatrix} -6.248627 & 44.782125 & -16.251788 & 123.354294 \\ 0 & -7.872595 & 7.165155 & -29.663640 \\ 0 & 0 & -0.129648 & 1.329901 \\ 0 & 0 & 0 & -0.512653 \end{bmatrix} \tag{l}$$

and we note that R, although not needed for future computations, was listed for completeness. Inserting Eqs. (i) and (k) into Eq. (6.228), we obtain

$$A_5 = Q^T A_3 Q = \begin{bmatrix} -1.562554 & -50.711921 & 61.768974 & 192.048645 \\ 0.166495 & 1.106605 & -2.279971 & -6.427138 \\ 0 & -0.008234 & 1.145103 & 2.725010 \\ 0 & 0 & -0.611457 & -1.189155 \end{bmatrix} \tag{m}$$

and we observe that convergence has begun in earnest, as the (3,2) element in A_5 is two orders of magnitude smaller than the (3,2) element in A_3.

At this point, the procedure is clear, so that we merely list the results of the iterations lending to convergence, as follows:

$$\begin{gathered}
\begin{matrix}\mu_5 \\ \bar{\mu}_5\end{matrix} = -0.022026 \pm 0.551395 \\
A_7 = \begin{bmatrix} -2.148580 & -50.602406 & -94.698131 & 178.043230 \\ 0.204670 & 1.703406 & 4.146521 & -8.221022 \\ 0 & -10.659614 \times 10^{-8} & -1.110470 & 2.796046 \\ 0 & 0 & -0.526066 & 1.055644 \end{bmatrix}
\end{gathered} \tag{n}$$

$$\begin{gathered}
\begin{matrix}\mu_7 \\ \bar{\mu}_7\end{matrix} = -0.027413 \pm 0.545795i \\
A_9 = \begin{bmatrix} -2.871503 & -50.536982 & -7.648625 & 201.362541 \\ 0.270420 & 2.426711 & 0.259732 & -12.100907 \\ 0 & -3.1 \times 10^{-13} & 0.191138 & 3.214531 \\ 0 & 0 & -0.107528 & -0.245946 \end{bmatrix}
\end{gathered} \tag{o}$$

$$\begin{gathered}
\begin{matrix}\mu_9 \\ \bar{\mu}_9\end{matrix} = -0.027404 \pm 0.545795i \\
A_{11} = \begin{bmatrix} -3.874946 & -50.410919 & -7.641901 & 201.080236 \\ 0.396483 & 3.429754 & 0.412657 & -16.125867 \\ 0 & 0 & 0.191138 & 3.214531 \\ 0 & 0 & -0.107528 & -0.245946 \end{bmatrix}
\end{gathered} \tag{p}$$

$$\begin{gathered}
\begin{matrix}\mu_{11} \\ \bar{\mu}_{11}\end{matrix} = -0.027404 \pm 0.545795i \\
A_{13} = \begin{bmatrix} -5.450540 & -50.129583 & -7.625002 & 200.468774 \\ 0.677820 & 5.005348 & 0.654432 & -22.485010 \\ 0 & 0 & 0.191138 & 3.214531 \\ 0 & 0 & -0.107528 & -0.245946 \end{bmatrix}
\end{gathered} \tag{q}$$

At this point, the shifts and the lower right corner 2 × 2 matrix have reached constant values, so that convergence to the first pair of complex conjugate eigenvalues has been achieved. Of course, the eigenvalues are equal to the shifts, or

$$\begin{matrix}\lambda_1 \\ \bar{\lambda}_1\end{matrix} = -0.027404 \pm 0.545795i \tag{r}$$

We note that these are the same values as those obtained in Example 6.15 by the power method, except that there they were labeled λ_3 and $\bar{\lambda}_3$, because they represent the subdominant pair.

In general, the iteration process continues with the $(n-2) \times (n-2)$ upper left corner matrix obtained by deleting the last two rows and columns from A_{13}. In this particular case, the upper left corner matrix is 2 × 2, so that no further iterations are needed. Hence, the eigenvalues of this 2 × 2 matrix are simply the other pair of eigenvalues, or

$$\begin{matrix}\lambda_3 \\ \bar{\lambda}_3\end{matrix} = -0.222596 \pm 2.578257i \tag{s}$$

which are precisely the values for the other pair of eigenvalues obtained in Example 6.15, where they were labeled λ_1 and $\bar{\lambda}_1$.

6.16 INVERSE ITERATION FOR COMPLEX EIGENSOLUTIONS

In Sec. 6.15, we have shown how to compute the eigenvalues of a real nonsymmetric matrix A by the QR method, with special emphasis being placed on the case in which some, or all the eigenvalues are complex. The problem of computing the eigenvectors of A remains. As demonstrated in Sec. 6.10, inverse iteration is able to produce the eigenvector $\mathbf{x}_r$ belonging to a known eigenvalue λ_r with extreme efficiency. Indeed, convergence to $\mathbf{x}_r$ is remarkably fast, quite often in two or three iteration steps. But, the algorithm described in Sec. 6.10 is predicated upon the eigensolutions being real, so that the question arises as to what happens when the eigensolutions are complex. From our experience, iterations with complex quantities are to be avoided, and inverse iteration is no exception. In this section, we develop an inverse iteration algorithm capable of producing eigenvectors belonging to given complex eigenvalues working with real quantities alone.

We consider a real nonsymmetric $2n \times 2n$ matrix A and introduce the notation

$$\lambda = \alpha + i\beta, \qquad \mathbf{x} = \mathbf{u} + i\mathbf{v} \tag{6.230}$$

Inserting Eqs. (6.230) into Eq. (6.183), we can write

$$A(\mathbf{u} + i\mathbf{v}) = (\alpha + i\beta)(\mathbf{u} + i\mathbf{v}) \tag{6.231}$$

so that, separating the real and imaginary parts, we obtain

$$\begin{aligned} A\mathbf{u} &= \alpha\mathbf{u} - \beta\mathbf{v} \\ A\mathbf{v} &= \alpha\mathbf{v} + \beta\mathbf{u} \end{aligned} \tag{6.232}$$

Equations (6.232) can be written in the compact form

$$B\mathbf{w} = \mathbf{0} \tag{6.233}$$

where $\mathbf{w} = \left[\mathbf{u}^T \ \mathbf{v}^T\right]^T$ is a real expanded $4n$-vector and

$$B = \begin{bmatrix} A - \alpha I & \beta I \\ -\beta I & A - \alpha I \end{bmatrix} \tag{6.234}$$

is a real expanded $4n \times 4n$ coefficient matrix.

Equation (6.233) forms the basis for the inverse iteration algorithm for complex eigenvalues using real quantities. By analogy with the ordinary inverse iteration described in Sec. 6.10, we express the iteration process in the form

$$B\mathbf{w}^{(p)*} = \mathbf{w}^{(p-1)}, \qquad \mathbf{w}^{(p)} = c^{(p)}\mathbf{w}^{(p)*}, \qquad p = 1, 2, \ldots \tag{6.235a, b}$$

where $c^{(p)}$ is a scaling factor. For convenience, we choose

$$c^{(p)} = \frac{1}{\max_i \left| w_i^{(p)*} \right|}, \qquad p = 1, 2, \dots \tag{6.236}$$

in which $\max_i \left| w_i^{(p)*} \right|$ denotes the component of $\mathbf{w}^{(p)*}$ of largest magnitude. The effect of the adopted scaling is to render the component of $\mathbf{w}^{(p)}$ of largest magnitude equal to 1 or to -1, and is designed to prevent the iterates from becoming too large. In this regard, it should be pointed out that the matrix B is close to being singular, so that $\mathbf{w}^{(p)*}$ is likely to be several orders of magnitude larger than $\mathbf{w}^{(p-1)}$, which is also the reason for the fast convergence. The process begins with an arbitrary initial choice $\mathbf{w}(0)$ and solving Eq. (6.235a) for $\mathbf{w}^{(1)*}$ by means of Gaussian elimination with back-substitution. Then, $\mathbf{w}^{(1)*}$ is normalized to $\mathbf{w}^{(1)}$ according to Eqs. (6.235b) and (6.236) and the process is repeated. If λ is close to an eigenvalue, say $\lambda = \lambda_r = \alpha_r + i\beta_r$, then

$$\lim_{p\to\infty} \left(\mathbf{u}^{(p)} + i\mathbf{v}^{(p)} \right) = \mathbf{x}_r \tag{6.237}$$

where $\mathbf{u}^{(p)}$ and $\mathbf{v}^{(p)}$ are the upper half and lower half of the iterate $\mathbf{w}^{(p)}$, respectively. In practice, only a few iteration steps are necessary.

Equation (6.237) only indicates in a qualitative manner that the process converges, but provides no clues as to when convergence occurs. Hence, we must develop a quantitative convergence criterion. To this end, we recall from Sec. 6.13 that two complex vectors represent the same eigenvector when the magnitude ratios of homologous components of the two vectors are the same and the phase angle difference between any pair of components of one vector is the same as the phase angle difference between homologous components of the other vector. To quantify this statement, we introduce the notation

$$\mathbf{x}^{(p)} = \mathbf{u}^{(p)} + i\mathbf{v}^{(p)} = \begin{bmatrix} \left|x_1^{(p)}\right| \underline{/\psi_1^{(p)}} \\ \left|x_2^{(p)}\right| \underline{/\psi_2^{(p)}} \\ \cdots\cdots\cdots\cdots \\ \left|x_{2n}^{(p)}\right| \underline{/\psi_{2n}^{(p)}} \end{bmatrix}, \qquad p = 1, 2, \dots \tag{6.238}$$

where $\left|x_i^{(p)}\right|$ denotes the magnitude of the ith component of $\mathbf{x}^{(p)}$ and $\underline{/\psi_i^{(p)}}$ denotes the corresponding phase angle ($i = 1, 2, \dots, 2n$). Then, the convergence criterion can be stated in the form

$$\left.\begin{aligned} \lim_{p\to\infty} \frac{\left|x_i^{(p)}\right|}{\left|x_{2n}^{(p)}\right|} &= r_i = \text{constant}, \\ \lim_{p\to\infty} \left(\psi_i^{(p)} - \psi_{2n}^{(p)} \right) &= \Delta\psi_i^{(p)} = \text{constant}, \end{aligned}\right\} \quad i = 1, 2, \dots, 2n-1 \tag{6.239}$$

The procedure described above yields the right eigenvector $\mathbf{x}_r$ belonging to the known eigenvalue λ_r. The same procedure, but with A replaced by A^T, can be used to obtain the left eigenvector $\mathbf{y}_r$.

Example 6.18

Consider the damped two-degree-of-freedom system of Example 6.15, $n = 2$, and compute the eigenvector $\mathbf{x}_1$ belonging to the dominant eigenvalue λ_1 by means of inverse iteration.

From Example 6.15, the coefficient matrix is

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -5 & 4 & -0.4 & 0.2 \\ 2 & -2 & 0.1 & -0.1 \end{bmatrix} \tag{a}$$

Also from Example 6.15, the dominant eigenvalue is

$$\lambda_1 = \alpha_1 + i\beta_1 = -0.222596 + 2.578257i \tag{b}$$

Hence, inserting Eqs. (a) and (b) into Eq. (6.234), we can write the expanded matrix

$$B = \begin{bmatrix} 0.222596 & 0 & 1 & 0 & 2.578257 & 0 & 0 & 0 \\ 0 & 0.222596 & 0 & 1 & 0 & 2.578257 & 0 & 0 \\ -5 & 4 & -0.177404 & 0.2 & 0 & 0 & 2.578257 & 0 \\ 2 & -2 & 0.1 & 0.122596 & 0 & 0 & 0 & 2.578257 \\ -2.578257 & 0 & 0 & 0 & 0.222596 & 0 & 1 & 0 \\ 0 & -2.578257 & 0 & 0 & 0 & 0.222596 & 0 & 1 \\ 0 & 0 & -2.578257 & 0 & -5 & 4 & -0.177404 & 0.2 \\ 0 & 0 & 0 & -2.578257 & 2 & -2 & 0.1 & 0.122596 \end{bmatrix} \tag{c}$$

We begin the iteration with the vector

$$\mathbf{w}^{(0)} = [1\ 1\ 1\ 1\ 1\ 1\ 1\ 1]^T \tag{d}$$

Introducing Eqs. (c) and (d) into Eq. (6.235a) with $p = 1$, using Gaussian elimination with back-substitution and normalizing according to Eqs. (6.235b) and (6.236), we obtain the normalized iterate

$$\mathbf{w}^{(1)} = [-0.230142\ 0.086898\ -1\ 0.440708\ 0.407729\ -0.178433\ -0.684124\ 0.263765]^T \tag{e}$$

so that, using Eq. (6.238) with $p = 1$, the corresponding complex vector is

$$\mathbf{x}^{(1)} = \begin{bmatrix} -0.230142 + 0.407729i \\ 0.086898 - 0.178433i \\ -1.000000 - 0.684124i \\ 0.440708 + 0.263765i \end{bmatrix} = \begin{bmatrix} 0.468197\angle 119.4425^\circ \\ 0.198468\angle 295.9664^\circ \\ 1.211621\angle 214.3770^\circ \\ 0.513606\angle 30.9009^\circ \end{bmatrix} \tag{f}$$

which permits us to compute the magnitude ratios and phase angle differences

$$\begin{aligned} |x_1^{(1)}|/|x_4^{(1)}| &= 0.911588, & \Delta\psi_1^{(1)} &= \psi_1^{(1)} - \psi_4^{(1)} = 88.5416^\circ \\ |x_2^{(1)}|/|x_4^{(1)}| &= 0.386421, & \Delta\psi_2^{(1)} &= \psi_2^{(1)} - \psi_4^{(1)} = 265.0655^\circ \\ |x_3^{(1)}|/|x_4^{(1)}| &= 2.359048, & \Delta\psi_3^{(1)} &= \psi_3^{(1)} - \psi_4^{(1)} = 183.4761^\circ \end{aligned} \tag{g}$$

Following the same pattern, we obtain for $p = 2$

$$\mathbf{x}^{(2)} = \begin{bmatrix} -0.413050 - 0.291785i \\ 0.182270 + 0.112843i \\ 0.844239 - 1.000000i \\ -0.331512 + 0.444819i \end{bmatrix} = \begin{bmatrix} 0.505716 \underline{/\ 215.2380^\circ} \\ 0.214373 \underline{/\ \ 31.7616^\circ} \\ 1.308717 \underline{/\ 310.1724^\circ} \\ 0.554765 \underline{/\ 126.6962^\circ} \end{bmatrix} \tag{h}$$

and

$$\begin{aligned} \left|x_1^{(2)}\right| / \left|x_4^{(2)}\right| &= 0.911586, & \Delta\psi_1^{(2)} &= \psi_1^{(2)} - \psi_4^{(2)} = 88.5418^\circ \\ \left|x_2^{(2)}\right| / \left|x_4^{(2)}\right| &= 0.386421, & \Delta\psi_2^{(2)} &= \psi_2^{(2)} - \psi_4^{(2)} = 265.0654^\circ \\ \left|x_3^{(2)}\right| / \left|x_4^{(2)}\right| &= 2.359048, & \Delta\psi_3^{(2)} &= \psi_3^{(2)} - \psi_4^{(2)} = 183.4762^\circ \end{aligned} \tag{i}$$

Comparing Eqs. (g) and (i), we conclude that convergence has been virtually achieved, so that we accept $\mathbf{x}_1 = \mathbf{x}^{(2)}$ as the eigenvector belonging to λ_1.

6.17 SYNOPSIS

In this chapter, we presented a large variety of iterative algorithms for the real symmetric eigenvalue problem. The power method has the advantage of simplicity, but convergence can be slow if the eigenvalues are not well spaced. It should be used only when a small number of dominant eigensolutions are desired. The Jacobi method has a certain air of elegance and is easy to understand. It is not particularly fast, however, and should be used only for moderate size problems. The two most attractive algorithms require that the matrix be in tridiagonal form. As tridiagonalization procedures, we single out Givens' method, Householder's method and Lanczos' method. The two algorithms for solving eigenvalue problems for matrices in symmetric tridiagonal form are Givens' method and the QR method, and both can produce only eigenvalues. Givens' method is based on the separation theorem and has the ability to target individual eigenvalues in a given range, but convergence is only linear. By contrast, the QR method has better than cubic convergence. However, before this remarkable convergence can be achieved, it is necessary to carry out shifts using the proper strategy. As far as the computation of eigenvectors belonging to known eigenvalues is concerned, inverse iteration has no peers. Another method with superior convergence characteristics is Rayleigh's quotient iteration. It also targets individual eigensolutions. Before it can be used, however, one must have a good guess of the eigenvector targeted, which for all practical purposes confines the usefulness of the method to the lowest vibration mode. Finally, simultaneous iteration, which can be regarded as an extension of the power method, iterates to several eigensolutions at a time. For a well-populated real symmetric matrix A, the most indicated approach to the full solution of the eigenvalue problem is to tridiagonalize A by means of Householder's method, use the QR method with shifts to compute the eigenvalues and inverse iteration to compute the eigenvectors.

In the case of nonsymmetric matrices, the choice of algorithms for solving the eigenvalue problem is significantly more limited than is the case of symmetric matrices. Here too, the power method has the advantage of simplicity, but should be considered only if a small number of dominant eigensolutions is required. The method of choice is once again the QR method, provided the matrix A is first reduced to Hessenberg form and the shifting strategy of Sec. 6.15, which obviates the problem of working with complex numbers, is used. Then, the eigenvectors are to be obtained by a version of inverse iteration capable of producing complex eigenvectors working with real iterates.

PROBLEMS

6.1 Solve the set of algebraic equations

$$\begin{aligned} 6x_1 + 5x_2 - x_3 + 3x_4 &= 2 \\ -3x_1 + x_2 + 3x_3 - 2x_4 &= -5.5 \\ 2x_1 - 2x_2 + x_3 - 6x_4 &= -7.5 \\ 4x_1 + x_2 - 2x_3 + 5x_4 &= 11.5 \end{aligned}$$

by Gaussian elimination with back-substitution.

6.2 Solve the set of algebraic equations

$$\begin{aligned} 1.2x_1 + 4.7x_2 + x_3 - 6\ x_4 &= 4.1 \\ 5.2x_1 + x_2 - 3x_3 + 2\ x_4 &= 16.1 \\ x_1 - x_2 + 4x_3 - 2\ x_4 &= -10.35 \\ 1.3x_1 + 2.2x_2 - x_3 - 5.5x_4 &= 3.625 \end{aligned}$$

by Gaussian elimination with back-substitution.

6.3 Solve Problem 6.1 by the Gauss-Jordan method.

6.4 Solve Problem 6.2 by the Gauss-Jordan method.

6.5 Verify that the real symmetric matrix

$$A = \begin{bmatrix} 1.44 & -2.76 & 0 & 0 \\ & 25.54 & 12.15 & 0 \\ \text{Symm} & & 20.25 & -2.88 \\ & & & 2.89 \end{bmatrix}$$

is positive definite. Then, carry out the Cholesky decomposition of A.

6.6 Verify that the real symmetric matrix

$$A = \begin{bmatrix} 12.25 & 4.2 & -7.525 & 2.87 \\ & 23.53 & 4.752 & -10.296 \\ \text{Symm} & & 35.1461 & -14.252 \\ & & & 23.5949 \end{bmatrix}$$

is positive definite. Then, carry out the Cholesky decomposition of A.

6.7 The mass and stiffness matrices of a three-degree-of-freedom system are

$$M = m\begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad K = k\begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Compute the natural frequencies and modes of vibration by means of the power method. Use the formulation given by Eqs. (6.72) and (6.73)

6.8 The mass and stiffness matrices of a four-degree-of-freedom system are

$$M = m\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad K = k\begin{bmatrix} 4 & -2 & 0 & 0 \\ -2 & 3 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Compute the natural frequencies and modes of vibration by means of the power method. Use the formulation given by Eqs. (6.72) and (6.73).

6.9 The mass and stiffness matrices of a four-degree-of-freedom system are

$$M = m\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}, \qquad K = k\begin{bmatrix} 3 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 3 & -2 \\ 0 & 0 & -2 & 5 \end{bmatrix}$$

Compute the natural frequencies and modes of vibration by means of the power method. Use the formulation given by Eqs. (6.72) and (6.73).

6.10 Solve Problem 6.7 by means of the threshold serial Jacobi method using a threshold value of 10^{-6}.

6.11 Solve Problem 6.8 by means of the threshold serial Jacobi method using a threshold value of 10^{-6}.

6.12 Solve Problem 6.9 by means of the threshold serial Jacobi method using a threshold value of 10^{-6}.

6.13 Use Givens' method to tridiagonalize the matrix $A = M^{1/2}K^{-1}M^{1/2}$, where M and K are as in Problem 6.8.

6.14 Use Givens' method to tridiagonalize the matrix $A = M^{1/2}K^{-1}M^{1/2}$, where M and K are as in Problem 6.9.

6.15 Tridiagonalize the matrix A of Problem 6.13 by means of Householder's method.

6.16 Tridiagonalize the matrix A of Problem 6.14 by means of Householder's method.

6.17 Tridiagonalize the matrix A of Problem 6.13 by means of Lanczos' method.

6.18 Tridiagonalize the matrix A of Problem 6.14 by means of Lanczos' method.

6.19 Use Givens' method (Sec. 6.8) to compute the eigenvalues of the tridiagonal matrix $M^{-1/2}KM^{-1/2}$, where M and K are as in Problem 6.7.

6.20 Solve Problem 6.19 with M and K as in Problem 6.8.

6.21 Solve Problem 6.19 with M and K as in Problem 6.9.

6.22 Compute the eigenvalues of the tridiagonal matrix obtained in Problem 6.13 (or Problem 6.15) by means of Givens' method (Sec. 6.8). Compare results with those obtained in Problem 6.20 and draw conclusions.

6.23 Compute the eigenvalues of the tridiagonal matrix obtained in Problem 6.14 (or Problem 6.16) by means of Givens' method (Sec. 6.8). Compare results with those obtained in Problem 6.9 and draw conclusions.

6.24 Solve Problem 6.19 by the QR method.

6.25 Solve Problem 6.20 by the QR method.

6.26 Solve Problem 6.21 by the QR method.

6.27 Solve Problem 6.22 by the QR method.

6.28 Solve Problem 6.23 by the QR method.

6.29 Use inverse iteration to compute the eigenvectors belonging to the eigenvalues obtained in Problem 6.24. Then, determine the actual modal vectors.

6.30 Use inverse iteration to compute the eigenvectors belonging to the eigenvalues obtained in Problem 6.27. Then, determine the actual modal vectors.

6.31 Use inverse iteration to compute the eigenvectors belonging to the eigenvalues obtained in Problem 6.28. Then, determine the actual modal vectors.

6.32 Compute the two lowest modes of vibration for the system of Problem 6.7 by means of Rayleigh's quotient iteration.

6.33 Compute the two lowest modes of vibration for the system of Problem 6.8 by means of Rayleigh's quotient iteration.

6.34 Compute the two lowest modes of vibration for the system of Problem 6.9 by means of Rayleigh's quotient iteration.

6.35 Solve Problem 6.32 by means of simultaneous iteration.

6.36 Solve Problem 6.33 by means of simultaneous iteration.

6.37 Solve Problem 6.34 by means of simultaneous iteration.

6.38 Solve Problem 6.32 by means of subspace iteration.

6.39 Solve Problem 6.33 by means of subspace iteration.

6.40 Solve Problem 6.34 by means of subspace iteration.

6.41 A damped three-degree-of-freedom system is defined by the mass, damping and stiffness matrices

$$M = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 0.4 & -0.2 & 0 \\ -0.2 & 0.3 & -0.1 \\ 0 & -0.1 & 0.1 \end{bmatrix}, \qquad K = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$$

Solve the eigenvalue problem by the power method (Sec. 6.13).

6.42 A damped three-degree-of-freedom system is defined by the mass, damping and stiffness matrices

$$M = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 0.4 & -0.2 & 0 \\ -0.2 & 0.3 & -0.1 \\ 0 & -0.1 & 2.1 \end{bmatrix}, \qquad K = \begin{bmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Solve the eigenvalue problem by the power method (Sec. 6.13).

6.43 Reduce the matrix A from Problem 6.41 to upper Hessenberg form.

6.44 Reduce the matrix A from Problem 6.42 to upper Hessenberg form.

6.45 Compute the eigenvalues of the upper Hessenberg matrix from Problem 6.43 by the QR method.

6.46 Compute the eigenvalues of the upper Hessenberg matrix from Problem 6.44 by the QR method.

6.47 Use inverse iteration (Sec. 6.16) to compute the right and left eigenvectors belonging to the eigenvalues obtained in Problem 6.45. Work with the matrix A from Problem 6.41.

6.48 Use inverse iteration (Sec. 6.16) to compute the right and left eigenvectors belonging to the eigenvalues obtained in Problem 6.46. Work with the matrix A from Problem 6.42.

BIBLIOGRAPHY

1. Clint, M. and Jennings, A., "The Evaluation of Eigenvalues and Eigenvectors of Real Symmetric Matrices by Simultaneous Iteration," *The Computer Journal*, Vol. 13, No. 1, 1970, pp. 76–80.
2. Francis, J.G.F., "The QR Transformation, Parts I and II," *The Computer Journal*, Vol. 4, 1961, pp. 265–271, 332–345.
3. Franklin, J. N., *Matrix Theory*, Prentice Hall, Englewood Cliffs, NJ, 1968.
4. Givens, J. W., "Numerical Computation of the Characteristic Values of a Real Symmetric Matrix," *Oak Ridge National Laboratory, Report ORNL-1574*, 1954.
5. Golub, G. H. and Van Loan, C. F., *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, 1989.
6. Hotelling, H., "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, Vol. 24, 1933, pp. 417–441, 498–520.
7. Jennings, A., "A Direct Iteration Method of Obtaining Latent Roots and Vectors of a Symmetric Matrix," *Proceedings of the Cambridge Philosophical Society*, Vol. 63, 1967, pp. 755–765.
8. Kublanovskaya, V. N., On Some Algorithms for the Solution of the Complete Eigenvalue Problem," *Z. Vycisl. Mat. i Mat. Fiz.*, Vol. 1, 1961, pp. 555–570; *USSR Comput. Math. Math. Phys.*, Vol. 3, pp. 637–657.
9. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan, New York, 1967.
10. Meirovitch, L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.
11. Meirovitch, L., *Elements of Vibration Analysis*, 2nd ed., McGraw-Hill, New York, 1986.
12. Stewart, G. W., *Introduction to Matrix Computations*, Academic Press, New York, 1973.
13. Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, Oxford University Press, London, 1988.

7

DISTRIBUTED-PARAMETER SYSTEMS

Mathematical models of vibrating systems are commonly divided into two broad classes, discrete, or lumped-parameter models, and continuous, or distributed-parameter models. In real life, systems can contain both lumped and distributed parameters. Until now, our study has been confined to discrete systems. In this chapter, our attention turns to systems with parameters distributed throughout the domain and in some cases with lumped parameters at boundaries. The emphasis is on theoretical developments and exact solutions. Because exact solutions are possible in only a limited number of cases, quite often the interest lies in approximate solutions. Such solutions are generally obtained through spatial discretization, which amounts to approximating distributed-parameter systems by discrete ones. Discretization methods are presented in Chapters 8 and 9. In this regard, it should be pointed out that systems containing both distributed parameters and lumped parameters dispersed throughout the system can be treated by some of the techniques developed in Chapters 8 and 9. The theory presented in this chapter not only provides a great deal of insight into the behavior of vibrating systems but the theory is essential to the approximate techniques developed in Chapters 8 and 9.

Discrete systems consist of aggregates of discrete components, such as masses and springs, with the masses assumed to be rigid and the springs assumed to be flexible but massless. The masses and the spring stiffnesses represent the system parameters, with the masses being concentrated at given points and connected by the springs, which explains why the parameters are referred to as lumped. The spatial position of each mass is identified by an index, and in general the number of masses coincides with the number of degrees of freedom of the system. In contrast, at each point of a continuous system there is both mass and stiffness, and these parameters are distributed over the entire system. The position of a point in a distributed-parameter system is identified by one, two, or three spatial coordinates, with the set of interior points defining a domain D and the set of points on the exterior of D defining the boundary S. Because there is an infinity of points in D, a distributed system can be regarded as having an infinite number of degrees of freedom.

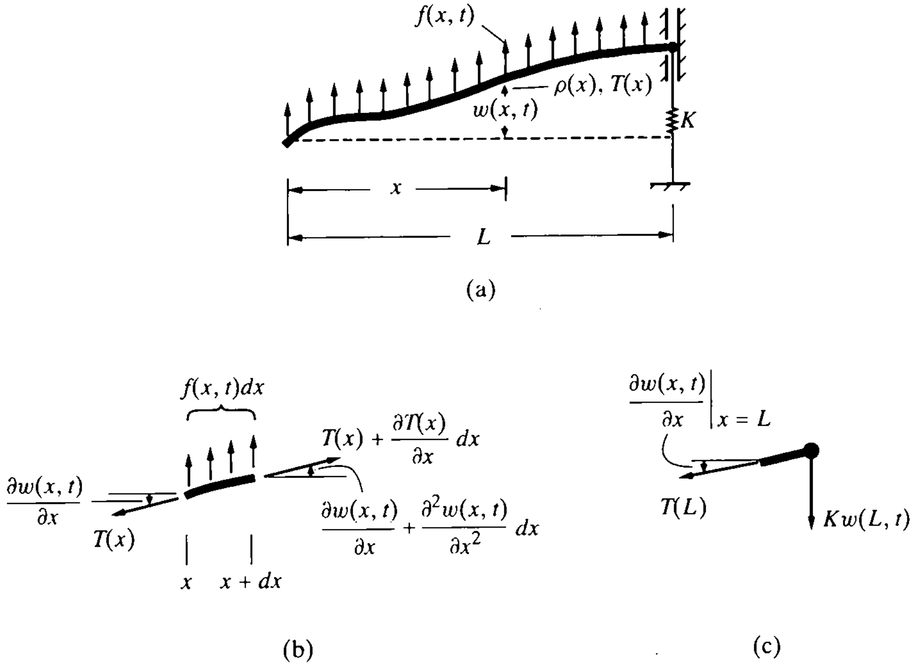
As can be expected, the mathematical formalism for distributed systems differs significantly from the formalism for discrete systems. For n-degree-of-freedom discrete systems, the motion is governed by n simultaneous ordinary differential equations. To solve these equations, it is necessary to solve the associated algebraic eigenvalue problem. The solution consists of n eigenvalues and eigenvectors, where the eigenvectors possess the orthogonality property. The orthogonal eigenvectors form a basis for an n-dimensional vector space, which can be used in conjunction with an expansion theorem to decouple the equations of motion into n independent second-order equations. The independent equations resemble the equation of motion for a single-degree-of-freedom system and can be solved with relative ease. In contrast, the motion of distributed-parameter systems is governed by boundary-value problems consisting of one, or several partial differential equations to be satisfied over D and an appropriate number of boundary conditions to be satisfied at every point of S. The solution of a boundary-value problem requires the solution of an associated differential eigenvalue problem, where the latter solution consists of an infinite set of eigenvalues and eigenfunctions. The eigenfunctions are orthogonal and can be used as a basis for an infinite-dimensional function space in conjunction with an expansion theorem to transform the boundary-value problem into an infinite set of independent second-order ordinary differential equations resembling entirely the independent equations for discrete systems, so that they can be solved with the same ease. Hence, whereas the mathematical formalism and the methods of solution for distributed systems are different from those for discrete systems, many concepts are entirely analogous.

In this chapter, we begin with the derivation of boundary-value problems for a variety of elastic members, such as strings in transverse vibration, rods in axial vibration, shafts in torsion and beams in bending. Subsequently, a generic boundary-value problem consisting of a Lagrange partial differential equation and suitable boundary conditions is derived. The free vibration problem leads naturally to the differential eigenvalue problem, which can be cast conveniently in differential operator form. Here, the concept of operator self-adjointness, which represents the counterpart of matrix symmetry in discrete systems, proves quite powerful in generalizing the developments to a large class of systems, covering most systems of interest to our study. In fact, the various elastic members mentioned in the beginning of this paragraph represent mere special cases of this general theory. In extending our discussion to two-dimensional systems, and in particular to membranes and plates, we encounter new concepts such as the shape of the boundary and degeneracy. Subjects such as the variational formulation and the integral formulation of the eigenvalue problem have certain implications in approximate methods of solution. Other topics of interest are distributed gyroscopic systems and distributed damped systems. The chapter concludes with an extensive discussion of systems with nonhomogeneous boundary conditions.

7.1 THE BOUNDARY-VALUE PROBLEM FOR STRINGS, RODS AND SHAFTS

Our interest lies in the vibration of systems with distributed parameters. The motion of such systems depends not only on time but also on the spatial position, which is defined by one, two, or three coordinates. Consistent with this, the domain of extension of the distributed-parameter system is one-, two-, or three-dimensional, respectively. Distributed-parameter systems are governed by boundary-value problems, which consist of differential equations of motion to be satisfied over all interior points of the domain and boundary conditions to be satisfied at points bounding the domain. Because there are at least two independent variables, the equations of motion are partial differential equations. Unlike discrete systems, for which the equations of motion tend to have the same form, the boundary-value problem tends to differ from one type of distributed system to another. Using operator notation, the various boundary-value problems for large classes of systems can be reduced to the same form, which enables us to draw general conclusions concerning all systems in a given class. In this section, we derive the boundary-value problem for a particular distributed system, and later in this chapter we generalize the formulation to large classes of systems.

We consider the string in transverse vibration shown in Fig. 7.1a. This is a one-dimensional distributed-parameter system with the domain of extension D : $0 < x < L$ and the boundary $S : x = 0, L$, where x is the spatial variable and L the length of the string. We denote the displacement in the transverse direction



$f(x, t)$
$\rho(x), T(x)$
$w(x, t)$
K
x
L
(a)
$f(x, t)dx$
$T(x) + \frac{\partial T(x)}{\partial x}dx$
$\frac{\partial w(x, t)}{\partial x}$
$T(x)$
$\frac{\partial w(x, t)}{\partial x} + \frac{\partial^2 w(x, t)}{\partial x^2}dx$
x
$x + dx$
(b)
$\frac{\partial w(x, t)}{\partial x}\Big|_{x = L}$
$T(L)$
$Kw(L, t)$
(c)

Figure 7.1 **(a)** String in transverse vibration **(b)** Free-body diagram for a differential element of string **(c)** Force diagram at $x = L$

of a typical point x of the string by the dependent variable $w(x, t)$, the force per unit length by $f(x, t)$, the mass per unit length of string by $\rho(x)$ and the tension by $T(x)$. The left end of the string is fixed and the right end is attached to a spring of stiffness K. For simplicity, we assume that the displacement is measured from the equilibrium position $w_{\text{eq}}(x)$, in which case a pretension in the string balances the force due to gravity. In all future discussions, we omit gravitational forces on the basis of this assumption. We propose to derive the boundary-value problem in two ways, first by means of Newton's second law and then by means of the extended Hamilton's principle.

To derive the boundary-value problem by means of Newton's second law, we refer to the free-body diagram of Fig. 7.1b and assume that the displacement $w(x, t)$ is sufficiently small that the sine and the tangent of the angle made by the string with respect to the x-axis is approximately equal to the slope of the displacement curve. Hence, summing up forces in the transverse direction, we obtain

$$\left[T(x) + \frac{\partial T(x)}{\partial x}dx\right]\left[\frac{\partial w(x,t)}{\partial x} + \frac{\partial^2 w(x,t)}{\partial x^2}dx\right] + f(x,t)\,dx - T(x)\frac{\partial w(x,t)}{\partial x}$$

$$= \rho(x)\,dx\frac{\partial^2 w(x,t)}{\partial t^2}, \qquad 0 < x < L \tag{7.1}$$

which, upon ignoring second-order terms in dx and dividing through by dx, can be reduced to

$$\frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right] + f(x,t) = \rho(x)\frac{\partial^2 w(x,t)}{\partial t^2}, \qquad 0 < x < L \tag{7.2}$$

Because the string is fixed at $x = 0$, the displacement must satisfy

$$w(x,t) = 0, \qquad x = 0 \tag{7.3a}$$

On the other hand, from Fig. 7.1c, we conclude that the transverse force balance at the right end of the string requires that

$$T(x)\frac{\partial w(x,t)}{\partial x} + Kw(x,t) = 0, \qquad x = L \tag{7.3b}$$

Equations (7.2) and (7.3) represent the *boundary-value problem*, in which Eq. (7.2) is the *partial differential equation of motion* and Eqs. (7.3) are the *boundary conditions.*

Next, we wish to derive the boundary-value problem by means of the extended Hamilton's principle, which can be expressed in the form

$$\int_{t_1}^{t_2} \left(\delta T - \delta V + \overline{\delta W}_{nc}\right) dt = 0, \qquad \delta w(x,t) = 0, \qquad t = t_1, t_2 \tag{7.4}$$

where

$$T(t) = \frac{1}{2}\int_0^L \rho(x)\left[\frac{\partial w(x,t)}{\partial t}\right]^2 dx \tag{7.5}$$

is the kinetic energy. The potential energy arises from the restoring forces due to the tension in the string and the elongation of the spring at $x = L$. To determine the potential energy due to the tension in the string, we denote the length of the differential element dx in displaced position by ds. Then, the potential energy is simply the sum of the work that must be performed by the tensile force to restore the string to the original horizontal position and the potential energy due to the end spring, or

$$V(t) = \int_0^L T(x)(ds - dx) + \frac{1}{2}Kw^2(L,t) \tag{7.6}$$

But, recognizing from Fig. 7.2 that $\partial w/\partial x << 1$, we can write

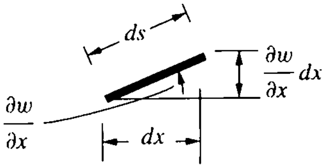
$$ds = \left[(dx)^2 + \left(\frac{\partial w}{\partial x}dx\right)^2\right]^{1/2} = \left[1 + \left(\frac{\partial w}{\partial x}\right)^2\right]^{1/2} dx \cong \left[1 + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2\right] dx \tag{7.7}$$

where we retained two terms only in the binomial expansion. Hence, inserting Eq. (7.7) into Eq. (7.6), we obtain the potential energy of the system

$$V(t) = \frac{1}{2}\int_0^L T(x)\left[\frac{\partial w(x,t)}{\partial x}\right]^2 dx + \frac{1}{2}Kw^2(L,t) \tag{7.8}$$

Moreover, the virtual work due to the nonconservative distributed force is simply

$$\overline{\delta W}_{nc}(t) = \int_0^L f(x,t)\,\delta w(x,t)\,dx \tag{7.9}$$



ds
$\frac{\partial w}{\partial x}dx$
$\frac{\partial w}{\partial x}$
dx

Figure 7.2 Differential element of string in displaced position

From Eq. (7.5), the variation in the kinetic energy is

$$\delta T(t) = \int_0^L \rho(x)\frac{\partial w(x,t)}{\partial t}\delta\left[\frac{\partial w(x,t))}{\partial t}\right] dx \tag{7.10}$$

Similarly, from Eq. (7.8), the variation in the potential energy has the form

$$\delta V(t) = \int_0^L T(x)\frac{\partial w(x,t)}{\partial x}\delta\left[\frac{\partial w(x,t)}{\partial x}\right] dx + Kw(L,t)\,\delta w(L,t) \tag{7.11}$$

Equation (7.10) contains the variation in the velocity, and we must transform the equation into one in terms of the virtual displacement. To this end, we assume that variations and differentiations with respect to time are interchangeable and carry out the following integration by parts with respect to t:

$$\begin{aligned}\int_{t_1}^{t_2} \delta T\, dt &= \int_{t_1}^{t_2}\int_0^L \rho(x)\frac{\partial w(x,t)}{\partial t}\delta\left[\frac{\partial w(x,t)}{\partial t}\right]dx\, dt \\ &= \int_0^L \left[\int_{t_1}^{t_2} \rho(x)\frac{\partial w(x,t)}{\partial t}\frac{\partial \delta w(x,t)}{\partial t}dt\right]dx \\ &= \int_0^L \left[\rho(x)\frac{\partial w(x,t)}{\partial t}\delta w(x,t)\bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2}\rho(x)\frac{\partial^2 w(x,t)}{\partial t^2}\delta w(x,t)\, dt\right]dx \\ &= -\int_{t_1}^{t_2}\int_0^L \rho(x)\frac{\partial^2 w(x,t)}{\partial t^2}\delta w(x,t)\, dx\, dt \end{aligned} \tag{7.12}$$

in which we used the end conditions on time in Eq. (7.4). Similarly, we perform an integration by parts with respect to x and rewrite Eq. (7.11) in the form

$$\begin{aligned}\delta V(t) &= \int_0^L T(x)\frac{\partial w(x,t)}{\partial x}\delta\left[\frac{\partial w(x,t)}{\partial x}\right]dx + Kw(L,t)\,\delta w(L,t) \\ &= \int_0^L T(x)\frac{\partial w(x,t)}{\partial x}\frac{\partial \delta w(x,t)}{\partial x}dx + Kw(L,t)\,\delta w(L,t) \\ &= T(x)\frac{\partial w(x,t)}{\partial x}\delta w(x,t)\bigg|_0^L \\ &\quad - \int_0^L \frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right]\delta w(x,t)\, dx + Kw(L,t)\,\delta w(L,t) \\ &= -\int_0^L \frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right]\delta w(x,t)\, dx \\ &\quad + \left[T(x)\frac{\partial w(x,t)}{\partial x} + Kw(x,t)\right]\delta w(x,t)\bigg|_{x=L} \\ &\quad - T(x)\frac{\partial w(x,t)}{\partial x}\delta w(x,t)\bigg|_{x=0} \end{aligned} \tag{7.13}$$

Hence, inserting Eqs. (7.9), (7.12) and (7.13) into Eq. (7.4), we obtain

$$\begin{aligned}\int_{t_1}^{t_2}\Bigg\langle -\int_0^L \left\{\rho(x)\frac{\partial^2 w(x,t)}{\partial t^2} - \frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right] - f(x,t)\right\}\delta w(x,t)\, dx \\ - \left[T(x)\frac{\partial w(x,t)}{\partial x} + Kw(x,t)\right]\delta w(x,t)\bigg|_{x=L}\end{aligned}$$

$$+ T(x)\frac{\partial w(x,t)}{\partial x}\delta w(x,t)\bigg|_{x=0}\Bigg\rangle dt = 0 \tag{7.14}$$

But, the virtual displacements are arbitrary, and hence they can be assigned values at will. We assume that either δw or $T\partial w/\partial x$ is zero at $x = 0$, that either δw or $(T\partial w/\partial x) + Kw$ is zero at $x = L$ and that δw is completely arbitrary at every point of the domain $0 < x < L$. Under these circumstances, we conclude that Eq. (7.14) can be satisfied if and only if the displacement $w(x,t)$ satisfies

$$\frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right] + f(x,t) = \rho(x)\frac{\partial^2 w(x,t)}{\partial t^2}, \qquad 0 < x < L \tag{7.15}$$

and is such that

$$T(x)\frac{\partial w(x,t)}{\partial x}\delta w(x,t) = 0 \qquad \text{at } x = 0 \tag{7.16a}$$

$$\left[T(x)\frac{\partial w(x,t)}{\partial x} + Kw(x,t)\right]\delta w(x,t) = 0 \qquad \text{at } x = L \tag{7.16b}$$

Conditions (7.16) can be satisfied in two ways each, where we note that the coefficient of $\delta w(x,t)$ in both Eq. (7.16a) and Eq. (7.16b) represents the transverse component of force. At $x = 0$ the transverse component of force cannot be zero for all t, so that $\delta w(x,t)\big|_{x=0}$ must be zero. Hence, Eq. (7.16a) is satisfied by

$$w(x,t) = 0, \qquad x = 0 \tag{7.17a}$$

On the other hand, $\delta w(x,t) \neq 0$ at $x = L$, so that $\delta w(x,t)\big|_{x=L}$ is arbitrary. It follows that Eq. (7.16b) can be satisfied if and only if

$$T(x)\frac{\partial w(x,t)}{\partial w} + Kw(x,t) = 0, \qquad x = L \tag{7.17b}$$

Equation (7.15) represents the partial differential equation and Eqs. (7.17) are the boundary conditions. Equations (7.15) and (7.17) constitute the boundary-value problem for the transverse vibration of the string shown in Fig. 7.1a, and they coincide with Eqs. (7.2) and (7.3) derived earlier by Newton's second law, respectively.

At this point, we wish to examine the question of boundary conditions more closely. Boundary condition (7.3a), or (7.17a), is geometric in nature and it indicates that the solution $w(x,t)$ of the differential equation, Eq. (7.2), or Eq. (7.15), must be zero at $x = 0$. Boundary conditions resulting from pure geometric compatibility are called *geometric*, *essential*, or *imposed boundary conditions*. On the other hand, boundary condition (7.3b), or (7.17b), states that the vertical component of the tension in the string must balance the spring force at $x = L$. Boundary conditions resulting from force balance are known as *natural*, *dynamic*, or *additional boundary conditions*. It is the satisfaction of the boundary conditions that renders the solution of the differential equation unique. It should be noted that geometric boundary conditions tend to be simpler than natural ones.

Boundary-value problems are classified according to order, which is determined by the order of the highest derivative with respect to the spatial coordinate in the differential equation of motion. The boundary-value problem defined by Eqs. (7.15) and (7.17) is of order two, and it represents one of the simplest boundary-value problems in vibrations. Because the order is two, there are two boundary conditions, one at each end. In the case of the string of Fig. 7.1a, one boundary condition is geometric and the other is natural. In general, there are other possibilities. For example, in the case of a string fixed at both ends, both boundary conditions are geometric. In the case of a free end, the slope to the displacement curve must be zero, provided the tension is not zero. Although this appears as a geometric boundary condition, it is really a natural one, as the transverse component of force, which is equal to $T(x)\partial w(x,t)/\partial x$, must be zero.

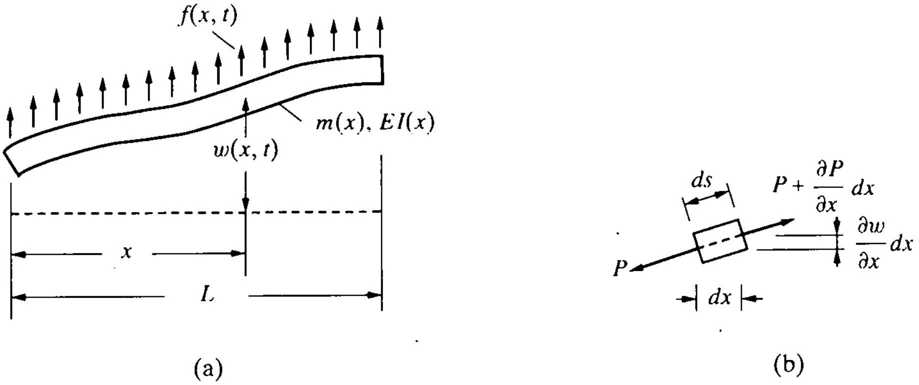
Other elastic members, such as shafts in torsion and rods in axial vibration, are also defined by boundary-value problems of order two. In fact, they are governed by boundary-value problems mathematically equivalent to that for strings in transverse vibration. The only difference lies in the nature of the displacement, excitation and parameters. Indeed, the structure of the differential equation is exactly the same, but the transverse displacement $w(x,t)$ must be replaced by the angular displacement $\theta(x,t)$ in the case of a shaft and by the axial displacement $u(x,t)$ in the case of a rod, the transverse force density $f(x,t)$ must be replaced by the distributed torque $m(x,t)$ and by the distributed axial force $f(x,t)$, the tension $T(x)$ must be replaced by the torsional stiffness $GJ(x)$, in which G is the shear modulus and $J(x)$ the polar area moment of inertia, and by the axial stiffness $EA(x)$, where E is the modulus of elasticity and $A(x)$ the cross-sectional area, and the mass density $\rho(x)$ must be replaced by the polar mass moment of inertia density $I(x)$ and by the mass per unit length $m(x)$, respectively.

Before turning our attention to other types of systems, we should point out that the reason for introducing second-order boundary-value problems via strings in transverse vibration, instead of shafts in torsional vibration or rods in axial vibration, is simply that transverse displacements are easier to visualize than angular or axial displacements. This point is made abundantly clear by the fact that angular displacements and axial displacement are commonly plotted as if they were transverse displacements, which at times can lead to confusion.

7.2 THE BOUNDARY-VALUE PROBLEM FOR BEAMS IN BENDING

An elastic member used most frequently in structures is the beam. Figure 7.3a shows a beam in bending vibration under the distributed transverse force $f(x,t)$. In addition, we assume that the beam is subjected to the axial force $P(x)$, as shown in Fig. 7.3b. We propose to derive the boundary-value problem by means of the extended Hamilton's principle, Eq. (7.4). To this end, we assume that the kinetic energy is due entirely to translation. Hence, denoting the transverse displacement by $w(x,t)$, we can write the kinetic energy expression in the form

$$T(t) = \frac{1}{2}\int_0^L m(x)\left[\frac{\partial w(x,t)}{\partial t}\right]^2 dx \tag{7.18}$$



f(x, t)
m(x), EI(x)
w(x, t)
x
L
(a)
ds
P
$P + \frac{\partial P}{\partial x}dx$
$\frac{\partial w}{\partial x}dx$
dx
(b)

Figure 7.3 **(a)** Beam in bending vibration **(b)** Axial forces acting on a beam differential element

where $m(x)$ is the mass per unit length of beam. The potential energy consists of two parts, one due to bending and one due to the axial force, where the latter is entirely analogous to the potential energy in a string, the integral in Eq. (7.8). The potential energy of a beam in bending can be found in any textbook on mechanics of materials. Moreover, using the analogy with the string, the combined potential energy can be shown to have the form

$$V(t) = \frac{1}{2}\int_0^L \left\{ EI(x)\left[\frac{\partial^2 w(x,t)}{\partial x^2}\right]^2 + P(x)\left[\frac{\partial w(x,t)}{\partial x}\right]^2 \right\} dx \tag{7.19}$$

where $EI(x)$ is the *bending stiffness*, or *flexural rigidity*, in which E is the Young's modulus of elasticity and $I(x)$ is the area moment of inertia about an axis normal to the plane defined by x and w. It should be noted here that the contribution of the gravitational forces to the potential energy can be ignored by measuring displacements from the equilibrium position. Finally, the virtual work due to nonconservative forces is simply

$$\overline{\delta W}_{nc} = \int_0^L f(x,t)\,\delta w(x,t)\,dx \tag{7.20}$$

We carry out the operations involved in the extended Hamilton's principle, Eq. (7.4), term by term. To this end, we assume that variations and differentiations are interchangeable, use Eq. (7.18) and write the variation in the kinetic energy in the form

$$\delta T = \int_0^L m\frac{\partial w}{\partial t}\delta\frac{\partial w}{\partial t}dx = \int_0^L m\frac{\partial w}{\partial t}\frac{\partial}{\partial t}\delta w\,dx \tag{7.21}$$

so that, integrating by parts, the first term in Eq. (7.4) yields

$$\int_{t_1}^{t_2}\delta T\,dt = \int_{t_1}^{t_2}\int_0^L m\frac{\partial w}{\partial t}\frac{\partial}{\partial t}\delta w\,dx\,dt = \int_0^L\int_{t_1}^{t_2} m\frac{\partial w}{\partial t}\frac{\partial}{\partial t}\delta w\,dt\,dx$$
$$= \int_0^L m\frac{\partial w}{\partial t}\delta w\Big|_{t_1}^{t_2} dx - \int_0^L\int_{t_1}^{t_2} m\frac{\partial^2 w}{\partial t^2}\delta w\,dt\,dx$$

$$= -\int_{t_1}^{t_2}\int_0^L m\frac{\partial^2 w}{\partial t^2}\delta w\,dx\,dt \tag{7.22}$$

where we considered the fact that $\delta w = 0$ at $t = t_1, t_2$. Moreover, we take the variation in the potential energy, Eq. (7.19), integrate by parts and obtain

$$\begin{aligned}
\delta V &= \int_0^L \left(EI\frac{\partial^2 w}{\partial x^2}\delta\frac{\partial^2 w}{\partial x^2} + P\frac{\partial w}{\partial x}\delta\frac{\partial w}{\partial x}\right)dx \\
&= \int_0^L \left(EI\frac{\partial^2 w}{\partial x^2}\frac{\partial^2}{\partial x^2}\delta w + P\frac{\partial w}{\partial x}\frac{\partial}{\partial x}\delta w\right)dx \\
&= EI\frac{\partial^2 w}{\partial x^2}\frac{\partial}{\partial x}\delta w\bigg|_0^L - \frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right)\delta w\bigg|_0^L + P\frac{\partial w}{\partial x}\delta w\bigg|_0^L \\
&\quad + \int_0^L \left[\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) - \frac{\partial}{\partial x}\left(P\frac{\partial w}{\partial x}\right)\right]\delta w\,dx
\end{aligned} \tag{7.23}$$

Finally, inserting Eqs. (7.20), (7.22) and (7.23) into Eq. (7.4) and grouping the terms in appropriate fashion, we obtain

$$\begin{aligned}
-\int_{t_1}^{t_2}\Bigg\{&\left[-\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right) + P\frac{\partial w}{\partial x}\right]\delta w\bigg|_0^L + EI\frac{\partial^2 w}{\partial x^2}\delta\frac{\partial w}{\partial x}\bigg|_0^L \\
&+ \int_0^L \left[m\frac{\partial^2 w}{\partial t^2} + \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) - \frac{\partial}{\partial x}\left(P\frac{\partial w}{\partial x}\right) - f\right]\delta w\,dx\Bigg\}dt = 0
\end{aligned} \tag{7.24}$$

At this point, we invoke the arbitrariness of the virtual displacements in a judicious manner. In particular, we assume that either δw or its coefficient in the boundary term is zero at $x = 0$ and $x = L$, that either $\delta(\partial w/\partial x)$ or its coefficient in the boundary term is zero at $x = 0$ and $x = L$ and that δw is entirely arbitrary over the domain $0 < x < L$. It follows that Eq. (7.24) can be satisfied if and only if

$$m\frac{\partial^2 w}{\partial t^2} + \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) - \frac{\partial}{\partial x}\left(P\frac{\partial w}{dx}\right) - f = 0, \qquad 0 < x < L \tag{7.25}$$

and, in addition, either

$$-\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right) + P\frac{\partial w}{\partial x} = 0 \qquad \text{at } x = 0, L \tag{7.26a}$$

or

$$w = 0 \qquad \text{at } x = 0, L \tag{7.26b}$$

and either

$$EI\frac{\partial^2 w}{\partial x^2} = 0 \qquad \text{at } x = 0, L \tag{7.27a}$$

or

$$\frac{\partial w}{\partial x} = 0 \qquad \text{at } x = 0, L \qquad (7.27b)$$

Equation (7.25) represents the equation of motion, a fourth-order partial differential equation to be satisfied at every point of the domain, and Eqs. (7.26) and (7.27) represent boundary conditions. We note that two boundary conditions must be satisfied at $x = 0$ and $x = L$, one from Eqs. (7.26) and one from Eqs. (7.27). The choice as to which of the two equations must be satisfied depends on the nature of the boundary. For example, we know that the displacement is zero at a pinned end, but the slope of the displacement curve must be different from zero. Hence, at a *pinned end* we must retain Eqs. (7.26b) and (7.27a) as the boundary conditions, or

$$w = 0; \qquad M = EI\frac{\partial^2 w}{\partial x^2} = 0 \qquad (7.28)$$

where M is identified as the bending moment. At a *free end*, we know that the displacement and slope must be different from zero, so that from Eqs. (7.26a) and (7.27a) the boundary conditions are

$$Q = -\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right) + P\frac{\partial w}{\partial x} = 0, \qquad M = EI\frac{\partial^2 w}{\partial x^2} = 0 \qquad (7.29)$$

where Q is recognized as a shearing force. Finally, at a *clamped end* the bending moment and shearing force are not zero, so that from Eqs. (7.26b) and (7.27b) we obtain the boundary conditions

$$w = 0, \qquad \partial w/\partial x = 0 \qquad (7.30)$$

Other boundary conditions than those described by Eqs. (7.28)–(7.30) are possible, but they require modifications in the formulation. For example, if an end is supported by a linear spring of stiffness K, then the term $\frac{1}{2}Kw^2$ must be added to the potential energy, and if an end is supported by a torsional spring of stiffness K_T, the term $\frac{1}{2}K_T(\partial w/\partial x)^2$ must be added. Clearly, the effect of such terms will be reflected in the boundary conditions, as shown in the next section.

The model of a beam in bending vibration considered in this section is the simplest possible and is known as an *Euler-Bernoulli beam.*

7.3 LAGRANGE'S EQUATION FOR DISTRIBUTED SYSTEMS. THE BOUNDARY-VALUE PROBLEM

In deriving boundary-value problems by means of the extended Hamilton's principle there are several steps that must be repeated time and again. In this regard, we recall from Chapter 2 that a similar situation exists for discrete systems, in which case it is possible to avoid the repetition by using the extended Hamilton's principle to derive Lagrange's equations and then derive the equations of motion by means of Lagrange's equations. Hence, it is only natural to seek a similar approach for distributed systems. Of course, the situation is more complicated in the case of distributed systems, because there are at least two independent variables instead of one. In view of this, the resulting formulation consists of a single Lagrange's equation, a partial differential equation, and associated boundary conditions. This formulation is quite general and covers a relatively large class of systems.

For convenience, we rewrite the extended Hamilton's principle, Eq. (7.4), as

$$\int_{t_1}^{t_2} \left(\delta L + \overline{\delta W}_{nc}\right) dt = 0, \qquad \delta w = 0, \quad t = t_1, t_2 \tag{7.31}$$

where $L = T - V$ is the Lagrangian. Moreover, we can simplify the notation by denoting derivatives wi th respect to the spatial variable x by primes and derivatives with respect to the time t by overdots whenever appropriate. This permits us to write the kinetic energy in the general functional form

$$T = \int_0^L \hat{T}(\dot{w})\, dx \tag{7.32}$$

in which the overcaret denotes a kinetic energy density. Similarly, the potential energy is assumed to have the form

$$V = V_0\left[w(0,t), w'(0,t)\right] + V_L\left[w(L,t), w'(L,t)\right] + \int_0^L \hat{V}(w, w', w'')\, dx \tag{7.33}$$

where the subscripts 0 and L refer to potential energy due to springs at the ends $x = 0$ and $x = L$, respectively, and the overcaret denotes a potential energy density. Hence, the Lagrangian can be expressed as

$$L = L_0\left[w(0,t),\ w'(0,t)\right] + L_L\left[w(L,t),\ w'(L,t)\right] + \int_0^L \hat{L}(w,\ w',\ w'',\ \dot{w})\, dx \tag{7.34}$$

in which L_0 and L_L are boundary Lagrangians and $\hat{L}$ is a Lagrangian density. Moreover, the virtual work is simply

$$\overline{\delta W}_{nc} = \int_0^L f\, \delta w\, dx \tag{7.35}$$

where $f = f(x,t)$ is the distributed force, and note that concentrated forces can be expressed as distributed by means of spatial Dirac delta functions.

The extended Hamilton's principle, Eq. (7.31), calls for the variation in the Lagrangian, which can be expressed in the form

$$\delta L = \delta L_0 + \delta L_L + \int_0^L \delta\hat{L}\, dx \tag{7.36}$$

where

$$\delta L_0 = \frac{\partial L_0}{\partial w(0,t)}\delta w(0,t) + \frac{\partial L_0}{\partial w'(0,t)}\delta w'(0,t) \tag{7.37a}$$

$$\delta L_L = \frac{\partial L_L}{\partial w(L,t)}\delta w(L,t) + \frac{\partial L_L}{\partial w'(L,t)}\delta w'(L,t) \tag{7.37b}$$

$$\delta\hat{L} = \frac{\partial \hat{L}}{\partial w}\delta w + \frac{\partial \hat{L}}{\partial w'}\delta w' + \frac{\partial \hat{L}}{\partial w''}\delta w'' + \frac{\partial \hat{L}}{\partial \dot{w}}\delta \dot{w} \tag{7.37c}$$

From Secs. 7.1 and 7.2, we recall that, before we can invoke the arbitrariness of the virtual displacements, it is necessary to carry out a number of integrations by parts with respect to x and t. This can be conveniently done term by term. First, we carry out integrations with respect to t, or

$$\int_{t_1}^{t_2} \frac{\partial \hat{L}}{\partial \dot{w}}\delta \dot{w} dt = \frac{\partial \hat{L}}{\partial \dot{w}}\delta w\bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right)\delta w\, dt = -\int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right)\delta w\, dt \tag{7.38}$$

where we considered the fact that δw is zero at $t = t_1, t_2$. Next, we carry out integrations with respect to x, which involve terms in $\hat{L}$. First, we have

$$\int_0^L \frac{\partial \hat{L}}{\partial w'}\delta w' dx = \int_0^L \frac{\partial \hat{L}}{\partial w'}\frac{\partial}{\partial x}\delta w\, dx = \frac{\partial \hat{L}}{\partial w'}\delta w\bigg|_0^L - \int_0^L \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'}\right)\delta w\, dx \tag{7.39a}$$

Similarly,

$$\int_0^L \frac{\partial \hat{L}}{\partial w''}\delta w''\, dx = \frac{\partial \hat{L}}{\partial w''}\delta w'\bigg|_0^L - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\delta w\bigg|_0^L + \int_0^L \frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''}\right)\delta w\, dx \tag{7.39b}$$

Introducing Eqs. (7.35)–(7.37) into Eq. (7.31), considering Eqs. (7.38) and (7.39) and collecting terms involving $\delta w(x, t)$, $\delta w(0, t)$, $\delta w(L, t)$, $\delta w'(0, t)$ and $\delta w'(L, t)$, we obtain

$$\begin{aligned}
\int_{t_1}^{t_2} \Bigg\{ & \frac{\partial L_0}{\partial w(0, t)}\delta w(0, t) + \frac{\partial L_0}{\partial w'(0, t)}\delta w'(0, t) \\
& + \frac{\partial L_L}{\partial w(L, t)}\delta w(L, t) + \frac{\partial L_L}{\partial w'(L, t)}\delta w'(L, t) \\
& + \int_0^L \left[\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'}\right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right) + f\right]\delta w\, dx \\
& + \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\delta w\bigg|_0^L + \frac{\partial \hat{L}}{\partial w''}dw'\bigg|_0^L \Bigg\} dt \\
= \int_{t_1}^{t_2} \Bigg\langle & \int_0^L \left[\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'}\right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right) + f\right]\delta w\, dx \\
& + \left\{\frac{\partial L_0}{\partial w(0, t)} - \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\bigg|_{x=0}\right\}\delta w(0, t)
\end{aligned}$$

$$+\left\{\frac{\partial L_L}{\partial w(L,t)}+\left[\frac{\partial \hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\bigg|_{x=L}\right\}\delta w(L,t)$$

$$+\left[\frac{\partial L_0}{\partial w'(0,t)}-\frac{\partial \hat{L}}{\partial w''}\bigg|_{x=0}\right]\delta w'(0,t)$$

$$+\left[\frac{\partial L_L}{\partial w'(L,t)}+\frac{\partial \hat{L}}{\partial w''}\bigg|_{x=L}\right]\delta w'(L,t)\Bigg\rangle dt = 0 \tag{7.40}$$

At this point, we invoke the arbitrariness of the virtual displacements. If we let $\delta w(0,t) = \delta w(L,t) = 0$ and $\delta w'(0,t) = \delta w'(L,t) = 0$, we conclude that Eq. (7.40) can be satisfied for all values of δw in the open domain $0 < x < L$ if and only if the coefficient of δw is zero, or

$$\frac{\partial \hat{L}}{\partial w}-\frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'}\right)+\frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''}\right)-\frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right)+f = 0, \qquad 0 < x < L \tag{7.41}$$

Moreover, by writing

$$\left\{\frac{\partial L_0}{\partial w(0,t)}-\left[\frac{\partial \hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\bigg|_{x=0}\right\}\delta w(0,t) = 0 \tag{7.42a}$$

$$\left[\frac{\partial L_0}{\partial w'(0,t)}-\frac{\partial \hat{L}}{\partial w''}\bigg|_{x=0}\right]\delta w'(0,t) = 0 \tag{7.42b}$$

$$\left\{\frac{\partial L_L}{\partial w(L,t)}+\left[\frac{\partial \hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\bigg|_{x=L}\right\}\delta w(L,t) = 0 \tag{7.43a}$$

$$\left[\frac{\partial L_L}{\partial w'(L,t)}+\frac{\partial \hat{L}}{\partial w''}\bigg|_{x=L}\right]\delta w'(L,t) = 0 \tag{7.43b}$$

we take into account that either $\delta w(0,t)$ or its coefficient is zero and either $\delta w'(0,t)$ or its coefficient is zero, and similar statements can be made about conditions at $x = L$.

Equation (7.41) represents the *Lagrange differential equation of motion* for the fourth-order distributed-parameter system with the Lagrangian given by Eq. (7.34). It is a partial differential equation to be satisfied at every point of the open domain $0 < x < L$. Moreover, from Eqs. (7.42), we conclude that *at* $x = 0$ the displacement must be such that either

$$\frac{\partial L_0}{\partial w(0,t)}-\left[\frac{\partial \hat{L}}{\partial w'}-\frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\bigg|_{x=0} = 0 \tag{7.44a}$$

or

$$w(0,t) = 0 \tag{7.44b}$$

and either

$$\frac{\partial L_0}{\partial w'(0,t)} - \left.\frac{\partial \hat{L}}{\partial w''}\right|_{x=0} = 0 \tag{7.45a}$$

or

$$w'(0,t) = 0 \tag{7.45b}$$

In addition, at $x = L$ either

$$\frac{\partial L_L}{\partial w(L,t)} + \left.\left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right)\right]\right|_{x=L} = 0 \tag{7.46a}$$

or

$$w(L,t) = 0 \tag{7.46b}$$

and either

$$\frac{\partial L_L}{\partial w'(L,t)} + \left.\frac{\partial \hat{L}}{\partial w''}\right|_{x=L} = 0 \tag{7.47a}$$

or

$$w'(L,t) = 0 \tag{7.47b}$$

For a given system, the solution $w(x,t)$ of the Lagrange equation, Eq. (7.41), must satisfy one of each of Eqs. (7.44)–(7.47), for a total of two equations at each end. The four equations represent the *boundary conditions*, and the choice of boundary conditions at each end is dictated by the physics of the problem.

It should be pointed out here that, although the formulation is for fourth-order systems, the formulation can be used also for second-order systems by simply deleting terms that do not apply.

Example 7.1

Derive the Lagrange equation and boundary conditions for the string of Fig. 7.1a.

This is only a second-order system, so that deleting from Eq. (7.41) terms involving derivatives higher than two the Lagrange equation reduces to

$$\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}}\right) + f = 0, \qquad 0 < x < L \tag{a}$$

Moreover, the physics of the problem dictates the boundary conditions

$$w(0,t) = 0 \tag{b}$$

$$\frac{\partial L_L}{\partial w(L,t)} + \left.\frac{\partial \hat{L}}{\partial w'}\right|_{x=L} = 0 \tag{c}$$

From Eqs. (7.5) and (7.8), we can write the Lagrangian density

$$\hat{L} = \hat{T} - \hat{V} = \frac{1}{2}\rho\dot{w}^2 - \frac{1}{2}T(w')^2 \tag{d}$$

and the boundary Lagrangians

$$L_0 = 0, \qquad L_L = -\frac{1}{2}Kw^2(L,t) \tag{e}$$

Inserting Eq. (d) into Eq. (a), we obtain the partial differential equation of motion

$$\frac{\partial}{\partial x}\left(T\frac{\partial w}{\partial x}\right) - \rho\frac{\partial^2 w}{\partial t^2} + f = 0, \qquad 0 < x < L \tag{f}$$

where we recognized that $\rho = \rho(x)$ does not depend on t. Moreover, introducing Eq. (d) and the second of Eqs. (e) into Eq. (c), we obtain the boundary condition at $x = L$

$$Kw(L,t) + T\frac{\partial w}{\partial x}\bigg|_{x=L} = 0 \tag{g}$$

The boundary condition at $x = 0$ remains in the form of Eq. (b).

It is easy to verify that the boundary-value problem given by Eqs. (f), (b) and (g) is identical to that obtained in Sec. 7.1

Example 7.2

Derive the Lagrange equation and boundary conditions for the rotating cantilever beam shown in Fig. 7.4a.

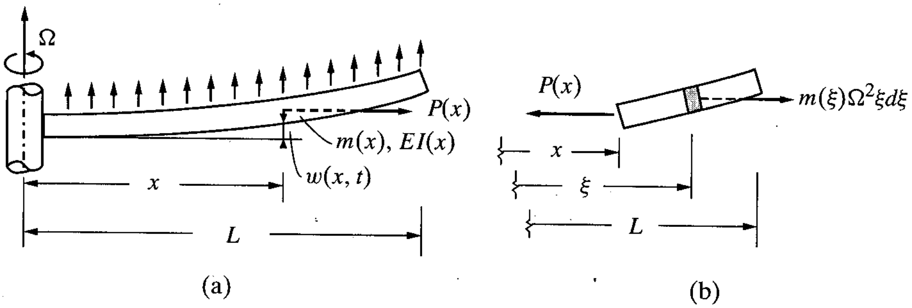


Figure 7.4 **(a)** Rotating cantilever beam **(b)** Axial force due to centrifugal effects

From Eqs. (7.18) and (7.19), the Lagrangian density has the expression

$$\hat{L} = \hat{T} - \hat{V} = \frac{1}{2}m\dot{w}^2 - \frac{1}{2}EI\left(w''\right)^2 - \frac{1}{2}P\left(w'\right)^2 \tag{a}$$

where, from Fig. 7.4b, the axial force has the form

$$P(x) = \int_x^L m\Omega^2\xi d\xi \tag{b}$$

Note that in the case at hand $L_0 = L_L = 0$.

The Lagrange differential equation for a fourth-order system is given in general form by Eq. (7.41). To obtain its explicit expression, we write

$$\frac{\partial\hat{L}}{\partial w} = 0, \qquad \frac{\partial\hat{L}}{\partial w'} = -Pw', \qquad \frac{\partial\hat{L}}{\partial w''} = -EIw'', \qquad \frac{\partial\hat{L}}{\partial\dot{w}} = m\dot{w} \tag{c}$$

Inserting Eqs. (c) into Eq. (7.41), we obtain the Lagrange equation in the explicit form

$$\frac{\partial}{\partial x}\left(P\frac{\partial w}{\partial x}\right) - \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) - m\frac{\partial^2 w}{\partial t^2} + f = 0, \qquad 0 < x < L \tag{d}$$

where P is given by Eq. (b).

To derive the boundary conditions, we observe that the displacement and rotation are zero at a clamped end. Hence, on physical grounds, we choose Eqs. (7.44b) and (7.45b) as the boundary conditions, so that

$$w = 0, \qquad w' = 0, \qquad x = 0 \tag{e}$$

On the other hand, the displacement and rotation at $x = L$ are not zero, so that we must choose Eqs. (7.46a) and (7.47a) as the boundary conditions. Observing from Eq. (b) that $P(L) = 0$, the boundary conditions at $x = L$ are

$$\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right) = 0, \qquad EI\frac{\partial^2 w}{\partial x^2} = 0, \qquad x = L \tag{f}$$

7.4 FREE VIBRATION OF CONSERVATIVE SYSTEMS. THE DIFFERENTIAL EIGENVALUE PROBLEM

In the absence of external forces, $f(x, t) = 0$, the boundary-value problem for the vibrating string of Fig. 7.1a, Eqs. (7.2) and (7.3), reduces to the partial differential equation of motion

$$\frac{\partial}{\partial x}\left[T(x)\frac{\partial w(x,t)}{\partial x}\right] = \rho(x)\frac{\partial^2 w(x,t)}{\partial t^2}, \qquad 0 < x < L \tag{7.48}$$

and the boundary conditions

$$w(x,t) = 0, \qquad x = 0 \tag{7.49a}$$

$$T(x)\frac{\partial w(x,t)}{\partial x} + Kw(x) = 0, \qquad x = L \tag{7.49b}$$

Equations (7.48) and (7.49) describe the free vibration of conservative second-order systems.

Our interest lies in the solution of Eqs. (7.48) and (7.49). In particular, we wish to explore the existence of solutions whereby the system executes *synchronous motions*, defined as motions in which every point performs the same motion in time. The physical implication is that synchronous motions are characterized by the fact that the ratio of the displacements corresponding to two different points of the string is constant. Mathematically, synchronous motions imply that the solution of Eqs. (7.48) and (7.49) is separable in the spatial variable and time, and hence it has the form

$$w(x,t) = W(x)F(t) \tag{7.50}$$

where $W(x)$ depends on the spatial position alone and $F(t)$ depends on time alone. Introducing Eq. (7.50) into Eqs. (7.48) and (7.49), we can write

$$\frac{d}{dx}\left[T(x)\frac{dW(x)}{dx}\right]F(t) = \rho(x)W(x)\frac{d^2F(t)}{dt^2}, \qquad 0 < x < L \tag{7.51}$$

$$W(0)F(t) = 0, \qquad \left[T(x)\frac{dW(x)}{dx} + KW(x)\right]\Bigg|_{x=L} F(t) = 0 \tag{7.52a, b}$$

Next, we divide Eq. (7.51) by $\rho W F$ and obtain

$$\frac{1}{\rho(x)W(x)}\frac{d}{dx}\left[T(x)\frac{dW(x)}{dx}\right] = \frac{1}{F(t)}\frac{d^2F(t)}{dt^2}, \qquad 0 < x < L \tag{7.53}$$

Observing that the left side of Eq. (7.53) depends on x alone and the right side on time alone and that F can be simply omitted from boundary conditions (7.52), we conclude that the solution is indeed separable in x and t. But, because the left side of Eq. (7.53) depends only on x and the right side only on t and, moreover, x and t are independent variables, the two sides of the equation must be equal to a constant, the same constant. In addition, the two sides are real, so that the constant must be real. For reasons that will become obvious shortly, we denote the constant by $-\lambda$, where λ is a positive constant, so that the right side of Eq. (7.53) yields

$$\ddot{F}(t) + \lambda F(t) = 0 \tag{7.54}$$

The solution of Eq. (7.54) has the exponential form

$$F(t) = Ae^{st} \tag{7.55}$$

Introducing Eq. (7.55) into Eq. (7.54) and dividing through by Ae^{st}, we obtain the *characteristic equation*

$$s^2 + \lambda = 0 \tag{7.56}$$

which has the roots

$$\begin{matrix} s_1 \\ s_2 \end{matrix} = \pm\sqrt{-\lambda} = \pm i\sqrt{\lambda} = \pm i\omega \tag{7.57}$$

It follows that the solution of Eq. (7.54) has the harmonic form

$$F(t) = A_1e^{i\omega t} + A_2e^{-i\omega t} = A_1e^{i\omega t} + \overline{A}_1e^{-i\omega t} = C\cos(\omega t - \phi) \tag{7.58}$$

where we let A_2 be equal to the complex conjugate $\overline{A}_1$ of A_1 in recognition of the fact that $F(t)$ must be real, and we note that the constants A_1 and $\overline{A}_1$ are related to the *amplitude* C and *phase angle* ϕ by

$$A_1 + \overline{A}_1 = C\cos\phi, \qquad i\left(A_1 - \overline{A}_1\right) = C\sin\phi \tag{7.59}$$

At this point our earlier choice of sign for the constant used in Eq. (7.53) is made clear by the fact that $F(t)$, as given by Eq. (7.58), represents harmonic oscillation, which is typical of the response of stable conservative systems. The choice will be justified mathematically in Sec. 7.5.

The question remains as to the displacement pattern, as well as to the frequency of oscillation. To answer these questions, we set the left side of Eq. (7.53) equal to $-\lambda$ and obtain the differential equation

$$-\frac{d}{dx}\left[T(x)\frac{dW(x)}{dx}\right] = \lambda\rho(x)W(x), \qquad \lambda = \omega^2, \qquad 0 < x < L \tag{7.60}$$

Moreover, we eliminate F from Eqs. (7.52) and obtain the boundary conditions

$$W(0) = 0; \; T(x)\frac{dW(x)}{dx} + KW(x) = 0, \qquad x = L \tag{7.61a, b}$$

Equations (7.60) and (7.61) represent the *differential eigenvalue problem* for the string shown in Fig. 7.1a. It can be described in words as *the problem of determining the constant* λ *so that the differential equation* (7.60) *admits nontrivial solutions satisfying boundary conditions* (7.61). Second-order problems of this type are known as *Sturm-Liouville problems*.

Next, we wish to derive the differential eigenvalue problem for the beam in bending discussed in Sec. 7.2. Letting $f(x,t) = 0$ in Eq. (7.25), assuming that the solution has the form given by Eq. (7.50) and following the approach used earlier in this section, we conclude once again that $F(t)$ is harmonic, as indicated by Eq. (7.58), and $W(x)$ satisfies the differential equation

$$\frac{d^2}{dx^2}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] - \frac{d}{dx}\left[P(x)\frac{dW(x)}{dx}\right] = \lambda m(x)W(x), \qquad \lambda = \omega^2, \qquad 0 < x < L \tag{7.62}$$

Moreover, from Eqs. (7.26) and (7.27), $W(x)$ must satisfy either

$$-\frac{d}{dx}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] + P(x)\frac{dW(x)}{dx} = 0 \qquad \text{at } x = 0, L \tag{7.63a}$$

or

$$W(x) = 0 \qquad \text{at } x = 0, L \tag{7.63b}$$

and either

$$EI(x)\frac{d^2W(x)}{dx^2} = 0 \qquad \text{at } x = 0, L \tag{7.64a}$$

or

$$\frac{dW(x)}{dx} = 0 \qquad \text{at } x = 0, L \tag{7.64b}$$

The differential eigenvalue problem consists of the differential equation, Eq. (7.62), to be satisfied over the domain $0 < x < L$ and two boundary conditions, one from Eqs. (7.63) and one from Eqs. (7.64), to be satisfied at $x = 0$ and $x = L$.

The differential eigenvalue problem is of vital importance in the vibration of distributed-parameter systems. Indeed, as in the case of discrete systems, its solution not only provides a great deal of information concerning the system characteristics but also can be used to derive the system response.

7.5 THE DIFFERENTIAL EIGENVALUE PROBLEM FOR SELF-ADJOINT SYSTEMS

In Secs. 7.3 and 7.4, we derived the differential eigenvalue problem for second-order and fourth-order systems. A cursory examination of the eigenvalue problem formulations permits us to conclude that the form differs from system to system, depending on the order of the system and the nature of the boundaries. In spite of the difference in appearance, various systems have many things in common. Hence, instead of discussing the individual systems separately, it is convenient to formulate the differential eigenvalue problem so as to apply to large classes of systems. This formulation can be best achieved by resorting to operator notation.

Let w be a function of one or two independent spatial variables x or x and y, so that in essence we confine ourselves to one- or two-dimensional problems, and consider the *homogeneous differential expression*

$$Lw = a_1 w + a_2 \frac{\partial w}{\partial x} + a_3 \frac{\partial w}{\partial y} + a_4 \frac{\partial^2 w}{\partial x^2} + a_5 \frac{\partial^2 w}{\partial x \partial y} + \dots \tag{7.65}$$

where the coefficients $a_1, a_2, \dots$ are known functions of the spatial variables x and y. We assume that Lw involves derivatives of w through order $2p$, where p is an integer, so that the differential expression is said to be of order $2p$. We can then define the *homogeneous differential operator* associated with Eq. (7.65) in the form

$$L = a_1 + a_2 \frac{\partial}{\partial x} + a_3 \frac{\partial}{\partial y} + a_4 \frac{\partial^2}{\partial x^2} + a_5 \frac{\partial^2}{\partial x \partial y} + \dots \tag{7.66}$$

and refer to L as being of order $2p$. If for the functions w_1 and w_2 the relation

$$L(c_1 w_1 + c_2 w_2) = c_1 L w_1 + c_2 L w_2 \tag{7.67}$$

holds, where c_1 and c_2 are constants, then the homogeneous differential operator L is said to be *linear*.

Next, we consider a generic eigenvalue problem and express the differential equation in the operator form

$$Lw = \lambda m w, \qquad x, y \text{ in } D \tag{7.68}$$

where L is a linear homogeneous differential operator of order $2p$, λ a parameter, m the mass density and D the domain of definition of Eq. (7.68). The operator L, referred to as *stiffness operator*, is of the type (7.66). Associated with the differential equation (7.68) there are p boundary conditions to be satisfied by the solution w at every point of the boundary S of the domain D. The boundary conditions are also expressed in operator form, as follows:

$$B_i w = 0, \qquad x, y \text{ in } S, \qquad i = 1, 2, \dots, p \tag{7.69}$$

where B_i are linear homogeneous differential operators of maximum order $2p - 1$. They are referred to as *boundary operators*. In the one-dimensional case, the domain D is a segment of the real line and the boundary S consists of the two points bounding D. In the two-dimensional case, the domain D is a plane and the boundary S consists of one or more closed curves bounding D. *The eigenvalue problem is defined as the problem of determining the values of the parameter λ for which there are nontrivial functions w satisfying the differential equation* (7.68) *and the boundary conditions* (7.69). Such parameters are called *eigenvalues* and the corresponding functions are called *eigenfunctions*. The eigenvalue problem defined by Eqs. (7.68) and (7.69) admits a *denumerably*, or *countably infinite set of eigenvalues* $\lambda_1, \lambda_2, \dots$ *and associated eigenfunctions* $w_1, w_2, \dots$. The implication is that *the eigenfunction w_r belongs to the eigenvalue λ_r* $(r = 1, 2, \dots)$.

One of the most important problems in the vibration of distributed-parameter systems involves the expansion of the response in terms of known functions, and in particular the expansion in terms of the system eigenfunctions. This expansion is based on some remarkable properties of the eigenfunctions. Before proceeding with the investigation of these properties, it will prove useful to introduce certain pertinent definitions concerning functions in general. In the first place, we assume that the functions considered are real and *piecewise smooth*, i.e., that they are piecewise continuous and possess piecewise continuous first derivatives in a given domain D. Then, for two such functions f and g, we define the *inner product of the two functions f and g over the domain D* as

$$(f, g) = \int_D fg\, dD \tag{7.70}$$

If the inner product vanishes, then the two functions f and g are said to be *orthogonal* over D. The square root of the inner product of a function f with itself is known as the *norm* of f, defined as

$$\|f\| = (f, f)^{1/2} = \left(\int_D f^2\, dD\right)^{1/2} \tag{7.71}$$

Orthogonal functions with unit norm are said to be *orthonormal*. The existence of the norm simply implies that $\|f\| < \infty$. A function whose norm exists is said to be *square summable* in D, which implies that f^2 is integrable in the Lebesgue sense (Ref. 1, p. 108). Functions f such that $\|f\| < \infty$ are said to have *finite energy*, and the space of such functions is denoted by $\mathcal{K}^0$, where the superscript indicates the order of the derivative required for finite energy, in this case zero.

A property of functions intimately related to orthogonality is *linear independence*. To define the concept, we consider a set of n functions $\phi_1, \phi_2, \dots, \phi_n$. Then, if a homogeneous linear relation with constant coefficients of the type

$$\sum_{i=1}^{n} c_i \phi_i = 0 \tag{7.72}$$

exists without all the coefficients $c_i (i = 1, 2, \dots, n)$ being identically zero, the set of functions $\phi_1, \phi_2, \dots, \phi_n$ is said to be *linearly dependent*. If Eq. (7.72) is satisfied only when all the coefficients c_i $(i = 1, 2, \dots, n)$ are identically zero, then the set of functions is said to be *linearly independent*. To explore the connection between orthogonality and linear independence, we assume that the set of functions $\phi_1, \phi_2, \dots, \phi_n$ is orthogonal, multiply Eq. (7.72) by ϕ_j, integrate over the domain D and obtain

$$\sum_{i=1}^{n} c_i \int_D \phi_i \phi_j\, dD = \sum_{i=1}^{n} c_i \|\phi_j\|^2 \delta_{ij} = c_j \|\phi_j\|^2 = 0, \qquad j = 1, 2, \dots, n \tag{7.73}$$

Because the norms $\|\phi_j\|$ cannot be zero, it follows from Eq. (7.73) that all the coefficients $c_j (j = 1, 2, \dots, n)$ must be zero. But, this is precisely the condition for the functions $\phi_1, \phi_2, \dots, \phi_n$ to be linearly independent, which proves that *orthogonal functions are linearly independent*. The converse is not true, as *independent functions are not necessarily orthogonal*. This is a matter of semantics, however, because *independent functions can be rendered orthogonal.* The procedure for rendering independent functions orthogonal is known as the *Gram-Schmidt orthogonalization process* (Ref. 1). It is common practice to normalize the functions during the process, so that the result is a set of orthonormal functions.

Next, we consider the problem of expanding any given function f in terms of known functions. To this end, we let $\phi_1, \phi_2, \ldots$ be an orthonormal system and express f as the *linear combination*

$$f = \sum_{r=1}^{\infty} c_r \phi_r \tag{7.74}$$

where the coefficients

$$c_r = (f, \phi_r) = \int_D f \phi_r \, dD, \qquad r = 1, 2, \ldots \tag{7.75}$$

are known as *components of f with respect to the orthonormal system $\phi_1, \phi_2, \ldots$.* This expansion is not unlike the expansion of a periodic function in terms of a Fourier series, and in fact it represents a generalization of the Fourier series expansion.

In vibrations there is considerable interest in approximating the given function f by means of *finite* series of the type

$$f = \sum_{r=1}^{n} d_r \phi_r \tag{7.76}$$

where ϕ_r are orthonormal functions, d_r are constant coefficients and n is fixed. The objective is to produce the "best" approximation of f, in the sense that the *mean square error*

$$M = \int_D \left(f - \sum_{r=1}^{n} d_r \phi_r \right)^2 dD \tag{7.77}$$

is as small as possible. To this end, we expand the right side of Eq. (7.77), consider Eqs. (7.71) and (7.75), as well as the orthonormality of ϕ_r $(r = 1, 2, \ldots, n)$, and write

$$\begin{aligned} M &= \int_D f^2 \, dD - 2 \sum_{r=1}^{n} d_r \int_D f \phi_r \, dD + \sum_{r=1}^{n} \sum_{s=1}^{n} d_r d_s \int_D \phi_r \phi_s \, dD \\ &= \|f\|^2 - 2 \sum_{r=1}^{n} d_r c_r + \sum_{r=1}^{n} d_r^2 = \|f\|^2 + \sum_{r=1}^{n} (d_r - c_r)^2 - \sum_{r=1}^{n} c_r^2 \end{aligned} \tag{7.78}$$

It is clear from Eq. (7.78) that M takes the smallest value when $d_r = c_r$ $(r = 1, 2, \ldots, n)$. An approximation of the type

$$f_n = \sum_{r=1}^{n} c_r \phi_r, \qquad n = 1, 2, \ldots \tag{7.79}$$

where $\phi_1, \phi_2, \ldots$ are orthonormal functions, is known as a *least squares approximation*, or *an approximation in the mean*. If, by choosing n sufficiently large, the mean square error satisfies the inequality

$$\int_D \left(f - \sum_{r=1}^{n} c_r \phi_r \right)^2 dD = \|f - f_n\|^2 < \epsilon \tag{7.80}$$

where ϵ is any arbitrarily small positive number, the set of functions $\phi_1, \phi_2, \ldots$ is said to be *complete*. Moreover, if

$$\lim_{n\to\infty} \|f - f_n\| = 0 \tag{7.81}$$

then the sequence $f_1, f_2, \ldots$ *converges in the mean* to f.

The completeness of a set of functions does not require that the functions be orthonormal, so that a system of functions is complete if every piecewise continuous function can be approximated in the mean to any desired degree of accuracy by means of a linear combination of the functions of the system. It should be pointed out that, whereas completeness is necessary for convergence, it gives no indication concerning the rate of convergence. The rate of convergence is often as important as convergence itself, as a given function can be approximated by more than one complete set. Clearly, *our interest lies in the set of functions capable of providing an accurate approximation with a number n of terms as small as possible.*

Conservative distributed structures represent a very large and important class of systems, namely, the class of *self-adjoint systems*. The eigenvalues and eigenfunctions of all systems belonging to this class possess very interesting and useful properties. To demonstrate these properties, we consider two $2p$ times differentiable trial functions u and v satisfying all the boundary conditions of the problem, Eq. (7.69), and introduce the inner product

$$(u, Lv) = \int_D uLv\, dD \tag{7.82}$$

Then, we say that the differential operator L is *self-adjoint* if

$$(u, Lv) = (v, Lu) \tag{7.83}$$

Self-adjointness implies certain mathematical symmetry and can be ascertained through integrations by parts with due consideration to the boundary conditions. This mathematical symmetry can be used to simplify the test for self-adjointness, Eq. (7.83). Indeed, if the left (right) side of Eq. (7.83) can be reduced to a symmetric form in u and v and their derivatives through p integrations by parts, then the operator L is self-adjoint, and the test can be regarded as successfully completed. *If the stiffness operator L is self-adjoint, then the system is said to be self-adjoint.*

The concept of self-adjointness can perhaps be best explained by invoking the analogy between distributed and discrete systems. Indeed, the stiffness operator L being self-adjoint corresponds to the stiffness matrix K being symmetric. In the particular case in which the eigenvalue problem is defined by Eqs. (7.68) and (7.69), the mass operator m is a mere function, which is self-adjoint by definition. It corresponds to the mass matrix M being symmetric. Hence, the system self-adjointness, which is implied by the self-adjointness of L, corresponds to the symmetry of the stiffness matrix K and mass matrix M. Because the mass matrix is positive definite, it further corresponds to the symmetry of the system matrix A. In this regard, we recall that in Sec. 4.8 we referred to an algebraic eigenvalue problem as self-adjoint if it is defined by a single real symmetric matrix A.

We denote the symmetric form in u and v resulting from integrations by parts of (u, Lv) by $[u, v]$ and refer to it as an *energy inner product.* For one-dimensional domains, the energy inner product has the general expression

$$[u, v] = \int_0^L \sum_{k=0}^{p} a_k \frac{d^k u}{dx^k}\frac{d^k v}{dx^k} dx + \sum_{\ell=0}^{p-1} b_\ell \frac{d^\ell u}{dx^\ell}\frac{d^\ell v}{dx^\ell}\bigg|_0^L = [v, u] \tag{7.84}$$

where a_k $(k = 0, 1, \ldots, p)$ and b_ℓ $(\ell = 0, 1, \ldots, p-1)$ are in general functions of x. It is clear from Eq. (7.84) that the mathematical symmetry consists of the fact that the functions u and v can trade positions without altering the result. Energy inner products for two-dimensional domains are discussed later in this chapter. The reason for the term "energy inner product" will become evident shortly.

If for any $2p$ times differentiable function u satisfying all the boundary conditions of the problem, Eqs. (7.69), the inequality

$$\int_D uLu\, dD \geq 0 \tag{7.85}$$

is true and the equality sign holds if and only if $u \equiv 0$, then the operator L is said to be *positive definite.* If the expression can be zero without u being identically zero, then the operator L is only *positive semidefinite. If the operator L is positive definite (semidefinite), then the system is positive definite (semidefinite).*

For $v = u$, Eq. (7.84) reduces to

$$[u, u] = \int_0^L \sum_{k=0}^{p} a_k \left(\frac{d^k u}{dx^k}\right)^2 dx + \sum_{\ell=0}^{p-1} b_\ell \left(\frac{d^\ell u}{dx^\ell}\right)^2 \bigg|_0^L \tag{7.86}$$

and we note that $[u, u]$ is a measure of the potential energy. Indeed, if the system executes harmonic motion with frequency ω, then $[u, u]$ is equal to twice the maximum potential energy, which explains why we referred earlier to $[u, v]$ as an energy inner product. Equation (7.86) can be used to define the *energy norm*

$$\|u\|_E = [u, u]^{1/2} \tag{7.87}$$

Next, we introduce the sequence of approximations

$$u_n = \sum_{r=1}^{n} c_r \phi_r, \qquad n = 1, 2, \ldots \tag{7.88}$$

where $\phi_1, \phi_2, \ldots$ are given functions from an independent set. Then, if by choosing n sufficiently large,

$$\|u - u_n\|_E < \epsilon \tag{7.89}$$

in which ϵ is an arbitrarily small positive number, the set of functions $\phi_1, \phi_2, \ldots$ is said to be *complete in energy*. Moreover, if

$$\lim_{n\to\infty} \|u - u_n\|_E = 0 \tag{7.90}$$

the sequence of approximations $u_1, u_2, \ldots$ is said to *converge in energy* to u.

In our study of vibrations, it is convenient to define two classes of functions. One class consists of functions that are $2p$ *times differentiable and satisfy all the boundary conditions of the problem*. This class was introduced earlier in conjunction with the self-adjointness definition. We refer to it as the class of *comparison functions* and denote it by $\mathcal{K}_B^{2p}$. It should be noted that the eigenfunctions are by definition comparison functions, but they represent only a very small subset of the class of comparison functions, as the comparison functions need not satisfy the differential equation. Examining Eq. (7.84), we conclude that the energy inner product is defined for functions outside the space $\mathcal{K}_B^{2p}$. Indeed, Eq. (7.84) is defined for functions that are only p times differentiable. Moreover, in integrating the left side of Eq. (7.83) by parts to obtain the energy inner product, Eq. (7.84), due consideration was given to the natural boundary conditions, in the sense that the higher-order derivatives in the natural boundary conditions were eliminated in favor of lower-order derivatives, such as those arising in geometric boundary conditions. As a result, the energy inner product is defined for functions that are only p times differentiable and satisfy only the geometric boundary conditions. We refer to p *times differentiable functions satisfying only the geometric boundary conditions* of the problem as *admissible functions*, and we denote this class of functions by $\mathcal{K}_G^p$. The comparison functions are by definition admissible functions, and in fact they constitute a small subset of the much larger class of admissible functions.

At this point, we turn our attention to the properties of the eigenvalues and eigenfunctions. To this end, we assume that the problem is self-adjoint and consider two distinct solutions λ_r, w_r and λ_s, w_s of the eigenvalue problem, Eqs. (7.68) and (7.69). Inserting these solutions into Eq. (7.68), we can write

$$Lw_r = \lambda_r m w_r, \qquad Lw_s = \lambda_s m w_s \tag{7.91a, b}$$

Multiplying Eq. (7.91a) by w_s and Eq. (7.91b) by w_r, subtracting the second from the first, integrating over domain D and considering Eqs. (7.82) and (7.83), we obtain

$$\int_D (w_s L w_r - w_r L w_s)\, dD = (\lambda_r - \lambda_s) \int_D m w_r w_s \, dD = 0 \tag{7.92}$$

But, by assumption, the eigenvalues λ_r and λ_s are distinct. Hence, Eq. (7.92) can be satisfied if and only if

$$\int_D m w_r w_s \, dD = 0, \qquad \lambda_r \neq \lambda_s, \qquad r, s = 1, 2, \ldots \tag{7.93}$$

Equation (7.93) represents the *orthogonality relation* for the eigenfunctions of distributed-parameter systems described by the eigenvalue problem given by Eqs. (7.68) and (7.69). Multiplying Eq. (7.91b) by w_r, integrating over D and using Eq. (7.93), it is easy to see that the eigenfunctions satisfy *a second orthogonality relation*, namely,

$$\int_D w_r L w_s \, dD = 0, \qquad \lambda_r \neq \lambda_s, \qquad r, s = 1, 2, \ldots \tag{7.94}$$

It should be stressed here that *the orthogonality of the eigenfunctions is a direct consequence of the system being self-adjoint.*

In the case of repeated eigenvalues, there are as many eigenfunctions belonging to the repeated eigenvalue as the multiplicity of the repeated eigenvalue, and these eigenfunctions are generally not orthogonal to one another, although they are independent and orthogonal to the remaining eigenfunctions of the system. But, as pointed out earlier in this section, independent functions can be orthogonalized by grouping them in proper linear combinations. Hence, *all the eigenfunctions of a self-adjoint system can be regarded as orthogonal, regardless of whether there are repeated eigenvalues or not.*

Because the eigenvalue problem, Eqs. (7.68) and (7.69), is homogeneous, *only the shape of the eigenfunctions is unique*, and *the amplitude is arbitrary*. This arbitrariness can be removed through normalization. A mathematically convenient normalization scheme is given by

$$\int_D m w_r^2 \, dD = 1, \qquad r = 1, 2, \ldots \tag{7.95a}$$

which implies that

$$\int_r w_r L w_r \, dD = \lambda_r, \qquad r = 1, 2, \ldots \tag{7.95b}$$

Then, Eqs. (7.93)–(7.95) can be combined into the *orthonormality relations*

$$\int_D m w_r w_s \, dD = \delta_{rs}, \qquad \int_D w_r L w_s \, dD = \lambda_r \, \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.96a, b}$$

where δ_{rs} is the Kronecker delta.

In Sec. 7.4 and in this section, we assumed on physical grounds that the eigenvalues and eigenfunctions are real. We propose to prove here mathematically that this is indeed so, provided the system is self-adjoint. To this end, we consider a complex solution λ, w of the eigenvalue problem, Eqs. (7.68) and (7.69). Because the eigenvalue problem is real, if the pair λ, w is a complex solution, then the complex conjugate pair $\overline{\lambda}$, $\overline{w}$ must also be a solution, so that Eq. (7.68) yields

$$Lw = \lambda m w, \qquad L\overline{w} = \overline{\lambda} m \overline{w} \tag{7.97a, b}$$

Multiplying Eq. (7.97a) by $\overline{w}$ and Eq. (7.97b) by w, subtracting the second from the first, integrating over D and invoking the self-adjointness of L, we obtain

$$\int_D (\overline{w} L w - w L \overline{w}) \, dD = \left(\lambda - \overline{\lambda}\right) \int_D m \overline{w} w \, dD = 0 \tag{7.98}$$

Introducing the notation

$$\frac{\lambda}{\overline{\lambda}} = \alpha \pm i\beta, \qquad \frac{w}{\overline{w}} = \text{Re}\, w \pm i\text{Im}\, w \tag{7.99}$$

we conclude that the integral on the right side of Eq. (7.98) is real and positive, so that the only alternative is

$$\lambda - \overline{\lambda} = \alpha + i\beta - (\alpha - i\beta) = 2i\beta = 0 \tag{7.100}$$

It follows that *the eigenvalues of a self-adjoint system are real.* As a corollary, *the eigenfunctions of a self-adjoint system are real.* Moreover, contrasting Eq. (7.95b) with inequality (7.85), we conclude that, *if the operator L is positive definite, all the eigenvalues are positive.* On the other hand, *if the operator L is only positive semidefinite, all the eigenvalues are nonnegative*, i.e., some are zero and the rest are positive.

The eigenfunctions w_r $(r = 1, 2, \dots)$ of a self-adjoint system constitute a *complete orthonormal set* of infinite dimension (see, for example, Ref. 1). The implication is that the eigenfunctions can be used as a *basis* for a *function space*, sometimes referred to as a *Hilbert space.* This fact can be stated formally as the following *expansion theorem for self-adjoint systems: Every function w with continuous Lw and satisfying the boundary conditions of the system can be expanded in an absolutely and uniformly convergent series in the eigenfunctions in the form*

$$w = \sum_{r=1}^{\infty} c_r w_r \tag{7.101}$$

where the coefficients c_r are such that

$$c_r = \int_D m w_r w \, dD, \qquad \lambda_r c_r = \int_D w_r L w \, dD, \; r = 1, 2, \dots \tag{7.102a, b}$$

The expansion theorem for self-adjoint distributed systems, Eqs. (7.101) and (7.102), represents the counterpart of the expansion theorem for conservative discrete systems, Eqs. (4.129) and (4.130), defined by symmetric matrices. The expansion theorem is made possible by the orthogonality of the eigenfunctions. Of course, there are many infinite sets of orthogonal functions, such as trigonometric functions, Bessel functions, etc., but none of these can be used as a basis for an expansion theorem for self-adjoint distributed systems, unless they happen to represent the system eigenfunctions. What is so remarkable about the eigenfunctions is that they are orthogonal not only with respect to the mass density m but also with respect to the stiffness operator L, as indicated by Eqs. (7.93) and (7.94).

The expansion theorem plays a pivotal role in the vibration of self-adjoint systems, as it permits the solution of the boundary-value problem by transforming it into an infinite set of *modal equations*, which are second-order ordinary differential equations for the time-dependent modal coordinates. The solution process is known as *modal analysis* and is entirely analogous to the one for discrete systems. In fact, the second-order differential equations look exactly like the modal equations for discrete systems, and can be solved by the techniques discussed in Chapter 3. The only difference is that in distributed systems the set is infinite and in discrete systems the set is finite.

The class of self-adjoint systems is extremely large and it includes essentially all the vibrating conservative systems discussed in this text. If it can be demonstrated that an individual system belongs to this class, then it can be assumed automatically that the system possesses the remarkable properties of self-adjoint systems discussed in this section. In effect, the various conservative systems considered in this text can be regarded as special cases. It should be pointed out that, provided a system is self-adjoint, the general properties hold, regardless of whether a closed-form solution to the differential eigenvalue problem exists or not. This fact can be of great value in developing approximate solutions.

Example 7.3

Consider the eigenvalue problem for the string in transverse vibration shown in Fig. 7.1a, demonstrate that the problem fits one of the generic formulations of this section, check whether the problem is self-adjoint and positive definite and draw conclusions.

From Sec. 7.4, Eqs. (7.60) and (7.61), the eigenvalue problem is given by the differential equation

$$-\frac{d}{dx}\left[T(x)\frac{dW(x)}{dx}\right] = \lambda\rho(x)W(x), \qquad 0 < x < L \tag{a}$$

and the boundary conditions

$$W = 0, \qquad x = 0 \tag{b}$$

$$T(x)\frac{dW(x)}{dx} + KW(x) = 0, \qquad x = L \tag{c}$$

The eigenvalue problem fits the generic formulation given by Eqs. (7.68) and (7.69), so that, comparing Eq. (a) with Eq. (7.68), we conclude that

$$L = -\frac{d}{dx}\left[T(x)\frac{d}{dx}\right], \qquad m = \rho \tag{d}$$

Because L is of order 2, $p = 1$. Moreover, comparing Eqs. (b) and (c) with Eqs. (7.69) we can write

$$B_1 = 1, \qquad x = 0 \tag{e}$$

$$B_1 = T(x)\frac{d}{dx} + K, \qquad x = L \tag{f}$$

To check for self-adjointness, we write

$$(u, Lv) = \int_0^L uLv\,dx = -\int_0^L u\frac{d}{dx}\left(T\frac{dv}{dx}\right)dx = -uT\frac{dv}{dx}\bigg|_0^L + \int_0^L \frac{du}{dx}T\frac{dv}{dx}dx \tag{g}$$

Upon considering boundary conditions (b) and (c), Eq. (g) reduces to the energy inner product

$$[u, v] = Ku(L)v(L) + \int_0^L T\frac{du}{dx}\frac{dv}{dx}dx = [v, u] \tag{h}$$

which is clearly symmetric, so that *the operator L, and hence the system, is self-adjoint.* We conclude immediately that *the eigenvalues are real and the eigenfunctions are orthogonal with respect to the mass density m and the stiffness operator L*.

Finally, we let $v = u$ in Eq. (h) and obtain the energy norm squared

$$[u, u] = \|u\|_E^2 = Ku^2(L) + \int_0^L T\left(\frac{du}{dx}\right)^2 dx > 0 \tag{i}$$

which, in view of Eq. (7.8), is recognized as twice the maximum potential energy. Clearly, the energy norm cannot be zero, except in the trivial case, so that *the operator L, and hence the system, is positive definite.* It follows that *all the eigenvalues are positive.*

Example 7.4

Consider the rotating cantilever beam in bending vibration shown in Fig. 7.4a and demonstrate that the eigenvalue problem fits one of the generic formulations of this section. Then, check whether the system is self-adjoint and positive definite and draw conclusions.

The boundary-value problem for a rotating cantilever beam was derived in Example 7.2. Letting $f = 0$ and using the procedure presented in Sec. 7.4, Eqs. (d)–(f) of Example 7.2 yield the eigenvalue problem defined by the differential equation

$$\frac{d^2}{dx^2}\left(EI\frac{d^2W}{dx^2}\right) - \frac{d}{dx}\left(P\frac{dW}{dx}\right) = \lambda mW, \qquad 0 < x < L \tag{a}$$

and the boundary conditions

$$W = 0, \qquad \frac{dW}{dx} = 0, \qquad x = 0 \tag{b}$$

$$EI\frac{d^2W}{dx^2} = 0, \qquad \frac{d}{dx}\left(EI\frac{d^2W}{dx^2}\right) = 0, \qquad x = L \tag{c}$$

Comparing Eq. (a) with Eq. (7.68), we conclude that

$$L = \frac{d^2}{dx^2}\left(EI\frac{d^2}{dx^2}\right) - \frac{d}{dx}\left(P\frac{d}{dx}\right) \tag{d}$$

Moreover, comparing Eqs. (b) and (c) with Eqs. (7.69), we can write

$$\begin{aligned} &B_1 = 1, \qquad B_2 = \frac{d}{dx}, \qquad x = 0 \\ &B_1 = EI\frac{d^2}{dx^2}, \qquad B_2 = \frac{d}{dx}\left(EI\frac{d^2}{dx^2}\right), \qquad x = L \end{aligned} \tag{e}$$

Next, we consider

$$\begin{aligned}(u, Lv) &= \int_0^L uLv\,dx = \int_0^L u\left[\frac{d^2}{dx^2}\left(EI\frac{d^2v}{dx^2}\right) - \frac{d}{dx}\left(P\frac{dv}{dx}\right)\right]dx \\ &= u\frac{d}{dx}\left(EI\frac{d^2v}{dx^2}\right)\Big|_0^L - \frac{du}{dx}EI\frac{d^2v}{dx^2}\Big|_0^L + \int_0^L \frac{d^2u}{dx^2}EI\frac{d^2v}{dx^2}dx \\ &\quad - uP\frac{dv}{dx}\Big|_0^L + \int_0^L \frac{du}{dx}P\frac{dv}{dx}dx \end{aligned} \tag{f}$$

and use boundary conditions (b) and (c) to obtain the symmetric energy inner product

$$[u, v] = \int_0^L \left(EI \frac{d^2u}{dx^2}\frac{d^2v}{dx^2} + P\frac{du}{dx}\frac{dv}{dx} \right) dx = [v, u] \tag{g}$$

It follows that *the operator L, and hence the system, is self-adjoint,* so that *the eigenvalues are real and the eigenfunctions are orthogonal with respect to the mass density m and the stiffness operator L.* Then, letting $v = u$ in Eq. (g), we obtain the energy norm squared

$$[u, u] = \|u\|_E^2 = \int_0^L \left[EI \left(\frac{d^2u}{dx^2} \right)^2 + P \left(\frac{du}{dx} \right)^2 \right] dx \geq 0 \tag{h}$$

The norm reduces to zero for $u =$ constant. In view of the first of boundary conditions (b), however, this constant must be zero. Because $u =$ constant $\neq 0$ is not a solution of the eigenvalue problem, the energy norm is positive definite, so that *the operator L, and hence the system, is positive definite,* which implies that *all the eigenvalues are positive.*

7.6 SOLUTION OF THE EIGENVALUE PROBLEM FOR STRINGS, RODS AND SHAFTS

As indicated at the end of Sec. 7.1, the behavior of strings in transverse vibration, rods in axial vibration and shafts in torsional vibration is described by mathematically equivalent second-order boundary-value problems, the difference between the various types lying in the nature of the parameters. This implies that the solution obtained for one of the three types of systems applies equally well for the remaining two. To emphasize this point, we shall solve the eigenvalue problem for all three types interchangeably using the formulation derived in Sec. 7.4 for strings.

i. Strings in transverse vibration

A problem of considerable interest in vibrations is that of *a string fixed at both ends.* From Sec. 7.4, the eigenvalue problem for a vibrating string is defined by the differential equation

$$-\frac{d}{dx}\left[T(x)\frac{dW(x)}{dx} \right] = \lambda\rho(x)W(x), \qquad \lambda = \omega^2, \qquad 0 < x < L \tag{7.103}$$

In the case at hand, the boundary conditions are

$$W(0) = 0, \qquad W(L) = 0 \tag{7.104a, b}$$

Using the method of Sec. 7.5, it is not difficult to verify that the system is self-adjoint and positive definite, so that the eigenvalues are real and positive and the eigenfunctions are real and orthogonal. Although this is one of the simplest examples of a distributed-parameter systems, no closed-form solution exists in the general case in which the tension $T(x)$ and the mass density $\rho(x)$ are arbitrary functions of the spatial position x. A closed-form solution does exist in the frequently encountered case in which *the tension is constant,* $T(x) = T =$ constant, and *the mass density is* *uniform*, $\rho(x) = \rho =$ constant. In this case, the differential equation, Eq. (7.103), can be rewritten as

$$\frac{d^2W(x)}{dx^2} + \beta^2 W(x) = 0, \qquad \beta^2 = \frac{\omega^2 \rho}{T}, \qquad 0 < x < L \tag{7.105}$$

The boundary conditions, Eqs. (7.104), do not depend on the system parameters, so that they remain the same. The solution of Eq. (7.105) is simply

$$W(x) = C_1 \sin\beta x + C_2 \cos\beta x \tag{7.106}$$

where C_1 and C_2 are constants. Solution (7.106) holds true for all strings with constant tension and uniform mass distribution, and it reduces to the eigenfunctions of a particular system only after the boundary conditions have been enforced.

Inserting solution (7.106) into boundary condition (7.104a), we obtain

$$W(0) = C_2 = 0 \tag{7.107}$$

Then, introducing solution (7.106) with $C_2 = 0$ into boundary condition (7.104b), we have

$$W(L) = C_1 \sin\beta L = 0 \tag{7.108}$$

Equation (7.108) can be satisfied in two ways. The first alternative is to set $C_1 = 0$, which corresponds to the trivial solution, so that it must be ruled out. Hence, we must opt for the second alternative, namely,

$$\sin\beta L = 0 \tag{7.109}$$

Equation (7.109) is known as the *characteristic equation*, or *frequency equation*, and has the denumerably infinite set of solutions

$$\beta_r = \frac{r\pi}{L}, \qquad r = 1, 2, \ldots \tag{7.110}$$

which represent the system *eigenvalues*. It should be pointed out that the term "eigenvalues" is used somewhat loosely here, as strictly speaking the eigenvalues of the system are λ_r, which are related to β_r by

$$\lambda_r = \beta_r^2 T/\rho, \qquad r = 1, 2, \ldots \tag{7.111}$$

Moreover, recalling that $\lambda = \omega^2$, Eqs. (7.110) and (7.111) can be combined to obtain the *natural frequencies*

$$\omega_r = \sqrt{\lambda_r} = \beta_r\sqrt{\frac{T}{\rho}} = r\pi\sqrt{\frac{T}{\rho L^2}}, \qquad r = 1, 2, \ldots \tag{7.112}$$

The frequency ω_1 is called the *fundamental frequency* and the higher frequencies are known as *overtones*. Overtones that are *integer multiples* of the fundamental frequency are called *higher harmonics*, in which case the fundamental frequency represents the *fundamental harmonic*. There are only a few vibrating systems with harmonic overtones, and most of them are used in musical instruments because they tend to produce pleasant sounds. In this regard, it should be mentioned that in a symphonic orchestra the group of instruments including the violin, viola, cello, etc., is commonly referred to as the "string section".

We conclude from the above that the two boundary conditions can be used to determine one of the constants and to derive the characteristic equation. The second constant of integration, C_1 in the case at hand, cannot be determined uniquely, so that the amplitude of the solution is arbitrary. This is consistent with the fact that the eigenvalue problem, Eqs. (7.103) and (7.104), is a homogeneous problem. In view of Eq. (7.110), the *eigenfunction*, or *natural mode*, belonging to β_r can be written in the form

$$W_r(x) = A_r \sin \frac{r\pi x}{L}, \qquad r = 1, 2, \dots \tag{7.113}$$

It is easy to verify that *the eigenfunctions are orthogonal* both with respect to the mass density ρ and with respect to the operator L, which in the case of constant tension reduces to

$$L = -T\frac{d^2}{dx^2} \tag{7.114}$$

The amplitudes A_r can be rendered unique through normalization. A convenient normalization process is given by

$$\int_0^L \rho W_r^2(x)\,dx = 1, \qquad r = 1, 2, \dots \tag{7.115}$$

which yields the *orthonormal set of eigenfunctions*, or *normal modes*

$$W_r(x) = \sqrt{\frac{2}{\rho L}} \sin \frac{r\pi x}{L}, \qquad r = 1, 2, \dots \tag{7.116}$$

The first three modes and natural frequencies are displayed in Fig. 7.5. We observe that there are points at which the displacement is zero. These points are referred to as *nodes* and they form a certain pattern. Indeed, excluding the end points, the mode $W_r(x)$ has $r - 1$ equidistant nodes occuring at the points $x_i = iL/r$ $(i = 1, 2, \dots, r-1)$.

The term "denumerably infinite set" introduced in Sec. 7.5 implies that the eigenvalues λ_r $(r = 1, 2, \dots)$ assume an infinite set of discrete values. Under these circumstances, the string is said to possess a *discrete frequency spectrum*. Whereas the shape of the modes is independent of the system parameters, the natural frequencies are proportional to the square root of the tension T, inversely proportional to the square root of the mass density ρ and inversely proportional to the length L, as can be observed from Eq. (7.112). In many string instruments, such as the violin, there are four strings, all differing in density. For all practical purposes, the density of each of the strings can be regarded as constant. Hence, the frequency spectrum for each string can be altered by changing the tension an d the length. The tension T is generally held constant. In fact, the process of tuning a violin consists of adjusting the tension so as to ensure a certain fundamental frequency. During performance, the violinist alters the length of the string so as to produce the notes demanded by the score. It should be pointed out that pleasing sounds are produced by enriching the fundamental harmonic with certain higher harmonics.

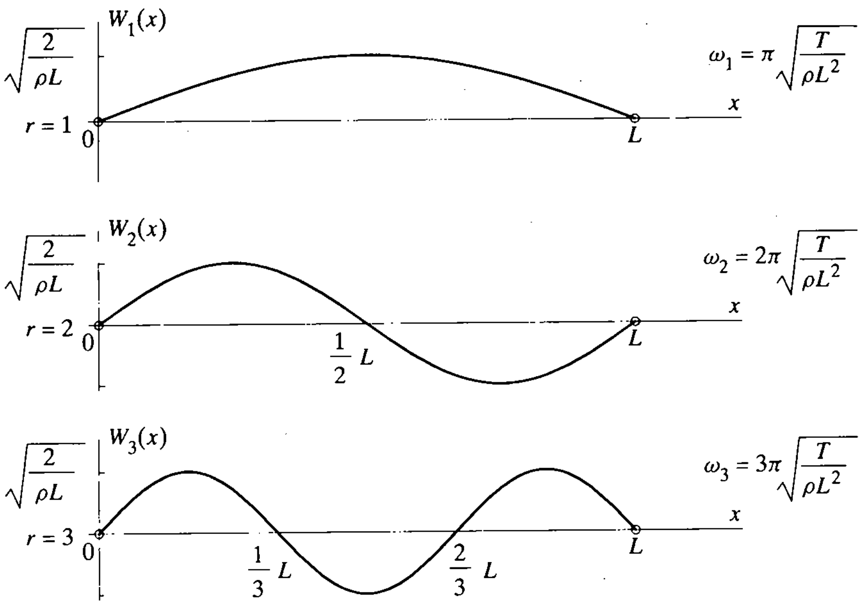

$W_1(x)$
$\sqrt{\frac{2}{\rho L}}$
$r = 1$
0
L
x
$\omega_1 = \pi\sqrt{\frac{T}{\rho L^2}}$
$W_2(x)$
$\sqrt{\frac{2}{\rho L}}$
$r = 2$
0
$\frac{1}{2}L$
L
x
$\omega_2 = 2\pi\sqrt{\frac{T}{\rho L^2}}$
$W_3(x)$
$\sqrt{\frac{2}{\rho L}}$
$r = 3$
0
$\frac{1}{3}L$
$\frac{2}{3}L$
L
x
$\omega_3 = 3\pi\sqrt{\frac{T}{\rho L^2}}$

Figure 7.5 First three modes of vibration of a uniform string fixed at both ends

As the length of a string increases, the natural frequencies draw closer and closer together. In fact, as L approaches infinity, we obtain a *continuous frequency spectrum*. At this point it is no longer meaningful to speak of natural frequencies and natural modes, and a different point of view must be adopted. Indeed, for infinitely long strings the motion can be regarded as consisting of *traveling waves*. The wave description of motion applies also to strings of finite length, except that in this case the waves are reflected from the boundaries, and the combination of incident and reflected waves gives rise to *standing waves*. It can be shown that the natural modes description of vibration is mathematically equivalent to the standing waves description (see Ref. 8, Sec. 8-2).

ii. Rods in axial vibration

Invoking the analogy discussed at the end of Sec. 7.1, the eigenvalue problem for a rod in axial vibration can be described by the differential equation

$$-\frac{d}{dx}\left[EA(x)\frac{dU(x)}{dx}\right] = \lambda m(x)U(x), \quad \lambda = \omega^2, \qquad 0 < x < L \quad (7.117)$$

where $EA(x)$ is the axial stiffness, in which E is Young's modulus and $A(x)$ is the cross-sectional area, and $m(x)$ is the mass density. The solution $U(x)$ is subject to given boundary conditions. We consider a rod *fixed at* $x = 0$ *and free at* $x = L$, as shown in Fig. 7.6, so that the boundary conditions are

$$U(0) = 0, \qquad EA(x)\frac{dU(x)}{dx}\bigg|_{x=L} = 0 \qquad (7.118a, b)$$

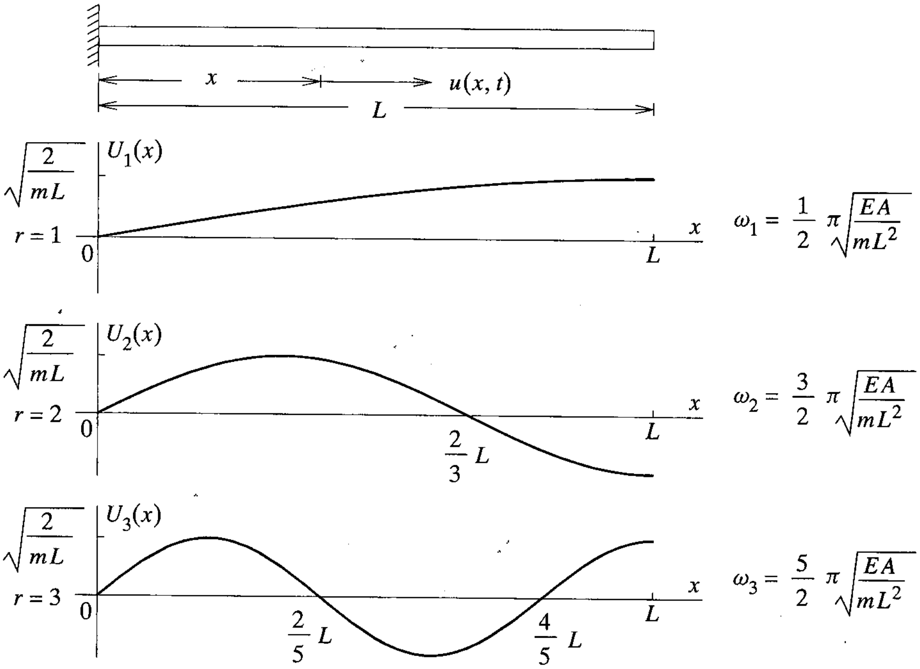


Figure 7.6 Rod in axial vibration fixed at $x = 0$ and free at $x = L$, and the first three modes of vibration

There is no difficulty in demonstrating that the problem is self-adjoint and positive definite.

For a *uniform* rod, $EA(x) = EA =$ constant, $m(x) = m =$ constant, the eigenvalue problem reduces to the differential equation

$$\frac{d^2U(x)}{dx^2} + \beta^2 U(x) = 0, \qquad \beta^2 = \frac{\omega^2 m}{EA}, \qquad 0 < x < L \tag{7.119}$$

and the boundary conditions

$$U(0) = 0, \qquad \left.\frac{dU(x)}{dx}\right|_{x=L} = 0 \tag{7.120a, b}$$

The differential equation is essentially the same as that for a uniform string, Eq. (7.105), so that the solution is

$$U(x) = C_1 \sin\beta x + C_2 \cos\beta x \tag{7.121}$$

Using boundary condition (7.120a), we conclude that $C_2 = 0$. Moreover, use of boundary condition (7.120b) yields the characteristic equation

$$\cos\beta L = 0 \tag{7.122}$$

Its solutions consist of the eigenvalues

$$\beta_r = \frac{(2r-1)\pi}{2L}, \qquad r = 1, 2, \ldots \tag{7.123}$$

so that the natural frequencies are

$$\omega_r = \beta_r\sqrt{\frac{EA}{m}} = \frac{(2r-1)\pi}{2}\sqrt{\frac{EA}{mL^2}}, \qquad r = 1, 2, \ldots \tag{7.124}$$

The eigenfunctions belonging to β_r are

$$U_r(x) = A_r \sin\frac{(2r-1)\pi x}{2L}, \qquad r = 1, 2, \ldots \tag{7.125}$$

and they are orthogonal. The coefficients A_r are arbitrary, and we propose to normalize them so as to satisfy

$$\int_0^L mU_r^2(x)\,dx = 1, \qquad r = 1, 2, \ldots \tag{7.126}$$

so that the eigenfunctions reduce to the orthonormal set

$$U_r(x) = \sqrt{\frac{2}{mL}}\sin\frac{(2r-1)\pi x}{2L}, \qquad r = 1, 2, \ldots \tag{7.127}$$

The first three modes are shown in Fig. 7.6. Note that, as customary, displacements have been plotted vertically, when they are in fact in the axial direction. Excluding the point $x = 0$, the rth mode, $U_r(x)$, has nodes at the points $x_i = 2iL/(2r-1)$ $(i = 1, 2, \ldots, r-1)$.

It will prove instructive to investigate a different case, namely, that in which *both ends* of the rod *are free* (Fig. 7.7). Of course, the differential equation remains the same, but the new boundary conditions are

$$EA(x)\frac{dU(x)}{dx} = 0, \qquad x = 0, L \tag{7.128}$$

Once again it can be verified that the problem is self-adjoint. To check for positive definiteness, we carry out an integration by parts with due consideration to the boundary conditions and obtain

$$\begin{aligned}\int_0^L U_r L U_r\,dx &= -\int_0^L U_r\frac{d}{dx}\left(EA\frac{dU_r}{dx}\right)dx\\ &= -U_r\left(EA\frac{dU_r}{dx}\right)\Big|_0^L + \int_0^L \frac{dU_r}{dx}\left(EA\frac{dU_r}{dx}\right)dx\\ &= \int_0^L EA\left(\frac{dU_r}{dx}\right)^2 dx \geq 0\end{aligned} \tag{7.129}$$

The last integral in (7.129) is equal to zero if U_r is constant. However, unlike our earlier experience in which end restraints required that the constant be zero, in the case at hand a nonzero constant solution is possible. It follows that *for a free-free rod in axial vibration the operator L is only positive semidefinite*, so that *the system is only positive semidefinite.*

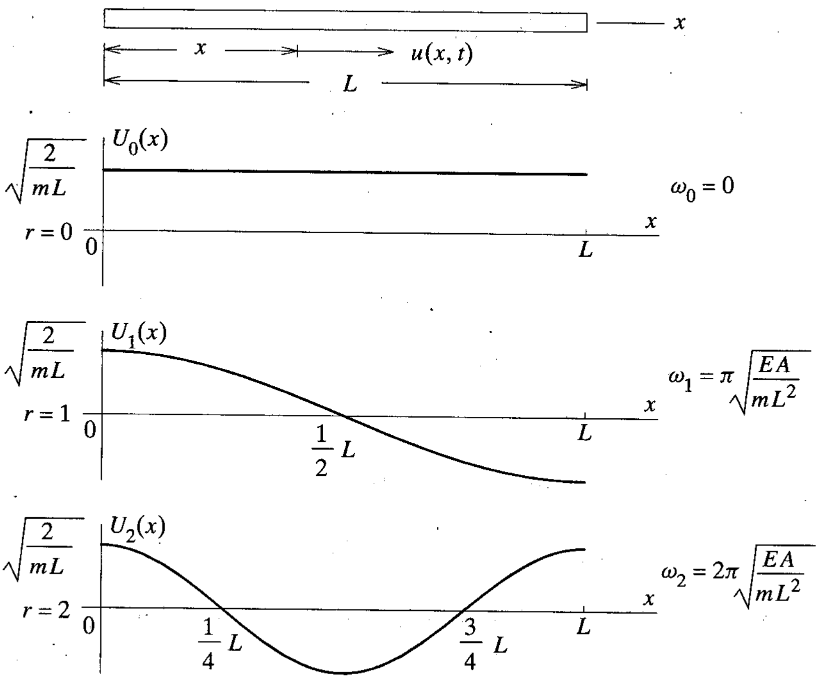


Figure 7.7 Rod in axial vibration free at both ends, the rigid-body mode and the first two elastic modes

In view of the fact that the system is positive semidefinite, *zero eigenvalues*, and hence *zero natural frequencies are possible.* The eigenfunctions belonging to zero eigenvalues represent *rigid-body modes.* To examine the question of rigid-body modes in more detail, we let $\lambda = \lambda_0 = 0$, $U = U_0$ in Eq. (7.117) and write

$$\frac{d}{dx}\left[EA(x)\frac{dU_0(x)}{dx}\right] = 0, \qquad 0 < x < L \tag{7.130}$$

Integrating with respect to x once and considering boundary conditions (7.128), we have

$$EA(x)\frac{dU_0(x)}{dx} = 0, \qquad 0 < x < L \tag{7.131}$$

Ignoring $EA(x)$ and integrating once more, we obtain

$$U_0(x) = A_0 = \frac{1}{\sqrt{mL}} \tag{7.132}$$

which represents the rigid-body mode with the zero natural frequency, $\omega_0 = 0$, and note that the mode has been normalized so that $\int_0^L mU_0^2 dx = 1$. Clearly, in the case under consideration *there is only one rigid-body mode.* Physically, the rigid-body mode represents displacement of the body as a whole, without elastic deformations. Rigid-body modes are typical of unrestrained systems, for which there are no forces or moments exerted by the supports. In the case at hand, we are concerned with forces in the longitudinal direction alone, and not with moments.

Next, we assume that the external excitations are zero and consider the vibration of the rod in the rth mode. Using Newton's second law, the equation of motion in the axial direction is

$$\begin{aligned} F(t) &= \int_0^L m(x)\frac{\partial^2 u_r(x,t)}{dt^2}dx \\ &= -\left[\int_0^L m(x)U_r(x)\,dx\right]c_r\omega_r^2\cos(\omega_r t - \phi_r) = 0, \\ &\qquad r = 1, 2, \ldots \end{aligned} \tag{7.133}$$

which can be interpreted as the orthogonality of the rigid-body mode to the elastic modes and rewritten in the form

$$\int_0^L m(x)U_0U_r(x)\,dx = 0, \qquad r = 1, 2, \ldots \tag{7.134}$$

Recalling that the system is self-adjoint and normalizing the elastic modes, we can extend the orthonormality relations so as to include the rigid-body mode, or

$$\int_0^L m(x)U_r(x)U_s(x)\,dx = \delta_{rs}, \qquad r, s = 0, 1, 2, \ldots \tag{7.135a}$$

$$\int_0^L U_r(x)LU_s(x)\,dx = \lambda_r\,\delta_{rs}, \qquad r, s = 0, 1, 2, \ldots \tag{7.135b}$$

At this point, we return to the solution of the eigenvalue problem, Eqs. (7.117) and (7.128). For a *uniform rod*, the differential equation is given once again by Eq. (7.119) and its solution by Eq. (7.121). Then, we conclude that, in contrast with the fixed-free case, in the free-free case the first of boundary conditions (7.128) yields $C_1 = 0$, whereas the second gives the characteristic equation

$$\sin\beta L = 0 \tag{7.136}$$

which leads to the eigenvalues

$$\beta_r = \frac{r\pi}{L}, \qquad r = 0, 1, 2, \ldots \tag{7.137}$$

and we note that $\beta_0 = 0$ is also an eigenvalue, corroborating our discussion of the rigid-body mode. Upon normalization, the eigenfunctions of a *free-free* rod are

$$\begin{aligned} U_0(x) &= A_0 = \frac{1}{\sqrt{mL}} \\ U_r(x) &= A_r\cos\frac{r\pi x}{L} = \sqrt{\frac{2}{mL}}\cos\frac{r\pi x}{L}, \qquad r = 1, 2, \ldots \end{aligned} \tag{7.138}$$

The first three modes are plotted in Fig. 7.7. We observe that the modes have nodes at the points $x_i = (2i-1)L/2r$ $(i = 1, 2, \ldots, r)$.

iii. Shafts in torsional vibration

Using once again the analogy discussed at the end of Sec. 7.1, we express the eigenvalue problem for a shaft in torsional vibration by means of the differential equation

$$-\frac{d}{dx}\left[GJ(x)\frac{d\Theta(x)}{dx}\right] = \lambda I(x)\Theta(x), \qquad \lambda = \omega^2, \qquad 0 < x < L \qquad (7.139)$$

where $GJ(x)$ is the torsional stiffness, in which G is the shear modulus and $J(x)$ is the area polar moment of inertia, and $I(x)$ is the mass polar moment of inertia density. The solution $\Theta(x)$ must satisfy boundary conditions yet to be specified. We consider a shaft *clamped at* $x = 0$ and *supported by a torsional spring of stiffness* K_T *at* $x = L$ (Fig. 7.8). A system analogous in both the differential equation and the boundary conditions was considered in Example 7.3 in the form of a string in transverse vibration. Hence, using the analogy with the string of Example 7.3, the boundary conditions are

$$\Theta(0) = 0; \; GJ(x)\frac{d\Theta(x)}{dx} + K_T\Theta(x) = 0, \qquad x = L \qquad (7.140a, b)$$

The system was shown in Example 7.3 to be self-adjoint and positive definite, so that the eigenvalues are real and positive and the eigenfunctions are real and orthogonal.

Under the assumption that *the shaft is uniform*, $GJ(x) = GJ =$ constant, $I(x) = I =$ constant, the solution once again has the form given by Eq. (7.121), except that $\Theta(x)$ replaces $U(x)$. Moreover, use of boundary condition (7.140a) results once again in $C_2 = 0$, so that

$$\Theta(x) = C_1 \sin\beta x \qquad (7.141)$$

On the other hand, boundary condition (7.140b) yields

$$GJ\beta C_1 \cos\beta L + K_T C_1 \sin\beta L = 0 \qquad (7.142)$$

The solution $C_1 = 0$ must be ruled out as representing the trivial solution. Hence, dividing through by C_1 and rearranging, we obtain the characteristic equation

$$\tan\beta L = -\frac{GJ}{K_T L}\beta L \qquad (7.143)$$

which is a transcendental equation in βL; its solutions consist of a denumerably infinite set of eigenvalues $\beta_r L (r = 1, 2, \ldots)$. The natural frequencies are related to the eigenvalues by

$$\omega_r = \beta_r L\sqrt{\frac{GJ}{IL^2}}, \qquad r = 1, 2, \ldots \qquad (7.144)$$

Belonging to the eigenvalues $\beta_r L$ are the eigenfunctions

$$\Theta_r(x) = A_r \sin\beta_r x, \qquad r = 1, 2, \ldots \qquad (7.145)$$

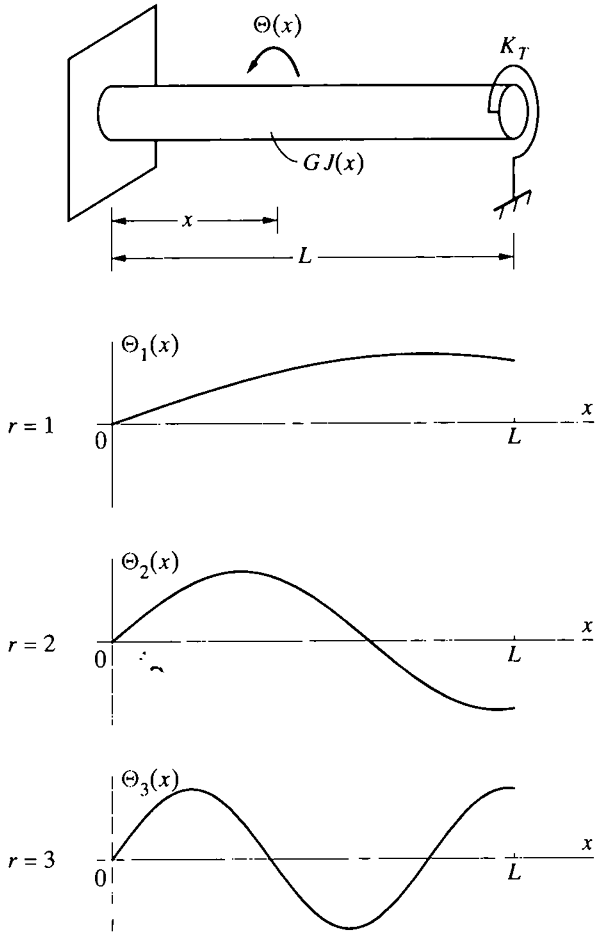


Figure 7.8 Shaft in torsional vibration clamped at $x = 0$ and supported by a spring at $x = L$, and the first three modes of vibration

where A_r are arbitrary amplitudes. The eigenfunctions are orthogonal and can be normalized so as to satisfy $\int_0^L I\Theta_r^2\,dx = 1$, in which case the coefficients A_r can be shown to have the values

$$A_r = 2\sqrt{\frac{\beta_r}{I\,(2\beta_r L - \sin 2\beta_r L)}}, \qquad r = 1, 2, \ldots \tag{7.146}$$

The first three eigenfunctions for a ratio $GJ/K_T L = 1$ are plotted in Fig. 7.8.

The solution of the characteristic equation, Eq. (7.143) must be obtained numerically for a given ratio $GJ/K_T L$ of parameters. If the eigenvalues need not be very accurate, a solution can also be obtained graphically, as shown in Fig. 7.9. We observe from the figure that, as $r \to \infty$, the eigenvalues $\beta_r L$ approach odd multiples of $\pi/2$ and the amplitudes of the eigenfunctions approach $\sqrt{2/IL}$, both

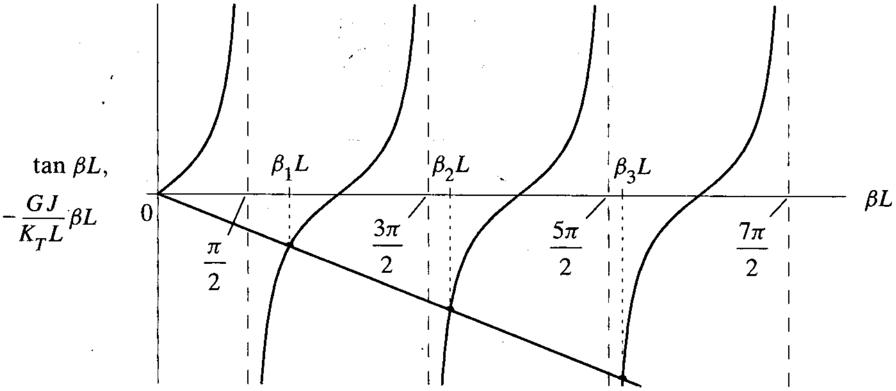


Figure 7.9 Graphical solution of the characteristic equation, Eq. (7.143).

eigenvalues and eigenfunctions being typical of a clamped-free system. Hence, the effect of the end spring K_T tends to diminish as the mode number increases.

In the fixed-fixed string, fixed-free rod and free-free rod discussed earlier in this section, the orthogonality of the modes was guaranteed by the system self-adjointness, and the same can be said about the fixed-spring supported shaft at hand. But, whereas in the first three cases orthogonality can be verified by inspection, this is not true in the present case. Verification of the orthogonality of the eigenfunctions given by Eq. (7.145) can be carried out by showing that the integral $\int_0^L \sin \beta_r x \sin \beta_s x \, dx$ is zero for $r \neq s$, which requires the use of the characteristic equation, Eq. (7.143). Of course, the fact that the system is self-adjoint makes this verification unnecessary.

7.7 SOLUTION OF THE EIGENVALUE PROBLEM FOR BEAMS IN BENDING

The differential eigenvalue problem for beams in bending was derived in Sec. 7.4. In this section, we wish to solve the problem for a number of cases lending themselves to closed-form solution.

We consider first the simplest case, namely, that of a *uniform beam hinged at both ends and with no axial force.* Under these circumstances, the differential equation, Eq. (7.62), can be rewritten in the form

$$\frac{d^4 W(x)}{dx^4} - \beta^4 W(x) = 0, \qquad \beta^4 = \frac{\omega^2 m}{EI}, \qquad 0 < x < L \tag{7.147}$$

and the boundary conditions, Eqs. (7.63b) and (7.64a), reduce to

$$W(0) = 0, \qquad W(L) = 0, \qquad \left.\frac{d^2 W(x)}{dx^2}\right|_{x=0} = 0, \qquad \left.\frac{d^2 W(x)}{dx^2}\right|_{x=L} = 0 \tag{7.148a–d}$$

The solution of Eq. (7.147) is

$$W(x) = C_1 \sin \beta x + C_2 \cos \beta x + C_3 \sinh \beta x + C_4 \cosh \beta x \tag{7.149}$$

and we note that solution (7.149) is valid for all uniform beams. Differences in the solution begin to appear only when the boundary conditions are enforced. Using boundary conditions (7.148a) and (7.148c), we obtain

$$W(0) = C_2 + C_4 = 0 \tag{7.150a}$$

$$\left.\frac{d^2W(x)}{dx^2}\right|_{x=0} = -\beta^2(C_2 - C_4) = 0 \tag{7.150b}$$

which yield

$$C_2 = C_4 = 0 \tag{7.151}$$

On the other hand, using boundary conditions (7.148b) and (7.148d), we have

$$W(L) = C_1 \sin\beta L + C_3 \sinh\beta L = 0 \tag{7.152a}$$

$$\left.\frac{d^2W(x)}{dx^2}\right|_{x=L} = -\beta^2(C_1 \sin\beta L - C_3 \sinh\beta L) = 0 \tag{7.152b}$$

Equations (7.152) have nontrivial solutions provided

$$C_3 = 0 \tag{7.153}$$

and

$$\sin\beta L = 0 \tag{7.154}$$

where the latter is recognized as the characteristic equation. Its solutions are the eigenvalues

$$\beta_r L = r\pi, \qquad r = 1, 2, \ldots \tag{7.155}$$

Belonging to these eigenvalues are the eigenfunctions

$$W_r(x) = \sqrt{\frac{2}{mL}} \sin\frac{r\pi x}{L}, \qquad r = 1, 2, \ldots \tag{7.156}$$

which were normalized so as to satisfy $\int_0^L m W_r^2\, dx = 1$. We observe that the eigenvalues and eigenfunctions are the same as for a uniform fixed-fixed string, so that the first three modes have the same shape as in Fig. 7.5. However, from Eq. (7.147), the natural frequencies are

$$\omega_r = (r\pi)^2 \sqrt{\frac{EI}{mL^4}}, \qquad r = 1, 2, \ldots \tag{7.157}$$

which are different from the natural frequencies of the fixed-fixed string.

A somewhat more involved case is the *uniform cantilever beam*, namely, *a beam with one end clamped and the other end free*, as shown in Fig. 7.10. In this case, the boundary conditions are

$$W(0) = 0, \quad \left.\frac{dW(x)}{dx}\right|_{x=0} = 0, \quad \left.\frac{d^2W(x)}{dx^2}\right|_{x=L} = 0, \quad \left.\frac{d^3W(x)}{dx^3}\right|_{x=L} = 0 \tag{7.158a-d}$$

Inserting solution (7.149) into boundary conditions (7.158a) and (7.158b), we obtain

$$W(0) = C_2 + C_4 = 0, \qquad \left.\frac{dW(x)}{dx}\right|_{x=0} = \beta\,(C_1 + C_3) = 0 \tag{7.159}$$

so that the solution reduces to

$$W(x) = C_1\,(\sin\beta x - \sinh\beta x) + C_2\,(\cos\beta x - \cosh\beta x) \tag{7.160}$$

Then, using boundary conditions (7.158c) and (7.158d), we arrive at the two simultaneous homogeneous equations

$$-\beta^2\,[C_1\,(\sin\beta L + \sinh\beta L) + C_2\,(\cos\beta L + \cosh\beta L)] = 0 \tag{7.161a}$$

$$-\beta^3\,[C_1\,(\cos\beta L + \cosh\beta L) - C_2\,(\sin\beta L - \sinh\beta L)] = 0 \tag{7.161b}$$

Equating the determinant of the coefficients to zero, we obtain the characteristic equation

$$\cos\beta L\cos h\beta L = -1 \tag{7.162}$$

The solutions, obtained numerically, are $\beta_1 L = 1.875$, $\beta_2 L = 4.694$, $\beta_3 L = 7.855, \ldots$. In addition, solving Eq. (7.161b) for C_2 in terms of C_1 and substituting into Eq. (7.160), we obtain the corresponding eigenfunctions

$$\begin{aligned} W_r(x) = A_r\,[&(\sin\beta_r L - \sinh\beta_r L)\,(\sin\beta_r x - \sinh\beta_r x) \\ &+ (\cos\beta_r L + \cosh\beta_r L)\,(\cos\beta_r x - \cosh\beta_r x)], \qquad r = 1, 2, \ldots \end{aligned} \tag{7.163}$$

where we introduced the notation $A_r = C_1/\,(\sin\beta_r L - \sinh\beta_r L)$. The system can be verified to be self-adjoint and positive definite, with the usual positivity of the eigenvalues and orthogonality of the eigenfunctions. The normalization ordinarily used in this text is not feasible. The first three natural modes and natural frequencies are displayed in Fig. 7.10. The mode $W_r(x)$ has $r - 1$ nodes $(r = 1, 2, \ldots, n)$, but their location can no longer be expressed as a rational fraction of L.

Another case of interest is the *free-free beam* (Fig. 7.11). From Sec. 7.4, the eigenvalue problem is defined by the differential equation

$$\frac{d^2}{dx^2}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] = \lambda m(x)W(x), \quad \lambda = \omega^2, \quad 0 < x < L \tag{7.164}$$

and the boundary conditions

$$\left.EI(x)\frac{d^2W(x)}{dx^2}\right|_{x=0} = 0, \qquad \left.\frac{d}{dx}\left[EI(x)\frac{d^2W(x)}{dx^2}\right]\right|_{x=0} = 0 \tag{7.165a, b}$$

$$\left.EI(x)\frac{d^2W(x)}{dx^2}\right|_{x=L} = 0, \qquad \left.\frac{d}{dx}\left[EI(x)\frac{d^2W(x)}{dx^2}\right]\right|_{x=L} = 0 \tag{7.165c, d}$$

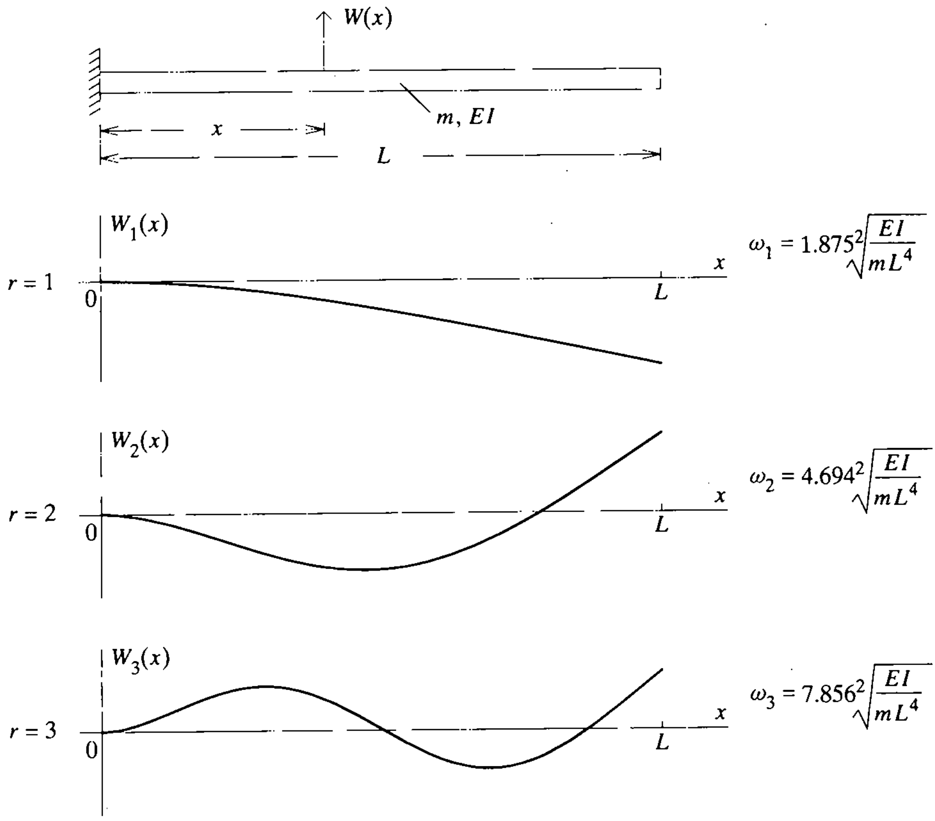

$W(x)$
m, EI
x
L
$W_1(x)$
$r = 1$
0
L
x
$\omega_1 = 1.875^2\sqrt{\frac{EI}{mL^4}}$
$W_2(x)$
$r = 2$
0
L
x
$\omega_2 = 4.694^2\sqrt{\frac{EI}{mL^4}}$
$W_3(x)$
$r = 3$
0
L
x
$\omega_3 = 7.856^2\sqrt{\frac{EI}{mL^4}}$

Figure 7.10 Cantilever beam and the first three modes of vibration

Using results obtained in Example 7.4, it is not difficult to show that the system is self-adjoint, so that the eigenfunctions are real and orthogonal. Moreover, the energy norm squared is given by

$$\|W\|_E^2 = \int_0^L EI(x)\left[\frac{d^2W(x)}{dx^2}\right]^2 dx \geq 0 \tag{7.166}$$

and it is easy to see that the integral is equal to zero if W is either a constant or a linear function of x. We observe that boundary conditions (7.165) permit such solutions. Under these circumstances, *the system is only positive semidefinite*, so that *the system admits* eigensolutions in the form of *rigid-body modes with zero natural frequencies*. To examine the nature of the rigid-body modes, we let $\lambda = \lambda_0 = 0$ in Eq. (7.164) and write

$$\frac{d^2}{dx^2}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] = 0, \qquad 0 < x < L \tag{7.167}$$

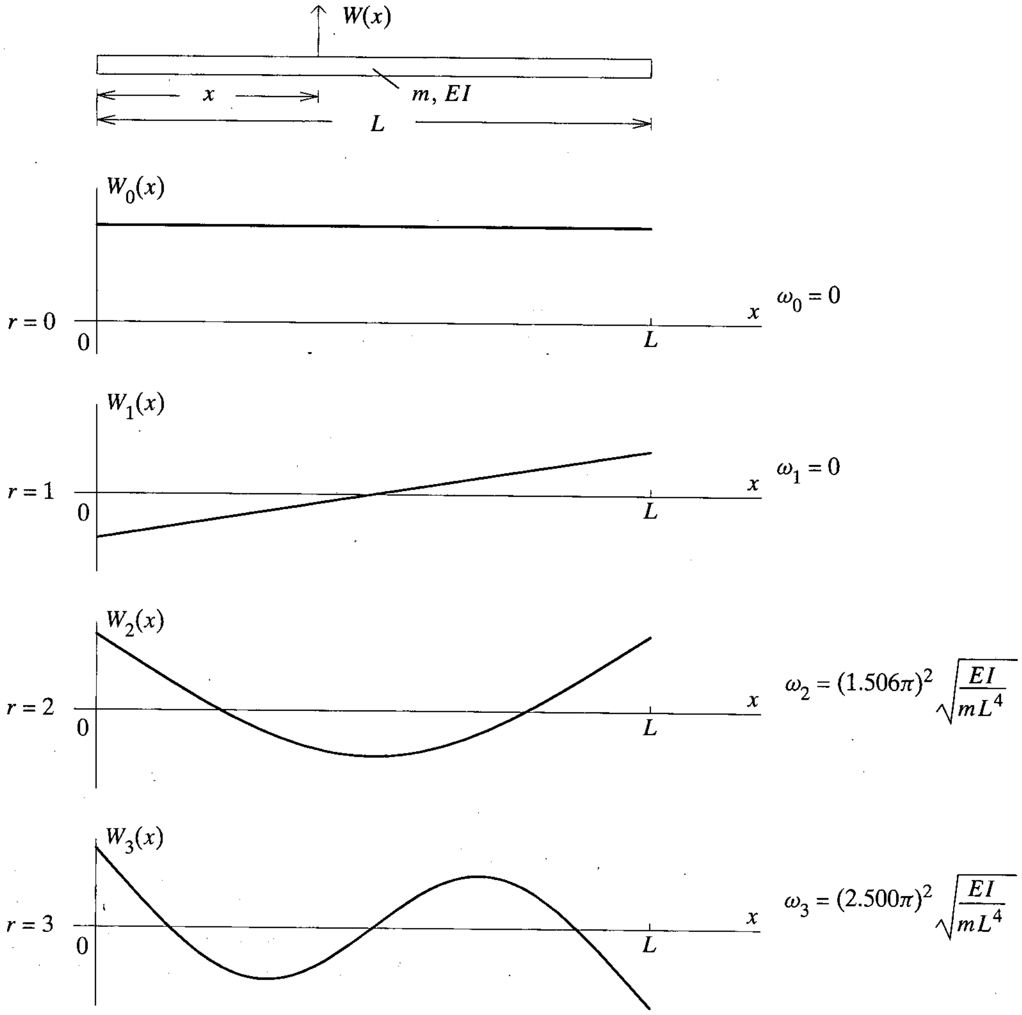


Figure 7.11 Free-free beam, the two rigid-body modes and the first two elastic modes

Integrating Eq. (7.167) once and using boundary conditions (7.165b) and (7.165d), we obtain

$$\frac{d}{dx}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] = 0, \qquad 0 < x < L \tag{7.168}$$

One more integration in conjunction with boundary conditions (7.165a) and (7.165c) yields

$$EI(x)\frac{d^2W(x)}{dx^2} = 0, \qquad 0 < x < L \tag{7.169}$$

Then, dividing by $EI(x)$ and integrating twice, we can write

$$W(x) = D_1 + D_2x \tag{7.170}$$

Because $W(x)$ contains two independent constants of integration, we conclude that there are two rigid-body modes. It is convenient to identify them as the transverse translation of the mass center C and rotation about C. Upon the usual normalization, the two rigid-body modes can be shown to have the form

$$W_0(x) = A_0 = \frac{1}{\sqrt{mL}} \tag{7.171a}$$

$$W_1(x) = A_1(x - x_C) = \frac{1}{\sqrt{I_C}}(x - x_C) \tag{7.171b}$$

where I_C is the mass moment of inertia of the beam about C and x_C is the distance between the left end and the mass center C.

Following the pattern established in Sec. 7.6, for zero resultants of external forces and moments about C, Newton's second law for the motion in the rth mode can be written as

$$\begin{aligned} F(t) &= \int_0^L m(x)\frac{\partial^2 w_r(x,t)}{\partial t^2}dx \\ &= -\left[\int_0^L m(x)W_r(x)dx\right]c_r\omega_r^2\cos(\omega_r t - \phi_r) = 0, \\ &\qquad r = 2, 3, \ldots \end{aligned} \tag{7.172a}$$

$$\begin{aligned} M_C(t) &= \int_0^L m(x)(x - x_C)\frac{\partial^2 w_r(x,t)}{\partial t^2}dx \\ &= -\left[\int_0^L m(x)(x - x_C)W_r(x)dx\right]c_r\omega_r^2\cos(\omega_r t - \phi_r) = 0, \\ &\qquad r = 2, 3, \ldots \end{aligned} \tag{7.172b}$$

In view of Eqs. (7.171), Eqs. (7.172) amount to the orthogonality of the rigid-body modes to the elastic modes. Upon normalization, the orthonormality relations for all modes are

$$\int_0^L m(x)W_r(x)W_s(x)\,dx = \delta_{rs}, \quad r = 0, 1, 2, \ldots \tag{7.173a}$$

$$\int_0^L W_r(x)\frac{d^2}{dx^2}\left[EI(x)\frac{d^2W_s(x)}{dx^2}\right]dx = \lambda_r\,\delta_{rs}, \quad r = 0, 1, 2, \ldots \tag{7.173b}$$

Next, we wish to solve the eigenvalue problem for the elastic modes of the *uniform free-free beam*. The solution is once again given by Eq. (7.149), whereas the boundary conditions, Eqs. (7.165), reduce to

$$\frac{d^2W(x)}{dx^2} = 0, \qquad \frac{d^3W(x)}{dx^3} = 0, \qquad x = 0, L \tag{7.174}$$

Inserting Eq. (7.149) into Eqs. (7.174) and using the same pattern as earlier in this section, we obtain the characteristic equation

$$\cos\beta L \cosh\beta L = 1 \tag{7.175}$$

which has the roots $\beta_0 L = \beta_1 L = 0$, $\beta_2 L = 1.506\pi$, $\beta_3 L = 2.500\pi, \ldots$, and note that for large r the eigenvalues approach $(2r-1)\pi/2$. The eigenfunctions belonging to $\beta_r L$ $(r = 2, 3, \ldots)$ are

$$\begin{aligned} W_r(x) = A_r\,[&(\cos\beta_r L - \cosh\beta_r L)(\sin\beta_r x + \sinh\beta_r x) \\ &- (\sin\beta_r L - \sinh\beta_r L)(\cos\beta_r x + \cosh\beta_r x)], \\ &\qquad r = 2, 3, \ldots \end{aligned} \tag{7.176}$$

The first four natural modes and natural frequencies are displayed in Fig. 7.11.

7.8 EXTENSIONS OF LAGRANGE'S EQUATION FOR DISTRIBUTED SYSTEMS

The Lagrange equation for distributed systems derived in Sec. 7.3 is somewhat limited, as it excludes certain effects that cannot always be ignored. The reason for this is mostly pedagogical, as the inclusion of these effects tends to raise the level of difficulty of the formulation. In this section, we propose to extend the Lagrange equation so as to include these effects. In carrying out the extension we follow a parallel path to the one of Sec. 7.3.

As in Sec. 7.3, we wish to derive the Lagrange equation by means of the extended Hamilton's principle, Eq. (7.31). The additional effects to be included here can all be accounted for in the kinetic energy, so that we replace Eq. (7.32) by

$$T = T_0\left[\dot{w}(0,t), \dot{w}'(0,t)\right] + T_L\left[\dot{w}(L,t), \dot{w}'(L,t)\right] + \int_0^L \hat{T}\left(\dot{w}, \dot{w}'\right) dx \tag{7.177}$$

and we note that the kinetic energy terms T_0 and T_L are designed to account for the effect of lumped boundary masses in translation and rotation and the kinetic energy density $\hat{T}$ includes the rotation of a differential element of mass, in addition to the translation considered in Eq. (7.32). On the other hand, the potential energy remains in the form given by Eq. (7.33). Combining Eqs. (7.33) and (7.177), the Lagrangian can be expressed as

$$\begin{aligned} L = L_0&\left[w(0,t),\ w'(0,t),\ \dot{w}(0,t),\ \dot{w}'(0,t)\right] \\ &+ L_L\left[w(L,t),\ w'(L,t),\ \dot{w}(L,t),\ \dot{w}'(L,t)\right] \\ &+ \int_0^L \hat{L}\left(w,\ w',\ w'',\ \dot{w},\ \dot{w}'\right) dx \end{aligned} \tag{7.178}$$

The virtual work remains as given by Eq. (7.35).

The extended Hamilton's principle requires the variation in the Lagrangian, which retains the form given by Eq. (7.36), except that the individual terms, Eqs. (7.37), must be augmented as follows:

$$\begin{aligned}\delta L_0 &= \frac{\partial L_0}{\partial w(0,t)}\delta w(0,t) + \frac{\partial L_0}{\partial w'(0,t)}\delta w'(0,t)\\ &\quad + \frac{\partial L_0}{\partial \dot{w}(0,t)}\delta \dot{w}(0,t) + \frac{\partial L_0}{\partial \dot{w}'(L,t)}\delta \dot{w}'(L,t) \qquad (7.179a)\\ \delta L_L &= \frac{\partial L_L}{\partial w(L,t)}\delta w(L,t) + \frac{\partial L_L}{\partial w'(L,t)}\delta w'(L,t)\\ &\quad + \frac{\partial L_L}{\partial \dot{w}(L,t)}\delta \dot{w}(L,t) + \frac{\partial L_L}{\partial \dot{w}'(L,t)}\delta \dot{w}'(L,t) \qquad (7.179b)\\ \delta \hat{L} &= \frac{\partial \hat{L}}{\partial w}\delta w + \frac{\partial \hat{L}}{\partial w'}\delta w' + \frac{\partial \hat{L}}{\partial w''}\delta w'' + \frac{\partial \hat{L}}{\partial \dot{w}}\delta \dot{w} + \frac{\partial \hat{L}}{\partial \dot{w}'}\delta \dot{w}' \qquad (7.179c)\end{aligned}$$

and we note that each of Eqs. (7.179a) and (7.179b) have two extra terms compared to Eqs. (7.37a) and (7.37b), and Eq. (7.179c) has one additional term. The next step in the use of the extended Hamilton's principle is to carry out integrations by parts with respect to x and t so as to produce variations in translational and rotational displacements alone. Many of these steps are given by Eqs. (7.38) and (7.39), so that it is necessary to carry out the integrations by parts only for the additional terms in Eqs. (7.179). To this end, we integrate with respect to t, recall that the variations vanish at $t = t_1, t_2$ and write

$$\begin{aligned}&\int_{t_1}^{t_2} \frac{\partial L_0}{\partial \dot{w}(0,t)}\delta \dot{w}(0,t)\,dt\\ &\quad = \frac{\partial L_0}{\partial \dot{w}(0,t)}\delta w(0,t)\bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}(0,t)}\right)\delta w(0,t)\,dt\\ &\quad = -\int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}(0,t)}\right)\delta w(0,t)\,dt \qquad (7.180a)\\ &\int_{t_1}^{t_2} \frac{\partial L_0}{\partial \dot{w}'(0,t)}\delta \dot{w}'(0,t)\,dt\\ &\quad = \frac{\partial L_0}{\partial \dot{w}'(0,t)}\delta w'(0,t)\bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}'(0,t)}\right)\delta w'(0,t)\,dt\\ &\quad = -\int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}'(0,t)}\right)\delta w'(0,t)\,dt \qquad (7.180b)\\ &\int_{t_1}^{t_2} \frac{\partial L_L}{\partial \dot{w}(L,t)}\delta \dot{w}(L,t)\,dt\end{aligned}$$

$$= \frac{\partial L_L}{\partial \dot{w}(L,t)} \delta w(L,t) \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial L_L}{\partial \dot{w}(L,t)} \right) \delta w(L,t)\, dt$$

$$= -\int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial L_L}{\partial \dot{w}(L,t)} \right) \delta w(L,t)\, dt \tag{7.180c}$$

$$\int_{t_1}^{t_2} \frac{\partial L_L}{\partial \dot{w}'(L,t)} \delta \dot{w}'(L,t)\, dt$$

$$= \frac{\partial L_L}{\partial \dot{w}'(L,t)} \delta w'(L,t) \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial L_L}{\partial \dot{w}'(L,t)} \right) \delta w'(L,t)\, dt$$

$$= -\int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial L_L}{\partial \dot{w}'(L,t)} \right) \delta w'(L,t)\, dt \tag{7.180d}$$

The next term involves integrations both with respect to x and t, as well as changes in the order of these integrations, as follows:

$$\int_{t_1}^{t_2} \int_0^L \frac{\partial \hat{L}}{\partial \dot{w}'} \delta \dot{w}' dx dt = \int_0^L \left(\int_{t_1}^{t_2} \frac{\partial \hat{L}}{\partial \dot{w}'} \delta \dot{w}' dt \right) dx$$

$$= \int_0^L \left[\frac{\partial \hat{L}}{\partial \dot{w}'} \delta w' \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \delta w' dt \right] dx$$

$$= -\int_{t_1}^{t_2} \left[\int_0^L \frac{\partial}{\partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \delta w' dx \right] dt$$

$$= -\int_{t_1}^{t_2} \left[\frac{\partial}{\partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \delta w \Big|_0^L - \int_0^L \frac{\partial^2}{\partial x \partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \delta w\, dx \right] dt \tag{7.181}$$

Introducing Eqs. (7.35) and (7.36) into Eq. (7.31), considering Eqs. (7.38), (7.39), (7.180) and (7.181) and collecting terms involving $\delta w(x,t)$, $\delta w(0,t)$, $\delta w(L,t)$, $\delta w'(0,t)$ and $\delta w'(L,t)$, we obtain the replacement of Eq. (7.40) in the form

$$\int_{t_1}^{t_2} \Bigg\{ \int_0^L \Bigg[\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x} \left(\frac{\partial \hat{L}}{\partial w'} \right) + \frac{\partial^2}{\partial x^2} \left(\frac{\partial \hat{L}}{\partial w''} \right)$$

$$- \frac{\partial}{\partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}} \right) + \frac{\partial^2}{\partial x \partial t} \left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) + f \Bigg] \delta w\, dx$$

$$+ \left[\frac{\partial L_0}{\partial w(0,t)} - \frac{\partial}{\partial t} \left(\frac{\partial L_0}{\partial \dot{w}(0,t)} \right) \right] \delta w(0,t)$$

$$+ \left[\frac{\partial L_0}{\partial w'(0,t)} - \frac{\partial}{\partial t} \left(\frac{\partial L_0}{\partial \dot{w}'(0,t)} \right) \right] \delta w'(0,t)$$

$$\begin{aligned}
&+ \left[\frac{\partial L_L}{\partial w(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}(L,t)} \right) \right] \delta w(L,t) \\
&+ \left[\frac{\partial L_L}{\partial w'(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}'(L,t)} \right) \right] \delta w'(L,t) \\
&+ \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''} \right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \right] \delta w \Big|_0^L + \frac{\partial \hat{L}}{\partial w''} \delta w' \Big|_0^L \Bigg\} dt \\
= \int_{t_1}^{t_2} \Bigg\langle \int_0^L &\Bigg[\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'} \right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''} \right) \\
&- \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}} \right) + \frac{\partial^2}{\partial x \partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) + f \Bigg] \delta w \\
&+ \Bigg\{ \frac{\partial L_0}{\partial w(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}(0,t)} \right) \\
&- \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''} \right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \right] \Bigg|_{x=0} \Bigg\} \delta w(0,t) \\
&+ \Bigg\{ \frac{\partial L_L}{\partial w(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}(L,t)} \right) \\
&+ \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''} \right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) \right] \Bigg|_{x=L} \Bigg\} \delta w(L,t) \\
&+ \left[\frac{\partial L_0}{\partial w'(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}'(0,t)} \right) - \frac{\partial \hat{L}}{\partial w''} \Big|_{x=0} \right] \delta w'(0,t) \\
&+ \Bigg[\frac{\partial L_L}{\partial w'(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}'(L,t)} \right) \\
&+ \frac{\partial \hat{L}}{\partial w''} \Big|_{x=L} \Bigg] \delta w'(L,t) \Bigg\rangle dt = 0
\end{aligned} \tag{7.182}$$

Then, invoking the arbitrariness of the virtual displacements in a manner similar to that in Sec. 7.3, we conclude that Eq. (7.182) can be satisfied for all δw over the open domain $0 < x < L$ if and only if

$$\frac{\partial \hat{L}}{\partial w} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w'} \right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial \hat{L}}{\partial w''} \right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}} \right) + \frac{\partial^2}{\partial x \partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'} \right) + f = 0, \qquad 0 < x < L \tag{7.183}$$

Moreover, by writing

$$\left\{\frac{\partial L_0}{\partial w(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}(0,t)}\right) - \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'}\right)\right]\bigg|_{x=0}\right\}\delta w(0,t) = 0 \tag{7.184a}$$

$$\left[\frac{\partial L_0}{\partial w'(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}'(0,t)}\right) - \frac{\partial \hat{L}}{\partial w''}\bigg|_{x=0}\right]\delta w'(0,t) = 0 \tag{7.184b}$$

and

$$\left\{\frac{\partial L_L}{\partial w(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}(L,t)}\right) + \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'}\right)\right]\bigg|_{x=L}\right\}\delta w(L,t) = 0 \tag{7.185a}$$

$$\left[\frac{\partial L_L}{\partial w'(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}'(L,t)}\right) + \frac{\partial \hat{L}}{\partial w''}\bigg|_{x=L}\right]\delta w'(L,t) = 0 \tag{7.185b}$$

we take into account that either $\delta w(0,t)$ or its coefficient is zero and either $\delta w'(0,t)$ or its coefficient is zero, and similar statements can be made about conditions at the end $x = L$.

Equation (7.183) represents *Lagrange's differential equation of motion* corresponding to the extended Lagrangian given by Eq. (7.178). Moreover, Eqs. (7.184) and (7.185) can be used to obtain a variety of possible boundary conditions. Indeed, from Eqs. (7.184) we conclude that at $x = 0$ either

$$\frac{\partial L_0}{\partial w(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}(0,t)}\right) - \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'}\right)\right]\bigg|_{x=0} = 0 \tag{7.186a}$$

or

$$w = 0 \tag{7.186b}$$

and either

$$\frac{\partial L_0}{\partial w'(0,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_0}{\partial \dot{w}'(0,t)}\right) - \frac{\partial \hat{L}}{\partial w''}\bigg|_{x=0} = 0 \tag{7.187a}$$

or

$$w' = 0 \tag{7.187b}$$

In addition, from Eqs. (7.185), at $x = L$ either

$$\frac{\partial L_L}{\partial w(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}(L,t)}\right) + \left[\frac{\partial \hat{L}}{\partial w'} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial w''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{w}'}\right)\right]\bigg|_{x=L} = 0 \tag{7.188a}$$

or

$$w = 0 \tag{7.188b}$$

and either

$$\frac{\partial L_L}{\partial w'(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{w}'(L,t)}\right) + \left.\frac{\partial \hat{L}}{\partial w''}\right|_{x=L} = 0 \tag{7.189a}$$

or

$$w' = 0 \tag{7.189b}$$

The Lagrange equation together with appropriate boundary conditions constitute a boundary-value problem. The formulation consisting of Eqs. (7.183)–(7.189) is suitable for fourth-order systems, but it can be used for second-order systems as well by merely omitting terms and boundary conditions that do not apply. In the case of a fourth-order system, the boundary-value problem consists of Lagrange's equation, Eq. (7.183), and two boundary conditions at each end, namely, one from each of Eqs. (7.186)–(7.189). On the other hand, the boundary-value problem for second-order systems consists of the differential equation, Eq. (7.183), with the third and fifth term removed and one boundary condition at each end, one from each of Eqs. (7.186) and Eqs. (7.188), where the fourth and fifth term are deleted from Eqs. (7.186a) and (7.188a).

Example 7.5

Derive the boundary-value problem for a shaft in torsional vibration with the left end clamped and with the right end supporting a disk of mass moment of inertia I_D, as shown in Fig. 7.12.

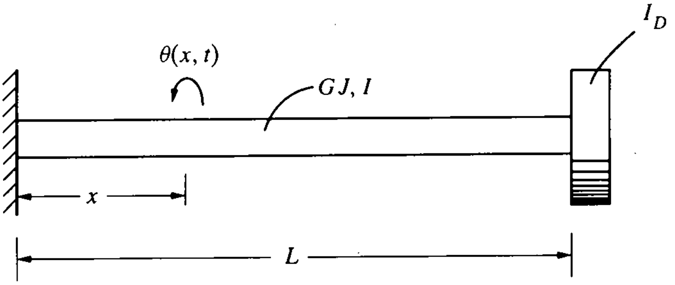


Figure 7.12 Shaft in torsional vibration clamped at $x = 0$ and with a disk at $x = L$

The kinetic energy has the expression

$$T(t) = \frac{1}{2}\int_0^L I(x)\left[\frac{\partial\theta(x,t)}{\partial t}\right]^2 dx + \frac{1}{2}I_D\dot{\theta}^2(L,t) \tag{a}$$

and the potential energy is simply

$$V(t) = \frac{1}{2}\int_0^L GJ(x)\left[\frac{\partial\theta(x,t)}{\partial x}\right]^2 dx \tag{b}$$

so that the Lagrangian can be written in the form

$$L = L_L + \int_0^L \hat{L}\,dx \tag{c}$$

where the boundary Lagrangian is given by

$$L_L = \frac{1}{2} I_D \dot{\theta}^2(L,t) \tag{d}$$

and the Lagrangian density by

$$\hat{L} = \hat{T} - \hat{V} = \frac{1}{2} I(x)\dot{\theta}^2(x,t) - \frac{1}{2} GJ(x)\left[\theta'(x,t)\right]^2 \tag{e}$$

This being a second-order system, Lagrange's equation, Eq. (7.183), reduces to

$$\frac{\partial \hat{L}}{\partial \theta} - \frac{\partial}{\partial x}\left(\frac{\partial \hat{L}}{\partial \theta'}\right) - \frac{\partial}{\partial t}\left(\frac{\partial \hat{L}}{\partial \dot{\theta}}\right) + m = 0, \qquad 0 < x < L \tag{f}$$

in which $m = m(x,t)$ is a distributed torque. Inserting Eq. (e) into Eq. (f) and recognizing that $I(x)$ does not depend on t, we obtain the partial differential equation of motion

$$\frac{\partial}{\partial x}\left(GJ\frac{\partial \theta}{\partial x}\right) - I\frac{\partial^2 \theta}{\partial t^2} + m = 0, \qquad 0 < x < L \tag{g}$$

In view of the fact that the left end is clamped, according to Eq. (7.186b), the boundary condition at $x = 0$ is

$$\theta(0,t) = 0 \tag{h}$$

Moreover, the boundary condition at $x = L$ is given by Eq. (7.188a) with appropriate deletions, or

$$\frac{\partial L_L}{\partial \theta(L,t)} - \frac{\partial}{\partial t}\left(\frac{\partial L_L}{\partial \dot{\theta}(L,t)}\right) + \left.\frac{\partial \hat{L}}{\partial \theta'}\right|_{x=L} = 0 \tag{i}$$

Hence, inserting Eqs. (d) and (e) into Eq. (i), the boundary condition at $x = L$ is

$$I_D \ddot{\theta}(L,t) + GJ(x)\theta'(x,t)\big|_{x=L} = 0 \tag{j}$$

We observe that, whereas there is nothing unusual about the differential equation, Eq. (g), and the boundary condition at $x = 0$, Eq. (h), the boundary condition at $x = L$, Eq. (j), depends on the acceleration $\ddot{\theta}(L,t)$. As a result, the associated differential eigenvalue problem does not fit the general mold defined by Eqs. (7.68) and (7.69), so that a more general formulation is required to accommodate the system of Fig. 7.12.

Example 7.6

Derive the boundary-value problem for the beam in bending vibration considered in Sec. 7.2 under the assumption that the kinetic energy of rotation is not negligible.

The various terms are the same as in Sec. 7.2, except that the kinetic energy, Eq. (7.18), must be replaced by

$$T(t) = \frac{1}{2}\int_0^L \left\{ m(x)\left[\frac{\partial w(x,t)}{\partial t}\right]^2 + J(x)\left[\frac{\partial^2 w(x,t)}{\partial t \partial x}\right]^2 \right\} dx \tag{a}$$

where $J(x)$ is the mass moment of inertia per unit length of beam. Hence, combining Eqs. (a) and (7.19), we can write the Lagrangian density

$$\hat{L} = \hat{T} - \hat{V} = \frac{1}{2}\left[m\dot{w}^2 + J\left(\dot{w}'\right)^2 - EI\left(w''\right)^2 - P\left(w'\right)^2\right] \tag{b}$$

and we note that there are no boundary Lagrangians.

Inserting Eq. (b) into Eq. (7.183), we obtain the explicit Lagrange equation

$$\frac{\partial}{\partial x}\left(P\frac{\partial w}{\partial x}\right)-\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right)-m\frac{\partial^2 w}{\partial t^2}+\frac{\partial}{\partial x}\left(J\frac{\partial^3 w}{\partial x\partial t^2}\right)+f=0, \qquad 0<x<L \tag{c}$$

Moreover, introducing Eq. (b) into Eqs. (7.186)–(7.189), we conclude that at $x = 0, L$ either

$$P\frac{\partial w}{\partial x}-\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right)+J\frac{\partial^3 w}{\partial x\partial t^2}=0 \tag{d}$$

must be satisfied or

$$w=0 \tag{e}$$

and either

$$EI\frac{\partial^2 w}{\partial x^2}=0 \tag{f}$$

or

$$w'=0 \tag{g}$$

for a total of two boundary conditions at each end.

From the differential equation, Eq. (c), we observe that there are two terms involving accelerations, and one of them involves spatial derivatives. Hence the mass density is no longer a simple function, but a differential expression. Moreover, from Eq. (d), we conclude that, if this boundary condition applies, then it depends on the angular acceleration. Note that the added terms involve J and are referred to as *rotatory inertia* terms. The differential eigenvalue problem for a beam in bending with the rotatory inertia included cannot be accommodated by the formulation of Sec. 7.5, and a generalization of the formulation is necessary.

7.9 GENERALIZATION OF THE DIFFERENTIAL EIGENVALUE PROBLEM FOR SELF-ADJOINT SYSTEMS

From Sec. 7.8, we conclude that the eigenvalue problem given by Eqs. (7.68) and (7.69) cannot accommodate a number of important problems. In view of this, we consider a more general eigenvalue problem and express the differential equation in the operator form

$$Lw=\lambda Mw, \qquad x, y \text{ in } D \tag{7.190}$$

where L and M are linear homogeneous differential operators of order $2p$ and $2q$, and are referred to as *stiffness operator and mass operator*, respectively, λ is a parameter and D is the domain of definition of Eq. (7.190). The operators L and M are of the type (7.66) and their order is such that $p > q$. Associated with the differential equation (7.190) there are p boundary conditions to be satisfied by the solution w at every point of the boundary S of the domain D. The boundary conditions are of the type

$$B_i w=0, \qquad x, y \text{ on } S, \qquad i=1,2,\ldots,k \tag{7.191a}$$

$$B_i w=\lambda C_i w, \qquad x, y \text{ on } S, \qquad i=k+1,\ k+2,\ldots,p \tag{7.191b}$$

where B_i and C_i are linear homogeneous differential *boundary operators* of maximum order $2p-1$ and $2q-1$, respectively.

Next, we consider two comparison functions u and v and state that *the stiffness operator L is self-adjoint* if

$$\int_D uLv\,dD + \sum_{i=k+1}^{p} \int_S uB_i v\,dS = \int_D vLu\,dD + \sum_{i=k+1}^{p} \int_S vB_i u\,dS \qquad (7.192a)$$

Moreover, *the mass operator M is self-adjoint* if

$$\int_D uMv\,dD + \sum_{i=k+1}^{p} \int_S uC_i v\,dS = \int_D vMu\,dD + \sum_{i=k+1}^{p} \int_S vC_i u\,dS \qquad (7.192b)$$

If L and M are self-adjoint, *the system*, or *the eigenvalue problem, is said to be self-adjoint.* As demonstrated in Sec. 7.5, self-adjointness can be ascertained through integration by parts with due consideration to the boundary conditions, and it implies certain mathematical symmetry. In fact, the concept of self-adjointness of the stiffness operator L and mass operator M in distributed systems is entirely analogous to the concept of symmetry of the stiffness matrix K and mass matrix M in discrete systems. This mathematical symmetry can be used to simplify the test for self-adjointness, as shown in Sec. 7.5. Indeed, if the left side of Eq. (7.192a), or of Eq. (7.192b), can be reduced to a symmetric form in u and v through integrations by parts, then the operator L, or operator M, is self-adjoint, and it is not really necessary to carry out the integrations on the right side of Eq. (7.192a), or Eq. (7.192b), as they are guaranteed to yield the same symmetric forms. For one-dimensional self-adjoint systems, we denote the symmetric result of the integrations by parts for the operator L by

$$\begin{aligned}[u, v]_P &= \int_D uLv\,dD + \sum_{i=k+1}^{p} \int_S uB_i v\,dS \\ &= \int_0^L \sum_{k=0}^{p} a_k \frac{d^k u}{dx^k}\frac{d^k v}{dx^k}dx + \sum_{\ell=0}^{p-1} b_\ell \frac{d^\ell u}{dx^\ell}\frac{d^\ell v}{dx^\ell}\bigg|_0^L \end{aligned} \qquad (7.193a)$$

and for the operator M by

$$\begin{aligned}[u, v]_K &= \int_D uMv\,dD + \sum_{i=k+1}^{p} \int_S uC_i v\,dS \\ &= \int_0^L \sum_{k=0}^{q} e_k \frac{d^k u}{dx^k}\frac{d^k v}{dx^u}dx + \sum_{\ell=0}^{q-1} f_\ell \frac{d^\ell u}{dx^\ell}\frac{d^\ell v}{dx^\ell}\bigg|_0^L \end{aligned} \qquad (7.193b)$$

where $[u, v]_P$ and $[u, v]_K$ will be referred to as *potential and kinetic energy inner products*, respectively, in which a_k, b_ℓ, e_k and f_ℓ are coefficients depending in general on x.

If for any comparison function u we have the inequality

$$\int_D uLu\,dD + \sum_{i=k+1}^{p} \int_S uB_i u\,dS \geq 0 \qquad (7.194a)$$

and the equality sign holds if and only if $u \equiv 0$, then *the operator L is said to be positive definite.* If the expression can be zero without u being identically zero, then *the operator L is only positive semidefinite.* Similarly, if

$$\int_D uMu\,dD + \sum_{i=k+1}^{p} \int_S uC_i u\,dS \geq 0 \tag{7.194b}$$

and the equality holds if and only if $u \equiv 0$, *the operator M is positive definite*, and if the expression is zero without u being identically zero, *the operator M is only positive semidefinite. If L and M are positive definite (semidefinite), then the system, or the eigenvalue problem, is positive definite (semidefinite).* Unless otherwise stated, *we will be concerned exclusively with systems for which M is positive definite.* Hence, the sign properties of the system are governed by the sign properties of the stiffness operator L.

For $v = u$, Eqs. (7.193) reduce to

$$[u, u]_P = \int_0^L \sum_{k=0}^{p} a_k \left(\frac{d^k u}{dx^k}\right)^2 dx + \sum_{\ell=0}^{p-1} b_\ell \left(\frac{d^\ell u}{dx^\ell}\right)^2 \Bigg|_0^L \tag{7.195a}$$

$$[u, u]_K = \int_0^L \sum_{k=0}^{q} e_k \left(\frac{d^k u}{dx^k}\right)^2 dx + \sum_{\ell=0}^{q-1} f_\ell \left(\frac{d^\ell u}{dx^\ell}\right)^2 \Bigg|_0^L \tag{7.195b}$$

and we note that $[u, u]_P$ and $[u, u]_K$ are measures of the potential and kinetic energy, respectively, which explains the terms of potential and kinetic energy inner products for $[u, v]_P$ and $[u, v]_K$ introduced earlier. Equations (7.195) can be used to define the *potential and kinetic energy norms*

$$\|u\|_P = [u, u]_P^{1/2}, \qquad \|u\|_K = [u, u]_K^{1/2} \tag{7.196a, b}$$

respectively.

Next, we introduce the sequence of approximations

$$u_n = \sum_{r=1}^{n} c_r \phi_r, \qquad n = 1, 2, \ldots \tag{7.197}$$

where $\phi_1, \phi_2, \ldots$ are given functions from an independent set. Then, if by choosing n sufficiently large,

$$\|u - u_n\|_P < \epsilon_P, \qquad \|u - u_n\|_K < \epsilon_K \tag{7.198a, b}$$

where ϵ_P and ϵ_K are arbitrarily small positive numbers, the set of functions ϕ_1, $\phi_2, \ldots$ is said to be *complete in energy*. Moreover, if

$$\lim_{n\to\infty} \|u - u_n\|_P = 0, \qquad \lim_{n\to\infty} \|u - u_n\|_K = 0 \tag{7.199a, b}$$

the sequence of approximations $u_1, u_2, \ldots$ is said to *converge in energy* to u.

Next, we wish to examine the properties of the eigenvalues and eigenfunctions. Assuming that the problem is self-adjoint and inserting two distinct solutions λ_r, w_r

and λ_s, w_s of the eigenvalue problem, Eqs. (7.190) and (7.191), into Eq. (7.190), we can write

$$Lw_r = \lambda_r M w_r, \qquad Lw_s = \lambda_s M w_s \tag{7.200a, b}$$

Multiplying Eq. (7.200a) by w_s and Eq. (7.200b) by w_r, subtracting the second from the first and integrating over domain D, we obtain

$$\int_D (w_s L w_r - w_r L w_s)\, dD = \int_D (\lambda_r w_s M w_r - \lambda_s w_r M w_s)\, dD \tag{7.201}$$

But, because the operators L and M are self-adjoint, we can use Eqs. (7.192) and (7.191b) to write

$$\begin{aligned}\int_D (w_s L w_r - w_r L w_s)\, dD &= \sum_{i=k+1}^{p} \int_S (w_r B_i w_s - w_s B_i w_r)\, dS \\ &= \sum_{i=k+1}^{p} \int_S (\lambda_s w_r C_i w_s - \lambda_r w_s C_i w_r)\, dS \end{aligned} \tag{7.202a}$$

and

$$\int_D w_s M w_r\, dD = \int_D w_r M w_s\, dD + \sum_{i=k+1}^{p} \int_S (w_r C_i w_s - w_s C_i w_r)\, dS \tag{7.202b}$$

so that, inserting Eqs. (7.202) into Eq. (7.201) and rearranging, we have

$$(\lambda_r - \lambda_s)\left(\int_D w_r M w_s\, dD + \sum_{i=k+1}^{p} \int_S w_r C_i w_s\, dS\right) = 0 \tag{7.203}$$

But, by assumption, the eigenvalues λ_r and λ_s are distinct. Hence, Eq. (7.203) can be satisfied if and only if

$$\int_D w_r M w_s\, dD + \sum_{i=k+1}^{p} \int_S w_r C_i w_s\, dS = 0, \qquad \lambda_r \neq \lambda_s, \qquad r, s = 1, 2, \ldots \tag{7.204}$$

Equation (7.204) represents the *orthogonality relation* for the eigenfunctions of distributed-parameter systems described by the eigenvalue problem given by Eqs. (7.190) and (7.191). Multiplying Eq. (7.200b) by w_r, integrating over D and using Eqs. (7.191b) and (7.204), it can be shown that the eigenfunctions satisfy *a second orthogonality relation*, namely,

$$\int_D w_r L w_s\, dD + \sum_{i=k+1}^{p} \int_S w_r B_i w_s\, dS = 0, \qquad \lambda_r \neq \lambda_s, \qquad r, s = 1, 2, \ldots \tag{7.205}$$

Clearly, the general orthogonality of the eigenfunctions solving the eigenvalue problem given by Eqs. (7.190) and (7.191) applies to self-adjoint systems alone. If an eigenvalue has multiplicity m, then there are exactly m eigenfunctions belonging to the repeated eigenvalue, and these eigenfunctions are generally not orthogonal to one another, although they are independent and orthogonal to the remaining eigenfunctions of the system. But, as pointed out in Sec. 7.5, independent functions can be orthogonalized by grouping them in proper linear combinations. Hence, *all the eigenfunctions of a self-adjoint system can be regarded as orthogonal, regardless of whether there are repeated eigenvalues or not.*

The eigenvalue problem, Eqs. (7.190) and (7.191), is homogeneous, so that only the shape of the eigenfunctions is unique, and the amplitude is arbitrary. This arbitrariness can be removed through normalization. A mathematically convenient normalization scheme is given by

$$\int_D w_r M w_r \, dD + \sum_{i=k+1}^{p} \int_S w_r C_i w_r \, dS = 1, \qquad r = 1, 2, \ldots \tag{7.206a}$$

which implies that

$$\int_r w_r L w_r \, dD + \sum_{i=k+1}^{p} \int_S w_r B_i w_r \, dS = \lambda_r, \qquad r = 1, 2, \ldots \tag{7.206b}$$

Then, Eqs. (7.204)–(7.206) can be combined into the *orthonormality relations*

$$\int_D w_r M w_s \, dD + \sum_{i=k+1}^{p} \int_S w_r C_i w_s \, dS = \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.207a}$$

$$\int_D w_r L w_s \, dD + \sum_{i=k+1}^{p} \int_S w_r B_i w_s \, dS = \lambda_r \, \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.207b}$$

In Sec. 7.5, we demonstrated that the eigenvalues and eigenfunctions of a self-adjoint system are real. We propose to prove here that the same is true for the more general eigenvalue problem. To this end, we consider a complex solution λ, w of the eigenvalue problem, Eqs. (7.190) and (7.191). Because all the operators are real, if λ, w are a complex solution of the eigenvalue problem, then the complex conjugates $\overline{\lambda}$, $\overline{w}$ must also be a solution, so that Eq. (7.190) yields

$$Lw = \lambda M w, \qquad L\overline{w} = \overline{\lambda} M \overline{w} \tag{7.208a, b}$$

Multiplying Eq. (7.208a) by $\overline{w}$ and Eq. (7.208b) by w, subtracting the second from the first and integrating over D, we obtain

$$\int_D (\overline{w} L w - w L \overline{w}) \, dD = \lambda \int_D \overline{w} M w \, dD - \overline{\lambda} \int_D w M \overline{w} \, dD \tag{7.209}$$

Letting $v = w$ and $u = \overline{w}$ in Eqs. (7.191b) and (7.192) and inserting the results into Eq. (7.209), we obtain after some manipulations

$$(\lambda - \overline{\lambda}) \left(\int_D \overline{w} M w \, dD + \sum_{i=k+1}^{p} \int_S \overline{w} C_i w \, dS \right) = 0 \tag{7.210}$$

Recalling Eqs. (7.99) and considering Eq. (7.192b), we conclude that the term in the second parentheses in Eq. (7.210) is real and positive, so that the only alternative is

$$\lambda - \bar{\lambda} = \alpha + i\beta - (\alpha - i\beta) = 2i\beta = 0 \tag{7.211}$$

Hence, as in Sec. 7.5, we conclude that *the eigenvalues of a self-adjoint system are real.* As a corollary, *the eigenfunctions of a self-adjoint system are real.* Moreover, considering inequality (7.194a), we conclude from Eq. (7.206b) that, *if the operator L is positive definite (semidefinite), all the eigenvalues are positive (nonnegative).*

Finally, we wish to extend the *expansion theorem for self-adjoint systems* of Sec. 7.5 as follows: *Every function w with continuous Lw and Mw and satisfying the boundary conditions of the system can be expanded in an absolutely and uniformly convergent series in the eigenfunctions in the form*

$$w = \sum_{r=1}^{\infty} c_r w_r \tag{7.212}$$

where the coefficients c_r are such that

$$c_r = \int_D w_r M w \, dD + \sum_{i=k+1}^{p} \int_S w_r C_i w \, dS \tag{7.213a}$$

$$\lambda_r c_r = \int_D w_r L w \, dD + \sum_{i=k+1}^{p} \int_S w_r B_i w \, dS \tag{7.213b}$$

We should note here that Eqs. (7.213) are based on the more general orthonormality relations, Eqs. (7.207). The expansion theorem just presented forms the basis for a modal analysis for self-adjoint systems with stiffness and mass operators L and M, respectively, and with boundary conditions depending on the eigenvalue λ. Although the expansion theorem, Eqs. (7.212) and (7.213), seems intimidating compared to the expansion theorem of Sec. 7.5, the process of using modal analysis to derive the modal equations remains essentially the same. Derivation of the system response by modal analysis is discussed later in this chapter.

Example 7.7

Derive the eigenvalue problem for the shaft in torsional vibration of Example 7.5 and show how it fits the formulation given by Eqs. (7.190) and (7.191).

From Example 7.5, the free vibration of the shaft, obtained by letting the distributed torque m be equal to zero, is described by the partial differential equation

$$\frac{\partial}{\partial x}\left[GJ(x)\frac{\partial \theta(x,t)}{\partial x}\right] = I(x)\frac{\partial^2 \theta(x,t)}{\partial t^2}, \qquad 0 < x < L \tag{a}$$

and the boundary conditions

$$\theta(0,t) = 0; \; I_D\frac{\partial^2 \theta(x,t)}{\partial t^2} + GJ(x)\frac{\partial \theta(x,t)}{\partial x} = 0, \qquad x = L \tag{b}$$

To derive the eigenvalue problem, we assume a solution in the form

$$\theta(x, t) = \Theta(x)F(t) \tag{c}$$

Introducing Eq. (c) into Eq. (a) and separating variables, we can write

$$\frac{1}{I(x)\Theta(x)}\frac{d}{dx}\left[GJ(x)\frac{d\Theta(x)}{dx}\right] = \frac{1}{F(t)}\frac{d^2F(t)}{dt^2}, \qquad 0 < x < L \tag{d}$$

Following the developments of Sec. 7.4, it can be shown that the function $F(t)$ is harmonic, and it satisfies

$$\frac{d^2F(t)}{dt^2} = -\lambda F(t), \qquad \lambda = \omega^2 \tag{e}$$

where ω is the frequency of oscillation, so that the left side of Eq. (d) yields the differential equation

$$-\frac{d}{dx}\left[GJ(x)\frac{d\Theta(x)}{dx}\right] = \lambda I(x)\Theta(x), \qquad 0 < x < L \tag{f}$$

Moreover, inserting Eqs. (c) and (e) into Eqs. (b), we obtain the boundary conditions

$$\Theta(0) = 0; \; GJ(x)\frac{d\Theta(x)}{dx} = \lambda I_D\Theta(x), \qquad x = L \tag{g}$$

Equations (f) and (g) constitute the desired differential eigenvalue problem.

Contrasting Eqs. (7.190) and (f) on the one hand and Eqs. (7.191) and (g) on the other hand, we conclude that the eigenvalue problem does fit the mold. The various operators can be identified as follows:

$$\begin{aligned} L &= -\frac{d}{dx}\left[GJ(x)\frac{d}{dx}\right], \qquad p = 1; \; M = I(x), \qquad q = 0 \\ B_1 &= 1, \; C_1 = 0 \quad \text{at } x = 0, \; k = 1 \\ B_1 &= GJ(x)\frac{d}{dx}, \; C_1 = I_D \quad \text{at } x = L, \qquad k = 0 \end{aligned} \tag{h}$$

A solution of the eigenvalue problem given by Eqs. (f) and (g) for a uniform shaft is presented in Sec. 7.10.

Example 7.8

Derive the eigenvalue problem for the beam in bending of Example 7.6 and show the relation with the formulation given by Eqs. (7.190) and (7.191). Assume that the beam is clamped at $x = 0$ and free at $x = L$.

Letting $f = 0$, the free vibration problem can be obtained from Example 7.6 in the form of the partial differential equation

$$\begin{aligned} &\frac{\partial}{\partial x}\left[P(x)\frac{\partial w(x,t)}{\partial x}\right] - \frac{\partial^2}{\partial x^2}\left[EI(x)\frac{\partial^2 w(x,t)}{\partial x^2}\right] - m(x)\frac{\partial^2 w(x,t)}{\partial t^2} \\ &\quad + \frac{\partial}{\partial x}\left[J(x)\frac{\partial^3 w(x,t)}{\partial x \partial t^2}\right] = 0, \qquad 0 < x < L \end{aligned} \tag{a}$$

and the boundary conditions

$$\begin{aligned} & w(x,t) = 0, \qquad \frac{\partial w(x,t)}{\partial x} = 0, \qquad x = 0 \\ & EI(x)\frac{\partial^2 w(x,t)}{\partial x^2} = 0,\ P(x)\frac{\partial w(x,t)}{\partial x} - \frac{\partial}{\partial x}\left[EI(x)\frac{\partial^2 w(x,t)}{\partial x^2}\right] \\ & \qquad + J(x)\frac{\partial^3 w(x,t)}{\partial x \partial t^2} = 0, \qquad x = L \end{aligned} \tag{b}$$

To derive the eigenvalue problem, we assume that the solution is separable in x and t, or

$$w(x,t) = W(x)F(t) \tag{c}$$

Inserting Eq. (c) into Eq. (a) and following the usual steps, we obtain

$$\frac{\dfrac{d^2}{dx^2}\left[EI(x)\dfrac{d^2W(x)}{dx^2}\right] - \dfrac{d}{dx}\left[P(x)\dfrac{dW(x)}{dx}\right]}{m(x)W(x) - \dfrac{d}{dx}\left[J(x)\dfrac{dW(x)}{dx}\right]} = -\frac{1}{F(t)}\frac{d^2F(t)}{dt^2} \tag{d}$$

Then, using the standard argument, we let both sides of Eq. (d) be equal to $\lambda = \omega^2$, so that $F(t)$ is harmonic with the frequency ω. Moreover, we obtain the differential equation

$$\begin{aligned} & \frac{d^2}{dx^2}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] - \frac{d}{dx}\left[P(x)\frac{dW(x)}{dx}\right] \\ & \qquad = \lambda\left\{m(x)W(x) - \frac{d}{dx}\left[J(x)\frac{dW(x)}{dx}\right]\right\}, \qquad 0 < x < L \end{aligned} \tag{e}$$

Similarly, introducing Eq. (c) into Eqs. (b), in conjunction with $\ddot{F}(t) = -\lambda F(t)$, and dividing through by $F(t)$, we obtain the boundary conditions

$$\begin{aligned} & W(x) = 0, \qquad \frac{dW(x)}{dx} = 0, \qquad x = 0 \\ & EI(x)\frac{d^2W(x)}{dx^2} = 0, \\ & -\frac{d}{dx}\left[EI(x)\frac{d^2W(x)}{dx^2}\right] + P(x)\frac{dW(x)}{dx} = \lambda J(x)W(x), \qquad x = L \end{aligned} \tag{f}$$

Equations (e) and (f) represent the differential eigenvalue problem for the system at hand.

Comparing Eqs. (e) and (f) with Eqs. (7.190) and (7.191), respectively, we can identify the various operators as follows:

$$\begin{aligned}
&L = \frac{d^2}{dx^2}\left[EI(x)\frac{d^2}{dx^2}\right] - \frac{d}{dx}\left[P(x)\frac{d}{dx}\right], \qquad p = 2\\
&M = m(x) - \frac{d}{dx}\left[J(x)\frac{d}{dx}\right], \qquad q = 1\\
&B_1 = 1, \quad B_2 = \frac{d}{dx}, \qquad C_1 = C_2 = 0, \quad k = 2, \quad x = 0, \\
&B_1 = EI(x)\frac{d^2}{dx^2}, \quad B_2 = -\frac{d}{dx}\left[EI(x)\frac{d^2}{dx^2}\right] + P(x)\frac{d}{dx},\\
&C_1 = 0, \quad C_2 = J(x), \quad k = 1, \quad x = L
\end{aligned} \tag{g}$$

Clearly, the differential eigenvalue problem, Eqs. (e) and (f), does fit the general formulation given by Eqs. (7.190) and (7.191).

Closed-form solutions to the eigenvalue problem given by Eqs. (e) and (f) do not exist.

7.10 SYSTEMS WITH BOUNDARY CONDITIONS DEPENDING ON THE EIGENVALUE

Let us return to the system shown in Fig. 7.12 and recall that the boundary-value problem was derived in Example 7.5 and the eigenvalue problem in Example 7.7. In this section, we consider the solution of the eigenvalue problem.

In the case in which the shaft is uniform, $GJ(x) = GJ =$ constant, $I(x) = I =$ constant, the differential equation, Eq. (f) of Example 7.7, reduces to the familiar form

$$\frac{d^2\Theta(x)}{dx^2} + \beta^2\Theta(x) = 0, \qquad \beta^2 = \frac{\lambda I}{GJ} = \frac{\omega^2 I}{GJ}, \qquad 0 < x < L \tag{7.214}$$

Moreover, the boundary conditions, Eqs. (g) of Example 7.7, become

$$\Theta(0) = 0, \qquad \left.\frac{d\Theta(x)}{dx}\right|_{x=L} = \frac{\lambda I_D}{GJ}\Theta(L) = \frac{\beta^2 I_D}{I}\Theta(L) \tag{7.215a, b}$$

so that boundary condition (7.215b) depends on the eigenvalue β. As in Sec. 7.6, the solution of Eq. (7.214) is

$$\Theta(x) = C_1 \sin\beta x + C_2 \cos\beta x \tag{7.216}$$

and boundary condition (7.215a) yields $C_2 = 0$. On the other hand, boundary condition (7.215b) leads to the characteristic equation

$$\tan\beta L = \frac{IL}{I_D}\frac{1}{\beta L} \tag{7.217}$$

which must be solved numerically for the eigenvalues $\beta_r L$ $(r = 1, 2, \ldots)$. If some accuracy can be sacrificed, then the solution can be obtained graphically, as shown in Fig. 7.13, in which the three lowest eigenvalues corresponding to $IL/I_D = 1$ were obtained. The natural modes are given by

$$\Theta_r(x) = A_r \sin \beta_r x, \qquad r = 1, 2, \ldots \tag{7.218}$$

and they are orthogonal. Using Eqs. (7.204) and (7.205) in conjunction with boundary conditions (7.215), the orthogonality relations can be shown to be

$$\int_0^L I\Theta_r(x)\Theta_s(x)\,dx + I_D\Theta_r(L)\Theta_s(L) = 0, \quad r, s = 1, 2, \ldots; r \neq s \tag{7.219a}$$

$$\int_0^L GJ\Theta_r'(x)\Theta_s'(x)dx = 0, \qquad r, s = 1, 2, \ldots;\ r \neq s \tag{7.219b}$$

The natural frequencies are related to the eigenvalues by

$$\omega_r = \beta_r L\sqrt{\frac{GJ}{IL^2}}, \qquad r = 1, 2, \ldots \tag{7.220}$$

The first three natural modes and natural frequencies are displayed in Fig. 7.14.

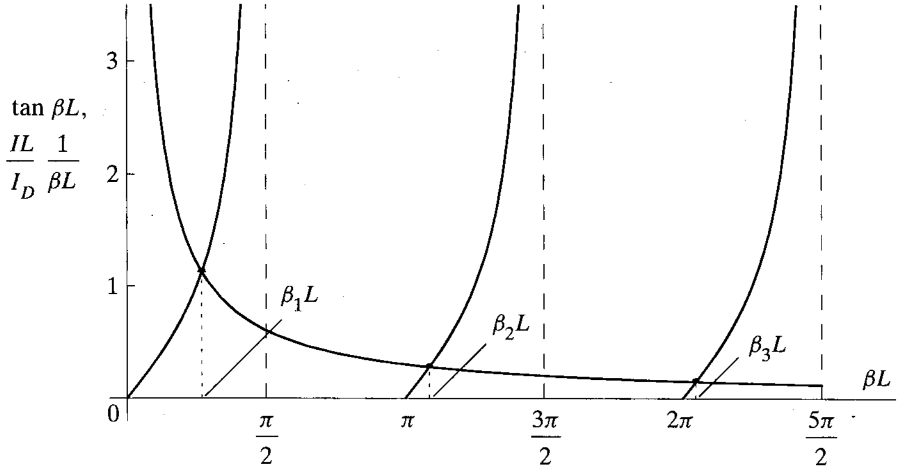


Figure 7.13 Graphical solution of the characteristic equation, Eq. (7.217)

From Eq. (7.217), as well as from Fig. 7.13, we observe that, as the eigenvalues $\beta_r L$ increase without bound, they tend to become integer multiples of π. Specifically,

$$\lim_{r\to\infty} \beta_r L = (r - 1)\pi \tag{7.221}$$

Inserting these values into Eq. (7.218), we conclude that the very high modes have nodes at $x = L$, which implies that the end disk is at rest for these modes, so that the end $x = L$ acts as if it were clamped.

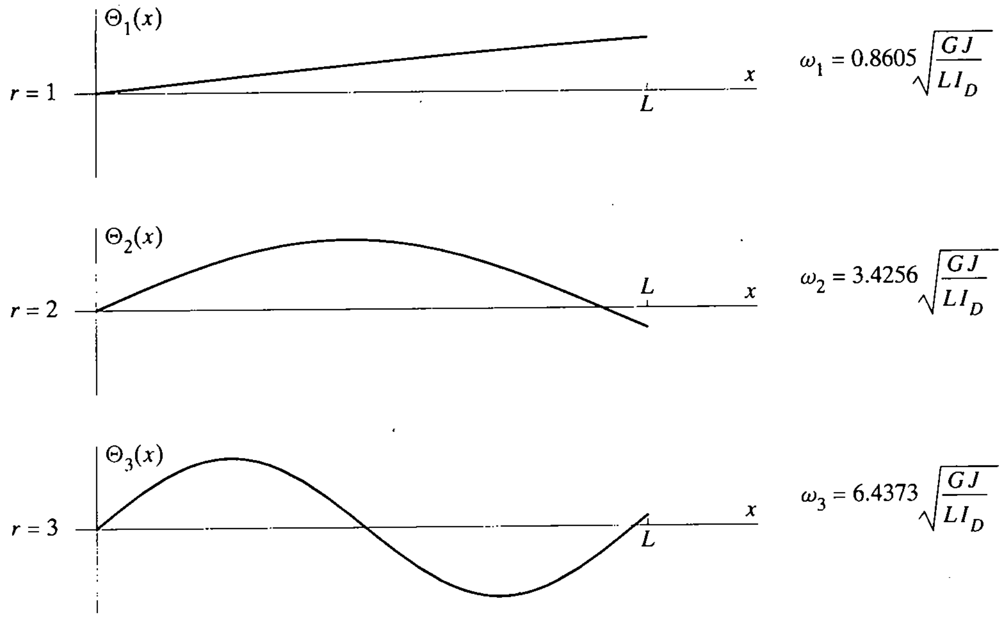

$\Theta_1(x)$
$r = 1$
x
L
$\omega_1 = 0.8605\sqrt{\frac{GJ}{LI_D}}$
$\Theta_2(x)$
$r = 2$
L
x
$\omega_2 = 3.4256\sqrt{\frac{GJ}{LI_D}}$
$\Theta_3(x)$
$r = 3$
x
L
$\omega_3 = 6.4373\sqrt{\frac{GJ}{LI_D}}$

Figure 7.14 The first three modes of vibration of a shaft in torsion clamped at $x = 0$ and with a disk at $x = L$

7.11 TIMOSHENKO BEAM

In Sec. 7.2, we derived the boundary-value problem for the simplest model of a beam in bending vibration, namely, the Euler-Bernoulli model, which is based on the elementary beam theory. Then, in Sec. 7.6 we refined the model by including the rotatory inertia effects. The model of Sec. 7.6 can be further refined by considering the shear deformation effects. The inclusion of the shear deformation presents us with a problem not encountered before. Indeed, because in this case the slope of the deflection curve is not equal to the rotation of the beam cross section, we are faced with the problem of two dependent variables. The model of a beam including both rotatory inertia and shear deformation effects is commonly referred to as a *Timoshenko beam*.

Our objective is to derive the boundary-value problem for the nonuniform beam in bending shown in Fig. 7.15a. To this end, we consider the differential element of Fig. 7.15b and denote the mass per unit length at any point x by $m(x)$, the cross-sectional area by $A(x)$ and the area and mass moments of inertia about an axis normal to the plane of motion and passing through point C by $I(x)$ and $J(x)$, respectively, where C represents the mass center of the differential element. From Fig. 7.15b, the total deflection $w(x, t)$ of the beam consists of two parts, one caused by bending and one by shear, so that the slope of the deflection curve at point x can be written in the form

$$\frac{\partial w(x, t)}{\partial x} = \psi(x, t) + \beta(x, t) \tag{7.222}$$

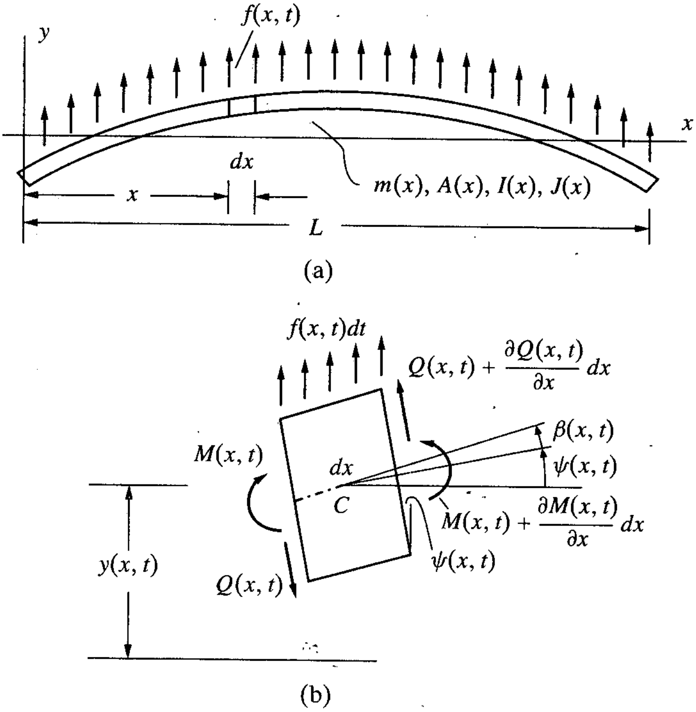


Figure 7.15 **(a)** Timoshenko beam **(b)** Timoshenko beam differential element

where $\psi(x, t)$ is the angle of rotation due to bending and $\beta(x, t)$ is the angle of distortion due to shear. As usual, the linear deflection and angular deflection are assumed small.

The relation between the bending moment and the bending deformation is

$$M(x, t) = EI(x)\frac{\partial \psi(x, t)}{\partial x} \tag{7.223}$$

and the relation between the shearing force and shearing deformation is given by

$$Q(x, t) = k'GA(x)\beta(x, t) \tag{7.224}$$

in which G is the shear modulus and k' is a numerical factor depending on the shape of the cross section. Because of shear alone, the element undergoes distortion but no rotation.

To formulate the boundary-value problem, we make use of the extended Hamilton's principle, Eq. (7.4), which requires the kinetic energy, potential energy and virtual work. The kinetic energy is due to translation and rotation and has the form

$$T(t) = \frac{1}{2}\int_0^L m(x)\left[\frac{\partial w(x, t)}{\partial t}\right]^2 dx + \frac{1}{2}\int_0^L J(x)\left[\frac{\partial \psi(x, t)}{\partial t}\right]^2 dx \tag{7.225}$$

where the mass moment of inertia density $J(x)$ is related to the area moment of inertia $I(x)$ by

$$J(x) = \rho I(x) = \frac{m(x)}{A(x)} I(x) = k^2(x)m(x) \tag{7.226}$$

in which ρ is the mass density and $k(x)$ is the radius of gyration about the neutral axis. The variation of the kinetic energy can be readily written as

$$\delta T = \int_0^L m\frac{\partial w}{\partial t}\delta\left(\frac{\partial w}{\partial t}\right) dx + \int_0^L k^2 m\frac{\partial \psi}{\partial t}\delta\left(\frac{\partial \psi}{\partial t}\right) dx \tag{7.227}$$

The potential energy has the expression

$$\begin{aligned} V(t) &= \frac{1}{2}\int_0^L M(x,t)\frac{\partial \psi(x,t)}{\partial x} dx + \frac{1}{2}\int_0^L Q(x,t)\beta(x,t)\, dx \\ &= \frac{1}{2}\int_0^L EI(x)\left[\frac{\partial \psi(x,t)}{\partial x}\right]^2 dx + \frac{1}{2}\int_0^L k'GA(x)\beta^2(x,t)\, dx \end{aligned} \tag{7.228}$$

so that the variation of the potential energy is simply

$$\begin{aligned} \delta V &= \int_0^L EI\frac{\partial \psi}{\partial x}\delta\left(\frac{\partial \psi}{\partial x}\right) dx + \int_0^L k'GA\beta\delta\beta dx \\ &= \int_0^L EI\frac{\partial \psi}{\partial x}\delta\left(\frac{\partial \psi}{\partial x}\right) dx + \int_0^L k'GA\left(\frac{\partial w}{\partial x} - \psi\right)\delta\left(\frac{\partial w}{\partial x} - \psi\right) dx \end{aligned} \tag{7.229}$$

The virtual work due to nonconservative forces is given by

$$\delta W_{nc}(t) = \int_0^L f(x,t)\,\delta w(x,t)\, dx \tag{7.230}$$

where f is the force density.

Introducing Eqs. (7.227), (7.229) and (7.230) into the extended Hamilton's principle, Eq. (7.4), we have

$$\begin{gathered} \int_{t_1}^{t_2} (\delta T - \delta V + \delta W_{nc})\, dt = \int_{t_1}^{t_2} \left\{ \int_0^L \left[m\frac{\partial w}{\partial t}\delta\left(\frac{\partial w}{\partial t}\right) + k^2 m\frac{\partial \psi}{\partial t}\delta\left(\frac{\partial \psi}{\partial t}\right) \right.\right. \\ \left.\left. - EI\frac{\partial \psi}{\partial x}\delta\left(\frac{\partial \psi}{\partial x}\right) - k'GA\left(\frac{\partial w}{\partial x} - \psi\right)\delta\left(\frac{\partial w}{\partial x} - \psi\right) + f\,\delta w\right] dx \right\} dt = 0, \\ \delta w = 0, \quad \delta\psi = 0, \quad t = t_1, t_2 \end{gathered} \tag{7.231}$$

and we note that we have two dependent variables, w and ψ. We carry out the operations involved in Eq. (7.231) term by term. Recalling that the order of the integrations with respect to x and t is interchangeable and that the variation and differentiation operators are commutative, we can perform the following integration by parts with respect to time:

$$\begin{aligned}\int_{t_1}^{t_2} m\frac{\partial w}{\partial t}\delta\left(\frac{\partial w}{\partial t}\right)dt &= \int_{t_1}^{t_2} m\frac{\partial w}{\partial t}\frac{\partial}{\partial t}\delta w\,dt \\ &= m\frac{\partial w}{\partial t}\delta w\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2}\frac{\partial}{\partial t}\left(m\frac{\partial w}{\partial t}\right)\delta w\,dt \\ &= -\int_{t_1}^{t_2} m\frac{\partial^2 w}{\partial t^2}\delta w\,dt \end{aligned} \tag{7.232}$$

where we took into account that δw vanishes at $t = t_1$ and $t = t_2$. In a similar fashion, we obtain

$$\int_{t_1}^{t_2} k^2 m\frac{\partial\psi}{\partial t}\delta\left(\frac{\partial\psi}{\partial t}\right)dt = -\int_{t_1}^{t_2} k^2 m\frac{\partial^2\psi}{\partial t^2}\delta\psi dt \tag{7.233}$$

On the other hand, integrations over the spatial variable yield

$$\begin{aligned}\int_0^L EI\frac{\partial\psi}{\partial x}\delta\left(\frac{\partial\psi}{\partial x}\right)dx &= \int_0^L EI\frac{\partial\psi}{\partial x}\frac{\partial}{\partial x}\delta\psi\,dx \\ &= \left(EI\frac{\partial\psi}{\partial x}\right)\delta\psi\Big|_0^L - \int_0^L\frac{\partial}{\partial x}\left(EI\frac{\partial\psi}{\partial x}\right)\delta\psi\,dx\end{aligned} \tag{7.234a}$$

$$\begin{aligned}&\int_0^L k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\delta\left(\frac{\partial w}{\partial x}-\psi\right)dx \\ &= \int_0^L k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\left(\frac{\partial}{\partial x}\delta w-\delta\psi\right)dx \\ &= \left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]\delta w\Big|_0^L \\ &\quad - \int_0^L\left\{\frac{\partial}{\partial x}\left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]\delta w + k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\delta\psi\right\}dx\end{aligned} \tag{7.234b}$$

Inserting Eqs. (7.232)–(7.234) into Eq. (7.231) and rearranging, we obtain

$$\begin{aligned}&\int_{t_1}^{t_2}(\delta T - \delta V + \delta W_{nc})\,dt \\ &= \int_{t_1}^{t_2}\left[\int_0^L\left\langle\left\{\frac{\partial}{\partial x}\left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right] - m\frac{\partial^2 w}{\partial t^2} + f\right\}\delta w\right.\right.\end{aligned}$$

$$+\left\{\left[\frac{\partial}{\partial x}\left(EI\frac{\partial\psi}{\partial x}\right)+k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]-k^2m\frac{\partial^2\psi}{\partial t^2}\right\}\delta\psi\Big\rangle dx$$

$$-\left(EI\frac{\partial\psi}{\partial x}\right)\delta\psi\Big|_0^L-\left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]\delta w\Big|_0^L\Bigg]dt=0 \qquad (7.235)$$

The virtual displacements $\delta\psi$ and δw are arbitrary and independent, so that they can be taken equal to zero at $x = 0$ and $x = L$ and arbitrary for $0 < x < L$. Hence, we must have

$$\frac{\partial}{\partial x}\left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]-m\frac{\partial^2 w}{\partial t^2}+f=0, \qquad 0<x<L \qquad (7.236a)$$

$$\frac{\partial}{\partial x}\left(EI\frac{\partial\psi}{\partial x}\right)+k'GA\left(\frac{\partial w}{\partial x}-\psi\right)-k^2m\frac{\partial^2\psi}{\partial t^2}=0, \quad 0<x<L \quad (7.236b)$$

In addition, if we write

$$\left(EI\frac{\partial\psi}{\partial x}\right)\delta\psi\Big|_0^L=0 \qquad (7.237a)$$

$$\left[k'GA\left(\frac{\partial w}{\partial x}-\psi\right)\right]\delta w\Big|_0^L=0 \qquad (7.237b)$$

we take into account the possibility that either $EI\,(\partial\psi/\partial x)$ or $\delta\psi$ on the one hand, and either $k'GA\,[(\partial w/\partial x)-\psi]$ or δw on the other hand vanishes at the ends $x = 0$ and $x = L$. Equations (7.236) are the differential equations of motion and Eqs. (7.237) represent the boundary conditions. The boundary-value problem consists of the differential equations, Eqs. (7.236) and two boundary conditions at each end to be chosen from Eqs. (7.237).

For a beam *clamped at both ends*, the deflection and rotation are zero, or

$$w(0,t)=0, \qquad \psi(0,t)=0 \qquad (7.238a, b)$$

$$w(L,t)=0, \qquad \psi(L,t)=0 \qquad (7.238c, d)$$

and note that it is the rotation that is zero and not the slope. All boundary conditions are geometric.

In the case of a *simply-supported* beam, i.e., a beam *pinned at both ends*, the boundary conditions are

$$w(0,t)=0, \qquad M(0,t)=EI\frac{\partial\psi}{\partial x}\Big|_{x=0}=0 \qquad (7.239a, b)$$

$$w(L,t)=0, \qquad M(L,t)=EI\frac{\partial\psi}{\partial x}\Big|_{x=L}=0 \qquad (7.239c, d)$$

so that there is one geometric and one natural boundary condition at each end.

For a beam cantilevered at $x = 0$, the boundary conditions at the clamped end are

$$w(0, t) = 0, \qquad \psi(0, t) = 0 \tag{7.240a, b}$$

At the free end neither the deflection nor the rotation is zero, so that we must have

$$M(L, t) = EI\frac{\partial \psi}{\partial x}\bigg|_{x=L} = 0, \qquad Q(L, t) = \left[k'GA\left(\frac{\partial w}{\partial x} - \psi\right)\right]_{x=L} = 0 \tag{7.240c,d}$$

which reflects the fact that both the bending moment and the shearing force vanish at a free end. Hence, at the clamped end we have two geometric boundary conditions and at the free end we have two natural boundary conditions.

Finally, in this case of a *free-free* beam, the boundary conditions are

$$M(0, t) = EI\frac{\partial \psi}{\partial x}\bigg|_{x=0} = 0, \qquad Q(0, t) = \left[k'GA\left(\frac{\partial w}{\partial x} - \psi\right)\right]_{x=0} = 0 \tag{7.241a, b}$$

$$M(L, t) = EI\frac{\partial \psi}{\partial x}\bigg|_{x=L} = 0, \qquad Q(L, t) = \left[k'GA\left(\frac{\partial w}{\partial x} - \psi\right)\right]_{x=L} = 0 \tag{7.241c, d}$$

and they are all natural.

The interesting part about the new formulation is that the mass density is no longer an operator, as in Example 7.8, but a mere function. Moreover, boundary conditions (7.240d), (7.241b) and (7.241d) do not depend on the acceleration, in contrast with the case in which the shear deformation is absent, as can be concluded from Eq. (d) of Example 7.6. Both differences are due to the fact that the rotation is no longer equal to the spatial derivative of the displacement. The simplification gained in the mass density expression and the boundary condition involving the shearing force is balanced by the complication arising from the fact that now there are two dependent variables.

Next, we wish to derive the eigenvalue problem. In view of our past experience, we assume that $f = 0$ and that the solution of the boundary-value problem is separable in x and t, or

$$w(x, t) = W(x)F(t), \qquad \psi(x, t) = \Psi(x)F(t) \tag{7.242}$$

where $F(t)$ is harmonic and it satisfies

$$\ddot{F}(t) = -\lambda F(t), \qquad \lambda = \omega^2 \tag{7.243}$$

Introducing Eqs. (7.242) and (7.243) into Eqs. (7.236) with $f = 0$, we obtain the ordinary differential equations

$$-\frac{d}{dx}\left[k'GA\left(\frac{dW}{dx} - \Psi\right)\right] = \lambda m W, \qquad 0 < x < L \tag{7.244a}$$

$$-\left\{\frac{d}{dx}\left(EI\frac{d\Psi}{dx}\right)+k'GA\left(\frac{dW}{dx}-\Psi\right)\right\}=\lambda k^2 m\Psi, \qquad 0<x<L \tag{7.244b}$$

The boundary conditions transform accordingly.

The eigenvalue problem is defined by two differential equations in terms of two dependent variables, instead of one differential equation in one variable, so that the traditional ways of checking for self-adjointness and positive definiteness do not apply in the case of a Timoshenko beam. In the following, we define new criteria. To this end, we introduce the displacement vector

$$\mathbf{y}(x)=[W(x)\ \ \Psi(x)]^T \tag{7.245}$$

as well as the stiffness and mass operator matrices

$$\mathcal{L}=-\begin{bmatrix}\frac{d}{dx}\left(k'GA\frac{d}{dx}\right) & -\frac{d}{dx}(k'GA\,\cdot) \\ k'GA\frac{d}{dx} & \frac{d}{dx}\left(EI\frac{d}{dx}\right)-k'GA\end{bmatrix}, \qquad \mathcal{M}=\begin{bmatrix} m & 0 \\ 0 & k^2m\end{bmatrix} \tag{7.246a,b}$$

where the dot indicates the implied position of Ψ, and write Eqs. (7.244) in the operator matrix form

$$\mathcal{L}\mathbf{y}=\lambda\mathcal{M}\mathbf{y} \tag{7.247}$$

Then, by analogy with the scalar definitions, the problem is self-adjoint if for any two vectors **u** and **v** of comparison functions

$$\int_0^L \mathbf{u}^T\mathcal{L}\mathbf{v}dx=\int_0^L \mathbf{v}^T\mathcal{L}\mathbf{u}dx \tag{7.248a}$$

$$\int_0^L \mathbf{u}^T\mathcal{M}\mathbf{v}dx=\int_0^L \mathbf{v}^T\mathcal{M}\mathbf{u}dx \tag{7.248b}$$

Moreover, the problem is positive definite if

$$\int_0^L \mathbf{u}^T\mathcal{L}\mathbf{u}dx\geq 0, \qquad \int_0^L \mathbf{u}^T\mathcal{M}\mathbf{u}dx\geq 0 \tag{7.249a, b}$$

and the equality sign holds true if and only if $\mathbf{u}\equiv\mathbf{0}$, and it is positive semidefinite if the equality sign holds true for some $\mathbf{u}\neq\mathbf{0}$. Because the mass operator matrix $\mathcal{M}$ is self-adjoint and positive definite by definition, the system self-adjointness and positive definiteness depend on the stiffness operator $\mathcal{L}$.

To check for self-adjointness, we insert Eq. (7.246a) into the left side of Eq. (7.248a), carry out suitable integrations by parts and obtain

$$\int_0^L \mathbf{u}^T\mathcal{L}\mathbf{v}dx=-\int_0^L \mathbf{u}^T\begin{bmatrix}\frac{d}{dx}\left(k'GA\frac{d}{dx}\right) & -\frac{d}{dx}(k'GA\,\cdot) \\ k'GA\frac{d}{dx} & \frac{d}{dx}\left(EI\frac{d}{dx}\right)-k'GA\end{bmatrix}\mathbf{v}dx$$

$$= -\mathbf{u}^T \begin{bmatrix} k'GA\dfrac{d}{dx} & -k'GA \\ 0 & EI\dfrac{d}{dx} \end{bmatrix} \mathbf{v}\Bigg|_0^L$$

$$+ \int_0^L \left(\frac{d\mathbf{u}^T}{dx} \begin{bmatrix} k'GA & 0 \\ 0 & EI \end{bmatrix} \frac{d\mathbf{v}}{dx} - \frac{d\mathbf{u}^T}{dx} \begin{bmatrix} 0 & k'GA \\ 0 & 0 \end{bmatrix} \mathbf{v} \right.$$

$$\left. -\mathbf{u}^T \begin{bmatrix} 0 & 0 \\ k'GA & 0 \end{bmatrix} \frac{d\mathbf{v}}{dx} + \mathbf{u}^T \begin{bmatrix} 0 & 0 \\ 0 & k'GA \end{bmatrix} \mathbf{v} \right) dx \qquad (7.250)$$

We observe that the integral on the right side of Eq. (7.250) is symmetric in **u** and **v**. Hence, *all systems for which the boundary term is zero are self-adjoint.* This is certainly the case with the systems with the boundary conditions given by Eqs. (7.238)–(7.241).

Before we proceed with the check for positive definiteness, we should state that the concept applies only to self-adjoint systems. Hence, assuming that the boundary term is zero and letting $\mathbf{v} = \mathbf{u}$ in Eq. (7.250), we have

$$\int_0^L \mathbf{u}^T \mathcal{L}\mathbf{u} dx = \int_0^L \left(\frac{d\mathbf{u}^T}{dx} \begin{bmatrix} k'GA & 0 \\ 0 & EI \end{bmatrix} \frac{d\mathbf{u}}{dx} - \frac{d\mathbf{u}^T}{dx} \begin{bmatrix} 0 & k'GA \\ 0 & 0 \end{bmatrix} \mathbf{u} \right.$$

$$\left. -\mathbf{u}^T \begin{bmatrix} 0 & 0 \\ k'GA & 0 \end{bmatrix} \frac{d\mathbf{u}}{dx} + \mathbf{u}^T \begin{bmatrix} 0 & 0 \\ 0 & k'GA \end{bmatrix} \mathbf{u} \right) dx$$

$$= \int_0^L \left[k'GA \left(\frac{du_1}{dx} - u_2 \right)^2 + EI \left(\frac{du_2}{dx} \right)^2 \right] dx \geq 0 \qquad (7.251)$$

where u_1 and u_2 are the components of **u**. It is easy to verify that the expression can be zero for the nontrivial case $u_1 =$ constant, $u_2 = 0$. But, when one of the ends is clamped or pinned, such as in the case of the systems with the boundary conditions given by Eqs. (7.238)–(7.240), this constant must be zero. It follows that in the three cases covered by boundary conditions (7.238)–(7.240), the operator $\mathcal{L}$ is positive definite, so that *the system is positive definite.* On the other hand, for a free-free beam with the boundary conditions given by Eqs. (7.241), $u_1 =$ constant, $u_2 = 0$ is a possible solution of the eigenvalue problem so that the operator $\mathcal{L}$ is only positive semidefinite, from which it follows that *the system is only positive semidefinite.* Consistent with this, the solution $u_1 =$ constant, $u_2 = 0$ *represents a rigid-body mode belonging to a zero eigenvalue.*

7.12 VIBRATION OF MEMBRANES

All distributed systems considered until now were one-dimensional, which implies that the description of their motion requires a single spatial variable. At this point, we turn our attention to two-dimensional systems whose motion is described in terms of two spatial coordinates. We confine ourselves to the case in which the domain D is planar, with the motion being measured normal to the nominal plane, and the boundary S consists of one or two nonintersecting curves. Two-dimensional problems introduce a new element into the boundary-value problem, namely, the shape of the boundary S. In two-dimensional problems, there are several choices of coordinates for describing the motion, such as rectangular, polar, elliptical, etc. The choice is generally not arbitrary but dictated by the shape of the boundary, because the boundary conditions for the most part involve derivatives along the normal direction n or along the tangent s to the boundary (Fig. 7.16). Hence, it is only natural to choose rectangular coordinates if the boundary S is a rectangle, polar coordinates if S is a circle, elliptical coordinates if S is an ellipse, etc. The question is not so clear when S has an irregular shape, in which case no closed-form solution can be expected. In this case, the choice of coordinates depends on the method used to produce an approximate solution.

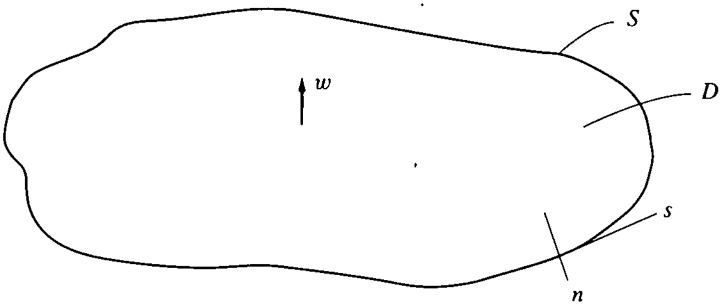


Figure 7.16 Two-dimensional distributed system

The simplest two-dimensional problem in vibrations is that of a membrane. Indeed, the membrane can be regarded as the two-dimensional counterpart of the string. The boundary-value problem can be obtained conveniently by means of the extended Hamilton's principle, Eq. (7.4). It is relatively easy to carry out the derivation in terms of rectangular coordinates. However, we opt for an approach valid for all types of coordinates. To this end, we consider a membrane fixed or free over a portion S_1 of the boundary S and supported by a distributed spring over the remaining portion S_2 and write the potential energy in the form

$$V = \frac{1}{2}\int_D T\nabla w \cdot \nabla w \, dD + \frac{1}{2}\int_{S_2} kw^2 \, dS \tag{7.252}$$

where w is the transverse displacement, ∇ a vector operator signifying the gradient, T the tension and k the distributed spring constant. For simplicity we assume that *the tension is constant.* The kinetic energy has the expression

$$T = \frac{1}{2}\int_D \rho \dot{w}^2 \, dD \tag{7.253}$$

in which ρ is the mass density, and the virtual work of the nonconservative forces is simply

$$\delta \overline{W}_{nc} = \int_D f \, \delta w \, dD \tag{7.254}$$

where f is the force density.

Next, we consider the variation in the potential energy, Eq. (7.252), in the form

$$\delta V = \int_D T\nabla w \cdot \delta\nabla w \, dD + \int_{S_2} kw\,\delta w\, dS = \int_D T\nabla w \cdot \nabla\delta w \, dD + \int_{S_2} kw\,\delta w\, dS \tag{7.255}$$

But, from Ref. 8, we can write the relation

$$u\nabla^2 v = u\nabla \cdot \nabla v = \nabla \cdot (u\nabla v) - \nabla u \cdot \nabla v \tag{7.256}$$

where $\nabla^2 = \nabla \cdot \nabla$ is the Laplacian, which is equal to the divergence of the gradient. Hence, letting $u = \delta w,\ v = w$, Eq. (7.255) becomes

$$\delta V = \int_D T\left[\nabla \cdot (\delta w \nabla w) - \delta w \nabla^2 w\right] dD + \int_{S_2} kw\,\delta w\, dS \tag{7.257}$$

At this point, we invoke the divergence theorem (Ref. 4)

$$\int_D \nabla \cdot \mathbf{A}\, dD = \int_S A_n\, dS \tag{7.258}$$

in which A_n is the component of the vector $\mathbf{A}$ along the exterior normal to boundary S, so that, letting $\mathbf{A} = \delta w \nabla w,\ dA_n = \dfrac{\partial w}{\partial n}\delta w$, Eq. (7.257) can be rewritten as

$$\begin{aligned}\delta V &= -\int_D T\nabla^2 w\,\delta w\, dD + \int_S T\frac{\partial w}{\partial n}\delta w\, dS + \int_{S_2} kw\,\delta w\, dS \\ &= -\int_D T\nabla^2 w\,\delta w\, dD + \int_{S_1} T\frac{\partial w}{\partial n}\delta w\, dS + \int_{S_2}\left(T\frac{\partial w}{\partial n} + kw\right)\delta w\, dS\end{aligned} \tag{7.259}$$

Moreover, by analogy with the one-dimensional case, Eq. (7.12),

$$\int_{t_1}^{t_2} \delta T\, dt = -\int_{t_1}^{t_2}\int_D \rho\ddot{w}\,\delta w\, dDdt \tag{7.260}$$

Inserting Eqs. (7.254), (7.259) and (7.260) into the extended Hamilton's principle, Eq. (7.4), we obtain

$$\begin{aligned}\int_{t_1}^{t_2}\Bigg[\int_D &\left(T\nabla^2 w - \rho\ddot{w} + f\right)\delta w\, dD \\ &+ \int_{S_1} T\frac{\partial w}{\partial n}\delta w\, dS + \int_{S_2}\left(T\frac{\partial w}{\partial n} + kw\right)\delta w\, dS\Bigg] dt = 0\end{aligned} \tag{7.261}$$

Finally, using the usual argument, we conclude that Eq. (7.261) can be satisfied for arbitrary δw in D and on S if and only if w satisfies the partial differential equation

$$T\nabla^2 w + f = \rho\ddot{w} \text{ in } D \tag{7.262}$$

and, in addition, either

$$T\frac{\partial w}{\partial n} = 0 \text{ on } S_1 \tag{7.263a}$$

or

$$w = 0 \text{ on } S_1 \tag{7.263b}$$

and

$$T\frac{\partial w}{\partial n} + kw = 0 \text{ on } S_2 \tag{7.264}$$

The boundary-value problem consists of the partial differential equation (7.262) to be satisfied over D and appropriate boundary conditions. For a membrane fixed at every point of S_1 and supported by a spring at every point of S_2, the boundary conditions consist of Eqs. (7.263b) and (7.264). On the other hand, if the membrane is free at every point of S_1, instead of being fixed, the boundary conditions consist of Eqs. (7.263a) and (7.264).

To derive the differential eigenvalue problem, we follow the established procedure, i.e., we let $f = 0$, assume that $w = WF$, where W depends on the spatial position alone and F depends on time alone and satisfies $\ddot{F} = -\lambda F$, eliminate the time dependence from the boundary-value problem, Eqs. (7.262)–(7.264), and obtain the partial differential equation

$$-T\nabla^2 W = \lambda\rho W, \qquad \lambda = \omega^2 \text{ over } D \tag{7.265}$$

Moreover, for a membrane fixed at every point of S_1 and spring-supported at every point of S_2, $S_1 + S_2 = S$, we obtain the boundary conditions

$$W = 0 \text{ on } S_1 \tag{7.266a}$$

$$T\frac{\partial W}{\partial n} + kW = 0 \text{ on } S_2 \tag{7.266b}$$

and if the membrane is free at every point of S_1, the boundary conditions

$$T\frac{\partial W}{\partial n} = 0 \text{ on } S_1 \tag{7.267a}$$

$$T\frac{\partial W}{\partial n} + kW = 0 \text{ on } S_2 \tag{7.267b}$$

The just derived eigenvalue problem is of the special type given by Eqs. (7.68) and (7.69) in which we identify the stiffness operator and the mass density

$$L = -T\nabla^2, \qquad m = \rho \text{ in } D \tag{7.268a, b}$$

Moreover, for a membrane fixed at all points of S_1 and spring-supported at all points of S_2, the boundary operators are

$$B_1 = 1 \text{ on } S_1 \tag{7.269a}$$

$$B_1 = T\frac{\partial}{\partial n} + k \text{ on } S_2 \tag{7.269b}$$

and if the membrane is free at all points of S_1, they are

$$B_1 = T\frac{\partial}{\partial n} \text{ on } S_1 \tag{7.270a}$$

$$B_1 = T\frac{\partial}{\partial n} + k \text{ on } S_2 \tag{7.270b}$$

The system is self-adjoint and positive definite for both types of boundary conditions. To verify self-adjointness, we consider two comparison functions u and v, use Eqs. (7.256) and (7.258), where in the latter $\mathbf{A} = u\nabla v$, and write

$$\begin{aligned}\int_D uLv\,dD &= -\int_D uT\nabla^2 v\,dD = -\int_D T\left[\nabla\cdot(u\nabla v) - \nabla u\cdot\nabla v\right]dD\\ &= \int_D T\nabla u\cdot\nabla v\,dD - \int_S Tu\frac{\partial v}{\partial n}dS\\ &= \int_D T\nabla u\cdot\nabla v\,dD + \int_{S_2} kuv\,dS\end{aligned} \tag{7.271}$$

which is symmetric in u and v. Hence, as anticipated, the operator L, and hence the system, is self-adjoint. It follows that the eigenvalues are real and the eigenfunctions are real and orthogonal. The eigenfunctions can be normalized so as to satisfy the orthonormality relations

$$\int_D \rho W_r W_s\,dD = \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.272a}$$

$$\begin{aligned}\int_D W_r L W_s\,dD &= -\int_D T W_r \nabla^2 W_s\,dD\\ &= \int_D T\nabla W_r\cdot\nabla W_s\,dD + \int_{S_2} kW_r W_s\,dS = \lambda_r\delta_{rs},\\ &\qquad r, s = 1, 2, \ldots\end{aligned} \tag{7.272b}$$

Moreover, to verify positive definiteness, we let $v = u$ in Eq. (7.271) and obtain

$$\begin{aligned}\int_D uLu\,dD &= \int_D T\nabla u\cdot\nabla u\,dD + \int_{S_2} ku^2\,dS\\ &= \int_D T\|\nabla u\|^2\,dD + \int_{S_2} ku^2\,dS = \|u\|_E^2 > 0\end{aligned} \tag{7.273}$$

where $\|u\|_E$ is the energy norm. Clearly, the energy norm is positive for nontrivial u, so that the operator L, and hence the system, is positive definite. It follows that all the eigenvalues are positive, a fact already taken into account when the time dependence was assumed to be harmonic.

In the above discussion, we carefully avoided reference to any particular set of coordinates. The deflection w can be given in terms of rectangular coordinates and time or curvilinear coordinates and time. Accordingly, the Laplacian ∇^2 can be expressed in terms of rectangular or curvilinear coordinates. As pointed out earlier in this section, the shape of the boundary dictates the choice of coordinates, because the only way we can deal effectively with boundary conditions is by formulating the problem in terms of coordinates capable of matching the shape of the boundary. In fact, there are only a few boundary shapes permitting closed-form solutions. We confine our discussion to rectangular and circular membranes.

i. Rectangular membranes

Under consideration is a rectangular membrane extending over a domain D defined by $0 < x < a$ and $0 < y < b$. The boundaries of the domain are the straight lines $x = 0, a$ and $y = 0, b$. If we assume that the mass density is constant, then Eq. (7.265) can be written in the form

$$\nabla^2 W(x, y) + \beta^2 W(x, y) = 0, \qquad \beta^2 = \frac{\rho\omega^2}{T}, \qquad x, y \text{ in } D \tag{7.274}$$

where the Laplacian in rectangular coordinates has the expression

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{7.275}$$

For a membrane *fixed at all boundaries*, the boundary conditions are

$$W(0, y) = 0, \qquad W(a, y) = 0 \tag{7.276a, b}$$

$$W(x, 0) = 0, \qquad W(x, b) = 0 \tag{7.276c, d}$$

and we note that they are all geometric boundary conditions. The differential equation, Eq. (7.274) together with the boundary conditions, Eqs. (7.276), constitute the eigenvalue problem.

Equation (7.274) can be solved by the method of separation of variables. To this end, we let the solution have the form

$$W(x, y) = X(x)Y(y) \tag{7.277}$$

Upon substitution in Eq. (7.274), we obtain

$$\frac{d^2X(x)}{dx^2}Y(y) + X(x)\frac{d^2Y(y)}{dy^2} + \beta^2 X(x)Y(y) = 0 \tag{7.278}$$

which can be divided through by $X(x)Y(y)$ to yield

$$\frac{1}{X(x)}\frac{d^2X(x)}{dx^2} + \frac{1}{Y(y)}\frac{d^2Y(y)}{dy^2} + \beta^2 = 0 \tag{7.279}$$

This leads to the equations

$$\frac{d^2X(x)}{dx^2} + \alpha^2 X(x) = 0 \tag{7.280a}$$

$$\frac{d^2Y(y)}{dy^2} + \gamma^2 Y(y) = 0 \tag{7.280b}$$

where

$$\alpha^2 + \gamma^2 = \beta^2 \tag{7.281}$$

The solution of Eq. (7.280a) is

$$X(x) = C_1 \sin \alpha x + C_2 \cos \alpha x \tag{7.282a}$$

and the solution of Eq. (7.280b) is

$$Y(y) = C_3 \sin \gamma y + C_4 \cos \gamma y \tag{7.282b}$$

so that, introducing Eqs. (7.282) into Eq. (7.277), we obtain

$$\begin{aligned} W(x, y) &= A_1 \sin \alpha x \sin \gamma y + A_2 \sin \alpha x \cos \gamma y \\ &\quad + A_3 \cos \alpha x \sin \gamma y + A_4 \cos \alpha x \cos \gamma y \end{aligned} \tag{7.283}$$

where A_1, A_2, A_3 and A_4, as well as α and γ, must be determined by using the boundary conditions.

Boundary condition (7.276a) gives

$$W(0, y) = A_3 \sin \gamma y + A_4 \cos \gamma y = 0 \tag{7.284}$$

which can hold true for all y, assuming that $\gamma \neq 0$, if and only if A_3 and A_4 are zero. Assuming that A_3 and A_4 are zero, boundary condition (7.276b) yields

$$W(a, y) = A_1 \sin \alpha a \sin \gamma y + A_2 \sin \alpha a \cos \gamma y = 0 \tag{7.285}$$

which can be satisfied if A_1 and A_2 are zero. This would give the trivial solution $W(x, y) = 0$, however, so that we must consider the other possibility, namely,

$$\sin \alpha a = 0 \tag{7.286a}$$

Similarly, boundary condition (7.276c) leads us to the conclusion that $A_2 = A_4 = 0$, whereas boundary condition (7.276d) gives

$$\sin \gamma b = 0 \tag{7.286b}$$

Equations (7.286) play the role of characteristic, or frequency equations, because together they define the eigenvalues of the system. Indeed, Eq. (7.286a) yields the infinite set of discrete roots

$$\alpha_m a = m\pi, \qquad m = 1, 2, \ldots \tag{7.287a}$$

and Eq. (7.286b) gives another infinite set of roots, or

$$\gamma_n b = n\pi, \qquad n = 1, 2, \ldots \tag{7.287b}$$

It follows from Eqs. (7.281) and (7.287) that the solution of the eigenvalue problem consists of the eigenvalues

$$\beta_{mn} = \sqrt{\alpha_m^2 + \gamma_n^2} = \pi \sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}, \qquad m, n = 1, 2, \ldots \tag{7.288}$$

and from Eq. (7.283) that the corresponding eigenfunctions are

$$W_{mn} = A_{mn} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} = \frac{2}{\sqrt{\rho ab}} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b}, \qquad m, n = 1, 2, \ldots \tag{7.289}$$

where the eigenfunctions have been normalized so as to satisfy $\int_0^a \int_0^b \rho W_{mn}^2 dx\, dy = 1 (m, n = 1, 2, \ldots)$. Moreover, from Eqs. (7.274) and (7.288), the natural frequencies are

$$\omega_{mn} = \beta_{mn} \sqrt{\frac{T}{\rho}} = \pi \sqrt{\left[\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2\right] \frac{T}{\rho}}, \qquad m, n = 1, 2, \ldots \tag{7.290}$$

We have shown earlier that the problem is self-adjoint and positive definite. From Eqs. (7.272), the orthonormality relations are

$$\int_0^a \int_0^b \rho W_{mn}(x, y) W_{rs}(x, y)\, dx\, dy = \delta_{mr}\delta_{ns} \tag{7.291a}$$

$$\begin{aligned} -&\int_0^a \int_0^b W_{mn}(x, y) T \nabla^2 W_{rs}(x, y)\, dx\, dy \\ &= \int_0^a \int_0^b T \left(\frac{\partial W_{mn}}{\partial x} \frac{\partial W_{rs}}{\partial x} + \frac{\partial W_{mn}}{\partial y} \frac{\partial W_{rs}}{\partial y} \right) dx\, dy = \lambda_{mn} \delta_{mr} \delta_{ns} \end{aligned} \tag{7.291b}$$

where $\lambda_{mn} = \omega_{mn}^2$. The first four eigenfunctions are plotted in Fig. 7.17. The nodes are straight lines; the number of nodal lines parallel to the x axis is $n - 1$ and the number parallel to the y axis is $m - 1$.

We note that some but not all of the higher natural frequencies are integer multiples of the fundamental frequency, $\omega_{mm} = m\omega_{11}$. For example, ω_{12} is not an integer multiple of ω_{11}. Hence, the sounds produced by vibrating membranes are not as pleasant as the sounds produced by strings or any other system with harmonic overtones.

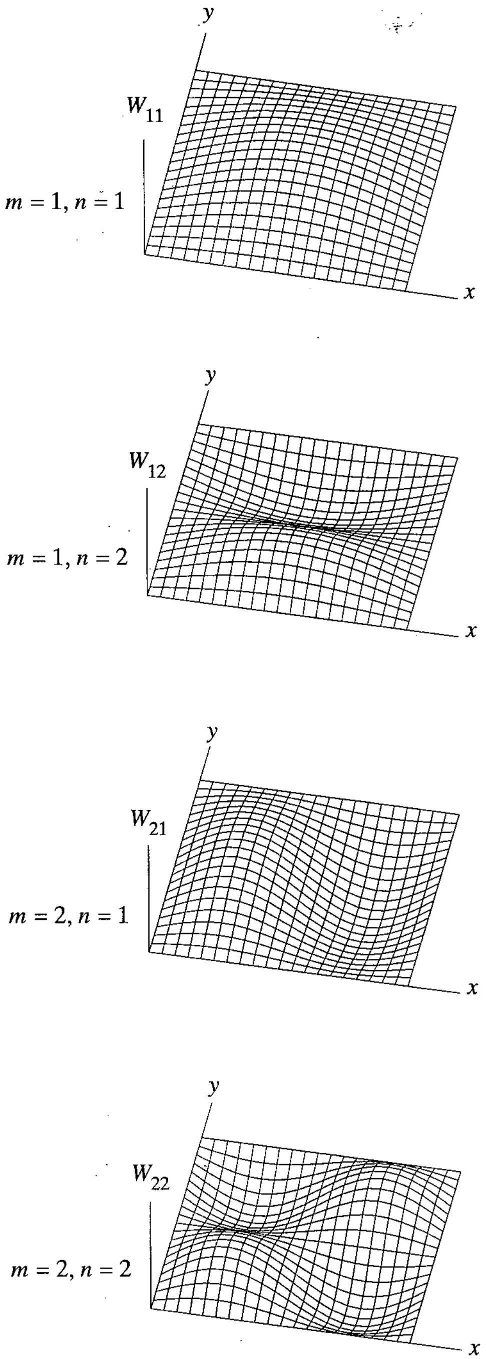
In the special case in which the ratio $R = a/b$ is a rational number, we have repeated natural frequencies $\omega_{mn} = \omega_{rs}$ if

$$m^2 + R^2 n^2 = r^2 + R^2 s^2 \tag{7.292}$$

For a ratio $R = 4/3$, we note that $\omega_{35} = \omega_{54}$, $\omega_{83} = \omega_{46}$, etc. For a square membrane, $a = b$, Eq. (7.292) reduces to

$$m^2 + n^2 = r^2 + s^2 \tag{7.293}$$

in which case we obtain repeated frequencies $\omega_{mn} = \omega_{nm}$. Hence, two distinct eigenfunctions W_{mn} and W_{nm} belong to the same eigenvalue, so that there are fewer eigenvalues than eigenfunctions. Such a case is said to be *degenerate*. As in the case of discrete systems, any linear combination of eigenfunctions belonging to repeated eigenvalues is also an eigenfunction. They are characterized by a large variety of



$$\omega_{11} = \pi\sqrt{\left(\frac{1}{a^2} + \frac{1}{b^2}\right)\frac{T}{\rho}}$$

$$\omega_{12} = \pi\sqrt{\left(\frac{1}{a^2} + \frac{4}{b^2}\right)\frac{T}{\rho}}$$

$$\omega_{21} = \pi\sqrt{\left(\frac{4}{a^2} + \frac{1}{b^2}\right)\frac{T}{\rho}}$$

$$\omega_{22} = 2\pi\sqrt{\left(\frac{1}{a^2} + \frac{1}{b^2}\right)\frac{T}{\rho}}$$

Figure 7.17 The first four modes of vibration of a uniform rectangular membrane fixed on all sides

nodal patterns. For the square membrane, the nodal lines are no longer straight lines except in special cases. The reader who wishes to pursue this subject further is referred to the text by Courant and Hilbert (Ref. 1, p. 302).

ii. Circular membranes

We consider a uniform circular membrane extending over a domain D defined by $0 < r < a$. The boundary of the domain is the circle S given by the equation $r = a$. Using the polar coordinates r and θ, the differential equation is

$$\nabla^2 W(r, \theta) + \beta^2 W(r, \theta) = 0, \qquad \beta^2 = \frac{\rho\omega^2}{T}, \qquad r, \theta \text{ in } D \tag{7.294}$$

where the Laplacian in polar coordinates is given by

$$\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} \tag{7.295}$$

Assuming a solution of the form

$$W(r, \theta) = R(r)\Theta(\theta) \tag{7.296}$$

Eq. (7.294) reduces to

$$\left(\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr}\right)\Theta + \frac{R}{r^2}\frac{d^2\Theta}{d\theta^2} + \beta^2 R\Theta = 0 \tag{7.297}$$

which can be separated into

$$\frac{d^2\Theta}{d\theta^2} + m^2\Theta = 0 \tag{7.298a}$$

$$\frac{d^2R}{dr^2} + \frac{1}{r}\frac{dR}{dr} + \left(\beta^2 - \frac{m^2}{r^2}\right)R = 0 \tag{7.298b}$$

where the constant m^2 has been assumed to be positive so as to obtain a harmonic solution for Θ. Furthermore, because the solution must be continuous, implying that the solution at $\theta = \theta_0$ must be identical to the solution at $\theta = \theta_0 + j2\pi$ $(j = 1, 2, \ldots)$ for any value θ_0, m must be an integer. Hence, the solution of Eq. (7.298a) is

$$\Theta_m(\theta) = C_{1m}\sin m\theta + C_{2m}\cos m\theta, \qquad m = 0, 1, 2, \ldots \tag{7.299}$$

Equation (7.298b), on the other hand, is a Bessel equation and its solution is

$$R_m(r) = C_{3m}J_m(\beta r) + C_{4m}Y_m(\beta r), \qquad m = 0, 1, 2, \ldots \tag{7.300}$$

where $J_m(\beta r)$ and $Y_m(\beta r)$ are Bessel functions of order m and of the first and second kind, respectively. The general solution can be written in the form

$$\begin{aligned} W_m(r, \theta) = {} & A_{1m}J_m(\beta r)\sin m\theta + A_{2m}J_m(\beta r)\cos m\theta \\ & + A_{3m}Y_m(\beta r)\sin m\theta + A_{4m}Y_m(\beta r)\cos m\theta, \\ & \qquad m = 0, 1, 2, \ldots \end{aligned} \tag{7.301}$$

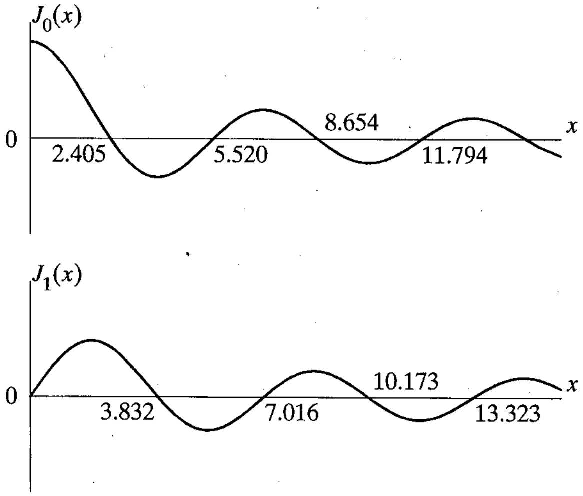


Figure 7.18 The zeros of the Bessel functions $J_0(x)$ and $J_1(x)$

Next, we consider a membrane *fixed* at the boundary $r = a$, so that the boundary condition at $r = a$ is

$$W_m(a, \theta) = 0, \quad m = 0, 1, 2, \ldots \tag{7.302}$$

At every interior point of the membrane *the displacement must be finite.* But Bessel functions of the second kind tend to infinity as the argument approaches zero. It follows that $A_{3m} = A_{4m} = 0$, so that Eq. (7.301) reduces to

$$W_m(r, \theta) = A_{1m} J_m(\beta r) \sin m\theta + A_{2m} J_m(\beta r) \cos m\theta, \quad m = 0, 1, 2, \ldots \tag{7.303}$$

At $r = a$, however, we have

$$W_m(a, \theta) = A_{1m} J_m(\beta a) \sin m\theta + A_{2m} J_m(\beta a) \cos m\theta = 0, \quad m = 0, 1, 2, \ldots \tag{7.304}$$

regardless of the value of θ. This can be satisfied only if

$$J_m(\beta a) = 0, \quad m = 0, 1, 2, \ldots \tag{7.305}$$

Equations (7.305) represent an infinite set of characteristic equations, or frequency equations, as for every m there is an infinite number of discrete solutions β_{mn} corresponding to the zeros of the Bessel functions J_m. As an illustration, the Bessel functions of zero and first order are plotted in Fig. 7.18. The intersections with the x-axis provide the roots $\beta_{mn}a$, from which we obtain the natural frequencies $\omega_{mn} = \beta_{mn}\sqrt{T/\rho}$. For each frequency ω_{mn} there are two modes, except when $m = 0$, for which we obtain only one mode. It follows that for $m \neq 0$ the natural modes are degenerate. The modes can be written as

$$W_{0n}(r, \theta) = A_{0n} J_0(\beta_{0n} r), \quad n = 1, 2, \ldots \tag{7.306a}$$

$$\begin{matrix} W_{mnc}(r, \theta) = A_{mnc} \\ W_{mns}(r, \theta) = A_{mns} \end{matrix} J_m(\beta_{mn} r) \begin{matrix} \cos m\theta, \\ \sin m\theta, \end{matrix} \quad m, n = 1, 2, \ldots \tag{7.306b}$$

The problem is self-adjoint and positive definite, so that the natural modes are orthogonal. From Eqs. (7.272), we can write the orthonormality relations

$$\int_D \rho W_{mn} W_{pq}\, dD = \int_0^{2\pi}\int_0^a \rho W_{mn} W_{pq} r dr d\theta = \delta_{mp}\delta_{nq} \tag{7.307a}$$

$$\begin{aligned}\int_D W_{mn} L W_{pq}\, dD &= -\int_0^{2\pi}\int_0^a W_{mn} T \nabla^2 W_{pq} r dr d\theta \\ &= \int_0^{2\pi}\int_0^a T\left(\frac{\partial W_{mn}}{\partial r}\frac{\partial W_{rs}}{\partial r} + \frac{1}{r^2}\frac{\partial W_{mn}}{\partial \theta}\frac{\partial W_{rs}}{\partial \theta}\right) r dr d\theta \\ &= \lambda_{mn}\delta_{mp}\delta_{nq}\end{aligned} \tag{7.307b}$$

where $\lambda_{mn} = \omega_{mn}^2$. To normalize the natural modes, we write (Ref. 7, p. 190)

$$\int_D \rho W_{0n}^2\, dD = \int_0^{2\pi}\int_0^a \rho A_{0n}^2 J_0^2(\beta_{0n} r)\, r\, dr\, d\theta = \pi\rho a^2 A_{0n}^2 J_1^2(\beta_{0n} a) = 1 \tag{7.308}$$

so that

$$A_{0n}^2 = \frac{1}{\pi\rho a^2 J_1^2(\beta_{0n} a)} \tag{7.309}$$

Also

$$\begin{aligned}\int_D \rho W_{mnc}^2\, dD &= \int_0^{2\pi}\int_0^a \rho A_{mnc}^2 J_m^2(\beta_{mn} r)\cos^2 m\theta\, r dr d\theta \\ &= \frac{\pi}{2}\rho A_{mnc}^2 a^2 J_{m+1}^2(\beta_{mn} a) = 1\end{aligned} \tag{7.310}$$

or

$$A_{mnc}^2 = \frac{2}{\pi\rho a^2 J_{m+1}^2(\beta_{mn} a)} \tag{7.311}$$

Similarly,

$$A_{mns}^2 = \frac{2}{\pi\rho a^2 J_{m+1}^2(\beta_{mn} a)} \tag{7.312}$$

The orthonormal modes take the form

$$W_{0n}(r,\theta) = \frac{1}{\sqrt{\pi\rho}\, a J_1(\beta_{0n} a)} J_0(\beta_{0n} r)\,, n = 1, 2, \ldots \tag{7.313a}$$

$$\begin{aligned}W_{mnc}(r,\theta) &= \\ W_{mns}(r,\theta) &= \end{aligned}\ \frac{\sqrt{2}}{\sqrt{\pi\rho}\, a J_{m+1}(\beta_{mn} a)} J_m(\beta_{mn} r) \begin{matrix}\cos m\theta, \\ \sin m\theta,\end{matrix}$$

$$m, n = 1, 2, \ldots \tag{7.313b}$$

They are plotted in Figs. 7.19 and 7.20. The nodal lines are circles r = constant and straight diametrical lines θ = constant. For $m = 0$, there are no diametrical nodes and there are $n - 1$ circular nodes. The first three modes W_{0n} are plotted in

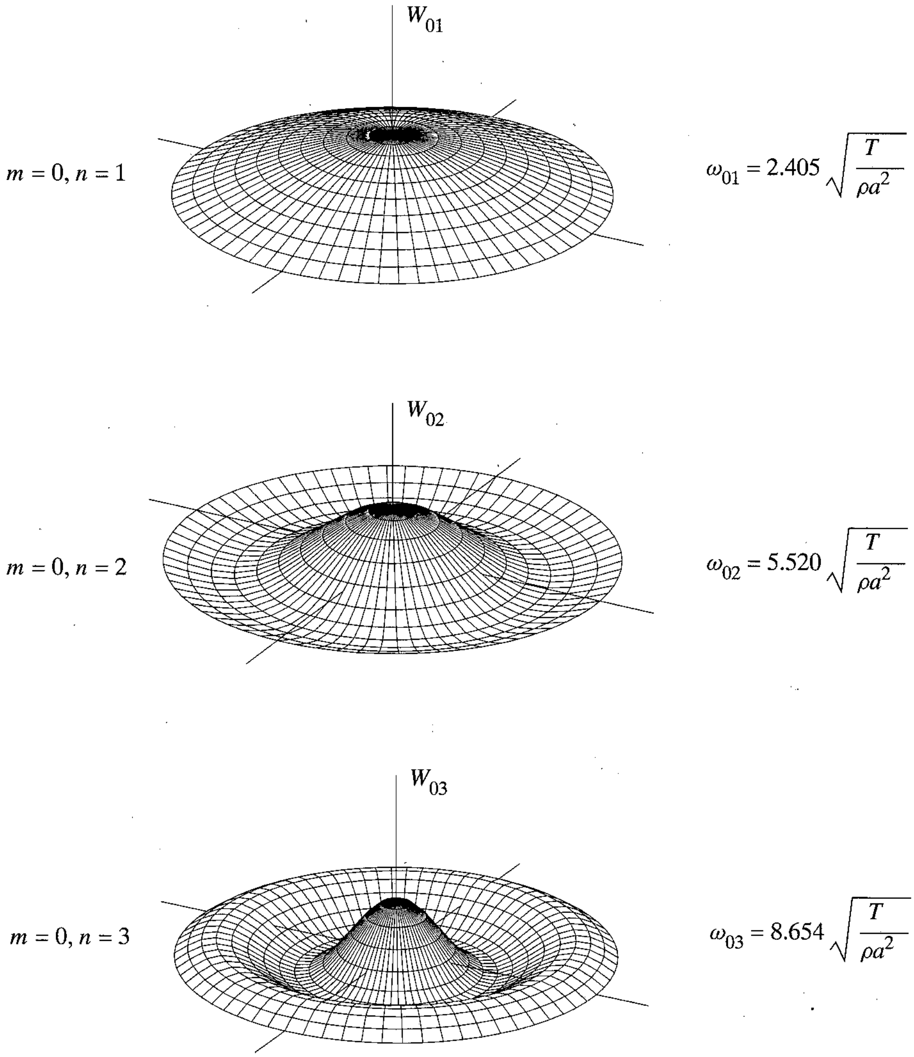

W_{01}
$m = 0, n = 1$
$\omega_{01} = 2.405\sqrt{\frac{T}{\rho a^2}}$
W_{02}
$m = 0, n = 2$
$\omega_{02} = 5.520\sqrt{\frac{T}{\rho a^2}}$
W_{03}
$m = 0, n = 3$
$\omega_{03} = 8.654\sqrt{\frac{T}{\rho a^2}}$

Figure 7.19 The three lowest symmetric modes of a uniform circular membrane fixed at $r = a$

Fig. 7.19. For $m = 1$ there is just one diametrical node and $n - 1$ circular nodes. The first two modes, W_{11c} and W_{12c}, are plotted in Fig. 7.20. In general, the mode W_{mn} has m equally spaced diametrical nodes and $n - 1$ circular nodes (the boundary is excluded) of radius $r_i = (\beta_{mi}/\beta_{mn})\,a$ $(i = 1, 2, \ldots, n - 1)$.

Note that, for very large arguments, we have the relation

$$\lim_{z\to\infty} J_m(z) = \sqrt{\frac{2}{\pi z}}\cos\left(z - \frac{2m+1}{4}\pi\right) \tag{7.314}$$

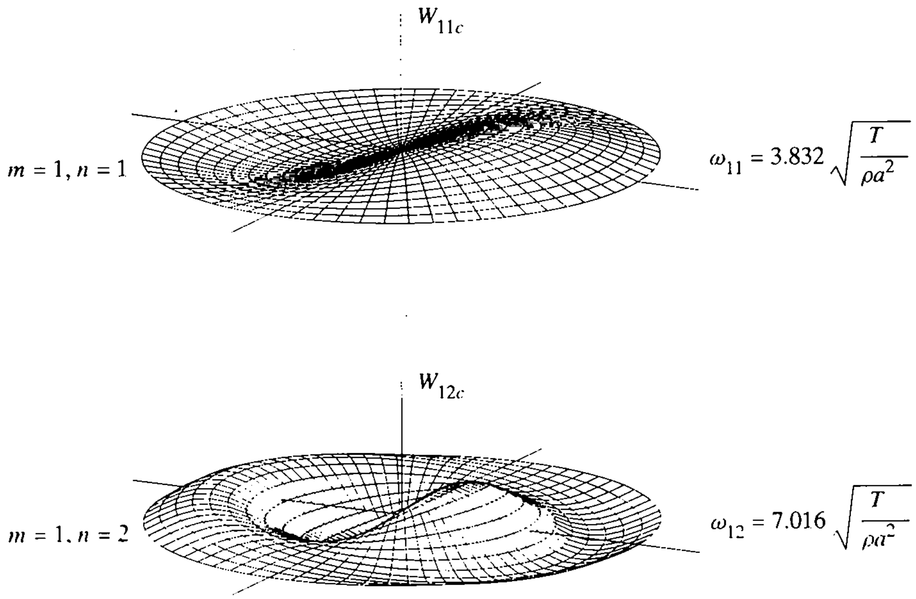

W_{11c}
$m = 1, n = 1$
$\omega_{11} = 3.832\sqrt{\frac{T}{\rho a^2}}$
W_{12c}
$m = 1, n = 2$
$\omega_{12} = 7.016\sqrt{\frac{T}{\rho a^2}}$

Figure 7.20 The two lowest antisymmetric modes of a uniform circular membrane fixed at $r = a$

so that the frequency equation, Eq. (7.305), leads us to the conclusion that, for very large n, the natural frequencies can be approximated by

$$\omega_{mn} = \left(\frac{m}{2} + n - \frac{1}{4}\right)\pi\sqrt{\frac{T}{\rho a^2}} \tag{7.315}$$

where both m and n are integers.

7.13 VIBRATION OF PLATES

In contrast to membranes, plates do have bending stiffness in a manner similar to beams in bending. There is one difference between beams and plates in bending, however. The beam is essentially a one-dimensional system. When a differential beam element bends, a portion of the material undergoes tension and the remaining portion undergoes compression, with the neutral axis acting as the dividing line between the two regions. The part in tension tends to contract laterally and the part in compression tends to expand. As long as the width of the beam is small, this lateral contraction and expansion is free to take place, and there are no lateral stresses. This is the essence of the ordinary beam theory. As the width of the beam increases, this effect tends to bend the cross section, so that a curvature is produced in the plane of the cross section, in addition to the curvature in the plane of bending. Let us now consider a plate and imagine for the moment that the plate is made up of individual parallel beams, obtained by dividing the plate by means of vertical planes, and focus our attention on material elements belonging to two such adjacent beams.

When undeformed, they share a lateral surface that is part of the dividing vertical plane. If allowed to behave like beams in bending, when the plate begins to bend these two adjacent beam elements would expand and contract laterally, so that in expanding each element would cross the dividing surface and occupy space belonging to the adjacent element, and in contracting each element would pull away from the dividing surface, resulting in a void in the material. In reality, this situation is not possible, so that internal lateral stresses must arise to prevent it from happening. Furthermore, in the case of plates one can think of two planes of bending, producing in general two distinct curvatures. In addition to bending, there is also twist present, because an element of plate area can be regarded as belonging to two orthogonal strips, so that bending of one strip can be looked upon as twisting of the orthogonal strip.

The *elementary theory of plates* is based on the following assumptions:

1. Deflections are small when compared with the plate thickness.
2. The normal stresses in the direction transverse to the plate can be ignored.
3. There is no force resultant on the cross-sectional area of a plate element. The middle plane of the plate does not undergo deformations during bending and can be regarded as a neutral plane.
4. Any straight line normal to the middle plane before deformation remains a straight line normal to the neutral plane during deformation.

These assumptions are reasonable for a relatively thin plate with no forces acting in the middle plane.

The boundary-value problem for a plate in bending vibration can be obtained by means of the extended Hamilton's principle, Eq. (7.4). To this end, we will find it convenient to begin with a description of the motion in terms of rectangular coordinates. The potential energy can be shown to have the expression (Ref. 15)

$$V = \frac{1}{2}\int_D D_E \left\{ \left(\nabla^2 w\right)^2 + 2(1-\nu)\left[\left(\frac{\partial^2 w}{\partial x \partial y}\right)^2 - \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2}\right]\right\} dD \tag{7.316}$$

where

$$D_E = \frac{Eh^3}{12\left(1-\nu^2\right)} \tag{7.317}$$

is the plate flexural rigidity, in which E is Young's modulus, h the plate thickness and ν Poisson's ratio. The kinetic energy is simply

$$T = \frac{1}{2}\int_D m\dot{w}^2\, dD \tag{7.318}$$

and the virtual work of the nonconservative forces is given by

$$\overline{\delta W} = \int_D f\,\delta w\; dD \tag{7.319}$$

where f is the force density. Note that, for simplicity, we assumed that there are no lumped masses and springs at the boundaries.

The variation in the potential energy has the form

$$\delta V = \int_D D_E \left\{ \nabla^2 w \, \delta \nabla^2 w + (1-\nu)\left[\frac{\partial^2 w}{\partial x \partial y} \delta \frac{\partial^2 w}{\partial x \partial y} + \frac{\partial^2 w}{\partial y \partial x} \delta \frac{\partial^2 w}{\partial y \partial x} - \frac{\partial^2 w}{\partial x^2} \delta \frac{\partial^2 w}{\partial x^2} - \frac{\partial^2 w}{\partial y^2} \delta \frac{\partial^2 w}{\partial y^2} \right] \right\} dD \qquad (7.320)$$

To render δV in a form involving variations in the displacement and the first partial derivatives of the displacement with respect to x and y alone, we use the relation (Ref. 8)

$$\nabla^2 u \nabla^2 v = u \nabla^4 v - \nabla \cdot \left(u \nabla \nabla^2 v \right) + \nabla \cdot \left(\nabla^2 v \nabla u \right) \qquad (7.321)$$

Then, assuming that the variation and differentiation processes are interchangeable, letting $u = \delta w$, $v = w$ and using the divergence theorem, Eq. (7.258), it can be shown that for *uniform flexural rigidity* Eq. (7.320) reduces to

$$\begin{aligned}\delta V = & \int_D D_E \nabla^4 w \, \delta w \, dD \\ & + \int_S D_E \left\{ \left[\frac{\partial}{\partial y}\left(\frac{\partial^2 w}{\partial y^2} + \nu \frac{\partial^2 w}{\partial x^2} \right) + (1-\nu)\frac{\partial^3 w}{\partial x^2 \partial y} \right] \delta w \, dx \right. \\ & - (1-\nu)\frac{\partial^2 w}{\partial x \partial y}\frac{\partial \delta w}{\partial x} dx - \left(\frac{\partial^2 w}{\partial y^2} + \nu \frac{\partial^2 w}{\partial x^2} \right) \frac{\partial \delta w}{\partial y} dx \\ & - \left[\frac{\partial}{\partial x}\left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2} \right) + (1-\nu)\frac{\partial^3 w}{\partial x \partial y^2} \right] \delta w \, dy \\ & \left. + (1-\nu)\frac{\partial^2 w}{\partial x \partial y}\frac{\partial \delta w}{\partial y} dy + \left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2} \right) \frac{\partial \delta w}{dx} dy \right\} \end{aligned} \qquad (7.322)$$

where $\nabla^4 = \nabla^2 \nabla^2$ is known as the *biharmonic operator*. Equation (7.322) can be expressed in terms of moments and forces by introducing the formulas (Ref. 13)

$$M_x = -D_E \left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2} \right) \qquad (7.323a)$$

$$M_y = -D_E \left(\frac{\partial^2 w}{\partial y^2} + \nu \frac{\partial^2 w}{\partial x^2} \right) \qquad (7.323b)$$

$$M_{xy} = -D_E (1-\nu) \frac{\partial^2 w}{\partial x \partial y} \qquad (7.323c)$$

$$Q_x = -D_E \left[\frac{\partial}{\partial x}\left(\frac{\partial^2 w}{\partial x^2} + \nu \frac{\partial^2 w}{\partial y^2} \right) + (1-\nu)\frac{\partial^3 w}{\partial x \partial y^2} \right]$$

$$= \frac{\partial M_x}{\partial x} + \frac{\partial M_{xy}}{\partial y} \tag{7.323d}$$

$$Q_y = -D_E\left[\frac{\partial}{\partial y}\left(\frac{\partial^2 w}{\partial y^2} + \nu\frac{\partial^2 w}{\partial x^2}\right) + (1-\nu)\frac{\partial^2 w}{\partial x^2 \partial y}\right]$$

$$= \frac{\partial M_y}{\partial y} + \frac{\partial M_{xy}}{\partial x} \tag{7.323e}$$

where M_x and M_y are bending moments, M_{xy} is a twisting moment and Q_x and Q_y are shearing forces. Inserting Eqs. (7.323) into Eq. (7.322), we obtain

$$\begin{aligned}\delta V &= \int_D D_E \nabla^4 w\, \delta w\, dD \\ &\quad + \int_S \left\{\left[-\left(\frac{\partial M_y}{\partial y} + \frac{\partial M_{xy}}{\partial x}\right)\delta w + M_{xy}\frac{\partial \delta w}{\partial x} + M_y \frac{\partial \delta w}{\partial y}\right] dx \right.\\ &\quad \left. + \left[\left(\frac{\partial M_x}{\partial x} + \frac{\partial M_{xy}}{\partial y}\right)\delta w - M_{xy}\frac{\partial \delta w}{\partial y} - M_x\frac{\partial \delta w}{\partial x}\right] dy\right\} \\ &= \int_D D_E \nabla^4 w\, \delta w\, dD + \int_S \left[\left(-Q_y\,\delta w + M_{xy}\frac{\partial \delta w}{\partial x} + M_y\frac{\partial \delta w}{\partial y}\right) dx \right.\\ &\quad \left. + \left(Q_x\,\delta w - M_{xy}\frac{\partial \delta w}{\partial y} - M_x\frac{\partial \delta w}{\partial x}\right) dy\right]\end{aligned} \tag{7.324}$$

At this point, we wish to express the boundary integral in terms of components normal and tangent to the boundary, n and s, respectively. To this end, we refer to Fig. 7.21 and write

$$\begin{aligned} dx &= -ds\,\sin\phi, \qquad dy = ds\,\cos\phi \\ \frac{\partial}{\partial x} &= \frac{\partial}{\partial n}\frac{\partial n}{\partial x} + \frac{\partial}{\partial s}\frac{\partial s}{\partial x} = \cos\phi\frac{\partial}{\partial n} - \sin\phi\frac{\partial}{\partial s} \\ \frac{\partial}{\partial y} &= \frac{\partial}{\partial n}\frac{\partial n}{\partial y} + \frac{\partial}{\partial s}\frac{\partial s}{\partial y} = \sin\phi\frac{\partial}{\partial n} + \cos\phi\frac{\partial}{\partial s}\end{aligned} \tag{7.325}$$

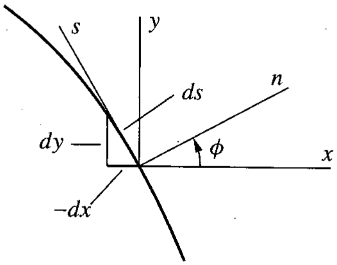


Figure 7.21 Tangential and normal directions at a plate boundary

Moreover, the moments and forces transform as follows (Ref. 13):

$$\begin{aligned} &M_x \cos^2\phi + 2M_{xy}\sin\phi\cos\phi + M_y\sin^2\phi = M_n \\ &(M_y - M_x)\sin\phi\cos\phi + M_{xy}(\cos^2\phi - \sin^2\phi) = M_{ns} \\ &Q_x\cos\phi + Q_y\sin\phi = Q_n \end{aligned} \tag{7.326}$$

Introducing Eqs. (7.325) and (7.326) into Eq. (7.324), we have

$$\delta V = \int_D D_E \nabla^4 w \delta w \, dD + \int_S \left(-M_n \frac{\partial \delta w}{\partial n} - M_{ns}\frac{\partial \delta w}{\partial s} + Q_n \delta w \right) ds \tag{7.327}$$

Equation (7.327) is still not in a form suitable for the derivation of the boundary-value problem. Indeed, the boundary integral would lead to three boundary conditions to be satisfied at every point of S, when in fact only two are called for. To resolve this apparent paradox, we carry out the following integration by parts

$$\int_S M_{ns}\frac{\partial \delta w}{\partial s} ds = M_{ns}\delta w \Big|_S - \int_S \frac{\partial M_{ns}}{\partial s}\delta w ds \tag{7.328}$$

If S is a closed smooth curve, then $M_{ns}\delta w\Big|_S = 0$, and Eq. (7.328) reduces to

$$\int_S M_{ns}\frac{\partial \delta w}{\partial s} ds = -\int_S \frac{\partial M_{ns}}{\partial s}\delta w \, ds \tag{7.329}$$

Inserting (7.329) into Eq. (7.327), we obtain the variation in the potential energy in the desired form

$$\delta V = \int_D D_E \nabla^4 w \, \delta w \, dD + \int_S \left[-M_n \, \delta\frac{\partial w}{\partial n} + \left(Q_n + \frac{\partial M_{ns}}{\partial s} \right) \delta w \right] ds \tag{7.330}$$

If S is not a smooth curve, as in the case in which the boundary is in the form of a polygon, the term $M_{ns}\delta w|_S$ gives rise to a so-called *corner condition* (Ref. 13). For a clamped corner, or a simply-supported corner, $\delta w = 0$ and for a free corner $M_{ns} = 0$. With the proviso that the term $M_{ns}\delta w|_S$ is either zero or is handled separately, we accept δV as given by Eq. (7.330).

Using Eq. (7.318) and recalling that δw vanishes at $t = t_1, t_2$, we can carry out an integration by parts with respect to t and write

$$\begin{aligned} \int_{t_1}^{t_2} \delta T \, dt &= \int_{t_1}^{t_2} \int_D m\dot{w}\,\delta\dot{w}\, dD\, dt = \int_D \int_{t_1}^{t_2} m\frac{\partial w}{\partial t}\frac{\partial}{\partial t}\delta w \, dt\, dD \\ &= \int_D \left[m\frac{\partial w}{\partial t}\delta w \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial}{\partial t}\left(m\frac{\partial w}{\partial t}\right) \delta w \, dt \right] dD \\ &= -\int_{t_1}^{t_2}\int_D m\ddot{w}\,\delta w \, dD\, dt \end{aligned} \tag{7.331}$$

Inserting Eqs. (7.319), (7.330) and (7.331) into the extended Hamilton's principle, Eq. (7.4), we obtain

$$\int_{t_1}^{t_2}\left\{-\int_D \left(D_E\nabla^4 w + m\ddot{w} - f\right)\delta w\, dD \right.$$
$$\left. + \int_S \left[M_n\,\delta\frac{\partial w}{\partial n} - \left(Q_n + \frac{\partial M_{ns}}{\partial s}\right)\delta w\right] ds\right\} dt = 0 \qquad (7.332)$$

Then, using the customary arguments concerning the arbitrariness of the virtual displacements, we conclude that Eq. (7.332) is satisfied if and only if

$$-D_E\nabla^4 w + f = m\ddot{w} \text{ in } D \qquad (7.333)$$

and either

$$M_n = 0 \text{ on } S \qquad (7.334a)$$

or

$$\frac{\partial w}{\partial n} = 0 \text{ on } S \qquad (7.334b)$$

and either

$$Q_{\text{eff}} = Q_n + \frac{\partial M_{ns}}{\partial s} = 0 \text{ on } S \qquad (7.335a)$$

where Q_{eff} denotes an "effective" shearing force, or

$$w = 0 \text{ on } S \qquad (7.335b)$$

The boundary-value problem consists of the partial differential equation Eq. (7.333) and two boundary conditions, one from Eqs. (7.334) and one from Eqs. (7.335).

The relations between the moments and shearing forces and deformations, in terms of normal and tangential coordinates (see Problem 7.43), are

$$M_n = -D_E\nabla^2 w + (1-\nu)\,D_E\left(\frac{1}{R}\frac{\partial w}{\partial n} + \frac{\partial^2 w}{\partial s^2}\right) \qquad (7.336a)$$

$$M_{ns} = -(1-\nu)\,D_E\left(\frac{\partial^2 w}{\partial n\partial s} - \frac{1}{R}\frac{\partial w}{\partial s}\right) \qquad (7.336b)$$

$$Q_n = -D_E\frac{\partial}{\partial n}\nabla^2 w \qquad (7.336c)$$

where the Laplacian has the form

$$\nabla^2 = \frac{\partial^2}{\partial n^2} + \frac{1}{R}\frac{\partial}{\partial n} + \frac{\partial^2}{\partial s^2} \qquad (7.337)$$

in which R is the radius of curvature of the boundary.

Boundary condition (7.335a) is associated with the name of Kirchhoff and is of some historical interest. Poisson believed that M_n, Q_n and M_{ns} must be independently zero at a free boundary, yielding a total of three boundary conditions, one too many for a fourth-order differential equation. Later, however, Kirchhoff cleared up the problem by pointing out that Q_n and M_{ns} are related as indicated by Eq. (7.335a). If variational principles are used to formulate the boundary-value problem, the boundary conditions are obtained both in the right number and in the correct form.

To derive the eigenvalue problem, we let $f = 0$ and assume a solution in the form

$$w = WF \tag{7.338}$$

where W depends on the spatial coordinates only and F is a time-dependent harmonic function of frequency ω. Then, following the usual steps involved in the separation of variables, the differential equation, Eq. (7.333), reduces to

$$D_E \nabla^4 W = \lambda m W, \qquad \lambda = \omega^2, \qquad \text{in } D \tag{7.339}$$

As mentioned earlier in this section, the boundary conditions to be satisfied at every point of S must be chosen from Eqs. (7.334) and (7.335) on the basis of physical considerations. For example, at a clamped edge, the displacement and slope must be zero. Hence, using Eq. (7.338) and dividing through by F, the boundary conditions for a *clamped edge* are simply

$$W = 0, \qquad \frac{\partial W}{\partial n} = 0 \tag{7.340a, b}$$

Moreover, from Eqs. (7.335b) and (7.334a), with due consideration to Eq. (7.336a), the boundary conditions for a *simply supported edge* become

$$W = 0, \qquad \nabla^2 W - (1 - \nu)\left(\frac{1}{R}\frac{\partial W}{\partial n} + \frac{\partial^2 W}{\partial s^2}\right) = 0 \tag{7.341a, b}$$

and, because W does not vary along the edge, Eqs. (7.341) assume the simplified form

$$W = 0, \qquad \frac{\partial^2 W}{\partial n^2} + \frac{\nu}{R}\frac{\partial W}{\partial n} = 0 \tag{7.342a, b}$$

Similarly using Eqs. (7.334a) and (7.335a), in conjunction with Eqs. (7.336), the boundary conditions along a *free edge* are

$$\nabla^2 W - (1 - \nu)\left(\frac{1}{R}\frac{\partial W}{\partial n} + \frac{\partial^2 W}{\partial s^2}\right) = 0 \tag{7.343a}$$

$$\frac{\partial}{\partial n}\nabla^2 W + (1 - \nu)\frac{\partial}{\partial s}\left(\frac{\partial^2 W}{\partial n \partial s} - \frac{1}{R}\frac{\partial W}{\partial s}\right) = 0 \tag{7.343b}$$

Again, we must recognize that the eigenvalue problem for the transverse vibration of uniform plates fits the pattern of Sec. 7.5. In this case, the stiffness operator and mass density are

$$L = D_E \nabla^4, \qquad M = m \tag{7.344}$$

and it follows that the eigenvalue problem is of the special type, in the sense that M is a mere function and the boundary conditions do not depend on λ.

Before proceeding to the solution of the eigenvalue problem for some cases of interest, we propose to derive a criterion for the self-adjointness of L, and hence of the system. To this end, we let u and v be two comparison functions, use Eq. (7.321), consider the divergence theorem, Eq. (7.258), and write

$$\begin{aligned}\int_D uLv\,dD &= \int_D D_E u\nabla^4 v\,dD \\ &= \int_D D_E\left[\nabla\cdot\left(u\nabla\nabla^2 v\right) - \nabla\cdot\left(\nabla^2 v\nabla u\right) + \nabla^2 u\nabla^2 v\right]dD \\ &= \int_S D_E\left(u\frac{\partial}{\partial n}\nabla^2 v - \nabla^2 v\frac{\partial u}{\partial n}\right)ds + \int_D D_E\nabla^2 u\nabla^2 v\,dD \qquad (7.345)\end{aligned}$$

If the integral over the boundary S vanishes, the right side of Eq. (7.345) is symmetric in u and v and the eigenvalue problem is self-adjoint. This is the case when the boundary points are simply supported, clamped, or free. Note that when some or all boundary points are supported by springs, the potential energy, Eq. (7.316), must be modified so as to include a boundary term of the type $\frac{1}{2}\int_S kw^2\,dS$. In this case, the integral over the boundary S in Eq. (7.345) does not vanish but is symmetric in u and v, so that the eigenvalue problem is once again self-adjoint.

For self-adjoint systems the eigenvalues are real and the eigenfunctions are real and orthogonal. We assume that the eigenfunctions have been normalized so as to satisfy the orthonormality conditions

$$\int_D mW_r W_s\,dD = \delta_{rs}, \qquad r, s = 1, 2, \ldots \qquad (7.346a)$$

$$\int_D D_e W_r \nabla^4 W_s\,dD = \lambda_r\,\delta_{rs}, \qquad r, s = 1, 2, \ldots \qquad (7.346b)$$

As in the case of vibration of thin membranes, the shape of the boundary dictates the type of coordinates to be used. For plates, however, the satisfaction of the boundary conditions turns out to be a much more formidable task than for membranes. Only rectangular and circular plates will be discussed here.

The boundary-value problem defined by the differential equation (7.333) and boundary conditions from Eqs. (7.334) and (7.335) and the eigenvalue problem defined by the differential equation (7.339) and boundary conditions from Eqs. (7.340)–(7.343) are for plates of constant flexural rigidity alone. If the flexural rigidity is not constant, such as when the plate thickness varies, additional terms must be included (Ref. 15). No closed-form solutions can be expected for variable-thickness plates.

The plate theory presented here ignores shear deformation and rotatory inertia effects and is known as the classical plate theory. An extension of the theory so as to include shear deformation in the static deflection of plates was carried out by Reissner (Ref. 12) and to include both shear deformation and rotatory inertia in the vibration of plates by Mindlin (Ref. 10). For a discussion of Mindlin's higher-order plate theory, see the monograph by Leissa (Ref. 6).

i. Rectangular plates

We consider a *uniform rectangular plate* extending over a domain D defined by $0 < x < a$ and $0 < y < b$. The boundaries of the domain are the straight lines $x = 0, a$ and $y = 0, b$. Equation (7.339), in rectangular coordinates, takes the form

$$\nabla^4 W(x, y) - \beta^4 W(x, y) = 0, \qquad \beta^4 = \frac{\omega^2 m}{D_E}, \ x, y \text{ in } D \tag{7.347}$$

where the biharmonic operator is given by

$$\nabla^4 = \nabla^2\nabla^2 = \frac{\partial^4}{\partial x^4} + 2\frac{\partial^4}{\partial x^2 \partial y^2} + \frac{\partial^4}{\partial y^4} \tag{7.348}$$

Equation (7.347) can be expressed in the operator form

$$\left(\nabla^4 - \beta^4\right) W(x, y) = \left(\nabla^2 + \beta^2\right)\left(\nabla^2 - \beta^2\right) W(x, y) = 0 \tag{7.349}$$

which permits us to write

$$\left(\nabla^2 - \beta^2\right) W = W_1, \qquad \left(\nabla^2 + \beta^2\right) W_1 = 0 \tag{7.350a, b}$$

Because β^2 is constant, the solution of Eq. (7.350a), and hence the solution of Eq. (7.347), is

$$W = W_1 + W_2 \tag{7.351}$$

where W_2 is the solution of the homogeneous equation

$$\left(\nabla^2 - \beta^2\right) W_2 = \left[\nabla^2 + (i\beta)^2\right] W_2 = 0 \tag{7.352}$$

We note here that the proportionality factor $-1/2\beta^2$ multiplying W_1 in Eq. (7.351) was omitted as irrelevant, because W_1 is obtained by solving a homogeneous equation, Eq. (7.350b). Equation (7.350b) resembles the equation for the vibration of a thin uniform membrane, whose general solution was obtained in Sec. 7.12 in the form of Eq. (7.283). Moreover, Eq. (7.352) has the same form as Eq. (7.350b), except that β is replaced by $i\beta$. Hence, the solution of Eq. (7.352) can be obtained from Eq. (7.283) by replacing the trigonometric functions by hyperbolic functions. It follows that the general solution of Eq. (7.347) is

$$\begin{aligned} W(x, y) = {} & A_1 \sin\alpha x \sin\gamma y + A_2 \sin\alpha x \cos\gamma y + A_3 \cos\alpha x \sin\gamma y \\ & + A_4 \cos\alpha x \cos\gamma y + A_5 \sinh\alpha_1 x \sinh\gamma_1 y \\ & + A_6 \sinh\alpha_1 x \cosh\gamma_1 y + A_7 \cosh\alpha_1 x \sinh\gamma_1 y \\ & + A_8 \cosh\alpha_1 x \cosh\gamma_1 y, \qquad \alpha^2 + \gamma^2 = \alpha_1^2 + \gamma_1^2 = \beta^2 \end{aligned} \tag{7.353}$$

We consider a *simply supported* plate. Because for a straight boundary the radius of curvature R is infinite the boundary conditions, Eqs. (7.342), reduce to

$$W = 0, \qquad \frac{\partial^2 W}{\partial x^2} = 0, \qquad x = 0, a \tag{7.354a, b}$$

$$W = 0, \qquad \frac{\partial^2 W}{\partial y^2} = 0, \qquad y = 0, b \tag{7.354c, d}$$

Upon using boundary conditions (7.354), we conclude that all the coefficients A_i, with the exception of A_1, vanish and, in addition, we obtain the two characteristic equations

$$\sin\alpha a = 0, \qquad \sin\gamma b = 0 \tag{7.355a, b}$$

Their solutions are

$$\alpha_m a = m\pi, \qquad m = 1, 2, \ldots \tag{7.356a}$$

$$\gamma_n b = n\pi, \qquad n = 1, 2, \ldots \tag{7.356b}$$

so that the natural frequencies of the system become

$$\omega_{mn} = \beta_{mn}^2 \sqrt{\frac{D_E}{m}} = \pi^2 \left[\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2\right] \sqrt{\frac{D_E}{m}}, \qquad m, n = 1, 2, \ldots \tag{7.357}$$

and the corresponding natural modes, normalized so that $\int_0^a \int_0^b m W_{mn}^2 dx\, dy = 1$, are

$$W_{mn}(x, y) = \frac{2}{\sqrt{mab}} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b}, \qquad m, n = 1, 2, \ldots \tag{7.358}$$

which are identical to the modes of the clamped rectangular membrane. However, the natural frequencies are different from those of the membrane.

It is easy to see that boundary conditions (7.354) render the boundary integral in Eq. (7.345) zero, so that the eigenvalue problem is self-adjoint. It follows immediately that the eigenfunctions are orthonormal, satisfying Eqs. (7.346).

A special class of eigenvalue problems for rectangular plates admitting closed-form solution is characterized by the fact that two opposing sides are simply supported. The interesting part is that attempts to obtain closed-form solutions by means of Eq. (7.353) do not bear fruit. An approach yielding results uses experience gained from the simply supported plate to assume a solution separable in x and y in which the part associated with the simply supported sides is given. To illustrate the approach, we consider a plate *simply supported at* $x = 0, a$ and *clamped at* $y = 0, b$. Then, consistent with results obtained for the plate simply supported on all sides, we assume a solution of the form (Ref. 14)

$$W_m(x, y) = Y_m(y) \sin\alpha_m x \tag{7.359}$$

in which, according to Eq. (7.356a), $\alpha_m = m\pi/a$ $(m = 1, 2, \ldots)$. Inserting Eq. (7.359) into Eq. (7.347), recalling Eq. (7.348) and dividing through by $\sin\alpha_m x$, we obtain

$$\frac{d^4 Y_m(y)}{dy^4} - 2\alpha_m^2 \frac{d^2 Y_m(y)}{dy^2} + (\alpha_m^4 - \beta_m^4) Y_m(y) = 0, \qquad 0 < y < b \tag{7.360}$$

where, in view of Eqs. (7.340), Y_m must satisfy the boundary conditions

$$Y_m = 0, \qquad \frac{dY_m}{dy} = 0, \qquad y = 0, b \tag{7.361a, b}$$

The solution of Eq. (7.360) has the exponential form

$$Y_m(y) = Ae^{s_m y} \tag{7.362}$$

Inserting Eq. (7.362) into Eq. (7.360) and dividing through by $e^{s_m y}$, we obtain the characteristic equation

$$s_m^4 - 2\alpha_m^2 s_m^2 + \alpha_m^4 - \beta_m^4 = 0 \tag{7.363}$$

which represents a quadratic equation in s_m^2. The solutions of Eq. (7.363) can be shown to be

$$s_{1m} = -s_{2m} = \gamma_{1m} = \sqrt{\beta_m^2 + \alpha_m^2}, \qquad s_{3m} = -s_{4m} = i\gamma_{2m} = i\sqrt{\beta_m^2 - \alpha_m^2} \tag{7.364}$$

so that solution (7.362) can be rewritten as

$$\begin{aligned} Y_m(y) &= A_1 e^{s_{1m} y} + A_2 e^{s_{2m} y} + A_3 e^{s_{3m} y} + A_4 e^{s_{4m} y} \\ &= C_1 \cosh \gamma_{1m} y + C_2 \sinh \gamma_{1m} y + C_3 \cos \gamma_{2m} y + C_4 \sin \gamma_{2m} y \end{aligned} \tag{7.365}$$

Introducing Eq. (7.365) into Eqs. (7.361), we obtain

$$\begin{aligned} Y_m(0) &= C_1 + C_3 = 0 \\ Y_m'(0) &= C_2\gamma_{1m} + C_4\gamma_{2m} = 0 \\ Y_m(b) &= C_1 \cosh \gamma_{1m} b + C_2 \sinh \gamma_{1m} b + C_3 \cos \gamma_{2m} b + C_4 \sin \gamma_{2m} b = 0 \\ Y_m'(b) &= C_1\gamma_{1m} \sinh \gamma_{1m} b + C_2\gamma_{1m} \cosh \gamma_{1m} b \\ &\quad - C_3\gamma_{2m} \sin \gamma_{2m} b + C_4\gamma_{2m} \cos \gamma_{2m} b = 0 \end{aligned} \tag{7.366}$$

where primes denote derivatives with respect to y. Equations (7.366) have a solution provided the determinant of the coeficients is zero, or

$$\begin{aligned} \Delta(\beta_m) &= \begin{vmatrix} 1 & 0 & 1 & 0 \\ 0 & \gamma_{1m} & 0 & \gamma_{2m} \\ \cosh \gamma_{1m} b & \sinh \gamma_{1m} b & \cos \gamma_{2m} b & \sin \gamma_{2m} b \\ \gamma_{1m} \sinh \gamma_{1m} b & \gamma_{1m} \cosh \gamma_{1m} b & -\gamma_{2m} \sin \gamma_{2m} b & \gamma_{2m} \cos \gamma_{2m} b \end{vmatrix} \\ &= 2\gamma_{1m}\gamma_{2m}(1 - \cosh \gamma_{1m} b \cos \gamma_{2m} b) \\ &\quad + (\gamma_{1m}^2 - \gamma_{2m}^2) \sinh \gamma_{1m} b \sin \gamma_{2m} b = 0, \qquad m = 1, 2, \dots \end{aligned} \tag{7.367}$$

Equations (7.367) with γ_{1m} and γ_{2m} given by Eqs. (7.364) represent an infinity of characteristic equations, one for every m, and each equation has an infinity of roots β_m. We identify these roots by $n = 1, 2, \dots$ and denote the double infinity of roots by β_{mn}^2 $(m, n = 1, 2, \dots)$. Then, inserting these values into Eqs. (7.364), we obtain

$$\gamma_{1mn} = \sqrt{\beta_{mn}^2 + \alpha_m^2}, \qquad \gamma_{2mn} = \sqrt{\beta_{mn}^2 - \alpha_m^2}, \qquad m, n = 1, 2, \dots \tag{7.368}$$

Moreover, solving Eqs. (7.366) for C_2, C_3 and C_4 in terms of C_1 and using Eqs. (7.368), we can write

$$\begin{aligned} Y_{mn}(y) = C_{mn}\Bigg[& \cosh \gamma_{1mn} y - \cos \gamma_{2mn} y \\ & - \frac{\cosh \gamma_{1mn} b - \cos \gamma_{2mn} b}{\sinh \gamma_{1mn} b - (\gamma_{1mn}/\gamma_{2mn}) \sin \gamma_{2mn} b} \Big(\sinh \gamma_{1mn} y \\ & - \frac{\gamma_{1mn}}{\gamma_{2mn}} \sin \gamma_{2mn} y \Big) \Bigg], \qquad m, n = 1, 2, \ldots \qquad (7.369) \end{aligned}$$

Finally, inserting Eq. (7.369) into Eq. (7.359), we obtain the desired eigenfunctions in the general form

$$W_{mn}(x, y) = Y_{mn}(y) \sin \alpha_m x, \qquad m, n = 1, 2, \ldots \qquad (7.370)$$

Moreover, from Eq. (7.347), we conclude that the natural frequencies are

$$\omega_{mn} = \beta_{mn}^2 \sqrt{D_E/m}, \qquad m, n = 1, 2, \ldots \qquad (7.371)$$

No confusion should arise from the fact that the symbol m denotes both the mass density and the first subscript in the natural frequencies and modes. We observe from Eq. (7.368) that the quantities γ_{1mn} and γ_{2mn}, defining the dependence of the eigenfunctions on y, are functions of the quantities α_m, defining the dependence of the eigenfunctions on x. By contrast, in the case of a plate simply supported on all sides, α_m and γ_n are independent of one another.

The eigenvalue problem has been solved numerically for a plate of sides ratio $a/b = 1.5$. Table 7.1 shows normalized natural frequenices for $m, n = 1, 2, 3$.

TABLE 7.1 Normalized Natural Frequencies

$$\overline{\omega}_{mn} = \omega_{mn} b^2 \sqrt{m/D_E}$$

n m	1	2	3
1	25.043584	65.007865	124.51603
2	35.103815	75.604983	135.61235
3	54.743071	94.585278	154.77570

ii. Circular plates

Now we consider a uniform circular plate extending over a domain D given by $0 < r < a$, where the boundary S of the domain is the circle $r = a$. Because the boundary is circular, we use the polar coordinates r and θ, so that the differential equation is

$$\nabla^4 W(r, \theta) - \beta^4 W(r, \theta) = 0, \qquad \beta^4 = \frac{\omega^2 m}{D_E}, \qquad r, \theta \text{ in } D \qquad (7.372)$$

where the biharmonic operator, in polar coordinates, has the form

$$\nabla^4 = \nabla^2\nabla^2 = \left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right)\left(\frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right) \tag{7.373}$$

Following the same pattern as for rectangular plates, Eq. (7.372) can be written in the operator form

$$\left(\nabla^2 + \beta^2\right) W_1(r,\theta) = 0, \qquad \left[\nabla^2 + (i\beta)^2\right] W_2(r,\theta) = 0 \tag{7.374a, b}$$

which must be satisfied over the domain D. Equation (7.374a) has precisely the same form as the equation for the vibration of circular membranes, Eq. (7.294), so that its solution is given by Eq. (7.301), although the definition of β is not the same as for membranes. Moreover, the solution of Eq. (7.374b) is obtained from Eq. (7.301) by replacing β by $i\beta$. The Bessel functions of imaginary argument, $J_m(ix)$ and $Y_m(ix)$, are called *modified or hyperbolic Bessel functions* and denoted by $I_m(x)$ and $K_m(x)$, respectively. The hyperbolic Bessel functions are not equal to the ordinary Bessel functions of imaginary argument but are proportional to them (Ref. 7, p. 113). It follows that the solution of Eq. (7.372), which is the sum of the solutions of Eqs. (7.374), has the form

$$\begin{aligned} W_m(r,\theta) = {} & [A_{1m}J_m(\beta r) + A_{3m}Y_m(\beta r) + B_{1m}I_m(\beta r) + B_{3m}K_m(\beta r)]\sin m\theta \\ & + [A_{2m}J_m(\beta r) + A_{4m}Y_m(\beta r) + B_{2m}I_m(\beta r) + B_{4m}K_m(\beta r)]\cos m\theta, \\ & m = 0, 1, 2, \ldots \end{aligned} \tag{7.375}$$

where $W_m(r,\theta)$ is subject to given boundary conditions.

As an example, we consider the case of a *clamped plate*, for which the boundary conditions are

$$W(a,\theta) = 0, \qquad \left.\frac{\partial W(r,\theta)}{\partial r}\right|_{r=a} = 0 \tag{7.376a, b}$$

In addition, *the solution must be finite at every interior point.* This immediately eliminates Bessel functions of the second kind, Y_m and K_m, which become infinite at $r = 0$. Hence, solution (7.375) reduces to

$$\begin{aligned} W_m(r,\theta) = {} & [A_{1m}J_m(\beta r) + B_{1m}I_m(\beta r)]\sin m\theta \\ & + [A_{2m}J_m(\beta r) + B_{2m}I_m(\beta r)]\cos m\theta, \qquad m = 0, 1, 2, \ldots \end{aligned} \tag{7.377}$$

Boundary condition (7.376a) yields

$$B_{1m} = -\frac{J_m(\beta a)}{I_m(\beta a)}A_{1m}, \qquad B_{2m} = -\frac{J_m(\beta a)}{I_m(\beta a)}A_{2m}, \qquad m = 0, 1, 2, \ldots \tag{7.378}$$

so that

$$W_m(r,\theta) = \left[J_m(\beta r) - \frac{J_m(\beta a)}{I_m(\beta a)}I_m(\beta r)\right](A_{1m}\sin m\theta + A_{2m}\cos m\theta), \qquad m = 0, 1, 2, \ldots \tag{7.379}$$

On the other hand, boundary condition (7.376b) leads to the set of characteristic equations

$$\left[\frac{d}{dr}J_m(\beta r) - \frac{J_m(\beta a)}{I_m(\beta a)}\frac{d}{dr}I_m(\beta r)\right]_{r=a} = 0, \quad m = 0, 1, 2, \ldots \tag{7.380}$$

But

$$\frac{d}{dr}J_m(\beta r) = \beta\left[J_{m-1}(\beta r) - \frac{m}{\beta r}J_m(\beta r)\right] \tag{7.381a}$$

$$\frac{d}{dr}I_m(\beta r) = \beta\left[I_{m-1}(\beta r) - \frac{m}{\beta r}I_m(\beta r)\right] \tag{7.381b}$$

so that the set of characteristic equations reduces to

$$I_m(\beta a)\,J_{m-1}(\beta a) - J_m(\beta a)\,I_{m-1}(\beta a) = 0, \quad m = 0, 1, 2, \ldots \tag{7.382}$$

For a given m, we must solve Eq. (7.382) numerically for the eigenvalues β_{mn}. The natural frequencies are related to the eigenvalues by

$$\omega_{mn} = \beta_{mn}^2\sqrt{\frac{D_E}{m}} \tag{7.383}$$

For each frequency ω_{mn} there are two corresponding natural modes, except for $m = 0$, for which there is just one mode. Hence, as for membranes, all modes for which $m \neq 0$ are *degenerate*. The natural modes can be written in the form

$$W_{0n}(r,\theta) = A_{0n}\left[I_0(\beta_{0n}a)\,J_0(\beta_{0n}r) - J_0(\beta_{0n}a)\,I_0(\beta_{0n}r)\right], \quad n = 1, 2, \ldots \tag{7.384a}$$

$$\begin{matrix} W_{mnc}(r,\theta) = A_{mnc} \\ W_{mns}(r,\theta) = A_{mns} \end{matrix}\left[I_m(\beta_{mn}a)\,J_m(\beta_{mn}r) - J_m(\beta_{mn}a)\,I_m(\beta_{mn}r)\right]\begin{matrix}\cos m\theta, \\ \sin m\theta,\end{matrix} \quad m, n = 1, 2, \ldots \tag{7.384b}$$

For $m = 0$, there are no diametrical nodes and there are $n - 1$ circular nodes. The modes W_{01} and W_{02} are plotted in Fig. 7.22. For $m = 1$, there is one diametrical node and $n - 1$ circular nodes. The mode W_{11c} is plotted in Fig. 7.23. Note that the overtones are not harmonic.

It is easy to see that boundary conditions (7.376) render the boundary integral in Eq. (7.345) equal to zero, so that the problem is self-adjoint. Consequently, the natural modes are orthogonal.

Defining the modified Bessel functions of the first kind by the relation $I_m(x) = i^{-m}J_m(ix)$ and using Eq. (7.314), it can be shown that, *for large argument*, the solutions of the characteristic equation, Eq. (7.382), tend to

$$a\beta_{mn} = \left(\frac{m}{2} + n\right)\pi \tag{7.385}$$

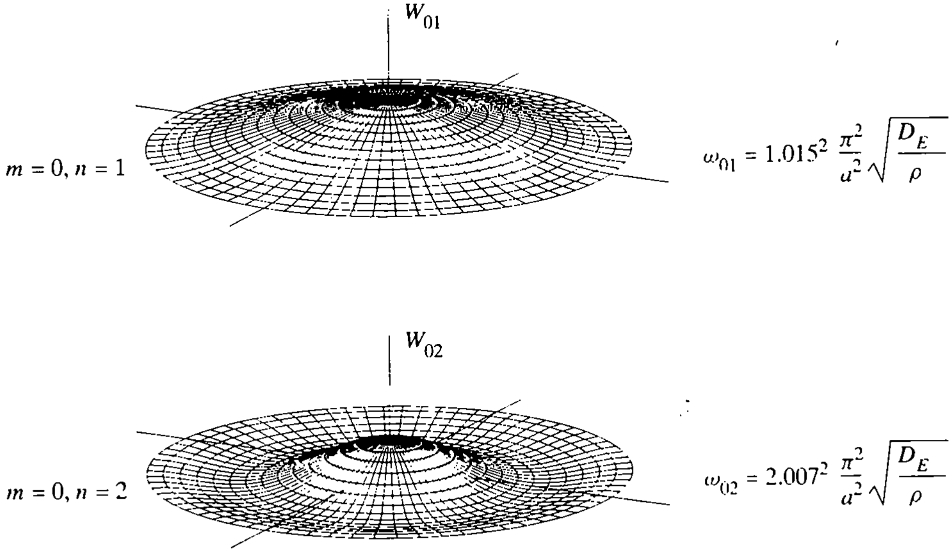

W_{01}
$m = 0, n = 1$
$\omega_{01} = 1.015^2 \frac{\pi^2}{a^2}\sqrt{\frac{D_E}{\rho}}$
W_{02}
$m = 0, n = 2$
$\omega_{02} = 2.007^2 \frac{\pi^2}{a^2}\sqrt{\frac{D_E}{\rho}}$

Figure 7.22 The two lowest symmetric modes of a uniform circular plate clamped at $r = a$

and, consequently, for large n, the natural frequencies tend to

$$\omega_{mn} = \left(\frac{m}{2} + n\right)^2 \frac{\pi^2}{a^2}\sqrt{\frac{D_E}{\rho}} \tag{7.386}$$

Because of the limitations of the elementary plate theory, however, this result is only of academic value.

In view of the fact that for circular plates the boundary $r = a$ is a closed smooth curve, the boundary integral in Eq. (7.345) simply vanishes. It follows that eigenvalue problems for circular plates are self-adjoint, so that the natural modes are orthogonal. Related to this is the fact that the boundary conditions for circular plates can be satisfied by working with the radial variable r alone, with the angle θ playing no role. This explains why eigenvalue problems for circular plates admit many more closed-form solutions than for rectangular plates (Ref. 6).

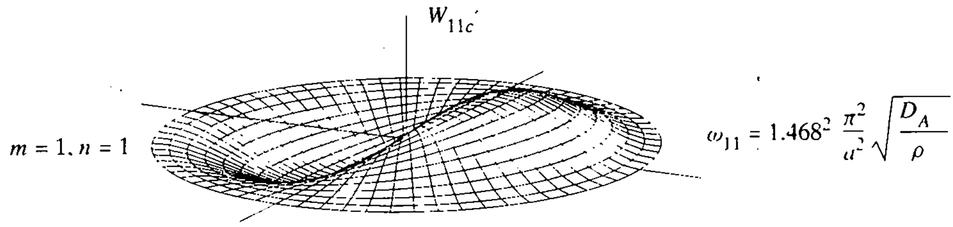

W_{11c}
$m = 1, n = 1$
$\omega_{11} = 1.468^2 \frac{\pi^2}{a^2}\sqrt{\frac{D_A}{\rho}}$

Figure 7.23 The lowest antisymmetric mode of a uniform circular plate clamped at $r = a$

7.14 VARIATIONAL FORMULATION OF THE DIFFERENTIAL EIGENVALUE PROBLEM

In Chapter 4, we demonstrated that the algebraic eigenvalue problem can be formulated as a variational problem consisting of rendering Rayleigh's quotient stationary. The extension of the approach to distributed systems can be advantageous at times, particularly when a closed-form solution to the differential eigenvalue problem proves elusive, and one must be content with an approximate solution.

Under consideration is a self-adjoint eigenvalue problem defined by the differential equation

$$Lw(x, y) = \lambda m(x, y)w(x, y), \qquad \lambda = \omega^2, \qquad x, y \text{ in } D \tag{7.387}$$

where L is a differential operator of order $2p$, and the boundary conditions

$$B_i w(x, y) = 0, \qquad i = 1, 2, \ldots, p, \qquad x, y \text{ on } S \tag{7.388}$$

in which B_i are boundary differential operators of maximum order $2p - 1$. Multiplication of Eq. (7.387) by w and integration over D yields

$$\int_D wLw\, dD = \lambda \int_D mw^2\, dD \tag{7.389}$$

which can be rewritten as

$$R(w) = \lambda = \omega^2 = \frac{\displaystyle\int_D wLw\, dD}{\displaystyle\int_D mw^2\, dD} \tag{7.390}$$

Equation (7.390) represents *Rayleigh's quotient* for a distributed system. Clearly, if w is an eigenfunction, say w_r, then Rayleigh's quotient is the associated eigenvalue λ_r. In general, w can be regarded as a trial function, and the question arises as to the behavior of Rayleigh's quotient as w varies over the κ_B^{2p} space (Sec. 7.5). To answer this question, we invoke the expansion theorem, Eq. (7.101), and write

$$w = \sum_{r=1}^{\infty} c_r w_r \tag{7.391}$$

where w_r $(r = 1, 2, \ldots)$ are orthonormal eigenfunctions satisfying Eqs. (7.96). Introducing Eq. (7.391) into Eq. (7.390) and using the orthonormality relations (7.96), we obtain

$$R(c_1, c_2, \ldots) = \frac{\displaystyle\sum_{r=1}^{\infty}\sum_{s=1}^{\infty} c_r c_s \int_D w_r L w_s\, dD}{\displaystyle\sum_{r=1}^{\infty}\sum_{s=1}^{\infty} c_r c_s \int_D m w_r w_s\, dD} = \frac{\displaystyle\sum_{r=1}^{\infty} c_r^2 \lambda_r}{\displaystyle\sum_{r=1}^{\infty} c_r^2} \tag{7.392}$$

and we note that Rayleigh's quotient is now a function of the coefficients $c_1, c_2, \ldots$. Hence, the first variation in Rayleigh's quotient is simply

$$\delta R = \sum_{i=1}^{\infty} \frac{\partial R}{\partial c_i} \delta c_i \tag{7.393}$$

If the first variation δR vanishes, then Rayleigh's quotient has a stationary value. Because the coefficients c_i are all independent, the stationarity conditions are

$$\frac{\partial R}{\partial c_i} = 0, \qquad i = 1, 2, \ldots \tag{7.394}$$

Introducing Eq. (7.392) into Eqs. (7.394), we obtain

$$\begin{aligned}\frac{\partial R}{\partial c_i} &= \frac{\left(\sum_{r=1}^{\infty} 2c_r \frac{\partial c_r}{\partial c_i} \lambda_r\right) \sum_{r=1}^{\infty} c_r^2 - \left(\sum_{r=1}^{\infty} 2c_r \frac{\partial c_r}{\partial c_i}\right) \sum_{r=1}^{\infty} c_r^2 \lambda_r}{\left(\sum_{r=1}^{\infty} c_r^2\right)^2} \\ &= \frac{2c_i \lambda_i \sum_{r=1}^{\infty} c_r^2 - 2c_i \sum_{r=1}^{\infty} c_r^2 \lambda_r}{\left(\sum_{r=1}^{\infty} c_r^2\right)^2} = \frac{2c_i \sum_{r=1}^{\infty} (\lambda_i - \lambda_r) c_r^2}{\left(\sum_{r=1}^{\infty} c_r^2\right)^2} = 0, \\ & \qquad i = 1, 2, \ldots \end{aligned} \tag{7.395}$$

If w coincides with one of the eigenfunctions, say $w = w_i$, then $c_r = c_i \delta_{ir}$ $(r = 1, 2, \ldots)$, where δ_{ir} is the Kronecker delta. It follows that every term in the series at the numerator of (7.395) is zero, except for the term corresponding to $r = i$, and this latter term vanishes because $\lambda_i - \lambda_r = 0$ for $r = i$. Hence, conditions (7.394) are satisfied when the trial function coincides with an eigenfunction, so that *Rayleigh's quotient has stationary points at the system eigenfunctions*. These are the only stationary points of Rayleigh's quotient. Letting $w = c_j w_j$ in Eq. (7.392), we conclude that

$$R(w_j) = \lambda_j, \qquad j = 1, 2, \ldots \tag{7.396}$$

so that *the stationary values of Rayleigh's quotient are precisely the system eigenvalues*. These results are to be expected, as they constitute *Rayleigh's principle* for self-adjoint distributed systems and they represent the counterpart of Rayleigh's principle demonstrated in Sec. 5.2 for symmetric discrete systems.

The variational characterization of the eigenvalue problem just presented is equivalent to a certain form of the differential eigenvalue problem, Eqs. (7.387) and (7.388). To show this, we multiply Eq. (7.387) by an admissible function v, integrate over the domain D and write

$$\int_D vLw \, dD = \lambda \int_D mvw \, dD \tag{7.397}$$

Then, integrating the left side of Eq. (7.397) by parts with due consideration to the boundary conditions, Eqs. (7.388), we can write the result in the form

$$[v, w] = \lambda \left(\sqrt{m}v, \sqrt{m}w\right) \tag{7.398}$$

where $[v, w]$ is an energy inner product, Eq. (7.84), and $\left(\sqrt{m}v, \sqrt{m}w\right)$ is a weighted inner product. Equation (7.398) represents the *weak form* of the eigenvalue problem and can be stated as follows: *Determine a scalar λ and a function w in the admissible space κ_G^p such that Eq. (7.398) is satisfied for all v in κ_G^p.*

Next, we integrate the numerator of Rayleigh's quotient, Eq. (7.390), by parts and rewrite the quotient as

$$R(w) = \lambda = \frac{[w, w]}{\left(\sqrt{m}w, \sqrt{m}w\right)} \tag{7.399}$$

Then, we consider an admissible function in the neighborhood of w and write it in the form $w + \epsilon v$, where v is a function from κ_G^p and ϵ is a small parameter. Replacing w in Eq. (7.399) by the varied function $w + \epsilon v$ and carrying out a binomial expansion of the denominator, we can write

$$\begin{aligned} R(w + \epsilon v) & \\ &= \frac{[w + \epsilon v, w + \epsilon v]}{\left(\sqrt{m}\,(w + \epsilon v)\,, \sqrt{m}\,(w + \epsilon v)\right)} \\ &= \frac{[w, w] + 2\epsilon\,[v, w] + \epsilon^2\,[v, v]}{\left(\sqrt{m}w, \sqrt{m}w\right) + 2\epsilon\left(\sqrt{m}v, \sqrt{m}w\right) + \epsilon^2\left(\sqrt{m}v, \sqrt{m}v\right)} \\ &= R(w) + 2\epsilon\frac{[v, w]\left(\sqrt{m}w, \sqrt{m}w\right) - [w, w]\left(\sqrt{m}v, \sqrt{m}w\right)}{\left(\sqrt{m}w, \sqrt{m}w\right)^2} + O\left(\epsilon^2\right) \\ &= R(w) + 2\epsilon\frac{[v, w] - \lambda\left(\sqrt{m}v, \sqrt{m}w\right)}{\left(\sqrt{m}w, \sqrt{m}w\right)} + O\left(\epsilon^2\right) \end{aligned} \tag{7.400}$$

For a given function v, $R\,(w + \epsilon v)$ depends only on ϵ. If the linear term in ϵ in Eq. (7.400) is zero, i.e., *if the first variation of R vanishes, then R has a stationary value at w*. For this to happen, the coefficient of ϵ must be zero, which is the same as satisfying Eq. (7.398). Hence, *rendering Rayleigh's quotient stationary is equivalent to solving the weak form of the eigenvalue problem.*

In our discussions of the differential eigenvalue problem we assumed implicitly that the eigenvalues are ordered so as to satisfy $\lambda_1 \leq \lambda_2 \leq \ldots$. Then, from Eq. (7.392), it is easy to see that *Rayleigh's quotient is an upper bound for the lowest eigenvalue*, or

$$R(w) \geq \lambda_1 \tag{7.401}$$

which also implies that the minimum value Rayleigh's quotient can take is λ_1, or

$$\lambda_1 = \min R(w) \tag{7.402}$$

Equation (7.402) is very important for two reasons. In the first place, because λ is proportional to ω^2, it characterizes the lowest natural frequency ω_1, which is the most important one. Then, the fact that λ_1 is a minimum value, as opposed to a mere stationary value, Eq. (7.402) forms the basis for certain methods for the computation of approximate solutions to the differential eigenvalue problem. In view of this, Eq. (7.402) alone is often referred to as *Rayleigh's principle.*

If the lower s eigenfunctions w_i are known, then the lower $s+1$ eigenvalues can be characterized *by constraining the trial function w to be orthogonal to w_i $(i = 1, 2, \ldots, s)$, in which case Rayleigh's quotient is an upper bound for λ_{s+1}*, or

$$R(w) \geq \lambda_{s+1}, \qquad (w, w_i) = 0, \qquad i = 1, 2, \ldots, s \tag{7.403}$$

This characterization is primarily of academic interest, as the eigenfunctions w_i $(i = 1, 2, \ldots, s)$ are not available.

By analogy with the approach used in Sec. 5.3, a characterization of λ_{s+1} independent of the eigenfunctions w_i $(i = 1, 2, \ldots, s)$ can be obtained by constraining the trial function w to be orthogonal to s independent, but otherwise arbitrary, functions v_i $(i = 1, 2, \ldots, s)$ and writing

$$\lambda_{s+1} = \max_{v_i} \min_{w} R(w), \qquad (w, v_i) = 0, \qquad i = 1, 2, \ldots, s \tag{7.404}$$

The *maximum-minimum characterization of the eigenvalues*, Eq. (7.404), represents the *Courant and Fischer maximin theorem for distributed systems* and can be stated as follows: *The eigenvalue λ_{s+1} of the system described by Eqs. (7.387) and (7.388) is the maximum value that can be given to* $\min R(w)$ *by the imposition of the s constraints* $(w, v_i) = 0$ $(i = 1, 2, \ldots, s)$, where the maximum is with respect to all sets containing $v_1, v_2, \ldots, v_s$ and the minimum is with respect to all functions in κ_G^p satisfying the imposed constraints.

The geometric interpretation of the stationarity of Rayleigh's quotient and of the Courant and Fischer maximin theorem for distributed systems is similar to that for discrete systems given in Secs. 5.2 and 5.3, respectively.

The variational approach presented in this section forms the basis for the classical Rayleigh-Ritz method and the finite element method for generating approximate solutions to the differential eigenvalue problem.

7.15 INTEGRAL FORMULATION OF THE EIGENVALUE PROBLEM

Up to this point, we have formulated the eigenvalue problem for distributed systems as a differential problem, consisting of one (or two) differential equation(s) and an appropriate number of boundary conditions. The differential formulation emerges naturally from the boundary-value problem, which is how the motion of distributed systems is described almost exclusively. The eigenvalue problem for distributed systems, however, can also be described in integral form. Whereas the differential form remains the preferred choice, a discussion of the integral form should prove rewarding.

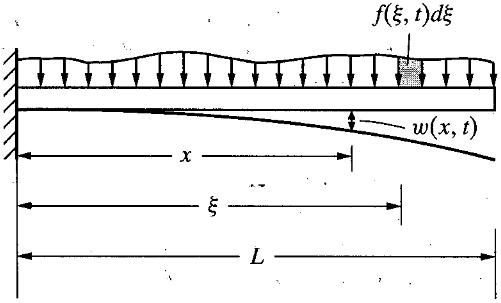


Figure 7.24 Cantilever beam in bending

The integral form of the eigenvalue problem is based on the concept of *flexibility influence function.* To introduce the concept, we consider a self-adjoint system, such as the cantilever beam of Fig. 7.24, and *define the flexibility influence function $a(x, \xi)$ as the displacement at point x due to a unit force at point ξ.* Then, the total displacement at point x due to the entire distributed force is simply

$$w(x, t) = \int_0^L a(x, \xi) f(\xi, t)\, d\xi \tag{7.405}$$

But, because displacements of elastic members increase linearly with forces, the system potential energy can be written as

$$V(t) = \frac{1}{2}\int_0^L w(x, t) f(x, t)\, dx = \frac{1}{2}\int_0^L \int_0^L a(x, \xi) f(\xi, t) f(x, t)\, dx\, d\xi \tag{7.406}$$

The potential energy expression can also be derived beginning with the displacement $w(\xi, t)$ instead of $w(x, t)$. Because the potential energy must be the same, irrespective of how it is derived, we conclude that *the flexibility influence function is symmetric in x and ξ*, or

$$a(x, \xi) = a(\xi, x) \tag{7.407}$$

Equation (7.407) represents *Maxwell's reciprocity theorem* and it states: *The displacement at point x due to a unit force at point ξ is equal to the displacement at point ξ due to a unit force at point x.* The symmetry of the flexibility influence function is consistent with the mathematical symmetry implied by the self-adjointness of the stiffness operator L. The flexibility influence function $a(x, \xi)$ is commonly known as a *Green's function.* Its expression differs from one elastic member to another, which is consistent with the fact that the expression for the differential operator L does. However, it should be observed that, whereas the self-adjointness of L depends not only on the expression for L but also on expressions for the boundary conditions, the boundary conditions are already built into the influence function $a(x, \xi)$. Moreover, whereas L is defined for positive definite as well as positive semidefinite systems, $a(x, \xi)$ is defined for positive definite systems alone.

In free vibration there are no external forces, so that the force density $f(\xi, t)$ in Eq. (7.405) is due entirely to inertial forces. Hence, denoting the mass density by $m(\xi)$, we have

$$f(\xi, t) = -m(\xi)\frac{\partial^2 w(\xi, t)}{\partial t^2} \tag{7.408}$$

But, as established in Sec. 7.4, free vibration of conservative systems is harmonic, so that

$$w(\xi, t) = w(\xi)\cos(\omega t - \phi) \tag{7.409}$$

where $w(\xi)$ is the vibration amplitude, ω the vibration frequency and ϕ an inconsequential phase angle. Hence, inserting Eqs. (7.408) and (7.409) into Eq. (7.405) and dividing through by $\cos(\omega t - \phi)$, we obtain the desired *integral form of the eigenvalue problem*

$$w(x) = \lambda \int_0^L a(x, \xi) m(\xi) w(\xi) d\xi, \qquad \lambda = \omega^2 \tag{7.410}$$

which represents a homogeneous linear integral equation, often referred to as a Fredholm homogeneous linear integral equation of the second kind (Ref. 16). It also represents a linear transformation in which the product $a(x, \xi)m(\xi)$ plays the role of the *kernel* of the transformation.

Equation (7.410) can be generalized by writing

$$w(P) = \lambda \int_D G(P, Q) m(Q) w(Q)\, dD(Q), \qquad \lambda = \omega^2 \tag{7.411}$$

where the position P is defined by one or two spatial coordinates, according to the nature of the problem. The function $G(P, Q)$ is a more general type of influence function, or Green's function. For a self-adjoint system, Green's function is symmetric in P and Q, $G(P, Q) = G(Q, P)$. The kernel $G(P, Q)m(Q)$ of the integral transformation (7.411) is not symmetric, unless $m(Q)$ is constant. It can be symmetrized, however, by introducing the function

$$v(P) = m^{1/2}(P) w(P) \tag{7.412}$$

and multiplying both sides of (7.411) by $m^{1/2}(P)$ to obtain

$$v(P) = \lambda \int_D K(P, Q) v(Q)\, dD(Q) \tag{7.413}$$

where the kernel $K(P, Q)$ is symmetric,

$$K(P, Q) = G(P, Q) m^{1/2}(P) m^{1/2}(Q) = K(Q, P) \tag{7.414}$$

For certain values λ_i, Eq. (7.413) has nontrivial solutions $v_i(P)$, which are related to the solutions $w_i(P)$ of Eq. (7.411) by Eq. (7.412). The values λ_i are the eigenvalues of the system and $w_i(P)$ are the associated eigenfunctions. Whereas the functions $v_i(P)$ are orthogonal in an ordinary sense, the functions $w_i(P)$ are orthogonal with respect to the function $m(P)$. To show this, we consider two distinct solutions of Eq. (7.413), or

$$v_i(P) = \lambda_i \int_D K(P, Q)v_i(Q)\,dD(Q) \tag{7.415a}$$

$$v_j(P) = \lambda_j \int_D K(P, Q)v_j(Q)\,dD(Q) \tag{7.415b}$$

Multiplying Eq. (7.415a) by $v_j(P)$, integrating over domain D and using Eq. (7.415b), we obtain

$$\begin{aligned}\int_D v_i(P)v_j(P)\,dD(P) &= \lambda_i \int_D v_j(P)\left[\int_D K(P, Q)v_i(Q)\,dD(Q)\right] dD(P)\\ &= \lambda_i \int_D v_i(Q)\left[\int_D K(Q, P)v_j(P)\,dD(P)\right] dD(Q)\\ &= \frac{\lambda_i}{\lambda_j}\int_D v_i(Q)v_j(Q)\,dD(Q)\end{aligned} \tag{7.416}$$

from which we obtain

$$(\lambda_i - \lambda_j)\int_D v_i(P)v_j(P)\,dD(P) = 0 \tag{7.417}$$

For two distinct eigenvalues, we obtain the orthogonality relation

$$\int_D v_i(P)v_j(P)\,dD(P) = 0, \qquad \lambda_i \neq \lambda_j \tag{7.418}$$

Introducing Eq. (7.412) in Eq. (7.418) and normalizing the eigenfunctions, we can write the orthonormality relations

$$\int_D m(P)w_i(P)w_j(P)\,dD(P) = \delta_{ij}, \qquad i, j = 1, 2, \ldots \tag{7.419}$$

where δ_{ij} is the Kronecker delta.

As for the differential eigenvalue problem, there is an expansion theorem concerning the eigenfunctions $w_i(P)$ according to which we represent a function satisfying the boundary conditions and possessing a continuous Lw by the infinite series

$$w(P) = \sum_{i=1}^{\infty} c_i w_i(P) \tag{7.420}$$

where the coefficients c_i are given by

$$c_i = \int_D m(P)w(P)w_i(P)\,dD(P), \qquad i = 1, 2, \ldots \tag{7.421}$$

There are several methods for solving Eq. (7.411). The description of these methods is beyond the scope of this text. We discuss here the iteration method, which is similar in principle to the matrix iteration method using the power method. The iteration process for the first eigenfunction is defined by

$$w_1^{(k+1)}(P) = \int_D G(P, Q)m(Q)w_1^{(k)}(Q)\,dD(Q), \qquad k = 1, 2, \ldots \tag{7.422}$$

To demonstrate convergence of the algorithm, we choose an initial trial function $w_1^{(1)}(P)$, which can be assumed to have the form of the series given by Eq. (7.420), insert it into Eq. (7.422) with $k = 1$, carry out the integration and write

$$\begin{aligned} w_1^{(2)}(P) &= \int_D G(P, Q)m(Q)w_1^{(1)}(Q)\,dD(Q) \\ &= \sum_{i=1}^{\infty} c_i \int_D G(P, Q)m(Q)w_i(Q)\,dD(Q) = \sum_{i=1}^{\infty} c_i \frac{w_i(P)}{\lambda_i} \end{aligned} \tag{7.423}$$

Using $w_1^{(2)}(P)$ as an improved trial function, we obtain

$$w_1^{(3)}(P) = \int_D G(P, Q)m(Q)w_1^{(2)}(Q)\,dD(Q) = \sum_{i=1}^{\infty} c_i \frac{w_i(P)}{\lambda_i^2} \tag{7.424}$$

In general, we have

$$w_1^{(p)}(P) = \sum_{i=1}^{\infty} c_i \frac{w_i(P)}{\lambda_i^{p-1}} \tag{7.425}$$

If the eigenvalues are such that $\lambda_1 < \lambda_2 < \lambda_3 < \ldots$, the first term in the series in Eq. (7.425) becomes increasingly large in comparison with the remaining ones and, as $p \to \infty$, $w_1^{(p)}(P)$ becomes proportional to the first eigenfunction, or

$$\lim_{p\to\infty} w_1^{(p)}(P) = w_1(P) \tag{7.426}$$

where the proportionality constant has been ignored as immaterial. After convergence has been reached, λ_1 is obtained as the ratio of two subsequent trial functions,

$$\lambda_1 = \lim_{p\to\infty} \frac{w_1^{(p)}(P)}{w_1^{(p+1)}(P)} \tag{7.427}$$

In practice, if the iterated functions are normalized by prescribing the value of the function at a given point, and this value is kept the same, then the normalization constant approaches λ_1 as p increases.

To obtain the second mode, we must insist that the trial function $w_2^{(1)}(P)$ used for iteration to the second mode be entirely free of the first mode. To this end, we use the iteration process

$$\varphi_2^{(k+1)} = \int_D G(P, Q)m(Q)w_2^{(k)}(Q)\,dD(Q), \qquad k = 1, 2, \ldots \tag{7.428}$$

and begin the iteration with the first trial function in the form

$$w_2^{(1)}(P) = \varphi_2^{(1)}(P) - a_1 w_1(P) \tag{7.429}$$

where $\varphi_2^{(1)}(P)$ is an arbitrarily chosen function and a_1 is a coefficient determined from the orthogonality requirement by writing

$$\begin{aligned}&\int_D m(P) w_2^{(1)}(P) w_1(P)\, dD(P)\\ &= \int_D m(P)\varphi_2^{(1)}(P) w_1(P)\, dD(P) - a_1 \int_D m(P)\,[w_1(P)]^2\, dD(P) = 0\end{aligned} \tag{7.430}$$

which yields

$$a_1 = \frac{\displaystyle\int_D m(P)\varphi_2^{(1)}(P) w_1(P)\, dD(P)}{\displaystyle\int_D m(P) w_1^2(P)\, dD(P)} \tag{7.431}$$

and if $w_1(P)$ is normalized so that $\int_D m w_1^2\, dD = 1$, then

$$a_1 = \int_D m(P)\varphi_2^{(1)}(P) w_1(P)\, dD(P) \tag{7.432}$$

Introducing $w_2^{(1)}(P)$ in Eq. (7.428) with $k = 1$ and performing the integration, we have

$$\varphi_2^{(2)}(P) = \int_D G(P, Q) m(Q) w_2^{(1)}(Q)\, dD(Q) \tag{7.433}$$

For the next iteration step, we use

$$w_2^{(2)}(P) = \varphi_2^{(2)}(P) - a_2 w_1(P) \tag{7.434}$$

where, for normalized $w_1(P)$, we have

$$a_2 = \int_D m(P)\varphi_2^{(2)}(P) w_1(P)\, dD(P) \tag{7.435}$$

In general, for the pth iteration step, we use

$$w_2^{(p)}(P) = \varphi_2^{(p)}(P) - a_p w_1(P) \tag{7.436}$$

Convergence is achieved when, as $p \to \infty$, $a_p \to 0$ and

$$\lim_{p\to\infty} w_2^{(p)}(P) = w_2(P) \tag{7.437}$$

$$\lambda_2 = \lim_{p\to\infty} \frac{w_2^{(p)}(P)}{w_2^{(p+1)}(P)} \tag{7.438}$$

Similarly, for the third mode, we use the trial function

$$w_3^{(1)}(P) = \varphi_3^{(1)}(P) - a_1 w_1(P) - b_1 w_2(P) \tag{7.439}$$

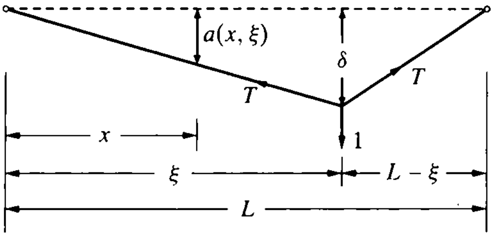


Figure 7.25 Displacement of a uniform string fixed at both ends due to a unit force

where $\varphi_3^{(1)}(P)$ is an arbitrary function and a_1 and b_1 are obtained by insisting that $w_3^{(1)}(P)$ be orthogonal to both $w_1(P)$ and $w_2(P)$. The same procedure is used to iterate to the third mode and, subsequently, to higher modes. In practice, a finite number p of iterations is needed for each mode.

As an illustration, we consider the free vibration of a string of uniformly distributed mass ρ, clamped at both ends and subjected to a constant tension T. The flexibility influence function $a(x, \xi)$ is obtained by applying a unit force at point ξ and calculating the deflection at point x (Fig. 7.25). For small angles α_1 and α_2, the equilibrium condition at the point of application of the load is

$$T\frac{\delta}{\xi} + T\frac{\delta}{L - \xi} = 1 \tag{7.440}$$

from which we obtain the influence function

$$a(x, \xi) = \delta\frac{x}{\xi} = \frac{x(L - \xi)}{TL}, \qquad \xi > x \tag{7.441a}$$

and it can be readily shown that

$$a(x, \xi) = \frac{\xi\,(L - x)}{TL}, \qquad \xi < x \tag{7.441b}$$

As expected, Green's function $G(x, \xi) = a(x, \xi)$ is symmetric in x and ξ, because the system is self-adjoint.

We use the iteration method to solve the eigenvalue problem. To this end, we assume

$$w_1^{(1)} = \frac{x}{L} \tag{7.442}$$

and obtain in sequence

$$\begin{aligned} w_1^{(2)}(x) &= \int_0^L G(x, \xi)\rho(\xi)w_1^{(1)}(\xi)d\xi \\ &= \frac{\rho}{TL^2}\left[\int_0^x \xi(L - x)\xi d\xi + \int_x^L x(L - \xi)d\xi\right] \\ &= \frac{\rho L^2}{6}\left(\frac{x}{L} - \frac{x^3}{L^3}\right) \end{aligned} \tag{7.443a}$$

$$\begin{aligned} w_1^{(3)}(x) &= \int_0^L G(x,\xi)\rho(\xi)w_1^{(2)}(\xi)d\xi \\ &= \frac{\rho L^2}{6T}\frac{7\rho L^2}{60T}\left(\frac{x}{L} - \frac{10}{7}\frac{x^3}{L^3} + \frac{3}{7}\frac{x^5}{L^5}\right) \end{aligned} \tag{7.443b}$$

$$\begin{aligned} w_1^{(4)}(x) &= \int_0^L G(x,\xi)\rho(\xi)w_1^{(3)}(\xi)\,d\xi \\ &= \frac{\rho L^2}{6T}\frac{7\rho L^2}{60T}\frac{31\rho L^2}{294T}\left(\frac{x}{L} - \frac{49}{31}\frac{x^3}{L^3} + \frac{21}{31}\frac{x^5}{L^5} - \frac{3}{31}\frac{x^7}{L^7}\right) \end{aligned} \tag{7.443c}$$

$$\begin{aligned} w_1^{(5)}(x) &= \int_0^L G(x,\xi)\rho(\xi)w_1^{(4)}(\xi)d\xi \\ &= \frac{\rho L^2}{6T}\frac{7\rho L^2}{60T}\frac{31\rho L^2}{294T}\frac{2667\rho L^2}{26040T}\left(\frac{x}{L} - \frac{4340}{2667}\frac{x^3}{L^3} + \frac{2058}{2667}\frac{x^5}{L^5}\right. \\ &\qquad \left. - \frac{420}{2667}\frac{x^7}{L^7} + \frac{35}{2667}\frac{x^9}{L^9}\right) \end{aligned} \tag{7.443d}$$

At this point we pause to check convergence. It can be easily verified that $w_1^{(5)}(x)$, as given by Eq. (7.443d), is almost proportional to $\sin \pi x/L$, which is the first natural mode. Moreover, letting $p = 4$ in Eq. (7.427) (and ignoring the limit), we have

$$\lambda_1 = \omega_1^2 \cong \frac{w_1^{(4)}(L/2)}{w_1^{(5)}(L/2)} = \frac{26040T}{2667\rho L^2} \tag{7.444}$$

so that

$$\omega_1 \cong 3.12\sqrt{\frac{T}{\rho L^2}} \tag{7.445}$$

The approximation is quite good, because the exact value of the first natural frequency is $\omega_1 = \pi\sqrt{T/\rho L^2}$. Hence, Eqs. (7.443d) and (7.445) can be accepted as representing the first natural mode and the first natural frequency, respectively. An interesting aspect of this iteration process is that, although $w_1^{(1)}$ violates the boundary condition at $x = L$, all iterates do satisfy both boundary conditions. It follows that multiplication by Green's function, in conjunction with integration over the domain, imposes the system boundary conditions on the iterates.

To obtain the second mode, we must use a trial function orthogonal to $w_1(x)$. This is left as an exercise to the reader.

It should be pointed out that Green's functions can be determined only for simple systems. Hence, the approach based on Green's functions has limited appeal for distributed-parameter systems. However, the approach proves useful in Sec. 8.1, where we use it for an approximate technique.

7.16 RESPONSE OF UNDAMPED DISTRIBUTED SYSTEMS

Using developments from Secs. 7.1, 7.2 and 7.5, we can write a typical boundary-value problem describing the behavior of vibrating undamped systems in the operator form

$$Lw(P,t) + m(P)\ddot{w}(P,t) = f(P,t), \qquad P \text{ in } D \tag{7.446}$$

where $w(P,t)$ is the displacement of a point P in the domain D, L a linear homogeneous self-adjoint stiffness differential operator of order $2p$, $m(P)$ the mass density and $f(P,t)$ the force density, and we note that any concentrated forces acting at $P = P_j$ can be treated as distributed by means of spatial Dirac delta functions defined by

$$\begin{aligned} \delta\left(P - P_j\right) &= 0, \qquad P \neq P_j \\ \int_D \delta\left(P - P_j\right) dD(P) &= 1 \end{aligned} \tag{7.447}$$

The solution $w(P,t)$ of Eq. (7.446) must satisfy the boundary conditions

$$B_i w(P,t) = 0, \qquad i = 1,2,\ldots,p, \qquad P \text{ on } S \tag{7.448}$$

where B_i are linear homogeneous boundary differential operators ranging in order from zero to $2p-1$ and S is the boundary of D. In addition, the solution is subject to the initial conditions

$$w(P,0) = w_0(P), \qquad \dot{w}(P,0) = v_0(P) \tag{7.449a, b}$$

In a manner analogous to that for discrete systems, the solution to the combined boundary-value problem and initial-value problem can be obtained conveniently by modal analysis. To this end, we must first solve the eigenvalue problem defined by the differential equation

$$Lw(P) = \lambda m(P)w(P), \qquad \lambda = \omega^2, \qquad P \text{ in } D \tag{7.450a}$$

and the boundary conditions

$$B_i w(P) = 0, \qquad i = 1,2,\ldots,p, \qquad P \text{ on } S \tag{7.450b}$$

The solution of Eqs. (7.450) consists of a denumerably infinite set of eigenvalues $\lambda_r = \omega_r^2$, where ω_r are the natural frequencies, and associated eigenfunctions $w_r(P)$ $(r = 1,2,\ldots)$. Because L is self-adjoint, the eigenvalues are real and the eigenfunctions are real and orthogonal. On the assumption that L is positive definite and that the eigenfunctions have been normalized, the orthonormality relations are

$$\int_D m(P)w_r(P)w_s(P)\,dD(P) = \delta_{rs}, \qquad r,s = 1,2,\ldots \tag{7.451a}$$

$$\int_D w_r(P)Lw_s(P)\,dD(P) = \omega_r^2\delta_{rs}, \qquad r,s = 1,2,\ldots \tag{7.451b}$$

where δ_{rs} is the Kronecker delta.

Using the expansion theorem, Eqs. (7.101) and (7.102), we can express the solution of Eq. (7.446) as a linear combination of the system eigenfunctions of the form

$$w(P,t) = \sum_{s=1}^{\infty} w_s(P)\eta_s(t) \tag{7.452}$$

where $\eta_s(t)$ are time-dependent generalized coordinates, referred to as *normal coordinates*, or *modal coordinates*, and playing the role of the expansion coefficients c_r. Strictly speaking, the expansion theorem is in terms of constant coefficients. To resolve this issue, we can conceive of expansion (7.452) being applied at the discrete times $t_1, t_2, \ldots$, resulting in constant coefficients $\eta_s(t_1), \eta_s(t_2), \ldots$. Then, if the times $t_1, t_2, \ldots$ are brought closer and closer together, the coefficients $\eta_s(t_1), \eta_s(t_2), \ldots$ change in a continuous manner with time, thus justifying Eq. (7.452). Inserting Eq. (7.452) into Eq. (7.446), multiplying through by $w_r(P)$, integrating over the domain D and using the orthonormality relations, Eqs. (7.451), we obtain the infinite set of independent equations

$$\ddot{\eta}_r(t) + \omega_r^2\eta_r(t) = N_r(t), \qquad r = 1,2,\ldots \tag{7.453}$$

known as *normal equations*, or *modal equations*, in which

$$N_r(t) = \int_D w_r(P)f(P,t)\,dD(P), \qquad r = 1,2,\ldots \tag{7.454}$$

are *generalized forces*, referred to as *modal forces*. Equations (7.453) are subject to the *initial generalized displacements and velocities*, or *initial modal displacements and velocities* $\eta_r(0)$ and $\dot{\eta}_r(0)$, respectively. They can be obtained by letting $t = 0$ in Eq. (7.452) multiplying through by $m(P)w_r(P)$, integrating over D and considering Eqs. (7.449a) and (7.451a). The result is

$$\eta_r(0) = \int_D m(P)w_r(P)w_0(P)\,dD(P), \qquad r = 1,2,\ldots \tag{7.455a}$$

In a similar fashion, we obtain

$$\dot{\eta}_r(0) = \int_D m(P)w_r(P)v_0(P)\,dD(P), \qquad r = 1,2,\ldots \tag{7.455b}$$

Equations (7.453) are identical in form to the modal equations for discrete systems, Eqs. (4.220), so that the solution is simply

$$\eta_r(t) = \int_0^t \left[\int_0^\tau N_r(\sigma)d\sigma\right]d\tau + \eta_r(0) + \dot{\eta}_r(0)t \tag{7.456a}$$

for rigid-body modes and

$$\eta_r(t) = \frac{1}{\omega_r}\int_0^t N_r(\tau)\sin\omega_r(t-\tau)d\tau + \eta_r(0)\cos\omega_r t + \frac{\dot{\eta}_r(0)}{\omega_r}\sin\omega_r t \tag{7.456b}$$

for elastic modes. The formal solution is completed by inserting Eqs. (7.456) into Eq. (7.452).

The boundary-value problem described by Eqs. (7.446) and (7.448) can be generalized so as to accommodate systems of the type discussed in Secs. 7.8 and 7.9. In particular, the differential equation can be generalized to

$$Lw(P,t) + M\ddot{w}(P,t) = f(P,t), \qquad P \text{ in } D \tag{7.457}$$

where the various quantities are as defined earlier in this section. The one exception is M, which in the case at hand is a linear homogenous self-adjoint mass differential operator of order $2q$, $q < p$, as opposed to a mere function in Eq. (7.446). Moreover, the boundary conditions are

$$B_i w(P,t) = 0, \qquad P \text{ on } S, \quad i = 1, 2, \ldots, k \tag{7.458a}$$

$$B_i w(P,t) + C_i \ddot{w}(P,t) = 0, \qquad P \text{ on } S, \quad i = k+1, k+2, \ldots, p \tag{7.458b}$$

The differential eigenvalue problem corresponding to Eqs. (7.457) and (7.458) is given by the differential equation

$$Lw(P) = \lambda M w(P), \qquad \lambda = \omega^2, \qquad P \text{ in } D \tag{7.459}$$

and the boundary conditions

$$B_i w(P) = 0, \qquad P \text{ on } S, \quad i = 1, 2, \ldots, k \tag{7.460a}$$

$$B_i w(P) = \omega^2 C_i w(P), \qquad P \text{ on } S, \quad i = k+1, k+2, \ldots, p \tag{7.460b}$$

Moreover, we recall from Sec. 7.9 that the orthonormality relations for the system eigenfunctions are given by

$$\int_D w_r(P) M w_s(P)\, dD + \sum_{i=k+1}^{p} \int_S w_r(P) C_i w_s(P)\, dS = \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.461a}$$

$$\int_D w_r(P) L w_s(P)\, dD + \sum_{i=k+1}^{p} \int_S w_r(P) B_i w_s(P)\, dS = \omega_r^2 \delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.461b}$$

Introducing solution (7.452) in Eq. (7.457), multiplying through by $w_r(P)$, integrating over D and using the orthonormality relations, we obtain

$$\begin{aligned} &\sum_{s=1}^{\infty} \left(\delta_{rs} - \sum_{i=k+1}^{p} \int_S w_r C_i w_s\, dS \right) \ddot{\eta}_s(t) \\ &\qquad + \sum_{s=1}^{\infty} \left(\omega_r^2 \delta_{rs} - \sum_{i=k+1}^{p} \int_S w_r B_i w_s\, dS \right) \eta(t) = \int_D w_r f\, dD, \\ &\qquad\qquad r = 1, 2, \ldots \end{aligned} \tag{7.462}$$

But, in view of Eq. (7.454) and boundary conditions (7.458b), Eqs. (7.462) reduce to the set of independent modal equations

$$\ddot{\eta}_r(t) + \omega_r^2 \eta_r(t) = N_r(t), \qquad r = 1, 2, \ldots \tag{7.463}$$

Equations (7.463) are the same as Eqs. (7.453), so that the solution is once again given by Eqs. (7.456). The conclusion is that, in spite of the intimidating appearance of the orthonormality relations, Eqs. (7.461), modal analysis for the solution of the general boundary-value problem, Eqs. (7.457) and (7.458), retains the same simplicity as for the common one, Eqs. (7.446) and (7.448).

Example 7.9

A uniform beam of mass density m, bending stiffness EI and length L is simply supported at both ends. Derive the response to the initial displacement

$$w(x, 0) = w_0(x) = A\left(\frac{x}{L} - 2\frac{x^3}{L^3} + \frac{x^4}{L^4}\right) \tag{a}$$

and note that $w_0(x)$ is symmetric with respect to $x = L/2$. The initial velocity is zero and there are no external forces. The normal modes and natural frequencies of a uniform simply supported beam are

$$w_r(x) = \sqrt{\frac{2}{mL}} \sin\frac{r\pi x}{L}, \qquad \omega_r = (r\pi)^2\sqrt{\frac{EI}{mL^4}}, \qquad r = 1, 2, \ldots \tag{b}$$

The response to the initial displacement, Eq. (a), is given by

$$w(x, t) = \sum_{r=1}^{\infty} w_r(x)\eta_r(t) \tag{c}$$

where, in the absence of initial velocities and external forces, the modal coordinates $\eta_r(t)$ are obtained from Eqs. (7.456) in the form

$$\eta_r(t) = \eta_r(0)\cos\omega_r t, \qquad r = 1, 2, \ldots \tag{d}$$

in which, using Eqs. (a) and (b), we have

$$\eta_r(0) = \int_0^L m(x)w_r(x)w_0(x)\,dx = A\sqrt{\frac{2m}{L}}\int_0^L \sin\frac{r\pi x}{L}\left(\frac{x}{L} - 2\frac{x^3}{L^3} + \frac{x^4}{L^4}\right)dx$$
$$= A\sqrt{2mL}\frac{24}{r^5\pi^5}[1 - (-1)^r], \qquad r = 1, 2, \ldots \tag{e}$$

When r is *even*,

$$\eta_r(0) = 0 \tag{f}$$

and when r is *odd*,

$$\eta_r(0) = \frac{48A}{r^5\pi^5}\sqrt{2mL} \tag{g}$$

Combining the above results, the response to the initial displacement, Eq. (a), is

$$w(x, t) = \frac{96A}{\pi^5}\sum_{r=1}^{\infty}\frac{1}{(2r-1)^5}\sin\frac{(2r-1)\pi x}{L}\cos\omega_r t \tag{h}$$

where

$$\omega_n = [(2r - 1)\pi]^2\sqrt{\frac{EI}{mL^4}}, \qquad r = 1, 2, \ldots \tag{i}$$

Examining Eq. (h), we note that the terms in the series are symmetric with respect to the middle of the beam. This should come as no surprise, because the initial displacement, Eq. (a), is symmetric. In fact, it represents the static deflection caused by a uniformly distributed load. We must also note that the amplitude of the second harmonic is only 0.41% of the amplitude of the first harmonic, so that motion resembles the first mode very closely. This is to be expected, because the initial displacement resembles the first mode.

Example 7.10

An unrestrained uniform rod lies at rest on a smooth horizontal surface (Fig. 7.26). Derive the response to an axial force in the form of a step function of magnitude F_0 applied at $x = 0$.

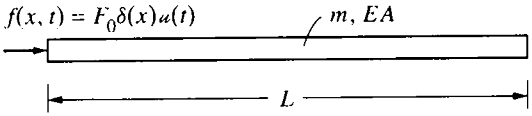


Figure 7.26 Unrestrained uniform rod in axial motion due to a force at $x = 0$

The longitudinal displacement of the rod can be written in the form

$$u(x,t) = \sum_{r=0}^{\infty} U_r(x)\eta_r(t) \tag{a}$$

where, from Eqs. (7.119), (7.137) and (7.138), the normal modes and natural frequencies are

$$\begin{aligned} U_0(x) &= \frac{1}{\sqrt{mL}}, & \omega_0 &= 0 \\ U_r(x) &= \sqrt{\frac{2}{mL}}\cos\frac{r\pi x}{L}, & \omega_r &= r\pi\sqrt{\frac{EA}{mL^2}}, \quad r = 1, 2, \ldots \end{aligned} \tag{b}$$

and we note the presence of one rigid-body mode.

The applied force can be written in the form of a distributed force as follows:

$$f(x,t) = F_0\delta(x)u(t) \tag{c}$$

where $\delta(x)$ is a spatial Dirac delta function applied at $x = 0$ and $u(t)$ is a unit step function applied at $t = 0$. The modal coordinates can be obtained from Eqs. (7.456) in the form

$$\begin{aligned} \eta_0(t) &= \int_0^t \left[\int_0^\tau N_0(\sigma)d\sigma\right] d\tau \\ \eta_r(t) &= \frac{1}{\omega_r}\int_0^t N_r(\tau)\sin\omega_r(t-\tau)d\tau, \quad r = 1, 2, \ldots, \end{aligned} \tag{d}$$

where from Eqs. (7.454), the modal forces are given by

$$N_r(t) = \int_0^L U_r(x)f(x,t)dx = U_r(0)F_0u(t), \quad r = 0, 1, 2, \ldots \tag{e}$$

Hence, using Eqs. (a), (b), (d) and (e), we obtain

$$\begin{aligned} u(x,t) &= U_0(x)U_0(0)F_0 \int_0^t \left[\int_o^\tau u(\sigma)d\sigma \right] d\tau \\ &\qquad + \sum_{r=1}^{\infty} \frac{U_r(x)U_r(0)}{\omega_r} F_0 \int_0^t u(\tau)\sin\omega_r(t-\tau)d\tau \\ &= \frac{F_0}{2mL}t^2 + F_0 \sum_{r=1}^{\infty} \frac{U_r(x)U_r(0)}{\omega_r^2}(1-\cos\omega_r t) \\ &= \frac{F_0}{2mL}t^2 + \frac{2F_0 L}{\pi^2 EA} \sum_{r=1}^{\infty} \frac{1}{r^2}\cos\frac{r\pi x}{L}(1-\cos\omega_r t) \end{aligned} \tag{f}$$

Furthermore,[1]

$$\sum_{r=1}^{\infty} \frac{1}{r^2}\cos\frac{r\pi x}{L} = \frac{\pi^2}{2L^2}\left[\frac{(L-x)^2}{2} - \frac{1}{6}L^2\right] \tag{g}$$

so that the general response is

$$u(x,t) = \frac{1}{2}\frac{F_0}{mL}t^2 + \frac{F_0}{EAL}\left[\frac{(L-x)^2}{2} - \frac{1}{6}L^2\right] - \frac{2F_0 L}{\pi^2 EA}\sum_{r=1}^{\infty}\frac{1}{r^2}\cos\frac{r\pi x}{L}\cos\omega_r t \tag{h}$$

The first term in Eq. (h) represents the rigid-body motion, and it is the only one to survive if the stiffness becomes infinitely large. The second term in Eq. (h) can be looked upon as the static deformation and the third term represents vibration. The first two terms can be interpreted as an average position about which the vibration takes place.

The same system can be looked upon as force-free with a nonhomogeneous boundary condition at the end $x = 0$. This approach is discussed in Sec. 7.19.

Example 7.11

Determine the response of a circular membrane of uniform thickness clamped at $r = a$ and subjected to the distributed force

$$f(r,\theta,t) = \begin{cases} f(t), & 0 \le r \le b \\ 0, & b < r \le a \end{cases} \tag{a}$$

as shown in Fig. 7.27. The membrane is at rest initially.

The response of the membrane can be written in the form of the series

$$\begin{aligned} w(r,\theta,t) &= \sum_{m=0}^{\infty}\sum_{n=1}^{\infty} W_{mn}(r,\theta)\eta_{mn}(t) \\ &= \sum_{n=1}^{\infty} W_{0n}(r)\eta_{0n}(t) + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} W_{mnc}(r,\theta)\eta_{mnc}(t) \\ &\qquad + \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} W_{mns}(r,\theta)\eta_{mns}(t) \end{aligned} \tag{b}$$

[1] See Peirce, B. O. and Foster, R. M. *A Short Table of Integrals*, 4th ed., Ginn, Boston, 1957, Formulas 889 and 891.

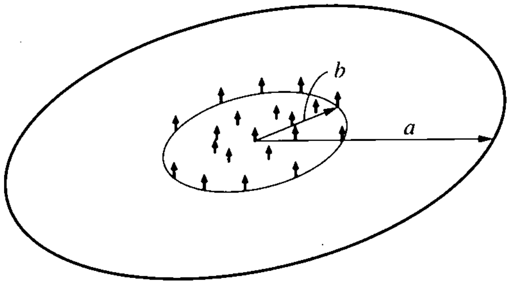


Figure 7.27 Uniform circular membrane fixed at $r = a$ with force distributed over the region $0 \le r \le b < a$

where W_{0n}, W_{nnc} and W_{mns} are the normal modes of a uniform membrane clamped at $r = a$, Eqs. (7.306). Moreover, the modal coordinates are given by

$$\begin{aligned} \eta_{0n}(t) &= \frac{1}{\omega_{0n}} \int_0^t N_{0n}(\tau) \sin \omega_{0n}(t - \tau) d\tau \\ \eta_{mnc}(t) &= \frac{1}{\omega_{mn}} \int_0^t N_{mnc}(\tau) \sin \omega_{mn}(t - \tau) d\tau \\ \eta_{mns}(t) &= \frac{1}{\omega_{mn}} \int_0^t N_{mns}(\tau) \sin \omega_{mn}(t - \tau) d\tau \end{aligned} \tag{c}$$

in which N_{0n}, N_{mnc} and N_{mns} are the modal forces. Inserting Eq. (a) into Eqs. (7.454) and considering Eqs. (7.306), the modal forces take the form

$$\begin{aligned} N_{0n}(t) &= \int_0^{2\pi} \int_0^a W_{0n}(r) f(r, \theta, t) r \, dr \, d\theta = \frac{2\pi f(t)}{\sqrt{\pi \rho} a J_1(\beta_{0n} a)} \frac{b}{\beta_{0n}} J_1(\beta_{0n} b) \\ N_{mnc}(t) &= \int_0^{2\pi} \int_0^a W_{mnc}(r, \theta) f(r, \theta, t) r \, dr \, d\theta = 0 \\ N_{mns}(t) &= \int_0^{2\pi} \int_0^a W_{mns}(r, \theta) f(r, \theta, t) r \, dr \, d\theta = 0 \end{aligned} \tag{d}$$

Hence,

$$\begin{aligned} \eta_{0n}(t) &= \frac{2\sqrt{\pi} b J_1(\beta_{0n} b)}{\sqrt{T} a \beta_{0n}^2 J_1(\beta_{0n} a)} \int_0^t f(\tau) \sin \omega_{0n}(t - \tau) d\tau \\ \eta_{mnc}(t) &= \eta_{mns}(t) = 0 \end{aligned} \tag{e}$$

where $\omega_{0n} = \beta_{0n}\sqrt{T/\rho}$. Introducing Eqs. (7.306) and (e) in the series (b), we obtain the transverse displacement of the membrane in the form

$$w(r, \theta, t) = \frac{2b}{a^2} \sqrt{\frac{1}{T\rho}} \sum_{n=1}^{\infty} \frac{J_1(\beta_{0n} b) J_0(\beta_{0n} r)}{\beta_{0n}^2 J_1^2(\beta_{0n} a)} \int_0^t f(\tau) \sin \omega_{0n}(t - \tau) d\tau \tag{g}$$

It appears that only Bessel functions of zero order, $m = 0$, participate in the motion. This is to be expected, owing to the nature of the load. The load is distributed uniformly over $0 \le r \le b$, so that there cannot be any trigonometric functions present, as modes with diametrical nodes cannot take part in the motion. Furthermore, the Bessel functions of order higher than zero have a zero at $r = 0$, so that, for continuity, they must be antisymmetric with respect to the vertical through $r = 0$, and hence ruled out.

Example 7.12

Obtain the response of a uniform rectangular plate extending over the domain $0 < x < a$, $0 < y < b$ and simply supported along the boundaries $x = 0, a$ and $y = 0, b$ to a concentrated force at the point $x = 3/4a,\ y = 1/2b$, as shown in Fig. 7.28. The plate is at rest initially.

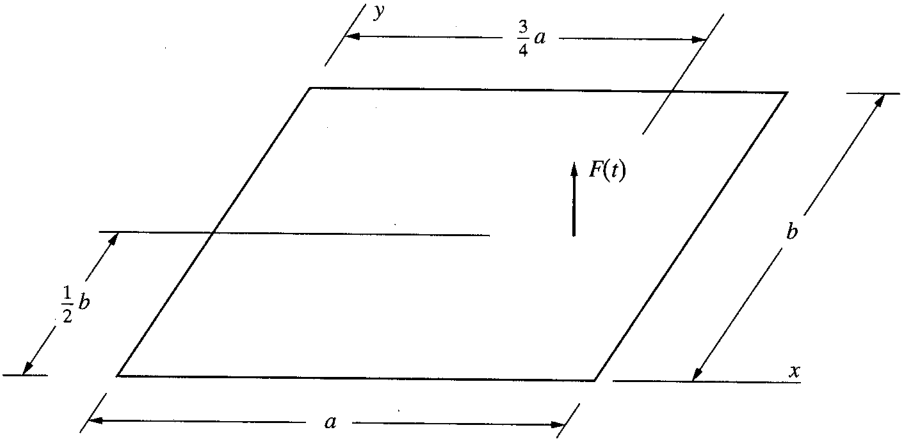


Figure 7.28 Uniform rectangular plate simply supported on all sides and subjected to a concentrated force

The force can be described mathematically as a distributed force given by

$$f(x, y, t) = F(t)\,\delta\left(x - \frac{3}{4}a,\ y - \frac{1}{2}b\right) \tag{a}$$

where $F(t)$ is the time-dependent amplitude of the force and $\delta\,(x - 3a/4,\ y - b/2)$ is a two-dimensional spatial Dirac delta function defined by

$$\delta\left(x - \frac{3}{4}a,\ y - \frac{1}{2}b\right) = 0, \qquad x \neq \frac{3}{4}a \text{ and/or } y \neq \frac{1}{2}b$$
$$\int_0^a \int_0^b \delta\left(x - \frac{3}{4}a,\ y - \frac{1}{2}b\right) dx\,dy = 1 \tag{b}$$

The normal modes of the simply supported uniform plate are

$$W_{mn}(x, y) = \frac{2}{\sqrt{\rho a b}} \sin\frac{m\pi x}{a} \sin\frac{n\pi y}{b}, \qquad m, n = 1, 2, \ldots \tag{c}$$

and the corresponding natural frequencies are

$$\omega_{mn} = \pi^2 \sqrt{\frac{D_E}{\rho}} \left[\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2\right], \qquad m, n = 1, 2, \ldots \tag{d}$$

where D_E is the plate flexural rigidity and ρ is the mass per unit area of plate.

Using the expansion theorem, the transverse displacement of the plate is

$$w(x, y, t) = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} W_{mn}(x, y)\, \eta_{mn}(t) \tag{e}$$

where $\eta_{mn}(t)$ are the modal coordinates having the expressions

$$\eta_{mn}(t) = \frac{1}{\omega_{mn}} \int_0^t N_{mn}(\tau) \sin \omega_{mn}(t - \tau)\, d\tau \tag{f}$$

in which $N_{mn}(t)$ are the modal forces given by

$$\begin{aligned} N_{mn}(t) &= \int_0^a \int_0^b W_{mn}(x, y) f(x, y, t)\, dx\, dy \\ &= \frac{2F(t)}{\sqrt{\rho a b}} \int_0^a \int_0^b \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \delta\left(x - \frac{3}{4}a,\ y - \frac{1}{2}b\right) dx\, dy \\ &= \frac{2F(t)}{\sqrt{\rho a b}} \sin \frac{3m\pi}{4} \sin \frac{n\pi}{2} \end{aligned} \tag{g}$$

Introducing Eq. (g) into Eq. (f), we obtain

$$\eta_{mn}(t) = \frac{2}{\omega_{mn}\sqrt{\rho a b}} \sin \frac{3m\pi}{4} \sin \frac{n\pi}{2} \int_0^t F(\tau) \sin \omega_{mn}(t - \tau) d\tau \tag{h}$$

so that, using Eq. (e), the response can be written in the form

$$w(x, y, t) = \frac{4}{\rho a b} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{\sin(3m\pi/4) \sin(n\pi/2)}{\omega_{mn}} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \times \int_0^t F(\tau) \sin \omega_{mn}(t - \tau) d\tau \tag{i}$$

where the frequencies ω_{mn} are given by Eq. (d).

It can be easily verified that, if m is an integer multiple of 4, the corresponding term in the series in Eq. (i) vanishes. This is consistent with the fact that a concentrated force applied at $x = 3a/4$ cannot excite the modes $\sin 4\pi x/a$, $\sin 8\pi x/a$, etc., which have nodes at that point. The same argument explains why all the terms for which n is an even number vanish.

7.17 DISTRIBUTED GYROSCOPIC SYSTEMS

In many cases of interest, the vibrating structure rotates about a given axis. If the structure has velocity components normal to the axis of rotation, then gyroscopic effects arise. We encountered gyroscopic effects for the first time in connection with discrete systems in Sec. 2.12 and then in Sec. 4.1. In this section, we consider such effects in connection with distributed systems.

We begin the study of distributed gyroscopic systems with the derivation of the boundary-value problem for a rotating elastic shaft simply supported at both ends, as shown in Fig. 7.29. The shaft rotates about axis X with the constant angular velocity Ω relative to the inertial axes XYZ. For convenience, we use a set of body axes xyz, with x coinciding with the rotation axis X, and playing the role of the spatial variable, and with axes y and z rotating together with the body with the same angular velocity Ω about axis $X = x$. Moreover, we let $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ be unit vectors along the rotating axes x, y and z, respectively. The shaft undergoes the bending displacements w_y and w_z in the y and z directions, respectively, so that the angular velocity vector and displacement vector can be written in the vector form

$$\boldsymbol{\omega} = \Omega \mathbf{i} \tag{7.464a}$$

$$\mathbf{w}(x,t) = w_y(x,t)\mathbf{j} + w_z(x,t)\mathbf{k} \tag{7.464b}$$

Then, using the analogy with Eq. (d) of Example 4.1, the velocity vector of a typical point on the shaft can be shown to be

$$\mathbf{v}(x,t) = \left(\dot{w}_y - \Omega w_z\right)\mathbf{j} + \left(\dot{w}_z + \Omega w_y\right)\mathbf{k} \tag{7.465}$$

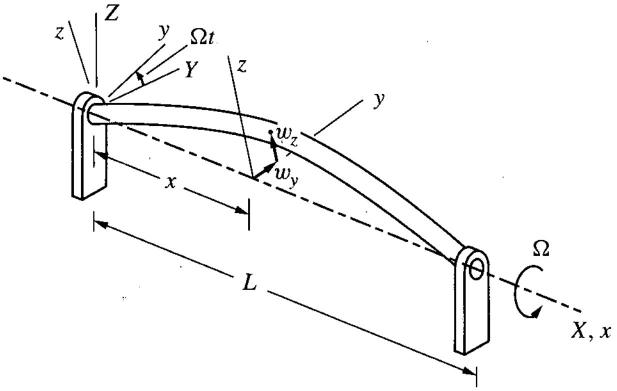


Figure 7.29 Rotating elastic shaft in bending simply supported at both ends

We propose to derive the boundary-value problem by means of Lagrange's equations for distributed systems. This requires an extension of the approach of Sec. 7.3 from systems defined by a single dependent variable to systems defined by two. To this end, we rewrite the extended Hamilton's principle, Eq. (7.31), in the form

$$\int_{t_1}^{t_2} \left(\delta L + \overline{\delta W}_{nc}\right) dt = 0, \qquad \delta w_y = \delta w_z = 0, \qquad t = t_1, t_2 \tag{7.466}$$

where $L = T - V$ is the Lagrangian, in which T is the kinetic energy and V the potential energy, and $\overline{\delta W}_{nc}$ is the virtual work performed by the nonconservative forces. Letting $m = m(x)$ be the mass density and using Eq. (7.465), the kinetic energy has the expression

$$\begin{aligned} T &= \frac{1}{2}\int_0^L m\mathbf{v}^T\mathbf{v}dx = \frac{1}{2}\int_0^L m\left[\left(\dot{w}_y - \Omega w_z\right)^2 + \left(\dot{w}_z + \Omega w_y\right)^2\right]dx \\ &= \int_0^L \hat{T}\left(w_y, w_z, \dot{w}_y, \dot{w}_z\right) dx \end{aligned} \tag{7.467}$$

in which $\hat{T}$ is the kinetic energy density. The potential energy is due to bending alone and has the form

$$V = \frac{1}{2}\int_0^L \left[EI_y\left(w_y''\right)^2 + EI_z\left(w_z''\right)^2\right]dx = \int_0^L \hat{V}\left(w_y'', w_z''\right)dx \tag{7.468}$$

where $\hat{V}$ is the potential energy density, in which EI_y and EI_z are bending stiffnesses. Finally, the virtual work performed by the nonconservative distributed forces is simply

$$\overline{\delta W}_{nc} = \int_0^L \left(f_y\,\delta w_y + f_z\,\delta w_z\right)dx \tag{7.469}$$

The boundary-value problem can be derived by inserting Eqs. (7.467)–(7.469) into Eq. (7.466) and carrying out the customary integrations by parts. All these operations have been performed in Sec. 7.3, however, and need not be repeated. Hence, using the analogy with Eq. (7.41), it is not difficult to show that Lagrange's differential equations of motion are

$$\frac{\partial\hat{L}}{\partial w_y} + \frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w_y''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}_y}\right) + f_y = 0, \qquad 0 < x < L \tag{7.470a}$$

$$\frac{\partial\hat{L}}{\partial w_z} + \frac{\partial^2}{\partial x^2}\left(\frac{\partial\hat{L}}{\partial w_z''}\right) - \frac{\partial}{\partial t}\left(\frac{\partial\hat{L}}{\partial\dot{w}_z}\right) + f_z = 0, \qquad 0 < x < L \tag{7.470b}$$

where $\hat{L} = \hat{T} - \hat{V}$ is the Lagrangian density. Moreover, one boundary condition for w_y and one for w_z must be selected at each end from

$$\frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w_y''}\right)\delta w_y\Big|_0^L = 0, \qquad \frac{\partial}{\partial x}\left(\frac{\partial\hat{L}}{\partial w_z''}\right)\delta w_z\Big|_0^L = 0 \tag{7.471a, b}$$

and one from

$$\frac{\partial\hat{L}}{\partial w_y''}\delta w_y'\Big|_0^L = 0, \qquad \frac{\partial\hat{L}}{\partial w_z''}\delta w_z'\Big|_0^L = 0 \tag{7.471c, d}$$

In view of Eqs. (7.467) and (7.468), the partial differential equations of motion have the explicit form

$$m\ddot{w}_y - 2m\Omega\dot{w}_z - m\Omega^2 w_y + \frac{\partial^2}{\partial x^2}\left(EI_y\frac{\partial^2 w_y}{\partial x^2}\right) = f_y, \quad 0 < x < L \tag{7.472a}$$

$$m\ddot{w}_z + 2m\Omega\dot{w}_y - m\Omega^2 w_z + \frac{\partial^2}{\partial x^2}\left(EI_z\frac{\partial^2 w_z}{\partial x^2}\right) = f_z, \quad 0 < x < L \tag{7.472b}$$

and, because the shaft is simply supported at both ends, the boundary conditions are

$$w_y = 0, \qquad EI_y\frac{\partial^2 w_y}{\partial x^2} = 0, \qquad x = 0, L \tag{7.473a, b}$$

$$w_z = 0, \qquad EI_z\frac{\partial^2 w_z}{\partial x^2} = 0, \qquad x = 0, L \tag{7.473c, d}$$

We observe that the boundary-value problem, Eqs. (7.472) and (7.473), is coupled by the gyroscopic terms alone.

To derive the eigenvalue problem, we let $f_y = f_z = 0$ in Eqs. (7.472) and assume a solution of the resulting homogeneous problem in the form

$$w_y(x,t) = w_y(x)e^{\lambda t}, \qquad w_z(x,t) = w_z(x)e^{\lambda t} \tag{7.474}$$

Inserting Eqs. (7.474) into Eqs. (7.472) and (7.473) and dividing through by $e^{\lambda t}$, we obtain the ordinary differential equations

$$\lambda^2 m w_y - 2\lambda m\Omega w_z + \frac{d^2}{dx^2}\left(EI_y\frac{d^2 w_y}{dx^2}\right) - m\Omega^2 w_y = 0, \quad 0 < x < L \tag{7.475a}$$

$$\lambda^2 m w_z + 2\lambda m\Omega w_y + \frac{d^2}{dx^2}\left(EI_z\frac{d^2 w_z}{dx^2}\right) - m\Omega^2 w_z = 0, \quad 0 < x < L \tag{7.475b}$$

as well as the boundary conditions

$$w_y = 0, \qquad EI_y\frac{d^2 w_y}{dx^2} = 0, \qquad x = 0, L \tag{7.476a, b}$$

$$w_z = 0, \qquad EI_z\frac{d^2 w_z}{dx^2} = 0, \qquad x = 0, L \tag{7.476c, d}$$

Closed-form solutions of the eigenvalue problem are not possible in general.

Next, we assume that the shaft is *uniform*, $m(x) = m =$ constant, $EI_y(x) = EI_y =$ constant and $EI_z(x) = EI_z =$ constant. Moreover, for simplicity, we assume that $I_y = I_z = I$. Then, dividing through by m, the differential equations can be rewritten as

$$\lambda^2 w_y - 2\lambda\Omega w_z + \frac{EI}{m}\frac{d^4 w_y}{dx^4} - \Omega^2 w_y = 0, \qquad 0 < x < L \tag{7.477a}$$

$$\lambda^2 w_z + 2\lambda\Omega w_y + \frac{EI}{m}\frac{d^4 w_z}{dx^4} - \Omega^2 w_z = 0, \qquad 0 < x < L \tag{7.477b}$$

and the boundary conditions reduce to

$$w_y = 0, \qquad \frac{d^2 w_y}{dx^2} = 0, \qquad x = 0, L \tag{7.478a, b}$$

$$w_z = 0, \qquad \frac{d^2 w_z}{dx^2} = 0, \qquad x = 0, L \tag{7.478c, d}$$

It is easy to verify that the eigenfunctions of the system are

$$w_{yj} = a_j \sin\frac{j\pi x}{L}, \qquad w_{zj} = b_j \sin\frac{j\pi x}{L}, \qquad j = 1, 2, \ldots \tag{7.479}$$

Introducing the solutions

$$w_y = \sum_{j=1}^{\infty} a_j \sin \frac{j\pi x}{L}, \qquad w_z = \sum_{j=1}^{\infty} b_j \sin \frac{j\pi x}{L} \tag{7.480}$$

in Eqs. (7.477), multiplying the resulting equations by $\sin k\pi x/L$, integrating over the domain $0 < x < L$ and considering the orthogonality of the eigenfunctions, we obtain the infinite set of pairs of homogeneous algebraic equations

$$\begin{aligned} \left(\lambda^2 + \omega_j'^2 - \Omega^2\right) a_j - 2\lambda\Omega b_j &= 0, \\ & \qquad j = 1, 2, \ldots \\ 2\lambda\Omega a_j + \left(\lambda^2 + \omega_j'^2 - \Omega^2\right) b_j &= 0, \end{aligned} \tag{7.481}$$

where

$$\omega_j' = (j\pi)^2 \sqrt{\frac{EI}{mL^4}}, \qquad j = 1, 2, \ldots \tag{7.482}$$

are recognized as the natural frequencies of the nonrotating shaft. The solution of Eqs. (7.481) can be verified to be

$$\lambda_{2j-1} = i\omega_{2j-1}, \quad \bar{\lambda}_{2j-1} = -i\omega_{2j-1}, \quad \lambda_{2j} = i\omega_{2j}, \quad \bar{\lambda}_{2j} = -i\omega_{2j}, \qquad j = 1, 2, \ldots \tag{7.483a}$$

$$\omega_{2j-1} = \Omega + \omega_j', \quad \omega_{2j} = \Omega - \omega_j', \quad j = 1, 2, \ldots \tag{7.483b}$$

$$b_{2j-1} = ia_{2j-1}, \quad \bar{b}_{2j-1} = -i\bar{a}_{2j-1}, \quad b_{2j} = ia_{2j}, \quad \bar{b}_{2j} = -i\bar{a}_{2j}, \qquad j = 1, 2, \ldots \tag{7.483c}$$

where overbars denote complex conjugates.

The *free vibration solution* can be obtained by inserting Eqs. (7.480) and (7.483) into Eqs. (7.474), with the result

$$w_y(x,t) = \sum_{j=1}^{\infty} \left(a_{2j-1}e^{\lambda_{2j-1}t} + \bar{a}_{2j-1}e^{\bar{\lambda}_{2j-1}t} + a_{2j}e^{\lambda_{2j}t} + \bar{a}_{2j}e^{\bar{\lambda}_{2j}t}\right) \sin \frac{j\pi x}{L} \tag{7.484a}$$

$$w_z(x,t) = \sum_{j=1}^{\infty} \left(b_{2j-1}e^{\lambda_{2j-1}t} + \bar{b}_{2j-1}e^{\bar{\lambda}_{2j-1}t} + b_{2j}e^{\lambda_{2j}t} + \bar{b}_{2j}e^{\bar{\lambda}_{2j}t}\right) \sin \frac{j\pi x}{L} \tag{7.484b}$$

Then, using Eqs. (7.483) and introducing the notation

$$a_{2j-1} = \frac{1}{2}A_{2j-1}e^{-i\phi_{2j-1}}, \qquad a_{2j} = \frac{1}{2}A_{2j}e^{-i\phi_{2j}}, \qquad j = 1, 2, \ldots \tag{7.485}$$

where A_{2j-1} and A_{2j} are real amplitudes and ϕ_{2j-1} and ϕ_{2j} are corresponding phase angles, quantities depending on the initial conditions, the response becomes

$$w_y(x,t) = \sum_{j=1}^{\infty} \left[A_{2j-1} \cos\left(\omega_{2j-1}t - \phi_{2j-1}\right) + A_{2j} \cos\left(\omega_{2j}t - \phi_{2j}\right)\right] \sin \frac{j\pi x}{L} \tag{7.486a}$$

$$w_z(x,t) = -\sum_{j=1}^{\infty} \left[A_{2j-1} \sin\left(\omega_{2j-1}t - \phi_{2j-1}\right) + A_{2j} \sin\left(\omega_{2j}t - \phi_{2j}\right)\right] \sin \frac{j\pi x}{L} \tag{7.486b}$$

Note that the results obtained here are consistent with those for discrete gyroscopic systems obtained in Sec. 4.7.

7.18 DISTRIBUTED DAMPED SYSTEMS

The systems considered in this chapter until now share one characteristic, namely, in the absence of external forces they are conservative. This implies that, if they are excited initially and then allowed to vibrate freely, the vibration will continue ad infinitum. But, conservative systems represent mathematical idealizations, and in practice all systems possess some degree of damping, so that free vibration dies out eventually. Nevertheless, the idealization is useful when damping is very small and the interest lies in time intervals too short for damping effects to become measurable. At this point, however, we consider the case in which damping is not negligible.

As established for discrete systems, and implied above, damping forces are nonconservative. Perhaps the simplest way to account for viscous damping forces is to treat them as a special type of nonconservative forces in the boundary-value problem for undamped systems of Sec. 7.16, Eqs. (7.457) and (7.458). We assume that the damping force at any point P is proportional to the velocity and opposite in direction to the velocity, or

$$f_d(P,t) = -C\dot{w}(P,t) \tag{7.487}$$

where C is a linear homogeneous differential operator of order $2p$. In fact, C is an operator similar to the operator L defined in Sec. 7.5. Inserting Eq. (7.487) into Eq. (7.457) and assuming that the force density $f(P,t)$ includes all nonconservative forces other than damping forces of the type defined by Eq. (7.487), we define the boundary-value problem for viscously damped systems as consisting of the differential equation

$$Lw(P,t) + C\dot{w}(P,t) + M\ddot{w}(P,t) = f(P,t), \qquad P \text{ in } D \tag{7.488}$$

and the boundary conditions

$$B_i w(P,t) = 0, \qquad P \text{ on } S, \qquad i = 1,2,\ldots,k \tag{7.489a}$$

$$B_i w(P,t) + C_i \dot{w}(P,t) = 0, \quad P \text{ on } S, \quad i = k+1, k+2, \ldots, p \tag{7.489b}$$

Closed-form solutions of the boundary-value problem for damped systems, Eqs. (7.488) and (7.489), are not possible in general due to difficulties in solving the eigenvalue problem. Under certain circumstances, however, the eigenfunctions of the undamped system can be used to decouple the modal equations, in a manner similar to that for discrete systems. To this end, we assume a solution of Eq. (7.488) in the form

$$w(P,t) = \sum_{s=1}^{\infty} w_s(P)\eta_s(t) \tag{7.490}$$

where $w_s(P)$ are the eigenfunctions of the undamped system, obtained by letting $C = 0$ in Eq. (7.488). The corresponding eigenvalue problem is given by Eqs. (7.459) and (7.460). Following the approach of Sec. 7.16, we introduce Eq. (7.490) in Eq. (7.488), multiply through by $w_r(P)$, consider the orthonormality relations, Eqs. (7.461), and obtain the modal equations

$$\ddot{\eta}_r(t) + \sum_{s=1}^{\infty} c_{rs}\dot{\eta}_s(t) + \omega_r^2\eta_r(t) = N_r(t), \qquad r = 1, 2, \ldots \tag{7.491}$$

where

$$c_{rs} = \int_D w_r(P)Cw_s(P)\,dD(P), \qquad r, s = 1, 2, \ldots \tag{7.492}$$

are damping coefficients and $N_r(t)$ are modal forces having the form given by Eqs. (7.454).

Equations (7.491) represent an infinite set of coupled ordinary differential equations, so that in general damping produces coupling of the modal equations. In the special case in which the damping operator C can be expressed as a linear combination of the stiffness operator L and mass operator M of the form

$$C = \alpha L + \beta M \tag{7.493}$$

where α and β are constant scalars, the damping coefficients, Eqs. (7.492), can be rewritten as

$$c_{rs} = c_r\,\delta_{rs} = 2\zeta_r\omega_r\,\delta_{rs}, \qquad r, s = 1, 2, \ldots \tag{7.494}$$

in which case Eqs. (7.491) reduce to the independent set

$$\ddot{\eta}_r(t) + 2\zeta_r\omega_r\dot{\eta}_r(t) + \omega_r^2\eta_r(t) = N_r(t), \qquad r = 1, 2, \ldots \tag{7.495}$$

and we note that the notation in Eq. (7.494) was chosen so as to render Eqs. (7.495) similar in structure to the modal equations of viscously damped systems. The solution of Eqs. (7.495) was obtained in Sec. 4.10 in the form of Eqs. (4.229). Damping of the type represented by Eq. (7.493) is known as *proportional damping*.

Several damping models in common use are merely special cases of proportional damping. We distinguish between external and internal damping. External damping generally carries the implication that the mass operator is a mere function, $M = m$, where m is the mass density. In this case, $\alpha = 0$ and the damping operator is assumed to be proportional to the mass density, $C = \beta m = c$, where c represents a viscous damping density function. This is the case of *distributed viscous damping*, a concept that raises as many questions as it answers. Internal damping is based on the assumption that the *material behaves viscoelastically*. A commonly used viscoelastic model is the *Kelvin-Voigt* model, whereby the normal stress is related to the strain and strain rate by

$$\sigma = E(\epsilon + c\dot{\epsilon}) = E\left(\frac{\partial u}{\partial x} + c\frac{\partial^2 u}{\partial t \partial x}\right) \tag{7.496}$$

where E is Young's modulus and c is a given constant. It is easy to verify that the Kelvin-Voigt model represents the distributed counterpart of the spring and dashpot in parallel model used repeatedly in Chapter 3.

In the case of a thin rod in axial vibration, the normal stresses are assumed to be distributed uniformly over the cross-sectional area, so that the axial force is related to the axial displacement by

$$F(x,t) = EA(x)\left[\frac{\partial u(x,t)}{\partial x} + c\frac{\partial^2 u(x,t)}{\partial t \partial x}\right] \tag{7.497}$$

It is not difficult to show that in this case

$$C = cL = -c\frac{\partial}{\partial x}\left(EA\frac{\partial}{\partial x}\right) \tag{7.498}$$

Comparing Eqs. (7.493) and (7.498), we conclude that the Kelvin-Voigt viscoelastic model is indeed a special case of proportional damping, in which $\alpha = c$, $\beta = 0$. Note that, in view of Eq. (7.497), any existing natural boundary conditions must be modified to include the contribution of viscosity to the force.

In the case of an Euler-Bernoulli beam in bending vibration, the assumption that cross-sectional areas remain planar during deformations, when used in conjunction with the Kelvin-Voigt model, implies that the bending moment is related to the bending displacement by

$$M(x,t) = EI(x)\left[\frac{\partial^2 w(x,t)}{\partial x^2} + c\frac{\partial^3 w(x,t)}{\partial t \partial x^2}\right] \tag{7.499}$$

so that once again we encounter a special case of proportional damping. This time

$$C = cL = c\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2}{\partial x^2}\right) \tag{7.500}$$

so that $\alpha = c$, $\beta = 0$. Here too, we must recognize that any existing natural boundary conditions must be modified to include the viscosity effect. In this regard, we should recognize that the shearing force Q is also involved, as it is equal to $-\partial M/\partial x$.

Finally, we consider *structural damping*. As indicated in Sec. 4.10, structural damping can be treated as viscous damping, provided the excitation is harmonic. To emphasize this point, we rewrite Eq. (7.488) in the form

$$Lw(P,t) + C\dot{w}(P,t) + M\ddot{w}(P,t) = f(P)e^{i\Omega t} \tag{7.501}$$

where $f(P)$ is a force density amplitude, generally a complex quantity, and Ω is the driving frequency. But, steady-state response is harmonic, so that

$$\dot{w}(P,t) = i\Omega w(P,t) \tag{7.502}$$

Inserting Eq. (7.502) into Eq. (7.501), we have

$$Lw(P,t) + i\Omega C w(P,t) + M(P)\ddot{w}(P,t) = f(P)e^{i\Omega t} \tag{7.503}$$

By analogy with the assumption made in Sec. 4.10 in connection with discrete systems, according to which the structural damping matrix is proportional to the stiffness matrix, it is customary to assume that the damping operator C is proportional to the stiffness operator L, or

$$C = \frac{\gamma}{\Omega}L \tag{7.504}$$

where γ is a *structural damping factor*. Introducing Eq. (7.504) in Eq. (7.503), we obtain

$$(1+i\gamma)\,Lw(P,t) + M\ddot{w}(P,t) = f(P)e^{i\Omega t} \tag{7.505}$$

Then, using Eq. (7.490) and following the same procedure as earlier in this section, we obtain the infinite set of decoupled modal equations

$$\ddot{\eta}_r(t) + (1+i\gamma)\,\omega_r^2\eta_r(t) = N_r e^{i\Omega t}, \qquad r = 1,2,\ldots \tag{7.506}$$

where $\omega_r (r = 1, 2, \ldots)$ are the natural frequencies of undamped oscillation and

$$N_r = \int_D w_r(P)f(P)\,dD(P), \qquad r = 1,2,\ldots \tag{7.507}$$

are constant modal force amplitudes. The solution of Eqs. (7.506) is simply

$$\eta_r(t) = \frac{N_r e^{i\Omega t}}{(1+i\gamma)\,\omega_r^2 - \Omega^2}, \qquad r = 1,2,\ldots \tag{7.508}$$

so that, inserting Eqs. (7.508) into Eq. (7.490), we obtain the response

$$w(P,t) = \sum_{r=1}^{\infty} \frac{N_r e^{i\Omega t}}{(1+i\gamma)\omega_r^2 - \Omega^2} w_r(P) \tag{7.509}$$

As in the case of discrete systems, the concept of structural damping should be used judiciously.

7.19 SYSTEMS WITH NONHOMOGENEOUS BOUNDARY CONDITIONS

In Sec. 7.16, we used modal analysis to obtain the response of boundary-value problems consisting of a nonhomogeneous differential equation of motion and homogeneous boundary conditions. According to this method, we first obtain the solution of the homogeneous boundary-value problem, which is done by separating the time and spatial dependence of the solution. This leads to an eigenvalue problem yielding the natural modes and the associated natural frequencies. Then, the solution of the nonhomogeneous differential equation is obtained by means of the expansion theorem in the form of a linear combination of normal modes.

In many cases, the boundary conditions are nonhomogeneous. In general, in these cases the approach of Sec. 7.16 will not work, and a different approach must be adopted. In this section, we modify the approach of Sec. 7.16 so as to enable us to use modal analysis. This modified approach is based on the fact that a boundary-value problem consisting of a homogeneous differential equation with nonhomogeneous boundary conditions can be transformed into a problem consisting of a nonhomogeneous differential equation with homogeneous boundary conditions (Ref. 8, p. 277). The latter problem can be solved by modal analysis. Actually, the approach can be used also when the differential equation is nonhomogeneous, in which case the nonhomogeneity of the differential equation becomes more involved.

We consider a one-dimensional system described by the differential equation of motion

$$Lw\,(x,t) + m\,(x)\,\frac{\partial^2 w\,(x,t)}{\partial t^2} = F(x,t),\ 0 < x < L \tag{7.510}$$

where L is a linear homogeneous differential operator of order $2p$, and by the nonhomogeneous boundary conditions

$$B_i w\,(x,t)\,\big|_{x=0} = e_i(t), \qquad i = 1, 2, \ldots, p \tag{7.511a}$$

$$B_j w\,(x,t)\,\big|_{x=L} = f_j(t), \qquad j = 1, 2, \ldots, p \tag{7.511b}$$

where B_i and B_j are linear homogeneous differential operators of order $2p - 1$ or lower. For simplicity, we assume that the initial conditions are zero, or

$$w(x,0) = 0, \qquad \left.\frac{\partial w(x,t)}{\partial t}\right|_{t=0} = 0 \tag{7.512}$$

It is not difficult to see that the differential equation of motion and the boundary conditions are nonhomogeneous. We attempt a solution of the problem by transforming it into a problem consisting of a nonhomogeneous differential equation with homogeneous boundary conditions. To this end, we assume a solution of the boundary-value problem described by Eqs. (7.510) and (7.511) in the form

$$w(x,t) = v(x,t) + \sum_{i=1}^{p} g_i(x)e_i(t) + \sum_{j=1}^{p} h_j(x)f_j(t) \tag{7.513}$$

where the functions $g_i(x)$ and $h_j(x)$ are chosen to render the boundary conditions for the variable $v(x,t)$ homogeneous. In this manner, we transform the boundary-value problem for the variable $w(x,t)$ into a boundary-value problem for the variable $v(x,t)$. The functions $g_i(x)$ and $h_j(x)$ are not unique and several choices may be acceptable. The corresponding results should be equivalent, however.

Introducing Eq. (7.513) in Eqs. (7.511), we obtain the boundary conditions

$$\begin{aligned} B_r w(x,t)\big|_{x=0} &= B_r v(x,t)\big|_{x=0} + \sum_{i=1}^{p} e_i(t) B_r g_i(x)\big|_{x=0} + \sum_{j=1}^{p} f_j(t) B_r h_j(x)\big|_{x=0} \\ &= e_r(t), \qquad r = 1,2,\dots,p \qquad (7.514a) \end{aligned}$$

$$\begin{aligned} B_s w(x,t)\big|_{x=L} &= B_s v(x,t)\big|_{x=L} + \sum_{i=1}^{p} e_i(t) B_s g_i(x)\big|_{x=L} + \sum_{j=1}^{p} f_j(t) B_s h_j(x)\big|_{x=L} \\ &= f_s(t), \qquad s = 1,2,\dots,p \qquad (7.514b) \end{aligned}$$

The functions $g_i(x)$ and $h_j(x)$ must be chosen so that the boundary conditions for $v(x,t)$ be homogeneous. Examination of Eqs. (7.514) reveals that, to satisfy these conditions, we must have

$$\begin{aligned} B_r g_i(x)\big|_{x=0} &= \delta_{ir}, \\ B_r h_j(x)\big|_{x=0} &= 0, \end{aligned} \quad i,j,r = 1,2,\dots,p \qquad (7.515a)$$

$$\begin{aligned} B_s g_i(x)\big|_{x=L} &= 0, \\ B_s h_j(x)\big|_{x=L} &= \delta_{js}, \end{aligned} \quad i,j,s = 1,2,\dots,p \qquad (7.515b)$$

as well as

$$B_i v(x,t)\big|_{x=0} = 0, \qquad i = 1,2,\dots,p \qquad (7.516a)$$

$$B_j v(x,t)\big|_{x=L} = 0, \qquad j = 1,2,\dots,p \qquad (7.516b)$$

Introducing Eq. (7.513) in Eq. (7.510), we obtain the nonhomogeneous differential equation

$$\begin{aligned} Lv(x,t) + m(x)\frac{\partial^2 v(x,t)}{\partial t^2} = F(x,t) &- \sum_{i=1}^{p} \left[e_i(t) L g_i(x) + \ddot{e}_i(t) m(x) g_i(x) \right] \\ &- \sum_{j=1}^{p} \left[f_j(t) L h_j(x) + \ddot{f}_j(t) m(x) h_j(x) \right] \end{aligned} \qquad (7.517)$$

where $v(x,t)$ is subject to homogeneous boundary conditions, Eqs. (7.516).

Using modal analysis, we first solve the eigenvalue problem consisting of the differential equation

$$Lv(x) = \omega^2 m(x) v(x), \qquad 0 < x < L \qquad (7.518)$$

and the boundary conditions

$$B_i v(x)\big|_{x=0} = 0, \qquad i = 1, 2, \dots p \tag{7.519a}$$

$$B_j v(x)\big|_{x=L} = 0, \qquad j = 1, 2, \dots, p \tag{7.519b}$$

The solution of the eigenvalue problem, Eqs. (7.518) and (7.519), yields an infinite set of natural modes $v_r(x)$ and associated natural frequencies ω_r. The modes are orthogonal and, in addition, we normalize them so as to satisfy the orthonormality relations

$$\int_0^L m(x) v_r(x) v_s(x)\, dx = \delta_{rs}, \qquad r, s = 1, 2, \dots \tag{7.520a}$$

$$\int_0^L v_r(x) L v_s(x)\, dx = \omega_r^2 \delta_{rs}, \qquad r, s = 1, 2, \dots \tag{7.520b}$$

Using the expansion theorem, we assume a solution of Eq. (7.517) in the form

$$v(x, t) = \sum_{s=1}^{\infty} v_s(x) \eta_s(t) \tag{7.521}$$

Introducing Eq. (7.521) in Eq. (7.517), we obtain

$$\begin{aligned}\sum_{s=1}^{\infty} \eta_s(t) L v_s(x) + \ddot{\eta}_s(t) m(x) v_s(x) \\ = F(x, t) - \sum_{i=1}^{p} \left[e_i\,(t)\, L g_i\,(x) + \ddot{e}_i\,(t)\, m\,(x)\, g_i\,(x) \right] \\ - \sum_{j=1}^{p} \left[f_j\,(t)\, L h_j\,(x) + \ddot{f}_j\,(t)\, m\,(x)\, h_j\,(x) \right]\end{aligned} \tag{7.522}$$

and, because $v_s(x)$ and ω_s satisfy Eq. (7.518), Eq. (7.522) reduces to

$$\begin{aligned}\sum_{s=1}^{\infty} \left[\ddot{\eta}_s(t) + \omega_s^2 \eta_s(t) \right] m(x) v_s(x) \\ = F(x, t) - \sum_{i=1}^{p} \left[e_i\,(t)\, L g_i\,(x) + \ddot{e}_i\,(t)\, m\,(x)\, g_i\,(x) \right] \\ - \sum_{j=1}^{p} \left[f_j\,(t)\, L h_j\,(x) + \ddot{f}_j\,(t)\, m\,(x)\, h_j\,(x) \right]\end{aligned} \tag{7.523}$$

Next, we multiply Eq. (7.523) through by $v_r(x)$, integrate with respect to x over the domain, introduce the notation

$$\left.\begin{aligned} G_{ri} &= \int_0^L m(x)v_r(x)g_i(x)\,dx, \\ G_{ri}^* &= \int_0^L v_r(x)Lg_i(x)\,dx, \end{aligned}\right\} \quad i = 1, 2, \ldots, p;\ r = 1, 2, \ldots \tag{7.524a}$$

$$\left.\begin{aligned} H_{rj} &= \int_0^L m(x)v_r(x)h_j(x)\,dx, \\ H_{rj}^* &= \int_0^L v_r(x)Lh_j(x)\,dx, \end{aligned}\right\} \quad j = 1, 2, \ldots, p;\ r = 1, 2, \ldots \tag{7.524b}$$

$$F_r(t) = \int_0^L v_r(x)F(x,t)\,dx, \qquad r = 1, 2, \ldots \tag{7.524c}$$

and obtain the infinite set of independent modal equations

$$\ddot{\eta}_r(t) + \omega_r^2\eta(t) = N_r(t), \qquad r = 1, 2, \ldots \tag{7.525}$$

where the modal forces have the form

$$N_r(t) = F_r(t) - \sum_{i=1}^{p}\left[G_{ri}^* e_i\,(t) + G_{ri}\ddot{e}_i\,(t)\right] - \sum_{i=1}^{p}\left[H_{rj}^* f_j\,(t) + H_{rj}\ddot{f}_j\,(t)\right],$$
$$r = 1, 2, \ldots \tag{7.526}$$

The solution of Eqs. (7.525), for zero initial conditions, is given in the form of the convolution integral

$$\eta_r(t) = \frac{1}{\omega_r}\int_0^t N_r(\tau)\sin\omega_r(t-\tau)d\tau, \qquad r = 1, 2, \ldots \tag{7.527}$$

Equations (7.527), when introduced in Eq. (7.521), yield the solution $v(x,t)$ of the transformed problem and, subsequently, using Eq. (7.513), the solution $w(x,t)$ of the original problem.

Example 7.13

Obtain the solution of the axial vibration problem of a uniform rod clamped at $x = 0$ and with a time-dependent tensile force $P(t)$ at $x = L$.

The differential equation of motion is

$$EA\frac{\partial^2 w(x,t)}{\partial x^2} = m\frac{\partial^2 w(x,t)}{\partial t^2} \tag{a}$$

and the boundary conditions are

$$w(0,t) = 0, \qquad EA\left.\frac{\partial w(x,t)}{\partial x}\right|_{x=L} = P(t) \tag{b}$$

so that we have a homogeneous differential equation with one homogeneous and one nonhomogeneous boundary condition. We assume a solution of Eq. (a) in the form

$$w(x,t) = v(x,t) + h(x)P(t) \tag{c}$$

so that the boundary conditions for $v(x, t)$ are

$$v(0,t) = -h(0)P(t), \qquad EA\frac{\partial v(x,t)}{\partial x}\bigg|_{x=L} = P(t) - P(t)EA\frac{dh(x)}{dx}\bigg|_{x=L} \tag{d}$$

To render boundary conditions (d) homogeneous, we must have

$$h(0) = 0, \qquad EA\frac{dh(x)}{dx}\bigg|_{x=L} = 1 \tag{e}$$

The second of boundary conditions (e) can be written as

$$\frac{dh(x)}{dx} = \frac{1}{EA}u[x - (L - \epsilon)] \tag{f}$$

where $u[x - (L - \epsilon)]$ is a spatial step function and ϵ is a small quantity. In view of the first of Eqs. (e), Eq. (f) has the solution

$$h(x) = \frac{1}{EA}[x - (L - \epsilon)]u[x - (L - \epsilon)] \tag{g}$$

and we note that $h(x)$ is zero over the domain $0 \le x \le L - \epsilon$.

The transformed problem consists of the nonhomogeneous differential equation

$$-EA\frac{\partial^2 v(x,t)}{\partial x^2} + m\frac{\partial^2 v(x,t)}{\partial t^2} = EA\frac{d^2h(x)}{dx^2}P(t) - mh(x)\ddot{P}(t) \tag{h}$$

and the homogeneous boundary conditions

$$v(0,t) = 0, \qquad EA\frac{\partial v(x,t)}{\partial x}\bigg|_{x=L} = 0 \tag{i}$$

The corresponding eigenvalue problem consists of the differential equation

$$-EA\frac{d^2v(x)}{dx^2} = \omega^2 mv(x), \qquad 0 < x < L \tag{j}$$

and the boundary conditions

$$v(0) = 0, \qquad EA\frac{dv(x)}{dx}\bigg|_{x=L} = 0 \tag{k}$$

Its solution was obtained in Sec. 7.6. The eigenfunctions are

$$v_r(x) = \sqrt{\frac{2}{mL}}\sin(2r - 1)\frac{\pi x}{2L}, \qquad r = 1, 2, \ldots \tag{l}$$

and the corresponding natural frequencies are

$$\omega_r = (2r - 1)\frac{\pi}{2}\sqrt{\frac{EA}{mL^2}}, \; r = 1, 2, \ldots \tag{m}$$

Using Eqs. (7.524b) and dropping the second subscript, we write

$$\begin{aligned} H_r &= \int_0^L m(x)v_r(x)h(x)dx = m\int_0^L v_r(x)h(x)\,dx \\ &= \frac{m}{EA}\sqrt{\frac{2}{mL}}\int_0^L \sin(2r - 1)\frac{\pi x}{2L}[x - (L - \epsilon)]u[x - (L - \epsilon)]\,dx = 0, \\ &\qquad\qquad r = 1, 2, \ldots \end{aligned} \tag{n}$$

$$\begin{aligned} H_r^* &= \int_0^L v_r(x)Lh(x)\,dx = -EA\int_0^L v_r(x)\frac{d^2h(x)}{dx^2}dx \\ &= -\sqrt{\frac{2}{mL}}\int_0^L \sin(2r-1)\frac{\pi x}{2L}\cdot\delta[x-(L-\epsilon)]\,dx \\ &= (-1)^r\sqrt{\frac{2}{mL}}\cos(2r-1)\frac{\pi\epsilon}{2L}, \qquad r = 1,2,\dots \qquad \text{(o)} \end{aligned}$$

where $\delta[x-(L-\epsilon)]$ is a spatial Dirac delta function. Because Eq. (a) is homogeneous and so is the first of boundary conditions (b), $F_r(t) = 0$ and $e(t) = 0$, so that Eqs. (7.526) yield the modal forces

$$N_r(t) = -H_r^*P(t) = (-1)^{r-1}\sqrt{\frac{2}{mL}}P(t)\cos(2r-1)\frac{\pi\epsilon}{2L}, \qquad r = 1,2,\dots \qquad \text{(p)}$$

which, when introduced in Eqs. (7.527), give the modal coordinates

$$\eta_r(t) = \frac{(-1)^{r-1}\cos(2r-1)(\pi\epsilon/2L)}{\omega_r}\sqrt{\frac{2}{mL}}\int_0^t P(\tau)\sin\omega_r(t-\tau)d\tau, \quad r = 1,2,\dots \qquad \text{(q)}$$

Finally, using Eq. (7.513), we obtain the desired solution

$$\begin{aligned} w(x,t) &= \sum_{r=1}^{\infty} v_r(x)\eta_r(t) + h(x)P(t) \\ &= \frac{2}{mL}\sum_{r=1}^{\infty}\frac{(-1)^{r-1}\cos(2r-1)(\pi\epsilon/2L)}{\omega_r}\sin(2r-1)\frac{\pi}{2}\frac{x}{L}\int_0^t P(\tau)\sin\omega_r(t-\tau)d\tau \\ &\qquad + \frac{P(t)}{EA}[x-(L-\epsilon)]u[x-(L-\epsilon)] \qquad \text{(r)} \end{aligned}$$

Although the last term in Eq. (r) is zero for $0 \le x \le L-\epsilon$ and small for $L-\epsilon \le x \le L$, it must be retained, because its derivatives are neither zero nor small for $L-\epsilon < x < L$; its presence ensures the satisfaction of the boundary condition at $x = L$.

7.20 SYNOPSIS

Mathematical models are not unique, and the same system can be modeled in different ways. In fact, modeling is more of an art than an exact science. In the first six chapters, we have been concerned with discrete, or lumped-parameter systems and in this chapter with distributed parameter systems. This division is very significant, as the mathematical techniques for the two types of systems differ drastically. Indeed, discrete systems possess a finite number of degrees of freedom and are governed by ordinary differential equations of motion, whereas distributed systems possess an infinite number of degrees of freedom and are described by boundary-values problems, consisting of partial differential equations and boundary conditions. Of course, the methods for solving ordinary differential equations are appreciably different from the methods for solving partial differential equations. Still, the difference between the two classes of models is more of form than substance. Indeed, we recall that discrete and distributed models share many of the same characteristics, so that in fact they exhibit similar behavior. This fact is helpful when distributed systems described by boundary-value problems do not admit closed-form solutions, which are more the rule rather than the exception, and the interest lies in approximating them by discrete systems.

In this chapter, the power of analytical mechanics, and in particular the power of the extended Hamilton's principle in deriving boundary-value problems for distributed-parameter systems, is amply demonstrated. The principle can be used to derive the boundary-value problem for a generic system in the form of a partial Lagrange's differential equation of motion and appropriate boundary conditions. As for discrete systems, the free vibration of conservative distributed systems leads to an eigenvalue problem, this time a differential eigenvalue problem, as opposed to an algebraic one. Conservative distributed systems represent a very large and important class of systems referred to as self-adjoint. The discrete counterpart of self-adjoint systems are systems described by real symmetric matrices. Self-adjoint systems possess real eigenvalues and real and orthogonal eigenfunctions, in the same way real symmetric matrices possess real eigenvalues and real and orthogonal eigenvectors. Not surprisingly, systems described by real symmetric matrices are also referred to at times as self-adjoint. Also by analogy with discrete systems, an expansion theorem exists for distributed systems as well. Undamped strings, rods, shafts and beams are all demonstrated to fall in the class of self-adjoint systems. The inclusion in the case of beam vibration of lumped masses at the boundary and rotatory inertia throughout complicates Lagrange's equation and the boundary conditions, as well as the eigenvalue problem, by causing some boundary conditions to depend on the eigenvalue. If shear deformation effects are also included, then a more accurate beam model, known as a Timoshenko beam, is obtained. Two-dimensional systems introduce the shape of the boundary as a factor affecting greatly the nature of the problem. If the shape of the boundary is relatively simple, such as rectangular or circular, then this factor controls the choice of coordinates used to describe the problem. Closed-form solutions are scarce for two-dimensional problems, even when the shape of the boundary is simple. If the shape is irregular, then closed-form solutions do not exist, and approximate solutions are the only viable alternative. In this chapter, we consider rectangular and circular membranes and plates and present some of the few closed-form solutions possible. Approximate solutions are considered in Chapters 8 and 9.

In the case of self-adjoint systems, the differential eigenvalue problem can also be formulated in a weak form by a variational approach, which amounts to rendering Rayleigh's quotient stationary. This approach provides the foundation for some important approximate techniques whereby distributed-parameter systems are discretized (in the spatial variables) by assuming a solution in the form of a series of admissible functions. Reference is made here to the Rayleigh-Ritz method discussed in Chapter 8, and in its premier form, the finite element method, presented in Chapter 9. The net result is to reduce differential eigenvalue problems to algebraic ones. The integral formulation of the eigenvalue problem also provides the basis for some approximate techniques, but on a much more modest scale.

Distributed-parameter system response follows the same pattern as for discrete systems, namely, solve the differential eigenvalue problem, assume the solution of the boundary-value problem in the form of an infinite series of eigenfunctions multiplied by time-dependent modal coordinates and use the orthogonality of the eigenfunctions to obtain an infinite set of independent ordinary differential equations, known as modal equations. The latter can be solved as usual.

From the above, we conclude that, whereas distributed systems differ from discrete systems in form, the basic ideas remain the same.

PROBLEMS

7.1 The n-degree-of-freedom system of Fig. 7.30 consists of n beads of mass m_i suspended on a string and subjected to the forces F_i $(i = 1, 2, \dots, n)$. The left end of the string is fixed and the right end is supported by a spring of stiffness k. The tension in the portion of the string of length Δx_i between the masses m_i and m_{i+1} is T_i. Derive the differential equations of motion for the transverse vibration of the system. Then, devise a limiting process by letting Δx_i approach zero so as to transform the equations of motion into the boundary-value problem derived in Sec. 7.1.

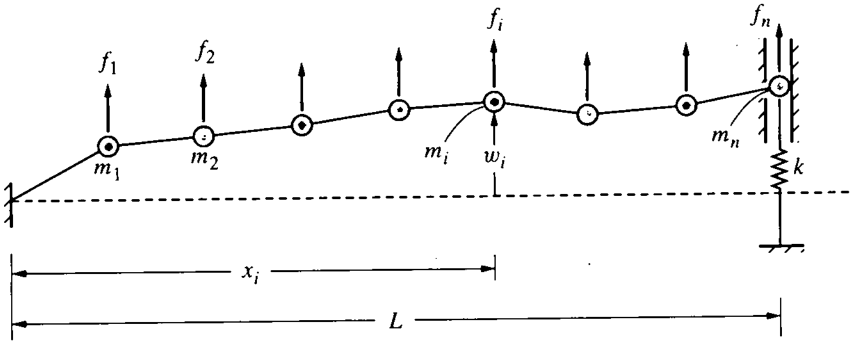


Figure 7.30 String with n masses in transverse vibration

7.2 A nonuniform rod in axial vibration has mass per unit length $m(x)$ and axial stiffness $EA(x)$, in which E is the modulus of elasticity and $A(x)$ the cross-sectional area. The left end is connected to a spring of stiffness k and the right end is free, as shown in Fig. 7.31. The rod is subjected to the force density $f(x, t)$. Derive the boundary-value problem in two ways, first by Newton's second law and then by the extended Hamilton's principle.

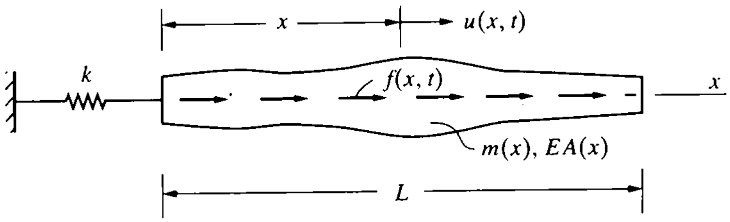


Figure 7.31 Rod in axial vibration connected to a spring at $x = 0$ and free at $x = L$

7.3 A nonuniform shaft in torsional vibration has polar mass moment of inertia per unit length $I(x)$ and torsional stiffness $GJ(x)$, in which G is the shear modulus and $J(x)$ is the area polar moment of inertia. The left end is supported by a torsional spring of stiffness k_1 and the right end by a torsional spring of stiffness k_2, as shown in Fig. 7.32. The shaft is subjected to the moment per unit length $m(x, t)$. Derive the boundary-value problem in two ways, first by Newton's second law and then by the extended Hamilton's principle.

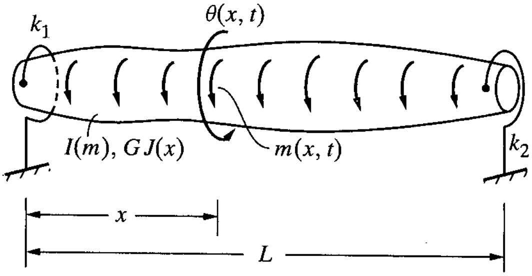


Figure 7.32 Shaft in torsional vibration supported by springs at both ends

7.4 A string of mass per unit length $\rho(x)$ hangs freely from a ceiling, as shown in Fig. 7.33. Derive the boundary-value problem for the transverse vibration of the string by the extended Hamilton's principle. Discuss the boundary condition at the lower end.

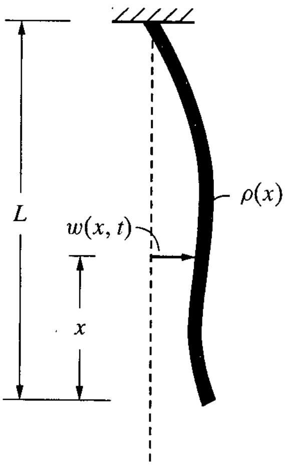


Figure 7.33 String in transverse vibration hanging freely from a ceiling

7.5 A beam of mass per unit length $m(x)$ and bending stiffness $EI(x)$, supported by springs of stiffness k_1 and k_2 at the two ends, is subjected to a distributed force $f(x, t)$, as shown in Fig. 7.34. Derive the boundary-value problem for the bending vibration of the beam in two ways, first by Newton's second law and then by the extended Hamilton's principle.

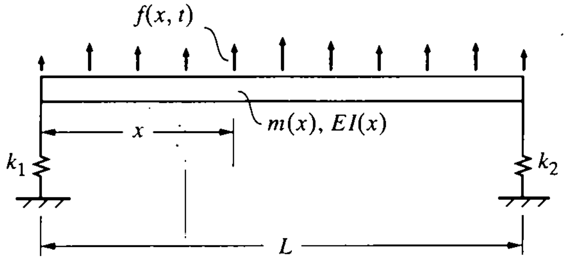


Figure 7.34 Beam in bending vibration supported by springs at both ends

7.6 A beam of mass per unit length $m(x)$ and bending stiffness $EI(x)$, free at both ends, lies on an elastic foundation of distributed stiffness $k(x)$, as shown in Fig. 7.35. Derive the boundary-value problem for the bending vibration of the beam.

$$\frac{\partial^2}{\partial x^2}\left[EI\frac{\partial^2 y}{\partial x^2}\right] + ky = -m\frac{\partial^2 y}{\partial t^2}$$ eq. of motion

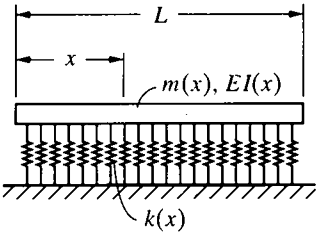


Figure 7.35 Beam in bending free at both ends lying on an elastic foundation

7.7 A beam of mass per unit length $m(x)$ and bending stiffness $EI(x)$, fixed at $x = 0$ and hinged at $x = L$ in a way that the bending slope is restrained by a spring of stiffness k (Fig. 7.36), is acted upon by the distributed force $f(x, t)$. Derive the boundary-value problem for the bending vibration of the beam.

$$\frac{\partial^2}{\partial x^2}\left[EI\frac{\partial^2 y}{\partial x^2}\right] + m\ddot{y} - f = .$$

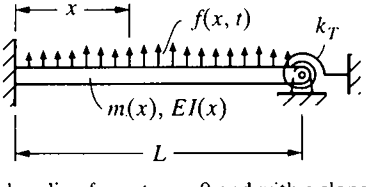


Figure 7.36 Beam in bending free at $x = 0$ and with a slope-restraining spring at $x = L$

7.8 A beam of circular cross section, capable of bending vibration about two orthogonal axes, rotates about the x-axis with the constant angular velocity Ω. The beam has mass per unit length $m(x)$, a disk of mass M at midspan and bending stiffness $EI(x)$, and is hinged at both ends, as shown in Fig. 7.37. Derive the boundary-value problem for the bending vibrations u_y and u_z of the beam about the rotating body axes y and z, respectively. Hint: Note that the lumped mass M of the disk can be treated as a distributed mass having the value $M\,\delta(x - L/2)$, where $\delta(x - L/2)$ is a spatial Dirac delta function.

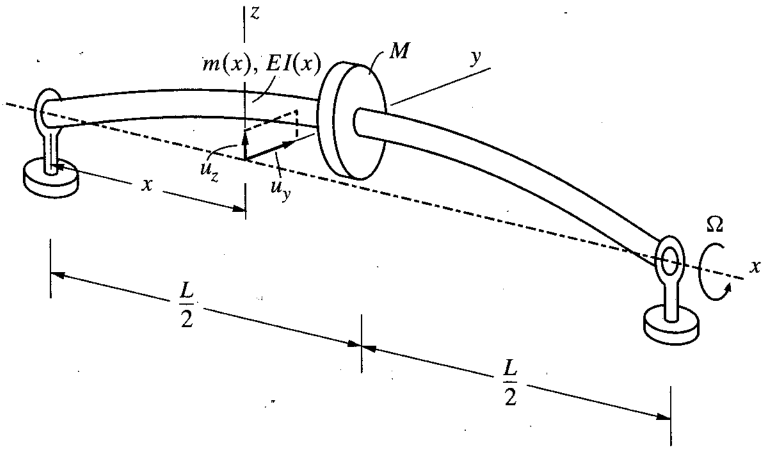


Figure 7.37 Rotating beam in bending hinged at both ends and with a disk at midspan

7.9 Derive the boundary-value problem for the rod of Problem 7.2 by the generic Lagrange equation of Sec. 7.3.

7.10 Derive the boundary-value problem for the shaft of Problem 7.3 by the generic Lagrange equation of Sec. 7.3.

7.11 Derive the boundary-value problem for the beam of Problem 7.5 by the generic Lagrange equation of Sec. 7.3.

7.12 Derive the boundary-value problem for the beam of Problem 7.7 by the generic Lagrange equation of Sec. 7.3.

7.13 Extend the generic formulation of Sec. 7.3 to the case of bending of a beam about two orthogonal axes. Then, use the formulation to derive the boundary-value problem for the rotating beam of Problem 7.8.

7.14 Derive the eigenvalue problem for the rod of Problem 7.2.

7.15 Derive the eigenvalue problem for the shaft of Problem 7.3.

7.16 Derive the eigenvalue problem for the string of Problem 7.4.

7.17 Derive the eigenvalue problem for the beam of Problem 7.5.

7.18 Derive the eigenvalue problem for the beam of Problem 7.6.

7.19 Derive the eigenvalue problem for the beam of Problem 7.7.

7.20 Cast the eigenvalue problem for the beam of Problem 7.14 in the generic form given by Eqs. (7.68) and (7.69), and then check whether the system is self-adjoint and positive definite.

7.21 Solve Problem 7.20 for the shaft of Problem 7.15.

7.22 Solve Problem 7.20 for the string of Problem 7.16.

7.23 Solve Problem 7.20 for the beam of Problem 7.17.

7.24 Solve Problem 7.20 for the beam of Problem 7.18.

7.25 Solve Problem 7.20 for the beam of Problem 7.19.

7.26 Assume that the rod of Problem 7.2 is uniform and solve the eigenvalue problem for the parameter ratio $EA/Lk = 1$. Plot the three lowest modes.

7.27 Assume that the shaft of Problem 7.3 is uniform and solve the eigenvalue problem for the parameters $k_1 = k_2 = k$, $GJ = 2kL$.

7.28 Solve the eigenvalue problem for a rod in axial vibration clamped at $x = 0$ and free at $x = L$ and with the following mass density and axial stiffness:

$$m(x) = 2m(1 - \frac{x}{L}), \qquad EA(x) = 2EA\left(1 - \frac{x}{L}\right)$$

Plot the three lowest modes. Hints: (1) A suitable transformation reduces the differential equation to a Bessel equation and (2) the boundary condition at $x = L$ is the unorthodox one that the displacement must be finite.

7.29 Assume that the mass density of the string of Problem 7.4 is constant and solve the eigenvalue problem. Plot the three lowest modes. Hints: (1) a suitable transformation reduces the differential equation to a Bessel equation and (2) the boundary condition at the lower end is that the displacement must be finite.

7.30 Assume that the beam of Problem 7.5 is uniform and solve the eigenvalue problem for the parameters $k_1 = k,\ k_2 = 2k,\ EI = 10kL^3$. Plot the three lowest modes.

7.31 Assume that the beam and the elastic foundation of Problem 7.6 are uniform and solve the eigenvalue problem. Plot the three lowest modes. Draw conclusions as to the effect of the elastic foundation on the eigensolutions.

7.32 Assume that the beam of Problem 7.7 is uniform and solve the eigenvalue problem for $EI = 5kL$. Plot the three lowest modes.

7.33 Derive the boundary-value problem for the shaft of Example 7.5 by Newton's second law.

7.34 Derive the boundary-value problem for the rotating beam of Example 7.2 by Newton's second law.

7.35 The beam shown in Fig. 7.38 has mass per unit length $m(x)$ and bending stiffness $EI(x)$. The left end is clamped and there is a lumped mass M with mass moment of inertia I_M at the right end. Derive the boundary-value problem by the approach of Sec. 7.4 on the assumption that the rotatory inertia of the beam is negligibly small.

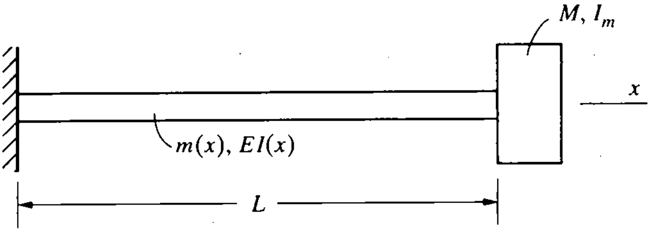


Figure 7.38 Beam in bending clamped at $x = 0$ and with a mass at $x = L$

7.36 Derive the eigenvalue problem for the system of Problem 7.35, verify that it fits the formulation of Sec. 7.9 by identifying the differential operators, and check the system self-adjointness and positive definiteness.

7.37 Assume that the mass and stiffness of the beam in the system of Problem 7.36 are distributed uniformly, solve the eigenvalue problem for the parameters ratios $M/mL = 2 \times 10^{-1}$ and $I_M/mL^3 = 5 \times 10^{-2}$ and plot the three lowest modes.

7.38 A uniform square membrane has repeated natural frequencies $\omega_{mn} = \omega_{nm}$. Any linear combination of the modes W_{mn} and W_{nm} is also a mode. Plot the nodal lines for the mode

$$W(x, y, c) = W_{13}(x, y) + cW_{31}(x, y)$$

for the values $c = 0, \frac{1}{2}, 1$.

7.39 Solve the eigenvalue problem for a uniform rectangular membrane fixed at $x = 0, a$ and free at $y = 0, b$. Assume that there are smooth vertical guides at $y = 0, b$ ensuring that the membrane tension is the same at every point and in every direction.

7.40 Solve the eigenvalue problem for a uniform rectangular membrane supported by a distributed spring of constant stiffness k at the boundaries $x = 0, a$ and fixed at the boundaries $y = 0, b$ for the parameters $b = 2a$, $T = 5ak$. The tension can be assumed to be constant at every point and in every direction, as in Problem 7.39.

7.41 Solve the eigenvalue problem for a uniform circular membrane supported at the boundary $r = a$ by a uniformly distributed spring of stiffness k for the case in which $T = 5ak$.

7.42 Solve the eigenvalue problem for a uniform annular membrane defined over the domain $b < r < a$ and fixed at the boundaries $r = b$ and $r = a$.

7.43 Use Eqs. (7.326) in conjunction with the relations

$$\frac{\partial}{\partial n} = \cos\phi\frac{\partial}{\partial x} + \sin\phi\frac{\partial}{\partial y}, \quad \frac{\partial}{\partial s} = -\sin\phi\frac{\partial}{\partial x} + \cos\phi\frac{\partial}{\partial y}, \quad \frac{\delta\phi}{\partial n} = 0, \quad \frac{\partial\phi}{\partial s} = \frac{1}{R}$$

to derive Eqs. (7.336).

7.44 Modify the derivation of the generic boundary-value problem for plate vibration of Sec. 7.13 so as to accommodate the case in which the displacement w at every point of the boundary is restrained by a distributed spring of stiffness k.

7.45 Repeat Problem 7.44 for the case in which the slope $\partial w/\partial n$ at every point of the boundary is restrained by a distributed spring of stiffness k.

7.46 Check the self-adjointness and positive definiteness of a rectangular plate with the following boundaries:

i. simply supported on all sides
ii. clamped on all sides
iii. free on all sides
iv. all sides supported as in Problem 7.44
v. all sides supported as in Problem 7.45
vi. any combination of sides as in cases i through v

7.47 A uniform rectangular plate is simply supported at the boundaries $y = 0, b$ and free at the boundaries $x = 0, a$. Solve the eigenvalue problem and plot the four lowest modes for the sides ratio $a/b = 1.5$.

7.48 Solve the eigenvalue problem for a uniform circular plate simply supported all around.

7.49 Calculate the value of Rayleigh's quotient for a uniform string fixed at both ends using the trial function $w = \frac{x}{L} - \left(\frac{x}{L}\right)^3$ and draw conclusions.

7.50 Calculate the value of Rayleigh's quotient for the string of Problem 7.22 using the trial function $w = \cos \pi x/2L$ and draw conclusions.

7.51 The natural frequencies and modes of vibration of a uniform string fixed at both ends are $\omega_r = r\pi\sqrt{T/\rho L^2}$ and $W_r(x) = \sqrt{2/\rho L}\sin r\pi x/L$ $(r = 1, 2, \ldots)$. Construct two trial functions so as to demonstrate that Rayleigh's quotient has a mere stationary value at the second mode.

7.52 Use Rayleigh's quotient to estimate the lowest natural frequency of the beam of Problem 7.37.

7.53 Use Rayleigh's quotient to estimate the lowest natural frequency of the membrane of Problem 7.41.

7.54 Use Rayleigh's quotient to estimate the lowest natural frequency of a uniform rectangular plate clamped on all sides.

7.55 Formulate the eigenvalue problem for a uniform rod in axial vibration fixed at $x = 0$ and free at $x = L$ in integral form. Then, use the iteration method of Sec. 7.15 to calculate the lowest eigenvalue.

7.56 Use modal analysis to derive the response of a uniform string fixed at both ends to the initial displacement shown in Fig. 7.39. Discuss the mode participation in the response.

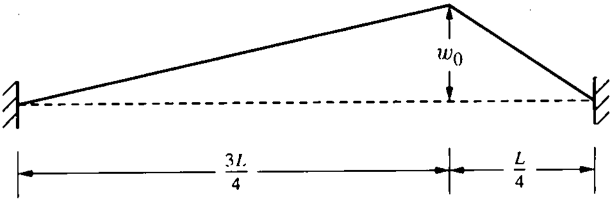


Figure 7.39 Initial displacement of a uniform string fixed at both ends

7.57 Derive the response of a uniform beam clamped at both ends to the initial velocity

$$v_0(x,0) = c\left[\frac{x}{L} - 2\left(\frac{x}{L}\right)^3 + \left(\frac{x}{L}\right)^4\right]$$

and discuss the mode participation.

7.58 A force $F(t)$ traveling on a bridge in the positive x direction at the constant velocity v can be treated as distributed by writing

$$f(x,t) = \begin{cases} F(t)\,\delta(x - vt), & 0 \le vt \le L \\ 0 \qquad\qquad\quad , & vt > L \end{cases}$$

where L is the length of the bridge. Derive the response to the traveling force if the bridge has the form of a uniform simply supported beam.

7.59 A concentrated moment of unit magnitude applied in the clockwise sense at $x = a$ can be represented by two unit impulses acting in opposite directions, as shown in Fig. 7.40. This generalized function, denoted by $\delta'(x - a)$, is called a *spatial unit doublet* and has units length^{-2}. Hence, a concentrated moment $M(t)$ acting in the counterclockwise sense at $x = a$ can be represented as the distributed force $f(x,t) = -M(t)\,\delta'(x - a)$. Use the unit doublet concept to determine the response of a uniform simply supported beam to a counterclockwise moment applied at the right end.

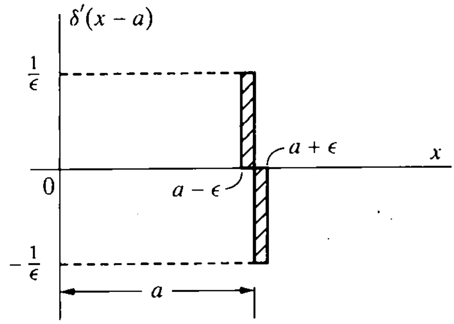


Figure 7.40 Spatial unit doublet at $x = a$

BIBLIOGRAPHY

1. Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, Vol. 1, Wiley, New York, Vol. 1, 1989.
2. Friedman, B., *Principles and Techniques of Applied Mathematics*, Wiley, New York, 1991.
3. Gould, S. H., *Variational Methods for Eigenvalue Problems: An Introduction to the Methods of Rayleigh, Ritz, Weinstein and Aronszajn*, Dover, New York, 1995.
4. Hildebrand, F. B., *Methods of Applied Mathematics*, Prentice Hall, Englewood Cliffs, NJ, 1960.
5. Lanczos, C., *The Variational Principles of Mechanics*, 4th ed., Dover, New York, 1986.
6. Leissa, A. W., *Vibration of Plates*, NASA SP-160, National Aeronautics and Space Administration, Washington, DC, 1969.
7. McLachlan, N. W., *Bessel Functions for Engineers*, Oxford University Press, New York, 1961.
8. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan, New York, 1967.
9. Meirovitch, L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, The Netherlands, 1980.
10. Mindlin, R. D., "Influence of Rotatory Inertia and Shear on Flexural Motions of Isotropic Elastic Plates," *Journal of Applied Mechanics*, Vol. 18, No. 1, 1951, pp. 31-38.
11. Morse, P. M., *Vibration and Sound*, Acoustical Society of America, New York, 1981.
12. Reissner, E., "The Effect of Transverse Shear Deformation on the Bending of Elastic Plates," *Journal of Applied Mechanics*, Vol. 12, 1945, p. A-69.
13. Shames, I. H. and Dym, C. L., *Energy and Finite Element Methods in Structural Mechanics*, McGraw-Hill, New York, 1985.
14. Soedel, W., *Vibrations of Shells and Plates*, 2nd ed., Marcel Dekker, New York, 1993.
15. Timoshenko, S. and Woinowsky-Krieger, S., *Theory of Plates and Shells*, 2nd ed., McGraw-Hill, New York, 1959.
16. Tricomi, F. G., *Integral Equations*, Dover, New York, 1985.

8

APPROXIMATE METHODS FOR DISTRIBUTED-PARAMETER SYSTEMS

Chapter 7 contains a wealth of information concerning the vibration of distributed-parameter systems, including a variety of formulations for boundary-value and differential eigenvalue problems, an all-encompassing discussion of the important class of self-adjoint systems and of the properties of the corresponding eigensolutions, solutions to some differential eigenvalue problems and system response. An overview of Chapter 7 leads to the unmistakable conclusion that it contains a preponderance of problem formulations and discussions of the general properties of the solutions, but only a small number of actual solutions to complex problems. The reason for this paucity of solutions lies in the fact that very few differential eigenvalue problems admit closed-form solutions. Indeed, closed-form solutions are possible only in relatively few cases, almost invariably (but not exclusively) involving uniformly distributed parameters and simple boundary conditions. The satisfaction of boundary conditions can be particularly difficult for two-dimensional problems. In many cases, even though closed-form solutions may be possible, the effort in obtaining them may be so great as to discourage all but the most tenacious investigators. Hence, quite often one must be content with an approximate solution.

The difficulty inherent in the solution of boundary-value problems lies in the dependence on spatial variables. Hence, it should come as no surprise that all approximate methods for distributed-parameter problems have one thing in common, namely, the elimination of the spatial dependence. This amounts to reducing a distributed system to a discrete one through spatial discretization. The spatial discretization methods can be divided into two broad classes, lumping procedures and series discretization methods. Lumping methods are physically motivated, intuitive in character. They all amount to lumping the distributed mass at given points of the domain of the system. On the other hand, the stiffness can be treated as distributed or it can be lumped also. Series discretization methods tend to be more abstract. In such methods, approximate solutions are assumed in the form of series of known space-dependent trial functions multiplied by undetermined coefficients. Certain integrations eliminate the spatial variables and reduce the problem to one of determining these coefficients. In all methods, the net result is to transform differential eigenvalue problems into algebraic ones.

In this chapter, we begin with the *lumped-parameter method using flexibility influence coefficients* whereby, as the name implies, the mass is lumped at discrete points; the stiffness is not lumped but described by means of influence coefficients. The method is applicable to one-dimensional and two-dimensional problems, although it may not be feasible to derive flexibility influence coefficients for the latter. Two other lumped-parameter methods considered are *Holzer's method* for torsional vibration and *Myklestad's method* for bending vibration, and we note that Holzer's method can be adapted to accommodate strings in transverse vibration and rods in axial vibration. These are chain, or step-by-step methods whereby both the mass and the stiffness are lumped and the description of the variables proceeds from one end of the elastic member to the other. The remaining methods in this chapter are all series discretization methods. They can be divided into two classes. The first class is based on variational principles and it amounts to minimization of Rayleigh's quotient. It is identified with the *Rayleigh-Ritz method* and is applicable to self-adjoint systems alone. The second is based on the idea of reducing the error caused by an approximate solution and is known as the *weighted residuals method.* It is in fact not one but a family of methods, the most important one being *Galerkin's method.* The weighted residuals methods are applicable to both self-adjoint and non-self-adjoint systems. Two other methods, component-mode synthesis and substructure synthesis, represent extensions of the Rayleigh-Ritz method to flexible multibody systems.

Conspicuous by its absence from this chapter is the finite element method, a method that rightfully belongs here. Indeed, as the other methods in this chapter, the finite element method seeks approximate solutions to differential eigenvalue problems (and other kinds of problems) not admitting closed-form solutions. Moreover, although the finite element method was believed in the beginning to be entirely new, it was demonstrated later to be another version of the Rayleigh-Ritz method. Still, the procedural details are sufficiently different and the body of literature on the subject has grown to such an extent that presentation of the finite element method in a separate chapter, namely Chapter 9, can be justified.

8.1 LUMPED-PARAMETER METHOD USING FLEXIBILITY INFLUENCE COEFFICIENTS

The lumped-parameter method is arguably the simplest method for the approximate solution of the eigenvalue problem for distributed systems. The approach is based on the integral formulation of the eigenvalue problem, Eq. (7.410). For convenience, we confine ourselves to one-dimensional domains, in which case the eigenvalue problem

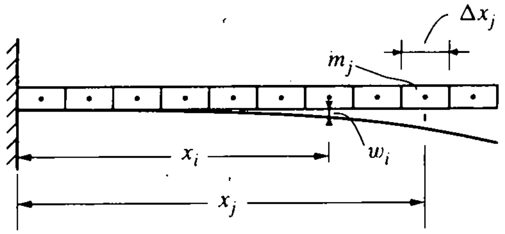


Figure 8.1 Mass lumping in a cantilever beam

has the integral form

$$w(x) = \omega^2 \int_0^L a(x, \xi) m(\xi) w(\xi)\, d\xi \tag{8.1}$$

where $a(x, \xi)$ is the flexibility influence function (Sec 7.15). Next, we consider the system of Fig. 8.1, divide the domain $0 < x < L$ into n small increments of length Δx_j and denote the center of these increments by $\xi = x_j$ $(j = 1, 2, \ldots, n)$. Moreover, we let $x = x_i$, as well as

$$w(x_i) = w_i, \qquad a(x_i, x_j) = a_{ij}, \qquad \int_{\Delta x_j} m(\xi)\, d\xi = m_j \tag{8.2}$$

and approximate Eq. (8.1) by

$$w_i = \omega^2 \sum_{j=1}^{n} a_{ij} m_j w_j, \qquad i = 1, 2, \ldots, n \tag{8.3}$$

Equation (8.3) is the discretized version of Eq. (8.1) and represents an algebraic eigenvalue problem, in which a_{ij} $(i, j = 1, 2, \ldots, n)$ are known as *flexibility influence coefficients* and denote *the displacement at x_i due to a unit force at x_j*. They are simply the discrete counterpart of the flexibility influence function $a(x, \xi)$ introduced in Sec. 7.15. Hence, by analogy, *Maxwell's reciprocity theorem* for lumped systems is

$$a_{ij} = a_{ji}, \qquad i, j = 1, 2, \ldots, n \tag{8.4}$$

indicating that *the flexibility influence coefficients are symmetric in i and j*.

Equations (8.3) can be cast in matrix form. To this end, we introduce the displacement vector $\mathbf{w} = [w_1\ w_2\ \ldots\ w_n]^T$, the *flexibility matrix* $A = [a_{ij}]$ and the *mass matrix* $M = \text{diag}\,(m_j)$, so that Eqs. (8.3) can be rewritten as

$$\mathbf{w} = \omega^2 A M \mathbf{w} \tag{8.5}$$

where we note that AM represents a special case of the dynamical matrix first encountered in Sec. 6.3, in the sense that here M is diagonal. We also note that the flexibility matrix A is the reciprocal of the stiffness matrix K, $A = K^{-1}$. This is of mere academic interest, however, because computation of stiffness coefficients for distributed-parameter systems is not feasible.

The matrix AM is in general not symmetric, and the most efficient computational algorithms are for real symmetric matrices. In this particular case, the problem can be symmetrized with ease by introducing the vector

$$\mathbf{u} = M^{1/2}\mathbf{w} \tag{8.6}$$

where $M^{1/2} = \text{diag}\,(\sqrt{m_j})$. Introducing Eq. (8.6) into Eq. (8.5) and premultiplying by $M^{1/2}/\omega^2$, we obtain an eigenvalue problem in the standard form

$$A'\mathbf{u} = \lambda\mathbf{u}, \qquad \lambda = 1/\omega^2 \tag{8.7}$$

where

$$A' = M^{1/2}AM^{1/2} \tag{8.8}$$

is a real symmetric matrix. The eigenvalue problem given by Eq. (8.7) can be solved by any of the algorithms presented in Chapter 6.

The accuracy of the results depends on the number and length of the increments Δx_j. Conceivably, the length of each individual increment can be varied to reflect the parameter nonuniformity, but in general the increments are taken equal in length, so that the question reduces to the number of increments. Unfortunately, there are no guidelines permitting a rational choice. Hence, whereas we can conclude on physical grounds that the approximations converge to the actual eigensolutions as $n \to \infty$, no quantitative convergence statement can be made.

The main appeal of the method is simplicity of the concepts. In fact, Eqs. (8.3) could have been obtained in a more direct manner by regarding the system as lumped from the beginning. But, whereas the concepts are simple, the implementation is not. The reason can be traced to the fact that, except for some simple cases, the evaluation of the influence coefficients a_{ij} can be quite difficult. This is often the case when the boundary conditions are complicated or when the problem is two-dimensional. Methods for obtaining influence coefficients are covered adequately in many textbooks on mechanics of materials.

Flexibility influence coefficients can be defined only when the potential energy is a positive definite function. Hence, the method just described is so restricted. However, the lumped-parameter method using influence coefficients can be extended to positive semidefinite systems. This amounts to eliminating the rigid-body motions from the formulation. To this end, we introduce a reference frame attached to the body in undeflected configuration and measure elastic displacements relative to the reference frame. The translation and rotation of the reference frame play the role of rigid-body modes, so that the elastic displacements are measured relative to the rigid-body modes. For convenience, we place the origin of the reference frame at the mass center C of the system. As an example, we consider the free vibration of the unrestrained beam in bending shown in Fig. 8.2 and denote the rigid-body translation of the origin C of the reference frame x, y by w_C, the rigid-body rotation of the reference frame by ψ_C, the elastic displacements of m_i relative to x, y by w_i and the total displacement of m_i relative to the inertial frame X, Y by W_i. We assume that the rotation ψ_C is relatively small, so that axes x, y are nearly parallel to axes X, Y. Under these circumstances, the absolute displacement of m_i has the form

$$W_i = w_C + x_i\psi_C + w_i, \qquad i = 1, 2, \ldots, n \tag{8.9}$$

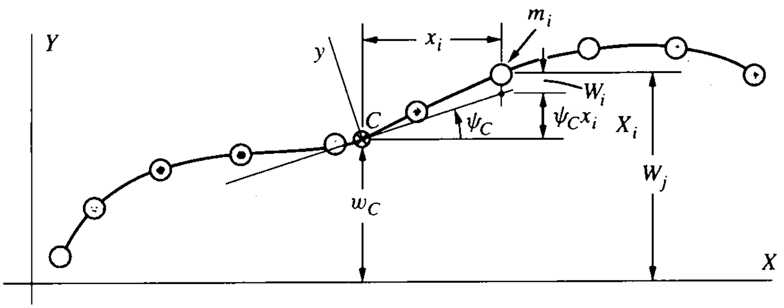


Figure 8.2 Lumped model of an unrestrained beam in bending

where x_i is the nominal position of m_i relative to C. To eliminate the rigid-body motions, we recognize that in free vibration the linear and angular momenta must vanish, or

$$\sum_{i=1}^{n} m_i \dot{W}_i = \sum_{i=1}^{n} m_i \left(\dot{w}_C + x_i \dot{\psi}_C + \dot{w}_i \right) = m \dot{w}_C + \mathbf{1}^T M \dot{\mathbf{w}} = 0 \tag{8.10a}$$

$$\sum_{i=1}^{n} m_i x_i \dot{W}_i = \sum_{i=1}^{n} m_i x_i \left(\dot{w}_C + x_i \dot{\psi}_C + \dot{w}_i \right) = I_C \dot{\psi}_C + \mathbf{x}^T M \dot{\mathbf{w}} = 0 \tag{8.10b}$$

where $m = \sum_{i=1}^{n} m_i$ is the total mass of the beam, $I_C = \sum_{i=1}^{n} m_i x_i^2$ the moment of inertia of the beam about C, $\mathbf{1} = [1\,1\,\ldots\,1]^T$, $\mathbf{x} = [x_1\,x_2\,\ldots\,x_n]^T$, $\dot{\mathbf{w}}$ the elastic velocity vector and M the diagonal mass matrix. Moreover, $\sum_{i=1}^{n} m_i x_i = 0$ by virtue of the fact that C is the mass center of the beam. Equations (8.10) yield simply

$$\dot{w}_C = -\frac{1}{m}\mathbf{1}^T M \dot{\mathbf{w}}, \qquad \dot{\psi}_C = -\frac{1}{I_C}\mathbf{x}^T M \dot{\mathbf{w}} \tag{8.11}$$

Next, we use Newton's second law, recognize that in free vibration all forces are internal and write the equation of motion for each of the lumped masses in the form

$$m_i \ddot{W}_i = m_i \left(\ddot{w}_C + x_i \ddot{\psi}_C + \ddot{w}_i \right) = -f_i = -\sum_{j=1}^{n} k_{ij} w_j, \qquad i = 1, 2, \ldots, n \tag{8.12}$$

where k_{ij} are the stiffness coefficients. Writing Eqs. (8.12) in matrix form and using Eqs. (8.11), we obtain the equations for the elastic motions alone

$$M' \ddot{\mathbf{w}} + K \mathbf{w} = 0 \tag{8.13}$$

in which

$$M' = M - \frac{1}{m} M \mathbf{1}\mathbf{1}^T M - \frac{1}{I_C} M \mathbf{x}\mathbf{x}^T M \tag{8.14}$$

is a modified mass matrix. Because free vibration is harmonic, $\ddot{\mathbf{w}} = -\omega^2\mathbf{w}$, Eq. (8.13) in conjunction with the usual operations yields the eigenvalue problem for the elastic modes

$$AM'\mathbf{w} = \lambda\mathbf{w}, \qquad \lambda = 1/\omega^2 \tag{8.15}$$

where $A = K^{-1}$ is the flexibility matrix, and we note that the stiffness coefficients are not really required. We also note that the flexibility matrix is block-diagonal, as it consists of two independent submatrices on the main diagonal, one for a cantilever beam extending to the right of C and one for a cantilever beam extending to the left of C.

Although the eigenvalue problem (8.15) is of order n, there are only $n-2$ valid solutions. The reason lies in the fact that the modified matrix M' is singular. Indeed, it is not difficult to verify that the rigid-body translation modal vector $\mathbf{1}$ and the rigid-body rotation modal vector $\mathbf{x}$ are in the nullspace (Appendix B) of M', and hence they satisfy $M'\mathbf{1} = \mathbf{0}$ and $M'\mathbf{x} = \mathbf{0}$. Of course, the system has a full complement of n eigenvectors, as the two missing eigenvectors are simply the two rigid-body modal vectors $\mathbf{1}$ and $\mathbf{x}$.

Equation (8.15) yields only the elastic part of the eigenvectors. To recover the contribution of the rigid-body modes to the elastic modes, we consider Eqs. (8.9), (8.11) and (8.14) and write in matrix form

$$\mathbf{W} = \mathbf{1}w_C + \mathbf{x}\psi_C + \mathbf{w} = \left(I - \frac{1}{m}\mathbf{1}\mathbf{1}^T M - \frac{1}{I_C}\mathbf{x}\mathbf{x}^T M\right)\mathbf{w} = M^{-1}M'\mathbf{w} \tag{8.16}$$

Finally, there is the question of lack of symmetry of the matrix AM'. Because M' is singular, the symmetrization process resulting in Eq. (8.8) is not possible. Fortunately, however, the flexibility matrix is positive definite. Hence, using the Cholesky decomposition, we can write

$$A = LL^T \tag{8.17}$$

Introducing Eq. (8.17) into Eq. (8.15), using the linear transformation

$$\mathbf{w} = L\mathbf{u}, \qquad L^{-1}\mathbf{w} = \mathbf{u} \tag{8.18a, b}$$

and premultiplying the result by L, we can reduce the eigenvalue problem to the standard form

$$M^*\mathbf{u} = \lambda\mathbf{u} \tag{8.19}$$

where

$$M^* = L^T M' L = M^{*T} \tag{8.20}$$

is a symmetric matrix. Because M^* is related to M' by an orthogonal transformation, the eigenvalue problem (8.19) retains the characteristics of the eigenvalue problem (8.15), i.e., it possesses only $n-2$ valid solutions. In this regard, we recognize that the nullspace of M^* consists of the two vectors $L^{-1}\mathbf{1}$ and $L^{-1}\mathbf{x}$. Upon solving eigenvalue problem (8.19), we must use Eq. (8.18a) in conjunction with the eigenvectors $\mathbf{u}_r$ to compute the elastic eigenvectors $\mathbf{w}_r$ for the original problem, Eq. (8.15).

Example 8.1

Obtain the natural frequencies and natural modes of vibration associated with the free-free beam shown in Fig. 8.3. The mass and stiffness distributions are

$$m(\xi) = \frac{4}{5}\frac{M}{L}\left(1 + \frac{\xi}{2L}\right), \qquad EI(\xi) = \frac{4}{5}EI\left(1 + \frac{\xi}{2L}\right) \tag{a}$$

where ξ is the distance from the left end, M the total mass and L the length of the beam.

The center of mass of the system is located at a distance $\bar{\xi}$ from the left end given by

$$\bar{\xi} = \frac{1}{M}\int_0^L \xi m(\xi)\,d\xi = \frac{8}{15}L \tag{b}$$

Next, we assume that the mass of the beam is lumped into n discrete masses such that the mass m_i $(i = 1, 2, \ldots, n)$ is equal to the mass in the segment $(i-1)(L/n) \le \xi \le i(L/n)$ and is located at the corresponding center of mass. Hence, the masses m_i have the values

$$m_i = \int_{(i-1)L/n}^{iL/n} m(\xi)\,d\xi = \frac{M}{5n^2}(4n + 2i - 1), \qquad i = 1, 2, \ldots n \tag{c}$$

and are located at distances ξ_i from the left end given by

$$\xi_i = \frac{1}{m_i}\int_{(i-1)L/n}^{iL/n} \xi m(\xi)\,d\xi = \frac{2L}{3n(4n + 2i - 1)}[3n(2i - 1) + 3i(i - 1) + 1], \qquad i = 1, 2, \ldots, n \tag{d}$$

When measured from the center of mass C of the beam, instead of the left end, these locations are

$$x_i = \xi_i - \bar{\xi}, \qquad i = 1, 2, \ldots, n \tag{e}$$

Similarly, in terms of the distance x from C, the stiffness has the form

$$EI(x) = \frac{2}{5}\frac{EI}{L}(2L + \bar{\xi} - x) \tag{f}$$

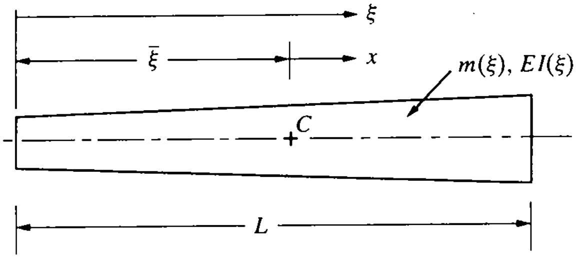


Figure 8.3 Nonuniform free-free beam in bending

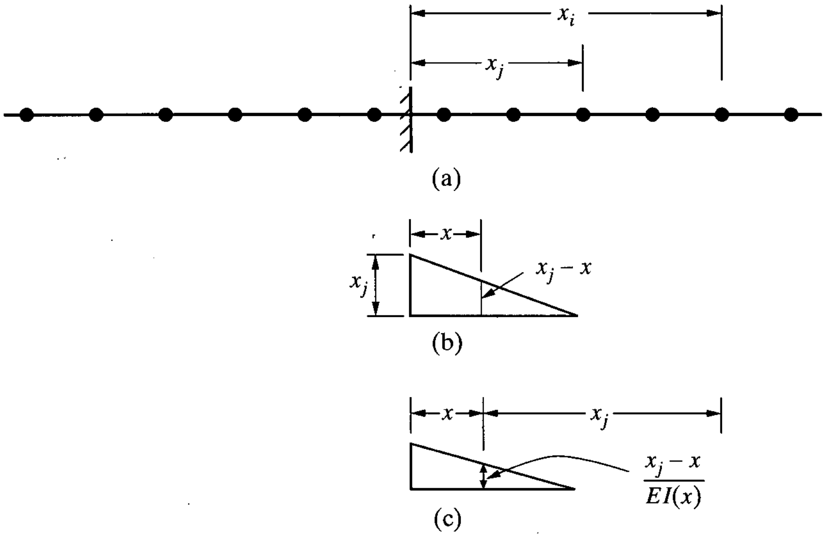


Figure 8.4 **(a)** Lumped-parameter model of a beam cantilevered on both sides **(b)** Bending moment diagram due to a unit force **(c)** Bending moment diagram divided by the bending stiffness

The influence coefficients a_{ij} can be determined by the moment-area method. To this end, we regard the beam as fixed at the mass center and cantilevered on each side, as shown in Fig. 8.4a. Figure 8.4b shows the bending moment diagram due to a unit load applied at $x = x_j$, and Fig. 8.4c shows the bending moment divided by the stiffness $EI(x)$. The slope of the deflection curve is zero at the fixed end, $x = 0$, so the moment-area method gives the deflection at the point x_i due to a unit load at x_j in the form of the moment with respect to point x_i of the area of the bending moment divided by $EI(x)$. Hence, for $x_i > x_j > 0$, we obtain

$$\begin{aligned} a_{ij} &= \int_0^{x_j} \frac{(x_j - x)(x_i - x)}{EI(x)} dx \\ &= \frac{5L}{2EI} \left\{ \left[\left(x_i + x_j + 2L + \bar{\xi} \right) \left(2L + \bar{\xi} \right) + x_i x_j \right] \mathrm{ln} \frac{x_j + 2L + \bar{\xi}}{2L + \bar{\xi}} \right. \\ &\quad \left. - \left(x_i + x_j + 2L + \bar{\xi} \right) x_j - \frac{1}{2} x_j^2 \right\}, \qquad i \leq j \qquad \text{(g)} \end{aligned}$$

and a similar expression can be written for the case in which $x_i < 0$ and $x_j < 0$. Furthermore, the influence coefficients are symmetric, $a_{ij} = a_{ji}$. For $x_i > 0$ and $x_j < 0$, or for $x_i < 0$ and $x_j > 0$, we have

$$a_{ij} = 0 \qquad \text{(h)}$$

The coefficients a_{ij} can be arranged in the form of the symmetric flexibility matrix A. The diagonal mass matrix M is obtained from Eq. (c) and the vector $\mathbf{x}$, as well as the mass moment of inertia I_C, from Eqs. (b), (d) and (e). This, in turn, allows us to evaluate the modified mass matrix M' according to Eq. (8.14). The natural frequencies and natural modes are obtained by solving the eigenvalue problem, Eq. (8.15). The solution consists of 18 eigenvalues ω_r^2 and purely elastic eigenvectors $\mathbf{w}_r$ $(r = 3, 4, \ldots, 20)$, i.e., excluding the contribution from the rigid-body modes. The eigenvectors can be inserted into Eq. (8.16) to obtain the absolute natural modes $\mathbf{W}_r$ ($r = 3, 4, \dots, 20$), in the sense that they are measured relative to the inertial space. It should be recalled that the remaining two modes are the rigid-body modes $\mathbf{W}_1 = \mathbf{1}$ and $\mathbf{W}_2 = \mathbf{x}$ with corresponding natural frequencies equal to zero. The elastic modes $\mathbf{W}_3$, $\mathbf{W}_4$ and $\mathbf{W}_5$ are displayed in Fig. 8.5.

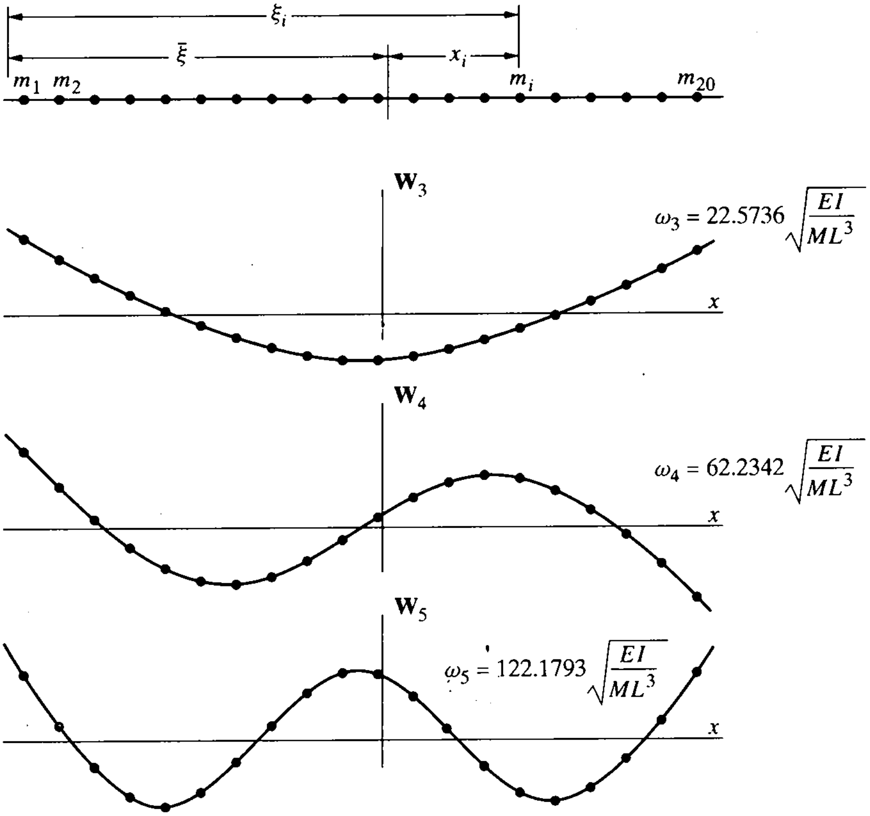

ξ_i
$\bar{\xi}$
x_i
m_1
m_2
m_i
m_{20}
$\mathbf{W}_3$
$\omega_3 = 22.5736\sqrt{\frac{EI}{ML^3}}$
x
$\mathbf{W}_4$
$\omega_4 = 62.2342\sqrt{\frac{EI}{ML^3}}$
x
$\mathbf{W}_5$
$\omega_5 = 122.1793\sqrt{\frac{EI}{ML^3}}$
x

Figure 8.5 Lumped-parameter model of a free-free beam in bending and the three lowest elastic modes

It should be noted that the natural frequencies and natural modes are reasonably close to the ones of a uniform free-free beam of total mass M and stiffness EI, as can be expected. We must also note that smaller displacements occur at the heavier, stiffer end of the beam, which agrees with the expectation.

8.2 HOLZER'S METHOD FOR TORSIONAL VIBRATION

In the lumped-parameter method using flexibility influence coefficients, the mass properties are approximated by lumping the distributed mass at individual discrete points. If the flexibility coefficients are evaluated by regarding the stiffness as distributed, as was done in Example 8.1, then the stiffness properties are accounted for exactly. This is the strength of the method, but also its main drawback, as the evaluation of flexibility coefficients for systems with distributed stiffness tends to be difficult. Even if the stiffness between any two adjacent lumps is assumed to be uniform, the situation is not significantly better, except for some simple systems, such as strings, rods and shafts characterized by second-order differential eigenvalue problems (Sec. 7.6).

Another lumped-parameter approach consists of lumping the mass at discrete points and regarding the portion between the lumped masses as being massless and of uniform stiffness, as suggested in the preceding paragraph, and does not rely on influence coefficients to characterize the stiffness properties. According to this approach, the eigenvalue problem is derived in a step-by-step process, advancing from one end of the member to the other. Clearly, this is a chain method, suitable for structures described by only one spatial variable. In this section, we apply the approach to a nonuniform shaft in torsional vibration, and in the next section we extend it to nonuniform beams in bending.

From mechanics of materials, the relation between the angle of twist $\theta(x, t)$ and the twisting moment $M(x, t)$ for a shaft in torsion is

$$\frac{\partial\theta(x,t)}{\partial x} = \frac{M(x,t)}{GJ(x)} \tag{8.21}$$

where $GJ(x)$ is the torsional stiffness, in which G is the shear modulus and $J(x)$ is the area polar moment of inertia of the cross section. Using the right-hand rule, M is positive if the vector indicating the sense of the moment is in the same direction as the normal to the cross section. For free vibration, the equation of motion is

$$\frac{\partial M(x,t)}{\partial x} = I(x)\frac{\partial^2\theta(x,t)}{\partial t^2} \tag{8.22}$$

where $I(x)$ is the mass moment of inertia density.

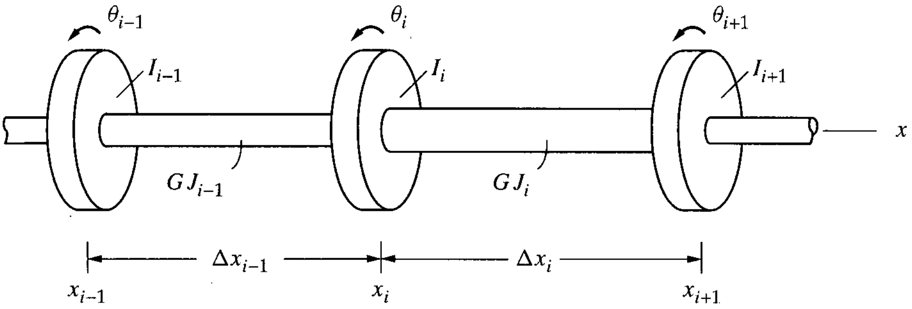


Figure 8.6 Lumped-parameter model of a shaft in torsion

Next, we consider a nonuniform shaft modeled as a lumped system consisting of a number of rigid disks connected by massless circular shafts of uniform stiffness, as shown in Fig. 8.6. Consistent with this, we approximate the differential expressions (8.21) and (8.22) by some recursive relations, which can be done through an incremental procedure combined with a good dose of ingenuity. It turns out that it is simpler to derive the recursive relations from the lumped model directly. To this end, we denote the angular displacement and torque *on the left side of disk* i by θ_i^L and M_i^L, respectively, and the same quantities *on the right side of disk* i by θ_i^R and M_i^R. Moreover, in keeping with the tradition, we refer to disk i as *station* i and to the segment between station i and station $i+1$ as *field* i. To derive the desired relations, we consider two free-body diagrams, one for station i and the other for field i; they are displayed in Figs. 8.7a and 8.7b, respectively. Because the disks are rigid, the rotations on both sides of disk i must be the same, or

$$\theta_i^R(t) = \theta_i^L(t) = \theta_i(t) \tag{8.23}$$

Then, referring to Fig. 8.7a, the moment equation of motion is

$$M_i^R(t) - M_i^L(t) = I_i\ddot{\theta}_i(t) \tag{8.24}$$

Moreover, considering the shaft segment in Fig. 8.7b, we can interpret

$$a_i = \frac{\Delta x_i}{GJ_i} \tag{8.25}$$

as a torsional flexibility coefficient representing the angular displacement at the right end of the shaft due to a unit torque at the same end, where J_i is the polar area moment of inertia of the cross section. Hence, the relation between the rotations at the two ends of field i is

$$\theta_{i+1}^L(t) = \theta_i^R(t) + a_i M_{i+1}^L(t) \tag{8.26}$$

Finally, considering the fact that the shaft segment has no inertia, we have

$$M_{i+1}^L(t) = M_i^R(t) \tag{8.27}$$

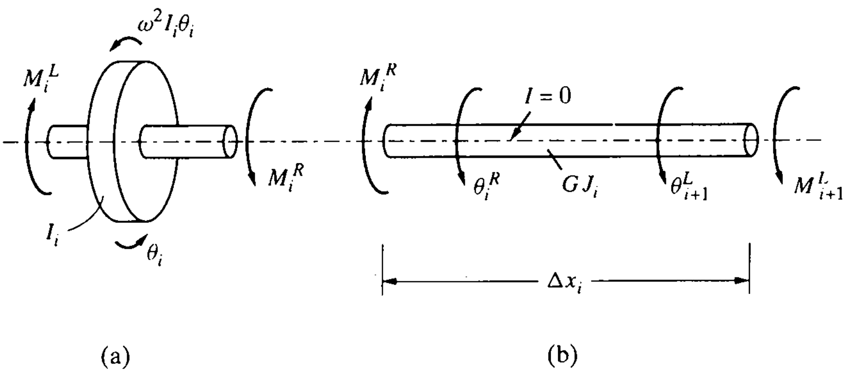


Figure 8.7 **(a)** Station i for a shaft in torsion **(b)** Field i for a shaft in torsion

At this point, we recall that free vibration is harmonic, so that $\ddot{\theta}_i(t) = -\omega^2\theta_i \cos(\omega t - \phi)$, where θ_i is a constant amplitude, ω the frequency of oscillation and ϕ an inconsequential phase angle. Introducing this expression into Eqs. (8.23) and (8.24), dividing through by $\cos(\omega t - \phi)$ and rearranging, we obtain relations in terms of amplitudes alone in the form

$$\theta_i^R = \theta_i^L, \qquad M_i^R = -\omega^2 I_i \theta_i^L + M_i^L \tag{8.28a, b}$$

and we observe that Eqs. (8.28) carry us across station i, i.e., they provide us with θ_i^R and M_i^R in terms of θ_i^L and M_i^L. Similarly, Eqs. (8.26) and (8.27) can be rewritten as

$$\theta_{i+1}^L = \theta_i^R + a_i M_i^R, \qquad M_{i+1}^L = M_i^R \tag{8.29a, b}$$

which carry us across the field i.

Equations (8.28) and (8.29) can be used recursively to relate the angle and torque at the left end to the angle and torque at the right end. The *method of Holzer* (Ref. 16) consists of using these recursive relations to solve the eigenvalue problem. The method represents a trial and error procedure, assuming values for θ_0^L and M_0^L consistent with the boundary condition at the left end and then assigning values for ω^2 repeatedly until the boundary condition at the right end is satisfied.

Instead of using Eqs. (8.28) and (8.29) on a recursive basis in conjunction with a trial and error procedure, it is possible to derive a characteristic equation in ω^2 and solve the equation by a root-finding algorithm. This approach is better explained by means of matrix notation. To this end, we express Eqs. (8.28) in the compact form

$$\mathbf{v}_i^R = T_{Si}\mathbf{v}_i^L \tag{8.30}$$

where $\mathbf{v}_i^R = \left[\theta_i^R \; M_i^R\right]^T$ and $\mathbf{v}_i^L = \left[\theta_i^L \; M_i^L\right]^T$ are referred to as *station vectors*[1] for the right side and left side of station i and

$$T_{Si} = \begin{bmatrix} 1 & 0 \\ -\omega^2 I_i & 1 \end{bmatrix} \tag{8.31}$$

is a *station transfer matrix* relating angular displacements and torques on both sides of station i. Similarly, Eqs. (8.29) can be expressed as

$$\mathbf{v}_{i+1}^L = T_{Fi}\mathbf{v}_i^R \tag{8.32}$$

where $\mathbf{v}_{i+1}^L = \left[\theta_{i+1}^L \; M_{i+1}^L\right]$ is a station vector for the left side of station $i+1$ and

$$T_{Fi} = \begin{bmatrix} 1 & a_i \\ 0 & 1 \end{bmatrix} \tag{8.33}$$

is a *field transfer matrix* relating the angular displacement and torque on the left end of field i (right side of station i) to the angular displacement and torque on the right end of field i (left side of station $i+1$). It should be stressed here that *the superscripts*

[1] In various discussions of the subject, the vectors $\mathbf{v}_i^R$ and $\mathbf{v}_i^L$ are referred to as "state vectors". In view of the fact that state vectors generally refer to vectors consisting of displacements and velocities, the term "station vectors" for $\mathbf{v}_i^R$ and $\mathbf{v}_i^L$ seems more appropriate.

R and L refer to the right side and left side of a station, not a field. Equations (8.30) and (8.32) can be combined into

$$\mathbf{v}_{i+1}^{L} = T_i \mathbf{v}_i^{L} \tag{8.34}$$

where

$$T_i = T_{Fi} T_{Si} = \begin{bmatrix} 1 - \omega^2 a_i I_i & a_i \\ -\omega^2 I_i & 1 \end{bmatrix} \tag{8.35}$$

is a transfer matrix relating the station vector on the left side of station $i + 1$ to the station vector on the left side of station i.

Equations (8.30), (8.32) and (8.34) can be used to derive an *overall transfer matrix* relating the station vector at the left boundary to the station vector at the right boundary. Embedded in this matrix is the characteristic polynomial. To illustrate the procedure, we consider the following cases:

1. *Clamped-free shaft.* In this case, we have n fields $i = 0, 1, \ldots, n-1$ and n stations, $i = 1, 2, \ldots n$, so that the recursive relations are

$$\begin{aligned} \mathbf{v}_1^L &= T_{F0}\mathbf{v}_0 \\ \mathbf{v}_2^L &= T_1\mathbf{v}_1^L = T_1 T_{F0}\mathbf{v}_0 \\ &\vdots \\ \mathbf{v}_n^L &= T_{n-1}\mathbf{v}_{n-1}^L = T_{n-1}T_{n-2}\cdots T_2 T_1 T_{F0}\mathbf{v}_0 \\ \mathbf{v}_n^R &= T_{Sn}\mathbf{v}_n^L = T_{Sn}T_{n-1}T_{n-2}\cdots T_2 T_1 T_{F0}\mathbf{v}_0 = T\mathbf{v}_0 \end{aligned} \tag{8.36}$$

where

$$T = T_{Sn}\left(\prod_{i=n-1}^{1} T_i\right) T_{F0} \tag{8.37}$$

is the *overall transfer matrix* for the case at hand. But, at the clamped end, we have the boundary condition

$$\theta_0 = 0 \tag{8.38a}$$

and at the right end the boundary condition is

$$M_n^R = 0 \tag{8.38b}$$

Hence, the last of Eqs. (8.36), in conjunction with Eqs. (8.38), yields

$$\begin{bmatrix} \theta_n^R \\ 0 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \begin{bmatrix} 0 \\ M_0 \end{bmatrix} \tag{8.39}$$

which requires that

$$T_{22} = T_{22}(\omega^2) = 0 \tag{8.40}$$

Equation (8.40) represents the *frequency equation*, in which T_{22} is a polynomial of degree n in ω^2. It has n roots, $\omega_1^2, \omega_2^2, \ldots, \omega_n^2$, which can be found by a root-finding technique, such as the *secant method*, the *Newton-Raphson method*, or *Graeffe's root squaring method* (Ref. 38).

The question remains as to how to determine the eigenvectors $\boldsymbol{\theta}_r$ belonging to the eigenvalue ω_r^2 $(r = 1, 2, \ldots, n)$. To this end, we return to the recursive relations, Eqs. (8.36), and write

$$\mathbf{v}_i^L = T_{i-1}(\omega_r^2)\, T_{i-2}(\omega_r^2) \cdots T_2(\omega_r^2)\, T_{F0}\mathbf{v}_0, \qquad r = 1, 2, \ldots, n \tag{8.41}$$

where we let for simplicity $M_0 = 1$, so that $\mathbf{v}_0 = [0\ 1]^T$. The eigenvector $\boldsymbol{\theta}_r$ has as its components the top component of $\mathbf{v}_i^L$ $(i = 1, 2, \ldots, n)$ in the mode r. Note that the bottom component can be used to compute the vector $\mathbf{M}_i^L$, which represents the torque vector on the left side of every disk in the mode r. Also note that, by taking $M_0 = 1$ arbitrarily, the eigenvector $\boldsymbol{\theta}_r$ has been normalized in a certain sense. As soon as $\boldsymbol{\theta}_r$ is computed, it can be normalized according to any other scheme, if desired.

2. *Free-free shaft.* In this case the shaft is only positive semidefinite. Hence, to approximate the shaft by an n-degree-of-freedom system, we model the system by $n+1$ stations $i = 0, 1, \ldots n$ and n in-between fields $i = 0, 1, \ldots, n-1$. Then, the relation between the station vectors to the left of station 0 and to the right of station n is simply

$$\mathbf{v}_n^R = T\mathbf{v}_0^L \tag{8.42}$$

where in this case the overall transfer matrix is

$$T = T_{Sn} \prod_{i=n-1}^{0} T_i \tag{8.43}$$

Because now the boundary conditions are

$$M_0^L = 0, \qquad M_n^R = 0 \tag{8.44a, b}$$

the frequency equation is

$$T_{21}(\omega^2) = 0 \tag{8.45}$$

where T_{21} is a polynomial of degree $n+1$ in ω^2. However, ω^2 can be factored out, so that T_{21} is the product of ω^2 and a polynomial of degree n in ω^2. Hence, there is one natural frequency equal to zero. This is consistent with the fact that the system is only positive semidefinite, so that there is one rigid-body mode with zero frequency and n elastic modes.

As the number of degrees of freedom of the system increases, the task of deriving the characteristic polynomial and finding its roots becomes more and more tedious. This task can be avoided altogether by returning to the idea of Holzer's method. For example, in the case of the clamped-free shaft, we begin with the arbitrary station vector $\mathbf{v}_0 = [0\ 1]^T$, choose some value for ω^2 and compute M_n^R by means of the recursive relations

$$\begin{aligned} \mathbf{v}_1^L &= T_{F0}\mathbf{v}_0 \\ \mathbf{v}_{i+1}^L &= T_i\mathbf{v}_i^L, \qquad i = 1, 2, \ldots, n-1 \\ \mathbf{v}_n^R &= T_{Sn}\mathbf{v}_n^L \end{aligned} \tag{8.46}$$

which are clearly based on Eqs. (8.36). If we begin with a very low value for ω^2, then the first value for which the bottom component of $\mathbf{v}_n^R$ becomes zero is the lowest eigenvalue ω_1^2. The procedure can be rendered more systematic by plotting the curve $M_n^R(\omega^2)$ versus ω^2. Then, the eigenvalues $\omega_1^2, \omega_2^2, \ldots, \omega_n^2$ are the points at which $M_n^R(\omega^2)$ intersects the axis ω^2.

Although the method was developed in connection with the torsional vibration of shafts, the approach can be clearly applied to the axial vibration of rods and the transverse vibration of strings.

8.3 MYKLESTAD'S METHOD FOR BENDING VIBRATION

Myklestad's method (Ref. 36) for the bending vibration of beams represents an extension of the ideas introduced in Sec. 8.2 for the torsional vibration of shafts. Although the extension appears natural, it is not trivial, as witnessed by the fact that it took over 20 years to complete. The presentation in this section parallels closely the one in Sec. 8.2. Here, however, we begin directly with the lumped-parameter system, instead of beginning with a distributed-parameter system and using an incremental procedure to derive a lumped-parameter model.

The differential eigenvalue problem for shafts in torsion is of degree two. Consistent with this, the station vectors are two-dimensional, with the components being the angular displacement and torsional moment. In contrast, the differential eigenvalue problem for beams in bending is of degree four, so that the station vectors must be of order four. Extrapolating from second-order problems, it is possible to conclude that the components of the station vectors must be the displacement, slope, bending moment and shearing force.

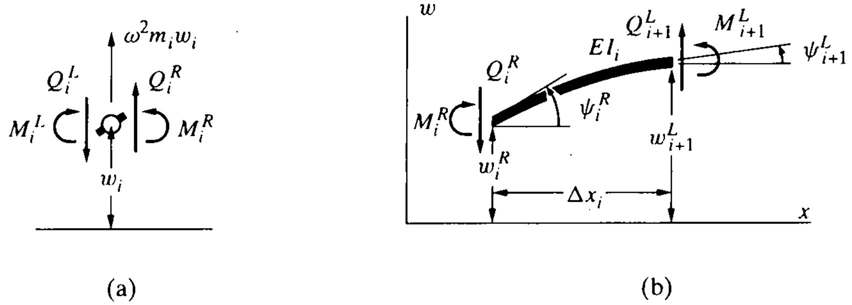


Figure 8.8 **(a)** Station i for a beam in bending **(b)** Field i for a beam in bending

By analogy with Sec. 8.2, we assume that a nonuniform Euler-Bernoulli beam is modeled as a set of lumped masses connected by massless uniform beams of length Δx_i. Free-body diagrams for a typical station i and field i are depicted in Figs. 8.8a and 8.8b, respectively. From Fig. 8.8a, due to continuity, we must have

$$w_i^R(t) = w_i^L(t) = w_i(t), \qquad \psi_i^R(t) = \psi_i^L(t) = \psi_i(t) \qquad (8.47a, b)$$

where ψ_i is the slope, i.e., the tangent to the deflection curve. Two other relations consist of the two equations of motion, one force and one moment equation. But, an Euler-Bernoulli beam implies that the rotatory inertia is negligibly small, so that the moment equation yields simply

$$M_i^R(t) = M_i^L(t) \tag{8.48}$$

On the other hand, the force equation is

$$Q_i^R(t) - Q_i^L(t) = m_i \ddot{w}_i(t) \tag{8.49}$$

Because beam segments possess flexibility, we can refer to Fig. 8.8b to obtain relations between translational and rotational displacements on the one hand and forces and moments on the other. To this end, it is convenient to *regard station i as clamped* and introduce several definitions of flexibility influence coefficients, as follows:

a_i^{wQ} is the translation at $i + 1$ due to a unit force at $i + 1$, $Q_{i+1}^L = 1$
a_i^{wM} is the translation at $i + 1$ due to a unit moment at $i + 1$, $M_{i+1}^L = 1$
$a_i^{\psi Q}$ is the rotation at $i + 1$ due to a unit force at $i + 1$, $Q_{i+1}^L = 1$
$a_i^{\psi M}$ is the rotation at $i + 1$ due to a unit moment at $i + 1$, $M_{i+1}^L = 1$.

Then, from Fig. 8.8b, we can write

$$w_{i+1}^L(t) = w_i^R(t) + \Delta x_i \psi_i^R(t) + a_i^{wM} M_{i+1}^L(t) + a_i^{wQ} Q_{i+1}^L(t) \tag{8.50a}$$

$$\psi_{i+1}^L(t) = \psi_i^R(t) + a_i^{\psi M} M_{i+1}^L(t) + a_i^{\psi Q} Q_{i+1}^L(t) \tag{8.50b}$$

Moreover, because beam segments are massless, we can write from Fig. 8.8b

$$M_{i+1}^L(t) = M_i^R(t) - \Delta x_i Q_i^R(t) \tag{8.51a}$$

$$Q_{i+1}^L(t) = Q_i^R(t) \tag{8.51b}$$

But, the right side of Eqs. (8.50) contains terms corresponding to both ends of the field. It is convenient, however, that all the terms on the right side of Eqs. (8.50) correspond to the right side only, so that we introduce Eqs. (8.51) into Eqs. (8.50) and obtain

$$w_{i+1}^L(t) = w_i^R(t) + \Delta x_i \psi_i^R(t) + a_i^{wM} M_i^R(t) + \left(a_i^{wQ} - \Delta x_i a_i^{wM}\right) Q_i^R(t) \tag{8.52a}$$

$$\psi_{i+1}^L(t) = \psi_i^R(t) + a_i^{\psi M} M_i^R(t) + \left(a_i^{\psi Q} - \Delta x_i a_i^{\psi M}\right) Q_i^R(t) \tag{8.52b}$$

At this point, following the pattern established in Sec. 8.2, we invoke the fact that free vibration is harmonic, eliminate the time dependence and express the equations in matrix form. In view of the fact that elimination of the time dependence is quite obvious, we proceed directly to the matrix formulation. To this end, we define the station vectors at i as $\mathbf{v}_i^R = \left[w_i^R \ \psi_i^R \ M_i^R \ Q_i^R\right]^T$ and $\mathbf{v}_i^L = \left[w_i^L \ \psi_i^L \ M_i^L \ Q_i^L\right]^T$, where the various components represent constant amplitudes, so that the time-independent version of Eqs. (8.47)–(8.49) can be written as

$$\mathbf{w}_i^R = T_{Si}\mathbf{w}_i^L \tag{8.53}$$

where T_{Si} is a station transfer matrix carrying us from the left side to the right side of station i and having the form

$$T_{Si} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\omega^2 m_i & 0 & 0 & 1 \end{bmatrix} \tag{8.54}$$

In a similar fashion, Eqs. (8.51) and (8.52) can be used to write a matrix expression carrying us from the left end to the right end of field i. Before writing this expression, however, we recall from mechanics of materials that the various flexibility coefficients defined earlier have the values

$$\begin{aligned} a_i^{wQ} &= \frac{(\Delta x_i)^3}{3EI_i} = \frac{a_i(\Delta x_i)^2}{3}, & a_i^{wM} &= \frac{(\Delta x_i)^2}{2EI_i} = \frac{a_i\Delta x_i}{2} \\ a_i^{\psi Q} &= \frac{(\Delta x_i)^2}{2EI_i} = \frac{a_i\Delta x_i}{2}, & a_i^{\psi M} &= \frac{\Delta x_i}{EI_i} = a_i \end{aligned} \tag{8.55}$$

where I_i is the area moment of inertia of the beam for field i. Then, inserting Eqs. (8.55) into Eqs. (8.51) and (8.52), we can write the expression relating the station vector at the left end of field i to the one at the right end in the compact matrix form

$$\mathbf{v}_{i+1}^L = T_{Fi}\mathbf{v}_i^R \tag{8.56}$$

where

$$T_{Fi} = \begin{bmatrix} 1 & \Delta x_i & a_i\Delta x_i/2 & -a_i(\Delta x_i)^2/6 \\ 0 & 1 & a_i & -a_i\Delta x_i/2 \\ 0 & 0 & 1 & -\Delta x_i \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{8.57}$$

is the associated field transfer matrix.

Following the process established in Sec. 8.2, it is possible to derive an overall transfer matrix relating the station vector $\mathbf{v}_n^R$ on the right of the beam to the station vector $\mathbf{v}_0^L$ on the left. Before the overall transfer matrix can be derived, we must specify the boundary conditions. Because the process is the same as in Sec. 8.2, except that here the matrix T is 4×4, the reader is referred to Sec. 8.2 for details.

As an illustration, we consider a *cantilever beam* in bending, in which case the relation between the station vectors $\mathbf{v}_0$ and $\mathbf{v}_n^R$ is given by the last of Eqs. (8.36), with the overall matrix T being given by Eq. (8.37). To derive the frequency equation, we must invoke the boundary conditions. The beam is clamped at the left end, so that the boundary conditions there are

$$w_0 = 0, \qquad \psi_0 = 0 \tag{8.58a}$$

On the other hand, the beam is free at the right end, so that the boundary conditions there are

$$M_n^R = 0, \qquad Q_n^R = 0 \tag{8.58b}$$

Hence, the last of Eqs. (8.36) for the case at hand has the form

$$\begin{bmatrix} w_n^R \\ \psi_n^R \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \\ T_{41} & T_{42} & T_{43} & T_{44} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ M_0 \\ Q_0 \end{bmatrix} \tag{8.59}$$

The satisfaction of the two bottom equations requires that

$$\det \begin{bmatrix} T_{33}(\omega^2) & T_{34}(\omega^2) \\ T_{43}(\omega^2) & T_{44}(\omega^2) \end{bmatrix} = 0 \tag{8.60}$$

which is recognized as the *frequency equation*, an equation of degree n in ω^2. Its solutions are the eigenvalues $\omega_1^2, \omega_2^2, \ldots, \omega_n^2$, which can be obtained by one of the methods mentioned in Sec. 8.2. Then, the eigenvectors $\mathbf{w}_r$ can be obtained by assuming arbitrarily that $\mathbf{v}_0 = [0\ 0\ M_0\ Q_0]^T = [0\ 0\ 1\ 1]^T$, using Eqs. (8.41) and retaining the top component w_i of the vectors $\mathbf{v}_i^L$ $(i = 1, 2, \ldots, n)$; these values of w_i represent the n components of the eigenvector $\mathbf{w}_r$. It should be pointed out that the second component ψ_i of the vector $\mathbf{v}_i^L$ represents the slope at station i. Although it may seem that a plot of the eigenvector $\mathbf{w}_r$ using both displacements and slopes is likely to be more accurate, for sufficiently large n, displacements alone should suffice.

Myklestad (Ref. 36) suggested a solution of the problem in tabular form, based on some scalar recursive formulas. Thomson (Ref. 45) was the first to set up the problem in matrix form using transfer matrices, without introducing the concept of station and field transfer matrices. The formulation presented here is closer to the treatment of Pestel and Leckie (Ref. 37), who applied the concept of transfer matrices to a large number of problems, including branched torsional systems and framed structures.

8.4 RAYLEIGH'S ENERGY METHOD

The lumped-parameter methods for approximating distributed systems presented in Secs. 8.1–8.3 are all characterized by the fact that the distributed mass is concentrated at given points. On the other hand, the methods differ in the treatment of the stiffness. Indeed, in the lumped-parameter method using influence coefficients the stiffness involves no approximation. In this regard, it should be noted that the use of flexibility influence coefficients instead of the flexibility influence function does not mean that the stiffness has been lumped. It simply means that the flexibility coefficients are merely obtained by evaluating the flexibility influence function at points coinciding with the nominal position of the lumped masses. By contrast, in Holzer's method and in Myklestad's method the stiffness is approximated by regarding it as uniform over the field between any two lumped masses. Satisfaction of the boundary conditions is not a major concern in lumped-parameter methods, as they are taken into account in the lumping process. Although convergence of the approximate eigenvalues to the actual ones can be assumed, no mathematical proof exists. Perhaps even more disquieting is that there are no clues as to the nature of convergence.

Beginning in this section, we turn our attention to an entirely different, and considerably more satisfying approach to the spatial discretization of distributed-parameter systems than the lumped-parameter approach. Indeed, this new approach addresses virtually all the concerns expressed in the preceding paragraph. The approach has a solid mathematical foundation permitting a clear statement concerning the nature of convergence, as well as the formulation of stability criteria. Moreover, the parameter discretization process is consistent, in the sense that the mass and stiffness distributions are treated in the same manner. The approach is based on Rayleigh's principle (Secs. 5.2 and 7.14).

To introduce the ideas, we consider first the free vibration of a conservative discrete system and write the kinetic energy and potential energy in the matrix form

$$T(t) = \frac{1}{2}\dot{\mathbf{q}}^T(t)M\dot{\mathbf{q}}(t), \qquad V(t) = \frac{1}{2}\mathbf{q}^T(t)K\mathbf{q}(t) \tag{8.61a, b}$$

where $\mathbf{q}(t) = [q_1(t)\, q_2(t)\, \dots\, q_n(t)]^T$ is the displacement vector, M the mass matrix and K the stiffness matrix. Both M and K are real symmetric and positive definite. As shown in Sec. 4.6, the free vibration of positive definite conservative systems is harmonic, so that the vector $\mathbf{q}(t)$ can be expressed as

$$\mathbf{q}(t) = \mathbf{u}\cos(\omega t - \phi) \tag{8.62}$$

where $\mathbf{u}$ is a constant vector, ω the frequency of oscillation and ϕ a phase angle. It follows that Eqs. (8.61) can be rewritten as

$$T(t) = \frac{1}{2}\omega^2\mathbf{u}^T M\mathbf{u}\sin^2(\omega t - \phi), \qquad V(t) = \frac{1}{2}\mathbf{u}^T K\mathbf{u}\cos^2(\omega t - \phi) \tag{8.63a, b}$$

When $\sin(\omega t - \phi) = \pm 1$, $\cos(\omega t - \phi) = 0$, the kinetic energy reaches the maximum value

$$T_{\text{max}} = \frac{1}{2}\omega^2\mathbf{u}^T M\mathbf{u} \tag{8.64a}$$

and the potential energy is zero. On the other hand, when $\cos(\omega t - \phi) = \pm 1$, $\sin(\omega t - \phi) = 0$, the potential energy reaches the maximum value

$$V_{\text{max}} = \frac{1}{2}\mathbf{u}^T K\mathbf{u} \tag{8.64b}$$

and the kinetic energy is zero. But, according to the conservation of energy principle (Sec. 2.5), we can write

$$E = T + V = T_{\text{max}} + 0 = 0 + V_{\text{max}} \tag{8.65}$$

where E is the total energy. Hence, inserting Eqs. (8.64) into Eq. (8.65) and introducing the *reference kinetic energy*

$$T_{\text{ref}} = \frac{1}{2}\mathbf{u}^T M\mathbf{u} \tag{8.66}$$

we obtain

$$\omega^2 = R(\mathbf{u}) = \frac{\mathbf{u}^T K \mathbf{u}}{\mathbf{u}^T M \mathbf{u}} = \frac{V_{\max}}{T_{\text{ref}}} \tag{8.67}$$

which is recognized as Rayleigh's quotient for discrete systems, first encountered in Sec. 5.2. In this regard, it should be noted that Rayleigh's quotient is expressed here for the first time as the ratio of one term proportional to the potential energy and another term proportional to the kinetic energy.

As demonstrated in Sec. 5.2, Rayleigh's quotient possesses the stationarity property, which can be stated in physical terms as follows: *The frequency of vibration of a conservative system oscillating about an equilibrium position has a stationary value in the neighborhood of a natural mode* (Ref. 39). This statement is known as *Rayleigh's principle.* As a special case, but by far the most important one, it can be stated: *The frequency of vibration of a conservative system has a minimum value in the neighborhood of the fundamental mode*, or

$$\omega_1^2 = \min R(\mathbf{u}) = \min \frac{\mathbf{u}^T K \mathbf{u}}{\mathbf{u}^T M \mathbf{u}} = \min \frac{V_{\max}}{T_{\text{ref}}} \tag{8.68}$$

where ω_1 is the lowest natural frequency. This statement alone is at times referred to as *Rayleigh's principle* (Ref. 13).

In many cases of practical interest, it is necessary to estimate the lowest natural frequency of a structure. In view of the stationarity property, Rayleigh's principle, Eq. (8.68), is ideally suited for the task. Indeed, if a trial vector $\mathbf{u}$ differing from the fundamental mode $\mathbf{u}_1$ by a small quantity of order $\boldsymbol{\epsilon}$ can be found, then Eq. (8.68) can be used to produce an estimate ω^2 differing from ω_1^2 by a small quantity of order ϵ^2. In this regard, it should be noted that if $\omega^2 = (1+\epsilon^2)\omega_1^2$, then $\omega \cong (1+\epsilon^2/2)\omega_1$, where the binomial approximation $(1+\epsilon^2)^{1/2} \cong 1+\epsilon^2/2$ has been used. This procedure for estimating the fundamental frequency is known as *Rayleigh's energy method.* Clearly, Rayleigh's energy method is applicable to discrete models of distributed-parameter systems, such as models derived by the lumped-parameter method using flexibility influence coefficients. We should note, however, that Rayleigh's quotient, as given by Eq. (8.68), involves the stiffness matrix, and it was pointed out in Sec. 8.1 that the evaluation of stiffness coefficients for distributed systems is not practical. This slight inconvenience can be overcome by recognizing that the force vector $\mathbf{f}$ is related to the displacement vector $\mathbf{u}$ by

$$\mathbf{f} = K\mathbf{u}, \qquad \mathbf{u} = K^{-1}\mathbf{f} = A\mathbf{f} \tag{8.69a, b}$$

where A is the flexibility matrix. Inserting Eqs. (8.69) into Eq. (8.68), we can rewrite Rayleigh's principle in the form

$$\omega_1^2 = \min \frac{\mathbf{f}^T A \mathbf{f}}{\mathbf{u}^T M \mathbf{u}} \tag{8.70}$$

The question remains as to how to obtain a vector $\mathbf{u}$ resembling the fundamental mode $\mathbf{u}_1$, as well as the associated force vector $\mathbf{f}$. Quite often a good choice for $\mathbf{u}$ is the static displacement vector due to loads proportional to the system lumped masses, which implies that the force vector $\mathbf{f}$ is proportional to the vector $[m_1\ m_2\ \ldots\ m_n]^T$.

Then, the static displacement vector **u** can be obtained by simply inserting **f** into Eq. (8.69b).

As indicated in the beginning of this section, our objective is to develop a method for the spatial discretization of distributed-parameter systems not involving parameter lumping. In this regard, it must be pointed out that, although we demonstrated Rayleigh's energy method on the basis of a discrete system, the method is equally applicable to distributed-parameter systems. By analogy with Eq. (8.68), Rayleigh's principle for distributed-parameter systems, Eq. (7.389), can be written in the form

$$\omega_1^2 = \min R(w) = \min \frac{V_{\max}}{T_{\text{ref}}} = \min \frac{[w, w]}{(\sqrt{m}w, \sqrt{m}w)} \tag{8.71}$$

where $[w, w]$ is an energy inner product (Sec. 7.5) and $(\sqrt{m}w, \sqrt{m}w)$ is a weighted inner product (Sec. 7.14). Here, however, the question of choosing a trial function w is more involved than in the discrete case. Of course, a function w resembling closely the lowest mode of vibration w_1 is always a good choice. Quite often, the static deflection curve due to a distributed load proportional to the mass density is likely to yield excellent estimates of w_1. Unfortunately, for complex mass and stiffness distributions the task of obtaining the static deflection curve is not trivial. Perhaps a good approach is to use as a trial function the first eigenfunction of a related but simpler system, such as one with uniform mass and stiffness distributions.

Example 8.2

Use Rayleigh's energy method to estimate the fundamental frequency of the tapered clamped-free rod in axial vibration shown in Fig. 8.9. The mass per unit length is given by

$$m(x) = 2m(1 - \frac{x}{L}) \tag{a}$$

and the stiffness distribution is

$$EA(x) = 2EA\left(1 - \frac{x}{L}\right) \tag{b}$$

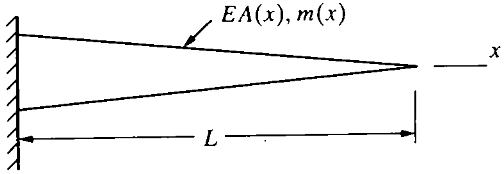


Figure 8.9 Tapered clamped-free rod in axial vibration

Using the analogy with the string in transverse vibration of Example 7.3, Rayleigh's quotient, Eq. (7.399), has the explicit expression

$$\omega^2 = R(U) = \frac{[U, U]}{(\sqrt{m}U, \sqrt{m}U)} = \frac{\int_0^L EA(x)\,[dU(x)/dx]^2\,dx}{\int_0^L m(x)U^2(x)\,dx} \tag{c}$$

As a trial function, we use the fundamental mode of a clamped-free uniform rod in axial vibration, which can be shown to be

$$U(x) = \sin\frac{\pi x}{2L} \tag{d}$$

It is clear that this trial function represents an admissible function for the problem at hand. Inserting Eq. (d) into the numerator and denominator of Rayleigh's quotient, we obtain

$$\begin{aligned}\int_0^L EA(x)\left[\frac{dU(x)}{dx}\right]^2 dx &= 2EA\left(\frac{\pi}{2L}\right)^2\int_0^L\left(1-\frac{x}{L}\right)\cos^2\frac{\pi x}{2L}dx\\ &= \left(1+\frac{\pi^2}{4}\right)\frac{EA}{2L}\end{aligned} \tag{e}$$

and

$$\int_0^L m(x)U^2(x)\,dx = 2m\int_0^L\left(1-\frac{x}{L}\right)\sin^2\frac{\pi x}{2L}dx = \left(1-\frac{4}{\pi^2}\right)\frac{mL}{2} \tag{f}$$

respectively. Hence, introducing Eqs. (e) and (f) in Eq. (c), we have

$$\omega^2 = \frac{\left(1+\pi^2/4\right)}{\left(1-4/\pi^2\right)}\frac{EA}{mL^2} = 5.8304\frac{EA}{mL^2} \tag{g}$$

from which we obtain the estimated fundamental frequency

$$\omega = 2.4146\sqrt{\frac{EA}{mL^2}} \tag{h}$$

As it turns out, the eigenvalue problem for the system under consideration can be solved in closed form (see Problem 7.28). The actual fundamental frequency has the value

$$\omega_1 = 2.4048\sqrt{\frac{EA}{mL^2}} \tag{i}$$

from which we conclude that Rayleigh's energy method yields an estimate about 0.4% higher than the actual fundamental frequency. This is a remarkable result, which can be attributed to the fact that the chosen trial function resembles the actual fundamental mode very closely.

8.5 THE RAYLEIGH-RITZ METHOD

According to Rayleigh's principle (Sec. 7.14), for a self-adjoint distributed-parameter system Rayleigh's quotient $R(w)$ has stationary values at the system eigenfunctions. Most importantly, the stationary value at the lowest eigenfunction w_1 is a minimum equal to the lowest eigenvalue λ_1, or

$$\lambda_1 = \min_w R(w) \tag{8.72}$$

where w is a trial function from the space $\mathcal{K}_B^{2p}$ of comparison functions or from the space $\mathcal{K}_G^p$ of admissible functions (Sec. 7.5), depending on the particular form of Rayleigh's quotient. This extremal property is very useful in estimating the lowest eigenvalue in cases in which no closed-form solution of the differential eigenvalue problem is possible. Indeed, Rayleigh's energy method (Sec. 8.4) consists of using Rayleigh's quotient in the form $R(w) = V_{\max}/T_{\text{ref}}$ in conjunction with an admissible function w differing from w_1 by a small quantity of order ϵ to obtain an estimate λ differing from λ_1 by a small quantity of order ϵ^2.

The extremal characterization can be extended to higher eigenvalues by restricting w to the space orthogonal to the lowest s eigenfunctions $w_i\,(i = 1, 2, \dots, s)$ and writing (Sec. 7.14)

$$\lambda_{s+1} = \min_{w} R(w), \qquad (w, w_i) = 0, \qquad i = 1, 2, \dots, s \tag{8.73}$$

Of course, if the objective is to obtain estimates of the eigenvalues $\lambda_2, \lambda_3, \dots, \lambda_{s+1}$, then this characterization has no practical value, as the eigenfunctions $w_1, w_2, \dots, w_s$ are generally not available.

A characterization independent of the lower eigenfunctions is provided by the Courant and Fischer maximin theorem (Sec. 7.14) in the form

$$\lambda_{s+1} = \max_{v_i} \min_{w} R(w), \qquad (w, v_i) = 0, \qquad i = 1, 2, \dots, s \tag{8.74}$$

where v_i $(i = 1, 2, \dots, s)$ are s independent, but otherwise arbitrary functions. Whereas the maximin theorem by itself does not represent a computational tool, it has significant implications in numerical solutions of the eigenvalue problem for distributed systems.

The Rayleigh-Ritz method is a technique for the computation of approximate solutions of the eigenvalue problem for self-adjoint distributed-parameter systems. It consists of replacing the eigenvalue problem for distributed systems by a sequence of algebraic eigenvalue problems. To introduce the ideas, it is convenient to specify the form of Rayleigh's quotient. Hence, from Eq. (7.390), we write

$$\lambda = R(w) = \frac{\int_D wLw\, dD}{\int_D mw^2\, dD} \tag{8.75}$$

where L is a self-adjoint differential operator of order $2p$, so that the trial function w must be from the space $\mathcal{K}_B^{2p}$. Next, we select a set of comparison functions $\phi_1(P)$, $\phi_2(P), \dots, \phi_n(P), \dots$ satisfying the two conditions: (i) any n members $\phi_1, \phi_2, \dots, \phi_n$ are linearly independent and (ii) the set of functions $\phi_1, \phi_2, \dots, \phi_n, \dots$ is complete (Sec. 7.5), where P denotes a nominal point in the domain D. Then, we determine min $R(w)$ not from the entire space $\mathcal{K}_B^{2p}$ but for functions of the form

$$w^{(n)}(P) = a_1\phi_1(P) + a_2\phi_2(P) + \dots + a_n\phi_n(P) = \sum_{i=1}^{n} a_i\phi_i(P) \tag{8.76}$$

The functions $\phi_i(P)$ are refered to as *coordinate functions* and they span a function space $\mathcal{R}_n$, referred to as a *Ritz space*. In fact, there is a sequence of Ritz spaces, $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_n$, each being a subspace of the next and with $\mathcal{R}_n$ being a subspace of $\mathcal{K}_B^{2p}$. For functions $w^{(n)}$ in $\mathcal{R}_n$, the coefficients a_i $(i = 1, 2, \dots, n)$ are constants yet to be determined. This amounts to approximating the variational problem for $R(w)$

by a sequence of variational problems for $R(w^{(n)})$ corresponding to $n = 1, 2, \ldots,$ or

$$\lambda^{(n)} = R(w^{(n)}) = \frac{(w^{(n)}, Lw^{(n)})}{\left(\sqrt{m}w^{(n)}, \sqrt{m}w^{(n)}\right)} \tag{8.77}$$

Of course, the case $n = 1$ merely represents Rayleigh's energy method, so that the variational approach applies to the cases in which $n \geq 2$. But, because the functions $\phi_1, \phi_2, \ldots, \phi_n$ are known, after carrying out the indicated integrations, Rayleigh's quotient reduces to a function of the undetermined coefficients, or

$$\lambda^{(n)}(a_1, a_2, \ldots, a_n) = R(a_1, a_2, \ldots, a_n) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}a_ia_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_ia_j} \tag{8.78}$$

where, because L is self-adjoint,

$$k_{ij} = k_{ji} = (\phi_i, L\phi_j) = \int_D \phi_i L\phi_j \, dD, \qquad i, j = 1, 2, \ldots, n \tag{8.79a}$$

$$m_{ij} = m_{ji} = (\sqrt{m}\phi_i, \sqrt{m}\phi_j) = \int_D m\phi_i\phi_j \, dD, \quad i, j = 1, 2, \ldots, n \tag{8.79b}$$

are symmetric *stiffness* and *mass coefficients*, respectively.

The condition for the stationarity of Rayleigh's quotient is simply

$$\delta\lambda^{(n)} = \delta R = \sum_{r=1}^{n} \frac{\partial R}{\partial a_r}\delta a_r = 0 \tag{8.80}$$

Observing that the coefficients $a_1, a_2, \ldots, a_n$ are independent, we conclude that Eq. (8.80) is satisfied if and only if the following conditions are satisfied:

$$\frac{\partial R}{\partial a_r} = 0, \qquad r = 1, 2, \ldots, n \tag{8.81}$$

Equations (8.81) involve the term

$$\begin{aligned}\frac{\partial}{\partial a_r}\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}a_ia_j &= \sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}\left(\frac{\partial a_i}{\partial a_r}a_j + a_i\frac{\partial a_j}{\partial a_r}\right) \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}\left(\delta_{ir}a_j + a_i\delta_{jr}\right) \\ &= \sum_{j=1}^{n} k_{rj}a_j + \sum_{i=1}^{n} k_{ir}a_i = 2\sum_{j=1}^{n} k_{rj}a_j, \qquad r = 1, 2, \ldots, n\end{aligned} \tag{8.82a}$$

where we considered the symmetry of the stiffness coefficients, as well as the fact that i and j are dummy indices. In the same fashion, we obtain

$$\frac{\partial}{\partial a_r}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_i a_j = 2\sum_{j=1}^{n} m_{rj}a_j, \qquad r = 1, 2, \ldots, n \tag{8.82b}$$

Hence, conditions (8.81) in conjunction with Eqs. (8.78) and (8.82) reduce to

$$\frac{\left(\frac{\partial}{\partial a_r}\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}a_i a_j\right)\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_i a_j\right) - \left(\sum_{i=1}^{n}\sum_{j=1}^{n} k_{ij}a_i a_j\right)\left(\frac{\partial}{\partial a_r}\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_i a_j\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_i a_j\right)^2}$$

$$= \frac{2}{\sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}a_i a_j}\left(\sum_{j=1}^{n} k_{rj}a_j - \lambda^{(n)}\sum_{j=1}^{n} m_{rj}a_j\right) = 0, \quad r = 1, 2, \ldots, n \tag{8.83}$$

which can be satisfied provided

$$\sum_{j=1}^{n} k_{rj}a_j = \lambda^{(n)}\sum_{j=1}^{n} m_{rj}a_j, \qquad r = 1, 2, \ldots, n \tag{8.84}$$

For $n = 1$, we obtain

$$\lambda_1^{(1)} = k_{11}/m_{11} \tag{8.85}$$

directly. On the other hand, letting $n = 2, 3, \ldots$, we obtain a sequence of algebraic eigenvalue problems of order n.

Before we proceed with a discussion of the eigenvalue problem, Eqs. (8.84), it will prove convenient to reformulate the problem in matrix form. To this end, we rewrite Eq. (8.76) as

$$w^{(n)}(P) = \boldsymbol{\phi}^T(P)\mathbf{a} \tag{8.86}$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \ldots\ \phi_n]^T$ is an n-vector with components depending on the spatial position P and $\mathbf{a} = [a_1\ a_2\ \ldots\ a_n]^T$ is a constant n-vector. Then, the sequence of algebraic eigenvalue problems, Eqs. (8.84), takes the form

$$K^{(n)}\mathbf{a} = \lambda^{(n)}M^{(n)}\mathbf{a} \tag{8.87}$$

in which

$$K^{(n)} = K^{(n)T} = \int_D \boldsymbol{\phi}L\boldsymbol{\phi}^T dD, \qquad M^{(n)} = M^{(n)T} = \int_D m\boldsymbol{\phi}\boldsymbol{\phi}^T dD \quad \text{(8.88a, b)}$$

are $n \times n$ symmetric *stiffness* and *mass matrices*. Each eigenvalue problem in the sequence represented by Eq. (8.87) is entirely analogous to that of a conservative n-degree-of-freedom discrete system, Eq. (4.81). Hence, *the Rayleigh-Ritz method* *is a discretization technique replacing a differential eigenvalue problem by a sequence of algebraic eigenvalue problems of increasing order.*

At this point, we consider Eq. (7.399) and write the second form of Rayleigh's quotient as

$$\lambda = R(w) = \frac{[w, w]}{(\sqrt{m}w, \sqrt{m}w)} \tag{8.89}$$

where $[w, w]$ is an energy inner product (Sec. 7.5) and $(\sqrt{m}w, \sqrt{m}w)$ is a weighted inner product (Sec. 7.14), and we note that this form of Rayleigh's quotient was used in Eq. (8.71) in conjunction with Rayleigh's energy method. Then, following the established pattern, we obtain the same eigenvalue problem as that given by Eq. (8.87), except that the stiffness matrix is given by

$$K^{(n)} = [\boldsymbol{\phi}, \boldsymbol{\phi}^T] \tag{8.90}$$

The mass matrix remains in the form of Eq. (8.88b). There is another difference, however. As discussed in Sec. 7.5, *the formulation of the eigenvalue problem based on* Eq. (8.89) *requires that the trial functions* $\phi_1, \phi_2, \dots, \phi_n$ *be only from the space* $\mathcal{K}_G^p$ *of admissible functions.*

The solution of the algebraic eigenvalue problem, Eq. (8.87) consists of n eigenvalues $\lambda_r^{(n)}$ and eigenvectors $\mathbf{a}_r$ $(r = 1, 2, \dots, n)$. The eigenvalues $\lambda_r^{(n)}$ provide approximations to the actual eigenvalues λ_r $(r = 1, 2, \dots, n)$. On the other hand, the eigenvectors $\mathbf{a}_r$ can be inserted into Eq. (8.86) to obtain the estimates

$$w_r^{(n)}(P) = \boldsymbol{\phi}^T(P)\mathbf{a}_r, \qquad r = 1, 2, \dots, n \tag{8.91}$$

of the eigenfunctions w_r. We refer to $\lambda_r^{(n)}$ as *Ritz eigenvalues* and to $w_r^{(n)}$ as *Ritz eigenfunctions*.

A question of particular interest is how the Ritz eigenvalues and eigenfunctions relate to the actual eigenvalues and eigenfunctions. In earlier discussions, it was implied that the Ritz eigenvalues are ordered so as to satisfy $\lambda_1^{(n)} \le \lambda_2^{(n)} \le \dots \le \lambda_n^{(n)}$, while the actual eigenvalues satisfy $\lambda_1 \le \lambda_2 \le \dots$. Because the coordinate functions $\phi_1, \phi_2, \dots, \phi_n$ are from a complete set, it can be assumed that the solution to the differential eigenvalue problem can be obtained by letting $n \to \infty$. For finite n, the approximate solution $w^{(n)}$ lies in the Ritz space $\mathcal{R}_n$, which can be interpreted as stating that the solution is subject to the constraints

$$a_{n+1} = a_{n+2} = \dots = 0 \tag{8.92}$$

But, according to Rayleigh's principle (Sec. 7.14), the lowest eigenvalue λ_1 is the minimum value Rayleigh's quotient can take as w varies over the space $\mathcal{K}_B^{2p}$. On the other hand, $\lambda_1^{(n)}$ is the minimum value Rayleigh's quotient can take for functions confined to the Ritz space $\mathcal{R}_n$. It follows that

$$\lambda_1 \le \lambda_1^{(n)} \tag{8.93}$$

To examine how the higher Ritz eigenvalues relate to the actual eigenvalues, we invoke the maximin theorem (Sec. 7.14). If we impose on the solution w of the actual system the requirement that it be orthogonal to the function v_1, then from Eq. (7.404) with $s = 1$ we can write

$$\lambda_2 = \max_{v_1} \min_{w} R(w), \qquad (w, v_1) = 0 \tag{8.94}$$

On the other hand, by imposing the same constraint on the Ritz system, we have

$$\lambda_2^{(n)} = \max_{v_1} \min_{w} R(w), \quad (w, v_1) = 0, \quad (w, \phi_j) = 0, \quad j = n+1, n+2, \ldots \tag{8.95}$$

Because the space of constraint $\mathcal{R}_{n-1}$ for calculating $\lambda_2^{(n)}$ is only a small subspace of the still infinite-dimensional space of constraint for calculating λ_2, we can write

$$\lambda_2 \leq \lambda_2^{(n)} \tag{8.96}$$

Inequality (8.96) can be generalized by writing

$$\lambda_r \leq \lambda_r^{(n)}, \qquad r = 1, 2, \ldots, n \tag{8.97}$$

Hence, the Ritz eigenvalues represent upper bounds for the actual eigenvalues.

Next, we address the question as to how the Ritz eigenvalues behave as the order n of the discrete model increases. To answer this question, we add one more term to series (8.76), so that Eq. (8.86) must be replaced by

$$w^{(n+1)}(P) = \boldsymbol{\phi}^T(P)\mathbf{a} \tag{8.98}$$

where now $\boldsymbol{\phi}$ and $\mathbf{a}$ are $(n+1)$-vectors. Consistent with this, Eq. (8.87) must be replaced by

$$K^{(n+1)}\mathbf{a} = \lambda^{(n+1)}M^{(n+1)}\mathbf{a} \tag{8.99}$$

and we observe that $K^{(n+1)}$ and $M^{(n+1)}$ are obtained through the addition of one row and one column to $K^{(n)}$ and $M^{(n)}$ without disturbing the elements of the latter two matrices. Hence, $K^{(n)}$ and $K^{(n+1)}$ on the one hand and $M^{(n)}$ and $M^{(n+1)}$ on the other hand possess the *embedding property*, or

$$K^{(n+1)} = \left[\begin{array}{c|c} K^{(n)} & \mathbf{k} \\ \hline \mathbf{k}^T & k \end{array}\right], \qquad M^{(n+1)} = \left[\begin{array}{c|c} M^{(n)} & \mathbf{m} \\ \hline \mathbf{m}^T & m \end{array}\right] \tag{8.100}$$

in which $\mathbf{k} = [k_{n+1,1}\ k_{n+1,2}\ \cdots\ k_{n+1,n}]^T$ and $\mathbf{m} = [m_{n+1,1}\ m_{n+1,2}\ \cdots\ m_{n+1,n}]^T$ are n-vectors and $k = k_{n+1,n+1}$ and $m = m_{n+1,n+1}$ are scalars. It follows from Sec. 5.4 that the two sets of eigenvalues corresponding to the two eigenvalue problems, Eqs. (8.87) and (8.99), satisfy the *separation theorem*, Eq. (5.73), which in the case at hand can be expressed as

$$\lambda_1^{(n+1)} \leq \lambda_1^{(n)} \leq \lambda_2^{(n+1)} \leq \lambda_2^{(n)} \leq \lambda_3^{(n+1)} \leq \cdots \leq \lambda_n^{(n+1)} \leq \lambda_n^{(n)} \leq \lambda_{n+1}^{(n+1)} \tag{8.101}$$

We observe that, by increasing the order of the eigenvalue problem from n to $n+1$, the n lowest newly computed eigenvalues decrease relative to the corresponding n previously computed eigenvalues, or at least they do not increase. At the same time, one more approximate eigenvalue at the higher end of the spectrum is obtained. But, inequalities (8.97) state that the approximate eigenvalues are higher than (or equal to) the corresponding actual eigenvalues. In view of the fact that the admissible functions $\phi_1, \phi_2, \ldots, \phi_n, \ldots$ are from a complete set, if follows from the preceding statements that, *as $n \to \infty$, the Ritz eigenvalues converge to the actual eigenvalues monotonically from above.* Hence, we can write

$$\lim_{n\to\infty} \lambda_r^{(n)} = \lambda_r, \qquad r = 1, 2, \ldots, n \tag{8.102}$$

The convergence process can be illustrated by means of the triangular array

$$\begin{array}{rcccccccccccc}
\lambda_1^{(1)} & \geq & \lambda_1^{(2)} & \geq & \lambda_1^{(3)} & \geq & \cdots & \geq & \lambda_1^{(n)} & \geq & \lambda_1^{(n+1)} & \geq\rightarrow & \lambda_1 \\
 & & \lambda_2^{(2)} & \geq & \lambda_2^{(3)} & \geq & \cdots & \geq & \lambda_2^{(n)} & \geq & \lambda_2^{(n+1)} & \geq\rightarrow & \lambda_2 \\
 & & & & \lambda_3^{(3)} & \geq & \cdots & \geq & \lambda_3^{(n)} & \geq & \lambda_3^{(n+1)} & \geq\rightarrow & \lambda_3 \\
 & & & & & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & & & & & & & & \lambda_n^{(n)} & \geq & \lambda_n^{(n+1)} & \geq\rightarrow & \lambda_n \\
 & & & & & & & & & & & & \vdots \\
 & & & & & & & & & & & & \infty
\end{array} \tag{8.103}$$

or more pictorially by the diagram of Fig. 8.10. Because each step of the process brings about a reduction in the computed eigenvalues, or at least not an increase, the sequence of approximate solutions $w^{(1)}, w^{(2)}, \ldots, w^{(n)}$ is referred to as a *minimizing sequence.*

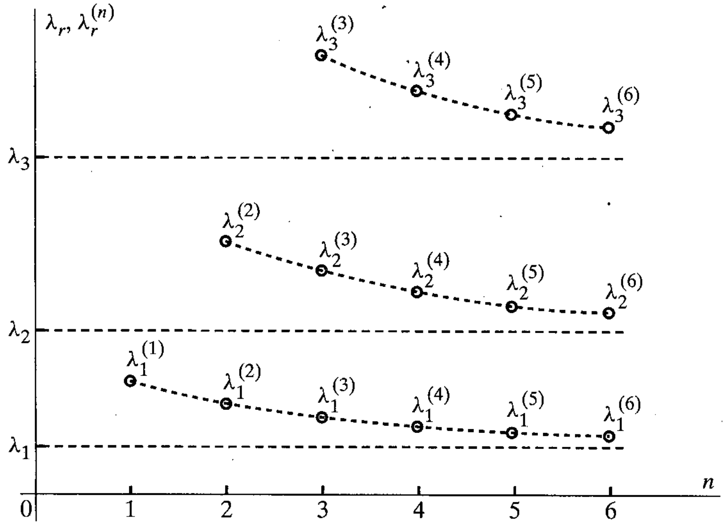


Figure 8.10 The three lowest computed eigenvalues versus the number of terms in the approximating series

The fact that the approximate eigenvalues computed by the Rayleigh-Ritz method converge to the actual eigenvalues as $n \to \infty$, provided the coordinate functions $\phi_1, \phi_2, \ldots, \phi_n, \ldots$ are from a complete set, is very reassuring and from a mathematical point of view perhaps sufficient. From a computational point of view, however, n cannot approach infinity. In fact, in computations the objective is to achieve convergence with as few terms as possible, so that the question of convergence rate is important. Before addressing this question, it is appropriate to mention a paradox characterizing not only the Rayleigh-Ritz method but all methods approximating distributed-parameter systems by discrete ones. It is no coincidence that we choose to discuss the paradox here, because the Rayleigh-Ritz method has the best convergence characteristics of all spatial discretization techniques. The paradox is that *no discrete model of a distributed system is able to yield a full set of accurate approximate eigenvalues*. As a rule of thumb, the lower computed eigenvalues tend to be more accurate than the higher eigenvalues and reach convergence first, as can be verified by means of Fig. 8.10. Increasing the order of the model will not change the state of affairs, because as the order is increased more eigenvalues are added, and these higher eigenvalues can have significant error, as shown in Fig. 8.10. In fact, at times the error in the higher computed eigenvalues can be so large as to render them meaningless. The situation is not as critical as it may appear, however, because in most practical cases the interest lies only in a given number of lower eigenvalues, as higher modes are difficult to excite and tend not to participate in the motion. To ensure that the lower eigenvalues of interest are accurate, the order of the discrete model must be at times more than twice as large as the number of these lower eigenvalues.

The rate of convergence of the Rayleigh-Ritz process depends on the quality of the trial functions $\phi_1, \phi_2, \ldots, \phi_n$, and in particular how well linear combinations of these functions can approximate the actual eigenfunctions. Of course, the actual eigenfunctions are not known a priori, so that an assessment of the choice of trial functions can be made only after an examination of the numerical results. Still, some guidelines for the selection of the trial functions can be formulated.

The issue of selecting the trial functions $\phi_1, \phi_2, \ldots, \phi_n$ is broadly connected with the particular form of Rayleigh's quotient used. As indicated earlier in this section, *in using Rayleigh's quotient in the form given by Eq. (8.75), the trial functions must be comparison functions*, i.e., they must be from the space $\mathcal{K}_B^{2p}$. By definition, comparison functions must be $2p$-times differentiable and satisfy all the boundary conditions of the problem. In problems involving natural boundary conditions, which tend to be more complicated than geometric boundary conditions (Sec. 7.1), comparison functions may not be readily available. In some cases, they can be generated by solving a related but simpler eigenvalue problem. As an example, we consider a nonuniform string in transverse vibration with one end fixed and with the other end attached to a spring, as shown in Fig. 7.1a, so that one boundary condition is geometric and the other is natural. A suitable set of comparison functions for this system consists of the eigenfunctions of a uniform string with the same boundary conditions. Another example is the rotating cantilever beam of Fig. 7.4a. The beam has two geometric boundary conditions at $x = 0$ and two natural boundary conditions at $x = L$. Even when the bending stiffness and the mass are distributed uniformly, the differential equation of motion contains terms depending on the spatial position, and no closed-form solution is possible. In this case, the eigenfunctions of a uniform nonrotating cantilever beam represent a suitable set of comparison functions. Clearly, eigenfunctions are independent and form a complete set by definition, albeit for a somewhat different system. In the first example it is relatively easy to generate a set of comparison functions and in the second example they are readily available; this is not the case in general. Indeed, there are many cases in which the natural boundary conditions cannot be satisfied exactly, particularly for two-dimensional systems. For this reason, *it is more common to base the Rayleigh-Ritz discretization process on the energy form of Rayleigh's quotient* given by Eq. (8.89), *in which case, the trial functions $\phi_1, \phi_2, \ldots, \phi_n$ need be only from the space of admissible functions*, i.e., from the space $\mathcal{K}_G^p$. We recall that functions from this space must be only p-times differentiable and satisfy only the geometric boundary conditions, a significantly larger space than the space of comparison functions. Independence and completeness are required of all sets of coordinate functions, so that they are assumed throughout. Differentiability is seldom a problem, so that admissible functions are only required to satisfy the geometric boundary conditions, which for the most part are very easy to satisfy. This opens the choice to a large number of sets of functions known to be independent, such as power series, trigonometric functions, Bessel functions, Legendre polynomials, Tchebycheff polynomials, etc. In fact, most of these sets are orthogonal in some sense, but this only guarantees independence, as they are not likely to be orthogonal with respect to the mass density of the particular system under consideration. Of course, independent functions can always be rendered orthogonal with respect to any mass distribution by the Gram-Schmidt orthogonalization process, which is akin to the process for discrete systems discussed in Appendix B, but the benefits of working with a diagonal mass matrix may pale compared with the effort required by the orthogonalization process. This is particularly true when the interest lies in shifting routine work from the analyst to the computer. Completeness in energy of a given set of functions is in general more difficult to ascertain, but it can be assumed for the most part. Of course, it is always possible to use comparison functions, as they are admissible by definition. In this case, the two forms of Rayleigh's quotient, Eqs. (8.75) and (8.89), yield exactly the same results, because the two forms can be derived from one another with due consideration to all boundary conditions. Even if comparison functions are used, Eq. (8.89) is still to be preferred, as it involves lower-order derivatives.

From the above discussion, it appears that using the eigenfunctions of a related but simpler system as admissible functions remains an attractive alternative, particularly in view of the fact that natural boundary conditions can be ignored. Care must be exercised in violating natural boundary conditions, however, as certain violations can slow down convergence significantly. We examine some of the implications of violating natural boundary conditions in Sec. 8.6.

The Rayleigh-Ritz theory is perhaps the best exponent of a variational approach used to approximate distributed-parameter systems by discrete ones. Indeed, the estimation of the Ritz eigenvalues is so satisfying mathematically that the associated theory has few equals. Unfortunately, the estimation of the Ritz eigenfunctions is a different matter. Although it may be reasonable to expect that the minimizing sequences representing the Ritz eigenfunctions $w_r^{(n)}$ converge to the actual eigenfunctions w_r, proof of convergence is not simple (Ref. 5). Moreover, because of the stationarity of Rayleigh's quotient, it can be stated that in general the Ritz eigenvalues approximate the actual eigenvalues one order of magnitude better than the

Ritz eigenfunctions approximate the actual eigenfunctions. In spite of that, it is common practice to assume that the eigenfunctions converge when the corresponding eigenvalues do.

In the case of two-dimensional problems, factors frustrating closed-form solutions are likely to frustrate approximate solutions by the Rayleigh-Ritz method as well. Indeed, if approximate solutions are at all possible, then they tend to be confined to systems with regular boundary shape, such as rectangular or circular. There is a significant difference in complexity between membranes and plates, making approximate solutions much more difficult for plates than for membranes. In the case of rectangular plates, approximate solutions can often be obtained in the form of products of beam eigenfunctions. A variety of such solutions can be found in Ref. 23. For two-dimensional systems with nonuniform mass and stiffness distributions, as for one-dimensional systems, the eigenfunctions of systems with the same boundary conditions but with uniform parameter distributions can serve as suitable admissible functions, even when these eigenfunctions are only approximations to the actual ones.

Before we conclude this section, perhaps a poignant historical note is in order. The approach was first used by Rayleigh on various occasions beginning in 1870 (Ref. 13) in connection with the vibration of air in organ pipes closed at one end and open at the other, but the approach did not receive much attention. The method became widely known as the *Ritz method* following publication of two papers by Ritz (Refs. 40 and 41). The wide attention received by these two papers can be attributed to two reasons, the "masterly" exposition of the theory by Ritz and the tragic circumstances under which Ritz wrote the papers (he was dying of consumption). In view of the fact that Ritz's work was independent of Rayleigh's, referring to the approach as the Rayleigh-Ritz method is quite appropriate. It is perhaps interesting to note that in Ref. 41 Ritz himself used products of beam eigenfunctions to solve the eigenvalue problem for a rectangular plate free on all sides.

Example 8.3

Derive and solve the appropriate algebraic eigenvalue problems for the tapered bar of Example 8.2 by means of the Rayleigh-Ritz method using Rayleigh's quotient in the form of Eq. (8.89). Use a minimizing sequence through $n = 3$ in terms of the admissible functions

$$\phi_i(x) = \sin\frac{(2i-1)\pi x}{2L}, \qquad i = 1, 2, \ldots, n \tag{a}$$

Then, verify the separation theorem.

Rayleigh's quotient for the problem at hand has the form

$$\lambda = \frac{[U, U]}{(\sqrt{m}U, \sqrt{m}U)} = \frac{\int_0^L EA(x)\,[dU(x)/dx]^2\,dx}{\int_0^L m(x)U^2(x)\,dx} \tag{b}$$

where, from Example 8.2,

$$EA(x) = 2EA\left(1 - \frac{x}{L}\right), \qquad m(x) = 2m\left(1 - \frac{x}{L}\right) \tag{c}$$

Note that, whereas the boundary condition at $x = 0$ is geometric, $w(0) = 0$, the boundary condition at $x = L$ would be natural. However, because the axial stiffness reduces to zero there, the boundary condition at $x = L$ is somewhat unorthodox and it amounts to the displacement being finite.

Using Eqs. (a)–(c), the stiffness coefficients for $i = j$ are

$$\begin{aligned} k_{ii} &= \int_0^L EA(x)(\phi_i')^2 dx \\ &= \frac{EA(2i-1)^2\pi^2}{2L^2}\int_0^L \left(1-\frac{x}{L}\right)\cos^2\frac{(2i-1)\pi x}{2L}dx \\ &= \frac{EA}{2L}\left[1+\frac{(2i-1)^2\pi^2}{4}\right] \end{aligned} \tag{d}$$

and for $i \neq j$ they are

$$\begin{aligned} k_{ij} &= \int_0^L EA(x)\phi_i'\phi_j'\,dx \\ &= \frac{EA(2i-1)(2j-1)\pi^2}{2L^2}\int_0^L \left(1-\frac{x}{L}\right)\cos\frac{(2i-1)\pi x}{2L}\cos\frac{(2j-1)\pi x}{2L}dx \\ &= \frac{EA(2i-1)(2j-1)}{4L^2}\left[\frac{1+(-1)^{i+j}}{(i+j-1)^2}+\frac{1-(-1)^{i-j}}{(i-j)^2}\right] \end{aligned} \tag{e}$$

Moreover, the mass coefficients for $i = j$ are

$$\begin{aligned} m_{ii} &= \int_0^L m(x)\phi_i^2(x)\,dx = 2m\int_0^L \left(1-\frac{x}{L}\right)\sin^2\frac{(2i-1)\pi x}{2L}dx \\ &= \frac{mL}{2}\left[1-\frac{4}{(2i-1)^2\pi^2}\right] \end{aligned} \tag{f}$$

and for $i \neq j$ they are

$$\begin{aligned} m_{ij} &= \int_0^L m(x)\phi_i(x)\phi_j(x)\,dx \\ &= 2m\int_0^L \left(1-\frac{x}{L}\right)\sin\frac{(2i-1)\pi x}{2L}\sin\frac{(2j-1)\pi x}{2L}dx \\ &= \frac{mL}{\pi^2}\left[\frac{1-(-1)^{i-j}}{(i-j)^2}-\frac{1+(-1)^{i+j}}{(i+j-1)^2}\right] \end{aligned} \tag{g}$$

For $n = 1$, we insert Eqs. (d) and (f) with $i = 1$ into Eq. (8.85) and obtain the estimate of the lowest eigenvalue

$$\lambda_1^{(1)} = \frac{k_{11}}{m_{11}} = \frac{\dfrac{EA}{2L}\left(1+\dfrac{\pi^2}{4}\right)}{\dfrac{mL}{2}\left(1-\dfrac{4}{\pi^2}\right)} = 5.8304\frac{EA}{mL^2} \tag{h}$$

from which we obtain the approximation to the lowest natural frequency

$$\omega_1^{(1)} = 2.4146\sqrt{\frac{EA}{mL^2}} \tag{i}$$

This is the result obtained in Example 8.2 by means of Rayleigh's energy method.

For $n = 2$, we use Eqs. (d)–(g) with $i, j = 1, 2$ to derive the 2×2 stiffness and mass matrices

$$K^{(2)} = \frac{EA}{8L}\begin{bmatrix} 4+\pi^2 & 12 \\ 12 & 4+9\pi^2 \end{bmatrix}, \qquad M^{(2)} = \frac{mL}{2\pi^2}\begin{bmatrix} \pi^2-4 & 4 \\ 4 & \pi^2-4/9 \end{bmatrix} \tag{j}$$

Inserting Eqs. (j) into Eq. (8.87) with $n = 2$, we obtain a 2×2 eigenvalue problem, which has the eigenvalues and eigenvectors

$$\begin{aligned} \lambda_1^{(2)} &= 5.7897\frac{EA}{mL^2}, \qquad \lambda_2^{(2)} = 30.5717\frac{EA}{mL^2} \\ \mathbf{a}_1 &= \begin{bmatrix} 1.0000 \\ -0.0369 \end{bmatrix}, \qquad \mathbf{a}_2 = \begin{bmatrix} 1.0000 \\ -1.5651 \end{bmatrix} \end{aligned} \tag{k}$$

and we note that the eigenvectors have been normalized so that the top component is unity. The eigenvalues can be used to compute an improved approximation to the lowest natural frequency and an estimate of the second natural frequency in the form

$$\omega_1^{(2)} = 2.4062\sqrt{\frac{EA}{mL^2}}, \qquad \omega_2^{(2)} = 5.5292\sqrt{\frac{EA}{mL^2}} \tag{l}$$

respectively. Moreover, inserting the eigenvectors into Eqs. (8.91), we obtain the approximate eigenfunctions

$$w_1^{(2)} = \sin\frac{\pi x}{2L} - 0.0369\sin\frac{3\pi x}{2L}, \qquad w_2^{(2)} = \sin\frac{\pi x}{2L} - 1.5651\sin\frac{3\pi x}{2L} \tag{m}$$

Following the same pattern for $n = 3$, we obtain the stiffness and mass matrices

$$\begin{aligned} K^{(3)} &= \frac{EA}{8L}\begin{bmatrix} 4+\pi^2 & 12 & 20/9 \\ 12 & 4+9\pi^2 & 60 \\ 20/9 & 60 & 4+25\pi^2 \end{bmatrix} \\ M^{(3)} &= \frac{mL}{2\pi^2}\begin{bmatrix} \pi^2-4 & 4 & -4/9 \\ 4 & \pi^2-4/9 & 4 \\ -4/9 & 4 & \pi^2-4/25 \end{bmatrix} \end{aligned} \tag{n}$$

The corresponding eigenvalue problem has the solutions

$$\begin{aligned} \lambda_1^{(3)} &= 5.7837\frac{EA}{mL^2}, \qquad \lambda_2^{(3)} = 30.4878\frac{EA}{mL^2}, \qquad \lambda_3^{(3)} = 75.0751\frac{EA}{mL^2} \\ \mathbf{a}_1 &= \begin{bmatrix} 1.0000 \\ -0.0319 \\ -0.0072 \end{bmatrix}, \qquad \mathbf{a}_2 = \begin{bmatrix} 1.0000 \\ -1.5540 \\ 0.0666 \end{bmatrix}, \qquad \mathbf{a}_3 = \begin{bmatrix} 1.0000 \\ -1.2110 \\ 2.0270 \end{bmatrix} \end{aligned} \tag{o}$$

and we note that $\lambda_1^{(3)}$ and $\lambda_2^{(3)}$ represent improved approximations to the actual eigenvalues λ_1 and λ_2 and $\lambda_3^{(3)}$ is a first estimate of the third actual eigenvalue λ_3. From Eqs. (o), we obtain the approximate natural frequencies

$$\omega_1^{(3)} = 2.4049\sqrt{\frac{EA}{mL^2}}, \qquad \omega_2^{(3)} = 5.5216\sqrt{\frac{EA}{mL^2}}, \qquad \omega_3^{(3)} = 8.6646\sqrt{\frac{EA}{mL^2}} \tag{p}$$

and approximate eigenfunctions

$$w_1^{(3)} = 0.9759\left(\sin\frac{\pi x}{2L} - 0.0319\sin\frac{3\pi x}{2L} - 0.0072\sin\frac{5\pi x}{2L}\right)$$

$$w_2^{(3)} = 0.3816\left(\sin\frac{\pi x}{2L} - 1.5540\sin\frac{3\pi x}{2L} + 0.0666\sin\frac{5\pi x}{2L}\right) \tag{q}$$
$$w_3^{(3)} = 0.2360\left(\sin\frac{\pi x}{2L} - 1.2110\sin\frac{3\pi x}{2L} + 2.0270\sin\frac{5\pi x}{2L}\right)$$

The approximate eigenfunctions, normalized so that $w_r^{(3)}(L) = 1 (r = 1, 2, 3)$, are displayed in Fig. 8.11.

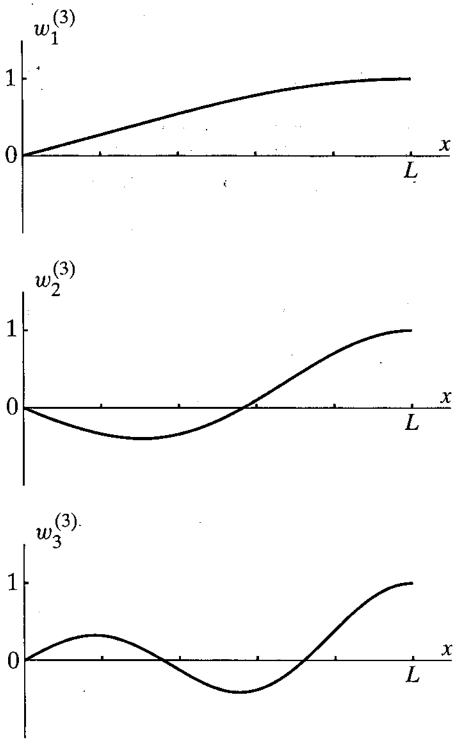


Figure 8.11 The three lowest approximate eigenfunctions for a tapered rod in axial vibration fixed at $x = 0$ and free at $x = L$

Omitting the parameter ratio EA/mL^2, the computed eigenvalues can be verified to satisfy the separation theorem as follows:

$$\begin{aligned} \lambda_1^{(2)} &= 5.7897 < \lambda_1^{(1)} = 5.8304 < \lambda_2^{(2)} = 30.5717 \\ \lambda_1^{(3)} &= 5.7837 < \lambda_1^{(2)} = 5.7897 < \lambda_2^{(3)} = 30.4878 \\ &< \lambda_2^{(2)} = 30.5717 < \lambda_3^{(3)} = 75.0751 \end{aligned} \tag{r}$$

Moreover, they form a triangular array as given by Eq. (8.103). To this end, we note that the differential eigenvalue problem admits a closed-form solution, which permits us to write

$$\begin{aligned}
\lambda_1^{(1)} = 5.8304 > \lambda_1^{(2)} = 5.7897 > \lambda_1^{(3)} &= 5.7837 > \ldots \rightarrow \lambda_1 = 5.7831 \\
\lambda_2^{(2)} = 30.5717 > \lambda_2^{(3)} &= 30.4878 > \ldots \rightarrow \lambda_2 = 30.4715 \\
\lambda_3^{(3)} &= 75.0751 > \ldots \rightarrow \lambda_3 = 74.8865
\end{aligned} \qquad \text{(s)}$$

From the above array, we conclude that the computed eigenvalues are remarkably accurate. Indeed, plots of the computed eigenvalues versus the number of admissible functions in the series for the approximate solution are relatively flat and close to the horizontal lines representing the asymptotes. The reason for this is that the admissible functions are in fact comparison functions and capable of approximating the actual eigenfunctions quite accurately, at least the first three. In general, such good accuracy with so few terms should not be expected.

8.6 THE CLASS OF QUASI-COMPARISON FUNCTIONS: AN ENHANCED RAYLEIGH-RITZ METHOD

As pointed out in Sec. 8.5, the Rayleigh-Ritz method is a technique for approximating a finite number of eigensolutions for a distributed-parameter system whereby the solution of a differential eigenvalue problem is replaced by a variational problem consisting of the minimization of Rayleigh's quotient. To this end, the solution is assumed to have the form of a minimizing sequence, with each term in the sequence consisting of a linear combination of trial functions, thus leading to a sequence of algebraic eigenvalue problems of increasing order. If the numerator of Rayleigh's quotient has the form of an inner product involving the stiffness operator L, then the trial functions must be from the space $\mathcal{K}_B^{2p}$ of comparison functions. A more common and more desirable version of Rayleigh's quotient is that in which the numerator represents a measure of the potential energy, in which case the trial functions need be from the space $\mathcal{K}_G^p$ of admissible functions alone.

The energy version of Rayleigh's quotient, Eq. (8.89), is equivalent to the version involving the stiffness operator, Eq. (8.75), only when the trial function w is from the space of comparison functions. Clearly, in using the energy version of Rayleigh's quotient in conjunction with admissible functions, the natural boundary conditions are violated, so that the question arises as to whether something that should not be sacrificed is in fact sacrificed. The answer depends on the character of the natural boundary conditions and what is potentially sacrificed is the speed of convergence.

The question of convergence speed is related to the completeness of the set of admissible functions. The concept of completeness is more qualitative than quantitative in nature (see Sec. 7.5). Whereas many will agree that $\epsilon = 10^{-6}$ is a small number, there is far less agreement as to what constitutes a sufficiently large number n of terms in the linear combination. It is precisely this number that defines the speed of convergence. A set of admissible functions can be complete in energy and still exhibit poor convergence characteristics. This can happen when eigenfunctions of a related simpler system are used as admissible functions for a system with natural boundary conditions more complicated than the free boundary of Example 8.5.

To investigate the convergence question raised above, we consider a nonuniform rod in axial vibration fixed at $x = 0$ and restrained by a spring at $x = L$, as shown in Fig. 8.12. Rayleigh's quotient for the problem at hand is

$$R(U) = \frac{[U, U]}{[\sqrt{m}U, \sqrt{m}U]} = \frac{\int_0^L EA(x)\,[dU(x)/dx]^2\,dx + kU^2(L)}{\int_0^L m(x)U^2(x)\,dx} \tag{8.104}$$

where the parameters are as follows:

$$EA(x) = \frac{6EA}{5}\left[1 - \frac{1}{2}\left(\frac{x}{L}\right)^2\right], \qquad m(x) = \frac{6m}{5}\left[1 - \frac{1}{2}\left(\frac{x}{L}\right)^2\right], \qquad k = \frac{EA}{L} \tag{8.105}$$

in which $EA(x)$ is the axial stiffness, $m(x)$ the mass density and k the spring constant. For future reference, the boundary conditions are as follows:

$$U(0) = 0, \qquad EA(x)\frac{dU(x)}{dx} + kU(x) = 0 \text{ at } x = L. \tag{8.106a, b}$$

and we note that Eq. (8.106a) represents a geometric boundary condition and Eq. (8.106b) a natural one.

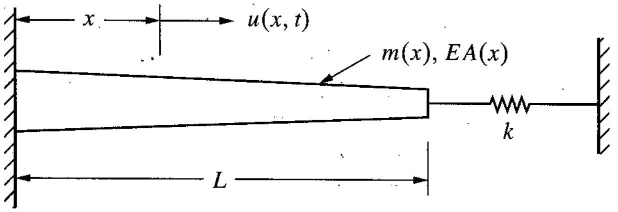


Figure 8.12 Nonuniform rod in axial vibration fixed at $x = 0$ and restrained by a spring at $x = L$

In accordance with the Rayleigh-Ritz method, we consider an approximate solution in the form

$$U^{(n)}(x) = \boldsymbol{\phi}^T(x)\mathbf{a} \tag{8.107}$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \dots\ \phi_n]^T$ is an n-vector of trial functions and $\mathbf{a} = [a_1\ a_2\ \dots\ a_n]^T$ an n-vector of undetermined coefficients. Inserting Eq. (8.107) into Eq. (8.104), we obtain the discretized Rayleigh quotient

$$R(\mathbf{a}) = \frac{\mathbf{a}^T K^{(n)}\mathbf{a}}{\mathbf{a}^T M^{(n)}\mathbf{a}} \tag{8.108}$$

in which

$$K^{(n)} = [\boldsymbol{\phi}, \boldsymbol{\phi}^T] = \int_0^L EA(x)\frac{d\boldsymbol{\phi}(x)}{dx}\frac{d\boldsymbol{\phi}^T(x)}{dx}\,dx + k\boldsymbol{\phi}(L)\boldsymbol{\phi}^T(L) \tag{8.109a}$$

$$M^{(n)} = \left(\sqrt{m}\boldsymbol{\phi}, \sqrt{m}\boldsymbol{\phi}^T\right) = \int_0^L m(x)\boldsymbol{\phi}(x)\boldsymbol{\phi}^T(x)\,dx \tag{8.109b}$$

are the stiffness matrix and mass matrix, respectively. Following the approach of Sec. 8.5, minimization of Rayleigh's quotient leads to the sequence of eigenvalue problems given by Eq. (8.87), which can be solved for the approximate eigenvalues $\lambda_r^{(n)}$ and eigenvectors $\mathbf{a}_r$ $(r = 1, 2, \ldots, n)$. The approximate eigenfunctions $U_r^{(n)}$ are obtained by inserting the eigenvectors into Eq. (8.107).

In view of the Rayleigh-Ritz theory, in using Rayleigh's quotient in the form of Eq. (8.104), the trial functions need be only admissible functions. We use as admissible functions the eigenfunctions of a uniform fixed-free rod, or

$$\phi_i(x) = \sin\frac{(2i-1)\pi x}{2L}, \qquad i = 1, 2, \ldots, n \tag{8.110}$$

Following introduction of Eq. (8.110) into Eqs. (8.109) and evaluation of the matrices $K^{(n)}$ and $M^{(n)}$, the eigenvalue problem (8.87) has been solved for $n = 1, 2, \ldots, 30$ (see Ref. 30). The resulting approximate natural frequencies $\omega_r^{(n)}$, related to the eigenvalues by $\omega_r^{(n)} = \sqrt{\lambda_r^{(n)}EA/mL^2}$, are displayed in Table 8.1. It is clear from Table 8.1 that convergence is painfully slow, as convergence to six significant digits accuracy has not been reached with $n = 30$ in the approximate solution, Eq. (8.107). The culprit can be easily identified as the inability to satisfy the natural boundary condition, Eq. (8.106b). Indeed, all admissible functions have zero derivative at $x = L$ and, according to Eq. (8.106b), the derivative at $x = L$ must be different from zero. In theory, for the derivative of a linear combination of n terms with zero derivative at $x = L$ to be different from zero there, the number n must approach infinity. This is unacceptable for an approximate solution, for which the number n must be not only finite but also as small as possible.

TABLE 8.1 The Three Lowest Approximate Natural Frequencies Using Admissible Functions

n	$\omega_1^{(n)}\sqrt{mL^2/EA}$	$\omega_2^{(n)}\sqrt{mL^2/EA}$	$\omega_3^{(n)}\sqrt{mL^2/EA}$
1	2.32965	—	—
2	2.27291	5.13905	—
3	2.25352	5.12823	8.13148
4	2.24369	5.12158	8.13028
5	2.23781	5.11727	8.12835
$\vdots$	$\vdots$	$\vdots$	$\vdots$
28	2.21920	5.10253	8.11855
29	2.21907	5.10242	8.11847
30	2.21895	5.10232	8.11840

Next, we wish to examine convergence with a solution in terms of comparison functions. To this end, we replace the admissible functions of Eq. (8.110) by the comparison functions

$$\phi_i(x) = \sin\beta_i x, \qquad i = 1, 2, \ldots, n \tag{8.111}$$

where, according to Eq. (8.106b), β_i must satisfy the transcendental equation

$$EA(L)\beta_i \cos \beta_i L + k \sin \beta_i L = 0, \qquad i = 1, 2, \ldots n \tag{8.112}$$

The approximate natural frequencies $\omega_r^{(n)}$ using comparison functions have been computed in Ref. 30 following the established pattern and the results are exhibited in Table 8.2. Clearly, the approximate natural frequencies computed by means of comparison functions have superior convergence characteristics compared with those computed by means of admissible functions. Indeed, $\omega_1^{(n)}$ reaches convergence with $n = 11$ and $\omega_2^{(n)}$ with $n = 18$. Whereas $\omega_3^{(n)}$ has not reached convergence yet with $n = 30$, convergence is not far away.

TABLE 8.2 The Three Lowest Approximate Natural Frequencies Using Comparison Functions

n	$\omega_1^{(n)}\sqrt{mL^2/EA}$	$\omega_2^{(n)}\sqrt{mL^2/EA}$	$\omega_3^{(n)}\sqrt{mL^2/EA}$
1	2.22297	—	—
2	2.21647	5.10630	—
3	2.21573	5.10070	8.12426
4	2.21559	5.09984	8.11790
5	2.21555	5.09964	8.11680
⋮	⋮	⋮	⋮
10	2.21553	5.09953	8.11633
11	2.21552	5.09953	8.11633
⋮		⋮	⋮
17		5.09953	8.11632
18		5.09952	8.11632
⋮			⋮
30			8.11632

From the preceding results, it is tempting to conclude that, when the system involves natural boundary conditions at boundaries that are not free, the argument has been settled in favor of using comparison functions. There is one problem with this conclusion, however. Whereas in the case at hand it was easy to generate comparison functions, this is not the case in general. In fact, in many cases it may not even be possible to generate comparison functions. Hence, the question is whether a way out of this seeming impasse exists. The answer is affirmative, but this requires breaking away from some of the thinking conditioned by the Rayleigh-Ritz method.

Before addressing the aspects of the Rayleigh-Ritz method in need of rethinking, it would help reviewing the points on which there is no dispute. There is general agreement that, when the differential equation cannot be satisfied exactly, approximate solutions are to be optimized by a variational process involving Rayleigh's quotient. Moreover, there is no question that the geometric boundary conditions must be satisfied. This leaves satisfaction of the natural boundary conditions as the only issue to be settled. The Rayleigh-Ritz method offers two choices, satisfy the natural boundary conditions exactly through the use of comparison functions, or abandon any attempt to satisfy them and use admissible functions. The first choice may not be an option and the second choice can lead to very poor convergence, as amply demonstrated here. But, the number of points inside the domain D of the system is infinitely larger than the number of points on the boundary S. If the boundary includes points at which the geometric integrity must be preserved, then the solution must reflect this requirement. On the other hand, if the boundary includes points involving force and moment balance, then there is no reason to insist that force and moment balance be satisfied at such boundary points while the differential equation is violated in the interior of the domain. It follows that the points inside D and the points on S involving natural boundary conditions should be afforded equal status. This implies that the same degree of approximation of the solution should extend to all points of the system, with the exception of boundary points involving the system geometry, which must be respected. This further implies that the degree of completeness required of the admissible functions should cover all the points in D and all the points on S in question. To this end, a new class of functions has been conceived in Ref. 30, namely, the class of *quasi-comparison functions*, defined as *linear combinations of admissible functions capable of approximating the differential equation and the natural boundary conditions to any degree of accuracy by merely increasing the number n of terms in the approximating solution.* In practice, *the linear combinations must be capable of satisfying the natural boundary conditions* by simply adjusting the coefficients a_i $(i = 1, 2, \ldots, n)$. This is not to say that the coefficients should be adjusted so as to satisfy the natural boundary conditions. On the contrary, the adjustment of the coefficients should be left to the variational process. The quasi-comparison functions can also be defined as linear combinations of admissible functions acting like comparison functions. It should be pointed out that there is a minimum number of admissible functions required before the linear combination becomes capable of satisfying all the boundary conditions of the system, including the natural boundary conditions. In another break with the Rayleigh-Ritz tradition, *the approximating solution $w^{(n)}$ must be constructed using members from different families of admissible functions*, each family having different dynamic characteristics. It is this variety of admissible functions that enhances the minimization process, thus permitting accurate approximations to the differential equation and the natural boundary conditions with only a relatively small number of terms. Such a feat cannot be duplicated with admissible functions from a single family, as in the ordinary Rayleigh-Ritz practice.

To illustrate this point, we return to the rod in axial vibration investigated earlier in this section and consider a set of quasi-comparison functions in the form

$$\phi_i(x) = \sin\frac{i\pi x}{2L}, \qquad i = 1, 2, \ldots, n \tag{8.113}$$

and we note that individually none of these admissible functions satisfies the natural boundary condition at $x = L$. However, as an example, the linear combination

$$w^{(2)} = \sin\frac{\pi x}{2L} + c\sin\frac{\pi x}{L} \tag{8.114}$$

can be made to satisfy the natural boundary condition, Eq. (8.106b), by merely adjusting the coefficient c. Following the same pattern as with the other two classes of functions, Eqs. (8.110) and (8.111), the three lowest approximate natural frequencies were computed in Ref. 30 using the quasi-comparison functions given by Eq. (8.113). They are given in Table 8.3. As can be concluded from Table 8.3, convergence is extremely rapid.

TABLE 8.3 The Three Lowest Approximate Natural Frequencies Using Quasi-Comparison Functions

n	$\omega_1^{(n)}\sqrt{mL^2/EA}$	$\omega_2^{(n)}\sqrt{mL^2/EA}$	$\omega_3^{(n)}\sqrt{mL^2/EA}$
1	2.32965	—	—
2	2.22359	5.98485	—
3	2.21615	5.10007	11.09264
4	2.21557	5.09957	8.15365
5	2.21555	5.09953	8.11632
6	2.21552	5.09952	8.11632
7			8.11632
8			8.11632
9			8.11632
10			8.11632
11			8.11632
12			8.11632
13			8.11631

Contrasting the results in Tables 8.2 and 8.3, we conclude that the solution in terms of quasi-comparison functions converges faster than even the solution in terms of comparison functions, except for small n. Although this may come as a surprise to some, the explanation lies in the fact that quasi-comparison functions are capable of approximating the actual solution throughout the domain $0 < x < L$ somewhat better than comparison functions, because a larger variety of functions permits a better minimization. As far as the poor results for small n are concerned, it should be recalled that there is a minimum number of admissible functions required before the linear combination becomes a quasi-comparison function. Consistent with the faster convergence of the eigenvalues, Ref. 30 gives a plot of the logarithm of the mean-square error of the first approximate eigenfunction showing that the error for the solution in terms of quasi-comparison functions drops faster with n than the corresponding error for comparison functions.

Finally, we should explain the statement that there is larger variety in the quasi-comparison functions than in comparison functions. We observe that the set of functions in Eq. (8.113) consists of two families. The functions corresponding to $i = 1, 3, \ldots$ are really the admissible functions given by Eq. (8.110). They represent the eigenfunctions of a uniform fixed-free rod and guarantee that the displacement is different from zero at $x = L$, although the slope there is zero. On the other hand, the functions corresponding to $i = 2, 4, \ldots$ represent the eigenfunctions of a uniform fixed-fixed rod and guarantee that the slope is different from zero at $x = L$, although the displacement there is zero. When the functions from the two families are combined to form an approximate solution, then the solution is such that both the displacement and force, where the latter is proportional to the slope, are different from zero at $x = L$. Hence, a linear combination of admissible functions from two families with different dynamic characteristics is able to satisfy the natural boundary condition at $x = L$ and can provide a better approximation throughout the domain $0 < x < L$ than admissible functions from a single family, or even than comparison functions. It should be pointed out that, although the two families represent eigenfunctions, they do not represent vibration modes in a physical sense, as a system cannot be fixed-free and fixed-fixed at the same time. In this regard, it should be stressed that the admissible functions used to generate quasi-comparison functions can be any functions with the proper characteristics, and need not be eigenfunctions of a related system at all (Ref. 30). In this particular example, the admissible functions must be such that at $x = 0$ any linear combination of these functions is zero and its derivative with respect to x is different from zero, thus guaranteeing that the displacement is zero and the force is not zero at $x = 0$. On the other hand, at $x = L$ both the linear combination and its derivative must be different from zero, thus ensuring that neither the displacement nor the force is zero at $x = L$. In the case of a fourth-order problem, such as a beam in bending, the required characteristics are more involved. This explains why, to secure the characteristics required to qualify as quasi-comparison functions, the linear combination of admissible functions must include members from two or more distinct families and must contain a minimum number of functions.

Care must be exercised in choosing the various families of admissible functions to ensure linear independence. Indeed, if each family forms a complete set, then a member from one family can be approximated by a sufficiently large linear combination of members from another family, which implies that, as the number of terms increases, the independence tends to be lost. This is actually the case with the families of fixed-free and fixed-fixed functions combining into the set given by Eq. (8.113). Because convergence was achieved with a relatively small number of terms, this dependence did not have an opportunity to materialize. The loss of independence can become a problem when computing higher eigenvalues, which require a larger number of terms in the linear combination. Related to this is the fact that the eigenfunctions of a uniform rod fixed at $x = 0$ and with a spring attached at $x = L$ tend to coincide with the eigenfunctions of a uniform fixed-free rod as the eigenfunction number increases. As this happens, the contribution of the eigenfunctions of the fixed-fixed rod to the accuracy of the approximate solution tends to wane, and in fact it can cause the mass matrix to become singular. If linear combinations from two families of admissible functions experience difficulties in yielding a desired number of accurate eigenvalues, then the possibility of using linear combinations from three or more families should be considered.

In the case of two-dimensional systems, the generation of quasi-comparison functions can encounter serious difficulties, and quite often may not even be possible. Still, the idea of approximating the solution to the differential eigenvalue problem by means of linear combinations from several families of admissible functions remains valid, even when they do not constitute quasi-comparison functions, and such linear combinations are likely to yield more accurate eigensolutions than linear combinations from a single family.

8.7 THE ASSUMED-MODES METHOD

The assumed-modes method is a procedure for the discretization of distributed systems closely related to the Rayleigh-Ritz method. In fact, at times it is referred to as such (Ref. 3). Although the motivation and details are different, the results are the same as those obtained by the Rayleigh-Ritz method using the energy form of Rayleigh's quotient. The main advantage of the assumed-modes method is that it is perhaps easier to grasp. Of course, the method is quite heuristic, so that if the interest lies in the finer points of analysis, such as convergence, then it is necessary to refer to the Rayleigh-Ritz theory.

In contrast with the Rayleigh-Ritz method, in the assumed-modes method we begin with the free vibration of distributed systems prior to the elimination of the time variable. First, by analogy with the Rayleigh-Ritz method, the system is discretized in space by means of a series of space-dependent trial functions multiplied by time-dependent generalized coordinates. Then, the equations of motion for the discretized system are derived by means of Lagrange's equations. The associated eigenvalue problem is precisely the same as that obtained by the Rayleigh-Ritz method.

To illustrate the approach, we assume an approximate solution in the form

$$w(P,t) \cong w^{(n)}(P,t) = \sum_{i=1}^{n} \phi_i(P)q_i(t) = \boldsymbol{\phi}^T(P)\mathbf{q}(t) \tag{8.115}$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \dots\ \phi_n]^T$ is an n-vector of trial functions depending on the spatial position P and $\mathbf{q} = [q_1\ q_2\ \dots\ q_n]^T$ is an n-vector of time-dependent generalized coordinates. Using Eq. (8.115), we discretize the kinetic energy as follows:

$$\begin{aligned} T(t) &= \frac{1}{2}\int_D m(P)\dot{w}^2(P,t)\,dD(P) \cong \frac{1}{2}\int_D m(P)\left[\dot{w}^{(n)}(P,t)\right]^2 dD(P) \\ &= \frac{1}{2}\int_D m(P)\dot{\mathbf{q}}^T(t)\boldsymbol{\phi}(P)\boldsymbol{\phi}^T(P)\dot{\mathbf{q}}(t)\,dD(P) = \frac{1}{2}\dot{\mathbf{q}}^T(t)M^{(n)}\dot{\mathbf{q}}(t) \end{aligned} \tag{8.116}$$

in which

$$M^{(n)} = \int_D m(P)\boldsymbol{\phi}(P)\boldsymbol{\phi}^T(P)\,dD(P) \tag{8.117}$$

is recognized as the mass matrix obtained by the Rayleigh-Ritz method, Eq. (8.88b). Moreover, by analogy with Eq. (7.86), the discretized potential energy has the generic form

$$V(t) = \frac{1}{2}[w(P,t), w(P,t)] \cong \frac{1}{2}\left[w^{(n)}(P,t), w^{(n)}(P,t)\right]$$
$$= \frac{1}{2}\left[\mathbf{q}^T(t)\boldsymbol{\phi}(P), \boldsymbol{\phi}^T(P)\mathbf{q}(t)\right] = \frac{1}{2}\mathbf{q}^T(t)K^{(n)}\mathbf{q}(t) \tag{8.118}$$

where

$$K^{(n)} = \left[\boldsymbol{\phi}(P), \boldsymbol{\phi}^T(P)\right] \tag{8.119}$$

is the same stiffness matrix as that obtained by the Rayleigh-Ritz method, Eq. (8.90).

Lagrange's equations of motion for conservative discrete systems can be written in the symbolic form

$$\frac{d}{dt}\left(\frac{\partial T}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial T}{\partial \mathbf{q}} + \frac{\partial V}{\partial \mathbf{q}} = \mathbf{0} \tag{8.120}$$

Hence, inserting Eqs. (8.116) and (8.118) into Eq. (8.120), we obtain the equations of motion

$$M^{(n)}\ddot{\mathbf{q}}(t) + K^{(n)}\mathbf{q}(t) = \mathbf{0} \tag{8.121}$$

Then, letting the solution be harmonic, so that $\mathbf{q}(t) = e^{i\omega^{(n)}t}\mathbf{a}$, $\omega^{(n)} = \sqrt{\lambda^{(n)}}$, Eq. (8.121) yields the algebraic eigenvalue problem

$$K^{(n)}\mathbf{a} = \lambda^{(n)}M^{(n)}\mathbf{a} \tag{8.122}$$

which is identical to that obtained by the Rayleigh-Ritz method, Eq. (8.87).

Although it may not be immediately obvious, the eigenvalue problem derived by the assumed-modes method, Eq. (8.122), can also be regarded as being based on a variational approach. In this regard, we observe that the generic Lagrange's equations themselves were derived by means of a variational approach, namely, Hamilton's principle.

Finally, there remains the question of the trial functions selection. The term "assumed modes" connotes certain eigenfunctions. But, the assumed-modes method is equivalent to the Rayleigh-Ritz method with Rayleigh's quotient in energy form, Eq. (8.89). Hence, the Rayleigh-Ritz theory applies equally well here, so that the same guidelines for the selection of the trial functions can be used for the assumed-modes method as for the Rayleigh-Ritz method. It follows from Sec. 8.5 that the trial functions need be admissible functions only, and need not be modes at all. In fact, in accordance to the enhanced Rayleigh-Ritz method of Sec. 8.6, improved accuracy can be realized by assuming solutions in the form of quasi-comparison functions.

8.8 THE METHOD OF WEIGHTED RESIDUALS

The Rayleigh-Ritz method is a technique for deriving approximate solutions to differential eigenvalue problems not permitting closed-form solutions. It is a variational approach based on the stationarity of Rayleigh's quotient, which restricts its use to self-adjoint systems. Whereas the class of self-adjoint systems is very large indeed, and it includes most systems of interest, there are many important systems not falling in this class. In view of this, a broader approach, applicable to both self-adjoint and non-self-adjoint systems is highly desirable. Such an approach is the *method of weighted residuals*, which is not just one method but an umbrella for a number of seemingly disparate approximate techniques. In contrast with the Rayleigh-Ritz method, the weighted residuals method works directly with the differential equation.

We are interested in eigenvalue problems of the type

$$Lw(x) = \lambda m(x) w(x) \tag{8.123}$$

where L is a generally non-self-adjoint differential operator of order $2p$ and m is the mass density. The solution $w(x)$ is subject to given boundary conditions. The assumption is that the problem does not admit a closed-form solution, so that we consider an approximate solution in the form

$$w(x) \cong w^{(n)}(x) = \sum_{j=1}^{n} a_j \phi_j(x) \tag{8.124}$$

in which $\phi_1, \phi_2, \ldots, \phi_n$ are n independent comparison functions from a complete set. Hence, we confine our approximate solution $w^{(n)}(x)$ to the n-dimensional subspace S_n of $\mathcal{K}_B^{2p}$, where S_n is referred to as the *trial space*. Because $w^{(n)}$ does not satisfy Eq. (8.123) exactly, there is an error at every point x. We refer to the error as a *residual* and denote it by

$$R(w^{(n)}, x) = Lw^{(n)} - \lambda^{(n)} m w^{(n)} \tag{8.125}$$

At the same time, we choose n independent functions $\psi_1(x), \psi_2(x), \ldots, \psi_n(x)$ from a different complete set and regard them as a basis for an n-dimensional subspace $\mathcal{T}_n$ of $\mathcal{K}^0$, referred to as the *test space*, and define the *weighted residual* as

$$\psi_i R = \psi_i \left(Lw^{(n)} - \lambda^{(n)} m w^{(n)} \right), \qquad i = 1, 2, \ldots, n \tag{8.126}$$

Clearly, our objective is to reduce the error to the largest extent possible. To this end, we insist that the coefficients a_j in Eq. (8.124) $(j = 1, 2, \ldots, n)$ be such that the integral of the weighted residual be zero for every i, or

$$\int_0^L \psi_i R dx = \int_0^L \psi_i \left(Lw^{(n)} - \lambda^{(n)} m w^{(n)} \right) dx = 0, \qquad i = 1, 2, \ldots, n \tag{8.127}$$

This is equivalent to requiring that the residual be orthogonal to every weighting function ψ_i $(i = 1, 2, \ldots, n)$. Inserting Eq. (8.124) into Eq. (8.127), we obtain the algebraic eigenvalue problem

$$(\psi_i, R) = \int_0^L \psi_i \left(\sum_{j=1}^n a_j L\phi_j - \lambda^{(n)} \sum_{j=1}^n a_j m\phi_j \right) dx$$

$$= \sum_{j=1}^n \left(k_{ij} - \lambda^{(n)} m_{ij} \right) a_j = 0, \qquad i = 1, 2, \ldots, n \tag{8.128}$$

where

$$k_{ij} = (\psi_i, L\phi_j) = \int_0^L \psi_i L\phi_j \, dx, \qquad i, j = 1, 2, \ldots, n \tag{8.129a}$$

$$m_{ij} = (\psi_i, m\phi_j) = \int_0^L \psi_i m\phi_j \, dx, \qquad i, j = 1, 2, \ldots, n \tag{8.129b}$$

are constant coefficients, generally nonsymmetric.

It remains to show that the solution of Eqs. (8.127) converges to the solution of the differential eigenvalue problem, Eq. (8.123). To this end, we recall that the weighting functions ψ_i $(i = 1, 2, \ldots, n)$ are from a complete set, and Eqs. (8.127) state that the residual R is orthogonal to every one of these functions. As the number n of comparison functions ϕ_j and weighting functions ψ_i is allowed to approach infinity, the only way the residual function R can be orthogonal to a complete set of functions ψ_i is for R itself to approach zero, or

$$\lim_{n\to\infty} R = \lim_{n\to\infty} (Lw^{(n)} - \lambda^{(n)} mw^{(n)}) = Lw - \lambda mw = 0 \tag{8.130}$$

Convergence arising from the vanishing of inner products, such as Eqs. (8.127), represents *weak convergence*.

It should be mentioned at this point that the weighting functions ψ_i $(i = 1, 2, \ldots, n)$ can actually be from the class $\mathcal{K}^{-1}$. It should also be mentioned that under certain circumstances the requirement that ϕ_j $(j = 1, 2, \ldots, n)$ be from the class of comparison functions can be relaxed. Indeed, integrating Eq. (8.129a) by parts and considering the boundary conditions, we conclude that ϕ_j can be from the class $\mathcal{K}_G^p$ of admissible functions. Then, as a result of the integrations by parts, the weighting functions ψ_i must also be from $\mathcal{K}_G^p$. Here too, the convergence can be vastly improved through the use of quasi-comparison functions (Refs. 14 and 32), instead of ordinary admissible functions.

The eigenvalue problem can be cast in matrix form. To this end, we introduce the n-vectors $\boldsymbol{\phi} = [\phi_1 \, \phi_2 \, \ldots \, \phi_n]^T$, $\boldsymbol{\psi} = [\psi_1 \, \psi_2 \, \ldots \, \psi_n]^T$ and $\mathbf{a} = [a_1 \, a_2 \, \ldots \, a_n]^T$, so that Eqs. (8.128) can be rewritten in the compact form

$$K^{(n)}\mathbf{a} = \lambda^{(n)} M^{(n)} \mathbf{a} \tag{8.131}$$

in which

$$K^{(n)} = (\boldsymbol{\psi}, L\boldsymbol{\phi}^T) = \int_0^L \boldsymbol{\psi} L \boldsymbol{\phi}^T dx, \qquad M^{(n)} = (\boldsymbol{\psi}, m\boldsymbol{\phi}^T) = \int_0^L m\boldsymbol{\psi}\boldsymbol{\phi}^T dx \tag{8.132a, b}$$

are nonsymmetric $n \times n$ matrices. It can be safely assumed that the matrix $M^{(n)}$ is nonsingular, so that Eq. (8.131) can be cast in the standard, single matrix form

$$A^{(n)}\mathbf{a} = \lambda^{(n)}\mathbf{a} \tag{8.133}$$

where

$$A^{(n)} = (M^{(n)})^{-1}K^{(n)} \tag{8.134}$$

is a nonsymmetric $n \times n$ matrix.

Because the coefficient matrix $A^{(n)}$ is nonsymmetric, the eigensolutions can be complex, although real solutions are possible, depending on the nature of the system. Methods for computing the eigensolutions of nonsymmetric matrices are presented in Secs. 6.13-6.16.

As indicated in the beginning of this section, the method of weighted residuals is a generic name for a family of methods based on the theory just presented. The various methods differ from one another in the nature of the test functions ψ_i. In the sequel, we discuss two methods of particular interest in vibrations.

i. Galerkin's method

Galerkin's method is the most widely used of the weighted residual methods. In fact, the method is better known under its own name than as a weighted residual method. In Galerkin's method, the weighting functions coincide with the trial functions. In vector form, we have

$$\boldsymbol{\psi} = \boldsymbol{\phi} \tag{8.135}$$

so that the coefficient matrices, Eqs. (8.132), become

$$K^{(n)} = (\boldsymbol{\phi}, L\boldsymbol{\phi}^T) = \int_0^L \boldsymbol{\phi} L\boldsymbol{\phi}^T dx, \qquad M^{(n)} = (\boldsymbol{\phi}, m\boldsymbol{\phi}^T) = \int_0^L m\boldsymbol{\phi}\boldsymbol{\phi}^T dx \tag{8.136a, b}$$

and we observe that, whereas $M^{(n)}$ is symmetric, $K^{(n)}$ is in general not symmetric, because L is non-self-adjoint.

The operator L in Eq. (8.136a) is of order $2p$ and, consistent with this, the trial functions $\phi_1, \phi_2, \ldots, \phi_n$ are from the space $\mathcal{K}_B^{2p}$ of comparison functions. As indicated earlier, the requirements on the trial functions can be lowered by integrating Eq. (8.136a) by parts p times with due consideration to the boundary conditions. Then, the trial functions need be from the space $\mathcal{K}_G^p$ of admissible functions alone. Even if these integrations are carried out, the matrix $K^{(n)}$ remains nonsymmetric because the operator L is non-self-adjoint. As an example, we consider the non-self-adjoint eigenvalue problem defined by the differential equation

$$-\frac{d}{dx}\left(s\frac{dw}{dx}\right) + r\frac{dw}{dx} = \lambda m w, \qquad 0 < x < L \tag{8.137}$$

and the boundary conditions

$$w(0) = 0, \qquad \left.\frac{dw}{dx}\right|_{x=L} = 0 \tag{8.138}$$

so that the order of the operator L is $2p = 2$. Hence, carrying out one integration by parts and considering the boundary conditions, Eqs. (8.138), a typical element of the matrix $K^{(n)}$ can be reduced as follows:

$$\begin{aligned} k_{ij} &= (\phi_i, L\phi_j) = \int_0^L \phi_i \left[-\frac{d}{dx}\left(s\frac{d\phi_j}{dx} \right) + r\frac{d\phi_j}{dx} \right] dx \\ &= -\phi_i s\frac{d\phi_j}{dx}\bigg|_0^L + \int_0^L \left(s\frac{d\phi_i}{dx}\frac{d\phi_j}{dx} + r\phi_i\frac{d\phi_j}{dx} \right) dx \\ &= \int_0^L \left(s\frac{d\phi_i}{dx}\frac{d\phi_j}{dx} + r\phi_i\frac{d\phi_j}{dx} \right) dx \end{aligned} \tag{8.139}$$

which is clearly not symmetric in ϕ_i and ϕ_j and their first derivative. We observe that the symmetry, and hence the self-adjointness, is destroyed by the term $r\phi_i d\phi_j/dx$.

In the special case in which the operator L is self-adjoint, p integrations by parts of Eq. (8.136a) with due consideration to the boundary conditions yield

$$K^{(n)} = [\boldsymbol{\phi}, \boldsymbol{\phi}^T] = K^{(n)T} \tag{8.140}$$

which is identical to Eq. (8.90) obtained by the Rayleigh-Ritz method. Clearly, now the trial functions $\phi_1, \phi_2, \ldots, \phi_n$ need be from the energy space $\mathcal{K}_G^p$ only, i.e., they need be admissible functions. Although the Galerkin method is based on a different idea than the Rayleigh-Ritz method, because it yields the same matrices $K^{(n)}$ and $M^{(n)}$ as the Rayleigh-Ritz method, *in the case of self-adjoint systems, the Galerkin and the Rayleigh-Ritz methods are equivalent.*

ii. The collocation method

The collocation method is another widely used weighted residuals method, although it is not commonly recognized as such. In this case, the weighting functions are spatial Dirac delta functions located at various preselected points x_i of the system, or

$$\psi_i(x) = \delta(x - x_i), \qquad i = 1, 2, \ldots, n \tag{8.141}$$

and we note that the Dirac delta functions are from the class $\mathcal{K}^{-1}$. This is permissible, because the functions ψ_i not only are not differentiated, but they are part of an integrand. Indeed, inserting Eqs. (8.141) into Eqs. (8.127), we obtain

$$\begin{aligned} \int_0^L \psi_i R\, dx &= \int_0^L \delta(x - x_i)(Lw^{(n)} - \lambda^{(n)} m w^{(n)})\, dx \\ &= Lw^{(n)}(x_i) - \lambda^{(n)} m(x_i) w^{(n)}(x_i) = 0, i = 1, 2, \ldots, n \end{aligned} \tag{8.142}$$

Introducing Eq. (8.124) into Eqs. (8.142), we obtain the algebraic eigenvalue problem given by Eq. (8.131), in which the matrices $K^{(n)}$ and $M^{(n)}$ have the elements

$$k_{ij} = \int_0^L \delta(x - x_i) L\phi_j dx = L\phi_j(x_i), \qquad i, j = 1, 2, \ldots, n \tag{8.143a}$$

and

$$m_{ij} = \int_0^L \delta(x - x_i) m \phi_j dx = m(x_i)\phi_j(x_i), \qquad i, j = 1, 2, \ldots, n \quad (8.143b)$$

respectively. It is obvious that the advantage of the collocation method over any other method is its simplicity, as manifested by the fact that *the evaluation of the coefficients k_{ij} and m_{ij} does not involve integrations.* The evaluation of k_{ij} requires differentiations and substitutions and the evaluation of m_{ij} involves mere substitutions, both relatively simple operations.

Equations (8.142) indicate that the differential equation is satisfied at the n preselected locations $x = x_i$ $(i = 1, 2, \ldots, n)$ throughout the domain $0 < x < L$, which explains the name of the method. Convergence can be argued in a heuristic fashion. To this end, we reinterpret the process defined by Eqs. (8.142) as driving the error to zero at the points $x = x_i$ $(i = 1, 2, \ldots, n)$. Convergence is achieved as $n \to \infty$, when the number of points at which the error has been annihilated becomes infinitely large, thus covering the entire domain $0 < x < L$.

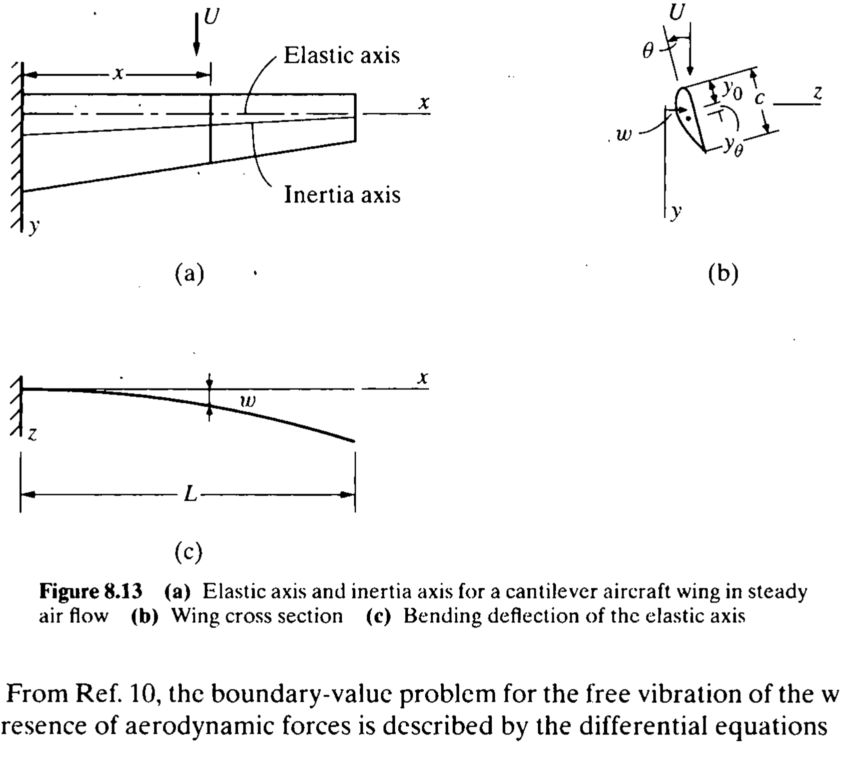
One drawback of the collocation method is that it requires the solution of a nonsymmetric eigenvalue problem. Indeed, matrices $K^{(n)}$ and $M^{(n)}$ are nonsymmetric, and remain so even when the operator L is self-adjoint. This disadvantage is mitigated somewhat in the case of self-adjoint systems, in that the approximate eigenvalues $\lambda_r^{(n)}$ $(r = 1, 2, \ldots, n)$ retain the self-adjointness characteristic of being real. The self-adjointness characteristics do not extend to the eigenvectors, which are not mutually orthogonal, albeit they are real. In this regard, we recall from Sec. 4.8 that in the case of nonsymmetric eigenvalue problems there are two sets of eigenvectors, right and left eigenvectors, and one set is orthogonal to the other set, i.e., they are biorthogonal. Another disadvantage of the collocation method is that it requires the use of comparison functions. It should be noted that, unlike Galerkin's method, carrying out p integrations by parts to obviate the use of comparison functions is not an option here, as Dirac delta functions cannot be differentiated as required.

Other weighted residual methods include the method of least squares, the method of subdomains and the method of moments. In the method of least squares, the objective is to minimize the norm of the residual squared. It has a serious drawback in that the resulting algebraic eigenvalue problem is of order $2n$, i.e., twice the order in the Galerkin method or the collocation method. The method of subdomains is somewhat similar to the collocation method, except that the weighting functions are defined over subdomains D_i of D rather than at points. The method has the same disadvantages as the collocation method without the advantage of simplicity, as the evaluation of the coefficients k_{ij} and m_{ij} requires integrations. In the method of moments, the weighting functions represent powers of x. The method is more suitable for boundary-layer problems, and there seem to be no applications from the area of vibrations. Details of these methods can be found in Ref. 28.

8.9 FLUTTER OF CANTILEVER WINGS

A classical example of non-self-adjoint problems consists of the combined bending and torsional vibration of a cantilever aircraft wing in steady air flow (Fig. 8.13).

Before we begin describing the problem, we should define the two axes shown in Fig. 8.13a, the inertia axis and the elastic axis. The inertia axis is defined as the locus of the mass centers of the cross sections and the elastic axis as the locus of the shear centers, where a shear center is a point such that a shearing force passing through it produces pure bending and a moment about it produces pure torsion. We denote the bending deflection of the elastic axis by $w(x, t)$ and the torsional rotation about the elastic axis by $\theta(x, t)$, where w is positive if it acts downward and θ is positive if the leading edge is up (Fig. 8.13b). The angle θ is referred to as the local angle of attack. We take the x-axis to coincide with the elastic axis, which is assumed to be straight, and denote the distance between the leading edge and the elastic axis at any point x by $y_0(x)$, the distance between the elastic axis and the inertia axis by $y_\theta(x)$ and the chord length by $c(x)$. The bending deflection of the elastic axis is shown in Fig. 8.13c. The speed of the air flow relative to the wing, denoted by U, is assumed to be constant.



(a) (b) (c)

Figure 8.13 **(a)** Elastic axis and inertia axis for a cantilever aircraft wing in steady air flow **(b)** Wing cross section **(c)** Bending deflection of the elastic axis

From Ref. 10, the boundary-value problem for the free vibration of the wing in the presence of aerodynamic forces is described by the differential equations

$$\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) + m\frac{\partial^2 w}{\partial t^2} + my_\theta\frac{\partial^2\theta}{\partial t^2} + \frac{\rho U^2}{2}c\frac{dC_L}{d\theta}\left[\theta + \frac{1}{U}\frac{\partial w}{\partial t}\right.$$

$$\left. + \frac{c}{U}\left(\frac{3}{4} - \frac{y_0}{c}\right)\frac{\partial\theta}{\partial t}\right] = 0, \qquad 0 < x < L \qquad (8.144a)$$

$$-\frac{\partial}{\partial x}\left(GJ\frac{\partial\theta}{\partial x}\right) + my_\theta\frac{\partial^2 w}{\partial t^2} + I_\theta\frac{\partial^2\theta}{\partial t^2} + \frac{\rho U^2}{2}c^2\left\{\frac{c\pi}{8U}\frac{\partial\theta}{\partial t}\right.$$

$$\left.+\left(\frac{1}{4}-\frac{y_0}{c}\right)\frac{dC_L}{d\theta}\left[\theta+\frac{1}{U}\frac{\partial w}{\partial t}+\frac{c}{U}\left(\frac{3}{4}-\frac{y_0}{c}\right)\frac{\partial \theta}{\partial t}\right]\right\} = 0,$$

$$0 < x < L \qquad (8.144b)$$

and the boundary conditions

$$w = 0, \quad \frac{\partial w}{\partial x} = 0, \quad \theta = 0 \quad \text{at } x = 0$$

$$EI\frac{\partial^2 w}{\partial x^2} = 0, \quad \frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right) = 0, \quad GJ\frac{\partial \theta}{\partial x} = 0 \quad \text{at } x = L \qquad (8.145)$$

where EI is the bending stiffness, GJ the torsional stiffness, m the mass per unit length, I_θ the mass moment of inertia per unit length, ρ the air density and C_L the local lift coefficient. The aerodynamic forces and moments were derived by means of the so-called quasi-steady "strip theory" whereby the local lift coefficient C_L is proportional to the instantaneous angle of attack θ. The derivative $dC_L/d\theta$ is assumed to be constant, with a theoretical value of 2π for incompressible flow and an experimental value of somewhat less than 2π. The quasi-steady assumption implies that the aerodynamic forces and moments depend only on the instantaneous deformations and prior history of the motion can be ignored (Ref. 10), which simplifies the equations of motion greatly. Indeed, the resulting equations of motion and boundary conditions are linear and homogeneous. Still, the system is non-self-adjoint.

The boundary-value problem admits a solution in the exponential form

$$w(x,t) = W(x)e^{\lambda t}, \qquad \theta(x,t) = \Theta(x)e^{\lambda t} \qquad (8.146)$$

where $W(x)$, $\Theta(x)$ and λ are in general complex. Inserting Eqs. (8.146) into Eqs. (8.144) and (8.145) and dividing through by $e^{\lambda t}$, we obtain the differential eigenvalue problem consisting of the differential equations

$$\frac{d^2}{dx^2}\left(EI\frac{d^2W}{dx^2}\right)+\frac{\rho U^2}{2}c\frac{dC_L}{d\theta}\Theta+\lambda\frac{\rho U}{2}c\frac{dC_L}{d\theta}\left[W+c\left(\frac{3}{4}-\frac{y_0}{c}\right)\Theta\right]$$

$$+\lambda^2 m\left(W+y_\theta\Theta\right) = 0, \qquad 0 < x < L \qquad (8.147a)$$

$$-\frac{d}{dx}\left(GJ\frac{d\Theta}{dx}\right)+\frac{\rho U^2}{2}c^2\left(\frac{1}{4}-\frac{y_0}{c}\right)\frac{dC_L}{d\theta}\Theta+\lambda\frac{\rho U}{2}c^2\left\{\left(\frac{1}{4}-\frac{y_0}{c}\right)\frac{dC_L}{d\theta}W\right.$$

$$\left.+c\left[\left(\frac{1}{4}-\frac{y_0}{c}\right)\left(\frac{3}{4}-\frac{y_0}{c}\right)\frac{dC_L}{d\theta}+\frac{\pi}{8}\right]\Theta\right\}+\lambda^2\left(my_\theta W+I_\theta\Theta\right) = 0,$$

$$0 < x < L \qquad (8.147b)$$

and the boundary conditions

$$W = 0, \qquad \frac{dW}{dx} = 0, \qquad \Theta = 0 \qquad \text{at } x = 0$$
$$EI\frac{d^2W}{dx^2} = 0, \qquad \frac{d}{dx}\left(EI\frac{d^2W}{dx^2}\right) = 0, \qquad GJ\frac{d\Theta}{dx} = 0 \qquad \text{at } x = L \tag{8.148}$$

The differential eigenvalue problem, Eqs. (8.147) and (8.148), has no closed-form solution, so that we consider an approximate solution by means of Galerkin's method. To this end, we assume a solution in the form

$$W(x) = \boldsymbol{\phi}_1^T(x)\mathbf{a}_1, \qquad \Theta(x) = \boldsymbol{\phi}_2^T(x)\mathbf{a}_2 \tag{8.149}$$

in which $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ are vectors of comparison functions and $\mathbf{a}_1$ and $\mathbf{a}_2$ are vectors of undetermined coefficients, where $\boldsymbol{\phi}_1$ and $\mathbf{a}_1$ are of dimension n_1 and $\boldsymbol{\phi}_2$ and $\mathbf{a}_2$ of dimension $n_2, n_1 + n_2 = n$. The vector $\boldsymbol{\phi}_1$ satisfies the boundary conditions

$$\boldsymbol{\phi}_1(0) = \mathbf{0}, \qquad \boldsymbol{\phi}_1'\Big|_{x=0} = \mathbf{0}, \qquad EI\boldsymbol{\phi}_1''\Big|_{x=L} = \mathbf{0}, \qquad (EI\boldsymbol{\phi}_1'')'\Big|_{x=L} = \mathbf{0} \tag{8.150a}$$

and the vector $\boldsymbol{\phi}_2$ satisfies the boundary conditions

$$\boldsymbol{\phi}_2(0) = \mathbf{0}, \qquad GJ\boldsymbol{\phi}_2'\Big|_{x=L} = \mathbf{0} \tag{8.150b}$$

in which primes denote the customary derivatives with respect to x. Inserting Eqs. (8.149) into Eqs. (8.147), premultiplying Eq. (8.147a) by $\boldsymbol{\phi}_1$ and Eq. (8.147b) by $\boldsymbol{\phi}_2$ and integrating over the length of the beam, we obtain the algebraic eigenvalue problem

$$\left[K + U^2H + \lambda UL + \lambda^2 M\right]\mathbf{a} = \mathbf{0} \tag{8.151}$$

where $\mathbf{a} = \left[\mathbf{a}_1^T\ \mathbf{a}_2^T\right]^T$ and the various matrices have the submatrices

$$K_{11} = \int_0^L \boldsymbol{\phi}_1(EI\boldsymbol{\phi}_1''^T)''dx = \int_0^L EI\boldsymbol{\phi}_1''\boldsymbol{\phi}_1''^T dx, \qquad K_{12} = 0$$

$$K_{21} = 0, \qquad K_{22} = -\int_0^L \boldsymbol{\phi}_2(GJ\boldsymbol{\phi}_2'^T)'\,dx = \int_0^L GJ\boldsymbol{\phi}_2'\boldsymbol{\phi}_2'^T\,dx$$

$$H_{11} = 0, \qquad H_{12} = \frac{\rho}{2}\frac{dC_L}{d\theta}\int_0^L c\boldsymbol{\phi}_1\boldsymbol{\phi}_2^T\,dx$$

$$H_{21} = 0, \qquad H_{22} = \frac{\rho}{2}\frac{dC_L}{d\theta}\int_0^L c^2\left(\frac{1}{4} - \frac{y_0}{c}\right)\boldsymbol{\phi}_2\boldsymbol{\phi}_2^T\,dx$$

$$L_{11} = \frac{\rho}{2}\frac{dC_L}{d\theta}\int_0^L c\boldsymbol{\phi}_1\boldsymbol{\phi}_1^T dx, \qquad L_{12} = \frac{\rho}{2}\frac{dC_L}{d\theta}\int_0^L c^2\left(\frac{3}{4} - \frac{y_0}{c}\right)\boldsymbol{\phi}_1\boldsymbol{\phi}_2^T dx$$

$$L_{21} = \frac{\rho}{2}\frac{dC_L}{d\theta}\int_0^L c^2\left(\frac{1}{4} - \frac{y_0}{c}\right)\boldsymbol{\phi}_2\boldsymbol{\phi}_1^T\,dx \tag{8.152}$$

$$L_{22} = \frac{\rho}{2}\int_0^L c^3\left[\left(\frac{1}{4} - \frac{y_0}{c}\right)\left(\frac{3}{4} - \frac{y_0}{c}\right)\frac{dC_L}{d\theta} + \frac{\pi}{8}\right]\boldsymbol{\phi}_2\boldsymbol{\phi}_2^T\,dx$$

$$M_{11} = \int_0^L m\boldsymbol{\phi}_1\boldsymbol{\phi}_1^T\,dx, \qquad M_{12} = \int_0^L my_\theta\boldsymbol{\phi}_1\boldsymbol{\phi}_2^T\,dx$$

$$M_{21} = \int_0^L my_\theta\boldsymbol{\phi}_2\boldsymbol{\phi}_1^T\,dx, \qquad M_{22} = \int_0^L I_\theta\boldsymbol{\phi}_2\boldsymbol{\phi}_2^T\,dx$$

The eigenvalue problem (8.151) can be expressed in the standard form

$$A\mathbf{x} = \lambda\mathbf{x} \tag{8.153}$$

in which $\mathbf{x} = [\mathbf{a}^T\ \lambda\mathbf{a}^T]^T$ and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}\left(K + U^2H\right) & -M^{-1}UL \end{array}\right] \tag{8.154}$$

The eigenvalue λ is a continuous function of the air speed U. When $U = 0$, the system is conservative and λ is pure imaginary. For $U \neq 0$, λ is in general complex, $\lambda = \alpha + i\omega$. It can be shown (Ref. 10) that for sufficiently small U and for $dC_L/d\theta < 2\pi$ the wing is losing energy to the surrounding air, so that the motion represents damped oscillation. This implies asymptotic stability, so that $\alpha < 0$. At some point, as U increases, α turns from negative to positive, as shown in Fig. 8.14, so that the motion turns from asymptotically stable to unstable. At the point $\alpha = 0$, at which the motion is merely stable and ready to become unstable, the air speed reaches the critical value U_{cr}. There can be more than one critical point but the lowest one is the most important, because in actual flight U increases from an initial zero value. There are two types of critical values, depending on the imaginary part ω. When $\alpha = 0$ and $\omega = 0$, so that $\lambda = 0$, the wing is said to be in critical *divergent* condition. When $\alpha = 0$ and $\omega \neq 0$ the wing is said to be in critical *flutter* condition. To compute U_{cr}, it is necessary to solve the eigenvalue problem repeatedly for increasing values of U, beginning with a small value. In the beginning all the eigenvalues will have negative real part. The first value of U for which the real part of one of the eigenvalues becomes zero is U_{cr}.

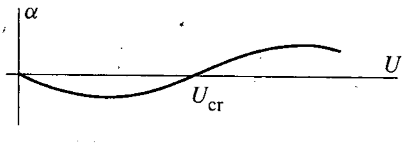


Figure 8.14 Eigenvalue real part versus the air speed

A first estimate of U_{cr} can be obtained by approximating W and Θ by means of a single term each, $n_1 = n_2 = 1$. Then, letting $\lambda = i\omega$ in Eq. (8.153) and premultiplying by block-diag $[I \ M]$, we can obtain U_{cr} from the determinantal equation

$$\det \begin{bmatrix} -i\omega & 0 & 1 & 0 \\ 0 & -i\omega & 0 & 1 \\ -k_{11} & -U_{\mathrm{cr}}^2 h_{12} & -(i\omega m_{11} + U_{\mathrm{cr}} l_{11}) & -(i\omega m_{12} + U_{\mathrm{cr}} l_{12}) \\ 0 & -(k_{22} + U_{\mathrm{cr}}^2 h_{22}) & -(i\omega m_{12} + U_{\mathrm{cr}} l_{21}) & -(i\omega m_{22} + U_{\mathrm{cr}} l_{22}) \end{bmatrix}$$
$$\begin{aligned} &= \omega^4 \left(m_{11}m_{22} - m_{12}^2\right) - i\omega^3 U_{\mathrm{cr}} \left[m_{11}l_{22} + m_{22}l_{11} - m_{12}(l_{12} + l_{21})\right] \\ &\quad - \omega^2 \left[U_{\mathrm{cr}}^2 \left(l_{11}l_{22} - l_{12}l_{21} - h_{12}m_{12} + h_{22}m_{11}\right) + k_{22}m_{11} + k_{11}m_{22}\right] \\ &\quad - i\omega U_{\mathrm{cr}} \left[U_{\mathrm{cr}}^2 \left(h_{12}l_{21} + h_{22}l_{11}\right) + k_{22}l_{11} - k_{11}l_{22}\right] + k_{11} \left(U_{\mathrm{cr}}^2 h_{22} + k_{22}\right) \\ &= 0 \end{aligned} \tag{8.155}$$

Equation (8.155) is complex, so that its satisfaction requires that both the real and imaginary part be zero, which permits a solution for both ω and U_{cr}. Indeed, equating the imaginary part to zero, we obtain

$$\omega^2 = \frac{U_{\mathrm{cr}}^2 \left(h_{12}l_{21} + h_{22}l_{11}\right) + k_{22}l_{11} - k_{11}l_{22}}{m_{12}(l_{12} + l_{21}) - (m_{11}l_{22} + m_{22}l_{11})} \tag{8.156}$$

Then, inserting Eq. (8.156) into the real part of Eq. (8.155), we obtain the quadratic equation in U_{cr}^2

$$aU_{\mathrm{cr}}^4 + bU_{\mathrm{cr}}^2 + c = 0 \tag{8.157}$$

where

$$\begin{aligned} a &= (h_{12}l_{21} + h_{22}l_{11}) \left\{(h_{12}l_{21} + h_{22}l_{11}) - (m_{11}m_{22} - m_{12}^2) - (l_{11}l_{22} - l_{12}l_{21} \right. \\ &\quad \left. - h_{12}m_{12} + h_{22}m_{11}) [m_{12}(l_{12} + l_{21}) - (m_{11}l_{22} + m_{22}l_{11})] \right\} \\ b &= 2(h_{12}l_{21} + h_{22}l_{11})(k_{22}l_{11} - k_{11}l_{22})(m_{11}m_{22} - m_{12}^2) \\ &\quad - [(h_{12}l_{21} + h_{22}l_{11})(k_{22}m_{11} + k_{11}m_{22}) \\ &\quad + (k_{22}l_{11} - k_{11}l_{22})(l_{11}l_{22} - l_{12}l_{21} - h_{12}m_{12} + h_{22}m_{11})] [m_{12}(l_{12} + l_{21}) \\ &\quad - (m_{11}l_{22} + m_{22}l_{11})] + k_{11}k_{22} [m_{12}(l_{12} + l_{21}) - (m_{11}l_{22} + m_{22}l_{11})]^2 \\ c &= (k_{22}l_{11} - k_{11}l_{22})^2 (m_{11}m_{22} - m_{12}^2) \\ &\quad - (k_{22}l_{11} - k_{11}l_{22}) (k_{22}m_{11} + k_{11}m_{22}) [m_{12}(l_{12} + l_{21}) \\ &\quad - (m_{11}l_{22} + m_{22}l_{11})] + k_{11}k_{22} [m_{12}(l_{12} + l_{21}) - (m_{11}l_{22} + m_{22}l_{11})]^2 \end{aligned} \tag{8.158}$$

The solution of Eq. (8.157) is simply

$$U_{\mathrm{cr}}^2 = -\frac{b}{2a} \pm \frac{1}{2a}\sqrt{b^2 - 4ac} \tag{8.159}$$

so that there are four values for U_{cr}. For flutter to occur, at least one of these values must be real and positive. Then, an approximation for the critical air speed U_{cr} is given by the smallest real positive value.

Additional insights into the problem of dynamic aeroelastic instability of cantilever aircraft wings can be gained from Ref. 8.

8.10 SYSTEM RESPONSE BY APPROXIMATE METHODS

In deriving the system response, we must distinguish between conservative and nonconservative systems. Indeed, the process is quite different in the two cases, as for conservative systems the response can be obtained in the configuration space and for general nonconservative systems it is necessary to cast the problem in the state space. In this section, we discuss both cases, as follows:

i. Conservative systems

We are concerned with the solution of a boundary-value problem consisting of the differential equation

$$Lw(P,t) + m(P)\ddot{w}(P,t) = f(P,t), \qquad P \text{ in } D \tag{8.160}$$

and the boundary conditions

$$B_i w(P,t) = 0, \qquad i = 1, 2, \ldots, p, \qquad P \text{ on } S \tag{8.161}$$

The various terms are as defined in Sec. 7.16. The solution of Eqs. (8.160) and (8.161) is subject to initial conditions in the form of the initial displacement $w(P,0) = w_0(P)$ and the initial velocity $\dot{w}(P,0) = v_0(P)$.

We consider the case in which the boundary-value problem, Eqs. (8.160) and (8.161), does not admit a closed-form solution, so that the interest lies in an approximate one. Assuming that the operator L is self-adjoint, we propose to derive an approximate solution in conjunction with the Rayleigh-Ritz method.

To this end, we assume an approximate solution of Eq. (8.160) in the form

$$w(P,t) \cong w^{(n)}(P,t) = \boldsymbol{\phi}^T(P)\mathbf{q}(t) \tag{8.162}$$

in which $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \ldots\ \phi_n]^T$ is an n-vector of comparison functions and $\mathbf{q} = [q_1\ q_2\ \ldots q_n]^T$ is an n-vector of generalized coordinates. Note that such a solution satisfies the boundary conditions (8.161) automatically. Inserting Eq. (8.162) into Eq. (8.160), premultiplying by $\boldsymbol{\phi}$ and integrating over the domain D, we obtain the spatially discretized system

$$M^{(n)}\ddot{\mathbf{q}}(t) + K^{(n)}\mathbf{q}(t) = \mathbf{Q}(t) \tag{8.163}$$

where

$$M^{(n)} = \int_D m\boldsymbol{\phi}\boldsymbol{\phi}^T\, dD, \qquad K^{(n)} = \int_D \boldsymbol{\phi}L\boldsymbol{\phi}^T\, dD \tag{8.164a, b}$$

are symmetric $n \times n$ mass and stiffness matrices and

$$\mathbf{Q}(t) = \int_D \boldsymbol{\phi} f \, dD \tag{8.165}$$

is an n-dimensional generalized force vector.

To obtain the solution of Eq. (8.163), we first solve the eigenvalue problem

$$K^{(n)}\mathbf{u} = \lambda^{(n)} M^{(n)}\mathbf{u} \tag{8.166}$$

by one of the approaches discussed in Chapter 6. The solution consists of the eigenvalues $\lambda_r^{(n)} = (\omega_r^{(n)})^2$, where $\omega_r^{(n)}$ are the approximate natural frequencies, and the eigenvectors $\mathbf{u}_r^{(n)}$ $(r = 1, 2, \ldots, n)$. The eigenvectors are orthogonal with respect to $M^{(n)}$ and $K^{(n)}$ and are assumed to be normalized so as to satisfy

$$\mathbf{u}_s^{(n)T} M^{(n)} \mathbf{u}_r^{(n)} = \delta_{rs}, \qquad \mathbf{u}_s^{(n)T} K^{(n)} \mathbf{u}_r^{(n)} = \lambda_r^{(n)} \delta_{rs}, \qquad r, s = 1, 2, \ldots, n \tag{8.167a, b}$$

At this point, we pause to consider questions of accuracy. We recall from Sec. 8.5 that not all eigensolutions $\lambda_r^{(n)}$, $\mathbf{u}_r^{(n)}$ $(r = 1, 2, \ldots, n)$ are accurate. In particular, the lower eigensolutions tend to be more accurate than the higher ones. In view of this, we assume that n is sufficiently large that the lowest m eigensolutions can be regarded as accurate. Moreover, we assume that m is sufficiently large that no mode higher than m is excited. Of course, this assumption must be verified by checking the extent of participation of the mode $m+1$. Then, we consider a solution of Eq. (8.163) in the truncated form

$$\mathbf{q}(t) = U_{\text{tr}} \boldsymbol{\eta}(t) \tag{8.168}$$

in which $U_{\text{tr}} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \ \mathbf{u}_m]$ is an $n \times m$ truncated modal matrix and $\boldsymbol{\eta}$ is a truncated m-vector of modal coordinates. Introducing Eq. (8.168) into Eq. (8.163), premultiplying by U_{tr}^T and using the orthonormality relations, Eqs. (8.167), we obtain the modal equation

$$\ddot{\boldsymbol{\eta}}(t) + \Lambda_{\text{tr}}^{(n)} \boldsymbol{\eta}(t) = \mathbf{N}(t) \tag{8.169}$$

where $\Lambda_{\text{tr}}^{(n)} = \text{diag}(\lambda_1^{(n)} \ \lambda_2^{(n)} \ \ldots \ \lambda_m^{(n)}) = \text{diag} \ [(\omega_1^{(n)})^2 \ (\omega_2^{(n)})^2 \ \ldots \ (\omega_m^{(n)})^2]$ is a truncated diagonal matrix of approximate natural frequencies squared and

$$\mathbf{N}(t) = U_{\text{tr}}^T \mathbf{Q}(t) \tag{8.170}$$

is the truncated m-vector of modal forces. Equation (8.169) represents a set of independent equations of the type examined in Sec. 4.10. Hence, from Sec. 4.10, we write the response

$$\eta_r(t) = \frac{1}{\omega_r^{(n)}} \int_0^t N_r(t - \tau) \sin \omega_r^{(n)} \tau \, d\tau + \eta_r(0) \cos \omega_r^{(n)} t + \frac{\dot{\eta}_r(0)}{\omega_r^{(n)}} \sin \omega_r^{(n)} t, \qquad r = 1, 2, \ldots, m \tag{8.171}$$

in which $\eta_r(0)$ and $\dot{\eta}_r(0)$ are initial modal displacement and velocity depending on the actual initial displacement $w_0(P)$ and initial velocity $v_0(P)$, respectively. To obtain the relation between the two types of quantities, we insert Eq. (8.168) into Eq. (8.162) and write the system response in the form

$$\begin{aligned} w(P,t) \cong w^{(n)}(P,t) &= \boldsymbol{\phi}^T(P)\mathbf{q}(t) = \boldsymbol{\phi}^T(P)U_{\text{tr}}\boldsymbol{\eta}(t) \\ &= \sum_{r=1}^{m} \boldsymbol{\phi}^T(P)\mathbf{u}_r\eta_r(t) = \sum_{r=1}^{m} w_r^{(n)}(P)\eta_r(t) \end{aligned} \tag{8.172}$$

where, by analogy with Eq. (8.91),

$$w_r^{(n)}(P) = \boldsymbol{\phi}^T(P)\mathbf{u}_r = \mathbf{u}_r^T\boldsymbol{\phi}(P) \tag{8.173}$$

are the approximate modes of the distributed system. Using Eqs. (8.164a) and (8.167a), these modes can be shown to satisfy the orthonormality relations

$$\int_D m(P)w_s^{(n)}(P)w_r^{(n)}(P)\,dD(P) = \delta_{rs}, \qquad r,s = 1,2,\ldots,n \tag{8.174}$$

But, letting $t = 0$ in Eq. (8.172), we can write

$$w(P,0) = w_0(P) \cong \sum_{r=1}^{m} w_r^{(n)}(P)\eta_r(0) \tag{8.175}$$

Hence, multiplying Eq. (8.175) by $m(P)w_s^{(n)}(P)$, integrating over the domain D and considering Eq. (8.174), we obtain

$$\eta_r(0) = \int_D m(P)w_r^{(n)}(P)w_0(P)\,dD(P), \qquad r = 1,2,\ldots m \tag{8.176a}$$

Similarly,

$$\dot{\eta}_r(0) = \int_D m(P)w_r^{(n)}(P)v_o(P)\,dD(P), \qquad r = 1,2,\ldots,m \tag{8.176b}$$

The formal approximate response of the system is obtained by inserting Eq. (8.171) in conjunction with Eqs. (8.176), as well as Eqs. (8.165) and (8.170), into Eq. (8.172).

ii. Nonconservative systems

In the general case of damping, the boundary-value problem can be described by the differential equation (Sec. 7.18)

$$Lw(P,t) + C\dot{w}(P,t) + m\ddot{w}(P,t) = f(P,t), \qquad P \text{ in } D \tag{8.177}$$

and the boundary conditions

$$B_iw(P,t) = 0, \qquad P \text{ on } S, \qquad i = 1,2,\ldots,k \tag{8.178a}$$

$$B_iw(P,t) + C_i\ddot{w}(P,t) = 0, \qquad P \text{ on } S, \qquad i = k+1,k+2,\ldots,p \tag{8.178b}$$

where the various terms are as defined in Sec. 7.18. Alternatively, boundary conditions (8.178b) can take the form

$$B_i w(P,t) + C_i \dot{w}(P,t) = 0, \qquad P \text{ on } S, \qquad i = k+1, k+2, \ldots, p \quad (8.178c)$$

and note that Eq. (8.178b) implies the presence of a lumped mass at S and Eq. (8.178c) the presence of a damper.

In Sec. 7.18, we discussed various cases of damping in which the classical modal analysis permitted a closed-form solution of Eqs. (8.177) and (8.178), where "classical" is in the sense that the eigenfunctions corresponding to the self-adjoint system, i.e., the undamped system, are capable of diagonalizing the damped system. The most important of these is the case of proportional damping. In this section, we consider the general case in which the classical modal analysis is not able to diagonalize the system, so that no closed-form solution is possible.

Even when the operator L is self-adjoint, general damping renders the system non-self-adjoint, which implies complex eigensolutions. As demonstrated in Refs. 14 and 32, approximate solutions of differential eigenvalue problems for non-self-adjoint systems in terms of quasi-comparison functions exhibit superior convergence characteristics. In view of this, we consider a solution of Eq. (8.177) by the Galerkin method in the form of the linear combination

$$w(P,t) \cong w^{(n)}(P,t) = \boldsymbol{\phi}^T(P)\mathbf{q}(t) \quad (8.179)$$

where $\boldsymbol{\phi}(P)$ is an n-vector of quasi-comparison functions and $\mathbf{q}(t)$ an n-vector of generalized coordinates. Inserting Eq. (8.179) into Eq. (8.177), premultiplying by $\boldsymbol{\phi}(P)$ and integrating over D, we obtain a set of discretized equations having the form

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{Q}(t) \quad (8.180)$$

in which

$$M = \int_D m\boldsymbol{\phi}\boldsymbol{\phi}^T\,dD, \qquad C = \int_D \boldsymbol{\phi}C\boldsymbol{\phi}^T\,dD, \qquad K = \int_D \boldsymbol{\phi}L\boldsymbol{\phi}^T\,dD \quad (8.181)$$

are a mass matrix, damping matrix and stiffness matrix, respectively, and

$$\mathbf{Q}(t) = \int_D \boldsymbol{\phi} f\,dD \quad (8.182)$$

is a generalized force vector. No confusion should arise from the fact that we used the same notation for the damping matrix and the damping operator.

To obtain a solution of Eq. (8.180), we use the approach of Sec. 4.10 and cast the equation in the state form

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{Q}(t) \quad (8.183)$$

where $\mathbf{x} = [\mathbf{q}^T\ \dot{\mathbf{q}}^T]^T$ is the state vector and

$$A = \left[\begin{array}{c|c} 0 & I \\ \hline -M^{-1}K & -M^{-1}C \end{array}\right], \qquad B = \left[\begin{array}{c} 0 \\ \hline M^{-1} \end{array}\right] \quad (8.184)$$

are coefficient matrices. As shown in Sec. 4.10, the solution of Eq. (8.183) is

$$\mathbf{x}(t) = \Phi(t)\mathbf{x}(0) + \int_0^t \Phi(t - \tau)B\mathbf{Q}(\tau)\,d\tau \tag{8.185}$$

where $\mathbf{x}(0) = [\mathbf{q}^T(0)\;\dot{\mathbf{q}}^T(0)]^T$ is the initial state vector and $\Phi(t-\tau) = \exp A(t-\tau)$ is the transition matrix. Note that the initial generalized displacement and velocity vectors can be obtained from the actual initial displacement $w(P,0)$ and velocity $\dot{w}(P,0)$ by writing

$$\mathbf{q}(0) = M^{-1}\int_D m\boldsymbol{\phi}w(P,0)\,dD, \qquad \dot{\mathbf{q}}(0) = M^{-1}\int_D m\boldsymbol{\phi}\dot{w}(P,0)\,dD \tag{8.186a, b}$$

The solution of Eq. (8.183) can also be obtained by means of a state version of the modal analysis (Sec. 4.10). A modal solution of Eq. (8.183) has the advantage of permitting truncation of inaccurate higher modes, provided there is some assurance that they are not excited.

8.11 COMPONENT-MODE SYNTHESIS

In the 1950s, it became apparent that the techniques for analyzing complex structures were woefully inadequate. This led to the independent development of two techniques, the finite element method and component-mode synthesis. The first to emerge was the finite element method (Ref. 46), initially conceived as a static analysis according to which the structure is divided into small subdomains, referred to as finite elements, and the deformation over each element is described in terms of interpolation functions. Since its early beginnings, the finite element method has grown significantly in scope, finding application in a large variety of engineering areas, as well as in applied mathematics. The entire Chapter 9 is devoted to the finite element method.

Following by a few years, Hurty (Refs. 18–20) developed the component-mode synthesis as a technique for the dynamic analysis of structures consisting of assemblages of substructures. The component-mode synthesis adopts a different point of view from the finite element method, as the modeling is carried out on a much larger scale. Indeed, the general idea is to describe the motion separately over each of the substructures, referred to as components, and then constrain the components to work together as a single structure. In fact, component-mode synthesis can be regarded as an extension of the assumed-modes method to flexible multibody systems. Indeed, as in the assumed-modes method, the motion of each component is described by a linear combination of modes multiplied by time-dependent generalized coordinates. Hurty divides the component modes into three types, rigid-body modes, constraint modes and normal modes. But, because each component is modeled separately, there are redundant coordinates, as points shared by two components undergo the same motions. The removal of redundant coordinates is carried out during an assembling process in which the constituent components are constrained to act as a whole structure.

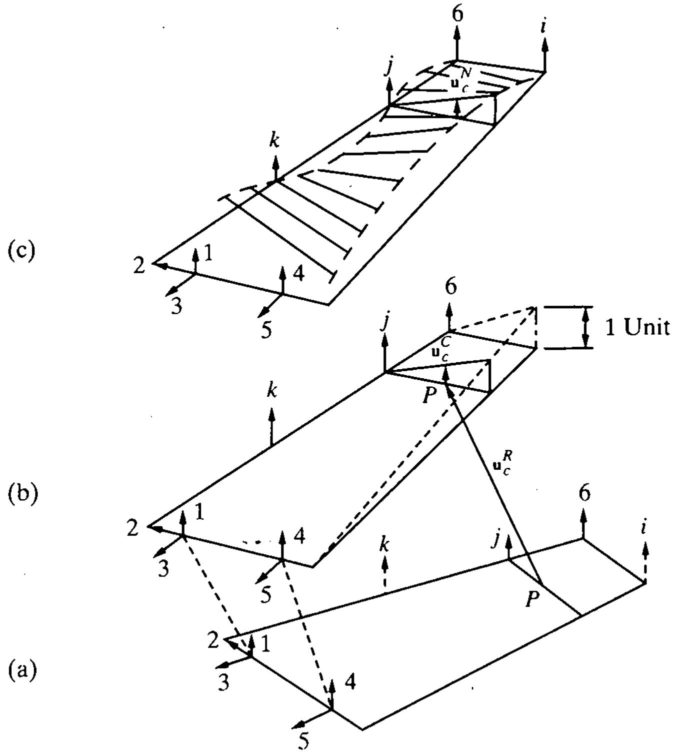


Figure 8.15 **(a)** Undisplaced, undeformed component **(b)** Component after it has undergone rigid-body displacement **(c)** Displacements relative to fixed-constraint motions

We consider a structure consisting of N components, concentrate on a typical component c $(c = 1, 2, \dots, N)$ and express the total displacement vector of an arbitrary point $P(x, y, z)$ in the form

$$\mathbf{u}_c(P, t) = \mathbf{u}_c^R(P, t) + \mathbf{u}_c^C(P, t) + \mathbf{u}_c^N(P, t) \tag{8.187}$$

where $\mathbf{u}_c^R$ is a rigid-body displacement vector, $\mathbf{u}_c^C$ a "constraint displacement" vector and $\mathbf{u}_c^N$ a displacement vector relative to the fixed constraints, as shown in Fig. 8.15. Figure 8.15a displays the undisplaced, undeformed component, with a set of constraints indicated by arrows. The constraints labeled 1–6 are regarded as statically determinate and the constraints i, j and k are considered as redundant. All the constraints are movable and the result of component c being attached to adjacent components that are themselves in motion. The points shared with a given adjacent component represent the *interface* with the component in question. Figure 8.15b shows the component after it has undergone six arbitrary rigid-body displacements, defined uniquely by the displacements of the six statically determinate constraints. These displacements are represented by the rigid-body displacement vector $\mathbf{u}_c^R$. The constraint displacement vector $\mathbf{u}_c^C$ results from the motion of the redundant coordinates relative to the rigid-body motions and it represents a linear combination of displacements obtained by letting each of the redundant constraints undergo an arbitrary displacement, as shown in Fig. 8.15b. In addition to these displacements, there is the displacement $\mathbf{u}_c^N$ of point P relative to the constraint motions, as depicted in Fig. 8.15c.

At this point, we begin the discretization process. To this end, we represent the three types of displacements as linear combinations of space-dependent functions multiplied by time-dependent generalized coordinates in the form

$$\begin{aligned} \mathbf{u}_c^R(P,t) &= \Phi_c^R(P)\boldsymbol{\zeta}_c^R(t), \\ \mathbf{u}_c^C(P,t) &= \Phi_c^C(P)\boldsymbol{\zeta}_c^C(t), \\ \mathbf{u}_c^N(P,t) &= \Phi_c^N(P)\boldsymbol{\zeta}_c^N(t) \end{aligned} \tag{8.188}$$

where Φ_c^R is in general a 3×6 matrix of rigid-body modes, Φ_c^C is a matrix of constraint modes with three rows and as many columns as the number of redundant constraints, a finite number, Φ_c^N is a matrix of "fixed-constraint" normal modes with three rows and a given finite number of columns and $\boldsymbol{\zeta}_c^R$, $\boldsymbol{\zeta}_c^C$ and $\boldsymbol{\zeta}_c^N$ are associated generalized displacement vectors. Equations (8.187) and (8.188) can be combined into the single expression

$$\mathbf{u}_c(P,t) = \Phi_c(P)\boldsymbol{\zeta}_c(t) \tag{8.189}$$

in which $\Phi_c = [\Phi_c^R \ \Phi_c^C \ \Phi_c^N]$ and $\boldsymbol{\zeta}_c = [(\boldsymbol{\zeta}_c^R)^T \ (\boldsymbol{\zeta}_c^C)^T \ (\boldsymbol{\zeta}_c^N)^T]^T$, and we observe that the net effect of Eq. (8.189) is to represent the motion of the distributed component by a finite number of degrees of freedom.

The next step is to derive the equations of motion for the discretized component. Using the analogy with the assumed-modes method (Sec. 8.7), we carry out this task by means of Lagrange's equations. To this end, we use Eq. (8.189) and write the component kinetic energy in the discretized form

$$T_c(t) = \frac{1}{2}\int_{D_c} m_c(P)\dot{\mathbf{u}}_c^T(P,t)\dot{\mathbf{u}}_c(P,t)\,dD_c = \frac{1}{2}\dot{\boldsymbol{\zeta}}_c^T(t)M_c\dot{\boldsymbol{\zeta}}_c(t) \tag{8.190}$$

where D_c is the domain of component c and

$$M_c = \int_{D_c} m_c(P)\Phi_c^T(P)\Phi_c(P)\,dD_c \tag{8.191}$$

is the corresponding component mass matrix. Assuming that the component is subjected to distributed viscous damping and denoting by $c_c(P)$ the distributed damping coefficient, we obtain the discretized Rayleigh's dissipation function (Sec. 4.1) for the component

$$\mathcal{F}_c = \frac{1}{2}\int_{D_c} c_c(P)\dot{\mathbf{u}}_c^T(P,t)\dot{\mathbf{u}}_c(P,t)\,dD_c = \frac{1}{2}\dot{\boldsymbol{\zeta}}_c^T(t)C_c\dot{\boldsymbol{\zeta}}_c(t) \tag{8.192}$$

in which

$$C_c = \int_{D_c} c_c(P)\Phi_c^T(P)\Phi_c(P)\,dD_c \tag{8.193}$$

is the associated component viscous damping matrix. Moreover, using the notation of Sec. 8.7, the component potential energy takes the discretized form

$$V_c(t) = \frac{1}{2}[\mathbf{u}_c^T, \mathbf{u}_c] = \frac{1}{2}\zeta_c^T(t)K_c\zeta_c(t) \tag{8.194}$$

where

$$K_c = [\Phi_c^T(P), \Phi_c(P)] \tag{8.195}$$

is the corresponding component stiffness matrix. Finally, assuming that the component is acted upon by the distributed force $\mathbf{f}_c(P, t)$, the discretized virtual work for the component can be written as

$$\overline{\delta W}_c(t) = \int_{D_c} \mathbf{f}_c^T(P, t)\delta\mathbf{u}_c(P, t)\, dD_c = \mathbf{Z}_c^T(t)\delta\zeta_c(t) \tag{8.196}$$

in which

$$\mathbf{Z}_c(t) = \int_{D_c} \Phi_c^T(P)\mathbf{f}(P, t)\, dD_c \tag{8.197}$$

is the associated generalized force vector for the component, and we note that $\mathbf{Z}_c$ excludes viscous damping forces.

By analogy with Eq. (4.14), Lagrange's equations of motion for the discretized component have the symbolic form

$$\frac{d}{dt}\left(\frac{\partial L_c}{\partial \dot{\zeta}_c}\right) - \frac{\partial L_c}{\partial \zeta_c} + \frac{\partial \mathcal{F}_c}{\partial \dot{\zeta}_c} = \mathbf{Z}_c \tag{8.198}$$

where $L_c = T_c - V_c$ is the component Lagrangian. Hence, using Eqs. (8.190), (8.192) and (8.194), we obtain the component equations of motion

$$M_c\ddot{\zeta}_c(t) + C_c\dot{\zeta}_c(t) + K_c\zeta_c(t) = \mathbf{Z}_c(t), \qquad c = 1, 2, \dots, N \tag{8.199}$$

Next, we turn our attention to the assembling process. To this end, we first collect all component equations into the "disjoint" set of equations

$$M^{(d)}\ddot{\zeta}(t) + C^{(d)}\dot{\zeta}(t) + K^{(d)}\zeta(t) = \mathbf{Z}(t) \tag{8.200}$$

where $\zeta = [\zeta_1^T\ \zeta_2^T\ \dots\ \zeta_N^T]^T$, $\mathbf{Z} = [\mathbf{Z}_1^T\ \mathbf{Z}_2^T\ \dots\ \mathbf{Z}_N^T]^T$ are disjoint displacement and force vectors and

$$M^{(d)} = \text{block-diag}[M_c], \qquad C^{(d)} = \text{block-diag}[C_c], \qquad K^{(d)} = \text{block-diag}[K_c] \tag{8.201}$$

are disjoint coefficient matrices. Of course, according to Eq. (8.200), the components still act independently of one another, as the vector ζ contains all the redundant coordinates. The assembling process is designed to cause the disjoint set of components to act as a single structure, which implies elimination of the redundant coordinates. If we assume that two adjacent components r and s are joined together so that there are no relative translations and rotations between the components at the interface, then we have

$$\mathbf{u}_r = \mathbf{u}_s, \qquad \boldsymbol{\theta}_r = \boldsymbol{\theta}_s \tag{8.202}$$

where $\boldsymbol{\theta}$ represents a rotation vector, which implies that the rotations are small. But, the translational and rotational displacements at the interfaces are related to the generalized displacement vector $\boldsymbol{\zeta}$ by means of equations of the type (8.189), in which P represents the position of the interface points. In view of this, we can combine Eqs. (8.202) corresponding to all interfaces into a single constraint equation having the general form

$$A\boldsymbol{\zeta} = \mathbf{0} \tag{8.203}$$

in which A is a $c \times m$ matrix, where c is the number of constraint equations. Then, dividing the vector $\boldsymbol{\zeta}$ into an n-vector $\mathbf{q}$ of independent variables and a vector $\mathbf{d}$ of dependent variables and partitioning the matrix A as follows:

$$A = \left[A_1 \;\vdots\; A_2 \right]$$

where A_1 is a $c \times n$ matrix and A_2 is a $c \times c$ nonsingular matrix, we can rewrite Eq. (8.203) as

$$A_1\mathbf{q} + A_2\mathbf{d} = \mathbf{0} \tag{8.205}$$

which yields

$$\mathbf{d} = -A_2^{-1}A_1\mathbf{q} \tag{8.206}$$

Equation (8.206) permits us to write a relation between the n-vector $\mathbf{q}(t)$ of independent generalized coordinates for the full structure and the m-vector $\boldsymbol{\zeta}(t)$, which includes redundant coordinates, in the form

$$\boldsymbol{\zeta}(t) = B\mathbf{q}(t) \tag{8.207}$$

in which

$$B = \begin{bmatrix} I \\ \hdashline -A_2^{-1}A_1 \end{bmatrix} \tag{8.208}$$

is an $m \times n$ matrix, where $n = m - c$ is the number of degrees of freedom of the model of the full structure. Introducing Eq. (8.207) into Eq. (8.200) and premultiplying by B^T, we obtain the *coupled* equations of motion of the full structure

$$M\ddot{\mathbf{q}}(t) + C\dot{\mathbf{q}}(t) + K\mathbf{q}(t) = \mathbf{Q}(t) \tag{8.209}$$

where

$$M = B^T M^{(d)} B, \qquad C = B^T C^{(d)} B, \qquad K = B^T K^{(d)} B \tag{8.210}$$

are $n \times n$ coefficient matrices and

$$\mathbf{Q}(t) = B^T\mathbf{Z}(t) \tag{8.211}$$

is an n-dimensional generalized force vector.

There are two main questions still to be addressed, namely, the choice of constraint modes and normal modes and the nature of the approximation. Hurty (Ref. 19) answers the first question, but does not address the second. In particular, the constraint modes are defined clearly in Ref. 19 as "... displacements produced by giving each redundant constraint in turn an arbitrary displacement while keeping all other constraints fixed." As far as the normal modes are concerned, they are defined in Ref. 19 somewhat more ambiguously by stating that "... it is convenient, although not necessary, to think of these as the 'fixed-constraint' natural modes of vibration of the structure." In practice, they have been widely interpreted as fixed-fixed component modes. The nature of the approximation is significantly more involved in the component-mode synthesis than in the assumed-modes method. The question is essentially how well a combination of the constraint modes and component normal modes can approximate the behavior of the component in the context of a full structure. The question of inaccuracies introduced by the use of constraint modes arises only in problems in which the interface is a line, such as when the components are two-dimensional, rather than a point, such as when the components are one-dimensional. Clearly, if the interface is a line, then there is an infinite number of interface points, and not a finite number. As a result, the boundary conditions internal to the structure at the interface cannot be satisfied exactly. To explain this point, it is convenient to conceive of a structure with the interface consisting of a finite number of points, where the points coincide with the location of the redundant constraints. We refer to this fictitious structure as an *intermediate structure* (Ref. 27), because the model lies somewhere between the disjoint structure and the fully coupled structure, where in the latter the full infinity of interface points is considered. It is obvious that *the results of the component-modes synthesis are valid for the intermediate structure*, and not necessarily for the actual structure. The component normal modes do not require elaboration, as fixed-fixed modes are generally well defined and the nature of the approximation is the same as in the Rayleigh-Ritz method applied to a single elastic member.

Since publication of Hurty's component-mode synthesis, there have been many attempts to enhance its accuracy. The issues of component modeling and of the manner in which the various components are made to work together as a whole structure have received a great deal of attention. Hurty's method makes a sharp distinction between determinate and indeterminate constraints. In reality, no such distinction exists, and all interface constraints should really receive equal treatment. This is the essence of an idea advanced by Craig and Bampton (Ref. 6), who suggested a simplification in the treatment of the component rigid-body modes by eliminating the separation of the boundary constraints into determinate and indeterminate ones. All constraint modes are defined as the mode shapes due to successive unit displacements at each of the interface points, with all other interface points being fixed. Craig and Bampton envision an entirely discretized structure, with the constraint modes being generated by matrix operations on a computer. The normal modes and the elimination of redundant coordinates remain essentially the same as envisioned by Hurty.

The method described above is generally referred to as a "fixed constraint mode" method, because the modes used to describe the motion correspond to fixed constraints. To account for motions caused by loads at unconstrained points, in developing a computer program for the method, Bamford (Ref. 1) introduced another class of displacement modes, referred to as "attachment modes" and defined as the static deflection of the component resulting from a unit force applied at one boundary coordinate while the remaining coordinates are force free. The possibility of using unconstrained modes has been suggested by Goldman (Ref. 12) and by Hou (Ref. 17). Some ill-conditioning problems have been experienced in using unconstrained modes. The use of unconstrained modes has also been proposed by Dowell (Refs. 7 and 21), who used Lagrange's multipliers to enforce continuity at interfaces. A somewhat different type of mode selection is advocated by MacNeal (Ref. 24) and Rubin (Ref. 42). Indeed, they use a low-frequency subset of the free-free component modes together with "residual modes," some shape functions designed to capture the contribution from the truncated normal modes. Truncation problems have been discussed by Kuhar and Stahle (Ref. 22), who present a condensation scheme similar to dynamic condensation (Ref. 28), and by Hintz (Ref. 15), who identified Hurty's elimination of redundant coordinates as a static condensation (Ref. 28).

A method developed by Benfield and Hruda (Ref. 2), and known as *component-mode substitution*, resembles both Hurty's component-mode synthesis and Gladwell's branch-mode analysis (Ref. 11). The interest in component-mode substitution lies in its connection with component-mode synthesis. The method differs from component-mode synthesis in that the component modes need not be constrained and can be free-free. More importantly, however, the component-mode substitution does not require that the generalized coordinates of the static constraint modes appear in the final formulation, thus reducing the number of degrees of freedom of the over all model. The efficiency of the method can be improved by applying stiffness and inertial loadings to the free interface coordinates of the component under consideration to account for the effect of the remaining components. It is this aspect of the method that is of interest here, as this procedure gives rise to so-called "loaded interface modes."

A comparison of the various choices of modes for Hurty's component-mode synthesis, focusing on the mode sets advocated by Craig-Bampton, MacNeal and Rubin and Benfield-Hruda is presented by Spanos and Tsuha (Ref. 43). In addition, they discuss the effect of controls on the reduction of the component order.

Finally, there is the question of the manner in which the redundant coordinates at interfaces are handled. In Hurty's component-mode synthesis, the redundant coordinates are eliminated by means of the linear transformation given by Eq. (8.207). On the other hand, Dowell (Ref. 7) and Klein and Dowell (Ref. 21) use Lagrange's multipliers for the same purpose.

The component-mode synthesis represents a sound heuristic, physically motivated approach to the dynamics of complex structures. With a proper choice of modes, the method should be capable of yielding reasonable results with a relatively small number of degrees of freedom. However, unlike the assumed-modes method, which could invoke the Rayleigh-Ritz theory to claim convergence, the component-mode synthesis cannot make such claims. Indeed, it can be argued at best that the eigenvalues of the model converge to the eigenvalues of the intermediate structure. The degree to which the intermediate structure approximates the actual structure is still an open question, and the answer depends on the extent to which the linear combination of constraint modes and normal modes is capable of satisfying the internal boundary conditions at the interfaces. This question is addressed in Sec. 8.12.

Because the finite element method and component-mode synthesis were developed with the same objective in mind, namely, to analyze complex structures, there is the perception that the two methods compete with one another. In this regard, it should be pointed out that, whereas the finite element method is capable of producing an accurate mathematical model of a complex structure, without many of the problems associated with the selection of suitable sets of modes, the model is likely to be of extremely large order. On the other hand, a model produced by component-mode synthesis is likely to be of significantly smaller order. In view of this, under certain circumstances the two methods can be regarded as complementing one another. Indeed, it is possible to produce the component normal modes by means of the finite element method and then use them in conjunction with the component-mode synthesis to reduce the number of degrees of freedom of the model. In fact, such an approach would be consistent with the process of replacing interface lines by interface points, as discussed above. A word of caution is in order, however, as the number of interface points in a finite element model tends to be much larger than the number of interface points in a component-mode synthesis, so that the concept of an intermediate structure still applies.

8.12 SUBSTRUCTURE SYNTHESIS

Many structures, such as fixed-wing aircraft, helicopters, flexible spacecraft, flexible robots, a variety of civil structures, etc., can be modeled as assemblages of interacting flexible bodies. Hurty's component-mode synthesis discussed in Sec. 8.10 consists of representing the motion of each of the constituent substructures by means of linear combinations of rigid-body modes, constraint modes and normal modes. Various other investigators use different sets of modes in an attempt to improve the convergence of the component-mode synthesis. Note that, to generate component modes, it is generally assumed that one must solve a component eigenvalue problem (see, for example, Ref. 43).

The component-mode synthesis is basically an extension of the assumed-modes method to flexible multibodies. In essence, the various approaches discussed in Sec. 8.10, whereby different sets of component modes are used, represent special cases of the Rayleigh-Ritz method (Refs. 26 and 27). Clearly, a proper choice of component modes can produce very good results. However, in the spirit of the Rayleigh-Ritz theory, approximate solutions can be constructed from the space of admissible functions, i.e., the functions need not be modes at all. In this regard, it should be mentioned that the practice of using component modes has practical implications when the various substructures are manufactured by different companies and the structural characteristics are provided in the form of component modes. In such cases, the component-mode synthesis can be used to generate a structural model for the fully assembled structure. Still, however defined, component modes represent mere subspaces of the much larger space of admissible functions, and component-mode synthesis is part of a larger picture.

In this section, we discuss a method for the modeling of flexible multibody systems developed in Ref. 31 for systems of the type shown in Fig. 8.16. The method represents an extension of the theory developed in Sec. 8.6 for single elastic members to flexible multibody systems. Because the theory is based on expansions of the solution over the individual substructures in terms of special classes of admissible functions, rather than modes, and the mathematical formulation of the equations of motion is entirely different from that in component-mode synthesis, the method is referred to as *substructure synthesis*.

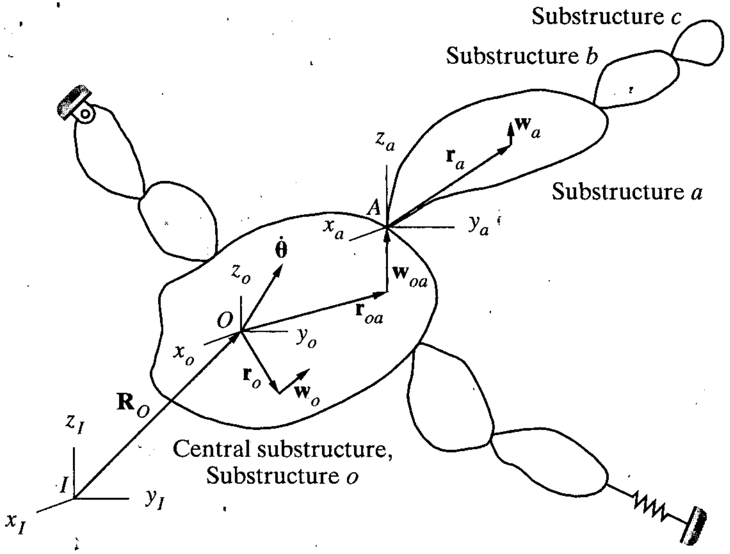


Figure 8.16 Flexible multibody system

In the Rayleigh-Ritz method, a minimizing sequence is constructed from the space of admissible functions or comparison functions, depending on the form of Rayleigh's quotient. If Rayleigh's quotient is in the form of Eq. (8.89), then admissible functions suffice. In the case of flexible multibody systems, boundary conditions cannot be defined independently of the motions of the adjacent substructures, so that comparison functions cannot be generated. Hence, the only alternative is the use of admissible functions, which include the various "substructure modes" as special cases. This, however, raises serious questions concerning the speed of convergence, as demonstrated in Sec. 8.6 for a single elastic member. This suggests the construction of an approximate solution over each of the flexible substructures from the space of quasi-comparison functions. Moreover, the geometric compatibility at interface points is ensured automatically by a kinematical procedure describing the motion of each point of the structure in a consistent manner, which obviates the question of constraints.

We propose to derive first the free-vibration equations of motion for flexible multibody systems of the type shown in Fig. 8.16. Then, the eigenvalue problem follows immediately from the free-vibration equations. Because the eigenvalue problem represents a linear problem, the interest lies in linear equations of motion, which implies that all displacements must be small, including the rotations. We derive the equations of motion by means of Lagrange's equations, which in the case of free vibration of undamped systems amounts to deriving the kinetic energy and potential energy. The latter two are fully defined by the mass matrix and stiffness matrix, respectively.

The task of deriving the equations of motion can be simplified appreciably by adopting a consistent kinematical procedure for describing the motion. To this end, we introduce an inertial set of axes $X_I Y_I Z_I$ with the origin at I, a set of body axes $x_o y_o z_o$ with the origin at O and attached to substructure o in the undeformed state, a set of body axes $x_a y_a z_a$ with the origin at A and attached to substructure a in the undeformed state, etc. The various axes are shown in Fig. 8.16. For simplicity, we limit the formulation to substructures of the type o and a. Extension of the formulation to substructures of the type b, c, etc., is obvious, but tends to be exceedingly laborious. To carry out the extension, we observe that the motion of b relates to the motion of a in the same manner as the motion of a relates to the motion of o. The position vector of typical points in o and a can be written as

$$\mathbf{R}_o = \mathbf{R}_O + \mathbf{r}_o + \mathbf{w}_o \tag{8.212a}$$

and

$$\mathbf{R}_a = \mathbf{R}_O + \mathbf{r}_{oa} + \mathbf{w}_{oa} + \mathbf{r}_a + \mathbf{w}_a, \qquad a = 1, 2, \ldots, N \tag{8.212b}$$

respectively, where $\mathbf{R}_O$ is the radius vector from I to O, $\mathbf{r}_o$ the radius vector from O to a typical point in o, $\mathbf{w}_o$ the elastic displacement vector of the typical point in o measured relative to axes $x_o y_o z_o$, $\mathbf{r}_{oa}$ the radius vector from O to A, $\mathbf{w}_{oa}$ the vector $\mathbf{w}_o$ evaluated at A, $\mathbf{r}_a$ the radius vector from A to a typical point in a and $\mathbf{w}_a$ the elastic displacement of the typical point in a measured relative to axes $x_a y_a z_a$. Note that all vectors are in terms of components along local axes.

The Lagrangian formulation requires the kinetic energy, which in turn requires the velocity of typical points in the various substructures. To derive expressions for these velocities, we assume that axes $x_o y_o z_o$ rotate with the angular velocity $\dot{\boldsymbol{\theta}}$ relative to the inertial space and that axes $x_a y_a z_a$ rotate with the angular velocity $\dot{\boldsymbol{\beta}}_a$ relative to axes $x_o y_o z_o$, due to the elastic motion at A. Recalling that we are interested in linearized equations of motion, the velocity vector for a typical point in substructure o and substructure a can be written as follows:

$$\dot{\mathbf{R}}_o = \dot{\mathbf{R}}_O + \tilde{r}_o^T \dot{\boldsymbol{\theta}} + \dot{\mathbf{w}}_o \tag{8.213a}$$

$$\dot{\mathbf{R}}_a = \dot{\mathbf{R}}_O + \left(\tilde{r}_{oa}^T + C_a^T \tilde{r}_a^T C_a\right) \dot{\boldsymbol{\theta}} + \dot{\mathbf{w}}_{oa} + C_a^T \tilde{r}_a^T C_a \dot{\boldsymbol{\beta}}_a + C_a^T \dot{\mathbf{w}}_a, \quad a = 1, 2, \ldots, N \tag{8.213b}$$

where $\dot{\mathbf{R}}_O$ is the velocity vector of O and C_a is a matrix of direction cosines between $x_a y_a z_a$ and $x_o y_o z_o$. Moreover, $\boldsymbol{\beta}_a = \nabla \times \mathbf{w}_{oa}$, and we note that, in writing the angular displacements due to elastic deformations in vector form, we take into account that these deformations are small. We also note that a tilde over a symbol denotes a skew symmetric matrix formed from the corresponding vector (Sec. 2.6).

The degrees of freedom of the system are associated with the rigid-body motions of the frame $x_o y_o z_o$ and the elastic motions of the substructures. We assume that the elastic displacements can be expressed in the form

$$\mathbf{w}_s(\mathbf{r}_s, t) = \Phi_s(\mathbf{r}_s)\mathbf{q}_s(t), \qquad s = o, a;\ a = 1, 2, \ldots, N \tag{8.214}$$

where Φ_s are matrices of trial functions and $\mathbf{q}_s$ are vectors of generalized displacements. Using Eqs. (8.213) in conjunction with Eqs. (8.214), the kinetic energy can be reduced to the form

$$\begin{aligned}
T &= \frac{1}{2}\int_{D_o} \rho_o \dot{\mathbf{R}}_o^T \dot{\mathbf{R}}_o \, dD_o + \frac{1}{2}\sum_{a=1}^{N}\int_{D_a} \rho_a \dot{\mathbf{R}}_a^T \dot{\mathbf{R}}_a \, dD_a \\
&= \frac{1}{2} m_t \mathbf{V}_O^T \mathbf{V}_O - \mathbf{V}_O^T \tilde{S}_t \boldsymbol{\omega} + \mathbf{V}_O^T \overline{\Phi}_t \dot{\mathbf{q}}_o + \mathbf{V}_O^T \sum_{a=1}^{N} C_a^T \overline{\Phi}_a \dot{\mathbf{q}}_o + \frac{1}{2}\boldsymbol{\omega}^T I_t \boldsymbol{\omega} \\
&\quad + \boldsymbol{\omega}^T \tilde{\Phi}_t \dot{\mathbf{q}}_o + \boldsymbol{\omega}^T \sum_{a=1}^{N} H_a \dot{\mathbf{q}}_a + \frac{1}{2}\dot{\mathbf{q}}_0^T M_t \dot{\mathbf{q}}_o + \dot{\mathbf{q}}_o^T \sum_{a=1}^{N} J_a \dot{\mathbf{q}}_a + \frac{1}{2}\sum_{a=1}^{N} \dot{\mathbf{q}}_a^T M_a \dot{\mathbf{q}}_a \\
&= \frac{1}{2}\dot{\mathbf{x}}^T M \dot{\mathbf{x}}
\end{aligned} \tag{8.215}$$

where $\mathbf{x} = \left[\mathbf{R}_O^T \; \boldsymbol{\theta}^T \; \mathbf{q}_o^T \; \mathbf{q}_1^T \; \mathbf{q}_2^T \ldots \mathbf{q}_N^T\right]^T$ is the configuration vector, in which we note that $\dot{\mathbf{R}}_o = \mathbf{V}_o$ and $\dot{\boldsymbol{\theta}} = \boldsymbol{\omega}$, and

$$M = \begin{bmatrix}
m_t I & \tilde{S}_t^T & \overline{\Phi}_t & C_1^T \overline{\Phi}_1 & C_2^T \overline{\Phi}_2 & \ldots & C_N^T \overline{\Phi}_N \\
\tilde{S}_t & I_t & \tilde{\Phi}_t & H_1 & H_2 & \ldots & H_N \\
\overline{\Phi}_t^T & \tilde{\Phi}_t^T & M_t & J_1 & J_2 & \ldots & J_N \\
\overline{\Phi}_1^T C_1 & H_1^T & J_1^T & M_1 & 0 & \ldots & 0 \\
\overline{\Phi}_2^T C_2 & H_2^T & J_2^T & 0 & M_2 & \ldots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\overline{\Phi}_N^T C_N & H_N^T & J_N^T & 0 & 0 & \ldots & M_n
\end{bmatrix} \tag{8.216}$$

is the mass matrix. The various quantities entering into Eq. (8.216) are as follows:

$$m_t = m_o + \sum_{a=1}^{N} m_a, \qquad \tilde{S}_t = \tilde{S}_o + \sum_{a=1}^{N}\left(m_a \tilde{r}_{oa} + C_a^T \tilde{S}_a C_a\right)$$

$$\overline{\Phi}_t = \overline{\Phi}_o + \sum_{a=1}^{N}\left(m_a \Phi_{oa} - C_a^T \tilde{S}_a C_a \Upsilon_{oa}\right)$$

$$I_t = I_o + \sum_{a=1}^{N}\left(C_a^T I_a C_a - m_a \tilde{r}_{oa}^2 - \tilde{r}_{oa} C_a^T \tilde{S}_a C_a - C_a^T \tilde{S}_a C_a \tilde{r}_{oa}\right)$$

$$\tilde{\Phi}_t = \tilde{\Phi}_o + \sum_{a=1}^{N}\left[\left(m_a \tilde{r}_{oa} + C_a^T \tilde{S}_a C_a\right)\Phi_{oa} + \left(C_a^T I_a C_a - \tilde{r}_{oa} C_a^T \tilde{S}_a C_a\right)\Upsilon_{oa}\right]$$

$$\begin{aligned}
M_t = M_o + \sum_{a=1}^{N}\Big(& m_a \Phi_{oa}^T \Phi_{oa} - \Phi_{oa}^T C_a^T \tilde{S}_a C_a \Upsilon_{oa} \\
& + \Upsilon_{oa}^T C_a^T \tilde{S}_a C_a \Phi_{oa} + \Upsilon_{oa}^T C_a^T I_a C_a \Upsilon_{oa}\Big)
\end{aligned}$$

$$H_s = C_s^T \tilde{\Phi}_s + \tilde{r}_{os} C_s^T \overline{\Phi}_s, \qquad J_s = \Gamma_s^T C_s^T \tilde{\Phi}_s + \Phi_{os}^T C_s^T \overline{\Phi}_s,$$
$$M_s = \int_{D_s} \rho_s \Phi_s^T \Phi_s \, dD_s \qquad s = o, a;\ a = 1, 2, \ldots, N \tag{8.217}$$

in which

$$\begin{aligned} m_s &= \int_{D_s} \rho_s \, dD_s, & \tilde{S}_s &= \int_{D_s} \rho_s \tilde{r}_s \, dD_s, \\ \Upsilon_{oa} &= \nabla \times \Phi_{oa}, & I_s &= \int_{D_s} \rho_s \tilde{r}_s \tilde{r}_s^T \, dD_s \\ \overline{\Phi}_s &= \int_{D_s} \rho_s \Phi_s \, dD_s, & \tilde{\Phi}_s &= \int_{D_s} \rho_s \tilde{r}_s \Phi_s \, dD_s, \\ \Phi_{os} &= \Phi_o(\mathbf{r}_{os}), & \Gamma_s &= \nabla \times \Phi_s(\mathbf{r}_{os}) \end{aligned} \tag{8.218}$$

The potential energy is assumed to be due entirely to the elastic deformations and can be written as

$$V = \frac{1}{2}\mathbf{q}_o^T K_o \mathbf{q}_o + \sum_{a=1}^{N} \frac{1}{2}\mathbf{q}_a^T K_a \mathbf{q}_a + \sum_{b=1}^{N} \frac{1}{2}\mathbf{u}_b^T K_b \mathbf{u}_b = \frac{1}{2}\mathbf{x}^T K \mathbf{x} \tag{8.219}$$

where

$$K_s = \left[\Phi_s^T, \qquad \Phi_s\right], \qquad s = o, a \tag{8.220}$$

are substructure stiffness matrices, K_b are boundary stiffness matrices due to the action of the springs at the boundary points B, and K is the overall stiffness matrix for the whole structure. Moreover,

$$\mathbf{u}_b \cong \mathbf{R}_O + \left(\tilde{r}_{oa}^T + C_a^T \tilde{r}_{ab}^T C_a\right)\boldsymbol{\theta} + \Phi_{oa}\mathbf{q}_o + C_a^T \Phi_{ab}\mathbf{q}_a \tag{8.221}$$

represents the displacement vector of point B, in which $\Phi_{ab} = \Phi_a(\mathbf{r}_{ab})$. The other quantities on the right side of Eq. (8.221) were defined earlier. The overall stiffness matrix can be written in the form

$$K = \begin{bmatrix} \kappa_{aa} & \kappa_{12} & \kappa_{13} & \kappa_{14}^1 & \kappa_{14}^2 & \cdots & \kappa_{14}^N \\ (\kappa_{12})^T & \kappa_{22} & \kappa_{23} & \kappa_{24}^1 & \kappa_{24}^2 & \cdots & \kappa_{24}^N \\ (\kappa_{13})^T & (\kappa_{23}^1)^T & K_o + \kappa_{33} & \kappa_{34}^1 & \kappa_{34}^2 & \cdots & \kappa_{34}^N \\ (\kappa_{14}^1)^T & (\kappa_{24}^1)^T & (\kappa_{34}^1)^T & K_1 + \kappa_{44}^1 & 0 & \cdots & 0 \\ (\kappa_{14}^2)^T & (\kappa_{24}^2)^T & (\kappa_{34}^2)^T & 0 & K_2 + \kappa_{44}^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (\kappa_{14}^N)^T & (\kappa_{24}^N)^T & (\kappa_{34}^N)^T & 0 & 0 & \cdots & K_N + \kappa_{44}^N \end{bmatrix} \tag{8.222}$$

in which

$$\kappa_{11} = \sum_{a=1}^{N} k_a, \qquad \kappa_{12} = -\sum_{a=1}^{N} k_a \left(\tilde{r}_{oa} + C_a^T \tilde{r}_{ab} C_a\right), \qquad \kappa_{13} = \sum_{a=1}^{N} k_a \Phi_{oa}$$

$$\kappa_{22} = -\sum_{a=1}^{N} \left(\tilde{r}_{oa} + C_a^T \tilde{r}_{ab} C_a\right) k_a \left(\tilde{r}_{oa} + C_a^T \tilde{r}_{ab} C_a\right),$$

$$\kappa_{23} = \sum_{a=1}^{N} \left(\tilde{r}_{oa} + C_a^T \tilde{r}_{ab} C_a\right) k_a \Phi_{oa} \tag{8.223}$$

$$\kappa_{33} = \sum_{a=1}^{N} \Phi_{oa}^T k_a \Phi_{oa}, \quad \kappa_{14}^a = k_a C_a^T \Phi_{ab}, \quad \kappa_{24}^a = \left(\tilde{r}_{oa} + C_a^T \tilde{r}_{ab} C_a\right) k_a C_a^T \Phi_{ab}$$

$$\kappa_{34}^a = \Phi_{oa}^T k_a C_a^T \Phi_{ab}, \qquad \kappa_{44}^a = \Phi_{ab}^T C_a k_a C_a^T \Phi_{ab}$$

As pointed out earlier in this section, Lagrange's equations for free vibrations are fully defined by the coefficient matrices in the kinetic energy and potential energy. Indeed, the equations of motion can be written simply as

$$M\ddot{\mathbf{x}}(t) + K\mathbf{x}(t) = \mathbf{0} \tag{8.224}$$

where M and K are given by Eqs. (8.216) and (8.222), respectively. Then, because free vibration of conservative systems is harmonic, $\mathbf{x}(t) = e^{i\omega t}\mathbf{x}$, Eq. (8.224) yields the eigenvalue problem

$$K\mathbf{x} = \lambda M\mathbf{x}, \qquad \lambda = \omega^2 \tag{8.225}$$

From Fig. 8.16, we observe that in the case of flexible multibody systems there are no geometric boundary conditions, except when a substructure is supported externally, so that, for the most part, the only boundary conditions characterizing a substructure are natural. It follows that, according to the Rayleigh-Ritz theory, the admissible functions need satisfy nothing. To be sure, they must be p times differentiable, but this requirement is satisfied routinely by virtually all choices. Hence, in theory, admissible functions corresponding to the modes of free-free substructures should be a suitable choice. However, this turns out not to be the case in general.

In the case of substructures in the form of beams in bending, there are four quantities entering into the boundary conditions, namely, displacement, slope, bending moment and shearing force, where the latter two involve second and third derivatives with respect to the spatial variable, respectively. In component-mode synthesis, geometric compatibility is enforced by means of constraint equations. The kinematical procedure used here ensures geometric compatibility at boundary points automatically, which is accomplished by defining the various sets of body axes so as to guarantee displacement and slope compatibility at boundary points common to any two substructures. This obviates the need for constraint equations enforcing such geometric compatibility. In addition, rigid-body motions are included in the displacement vector of the central substructure o. But, this substructure synthesis goes beyond ensuring geometric compatibility. Indeed, through the use of quasi-comparison functions, provisions are also made for balancing to a large degree the bending moment and shearing force at boundary points between any two substructures, thus satisfying the natural boundary conditions approximately at these points. In this regard, a distinction must be made between substructure o and substructures $a(a = 1, 2, \dots, N)$. Indeed, for substructure o, it is necessary to make provisions for nonzero displacement, slope, bending moment and shearing force at the boundary points A. On the other hand, because displacement and slope compatibility are guaranteed automatically by the kinematical procedure, for substructures of type a it is only necessary to make provisions for nonzero bending moment and shearing force at points A. All this is done through a judicious choice of admissible functions. For example, linear combinations of free-free functions alone do not qualify as quasi-comparison functions for substructure o, because they are characterized by zero bending moment and shearing force at boundary points. However, a set including free-free functions and clamped-clamped functions can qualify as quasi-comparison functions if the set provides for nonzero displacement, slope, bending moment and shearing force of arbitrary magnitude at all boundary points. This implies that there must be a minimum number of functions in the set. As an example, we consider substructure o as a typical beam in bending and confine ourselves to transverse displacements only, so that there are four arbitrary quantities at each end, for a total of eight. Hence, in addition to one rigid-body translation and one rigid-body rotation, it is necessary to include six shape functions in the set, perhaps three free-free functions and three clamped-clamped functions. This choice of shape functions is far from being unique. Another choice that can prove quite suitable, and one likely to cause initial skepticism, is a set of clamped-free and free-clamped functions. It should be reiterated that the preceding functions are not modes at all, as there is no conceivable substructure that can be free-free and clamped-clamped, or clamped-free and free-clamped at the same time. Going one step further, in the numerical example to follow, we demonstrate that sine and cosine functions can constitute a suitable set from which to construct quasi-comparison functions.

The fact that a given choice of quasi-comparison functions for the substructures is capable of providing both for the satisfaction of geometric compatibility and for the matching of bending moment and shearing force at boundary points does not mean that the natural boundary conditions will actually be satisfied exactly. Indeed, in rendering the Rayleigh quotient in terms of the energy inner product stationary, the natural boundary conditions will be satisfied only approximately, and so will the differential equations. The substructure synthesis process tends to reduce the error at all points of the structure, regardless of whether they are boundary points or points in the interior of the substructures. In using quasi-comparison functions, instead of mere admissible functions, the process is given the chance to reduce errors at all points of the structure, resulting in superior convergence characteristics.

As mentioned on several occasions, the substructure synthesis is in fact a Rayleigh-Ritz method. The main difference between the substructure synthesis presented here and the classical Rayleigh-Ritz method is that here the admissible functions are local in the sense that they are defined over the domain of a given substructure, whereas in the classical Rayleigh-Ritz method they are global, in the sense that they are defined over the entire structure. Of course, the fact that this substructure synthesis is a Rayleigh-Ritz method has many implications. The most important of these is that most of the theory developed in conjunction with the Rayleigh-Ritz method is valid for the present substructure synthesis as well. Having the backing of the Rayleigh-Ritz theory permits us to draw some immediate conclusions concerning the convergence of substructure synthesis. To this end, we assume that the number of degrees of freedom of the system is n. Then, as the number of admissible functions entering into the comparison functions increases, we can state that

$$\lim_{n\to\infty} \lambda_r^n = \lambda_r \tag{8.226}$$

or the computed eigenvalues converge to the actual eigenvalues. Moreover, they approach the actual eigenvalues from above. Although the Rayleigh-Ritz method does not permit a similar statement concerning the computed eigenvectors, and hence concerning the approximate eigenfunctions, it is safe to say that, as the number of admissible functions entering into the quasi-comparison functions increases, the error in satisfying the differential equation and the boundary conditions tends to decrease. The rate of convergence depends largely on the choice of quasi-comparison functions, and tends to be faster for eigenvalues than for eigenvectors. This can be attributed to the stationarity of Rayleigh's quotient, which implies that, if an approximate eigenvector differs from the corresponding actual eigenvector by a small quantity of first order, the approximate eigenvalue differs from the corresponding actual eigenvalue by a small quantity of second order.

From the preceding discussion, we conclude that the substructure synthesis presented here is different philosophically from the component-mode synthesis. In fact, it is closer in nature to the hierarchical finite element method (Ref. 29). Indeed, both this substructure synthesis and the hierarchical finite element method describe the motion in terms of local admissible functions. In the first, the local functions are admissible functions capable of yielding quasi-comparison functions defined over entire substructures, and in the second, the local functions are polynomials defined over finite elements. Moreover, in substructure synthesis convergence is achieved by increasing the number of admissible functions entering into the quasi-comparison function for a given substructure, and in the hierarchical finite element method convergence is achieved by increasing the number and degree of polynomials for a given finite element. The latter is in contrast with the ordinary finite element method, in which convergence is achieved by keeping the number of polynomials constant and refining the finite element mesh.

The substructure synthesis method described in this section is suitable for structures for which the substructures represent one-dimensional elastic members.

Example 8.4

The theory just developed is applied to the structure shown in Fig. 8.17 (Ref. 31). The structure consists of three substructures, the central substructure o and two substructures of type a, with the supports being mounted on springs. The central substructure is a uniform beam, and the other two substructures are tapered beams, as shown in Fig. 8.17.

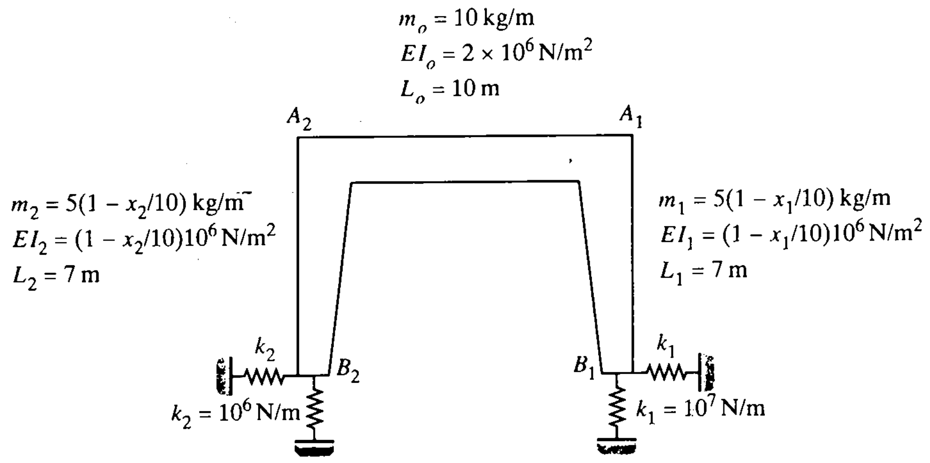

$m_o = 10$ kg/m
$EI_o = 2 \times 10^6$ N/m^2
$L_o = 10$ m
A_2
A_1
$m_2 = 5(1 - x_2/10)$ kg/m
$EI_2 = (1 - x_2/10)10^6$ N/m^2
$L_2 = 7$ m
$m_1 = 5(1 - x_1/10)$ kg/m
$EI_1 = (1 - x_1/10)10^6$ N/m^2
$L_1 = 7$ m
k_2
B_2
B_1
k_1
$k_2 = 10^6$ N/m
$k_1 = 10^7$ N/m

Figure 8.17 Structure consisting of three substructures

We propose to describe the motion of the substructures in five different ways. In case 1, we represent the motion of all three substructures by means of mere admissible functions. To this end, we use free-free functions for the central substructure and clamped-free functions for the remaining two substructures, all functions corresponding to modes for the associated uniform beams. As pointed out earlier, the boundary conditions for the central substructure cannot be satisfied with a finite number of free-free functions. Similarly, the boundary conditions at points B_1 and B_2 in the case of substructures of type a cannot be satisfied with a finite number of clamped-free functions. Note that the geometric boundary conditions at points A_1 and A_2 are satisfied automatically by the kinematical procedure implied by Eqs. (8.212)–(8.214). In case 2, the motion of the central substructure is represented by means of an improved set of admissible functions and the motion of the remaining two substructures is represented by means of quasi-comparison functions. In particular, for the central substructure, we use a combination of free-free and pinned-pinned functions. Although this combination of functions represents an improvement over case 1, it still falls in the class of mere admissible functions, as the bending moment at the boundaries remains zero with a finite number of terms. For the other two substructures, we use combinations of clamped-free and clamped-pinned functions, so that these combinations of functions qualify as quasi-comparison functions. In cases 3–5, the motion of all three substructures is represented by quasi-comparison functions. In case 3, we use combinations of free-free and clamped-clamped functions for the central substructure, so that both boundary conditions at each end can be satisfied with a finite number of functions; the quasi-comparison functions for the other two substructures remain as in case 2. Cases 4 and 5 differ from case 3 in that in case 4 the motion of the central substructure is represented by clamped-free and free-clamped functions, and in case 5 by pinned-pinned and cosine shape functions. The motion of the other two substructures is represented in cases 3, 4 and 5 as in case 2. Figures 8.18a–e show the various functions for cases 1–5.

The various types of functions used have closed-form expressions. The free-free functions have the expression

$$\phi_i(x) = \cosh\beta_i x + \cos\beta_i x - \sigma_i(\sinh\beta_i x + \sin\beta_i x) \tag{a}$$

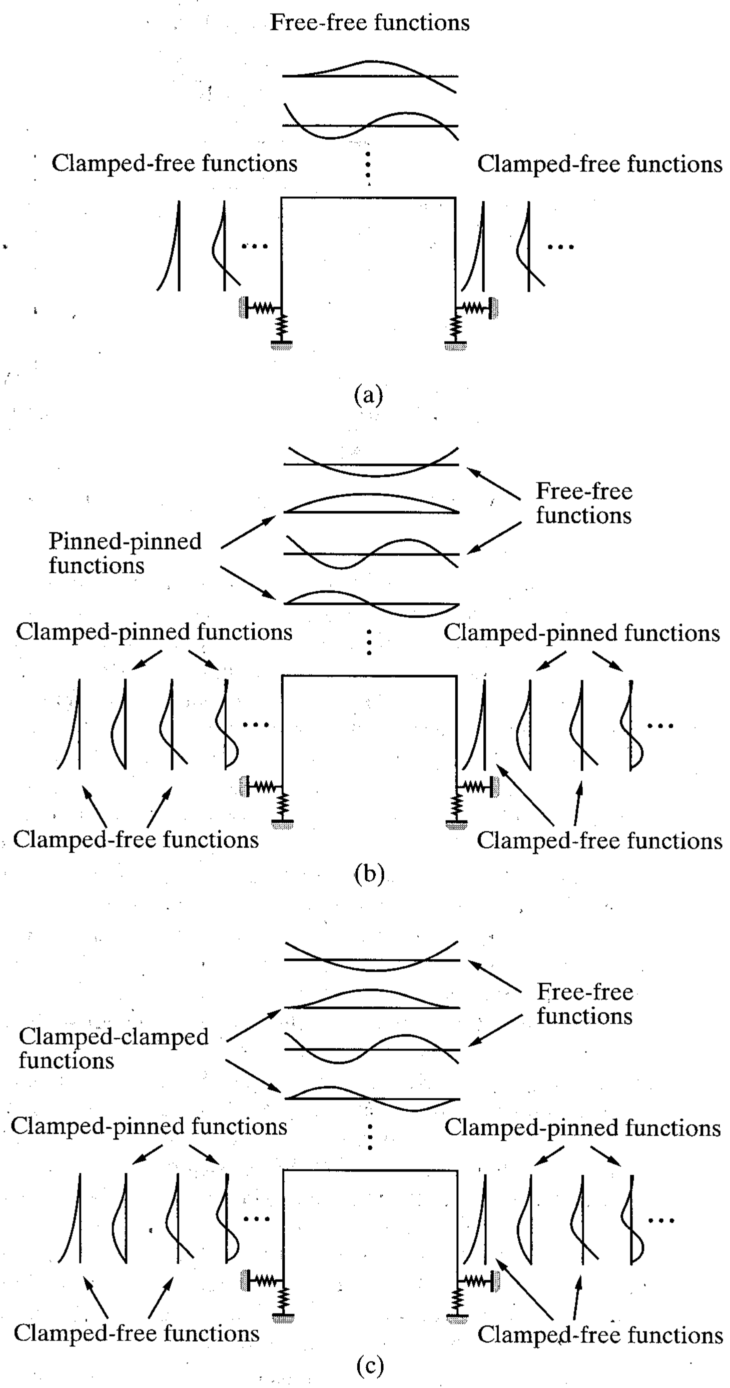


Figure 8.18 **(a)** Admissible functions for all three substructures **(b)** Improved admissible functions for beam and quasi-comparison functions for columns **(c)** Quasi-comparison functions for all three substructures

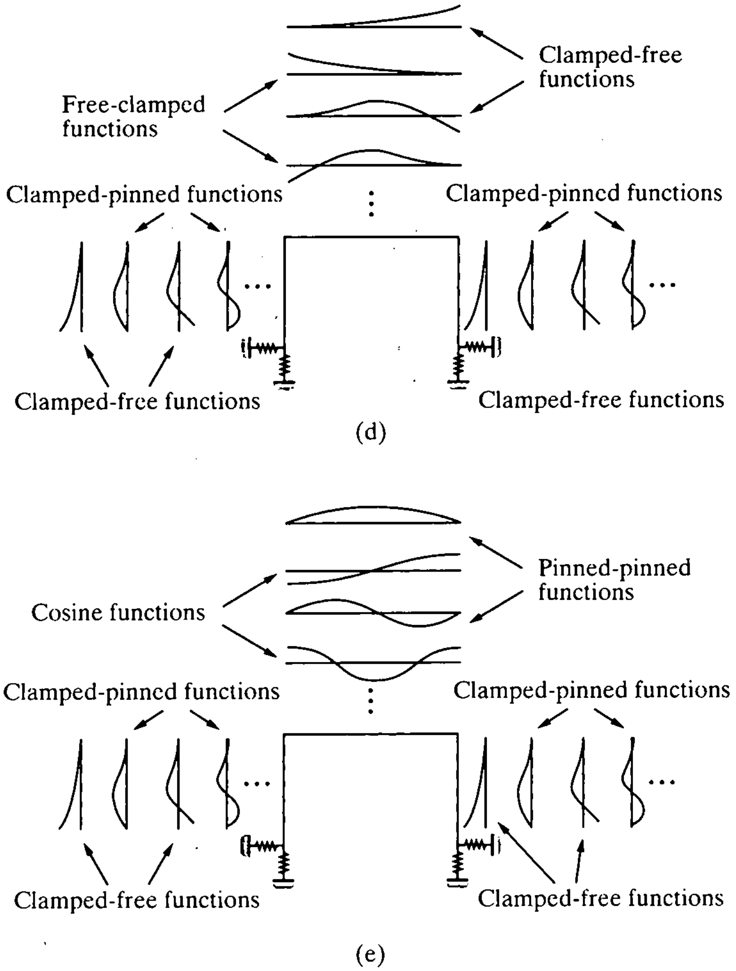


Figure 8.18 (Continued) **(d)** Quasi-comparison functions for all three substructures **(e)** Quasi-comparison functions for all three substructures

The clamped-free, clamped-clamped and clamped-pinned functions are given by

$$\phi_i = \cosh \beta_i x - \cos \beta_i x - \sigma_i (\sinh \beta_i x - \sinh \beta i x) \tag{b}$$

where the constants β_i and σ_i vary from case to case; their numerical values are given by Blevins.[2] On the other hand, the pinned-pinned functions have the form

$$\phi_i(x) = \sin \frac{i\pi x}{L} \tag{c}$$

and the cosine functions are

$$\phi_i(x) = \cos \frac{i\pi x}{L} \tag{d}$$

[2] Blevins, R.D., *Formulas for Natural Frequency and Mode Shape*, Van Nostrand Reinhold, New York, 1979.

Finally, the stiffness matrices for beams in bending have the entries

$$k_{oij} = \int_0^{l_o} EI_0(x)\phi_i''(x)\phi_j''(x)\,dx, \qquad k_{aij} = \int_0^{l_a} EI_a(x)\phi_i''(x)\phi_j''(x)\,dx, \qquad a = 1, 2 \tag{e}$$

where primes denote differentiations with respect to x. The numerical values for the various parameters are shown in Fig. 8.17.

The eigenvalue problem for the system shown in Fig. 8.17 was solved for the five cases just described and the three lowest natural frequencies are listed in Tables 8.4–8.6. The results are in agreement with the expectations. In case 1, in which the motion is expressed in terms of one type of admissible functions only for each substructure, the convergence is unsatisfactory. Using a 27-degree-of-freedom model, the computed natural frequencies are relatively far from the actual ones, and improvement with the addition of degrees of freedom is very slow. At this point, there is no indication how many degrees of freedom will be necessary for convergence. In case 2, in which the motion of substructure o is described by means of more suitable admissible functions than in case 1 and the motion of substructures a is represented by means of quasi-comparison functions, the results are significantly better than in case 1, although convergence is still elusive. In cases 3–5, in which the motion of all substructures is represented by means of quasi-comparison functions, convergence is relatively rapid, with the results of case 4 being better than those of cases 3 and 5. Clearly, the results are far superior to those obtained in the first two cases. As pointed out earlier, there is a minimum number of terms necessary before the linear combinations of admissible functions become quasi-comparison functions. Hence, in cases 3–5, the results for small numbers of degrees of freedom are not meaningful.

TABLE 8.4 First Natural Frequency

DoF	Case 1	Case 2	Case 3	Case 4	Case 5
6	1.75746	1.75746	1.75746	1.69875	1.75745
9	1.57714	1.65474	1.65474	1.47335	1.47530
12	1.56932	1.56538	1.56538	1.47102	1.47084
15	1.53536	1.49632	1.47391	1.47073	1.47077
18	1.53462	1.49632	1.47391	1.47073	1.47077
21	1.52395	1.49632	1.47391	1.47073	1.47073
24	1.52339	1.48726	1.47157	1.47073	1.47073
27	1.51702	1.48138	1.47073	1.47073	1.47073

To verify how well the natural boundary condition at points A is satisfied, we define the error in the bending moment as

$$\epsilon_M^n = M(A-) - M(A+) \tag{f}$$

where $M(A-)$ denotes the bending moment at point A_2 computed from the solution for substructure o and $M(A+)$ is the same quantity corresponding to substructure $a = 2$. Figure 8.19 shows plots of ϵ_M^n versus n for all five cases just discussed. It is clear that the solutions in terms of quasi-comparison functions are far superior to those in terms of mere admissible functions.

TABLE 8.5 Second Natural Frequency

DoF	Case 1	Case 2	Case 3	Case 4	Case 5
6	8.59485	8.59485	8.59485	11.62478	8.71827
9	8.29244	8.25873	8.23772	9.76602	8.33865
12	8.26215	8.24909	8.22873	8.22884	8.32848
15	8.25375	8.24867	8.22834	8.21273	8.21143
18	8.24792	8.23106	8.20832	8.21197	8.20475
21	8.24659	8.22109	8.20449	8.20455	8.20475
24	8.24418	8.22108	8.20447	8.20443	8.20475
27	8.24369	8.22107	8.20446	8.20443	8.20444

TABLE 8.6 Third Natural Frequency

DoF	Case 1	Case 2	Case 3	Case 4	Case 5
6	26.86095	26.86095	26.86095	28.98808	26.92279
9	21.78531	21.91468	21.35960	20.37671	20.24966
12	21.30582	21.11908	20.98775	20.18728	19.93381
15	20.66517	20.11777	20.03998	19.94335	19.93298
18	20.61882	20.11582	20.03894	19.93163	19.93137
21	20.46641	20.11493	20.03876	19.93120	19.93120
24	20.45673	20.05149	19.95669	19.93111	19.93111
27	20.37866	20.00927	19.93139	19.93102	19.93102

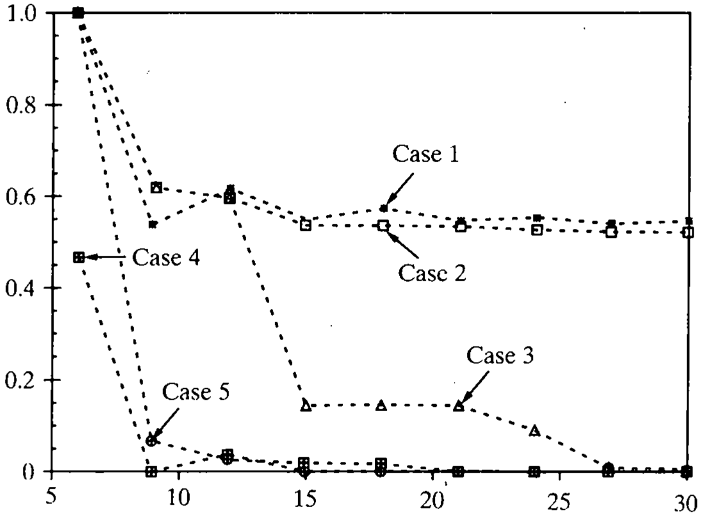


Figure 8.19 Bending moment error at point A_2 versus the number of degrees of freedom of the model

8.13 SYNOPSIS

The spatial discretization methods can be divided broadly into lumping procedures and series discretization methods. Lumping techniques appeal to physical intuition and tend to be easy to understand. Indeed, the system parameters, i.e., the mass distribution, or the stiffness distribution, or both, are lumped at given points of the system. For the most part this is a heuristic process, with no analytical guidelines to call upon. The best known methods are the lumped-parameter method using influence coefficients, Holzer's method for torsional vibration of shafts (and hence for the transverse vibration of strings and axial vibration of rods) and Myklestad's method for bending vibration of beams. In the first, the mass is lumped at discrete points, thus yielding a diagonal mass matrix. On the other hand, the stiffness properties are accounted for in an "exact" manner through flexibility influence coefficients. This being a discretized version of the integral form of the eigenvalue problem, superior results can be expected for a sufficiently large number of discrete points. Moreover, at least in theory, the method is applicable to all types of vibrating systems. In practice, the determination of the influence coefficients can cause serious difficulties, particularly for systems with nonuniform stiffness distribution and/or complex boundary conditions, or for two-dimensional problems. In Holzer's method, both the mass and stiffness are lumped. The mass is lumped into rigid disks at discrete points and the shaft segments between these points are assumed to be massless and to have uniform torsional stiffness. The computation of the natural frequencies and modes of vibration can be carried out in a systematic manner by means of transfer matrices. Myklestad's method extends the ideas to the bending vibration of beams.

Lumped-parameter methods lack mathematical rigor and their convergence is not easy to judge. By contrast, series discretization methods do not suffer from these drawbacks. They tend to be more abstract, however. In the case of conservative systems, the discretization process is embedded into a variational approach with its origin in Rayleigh's principle, which states that the frequency of vibration has a minimum in the neighborhood of the fundamental mode. The fundamental mode is actually not known, and any guess for the fundamental mode yields a frequency of vibration larger than the lowest natural frequency. It is therefore natural to attempt to improve the guess of the fundamental mode, thus lowering the estimate of the lowest natural frequency. This improved guess is in the form of a series of admissible functions with undetermined coefficients and the coefficients are determined so as to render Rayleigh's quotient stationary. The procedure is commonly known as the Rayleigh-Ritz method. The main drawback of the method lies in the difficulty of coming up with suitable admissible functions, making the approach more of an art than a method. The weighted residuals method is really a family of series discretization procedures based on the idea of reducing the approximation error. It is not a variational approach, so that it is applicable to both self-adjoint and non-self-adjoint systems. By far the best known of the weighted residuals methods is Galerkin's method, which is equivalent to the Rayleigh-Ritz method for self-adjoint systems. The convergence of approximations for both self-adjoint and non-self-adjoint systems can be improved through the use of quasi-comparison functions.

As originally envisioned by Hurty, the component-mode synthesis is an extension of the Rayleigh-Ritz method to complex structures with identifiable substructures, referred to as "components." The term "mode" can be traced to the wide assumption that the admissible functions used to represent the motion of the individual components must be some loosely defined modes of vibration. Substructure synthesis represents an extension of the enhanced Rayleigh-Ritz method of Sec. 8.6 to structures consisting of chains of substructures. A consistent kinematical procedure obviates the problem of using constraints to force substructures to work together as a single structure. The method is able to accommodate substructures rotating relative to one another. Here too, the use of quasi-comparison functions can improve convergence.

Whereas the Rayleigh-Ritz and the Galerkin methods have many advantages over other approximate techniques, they also have serious shortcomings. In particular, the applicability of the methods is confined to relatively simple systems, such as one-dimensional ones and two-dimensional ones with rectangular and circular boundaries, although the formulation can be modified so as to accommodate systems with trapezoidal boundaries. The component-mode synthesis and substructure synthesis can extend the usefulness of the series discretization approach, but the above limitations still apply to the individual substructures. Moreover, the question of generating suitable admissible functions is not entirely settled. In addition, the computation of the mass and stiffness matrices generally requires extensive numerical integrations. Many of these shortcomings can be attributed to the fact that the admissible functions are global functions, in the sense that they extend over the entire domain of the elastic member, or of the substructure. Another series discretization method, the finite element method, does not have these drawbacks by virtue of the fact that it uses local admissible functions, defined over small subdomains of the structure. Because these subdomains are small, the local admissible functions can be chosen in the form of low-degree polynomials, and the finite element mesh can be constructed so as to accommodate boundaries with very complex geometry. Finally, the process of computing the mass and stiffness matrices can be automated, relieving the analyst from many computer coding chores. All these attributes make the finite element method a very versatile one.

It should be pointed out here that the Rayleigh-Ritz theory does not actually require that the admissible functions be global, although this has generally been the practice, and local functions are indeed admissible, provided they satisfy the differentiability requirements. Hence, although not conceived originally as such, the finite element method does represent another version of the Rayleigh-Ritz method differing from the classical one presented in this chapter in the nature of the admissible functions. The identification of the finite element method as a Rayleigh-Ritz method was very fortunate indeed, as the mathematical foundation of the Rayleigh-Ritz method could be extended instantly to the finite element method, a foundation lacking originally in the heuristically developed finite element method. Due to its extreme versatility, the finite element method has become the method of choice in many areas of engineering analysis, reaching far beyond the original structural applications. In recognition of its dominant role in vibrations, the entire Chapter 9 is devoted to the finite element method.

PROBLEMS

8.1 Formulate the eigenvalue problem for a uniform string fixed at both ends by means of the lumped-parameter method using flexibility influence coefficients. Solve the eigenvalue problem for $n = 20$, compare the results with the exact solution obtained in Sec. 7.6 and draw conclusions concerning the accuracy of the approximate solution.

8.2 A shaft in torsional vibration fixed at both ends has the mass polar moment of inertia and torsional stiffness distributions

$$I(x) = I\left[\frac{3}{4} + \frac{x}{L} - \left(\frac{x}{L}\right)^2\right], \qquad GJ(x) = GJ\left[\frac{3}{4} + \frac{x}{L} - \left(\frac{x}{L}\right)^2\right]$$

Formulate and solve the eigenvalue problem by means of the lumped-parameter method using flexibility influence coefficients for the case in which $n = 20$.

8.3 A shaft in torsional vibration fixed at $x = 0$ and with a torsional spring of stiffness $k = GJ/L$ at $x = L$ has the mass polar moment of inertia and torsional stiffness distributions

$$I(x) = \left[1 - \frac{1}{2}\left(\frac{x}{L}\right)^2\right], \qquad GJ(x) = GJ\left[1 - \frac{1}{2}\left(\frac{x}{L}\right)^2\right]$$

Formulate and solve the eigenvalue problem by means of the lumped-parameter method using flexibility influence coefficients for the case in which $n = 20$.

8.4 A cantilever beam clamped at $x = 0$ has rectangular cross section of unit width and height varying according to $h(x) = h(1-2x/3L)$. Formulate and solve the eigenvalue problem for the bending vibration of the beam by means of the lumped-parameter method using flexibility influence coefficients for the case in which $n = 20$.

8.5 A simply supported beam has the mass and stiffness distributions

$$m(x) = m\left[1 + 12\frac{x}{L} - 12\left(\frac{x}{L}\right)^2\right], \qquad EI(x) = EI\left[1 + 12\frac{x}{L} - 12\left(\frac{x}{L}\right)^2\right]$$

Formulate and solve the eigenvalue problem by means of the lumped-parameter method using flexibility influence coefficients for the case in which $n = 20$.

8.6 Solve Problem 8.2 by means of Holzer's method, compare results and draw conclusions concerning the relative accuracy; provide arguments in support of your conclusions.

8.7 Solve Problem 8.3 by means of Holzer's method, compare results, and draw conclusions concerning the relative accuracy; provide arguments in support of your conclusions.

8.8 Solve Problem 8.4 by means of Myklestad's method, compare results and draw conclusions concerning the relative accuracy; provide arguments in support of your conclusions.

8.9 Solve Problem 8.5 by means of Myklestad's method, compare results and draw conclusions concerning the relative accuracy; provide arguments in support of your conclusions.

8.10 Estimate the lowest natural frequency of the string of Problem 8.1 by means of Rayleigh's energy method using the static displacement curve as a trial function. Compare the estimate with the exact solution obtained in Sec. 7.6 and draw conclusions as to the accuracy of the estimate.

8.11 Estimate the lowest natural frequency of the shaft of Problem 8.2 by means of Rayleigh's energy method using as a trial function the static displacement curve due to a distributed torque proportional to the mass polar moment of inertia. Compare the estimate with results obtained in Problems 8.2 and 8.6 and draw conclusions concerning the relative accuracy of the estimate.

8.12 Solve Problem 8.11 with Problems 8.3 and 8.7 replacing Problems 8.2 and 8.6, respectively.

8.13 Solve Problem 8.11 with Problems 8.4 and 8.8 replacing Problems 8.2 and 8.6, respectively.

8.14 Solve Problem 8.11 with Problems 8.5 and 8.9 replacing Problems 8.2 and 8.6, respectively.

8.15 Estimate the lowest natural frequency of the beam of Problem 7.32 by means of Rayleigh's energy method using the static displacement curve as a trial function. Compare the estimate with the exact solution obtained in Problem 7.32 and draw conclusions concerning the accuracy of the estimate.

8.16 Solve Problem 8.2 by the Rayleigh-Ritz method using the eigenfunctions of a uniform shaft fixed at both ends as comparison functions for $n = 1, 2, \dots, 6$. Construct an array as in Eq. (8.103) and draw conclusions.

8.17 Solve Problem 8.3 by the Rayleigh-Ritz method using the eigenfunctions of a uniform fixed-free shaft as admissible functions for $n = 1, 2, \dots, 6$. Construct an array as in Eq. (8.103) and draw conclusions.

8.18 Solve Problem 8.5 by the Rayleigh-Ritz method using the eigenfunctions of a uniform simply supported beam as admissible functions for $n = 1, 2, \dots, 6$. Construct an array as in Eq. (8.103) and draw conclusions.

8.19 Solve Problem 7.32 by the Rayleigh-Ritz method using the eigenfunctions of a clamped-pinned beam (without the spring) as admissible functions for $n = 1, 2, \dots, 6$. Construct an array as in Eq. (8.103) and draw conclusions.

8.20 Solve Problem 7.40 by the Rayleigh-Ritz method using the eigenfunctions of a uniform membrane free at $x = 0, a$ and fixed at $y = 0, b$ as admissible functions for $m = 1, 2$; $n = 1, 2, 3$. Compare the results with those obtained in Problem 7.40 and draw conclusions.

8.21 Solve Problem 7.41 by means of the Rayleigh-Ritz method using admissible functions in the form of the products $f_i(r)g_j(\theta)$, where $f_i(r) = (r/a)^i$ and $g_j(\theta)$ are trigonometric functions, for $i = 0, 1, \dots 8$; $j = 0, 1, 2, 3$. Compare the natural frequencies with those obtained in Problem 7.41 and draw conclusions.

8.22 Solve Problem 7.47 by the Rayleigh-Ritz method using products of beam eigenfunctions as admissible functions.

8.23 A rectangular plate simply supported at the boundaries $x = 0, a$ and $y = 0, b$ has the mass density and flexural rigidity

$$m(x, y) = m\left[1 + 0.25\frac{x}{a}\left(1 - \frac{x}{a}\right)\frac{y}{b}\left(1 - \frac{y}{b}\right)\right]$$

$$D_E(x, y) = D_E\left[1 + 0.75\frac{x}{a}\left(1 - \frac{x}{a}\right)\frac{y}{b}\left(1 - \frac{y}{b}\right)\right].$$

Solve the eigenvalue problem by the Rayleigh-Ritz method and give the values of the four lowest natural frequencies and the expressions of the four associated natural modes.

8.24 Solve Problem 8.17 by the enhanced Rayleigh-Ritz method using quasi-comparison functions of your own choice. Compare convergence of the three lowest natural frequencies obtained here with the convergence of the three lowest natural frequencies obtained in Problem 8.17 and draw conclusions.

8.25 Solve Problem 8.19 by the enhanced Rayleigh-Ritz method using quasi-comparison functions from two families, clamped-pinned and clamped-clamped functions. Compare convergence of the three lowest natural frequencies obtained here with the convergence of the three lowest natural frequencies obtained in Problem 8.19 and draw conclusions.

8.26 Solve Problem 8.25 by the enhanced Rayleigh-Ritz method using quasi-comparison functions from two families, clamped-pinned functions and a second family of functions of your own choice.

8.27 Solve Problem 7.40 by the enhanced Rayleigh-Ritz method using quasi-comparison functions of your own choice. Compare the results with the results obtained in Problem 8.20 and draw conclusions.

8.28 The boundary-value problem for a given distributed-parameter system is defined by the differential equation

$$Lw + C\dot{w} + m\ddot{w} = 0$$

and suitable boundary conditions, where L is a stiffness operator and C a damping operator. Use Galerkin's method to derive an algebraic eigenvalue problem approximating the associated differential eigenvalue problem.

8.29 Assume that the system of Problem 8.3 is subjected to damping of the Kelvin-Voigt type with $c = 0.1$ (Sec. 7.18). Then, use the formulation of Problem 8.28, let the system parameters be uniformly distributed, solve the algebraic eigenvalue problem for $n = 3, 4, 5, 6$ and discuss the behavior of the eigenvalues.

8.30 Solve Problem 8.29 with Problem 7.32 replacing Problem 8.3.

8.31 Solve Problem 8.16 by the collocation method. Determine the required number n of terms in the approximation to match the accuracy of the lowest eigenvalue obtained in Problem 8.16 for $n = 6$.

8.32 Solve Problem 8.31 with Problem 8.17 replacing Problem 8.16.

8.33 Solve Problem 8.31 with Problem 8.19 replacing Problem 8.16.

8.34 Derive the response of the system of Problem 8.16 with $n = 6$ to the concentrated torque $M_0 u(t)$ applied at $x = L/2$, where $u(t)$ is the unit step function. Discuss the mode participation in the response.

8.35 Derive the response of the system of Problem 8.19 with $n = 6$ to the distributed force $f(x, t) = f_0(1 - x/L)\delta(t)$, where $\delta(t)$ is the unit impulse.

8.36 Derive the response of the system of Problem 8.29 with $n = 6$ to the distributed torque $m(x, t) = m_0 r(t)$, where $r(t)$ is the unit ramp function (Sec. 1.7).

8.37 Derive the response of the system of Problem 8.30 with $n = 6$ to the concentrated force $F_0[u(t) - u(t - T)]$ applied at $x = L/2$, where $u(t)$ is the unit step function.

8.38 Derive the response of the system of Problem 8.31 with $n = 6$ to the distributed torque $m(x, t) = m_0[r(t) - r(t - T)]$, where $r(t)$ is the unit ramp function (Sec. 1.7). Discuss the mode participation in the response.

8.39 Derive the response of the system of Problem 8.33 with $n = 6$ to the distributed torque $m(x, t) = m_0(1 - x/2L)[r(t) - r(t - T) - Tu(t - 2T)]$, where $r(t)$ is the unit ramp function (Sec. 1.7) and $u(t)$ the unit step function.

8.40 Derive the response of the plate of Problem 8.23 to the force $f(x, y, t) = f_0[r(t) - r(t - T)]$ distributed uniformly over the rectangular area defined by $a/2 < x < 3a/4$, $b/4 < y < 3b/4$, where $r(t)$ is the unit ramp function (Sec. 1.7).

BIBLIOGRAPHY

1. Bamford, R. M., "A Modal Combination Program for Dynamic Analysis of Structures," *TM 33-290*, Jet Propulsion Laboratory, Pasadena, CA, July 1967.
2. Benfield, W. A. and Hruda, R. F., "Vibration Analysis of Structures by Component Mode Substitution," *AIAA Journal*, Vol. 9, No. 7, 1971, pp. 1255–1261.
3. Bisplinghoff, R. L., Ashley, H. and Halfman, R. L., *Aeroelasticity*, Addison-Wesley, Reading, MA, 1957.
4. Collatz, L., *The Numerical Treatment of Differential Equations*, Springer-Verlag, New York, 1966.
5. Courant, R. and Hilbert, D., *Methods of Mathematical Physics*, Vol. 1, Wiley, York, 1989.
6. Craig, R. R. Jr. and Bampton, M.C.C., "Coupling of Substructures for Dynamic Analysis," *AIAA Journal*, Vol. 6, No. 7, 1968, pp. 1313–1319.
7. Dowell, E. H., "Free Vibration of an Arbitrary Structure in Terms of Component Modes," *Journal of Applied Mechanics*, Vol. 39, No. 3, 1972, pp. 727–732.
8. Dowell, E. H. et al., *A Modern Course in Aeroelasticity*, 2nd ed., Kluwer, Dordrecht, The Netherlands, 1989.
9. Finlayson, B. A., *The Method of Weighted Residuals and Variational Principles*, Academic Press, New York, 1972.
10. Fung, Y. C., *Theory of Aeroelasticity*, Dover, New York, 1969.
11. Gladwell, G.M.L., "Branch Mode Analysis of Vibrating Systems," *Journal of Sound and Vibration*, Vol. 1, 1964, pp. 41–59.
12. Goldman, R. L., "Vibration Analysis by Dynamic Partitioning," *AIAA Journal*, Vol. 7, No. 6, 1969, pp. 1152–1154.
13. Gould, S. H., *Variational Methods for Eigenvalue Problems: An Introduction to the Methods of Rayleigh, Ritz, Weinstein and Aronszajn*, Dover, New York, 1995.
14. Hagedorn, P., "The Rayleigh-Ritz Method With Quasi-Comparison Functions in Nonself-Adjoint Problems," *Journal of Vibration and Acoustics*, Vol. 115, July 1993, pp. 280–284.
15. Hintz, R. M., "Analytical Methods in Component Modal Synthesis," *AIAA Journal*, Vol. 13, No. 8, 1975, pp. 1007–1016.
16. Holzer, H., *Die Berechnung der Drehschwingungen*, Springer-Verlag, Berlin, Germany, 1921.
17. Hou, S. N., "Review of a Modal Synthesis Technique and a New Approach," *Shock and Vibration Bulletin*, No. 40, Part 4, 1969, pp. 25–30.
18. Hurty, W. C., "Vibrations of Structural Systems by Component-Mode Synthesis," *Journal of Engineering Mechanics Division, ASCE*, Vol. 86, August 1960, pp. 51–69.
19. Hurty, W. C., "Dynamic Analysis of Structural Systems Using Component Modes," *AIAA Journal*, Vol. 3, No. 4, 1965, pp. 678–685.
20. Hurty, W. C., "A Criterion for Selecting Natural Modes of a Structure," *TM33–364*, Jet Propulsion Laboratory, Pasadena, CA, January 1967.
21. Klein, L. R. and Dowell, E. H., "Analysis of Modal Damping by Component Modes Using Lagrange Multipliers," *Journal of Applied Mechanics*, Vol. 41, 1974, pp. 527–528.
22. Kuhar, E. J. and Stahle, L. V., "Dynamic Transformation Method for Modal Synthesis," *AIAA Journal*, Vol. 12, No. 5, 1974, pp. 672–678.
23. Leissa, A. W., *Vibration of Plates*, NASA SP-160, National Aeronautics and Space Administration, Washington, DC, 1969.
24. MacNeal, R. H., "A Hybrid Method of Component Mode Synthesis," *Computers and Structures*, Vol. 1, No. 4, 1971, pp. 581–601.

25. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan, New York, 1967.

26. Meirovitch, L. and Hale, A. L., "Synthesis and Dynamic Characteristics of Large Structures with Rotating Substructures," *Proceedings of the IUTAM Symposium on the Dynamics of Multibody Systems* (K. Magnus, editor), Springer-Verlag, Berlin, West Germany, 1978, pp. 231–244.

27. Meirovitch, L. and Hale, A. L., "On the Substructure Synthesis Method," *AIAA Journal*, Vol. 19, No. 7, 1981, pp. 940–947.

28. Meirovitch L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

29. Meirovitch, L., *Elements of Vibration Analysis*, 2nd ed., McGraw-Hill, New York, 1986.

30. Meirovitch, L. and Kwak, M. K., "On the Convergence of the Classical Rayleigh-Ritz Method and the Finite Element Method," *AIAA Journal*, Vol. 28, No. 8, 1990, pp. 1509–1516.

31. Meirovitch, L. and Kwak, M. K., "Rayleigh-Ritz Based Substructure Synthesis for Flexible Multibody Systems," *AIAA Journal*, Vol. 29, No. 10, 1991, pp. 1709–1719.

32. Meirovitch, L. and Hagedorn, P., "A New Approach to the Modelling of Distributed Non-Self-Adjoint Systems," *Journal of Sound and Vibration*, Vol. 178, No. 2, 1994, pp. 227–241.

33. Meirovitch, L. and Seitz, T. J., "Structural Modeling for Optimization of Low Aspect Ratio Wings," *Journal of Aircraft*, Vol. 32, No. 5, 1995.

34. Mikhlin, S. G., *Variational Methods in Mathematical Physics*, Pergamon Press, New York, 1964.

35. Mikhlin, S. G. and Smolitskiy, K. L., *Approximate Methods for Solution of Differential and Integral Equations*, American Elsevier, New York, 1967.

36. Myklestad, N. O., "A New Method for Calculating Normal Modes of Uncoupled Bending Vibration of Airplane Wings and Other Types of Beams," *Journal of Aeronautical Sciences*, Vol. 11, 1944, pp. 153–162.

37. Pestel, E. C. and Leckie, F. A., *Matrix Methods in Elastomechanics*, McGraw-Hill, New York, 1963.

38. Ralston, A., *A First Course in Numerical Analysis*, 2nd ed., McGraw-Hill, New York, 1990.

39. Rayleigh, Lord, *The Theory of Sound*, Dover, New York, 1945.

40. Ritz, W., "Über eine neue Methode zur Lösung gewisser Variationsprobleme die mathematischen Physik," *Journal für die reine und angewandte Mathematik*, Vol. 135, 1909, pp. 1–61.

41. Ritz, W., "Theorie der Transversalschwingungen einer quadratischen Platte mit freie Rändern," *Annalen der Physik*, Vol. 38, 1909.

42. Rubin, S., "Improved Component-Mode Representation for Structural Dynamic Analysis," *AIAA Journal*, Vol. 13, No. 8, 1975, pp. 995–1006.

43. Spanos, J. T. and Tsuha, W. S., "Selection of Component Modes for Flexible Multibody Simulation," *Journal of Guidance, Control and Dynamics*, Vol. 14, No. 2, 1991, pp. 278–286.

44. Strang, G. and Fix, G. J., *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

45. Thomson, W. T., "Matrix Solution for the Vibration of Nonuniform Beams," *Journal of Applied Mechanics*, Vol. 17, 1950, pp. 337–339.

46. Turner, M. J., Clough, R. W., Martin, H. C. and Topp, L. J., "Stiffness and Deflection Analysis of Complex Structures," *Journal of Aeronautical Sciences*, Vol. 23, 1956, pp. 805–823.

9

THE FINITE ELEMENT METHOD

The finite element method must be regarded as the most successful technique of structural analysis. Originally conceived by Turner, Clough, Martin and Topp in the mid 1950s (Ref. 26) as a procedure for static stress analysis of complex structures, the method has been expanding rapidly into many engineering areas. The phenomenal success of the finite element method can be attributed to a large extent to timing, as at the same time the finite element method was being developed so were increasingly powerful digital computers. In fact, in many ways the computer has helped greatly with the development of the method.

Although developed independently of the Rayleigh-Ritz method, the finite element method was demonstrated later to be a Rayleigh-Ritz method. To distinguish between the two, we refer to the original one as the *classical Rayleigh-Ritz method.* We recall from Sec. 8.5 that the classical Rayleigh-Ritz method represents a variational approach whereby a distributed system is approximated by a discrete one by assuming a solution of the differential eigenvalue problem as a finite series of admissible functions. The wide use of the classical Rayleigh-Ritz method has been limited by the inability to generate suitable admissible functions for a large number of problems. This inability can be attributed to the traditional manner in which the method has been applied rather than to inherent flaws in the method itself. Indeed, a major source of difficulties is due to the insistence on using global admissible functions, and there is nothing in the theory requiring that the functions be global. In this regard, it should be noted that systems with complex boundary conditions, or complex geometry, cannot be accommodated easily by global admissible functions. Such cases are quite common in two- and three-dimensional structures. Moreover, global admissible functions tend to have complicated expressions, difficult to handle on a routine basis. The basic difference between the classical Rayleigh-Ritz method and the finite element method lies in the fact that in the latter an approximate solution is constructed using local admissible functions defined over small subdomains of the structure. In this regard it should perhaps be mentioned that, in a paper generally regarded as the forerunner of the finite element method, Courant (Ref. 8) used a variational approach in conjunction with linear admissible functions defined over small triangular subdomains to produce an approximate solution to St. Venant's torsion problem, thus preceding the development of the finite element method by over a decade. The reason why Courant's work did not attract more attention can be attributed to poor timing. Indeed, in the early 1940s, computers capable of solving large sets of equations of equilibrium, or equations of motion, did not exist, so that the method was not practical then.

The concept of local functions defined over small subdomains carries enormous implications, and is the key to the success of the finite element method. In the first place, because the subdomains are small, good approximations can be realized with local admissible functions in the form of low-degree polynomials. These low-degree polynomials, often referred to as interpolation functions, not only make the computation of the stiffness and mass matrices appreciably easier, but also eliminate the troublesome task of choosing suitable admissible functions, as given classes of problems call for certain choices of polynomials. Perhaps more important is the fact that the computations lend themselves to automation. Indeed, the computer is not only able to solve the discretized equations of equilibrium, or equations of motion, but also to carry out such diverse tasks as the formulation of the equations by making decisions concerning the finite element mesh and the assembly of the stiffness and mass matrices. Finally, the finite element method has no equal in its ability to accommodate systems with complicated geometries and parameter distributions. Of course, the geometry can be a serious concern in two- and three-dimensional problems. To match a given irregular boundary, or to accommodate parameter nonuniformities, not only the size of the finite elements can be changed but also their shape. This extreme versatility, coupled with the fact that many powerful computer codes based on the method have become available, has made the finite element method the method of choice for static and dynamic analysis of structures.

In this chapter, we begin with the presentation of the finite element method as a Rayleigh-Ritz method. Then, the procedure for determining element stiffness and mass matrices and for assembling them into global stiffness and mass matrices is demonstrated by means of second-order systems, such as strings in transverse vibration, first using linear and then higher-degree interpolation functions. The approach is subsequently extended to fourth-order systems, such as beams in bending. The real power of the finite element method becomes evident in two-dimensional systems, such as membranes and plates. Here we encounter finite elements of various shapes, such as triangular, rectangular and quadrilateral elements, as well as elements with curved boundaries. Another version of the method, known as the hierarchical finite element method, combines some of the best features of the finite element and classical Rayleigh-Ritz methods. The chapter concludes with a discussion of the system response.

9.1 THE FINITE ELEMENT METHOD AS A RAYLEIGH-RITZ METHOD

The *finite element method* is a technique for the spatial discretization of distributed-parameter systems. It consists of dividing the domain D of the system into a set of subdomains and describing the motion over each of these subdomains by means of a linear combination of trial functions. The subdomains are called *finite elements*, the set of finite elements is known as the *mesh* and the trial functions are referred to as *interpolation functions*. To introduce the ideas, we consider first a one-dimensional domain, such as that shown in Fig. 9.1, and denote the number of elements by n and the length of the elements by h, so that $nh = L$, where L is the length of the domain. For the sake of this discussion, we assumed that the elements are equal in length, although in general their length can vary, depending on the nature of the problem. The boundary points between two elements are known as *nodes* and their displacements as *nodal displacements*.[1] For example, node j lies between the elements $j - 1$ and j at a distance jh from the left end and has the displacement a_j. Figure 9.1 shows a displacement profile $w^{(n)}(x)$ approximating the actual displacement curve $w(x)$, in which the displacement $w^{(n)}(x)$ at any point x between two typical nodes $(j - 1)h$ and jh varies linearly from a_{j-1} at $(j - 1)h$ to a_j at jh. But, observing that the displacement profile can be generated by a superposition of triangles of width $2h$ and height a_j, except for the last triangle which is of width h, we can express $w^{(n)}(x)$ as the linear combination

$$w^{(n)}(x) = \sum_{j=1}^{n} a_j \phi_j(x) \tag{9.1}$$

where $\phi_j(x)$ are the *roof functions* shown in Fig. 9.2. All roof functions have *unit amplitude* and extend over two elements, $(j - 1)h \leq x \leq (j + 1)h$, with the exception of the function $\phi_n(x)$, which extends over the single element $(n - 1)h \leq x \leq nh = L$. We note with interest that the roof functions are *nearly orthogonal*, as ϕ_j is orthogonal to all other functions, except ϕ_{j-1} and ϕ_{j+1}, which are the only two functions overlapping ϕ_j. The near orthogonality has important computational implications.

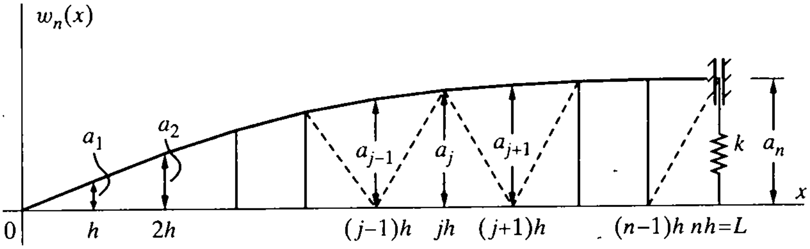


Figure 9.1 Displacement profile approximated by the finite element method

[1] This is an unfortunate term, in view of the fact that in vibrations nodes are defined as points of zero displacement. Nevertheless, the term is entrenched in finite element literature, so that we adopt it.

Comparing Eq. (9.1) with Eq. (8.76), we conclude that the finite element approximation has the same form as the Rayleigh-Ritz approximation. In fact, for second-order systems, such as strings, rods and shafts, not only that *the finite element solution* has the same form, but it *is a Rayleigh-Ritz solution*, provided the eigenvalue problem is formulated in variational form, as root functions are admissible for such systems. Moreover, observing that $w^{(n)}(x)$ is zero and $dw^{(n)}(x)/dx$ is different from zero at $x = 0$ and both $w^{(n)}(x)$ and $dw^{(n)}(x)/dx$ are different from zero at $x = L$, we conclude that $w^{(n)}(x)$ represents not a mere linear combination of admissible functions but a quasi-comparison function for second-order systems fixed at $x = 0$ and supported by a spring at $x = L$. Hence, *Eq.* (9.1) *represents a solution for the enhanced version of the Rayleigh-Ritz method* presented in Sec. 8.6. It follows that the finite element method can be based on the same mathematical foundation as the Rayleigh-Ritz method, although significant procedural differences remain. In fact, the enormous success of the finite element method can be attributed to these procedural differences. In recognition of the fact that the finite element method is a Rayleigh-Ritz method, we refer to the approach presented in Secs. 8.5 and 8.6 as the *classical Rayleigh-Ritz* method. Before we proceed with the details of the finite element method, a comparison with the classical Rayleigh-Ritz method should prove quite rewarding.

For second-order systems, the variational approach requires that the trial functions $\phi_j(x)$ be from the energy space $\mathcal{K}_G^1$, i.e., they must be mere admissible functions. In the case at hand, the functions ϕ_j must be merely continuous. This rules out piecewise constant functions, but piecewise linear functions are admissible, provided they contain no discontinuities. Moreover, these piecewise linear functions need not be defined over the entire domain $D : 0 \leq x \leq L$, but only over certain subdomains $D_j : (j-1)h \leq x \leq jh$, and can be identically zero everywhere else. We refer to such a basis as a *local basis*, in contrast with the classical Rayleigh-Ritz method, which uses *global bases*. Clearly, the roof functions of Fig. 9.2 represent a local basis from the energy space $\mathcal{K}_G^1$. In fact, *they represent the simplest set of admissible functions*. Hence, *the appeal of the finite element method can be attributed to the fact that the admissible functions constitute a local basis of the simplest form permitted by the Rayleigh-Ritz theory.*

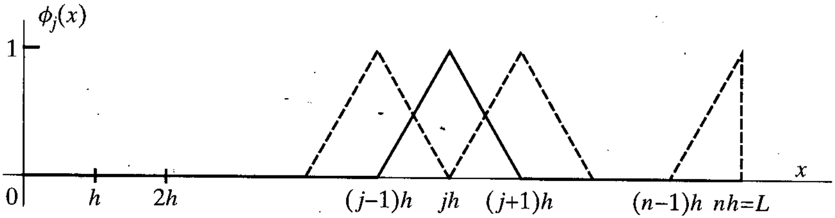


Figure 9.2 Roof functions as admissible functions

It was mentioned earlier that the roof functions are nearly orthogonal. To this should be added that *the near orthogonality holds regardless of any weighting functions*, as orthogonality is simply the result of absence of overlap between any two roof functions. The fact that all roof functions have unit amplitude has significant physical implications. Indeed, *this gives the coefficients a_j in series* (9.1) *a great deal of physical significance, as it renders a_j equal to the approximate displacement $w^{(n)}(jh)$ of the node $x = x_j = jh$.* By contrast, in the classical Rayleigh-Ritz method, the coefficients a_j represent abstract quantities, not unlike the coefficients in a Fourier series expansion. Another difference between the finite element method and the classical Rayleigh-Ritz method lies in the nature of the convergence. In the classical Rayleigh-Ritz method, the addition of another admissible function ϕ_{n+1} to series (9.1) enlarges the Ritz space from $\mathcal{R}_n$ to $\mathcal{R}_{n+1}$ without affecting $\mathcal{R}_n$, where $\mathcal{R}_n$ is a subspace of $\mathcal{R}_{n+1}$. The implication is that the mass matrices $M^{(n)}$ and $M^{(n+1)}$ on the one hand and the stiffness matrices $K^{(n)}$ and $K^{(n+1)}$ on the other hand possess the embedding property, Eqs. (8.100). As a result, the two sets of eigenvalues computed by means of the classical Rayleigh-Ritz method satisfy the separation theorem, inequalities (8.101), which guarantees monotonic convergence from above. By contrast, in the finite element method, the addition of another function ϕ_{n+1} can be done only by refining the mesh, which amounts to dividing the domain into $n + 1$ elements. As a result, *the entire set of n admissible functions ϕ_j ($j = 1, 2, \ldots, n$) changes*, in the sense that the roof functions are now defined over smaller subdomains, even though the amplitudes remain unity. In view of this, the matrices $M^{(n)}$, $M^{(n+1)}$, $K^{(n)}$ and $K^{(n+1)}$ do not possess the embedding property, and there is no mathematical proof that the separation theorem holds (Ref. 14). This does not mean that it does not hold, or that the finite element method does not converge. Indeed, the method does converge (Ref. 22), provided the elements satisfy certain conditions, but proof of convergence is not an easy matter, particularly for two-dimensional domains, where mesh refinement presents many options.

The above considerations may seem trivial when we consider the real reasons why the finite element method has gained such universal acceptance. Among these reasons we cite the virtually routine choice of admissible functions, the minimal effort in producing the mass and stiffness matrices and versatility. Although the classical Rayleigh-Ritz method has many attributes from a mathematical point of view, the method has weaknesses from a practical point of view. In particular, the selection of admissible functions is a constant source of consternation, although the development of the class of quasi-comparison functions can at times mitigate the situation. Still, the selection must be regarded as an art rather than a well-established process. By contrast, in the finite element method the selection process is relatively routine, as there is by now an established inventory of interpolation functions ready to be used for most systems of interest. The finite element method has also a clear edge in the computation of the mass and stiffness matrices. In the classical Rayleigh-Ritz method, the computation involves integrations of relatively complicated functions over the entire domain. By contrast, the generation of the mass and stiffness matrices is relatively routine in the finite element method, as it consists of assembling predetermined element matrices. To put it in simple terms, the comparison is between a custom-made process requiring a great deal of experience and physical insight and an automated, mass-production process, whereby full advantage is taken of the power of the digital computer. In fact, the wide acceptance of the finite element method can be attributed in large part to the development of increasingly powerful computers permitting numerical solutions to very complex problems. Finally, whereas the classical Rayleigh-Ritz method is essentially a structural method capable of treating linear systems with relatively uncomplicated geometry and parameter distributions, the finite element method, developed originally as a method for analyzing stresses in complicated aircraft structures, has evolved into a technique of great versatility, as it can be applied to a large variety of linear and nonlinear engineering problems.

The comparison is not as one-sided as it may seem, however. In the first place, the finite element method has a large drawback in that it requires a significantly larger number of degrees of freedom than the classical Rayleigh-Ritz method to achieve comparable accuracy (see, for example, Ref. 15). Moreover, the development of the component-mode synthesis (Sec. 8.11) and substructure synthesis (Sec. 8.12) has extended the usefulness of the concepts underlying the classical Rayleigh-Ritz method. Perhaps the single most important argument for an in-depth study of the classical Rayleigh-Ritz method is that so much of its mathematical theory extends to the finite element method.

9.2 SECOND-ORDER PROBLEMS. LINEAR INTERPOLATION FUNCTIONS

In Sec. 9.1, we introduced the general ideas behind the finite element method as applied to eigenvalue problems without entering into any procedural details. In particular, a question of interest is how to generate the mass and stiffness matrices for a given set of trial functions. Of course, in the Rayleigh-Ritz method, the mass matrix involves the evaluation of weighted inner products, Eq. (8.88b), and the stiffness matrix requires energy inner products, Eq. (8.90). In the classical Rayleigh-Ritz method, the process is carried out in one step. By contrast, in the finite element method, the process is carried out in two steps, namely, the evaluation of element matrices and the assembly of these matrices to obtain global matrices. In this section, we demonstrate the process for the second-order system shown in Fig. 9.1.

From Eq. (8.89), Rayleigh's quotient can be written the form

$$R = \frac{[w, w]}{(\sqrt{m}w, \sqrt{m}w)} = \frac{N}{D} = \frac{\sum_{j=1}^{n} N_j}{\sum_{j=1}^{n} D_j} \tag{9.2}$$

where N and D simply denote the numerator and denominator, respectively, and N_j and D_j represent the contributions from element j. We assume that the system of Fig. 9.1 represents a string in transverse vibration, so that, using the analogy with Eq. (8.104) for a rod in axial vibration, the element numerator and denominator can be written as

$$N_j = [w, w]_j = \int_{(j-1)h}^{jh} T(x) \left[\frac{dw(x)}{dx}\right]^2 dx + \delta_{nj} k w^2(L), \qquad j = 1, 2, \ldots, n \tag{9.3a}$$

$$D_j = (\sqrt{m}w, \sqrt{m}w)_j = \int_{(j-1)h}^{jh} \rho(x) w^2(x)\, dx, \qquad j = 1, 2, \ldots, n \tag{9.3b}$$

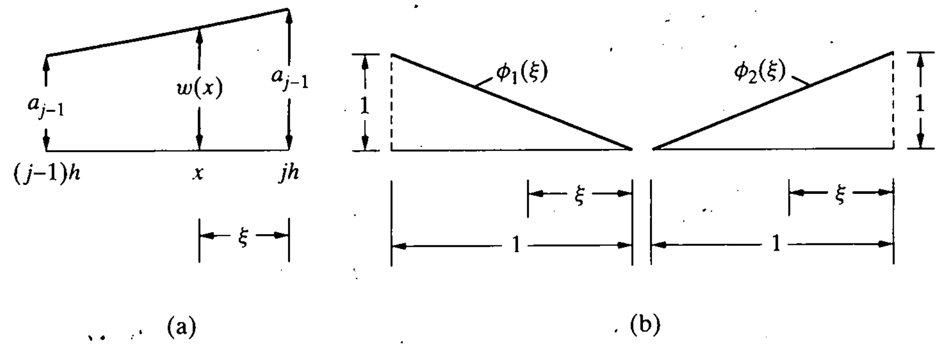


Figure 9.3 **(a)** String displacement over element j showing local coordinate **(b)** Linear interpolation functions

in which $T(x)$ is the tension in the string, k the end spring and $\rho(x)$ the mass density. At this point, we propose to simplify the evaluation of N_j and D_j greatly. To this end, we refer to Fig. 9.3a and introduce the local nondimensional coordinate[2]

$$\xi = \frac{jh - x}{h} = j - \frac{x}{h} \tag{9.4}$$

Then, we express the displacement w inside element j in terms of the nodal coordinates a_{j-1} and a_j in the form

$$w(\xi) = \phi_1(\xi)a_{j-1} + \phi_2(\xi)a_j = \boldsymbol{\phi}^T\mathbf{a}_j \tag{9.5}$$

where ϕ_1 and ϕ_2 are *trial functions*, also known as *shape functions*, or *interpolation functions*, *elements*, and $\boldsymbol{\phi} = [\phi_1 \ \phi_2]^T$ is an associated vector. Moreover, $\mathbf{a}_j = [a_{j-1} \ a_j]^T$ is known as a *nodal vector.* In the case at hand, *the elements are linear* and can be expressed as

$$\phi_i(\xi) = c_{i1} + c_{i2}\xi, \qquad i = 1, 2 \tag{9.6}$$

in which c_{i1} and c_{i2} are constants yet to be determined. They can be determined by considering the fact that the displacement $w(\xi)$ at a given node must be equal to the corresponding nodal displacement. To this end, we use Eq. (9.5) and write

$$\begin{aligned} w(1) &= \phi_1(1)a_{j-1} + \phi_2(1)a_j = a_{j-1} \\ w(0) &= \phi_1(0)a_{j-1} + \phi_2(0)a_j = a_j \end{aligned} \tag{9.7}$$

from which we conclude that ϕ_1 and ϕ_2 must satisfy the *end conditions*

$$\phi_1(1) = 1, \qquad \phi_1(0) = 0 \tag{9.8a}$$

and

$$\phi_2(1) = 0, \qquad \phi_2(0) = 1 \tag{9.8b}$$

[2] Referred to in finite element terminology as a *natural*, or *normal coordinate*, another unfortunate term in view of the fact that in vibrations the term is used for decoupled modal coordinates.

respectively. Inserting Eqs. (9.8a) and (9.8b) into Eqs. (9.6), in sequence, we obtain two pairs of algebraic equations having the solution

$$c_{11} = 0, \qquad c_{12} = 1 \tag{9.9a}$$

$$c_{21} = 1, \qquad c_{22} = -1 \tag{9.9b}$$

so that the interpolation functions have the form

$$\phi_1 = \xi, \qquad \phi_2 = 1 - \xi \tag{9.10}$$

The interpolation functions ϕ_1 and ϕ_2 are displayed in Fig. 9.3b. Note that, although we could have written the expressions for ϕ_1 and ϕ_2 directly from Fig. 9.3a, we chose to go through the process defined by Eqs. (9.6)–(9.10) in order to introduce the general procedure for generating interpolation functions.

The preceding derivation of the interpolation functions can be cast in matrix form. This may seem as a frivolous exercise, and for the simple task of generating two linear interpolation functions it is. However, the same procedure can be used for significantly more involved cases, and this simple example permits us to illustrate the ideas in an effective manner. Inserting Eqs. (9.8) into Eqs. (9.6), we obtain two nonhomogeneous algebraic equations for the constants c_{i1} and c_{i2}, which can be written in the matrix form

$$A\mathbf{c}_i = \mathbf{e}_i, \qquad i = 1, 2 \tag{9.11}$$

where

$$A = \begin{bmatrix} 1 & \xi_1 \\ 1 & \xi_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{9.12}$$

in which ξ_1 and ξ_2 are the values of ξ at the first and second node, respectively, i.e., $\xi = \xi_1 = 1$ and $\xi = \xi_2 = 0$. Moreover, $\mathbf{c}_i = [c_{i1} \ c_{i2}]^T$ are two-dimensional vectors of constants and $\mathbf{e}_i$ are the two-dimensional standard unit vectors, $\mathbf{e}_1 = [1 \ 0]^T$ and $\mathbf{e}_2 = [0 \ 1]^T$. The solution of Eq. (9.11) is simply

$$\mathbf{c}_i = A^{-1}\mathbf{e}_i, \qquad i = 1, 2 \tag{9.13}$$

from which we conclude that $\mathbf{c}_1$ and $\mathbf{c}_2$ are the first and second column of A^{-1}, respectively. Hence, we can write

$$[\mathbf{c}_1 \ \mathbf{c}_2] = \begin{bmatrix} c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} = A^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \tag{9.14}$$

which agrees with Eqs. (9.9).

At this point, we are ready to evaluate N_j and D_j. In the process, we derive the stiffness and mass matrices defining the eigenvalue problem. To this end, we must perform the integrations indicated in Eqs. (9.3). But, in view of the fact that the displacement w, Eq. (9.5), is in terms of the local coordinate ξ, we must first carry out the transformation from x to ξ. Hence, from Eq. (9.4), we transform the differential element, the derivative with respect to x and the limits of integration, as follows:

$$dx = -h\,d\xi, \qquad \frac{d}{dx} = \frac{d}{d\xi}\frac{d\xi}{dx} = -\frac{1}{h}\frac{d}{d\xi} \tag{9.15a, b}$$

$$x = jh \rightarrow \xi = 0 \tag{9.15c}$$

$$x = (j-1)h \rightarrow \xi = 1 \tag{9.15d}$$

Introducing Eqs. (9.5) and (9.15) into Eqs. (9.3), we obtain

$$\begin{aligned} N_j &= -\frac{1}{h}\int_1^0 T_j(\xi)\mathbf{a}_j^T \frac{d\boldsymbol{\phi}(\xi)}{d\xi}\frac{d\boldsymbol{\phi}'^T(\xi)}{d\xi}\mathbf{a}_j\,d\xi + \delta_{jn}k\mathbf{a}_j^T\boldsymbol{\phi}(0)\boldsymbol{\phi}^T(0)\mathbf{a}_j \\ &= \mathbf{a}_j^T K_j \mathbf{a}_j, \qquad j = 1, 2, \ldots, n \end{aligned} \tag{9.16a}$$

$$D_j = -h\int_1^0 \rho_j(\xi)\mathbf{a}_j^T\boldsymbol{\phi}(\xi)\boldsymbol{\phi}^T(\xi)\mathbf{a}_j\,d\xi = \mathbf{a}_j^T M_j \mathbf{a}_j, \quad j = 1, 2, \ldots, n \tag{9.16b}$$

where

$$K_j = \frac{1}{h}\int_0^1 T_j(\xi)\boldsymbol{\phi}'(\xi)\boldsymbol{\phi}'^T(\xi)\,d\xi + \delta_{jn}k\boldsymbol{\phi}(0)\boldsymbol{\phi}^T(0), \qquad j = 1, 2, \ldots, n \tag{9.17a}$$

are *element stiffness matrices*, in which $T_j(\xi)$ is the tension over element j and primes denote derivatives with respect to ξ, and

$$M_j = h\int_0^1 \rho_j(\xi)\boldsymbol{\phi}(\xi)\boldsymbol{\phi}^T(\xi)\,d\xi, \qquad j = 1, 2, \ldots, n \tag{9.17b}$$

are *element mass matrices*, in which $\rho_j(\xi)$ is the mass density over element j. Hence, inserting Eqs. (9.10) into Eqs. (9.17), the element stiffness and mass matrices become

$$K_j = \frac{1}{h}\int_0^1 T_j(\xi)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} d\xi + \delta_{jn}k\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad j = 1, 2, \ldots, n \tag{9.18a}$$

and

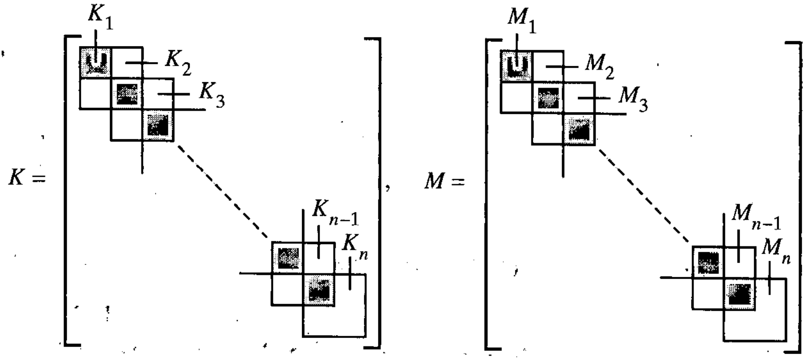
$$M_j = h\int_0^1 \rho_j(\xi)\begin{bmatrix} \xi^2 & \xi(1-\xi) \\ \xi(1-\xi) & (1-\xi)^2 \end{bmatrix} d\xi, \qquad j = 1, 2, \ldots, n \tag{9.18b}$$

respectively. It should be pointed out that, due to the fact that $a_0 = 0$, we must remove the first row and column from matrices K_1 and M_1. For computer programming purposes, it is perhaps simpler if the first row and column are removed after the assembly process.

The assembly process consists of inserting Eqs. (9.16) into Eq. (9.2) and writing the Rayleigh quotient in the discrete form

$$R = \frac{\sum_{j=1}^n \mathbf{a}_j^T K_j \mathbf{a}_j}{\sum_{j=1}^n \mathbf{a}_j^T M_j \mathbf{a}_j} = \frac{\mathbf{a}^T K \mathbf{a}}{\mathbf{a}^T M \mathbf{a}} \tag{9.19}$$

where $\mathbf{a} = [a_1\, a_2\, \ldots\, a_n]^T$ is the *nodal vector*, from which the component $a_0 = 0$ was excluded, and K and M are the *global stiffness* and *mass matrices*, from which the first row and column were deleted. In carrying out the assembly of the element matrices, we observe that the nodal displacement a_j appears as the bottom component in $\mathbf{a}_j$ and as the top component in $\mathbf{a}_{j+1}$. Consistent with this, there are two element matrix entries corresponding to a_j, the entry (2,2) of K_j and M_j and the entry (1, 1) of K_{j+1} and M_{j+1}. The assembly consists of elimination of the duplication of a_j from the nodal vector $\mathbf{a}$ and adding correspondingly the (2,2) entries in K_j and M_j to the (1,1) entries in K_{j+1} and M_{j+1} $(j = 1, 2, \ldots, n-1)$. The resulting global matrices K and M are real symmetric and positive definite; they are displayed schematically in the form



(9.20)

where we note that the shaded areas denote entries representing the sum of (1, 1) and (2, 2) entries as described above. We note that the near orthogonality of the roof functions is responsible for the *banded* nature of the stiffness and mass matrices, where the *half-bandwidth* is equal to one. A matrix is said to have half-bandwidth t if the entries (r, s) are zero for all r and s satisfying the inequality $s > r + t$.

By analogy with the classical Rayleigh-Ritz method (Sec. 8.5), the requirement that Rayleigh's quotient, Eq. (9.19), be stationary yields the algebraic eigenvalue problem

$$K\mathbf{a} = \lambda M\mathbf{a} \tag{9.21}$$

Moreover, because both K and M are real symmetric positive definite matrices, the eigenvalue problem can be reduced to one in terms of a single real symmetric positive definite matrix A (Sec. 4.6), which can be solved with ease by any of the methods discussed in Chapter 6.

For sufficiently small h, the parameters $T_j(\xi)$ and $\rho_j(\xi)$ can be regarded as being constant over the width of the element, although they can still vary from element to element, so that $T_j(\xi) \cong T_j =$ constant, $\rho_j(\xi) \cong \rho_j =$ constant $(j = 1, 2, \ldots, n)$. Under these circumstances, following the integrations indicated in Eqs. (9.18), the global stiffness and mass matrices, Eqs. (9.20), can be shown to have the form

$$K = \frac{1}{h} \times$$

$$\begin{bmatrix} T_1+T_2 & -T_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & T_2+T_3 & -T_3 & 0 & \cdots & 0 & 0 & 0 \\ & & T_3+T_4 & -T_4 & \cdots & 0 & 0 & 0 \\ & & & T_4+T_5 & \cdots & 0 & 0 & 0 \\ & & & & \cdots & \cdots & \cdots & \cdots \\ & & \text{symm} & & & T_{n-2}+T_{n-1} & -T_{n-1} & 0 \\ & & & & & & T_{n-1}+T_n & -T_n \\ & & & & & & & T_n+kh \end{bmatrix} \tag{9.22a}$$

and

$$M = \frac{h}{6}\times$$

$$\begin{bmatrix} 2(\rho_1+\rho_2) & \rho_2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 2(\rho_2+\rho_3) & \rho_3 & 0 & \cdots & 0 & 0 & 0 \\ & & 2(\rho_3+\rho_4) & \rho_4 & \cdots & 0 & 0 & 0 \\ & & & 2(\rho_4+\rho_5) & \cdots & 0 & 0 & 0 \\ & & & & \cdots & \cdots & \cdots & \cdots \\ & & \text{symm} & & & 2(\rho_{n-2}+\rho_{n-1}) & \rho_{n-1} & 0 \\ & & & & & & 2(\rho_{n-1}+\rho_n) & \rho_n \\ & & & & & & & 2\rho_n \end{bmatrix} \tag{9.22b}$$

respectively. Of course, for a uniform string under constant tension, the stiffness and mass matrices assume the familiar form

$$K = \frac{T}{h}\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ & 2 & -1 & \cdots & 0 & 0 \\ & & 2 & \cdots & 0 & 0 \\ & & & \cdots & \cdots & \cdots \\ & \text{symm} & & & 2 & -1 \\ & & & & & 1+kh/T \end{bmatrix}, \quad M = \frac{\rho h}{6}\begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 \\ & 4 & 1 & \cdots & 0 & 0 \\ & & 4 & \cdots & 0 & 0 \\ & & & \cdots & \cdots & \cdots \\ \text{symm} & & & & 4 & 1 \\ & & & & & 2 \end{bmatrix} \tag{9.23a,b}$$

The formulation presented in this section applies to all second-order systems, including not only strings in transverse vibration but also rods in axial vibration and shafts in torsional vibration.. The formulation is in terms of linear elements, which are the simplest elements satisfying the differentiability conditions required of admissible functions for second-order systems. Indeed, the displacement profile consists of a concatenation of linear segments and the transverse force profile is sectionally constant. This implies that the transverse force density, which is equal to $\partial(T\partial w/\partial x)/\partial x$, is a collection of spatial Dirac delta functions. Clearly, this may be mathematically acceptable, but is hard to reconcile with the physics of the problem, which dictates that the transverse force density be continuous throughout the domain. Of course, the problem disappears as $n \to \infty$. Not surprisingly, convergence of a solution in terms of linear elements is relatively slow, as shown in Example 9.1.

Still, such solutions are almost universally used. Convergence can be expedited by means of higher-order interpolation functions, as shown in Sec. 9.3. The question of convergence of the finite element method is examined later in this chapter.

Example 9.1

Solve the eigenvalue problem for the rod in axial vibration discussed in Sec. 8.6 by means of the finite element method in terms of linear interpolation functions. Compare the results with those obtained in Sec. 8.6 by means of the classical Rayleigh-Ritz method using mere admissible functions and the enhanced classical Rayleigh-Ritz method using quasi-comparison functions and draw conclusions.

The rod in axial vibration of Sec. 8.6 is entirely analogous to the string in transverse vibration discussed in this section. Hence, the element stiffness and mass matrices remain as given by Eqs. (9.18), except that we must replace the tension by the axial stiffness and change the notation for the mass density. Inserting Eq. (9.4) into Eqs. (8.105), the system parameters transform into

$$\begin{aligned} EA(\xi) &= \frac{6EA}{5}\left[1 - \frac{1}{2}\left(\frac{h}{L}\right)^2 (j-\xi)^2\right], \\ m(\xi) &= \frac{6m}{5}\left[1 - \frac{1}{2}\left(\frac{h}{L}\right)^2 (j-\xi)^2\right], \qquad k = \frac{EA}{L} \end{aligned} \tag{a}$$

so that, introducing Eqs. (a), into Eqs. (9.18) and carrying out the indicated integrations, we obtain the element stiffness matrices

$$K_j = \frac{6EA}{5L}\left[n - \frac{1}{6n}\left(1 - 3j + 3j^2\right)\right]\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \delta_{jn}\frac{EA}{L}\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{b}$$

and the element mass matrices

$$M_j = \frac{mL}{5n}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} - \frac{mL}{100n}\begin{bmatrix} 2(6 - 15j + 10j^2) & 3 - 10j + 10j^2 \\ 3 - 10j + 10j^2 & 2(1 - 5j + 10j^2) \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{c}$$

in which we set $L/h = n$. It should be pointed out here that the element stiffness and mass matrices, Eqs. (b) and (c), are "exact," in the sense that their entries were computed using the actual expressions for the system parameters, Eqs. (a). This was done because this example involves such small numbers n of elements that the assumption of constant parameter values over the elements would cause gross errors. In ordinary finite element practice, the parameters are taken as constant over the elements as a rule, because the number n of elements tends to be large and the width of the elements tends to be small.

Recalling that the first row and column of K_1 and M_1 must be omitted and using the scheme (9.20), we obtain the global stiffness matrix

$$K = \frac{6EAn}{5L}\left[\begin{array}{cccccc} 2 & -1 & 0 & \ldots & 0 & 0 \\ & 2 & -1 & \ldots & 0 & 0 \\ & & 2 & \ldots & 0 & 0 \\ & \text{symm} & & \ldots & \ldots & \ldots \\ & & & & 2 & -1 \\ & & & & & 1 + 1/6n \end{array}\right].$$

$$- \frac{EA}{5Ln}\begin{bmatrix} 8 & -7 & 0 & \dots & 0 & 0 \\ & 26 & -19 & \dots & 0 & 0 \\ & & 56 & \dots & 0 & 0 \\ \text{symm} & & & \dots\dots\dots\dots & \dots\dots\dots\dots & \dots\dots\dots\dots \\ & & & & 2(4 - 6n + 3n^2) & -(1 - 3n + 3n^2) \\ & & & & & 1 - 3n + 3n^2 \end{bmatrix} \quad \text{(d)}$$

and the global mass matrix

$$M = \frac{mL}{5n}\begin{bmatrix} 4 & 1 & 0 & \dots & 0 & 0 \\ & 4 & 1 & \dots & 0 & 0 \\ & & 4 & \dots & 0 & 0 \\ \text{symm} & & & \dots\dots & \dots & \dots \\ & & & & 4 & 1 \\ & & & & & 2 \end{bmatrix}$$

$$- \frac{mL}{100n^3}\begin{bmatrix} 44 & 23 & 0 & \dots & 0 & 0 \\ & 164 & 63 & \dots & 0 & 0 \\ & & 364 & \dots & 0 & 0 \\ \text{symm} & & & \dots\dots\dots\dots & \dots\dots\dots\dots & \dots\dots\dots\dots \\ & & & & 2(37 - 50n + 20n^2) & 3 - 10n + 10n^2 \\ & & & & & 2(1 - 5n + 10n^2) \end{bmatrix} \quad \text{(e)}$$

The eigenvalue problem associated with the matrices K and M has been solved in Ref. 15 for $n = 1, 2, \dots, 30$, and the first three approximate natural frequencies are listed in Table 9.1. Contrasting the results of Table 9.1 with the results of Table 8.1, we arrive to mixed conclusions. Indeed, the finite element method using linear elements shows faster convergence to ω_1 and slower convergence to ω_2 and ω_3 than the classical Rayleigh-Ritz method using mere admissible functions. On the other hand, from Tables 8.3 and 9.1, we conclude that convergence of the finite element method using linear

TABLE 9.1 First, Second, and Third Natural Frequencies Computed by the Finite Element Method Using Linear Elements

n	$\omega_1^{(n)}\sqrt{mL^2/EA}$	$\omega_2^{(n)}\sqrt{mL^2/EA}$	$\omega_3^{(n)}\sqrt{mL^2/EA}$
1	2.67261	—	—
2	2.32551	6.27163	—
3	2.26469	5.68345	9.88405
4	2.24326	5.43128	9.41491
5	2.23330	5.31181	8.98398
⋮	⋮	⋮	⋮
28	2.21609	5.10625	8.14407
29	2.21605	5.10580	8.14218
30	2.21602	5.10539	8.14049

elements is very slow compared to the enhanced Rayleigh-Ritz method using quasi-comparison functions. Indeed, in the enhanced Rayleigh-Ritz method using quasi-comparison functions $\omega_1^{(n)}$ and $\omega_2^{(n)}$ reach convergence to ω_1 and ω_2, respectively, with $n = 6$ and $\omega_3^{(n)}$ achieves convergence to ω_3 with $n = 13$. By contrast, the finite element method is still not close to convergence with $n = 30$.

9.3 HIGHER-DEGREE INTERPOLATION FUNCTIONS

As pointed out on several occasions, one of the reasons for the wide appeal of the finite element method is the simplicity of the admissible functions. Indeed, as a rule, they are low-degree polynomials. For second-order systems, the lowest degree admissible is the first, i.e., the interpolation functions are linear. But, as discovered in Example 9.1, there is a price to be paid for this simplicity in that convergence tends to be slow. Hence, the question arises as to whether convergence can be accelerated by using higher-degree polynomials. In this section, we address the question by considering quadratic and cubic interpolation functions. Higher-degree polynomials can be generated by means of Lagrange's interpolation formula (Ref. 11). Quadratic and cubic polynomials can be generated just as easily by means of the approach of Sec. 9.2.

Quadratic interpolation functions, or *quadratic elements*, can be generated by means of the quadratic polynomials

$$\phi_i = c_{i1} + c_{i2}\xi + c_{i3}\xi^2, \qquad i = 1, 2, 3 \tag{9.24}$$

but we run immediately into a problem when we try to determine the coefficients c_{i1}, c_{i2} and c_{i3} $(i = 1, 2, 3)$. Indeed, there are three coefficients to be determined for every element and only two nodes available for producing the three necessary conditions. It follows that another node must be created. For simplicity, we choose the location of the third node at $\xi = 1/2$ and denote the corresponding nodal displacement by $a_{j-1/2}$. Clearly, the point $\xi = 1/2$ represents an *internal node*, which makes the points $\xi = 0$ and $\xi = 1$ *external nodes*. Following the pattern established in Sec. 9.2, we express the approximate displacement in the form

$$w(\xi) = \phi_1(\xi)a_{j-1} + \phi_2(\xi)a_{j-1/2} + \phi_3(\xi)a_j = \boldsymbol{\phi}^T \mathbf{a}_j \tag{9.25}$$

where the vector $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \phi_3]^T$ of interpolation functions and the nodal vector $\mathbf{a}_j = [a_{j-1}\ a_{j-1/2}\ a_j]^T$ are now three-dimensional.

The determination of the coefficients c_{i1}, c_{i2}, c_{i3} $(i = 1, 2, 3)$ to be used in Eqs. (9.24) follows the pattern established in Sec. 9.2, Eqs. (9.11)–(9.14). To this end, we observe that the values of ξ at the nodes, taken in sequence, are $\xi_1 = 1$, $\xi_2 = 1/2$ and $\xi_3 = 0$ and write

$$A = \begin{bmatrix} 1 & \xi_1 & \xi_1^2 \\ 1 & \xi_2 & \xi_2^2 \\ 1 & \xi_3 & \xi_3^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1/2 & 1/4 \\ 1 & 0 & 0 \end{bmatrix} \tag{9.26}$$

so that

$$[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3] = \begin{bmatrix} c_{11} & c_{21} & c_{31} \\ c_{12} & c_{22} & c_{32} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} = A^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ -1 & 4 & -3 \\ 2 & -4 & 2 \end{bmatrix} \tag{9.27}$$

Inserting the values of the constants c_{i1}, c_{i2} and c_{i3} $(i = 1, 2, 3)$ from Eq. (9.27) into Eqs. (9.24), we obtain the *quadratic interpolation functions*, or *quadratic elements*

$$\phi_1 = \xi(2\xi - 1), \qquad \phi_2 = 4\xi(1 - \xi), \qquad \phi_3 = (1 - \xi)(1 - 2\xi) \tag{9.28}$$

The functions ϕ_i $(i = 1, 2, 3)$ are displayed in Fig. 9.4.

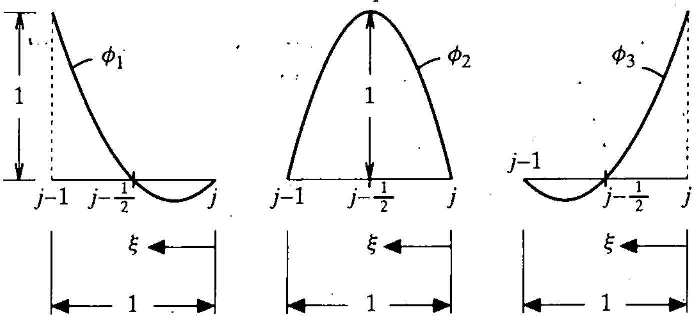


Figure 9.4 Quadratic interpolation functions

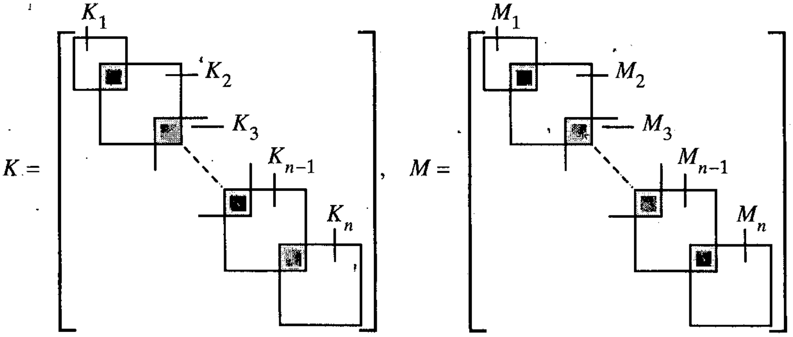
The stiffness and mass matrices remain in the general form of Eqs. (9.17a) and (9.17b), respectively. Inserting Eqs. (9.28) into Eqs. (9.17), we obtain the element stiffness and mass matrices in the more explicit form

$$K_j = \frac{1}{h}\int_0^1 T_j(\xi)\begin{bmatrix} (4\xi - 1)^2 & 4(4\xi - 1)(1 - 2\xi) & (4\xi - 1)(4\xi - 3) \\ & 16(1 - 2\xi)^2 & 4(1 - 2\xi)(4\xi - 3) \\ \text{symm} & & (4\xi - 3)^2 \end{bmatrix} d\xi + \delta_{jn}k\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{9.29a}$$

and

$$M_j = h\int_0^1 \rho_j(\xi)\begin{bmatrix} \xi^2(2\xi - 1)^2 & 4\xi^2(2\xi - 1)(1 - \xi) & -\xi(1 - \xi)(1 - 2\xi)^2 \\ & 16\xi^2(1 - \xi)^2 & 4\xi(1 - \xi)^2(1 - 2\xi) \\ \text{symm} & & (1 - \xi)^2(1 - 2\xi)^2 \end{bmatrix} d\xi, \qquad j = 1, 2, \ldots, n \tag{9.29b}$$

respectively. Using the assembly process described in Sec. 9.2 and recalling once again that $a_0 = 0$, the *global stiffness* and *mass matrices* can be displayed in the schematic form



(9.30)

The shaded matrix elements correspond to the external nodes and the elements in-between correspond to internal nodes. Both matrices are banded, with a half-bandwidth equal to two.

For uniform tension, the element stiffness matrices reduce to

$$K_j = \frac{T}{3h}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix} + \delta_{jn}k\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{9.31a}$$

which are all the same except K_n, and for uniform mass density the element mass matrices become

$$M_j = \frac{h\rho}{30}\begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{9.31b}$$

and these are all the same. Using the scheme given by Eqs. (9.30), the global stiffness matrix and global mass matrix can be shown to have the form

$$K = \frac{T}{3h}\begin{bmatrix} 16 & -8 & 0 & \ldots & 0 & 0 & 0 \\ & 14 & -8 & \ldots & 0 & 0 & 0 \\ & & 16 & \ldots & 0 & 0 & 0 \\ & & & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ & \text{symm} & & & 14 & -8 & 1 \\ & & & & & 16 & -8 \\ & & & & & & 7 + 3kh/T \end{bmatrix} \tag{9.32a}$$

and

$$M = \frac{h\rho}{30}\begin{bmatrix} 16 & 2 & 0 & \ldots & 0 & 0 & 0 \\ & 8 & 2 & \ldots & 0 & 0 & 0 \\ & & 16 & \ldots & 0 & 0 & 0 \\ & & & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ & \text{symm} & & & 8 & 2 & -1 \\ & & & & & 16 & 2 \\ & & & & & & 4 \end{bmatrix} \tag{9.32b}$$

respectively.

The same procedure can be used to derive the *cubic interpolation functions*. To this end, we begin with the polynomials

$$\phi_i = c_{i1} + c_{i2}\xi + c_{i3}\xi^2 + c_{i4}\xi^3, \qquad i = 1, 2, 3, 4 \tag{9.33}$$

and observe that now we must have four nodes, *two external and two internal nodes*. Then, assuming that the approximate displacement has the expression

$$w(\xi) = \phi_1(\xi)a_{j-1} + \phi_2(\xi)a_{j-2/3} + \phi_3(\xi)a_{j-1/3} + \phi_4(\xi)a_j = \boldsymbol{\phi}^T\mathbf{a}_j \tag{9.34}$$

where the four-dimensional vectors $\boldsymbol{\phi}$ and $\mathbf{a}_j$ are obvious, and following the established pattern, we obtain the *cubic elements* (Problem 9.7)

$$\begin{aligned} \phi_1 &= \frac{1}{2}\xi(2 - 9 + 9\xi^2), \qquad & \phi_2 &= -\frac{9}{2}\xi(1 - 4\xi + 3\xi^2) \\ \phi_3 &= \frac{9}{2}\xi(2 - 5\xi + 3\xi^2), \qquad & \phi_4 &= 1 - \frac{11}{2}\xi + 9\xi^2 - \frac{9}{2}\xi^3 \end{aligned} \tag{9.35}$$

They are shown in Fig. 9.5. The calculation of the 4 × 4 element stiffness and mass matrices for cubic elements and given system parameters can be carried out by inserting Eqs. (9.35) into Eqs. (9.17) and performing the indicated integrations (Problem 9.8).

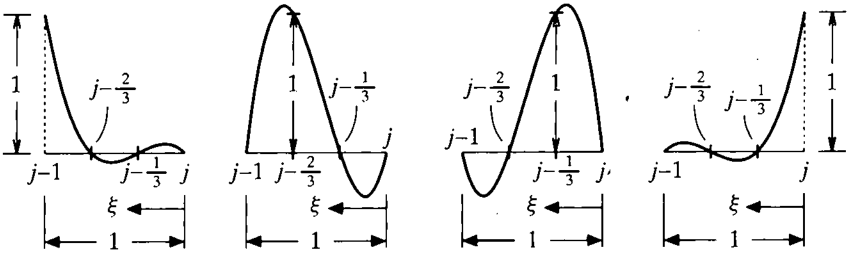


Figure 9.5 Cubic interpolation functions

In general, approximate eigenvalues computed on the basis of quadratic interpolation functions tend to be more accurate than those computed using linear interpolation functions, and the same can be said for approximate eigenvalues based on cubic elements compared to those based on quadratic elements.

Example 9.2

Solve the eigenvalue problem for the rod in axial vibration of Example 9.1 by means of the finite element method in two ways, first using quadratic interpolation functions and then using cubic interpolation functions. Compare the results with those obtained in Sec. 8.6 by the classical Rayleigh-Ritz method in conjunction with quasi-comparison functions and in Example 9.1 by the finite element method using linear interpolation functions and draw conclusions.

The 3 × 3 element stiffness and mass matrices for quadratic elements are obtained by inserting the system parameters given by Eqs. (a) of Example 9.1 into Eqs. (9.29) with $T_j(\xi)$ and $\rho_j(\xi)$ replaced by $EA_j(\xi)$ and $m_j(\xi)$, respectively, and carrying out the indicated integrations. Then, the global stiffness and mass matrices are assembled by

TABLE 9.2 First Three Natural Frequencies Computed by the Finite Element Method Using Quadratic (FEMQ) and Cubic (FEMC) Interpolation Functions

n	$\omega_1^{(n)}\sqrt{mL^2/EA}$		$\omega_2^{(n)}\sqrt{mL^2/EA}$		$\omega_3^{(n)}\sqrt{mL^2/EA}$	
	FEMQ	FEMC	FEMQ	FEMC	FEMQ	FEMC
1	—	—	—	—	—	—
2	2.23433	—	6.27984	—	—	—
3	—	2.1558	—	5.25278	—	10.76486
4	2.21705	—	5.18817	—	9.13233	—
5	—	—	—	—	—	—
6	2.21584	2.21553	5.12170	5.10352	8.30344	8.21161
7	—	—	—	—	—	—
8	2.21563	—	5.10695	—	8.18985	—
9	—	2.21553	—	5.09988	—	8.12705
10	2.21557	—	5.10265	—	8.14851	—
11	—	—	—	—	—	—
12	2.21555	2.21552	5.10105	5.09959	8.13242	8.11835
13	—		—	—	—	—
14	2.21554		5.10036	—	8.12520	—
15	—		—	5.09954	—	8.11688
16	2.21553		5.10002	—	8.12160	—
17	—		—	—	—	—
18	2.21553		5.09983	5.09953	8.11965	8.11651
19	—		—	—	—	—
20	2.21553		5.09973	—	8.11852	—
21	—		—	5.09953	—	8.11640
22	2.21553		5.09966	—	8.11783	—
23	—		—	—	—	—
24	2.21553		5.09962	5.09953	8.11739	8.11635
25	—		—	—	—	—
26	2.21553		5.09960	—	8.11710	—
27	—		—	5.09953	—	8.11634
28	2.21553		5.09958	—	8.11690	—
29	—		—	—	—	—
30	2.21552		5.09956	5.09953	8.11676	8.11633

using the pattern exhibited in Eqs. (9.30). We note that, as the number of finite elements increases by one the dimension of the global matrices increases by two.

In a similar fashion, the 4 × 4 element stiffness and mass matrices for cubic elements are computed by introducing the same system parameters and Eq. (9.35) into

Eqs. (9.17) and performing the prescribed integrations. Then, the global stiffness and mass matrices are assembled in the customary way. But, because in the case of cubic elements there are two internal nodes for every external node, the global matrices are characterized by two nonshaded entries on the main diagonal separating any pair of shaded entries. Consistent with this, as the number of finite elements increases by one, the dimension of the global matrices increases by three. The derivation of the global matrices follows the established pattern and is the subject of Problem 9.8.

The eigenvalue problems corresponding to quadratic and cubic elements have been solved in Ref. 15, and the results are displayed in Table 9.2. In the first place, we note that, in using quadratic elements, the number of degrees of freedom increases by two at a time, and so does the number of computed natural frequencies. In the case of cubic elements, the number of computed natural frequencies increases by three at a time. This explains the empty spaces in Table 9.2. A comparison of Tables 9.1 and 9.2 confirms the expectation that the finite element model based on quadratic elements has better convergence characteristics than the model based on linear elements, and the model using cubic elements converges faster than the model using quadratic elements. This statement must be tempered by the realization that the relatively good results obtained by the finite element method in conjunction with cubic interpolation functions fall far short of the results obtained by the enhanced Rayleigh-Ritz method using quasi-comparison functions. Indeed, a comparison of Tables 8.3 and 9.2 reveals convergence of the first three natural frequencies computed by the enhanced Rayleigh-Ritz method with six-, six- and thirteen-degree-of-freedom models. By contrast, the first natural frequency computed by the finite element method using cubic elements converges with a twelve-degree-of-freedom model, whereas the second and third natural frequencies have not achieved convergence with a thirty-degree-of-freedom model.

9.4 BEAMS IN BENDING VIBRATION

According to the Rayleigh-Ritz theory, the lowest degree polynomials admissible for beams in bending are quadratic, so that we consider them as possible candidates for interpolation functions. To this end, we recognize that in bending both the displacement and the slope must be continous at nodal points. Because there are two nodes, it follows that every interpolation function must satisfy four end conditions. But, as shown in Sec. 9.3, quadratic interpolation functions are defined by only three constants, from which it follows that second-degree polynomials cannot be used as interpolation functions. Hence, the lowest-degree polynomials admissible are cubic, which are defined by four constants.

The derivation of the interpolation functions for beams in bending can be carried out by considering the typical finite element shown in Fig. 9.6, in which w_{j-1}, θ_{j-1} and w_j, θ_j denote the translation and rotation of the beam at the nodal points $j-1$ and j, respectively. Using the same approach as in Secs. 9.2 and 9.3, we express the displacement at point ξ in the form

$$w(\xi) = \phi_1(\xi)w_{j-1} + \phi_2(\xi)h\theta_{j-1} + \phi_3(\xi)w_j + \phi_4(\xi)h\theta_j = \boldsymbol{\phi}^T\mathbf{a}_j \qquad (9.36)$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \phi_3\ \phi_4]^T$ and $\mathbf{a}_j = [w_{j-1}\ h\theta_{j-1}\ w_j\ h\theta_j]^T$. Note that we multiplied the rotations θ_{j-1} and θ_j by h so as to ensure that the interpolation functions ϕ_1, ϕ_2, ϕ_3 and ϕ_4 are all dimensionless.

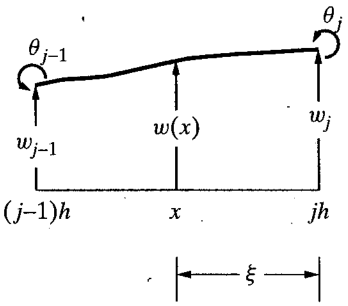


Figure 9.6 Beam displacement over element j

As in Sec. 9.3, cubic elements have the form

$$\phi_i(\xi) = c_{i1} + c_{i2}\xi + c_{i3}\xi^2 + c_{i4}\xi^3, \qquad i = 1, 2, 3, 4 \tag{9.37}$$

in which c_{i1}, c_{i2}, c_{i3} and c_{i4} $(i = 1, 2, 3, 4)$ are coefficients to be determined. To this end, we replace the rotations θ_{j-1} and θ_j by the slopes of the displacement curve at the nodes $j-1$ and j, respectively, recall Eq. (9.15b) and write

$$\theta = \frac{dw}{dx} = \frac{dw}{d\xi}\frac{d\xi}{dx} = -\frac{1}{h}\frac{dw}{d\xi} = -\frac{1}{h}w' \tag{9.38}$$

Then, following the pattern of Secs. 9.2 and 9.3 and considering Eq. (9.38), we can write

$$A = \begin{bmatrix} 1 & \xi_1 & \xi_1^2 & \xi_1^3 \\ 0 & -1 & -2\xi_1 & -3\xi_1^2 \\ 1 & \xi_2 & \xi_2^2 & \xi_2^3 \\ 0 & -1 & -2\xi_2 & -3\xi_2^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & -1 & -2 & -3 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \tag{9.39}$$

in which we substituted $\xi_1 = 1$ and $\xi_2 = 0$. Moreover, we observe that the second row of A is the negative of the derivative of the first row with respect to ξ_1 and an analogous statement can be made about the fourth and third rows. Hence, the desired coefficients are obtained as the elements of the columns of A^{-1}, or

$$[\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3\ \mathbf{c}_4] = A^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ 3 & 1 & -3 & 2 \\ -2 & -1 & 2 & -1 \end{bmatrix} \tag{9.40}$$

Inserting the elements of the columns of A^{-1} into Eqs. (9.37), we obtain the interpolation functions

$$\phi_1 = 3\xi^2 - 2\xi^3, \quad \phi_2 = \xi^2 - \xi^3, \quad \phi_3 = 1 - 3\xi^2 + 2\xi^3, \quad \phi_4 = -\xi + 2\xi^2 - \xi^3 \tag{9.41}$$

The interpolation functions given by Eqs. (9.41) are known as *Hermite cubics*. They are displayed in Fig. 9.7.

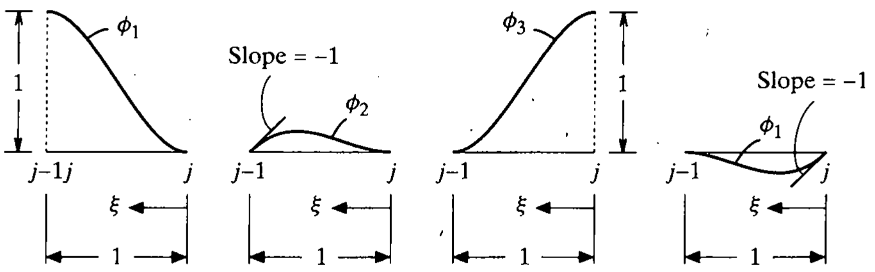


Figure 9.7 Hermite cubics

The derivation of the global stiffness and mass matrices follows a pattern similar to that established in Sec. 9.2. To demonstrate the process, we consider a beam in bending subjected to an axial force and recall the definition of Rayleigh's quotient

$$R = \frac{[w, w]}{\left(\sqrt{m}w, \sqrt{m}w\right)} = \frac{\sum_{j=1}^{n} N_j}{\sum_{j=1}^{n} D_j} \tag{9.42}$$

first encountered in Sec. 9.2, where, from Secs. 7.2 and 7.5,

$$N_j = [w, w]_j = \int_{(j-1)h}^{jh} \left\{ EI(x)\left[\frac{d^2w(x)}{dx^2}\right]^2 + P(x)\left[\frac{dw(x)}{dx}\right]^2 \right\} dx, \quad j = 1, 2, \dots, n \tag{9.43a}$$

$$D_j = \left(\sqrt{m}w, \sqrt{m}w\right)_j = \int_{(j-1)h}^{jh} m(x)w^2(x)\,dx, \quad j = 1, 2, \dots, n \tag{9.43b}$$

Then, changing variables from x to the local coordinate ξ according to Eqs. (9.15) and using Eq. (9.36), we have

$$N_j = \mathbf{a}_j^T K_j \mathbf{a}_j, \qquad D_j = \mathbf{a}_j^T M_j \mathbf{a}_j, \qquad j = 1, 2, \dots, n \tag{9.44a, b}$$

in which

$$K_j = \frac{1}{h^3}\int_0^1 \left[EI_j(\xi)\boldsymbol{\phi}''\boldsymbol{\phi}''^T + h^2 P_j(\xi)\boldsymbol{\phi}'\boldsymbol{\phi}'^T\right] d\xi, \qquad j = 1, 2, \dots, n \tag{9.45a}$$

$$M_j = h\int_0^1 m_j(\xi)\boldsymbol{\phi}\boldsymbol{\phi}^T d\xi, \qquad j = 1, 2, \dots, n \tag{9.45b}$$

are the element stiffness matrix and element mass matrix, respectively.

Finally, there is the assembly process, which requires the system boundary conditions. As an illustration, we consider a beam clamped at $x = 0$ and free at $x = L$. In this case, the assembly process yields global stiffness and mass matrices of the form depicted in Eqs. (9.20), except that now matrices K_1 and M_1 are 2×2 and matrices K_j and M_j are 4×4 $(j = 2, 3, \dots, n)$.

Example 9.3

Use the finite element method to derive the global stiffness and mass matrices for a uniform helicopter blade rotating with the constant angular velocity Ω.

From Example 7.4, the energy inner product has the expression

$$[w, w] = \int_0^L \left\{ EI \left[\frac{d^2 w(x)}{dx^2} \right]^2 + P(x) \left[\frac{dw(x)}{dx} \right]^2 \right\} dx \tag{a}$$

where, from Example 7.2, the axial force is given by

$$P(x) = \int_x^L m\Omega^2 \zeta \, d\zeta = \frac{1}{2} m\Omega^2 L^2 \left[1 - \left(\frac{x}{L} \right)^2 \right] \tag{b}$$

The weighted inner product is simply

$$\left(\sqrt{m} w, \sqrt{m} w \right) = \int_0^L m w^2(x) dx \tag{c}$$

Hence, using Eqs. (9.45), the element stiffness and mass matrices have the form

$$K_j = \frac{1}{h^3} \left\{ EI \int_0^1 \boldsymbol{\phi}'' \boldsymbol{\phi}''^T \, d\xi + \frac{1}{2} m\Omega^2 L^2 h^2 \int_0^1 \left[1 - \left(\frac{h}{L} \right)^2 (j - \xi)^2 \right] \boldsymbol{\phi}' \boldsymbol{\phi}'^T \, d\xi \right\}, \quad j = 1, 2, \ldots, n \tag{d}$$

and

$$M_j = hm \int_0^L \boldsymbol{\phi} \boldsymbol{\phi}^T d\xi, \qquad j = 1, 2, \ldots, n \tag{e}$$

respectively, where, from Eqs. (9.41), the vector of interpolation functions is

$$\boldsymbol{\phi} = \left[3\xi^2 - 2\xi^3 \quad \xi^2 - \xi^3 \quad 1 - 3\xi^2 + 2\xi^3 \quad -\xi + 2\xi^2 - \xi^3 \right]^T \tag{f}$$

Inserting Eqs. (f) into Eqs. (d) and (e) and carrying out the appropriate integrations, we obtain the explicit element stiffness matrices

$$\begin{aligned} K_j = {} & \frac{EI}{h^3} \begin{bmatrix} 12 & 6 & -12 & 6 \\ & 4 & -6 & 2 \\ \text{symm} & & 12 & -6 \\ & & & 4 \end{bmatrix} \\ & + \frac{1}{2} m\Omega^2 L n \left(\frac{1}{30} \left[1 - \left(\frac{j}{n} \right)^2 \right] \begin{bmatrix} 36 & -3 & -36 & 3 \\ & 4 & -3 & 1 \\ \text{symm} & & 36 & -3 \\ & & & 4 \end{bmatrix} \right. \\ & \left. - \frac{j}{30n} \begin{bmatrix} 36 & 0 & -36 & 6 \\ & 6 & 0 & -1 \\ \text{symm} & & 36 & -6 \\ & & & 2 \end{bmatrix} + \frac{1}{210n^2} \begin{bmatrix} 72 & -6 & -72 & -15 \\ & 18 & 6 & -3 \\ \text{symm} & & 72 & 15 \\ & & & 4 \end{bmatrix} \right), \\ & \qquad j = 1, 2, \ldots, n \end{aligned} \tag{g}$$

and element mass matrices

$$M_j = \frac{mL}{420n} \begin{bmatrix} 156 & 22 & 54 & -13 \\ & 4 & 13 & -3 \\ \text{symm} & & 156 & -22 \\ & & & 4 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{h}$$

Then, recalling that we must delete the first two rows and columns from K_1 and M_1, the global stiffness matrix has the form

$$
K = \frac{EI}{h^3}\left[\begin{array}{ccccccccccc}
24 & 0 & -12 & 6 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & 8 & -6 & 2 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & & 24 & 0 & -12 & 6 & \ldots & 0 & 0 & 0 & 0 \\
 & & & 8 & -6 & 2 & \ldots & 0 & 0 & 0 & 0 \\
 & & & & \multicolumn{7}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
 & & & & & & & 24 & 0 & -12 & 6 \\
 & & \text{symm} & & & & & & 8 & -6 & 2 \\
 & & & & & & & & & 12 & -6 \\
 & & & & & & & & & & 4
\end{array}\right]
+ \frac{m\Omega^2 Ln}{60}\left[\begin{array}{ccccccccccc}
72 & -6 & -36 & 3 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & 8 & -3 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & & 72 & -6 & -36 & 3 & \ldots & 0 & 0 & 0 & 0 \\
 & & & 8 & -3 & 1 & \ldots & 0 & 0 & 0 & 0 \\
 & & & & \multicolumn{7}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
 & & & & & & & 72 & -6 & -36 & 3 \\
 & & & \text{symm} & & & & & 8 & -3 & 1 \\
 & & & & & & & & & 36 & -3 \\
 & & & & & & & & & & 4
\end{array}\right]
$$

$$
- \frac{m\Omega^2 L}{n}\left[\begin{array}{ccccccccccc}
180 & -15 & -144 & 12 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & 20 & -12 & 4 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & & 468 & -39 & -324 & 27 & \ldots & 0 & 0 & 0 & 0 \\
 & & & 52 & -27 & 9 & \ldots & 0 & 0 & 0 & 0 \\
 & & & & \multicolumn{7}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
 & & & & & & & 36(2n^2-2n+1) & -3(2n^2-2n+1) & -36 & 3 \\
 & & & \text{symm} & & & & & 4(2n^2-2n+1) & -3 & 1 \\
 & & & & & & & & & 36 & -3 \\
 & & & & & & & & & & 4
\end{array}\right]
$$

$$
- \frac{1}{30n}\left[\begin{array}{ccccccccccc}
108 & -6 & -72 & 12 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & 14 & 0 & -2 & 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\
 & & 180 & -12 & -108 & 18 & \ldots & 0 & 0 & 0 & 0 \\
 & & & 22 & 0 & -3 & \ldots & 0 & 0 & 0 & 0 \\
 & & & & \multicolumn{7}{c}{\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots} \\
 & & & & & & & 36(2n-1) & -6(n-1) & -36n & 6n \\
 & & & \text{symm} & & & & & 2(4n-1) & 0 & -n \\
 & & & & & & & & & 36n & -6n \\
 & & & & & & & & & & 2n
\end{array}\right]
$$

$$+\frac{1}{210n^2}\left[\begin{array}{cccccccllll}
144 & 9 & -72 & -15 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
 & 22 & 6 & -3 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
 & & 144 & 9 & -72 & -15 & \cdots & 0 & 0 & 0 & 0 \\
 & & & 22 & 6 & -3 & \cdots & 0 & 0 & 0 & 0 \\
 & & & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & & & & & & & 144 & 9 & -72 & -15 \\
 & & & \text{symm} & & & & & 22 & 6 & -3 \\
 & & & & & & & & & 72 & 15 \\
 & & & & & & & & & & 4
\end{array}\right] \qquad \text{(i)}$$

and the global mass matrix is

$$M = \frac{mL}{420n}\left[\begin{array}{ccccccccccc}
312 & 0 & 54 & -13 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
 & 8 & 13 & -3 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 \\
 & & 312 & 0 & 54 & -13 & \cdots & 0 & 0 & 0 & 0 \\
 & & & 8 & 13 & -3 & \cdots & 0 & 0 & 0 & 0 \\
 & & & & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
 & & & & & & & 312 & 0 & 54 & -13 \\
 & & & \text{symm} & & & & & 8 & 13 & -3 \\
 & & & & & & & & & 156 & -22 \\
 & & & & & & & & & & 4
\end{array}\right] \qquad \text{(j)}$$

9.5 VIBRATION OF MEMBRANES. TRIANGULAR ELEMENTS

The transverse vibration of membranes is described by two-dimensional boundary-value problems, which are materially more complex than one-dimensional ones. Indeed, as explained in Sec. 7.12, an important consideration in two-dimensional problems is the shape of the boundary. In the relatively few cases in which closed-form solutions are possible, the boundary shape dictates the choice of coordinates. More often than not, however, the shape of the boundary places closed-form solutions beyond reach, making approximate solutions a virtual necessity. It is here that the versatility of the finite element method becomes evident, as it permits solutions where other methods fail. In using the finite element method, the choice of coordinates tends to fade as an issue, at least in the classical sense.

In seeking a closed-form solution, a smooth boundary can be taken as an indication that the solution may be smooth. In contrast, in approximate solutions by the finite element method, a smooth boundary creates a new problem in that the choice of the finite element mesh must not only permit an accurate approximate solution but also the perimeter of the bounding polygon must approximate well the boundary itself. Indeed, when the boundary is smooth it is generally not feasible to devise a finite element mesh covering the domain D exactly, so that there is a difference between D and the domain $D^{(n)}$ covered by the finite element mesh, where the latter is referred to as the *finite element domain*. An important question relates to the shape of the elements minimizing the difference $D - D^{(n)}$ between the actual domain bounded by a smooth curve and the finite element domain bounded by a polygon. Experience shows that triangular elements are particularly suited to the task of filling tightly domains with smooth boundaries, thus minimizing $D - D^{(n)}$. Of course, this implies the use of increasingly smaller elements as we approach the boundary. This is demonstrated in Fig. 9.8, in which $D - D^{(n)}$ represents the union

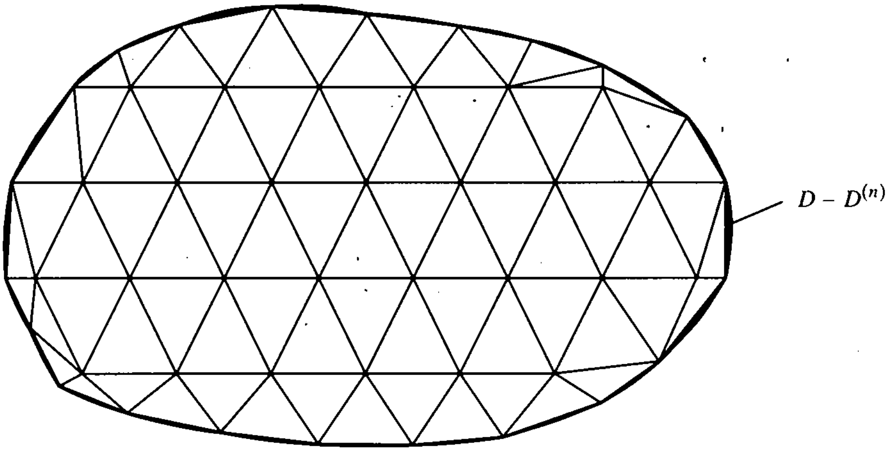


Figure 9.8 Finite element mesh for a two-dimensional domain

of the small shaded areas. It is clear that, as the number n of elements increases, the difference $D - D^{(n)}$ tends to disappear.

The problem of approximating a domain D with a smooth boundary S by a domain $D^{(n)}$ with a polygonal boundary $S^{(n)}$ can be regarded as the problem of approximating S by $S^{(n)}$. The latter problem is very old indeed, as mathematicians in ancient times were able to calculate π with a high degree of accuracy by calculating the perimeter of a polygon with equal sides inscribed in a circle and progressively increasing the number of sides. This purely geometric idea was given physical content by Courant (Ref. 8), who used triangular elements to produce an approximate solution to the plane torsion problem for multiconnected domains, thus anticipating the finite element method by over a decade.

Triangular elements are equally useful in cases in which the boundaries possess corners. There are cases, however, in which quadrilateral elements may prove superior, as they are able to produce a finite element mesh with fewer elements.

Another question arising in two-dimensional problems not encountered in one-dimensional ones relates to the choice of a numbering scheme for the elements and the nodal points. In the one-dimensional case, the nodes are along a straight line and the numbering of the elements and nodes progresses uninterruptedly from one boundary point to the other. The resulting mass and stiffness matrices are banded, which is due to the fact that the interpolation functions are nearly orthogonal. Whereas a similar situation exists in the case of two-dimensional problems, the bandedness is not guaranteed. Indeed, in two-dimensional problems there is a large variety of choices in the numbering of the elements and nodes, and the bandwidth tends to differ from choice to choice. Of course, the most desirable numbering scheme is the one minimizing the bandwidth of the mass and stiffness matrices.

The procedure for obtaining the stiffness and mass matrices for systems defined over two-dimensional domains parallels the procedure for one-dimensional domains, except that certain details are different. In particular, we write once again Rayleigh's quotient in the form given by Eq. (9.2), where $N_j = [w, w]_j$ is the element energy inner product and $D_j = \left(\sqrt{m}w, \sqrt{m}w\right)_j$ is the element weighted inner product. Then, we use $[w, w]_j$ and $\left(\sqrt{m}w, \sqrt{m}w\right)_j$ to derive element stiffness and mass matrices K_j and M_j, respectively. Finally, we assemble the element stiffness and mass matrices to obtain global stiffness and mass matrices. Differences arise in the determination of local, natural coordinates (in the finite element sense) and in the assembly process, both being more involved for two-dimensional problems than for one-dimensional ones. Perhaps this is the time to explain a statement made earlier in this section that the choice of coordinates is not an issue when the finite element method is applied to two-dimensional problems. Indeed, the shape of the boundary plays no particular role in choosing coordinates and, almost as a rule, $[w, w]_j$ and $(\sqrt{m}w, \sqrt{m}w)_j$ are expressed in terms of rectangular coordinates prior to the transformation to natural coordinates. This is the case even for boundaries with nice analytical form, such as circular and elliptic.

Next, we consider the problem of deriving stiffness and mass matrices for a membrane in transverse vibration by means of the finite element method. From Eq. (7.252), we conclude that the elements must be from $\mathcal{K}_G^1$, so that linear elements are admissible. In fact, they are the simplest elements admissible. *Linear triangular elements* have some very desirable features. One of them is that the plane

$$w(x, y) = a_1 + a_2 x + a_3 y \tag{9.46}$$

defining the displacement at any point x, y of the element, is determined uniquely by the values w_1, w_2 and w_3 of the nodal displacements, namely, the values of w at the three vertices of the triangle (Fig. 9.9). Moreover, we will show that the value of w along an element edge reduces to a linear function of a single variable, a local coordinate defined uniquely by the nodal displacements of the two end points of the edge, which implies that the nodal displacement at the third point does not affect the value of w along the edge in question. It follows that *the continuity of w across the edge is guaranteed by continuity at the nodal points.*

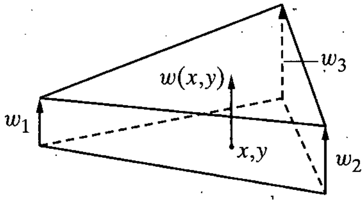


Figure 9.9 Planar displacement over triangular element

Using the analogy with Eq. (9.1), and dropping the superscript (n) for simplicity, we write the finite element solution in the general form

$$w(x, y) = \sum_{j=1}^{n} a_j \phi_j(x, y) \tag{9.47}$$

where $\phi_j(x, y)$ are trial functions. For linear elements, the trial functions have the form of *pyramid functions*, such as that depicted in Fig. 9.10, and they represent the two-dimensional counterpart of the roof functions of Fig. 9.2. We take the height of the pyramid equal to unity, so that *the coefficients a_j can be identified as the displacement of the membrane at the nodal points.*

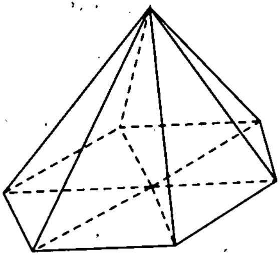


Figure 9.10 Pyramid function

At this point, we turn our attention to the derivation of local coordinates. To this end, we consider the triangular element shown in Fig. 9.11 and denote the vectors from the origin O of the rectangular system x, y to the vertices 1, 2 and 3 by $\mathbf{r}_1$, $\mathbf{r}_2$ and $\mathbf{r}_3$, respectively. Then, letting $P(x, y)$ be a typical point inside the triangle and denoting the radius vector from O to P by $\mathbf{r}$, the vectors from 1, 2 and 3 to P are simply $\mathbf{r} - \mathbf{r}_1$, $\mathbf{r} - \mathbf{r}_2$ and $\mathbf{r} - \mathbf{r}_3$, respectively. Because the three vectors lie in the same plane, they must satisfy the relation

$$c_1(\mathbf{r} - \mathbf{r}_1) + c_2(\mathbf{r} - \mathbf{r}_2) + c_3(\mathbf{r} - \mathbf{r}_3) = \mathbf{0} \tag{9.48}$$

which can be solved for $\mathbf{r}$ with the result

$$\mathbf{r} = \frac{c_1\mathbf{r}_1 + c_2\mathbf{r}_2 + c_3\mathbf{r}_3}{c_1 + c_2 + c_3} = \sum_{i=1}^{3} \xi_i \mathbf{r}_i \tag{9.49}$$

where

$$\xi_i = c_i \Big/ \sum_{j=1}^{3} c_j, \qquad i = 1, 2, 3 \tag{9.50}$$

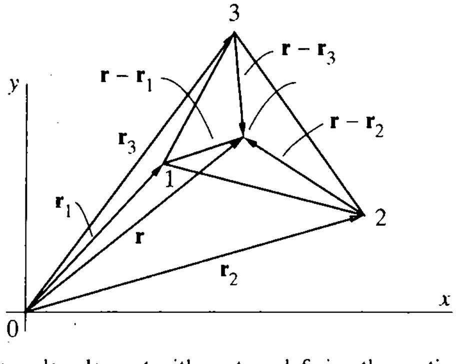


Figure 9.11 Triangular element with vectors defining the vertices and point $P(x, y)$

Equations (9.49) and (9.50) can be written in the more explicit form

$$\begin{aligned} x_1\xi_1 + x_2\xi_2 + x_3\xi_3 &= x \\ y_1\xi_1 + y_2\xi_2 + y_3\xi_3 &= y \\ \xi_1 + \xi_2 + \xi_3 &= 1 \end{aligned} \tag{9.51}$$

Equations (9.51) represent three equations in the unknowns ξ_1, ξ_2 and ξ_3 and have the solution

$$\begin{aligned} \xi_1(x,y) &= \frac{1}{\Delta}[x_2y_3 - x_3y_2 + (y_2 - y_3)x + (x_3 - x_2)y] \\ \xi_2(x,y) &= \frac{1}{\Delta}[x_3y_1 - x_1y_3 + (y_3 - y_1)x + (x_1 - x_3)y] \\ \xi_3(x,y) &= \frac{1}{\Delta}[x_1y_2 - x_2y_1 + (y_1 - y_2)x + (x_2 - x_1)y] \end{aligned} \tag{9.52}$$

where

$$\Delta = (x_2y_3 - x_3y_2) + (x_3y_1 - x_1y_3) + (x_1y_2 - x_2y_1) \tag{9.53}$$

The functions ξ_1, ξ_2 and ξ_3 possess some very interesting and useful properties. We observe that for $x = x_1,\ y = y_1$ Eqs. (9.52) yield $\xi_1 = 1,\ \xi_2 = \xi_3 = 0$, for $x = x_2,\ y = y_2$ they yield $\xi_1 = 0,\ \xi_2 = 1,\ \xi_3 = 0$, and for $x = x_3$, $y = y_3$ they give $\xi_1 = \xi_2 = 0,\ \xi_3 = 1$. It follows that the vertices (x_1, y_1), (x_2, y_2) and (x_3, y_3) of the triangle are defined by the triplets $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$, respectively. This suggests that the functions $\xi_i\,(i = 1, 2, 3)$ can be used as coordinates, as shown in Fig. 9.12. Indeed, they represent the local coordinates mentioned earlier in this section. For example, the edge 2-3 is defined as $\xi_1 = 0$ and the vertex 1 as $\xi_1 = 1$. A line parallel to the edge 2-3 is described by $\xi_1 = c =$ constant, where the constant c is proportional to the distance between 2-3 and the line in question.

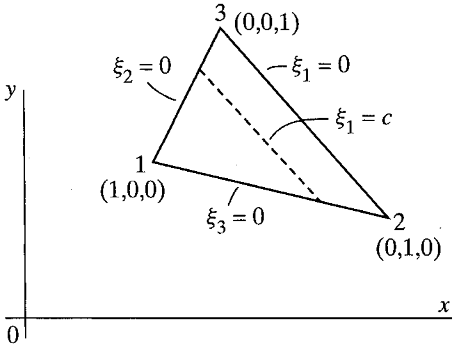


Figure 9.12 Local coordinates for triangular element

Similar geometric interpretations can be given to ξ_2 and ξ_3. The coordinates ξ_1, ξ_2 and ξ_3 can be identified as the *natural coordinates* for the triangular element. The natural coordinates ξ_1, ξ_2, ξ_3 can be given a different geometric interpretation by observing that $\Delta = 2A$, where A is the area of the triangular element. Letting $x_i = x,\ y_i = y\ (i = 1, 2, 3)$ in Eq. (9.53), in sequence, we obtain

$$\begin{aligned}(x_2y_3 - x_3y_2) + (x_3y - xy_3) + (xy_2 - x_2y) &= 2A_1\\ (xy_3 - x_3y) + (x_3y_1 - x_1y_3) + (x_1y - xy_1) &= 2A_2\\ (x_2y - xy_2) + (xy_1 - x_1y) + (x_1y_2 - x_2y_1) &= 2A_3\end{aligned} \tag{9.54}$$

where A_i is the area of the triangle formed by point $P(x, y)$ with the side opposite to vertex i $(i = 1, 2, 3)$, as shown in Fig. 9.13. Then, comparing Eqs. (9.52) and (9.54), we conclude that

$$\begin{aligned}\xi_1(x, y) &= \frac{1}{2A}[(y_2 - y_3)x + (x_3 - x_2)y + x_2y_3 - x_3y_2] = \frac{A_1}{A}\\ \xi_2(x, y) &= \frac{1}{2A}[(y_3 - y_1)x + (x_1 - x_3)y + x_3y_1 - x_1y_3] = \frac{A_2}{A}\\ \xi_3(x, y) &= \frac{1}{2A}[(y_1 - y_2)x + (x_2 - x_1)y + x_1y_2 - x_2y_1] = \frac{A_3}{A}\end{aligned} \tag{9.55}$$

In view of Eqs. (9.55), ξ_1, ξ_2 and ξ_3 are also called *area coordinates*.

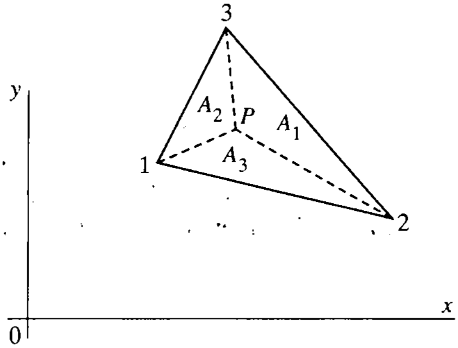


Figure 9.13 Area coordinates for triangular element

By analogy with the linear interpolation functions for one-dimensional domains, the natural coordinates ξ_1, ξ_2, ξ_3 can be used as linear elements for our membrane, the lowest-degree polynomials admissible. Letting

$$\phi_i = \xi_i, \qquad i = 1, 2, 3 \tag{9.56}$$

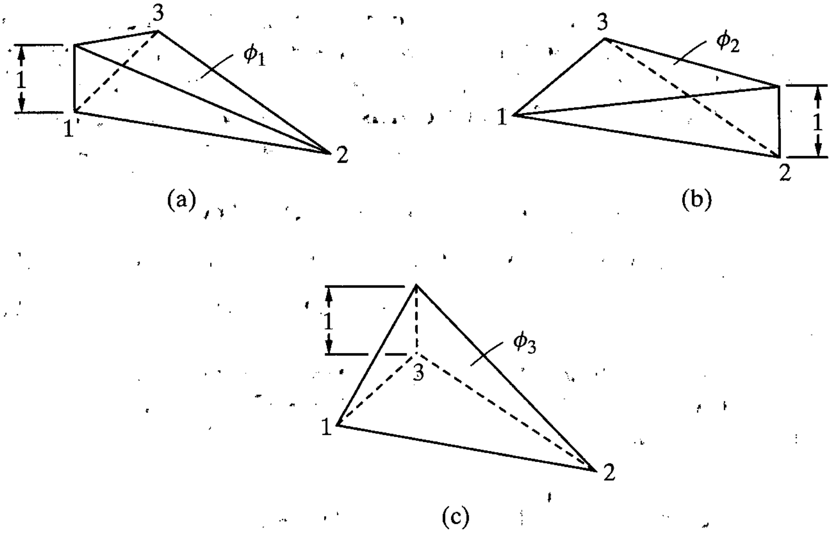


Figure 9.14 Linear interpolation functions for triangular element

and referring to Fig. 9.12, we can depict ϕ_1, ϕ_2 and ϕ_3 as the pyramid sections shown in Figs. 9.14a, b and c, respectively, so that they really represent linear interpolation functions for the triangular element. The analogy can be extended by representing the displacement at any point ξ_1, ξ_2, ξ_3 inside the triangle as a linear combination of the interpolation functions ϕ_1, ϕ_2 and ϕ_3 multiplied by the nodal displacements w_1, w_2 and w_3, respectively, or

$$w(\xi_1, \xi_2, \xi_3) = \sum_{i=1}^{3} \phi_i w_i = \boldsymbol{\phi}^T \mathbf{w} \tag{9.57}$$

where $\boldsymbol{\phi} = [\xi_1 \; \xi_2 \; \xi_3]^T$ and $\mathbf{w} = [w_1 \; w_2 \; w_3]^T$. The function $w(\xi_1, \xi_2, \xi_3)$ is displayed in Fig. 9.15. Regarding Fig. 9.15 as representing the displacement over a given element D_j, the displacement of the whole membrane can be expressed in the form

$$w(\xi_1, \xi_2, \xi_3) = \boldsymbol{\phi}^T \mathbf{w}_j \quad x, y \text{ in } D_j \tag{9.58}$$

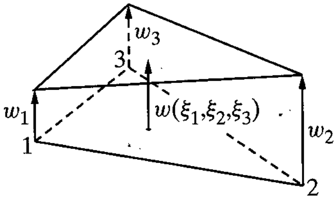


Figure 9.15 Membrane displacement in terms of linear interpolation functions

where $\mathbf{w}_j$ is the nodal vector corresponding to the jth finite element. Hence, by analogy with the piecewise linear string displacement profile of Fig. 9.1, the displacement of the membrane consists of a surface made up of flat triangular surfaces joined along the edges. Moreover, the components of the nodal vectors $\mathbf{w}_j$ represent the actual displacements of the membrane at the nodal points in question. It should be remarked here that this solution of the membrane vibration problem is similar to Courant's solution of the plane torsion problem (Ref. 8).

As suggested earlier in this section, we can write the energy inner product in the form

$$[w, w] = \sum_{j=1}^{n} [w, w]_j \tag{9.59}$$

where, using results from Sec. 7.12 and assuming that the tension T is constant,

$$[w, w]_j = T \int_{A_j} \left[\left(\frac{\partial w}{\partial x} \right)^2 + \left(\frac{\partial w}{\partial y} \right)^2 \right] dA_j \tag{9.60}$$

in which A_j is the area of the jth element. Moreover, for constant mass density ρ, the weighted inner product can be written as

$$\left(\sqrt{m}w, \sqrt{m}w\right) = \sum_{j=1}^{n} \left(\sqrt{m}w, \sqrt{m}w\right)_j \tag{9.61}$$

where

$$\left(\sqrt{m}w, \sqrt{m}w\right)_j = \rho \int_{A_j} w^2 dA_j \tag{9.62}$$

To evaluate the integrals in Eqs. (9.60) and (9.62), it is necessary to carry out a transformation from rectangular coordinates to natural coordinates. Hence, using Eqs. (9.55) and (9.58), we can write

$$\begin{aligned} \frac{\partial w}{\partial x} &= \frac{\partial w}{\partial \xi_1}\frac{d\xi_1}{\partial x} + \frac{\partial w}{\partial \xi_2}\frac{\partial \xi_2}{\partial x} + \frac{\partial w}{\partial \xi_3}\frac{\partial \xi_3}{\partial x} \\ &= \frac{1}{2A_j}\left[\frac{\partial \boldsymbol{\phi}^T}{\partial \xi_1}(y_2 - y_3) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_2}(y_3 - y_1) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_3}(y_1 - y_2) \right] \mathbf{w}_j \\ &= \frac{1}{2A_j}[y_2 - y_3 \quad y_3 - y_1 \quad y_1 - y_2]\,\mathbf{w}_j \end{aligned} \tag{9.63a}$$

$$\begin{aligned} \frac{\partial w}{\partial y} &= \frac{\partial w}{\partial \xi_1}\frac{\partial \xi_1}{\partial y} + \frac{\partial w}{\partial \xi_2}\frac{\partial \xi_2}{\partial y} + \frac{\partial w}{\partial \xi_3}\frac{\partial \xi_3}{\partial y} \\ &= \frac{1}{2A_j}\left[\frac{\partial \boldsymbol{\phi}^T}{\partial \xi_1}(x_3 - x_2) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_2}(x_1 - x_3) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_3}(x_2 - x_1) \right] \mathbf{w}_j \\ &= \frac{1}{2A_j}[x_3 - x_2 \quad x_1 - x_3 \quad x_2 - x_1]\,\mathbf{w}_j \end{aligned} \tag{9.63b}$$

so that, introducing Eqs. (9.63) into Eq. (9.60), we obtain

$$[w, w]_j = \mathbf{w}_j^T K^{(j)} \mathbf{w}_j \tag{9.64}$$

where

$$\begin{aligned} K^{(j)} &= \frac{T}{4A_j^2} \int_{A_j} \left\{ \begin{bmatrix} y_2 - y_3 \\ y_3 - y_1 \\ y_1 - y_2 \end{bmatrix} \begin{bmatrix} y_2 - y_3 \\ y_3 - y_1 \\ y_1 - y_2 \end{bmatrix}^T + \begin{bmatrix} x_3 - x_2 \\ x_1 - x_3 \\ x_2 - x_1 \end{bmatrix} \begin{bmatrix} x_3 - x_2 \\ x_1 - x_3 \\ x_2 - x_1 \end{bmatrix}^T \right\} dA_j \\ &= \frac{T}{4A_j} \begin{bmatrix} (y_2 - y_3)^2 + (x_3 - x_2)^2 & (y_2 - y_3)(y_3 - y_1) + (x_3 - x_2)(x_1 - x_3) & (y_2 - y_3)(y_1 - y_2) + (x_3 - x_2)(x_2 - x_1) \\ & (y_3 - y_1)^2 + (x_1 - x_3)^2 & (y_3 - y_1)(y_1 - y_2) + (x_1 - x_3)(x_2 - x_1) \\ \text{symm} & & (y_1 - y_2)^2 + (x_2 - x_1)^2 \end{bmatrix} \end{aligned} \tag{9.65}$$

is the *element stiffness matrix*, and we note that the integral was trivially evaluated because the integrand did not depend on the natural coordinates. On the other hand, the integral in the element mass matrix does involve the natural coordinates, and its evaluation can be rendered routine by means of the formula (Ref. 11)

$$\int_{A_j} \xi_1^m \xi_2^n \xi_3^p \, dA_j = \frac{m!n!p!}{(m + n + p + 2)!} 2A_j \tag{9.66}$$

Inserting Eq. (9.58) into Eq. (9.62), we have

$$(\sqrt{m}w, \sqrt{m}w)_j = \mathbf{w}_j^T M^{(j)} \mathbf{w}_j \tag{9.67}$$

where $M^{(j)}$ is the *element mass matrix*, which, upon using Eq. (9.66), can be written in the explicit form

$$\begin{aligned} M^{(j)} &= \rho \int_{A_j} \boldsymbol{\phi}\boldsymbol{\phi}^T dA_j = \rho \int_{A_j} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{bmatrix}^T dA_j \\ &= \rho \int_{A_j} \begin{bmatrix} \xi_1^2 & \xi_1\xi_2 & \xi_1\xi_3 \\ \xi_1\xi_2 & \xi_2^2 & \xi_2\xi_3 \\ \xi_1\xi_3 & \xi_2\xi_3 & \xi_3^2 \end{bmatrix} dA_j \\ &= \frac{\rho A_j}{12} \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \end{aligned} \tag{9.68}$$

The general ideas behind the assembly process for two-dimensional domains are essentially the same as for one-dimensional domains, but the details are more involved, which can be attributed to the fact that there is no longer a simple correspondence between the node number and element number. To introduce the ideas, we consider part of a uniform membrane consisting of four triangular elements, as shown in Fig. 9.16, in which the encircled numbers represent the element number, the outside numbers the global node number and the smaller size inside numbers the local node numbers for each element. The membrane is free on all sides. Inserting the coordinates corresponding to the local node numbers into Eq. (9.65), we obtain two types of element stiffness matrices, as follows:

$$K^{(j)} = \frac{T}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \qquad j = 1, 3$$
$$K^{(j)} = \frac{T}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \qquad j = 2, 4 \tag{9.69}$$

where we note that, to make the assembly process easier to implement, we identify the element number by a superscript. On the other hand, letting $A_j = h^2/2$ in Eq. (9.68), we obtain the element mass matrices

$$M^{(j)} = \frac{mh^2}{24}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \qquad j = 1, 2, 3, 4 \tag{9.70}$$

Because the motion of the membrane is defined by the global nodal displacements and the entries of the element stiffness and mass matrices correspond to local nodal displacements, it is necessary to develop a scheme for placing the element entries in the proper position in the global matrices. To this end, we define the *connectivity array* $C = [c_{jk}]$, where the subscript j identifies the element number and the subscript k the global node numbers listed in the order specified by the local nodes. The connectivity array for the system of Fig. 9.16 is simply

$$C = \begin{bmatrix} 1 & 4 & 5 \\ 1 & 5 & 2 \\ 2 & 5 & 6 \\ 2 & 6 & 3 \end{bmatrix} \tag{9.71}$$

As an illustration of the use of the connectivity array, we construct the 6 × 6 global stiffness matrix for the system of Fig. 9.16. The first row in C represents the first element and it instructs us to place the entries $k_{11}^{(1)}$, $k_{12}^{(1)}$, $k_{13}^{(1)}$, $k_{22}^{(1)}$, $k_{23}^{(1)}$ and $k_{33}^{(1)}$ of the element stiffness matrix $K^{(1)}$ in the positions (1, 1), (1, 4), (1, 5), (4, 4), (4, 5) and (5, 5) of the global stiffness matrix K. Of course, placement of the symmetric entries is implied. Repeating the process for the remaining three rows of C, we obtain the global stiffness matrix

$$K = \begin{bmatrix} k_{11}^{(1)} + k_{11}^{(2)} & k_{13}^{(2)} & 0 & k_{12}^{(1)} & k_{13}^{(1)} + k_{12}^{(2)} & 0 \\ & k_{33}^{(2)} + k_{11}^{(3)} + k_{11}^{(4)} & k_{13}^{(4)} & 0 & k_{23}^{(2)} + k_{12}^{(3)} & k_{13}^{(3)} + k_{12}^{(4)} \\ \text{symm} & & k_{33}^{(4)} & 0 & 0 & k_{23}^{(4)} \\ & & & k_{22}^{(1)} & k_{23}^{(1)} & 0 \\ & & & & k_{33}^{(1)} + k_{22}^{(2)} + k_{22}^{(3)} & k_{23}^{(3)} \\ & & & & & k_{33}^{(3)} + k_{22}^{(4)} \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -1 & 0 & -1 & 0 & 0 \\ & 4 & -1 & 0 & -2 & 0 \\ & & 2 & 0 & 0 & -1 \\ & \text{symm} & & 2 & -1 & 0 \\ & & & & 4 & -1 \\ & & & & & 2 \end{bmatrix} \tag{9.72}$$

The same process applies to the construction of the global mass matrix.

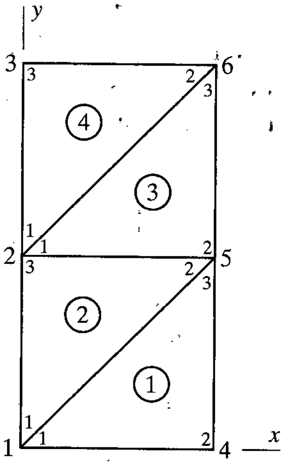


Figure 9.16 Four triangular elements with numbering scheme

The generation of the finite element mesh amounts to dividing the domain D into triangular domains D_j in such a way that no vertex of one triangle lies on the edge of another triangle. It is common practice to begin with a coarse mesh and refine the mesh so as to improve the accuracy of the eigensolutions. A mesh refinement guaranteeing that no vertex of one triangle lies on the edge of another consists of dividing each triangle into four similar triangles obtained by joining the midpoints of the edges. With each refinement of the mesh, a check verifying that a reduction in h leads indeed to a suitable reduction in the approximate eigenvalues is highly desirable.

As an alternative to refining the finite element mesh, the accuracy can be improved by refining the elements. This implies higher-degree polynomials, such as quadratic, cubic, etc. The various polynomials can be arranged in the so-called *Pascal's triangle* displayed in Fig. 9.17. Of course, the top two rows represent the linear polynomial, the top three the quadratic, etc. We consider first the *quadratic elements*

$$w(x, y) = a_1 + a_2 x + a_3 y + a_4 x^2 + a_5 xy + a_6 y^2 \tag{9.73}$$

Because there are now six coefficients, we need six nodes. We choose three nodes at the vertices of the triangle and three at the midpoints of the edges, as shown in Fig. 9.18.

1	constant	
$x \quad y$	linear	3 terms
$x^2 \quad xy \quad y^2$	quadratic	6 terms
$x^3 \quad x^2y \quad xy^2 \quad y^3$	cubic	10 terms
$x^4 \quad x^3y \quad x^2y^2 \quad xy^3 \quad y^4$	quartic	15 terms
....	....	....

Figure 9.17 Pascal's triangle

As in the case of linear elements, it is convenient to use local, natural coordinates in the form of the area coordinates ξ_1, ξ_2, ξ_3, instead of global rectangular coordinates. Recalling that the natural coordinates satisfy the relation $\xi_1+\xi_2+\xi_3 = 1$, only two of the coordinates should be regarded as independent. It does not matter which of the three coordinates are chosen as the independent ones, as the final result is the same. Hence, we choose ξ_1 and ξ_2 as the independent coordinates and express the interpolation functions in the form

$$\phi_i(\xi_1, \xi_2\, \xi_3) = c_{i1} + c_{i2}\xi_1 + c_{i3}\xi_2 + c_{i4}\xi_1^2 + c_{i5}\xi_1\xi_2 + c_{i6}\xi_2^2, \qquad i = 1, 2, \ldots, 6 \tag{9.74}$$

where the dependence on ξ_3 is only implicit. Equations (9.74) require six conditions on each interpolation function, which can be generated by considering the six nodes shown in Fig. 9.18 and insisting that each ϕ_i be equal to 1 at node i and equal to zero at the remaining five nodes. Then, if we denote the values of ξ_1 and ξ_2 at the node k by $\xi_{1,k}$ and $\xi_{2,k}$, respectively, we can extend the procedure given by Eqs. (9.11)–(9.14) to the case of two variables and write

$$A = \begin{bmatrix} 1 & \xi_{1,1} & \xi_{2,1} & \xi_{1,1}^2 & \xi_{1,1}\xi_{2,1} & \xi_{2,1}^2 \\ 1 & \xi_{1,2} & \xi_{2,2} & \xi_{1,2}^2 & \xi_{1,2}\xi_{2,2} & \xi_{2,2}^2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & \xi_{1,6} & \xi_{2,6} & \xi_{1,6}^2 & \xi_{1,6}\xi_{2,6} & \xi_{2,6}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1/2 & 1/2 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 1/2 & 0 & 0 & 1/4 \\ 1 & 1/2 & 0 & 1/4 & 0 & 0 \end{bmatrix} \tag{9.75}$$

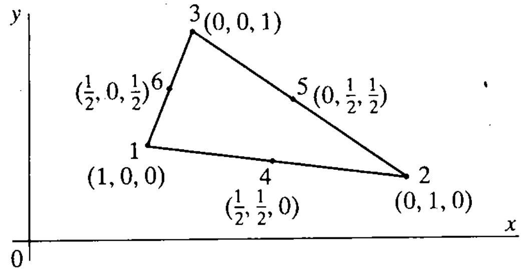

y
3(0, 0, 1)
$(\frac{1}{2}, 0, \frac{1}{2})$6
5$(0, \frac{1}{2}, \frac{1}{2})$
1
(1, 0, 0)
4
$(\frac{1}{2}, \frac{1}{2}, 0)$
2
(0, 1, 0)
x
0

Figure 9.18 Triangular element with six nodes

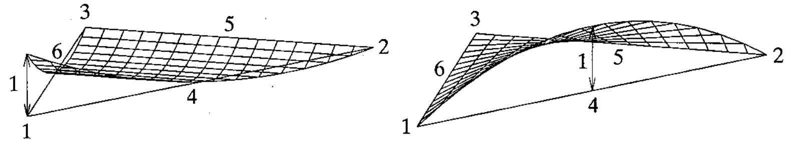
so that

$$[\mathbf{c}_1\ \mathbf{c}_2\ \dots\ \mathbf{c}_6] = A^{-1} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & -3 & 0 & 0 & 4 \\ 0 & -1 & -3 & 0 & 4 & 0 \\ 2 & 0 & 2 & 0 & 0 & -4 \\ 0 & 0 & 4 & 4 & -4 & -4 \\ 0 & 2 & 2 & 0 & -4 & 0 \end{bmatrix} \tag{9.76}$$

Inserting the elements in the columns of Eq. (9.76) into Eqs. (9.74) and recalling that $\xi_1 + \xi_2 + \xi_3 = 1$, we obtain the interpolation functions

$$\begin{aligned} \phi_1 &= -\xi_1 + 2\xi_1^2 = \xi_1(2\xi_1 - 1) \\ \phi_2 &= -\xi_2 + 2\xi_2^2 = \xi_2(2\xi_2 - 1) \\ \phi_3 &= 1 - 3\xi_1 - 3\xi_2 + 2\xi_1^2 + 4\xi_1\xi_2 + 2\xi_2^2 = \xi_3(2\xi_3 - 1) \\ \phi_4 &= 4\xi_1\xi_2 \\ \phi_5 &= 4\xi_2 - 4\xi_1\xi_2 - 4\xi_2^2 = 4\xi_2\xi_3 \\ \phi_6 &= 4\xi_1 - 4\xi_1^2 - 4\xi_1\xi_2 = 4\xi_1\xi_3 \end{aligned} \tag{9.77}$$

The functions ϕ_1 and ϕ_4 are displayed in Figs. 9.19a and 9.19b, respectively. The functions ϕ_2 and ϕ_3 are similar to ϕ_1, except that the unit displacement is at the nodes 2 and 3, respectively, and the same statement can be made concerning the similarity between the functions ϕ_5 and ϕ_6 and the function ϕ_4.



(a) (b)

Figure 9.19 Quadratic interpolation functions

Cubic elements have the form

$$\begin{aligned} \phi_i(\xi_1, \xi_2, \xi_3) = {} & c_{i1} + c_{i2}\xi_1 + c_{i3}\xi_2 + c_{i4}\xi_1^2 + c_{i5}\xi_1\xi_2 + c_{i6}\xi_2^2 + c_{i7}\xi_1^3 \\ & + c_{i8}\xi_1^2\xi_2 + c_{i9}\xi_1\xi_2^2 + c_{i10}\xi_2^3, \qquad i = 1, 2, \dots, 10 \end{aligned} \tag{9.78}$$

so that the triangular element must have ten nodes. An element satisfying this requirement is depicted in Fig. 9.20. Using the same process as for quadratic elements, the interpolation functions can be shown to be (Prob. 9.17)

$$
\begin{aligned}
\phi_i &= \frac{1}{2}\xi_i(3\xi_i - 1)(3\xi_i - 2), \qquad i = 1, 2, 3 \\
\phi_4 &= \frac{9}{2}\xi_2\xi_1(3\xi_1 - 1), \qquad \phi_5 = \frac{9}{2}\xi_1\xi_2(3\xi_2 - 1), \qquad \phi_6 = \frac{9}{2}\xi_3\xi_2(3\xi_2 - 1) \\
\phi_7 &= \frac{9}{2}\xi_2\xi_3(3\xi_3 - 1), \qquad \phi_8 = \frac{9}{2}\xi_1\xi_3(3\xi_3 - 1), \qquad \phi_9 = \frac{9}{2}\xi_3\xi_1(3\xi_1 - 1) \\
\phi_{10} &= 27\xi_1\xi_2\xi_3
\end{aligned}
\tag{9.79}
$$

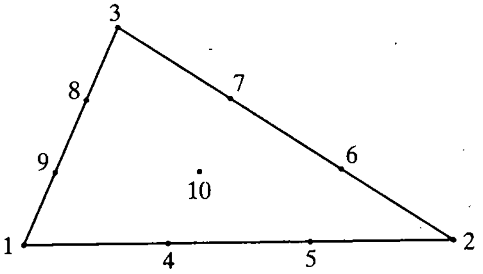


Figure 9.20 Triangular element with ten nodes

Example 9.4

Derive the global stiffness and mass matrices for a $2h \times 3h$ rectangular membrane. The membrane is free on all sides.

We divide the membrane into 12 triangular elements, as shown in Fig. 9.21, and note that the membrane of Fig. 9.16 represents a mere one third of that of Fig. 9.21. In fact, the numbering of the elements, global and local nodes in Fig. 9.21 is entirely consistent with that of Fig. 9.16. It can be verified that the element stiffness matrices are still of two types, as in Eqs. (9.69), or

$$
K^{(j)} = \frac{T}{2}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, \quad j \text{ odd}, \qquad K^{(j)} = \frac{T}{2}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 2 \end{bmatrix}, \quad j \text{ even} \tag{a}
$$

Moreover, from Eqs. (9.70), the element stiffness matrices are

$$
M^{(j)} = \frac{mh^2}{24}\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \qquad j = 1, 2, \ldots, 12 \tag{b}
$$

From Fig. 9.21, the connectivity array is

$$
C = \begin{bmatrix} 1 & 1 & 2 & 2 & 4 & 4 & 5 & 5 & 7 & 7 & 8 & 8 \\ 4 & 5 & 5 & 6 & 7 & 8 & 8 & 9 & 10 & 11 & 11 & 12 \\ 5 & 2 & 6 & 3 & 8 & 5 & 9 & 6 & 11 & 8 & 12 & 9 \end{bmatrix}^T \tag{c}
$$

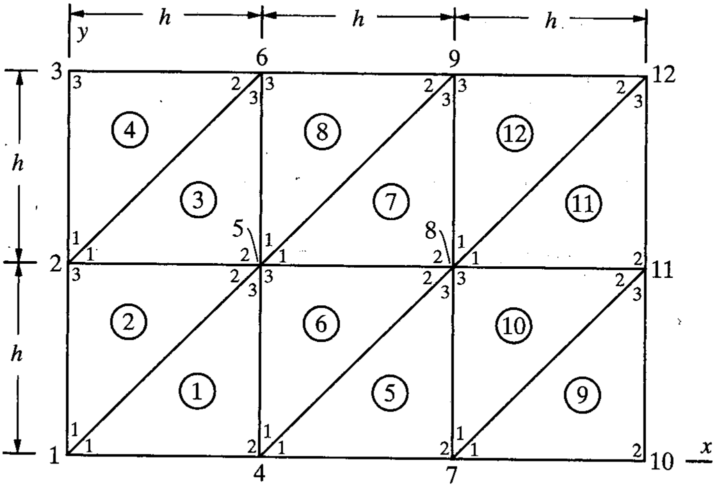


Figure 9.21 Membrane with 12 triangular elements

Hence, following the asembly procedure established in this section, we obtain the global stiffness matrix

$$K = \frac{T}{2}\begin{bmatrix} 2 & -1 & 0 & -1 & & & & & & & & \\ & 4 & -1 & 0 & -2 & & & & & & & \\ & & 2 & 0 & 0 & -1 & & & & & & \\ & & & 4 & -2 & 0 & -1 & & & & & \\ & & & & 8 & -2 & 0 & -2 & & & & \\ & & & & & 4 & 0 & 0 & -1 & & & \\ & & & & & & 4 & -2 & 0 & -1 & & \\ & & & \text{symm} & & & & 8 & -2 & 0 & -2 & \\ & & & & & & & & 4 & 0 & 0 & -1 \\ & & & & & & & & & 2 & -1 & 0 \\ & & & & & & & & & & 4 & -1 \\ & & & & & & & & & & & 2 \end{bmatrix} \tag{d}$$

and global mass matrix

$$M = \frac{mh^2}{24}\begin{bmatrix} 4 & 1 & 0 & 1 & 2 & & & & & & & \\ & 4 & 1 & 0 & 2 & 0 & & & & & & \\ & & 4 & 0 & 2 & 1 & 0 & & & & & \\ & & & 6 & 2 & 0 & 1 & 2 & & & & \\ & & & & 12 & 2 & 0 & 2 & 0 & & & \\ & & & & & 6 & 0 & 2 & 1 & 0 & & \\ & & & & & & 6 & 2 & 0 & 1 & 2 & \\ & & & & & & & 12 & 2 & 0 & 2 & 0 \\ & & & \text{symm} & & & & & 6 & 0 & 2 & 1 \\ & & & & & & & & & 2 & 1 & 0 \\ & & & & & & & & & & 8 & 1 \\ & & & & & & & & & & & 2 \end{bmatrix} \tag{e}$$

where the elements of K and M not listed are all zero. It is easy to verify that the stiffness matrix is singular, which is consistent with the fact that all sides of the membrane are free.

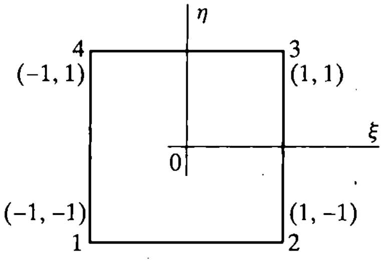


Figure 9.22 Rectangular element with natural coordinates

Both K and M are banded with half-bandwidth equal to three and four, respectively. For banded matrices, it is more efficient to list the entries by the nonzero diagonals.

9.6 RECTANGULAR ELEMENTS

When the membrane boundary possesses corners, *quadrilateral elements* may be able to yield a more economical finite element mesh. In this section, we consider the simplest quadrilateral elements, namely, *rectangular elements*. Figure 9.22 shows a typical rectangular element together with the corresponding natural coordinates ξ, η. Regarding the four corners as nodal points, the lowest-degree polynomials admissible are the *bilinear*, given by

$$\phi_i(\xi, \eta) = c_{i1} + c_{i2}\xi + c_{i3}\eta + c_{i4}\xi\eta, \qquad i = 1, 2, 3, 4 \tag{9.80}$$

Hence, following the standard procedure, we obtain the bilinear interpolation functions (Problem 9.19)

$$\begin{aligned} \phi_1 &= \frac{1}{4}(1-\xi)(1-\eta), & \phi_2 &= \frac{1}{4}(1+\xi)(1-\eta) \\ \phi_3 &= \frac{1}{4}(1+\xi)(1+\eta), & \phi_4 &= \frac{1}{4}(1-\xi)(1+\eta) \end{aligned} \tag{9.81}$$

The function ϕ_1 is shown in Fig. 9.23. The other three functions are relatively easy to visualize.

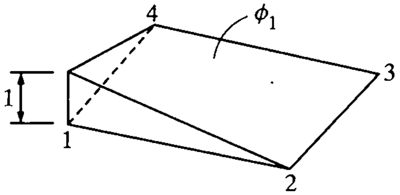


Figure 9.23 Bilinear interpolation function for rectangular element

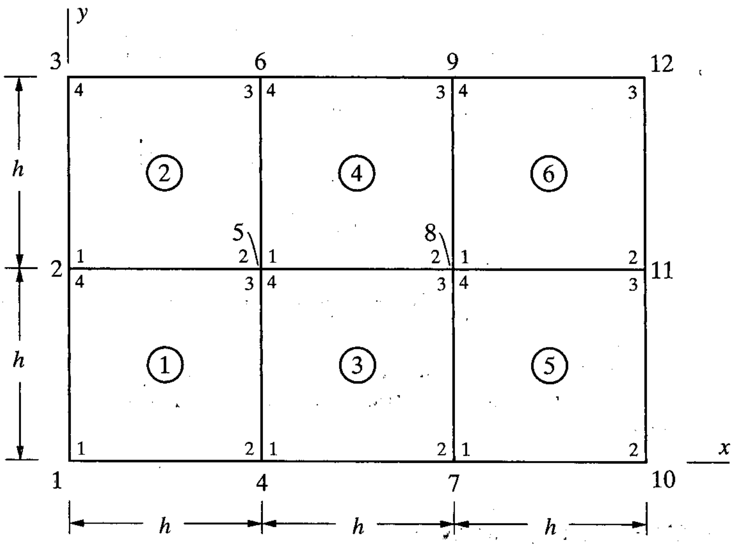


Figure 9.24 Membrane with six rectangular elements

Next, we derive the stiffness and mass matrices for the membrane of Example 9.4 by means of the bilinear elements. We begin with the derivation of the element stiffness and mass matrices. To this end, we express the displacement of the membrane at any point of the rectangle in the form

$$w(\xi, \eta) = \boldsymbol{\phi}^T \mathbf{w}_j \tag{9.82}$$

where $\boldsymbol{\phi} = [\phi_1\ \phi_2\ \phi_3\ \phi_4]^T$ is a four-dimensional vector of interpolation functions and $\mathbf{w}_j = [w_1\ w_2\ w_3\ w_4]^T$ is the corresponding nodal vector for element j. We recall from Sec. 9.5 that the stiffness matrix involves the terms $\partial w/\partial x$ and $\partial w/\partial y$. Because w is in terms of the natural coordinates ξ, η, we must first write the relations between the rectangular coordinates x, y and the local, natural coordinates ξ, η for each element. From Fig. 9.24, we have simply

$$\begin{aligned} x &= \frac{h}{2}(j + \xi), \qquad y = \frac{h}{2}(1 + \eta), \qquad j = 1, 3, 5 \\ x &= \frac{h}{2}(j - 1 + \xi), \qquad y = \frac{h}{2}(3 + \eta), \qquad j = 2, 4, 6 \end{aligned} \tag{9.83}$$

Hence, using Eqs. (9.81)–(9.83), we can write

$$\begin{aligned} \frac{\partial w}{\partial x} &= \frac{\partial \xi}{\partial x}\frac{\partial w}{\partial \xi} = \frac{2}{h}\frac{\partial \boldsymbol{\phi}^T}{\partial \xi}\mathbf{w}_j \\ &= \frac{1}{2h}[-(1-\eta) \quad 1-\eta \quad 1+\eta \quad -(1+\eta)]^T\,\mathbf{w}_j \end{aligned} \tag{9.84a}$$

$$\frac{\partial w}{\partial y} = \frac{\partial \eta}{\partial y}\frac{\partial w}{\partial \eta} = \frac{2}{h}\frac{\partial \boldsymbol{\phi}^T}{\partial \eta}\mathbf{w}_j$$

$$= \frac{1}{2h}[-(1-\xi) \quad -(1+\xi) \quad 1+\xi \quad 1-\xi]^T \mathbf{w}_j \tag{9.84b}$$

Introducing Eqs. (9.84) into Eq. (9.60) and recalling Eq. (9.64), we obtain the element stiffness matrices

$$K^{(j)} = \frac{T}{4h^2}\int_{A_j}\left(\frac{\partial\boldsymbol{\phi}}{\partial\xi}\frac{\partial\boldsymbol{\phi}^T}{\partial\xi} + \frac{\partial\boldsymbol{\phi}}{\partial\eta}\frac{\partial\boldsymbol{\phi}^T}{d\eta}\right)dA_j$$

$$= \frac{T}{16}\int_{-1}^{1}\int_{-1}^{1}\left\{\begin{bmatrix}-(1-\eta)\\ 1-\eta\\ 1+\eta\\ -(1+\eta)\end{bmatrix}\begin{bmatrix}-(1-\eta)\\ 1-\eta\\ 1+\eta\\ -(1+\eta)\end{bmatrix}^T\right.$$

$$\left. + \begin{bmatrix}-(1-\xi)\\ -(1+\xi)\\ 1+\xi\\ 1-\xi\end{bmatrix}\begin{bmatrix}-(1-\xi)\\ -(1+\xi)\\ 1+\xi\\ 1-\xi\end{bmatrix}^T\right\} d\xi\, d\eta$$

$$= \frac{T}{6}\begin{bmatrix}4 & -1 & -2 & -1\\ & 4 & -1 & -2\\ \text{symm} & & 4 & -1\\ & & & 4\end{bmatrix} \tag{9.85}$$

Moreover, inserting Eqs. (9.81)–(9.83) into Eq. (9.62) and recalling Eq. (9.67), we obtain the element mass matrix

$$M^{(j)} = \rho\int_{A_j}\boldsymbol{\phi}\boldsymbol{\phi}^T dA_j$$

$$= \frac{\rho h^2}{64}\int_{-1}^{1}\int_{-1}^{1}\begin{bmatrix}(1-\xi)(1-\eta)\\ (1+\xi)(1-\eta)\\ (1+\xi)(1+\eta)\\ (1-\xi)(1+\eta)\end{bmatrix}\begin{bmatrix}(1-\xi)(1-\eta)\\ (1+\xi)(1-\eta)\\ (1+\xi)(1+\eta)\\ (1-\xi)(1+\eta)\end{bmatrix}^T d\xi\, d\eta$$

$$= \frac{\rho h^2}{36}\begin{bmatrix}4 & 2 & 1 & 2\\ & 4 & 2 & 1\\ \text{symm} & & 4 & 2\\ & & & 4\end{bmatrix} \tag{9.86}$$

The assembly process is the same as that for triangular elements described in Sec. 9.5. Hence, using the numbering scheme of Fig. 9.24, we obtain the connectivity array

$$C = \begin{bmatrix}1 & 2 & 4 & 5 & 7 & 8\\ 4 & 5 & 7 & 8 & 10 & 11\\ 5 & 6 & 8 & 9 & 11 & 12\\ 2 & 3 & 5 & 6 & 8 & 9\end{bmatrix} \tag{9.87}$$

which permits us to derive the global stiffness matrix

$$K = \frac{T}{6}\left[\begin{array}{cccccccccccc}
4 & -1 & 0 & -1 & -2 & & & & & & & \\
& 8 & -1 & -2 & -2 & -2 & & & & & & \\
& & 4 & 0 & -2 & -1 & 0 & & & & & \\
& & & 8 & -2 & 0 & -1 & -2 & & & & \\
& & & & 16 & -2 & -2 & -2 & -2 & & & \\
& & & & & 8 & 0 & -2 & -1 & 0 & & \\
& & & & & & 8 & -2 & 0 & -1 & -2 & \\
& & & \text{symm} & & & & 16 & -2 & -2 & -2 & -2 \\
& & & & & & & & 8 & 0 & -2 & -1 \\
& & & & & & & & & 4 & -1 & 0 \\
& & & & & & & & & & 8 & -1 \\
& & & & & & & & & & & 4
\end{array}\right] \tag{9.88}$$

and the global mass matrix

$$M = \frac{\rho h^2}{36}\left[\begin{array}{cccccccccccc}
4 & 2 & 0 & 2 & 1 & & & & & & & \\
& 8 & 2 & 1 & 4 & 1 & & & & & & \\
& & 4 & 0 & 1 & 2 & 0 & & & & & \\
& & & 8 & 4 & 0 & 2 & 1 & & & & \\
& & & & 16 & 4 & 1 & 4 & 1 & & & \\
& & & & & 8 & 0 & 1 & 2 & 0 & & \\
& & & & & & 8 & 4 & 0 & 2 & 1 & \\
& & & \text{symm} & & & & 16 & 4 & 1 & 4 & 1 \\
& & & & & & & & 8 & 0 & 1 & 2 \\
& & & & & & & & & 4 & 2 & 0 \\
& & & & & & & & & & 8 & 2 \\
& & & & & & & & & & & 4
\end{array}\right] \tag{9.89}$$

Comparing the stiffness matrix computed in Example 9.4 by means of triangular elements, Eq. (d), with that computed here using rectangular elements, Eq. (9.88), we conclude that the first has half-bandwidth equal to three and the second has half-bandwidth equal to four. Balancing this relatively minor disadvantage is the fact that the latter matrix was easier to compute. The mass matrices in both cases have half-bandwidth equal to four and the amount of work required for deriving the matrices is about the same.

Quadratic rectangular elements are given by

$$\phi_i(\xi,\eta) = c_{i1} + c_{i2}\xi + c_{i3}\eta + c_{i4}\xi\eta + c_{i5}\xi^2 + c_{i6}\eta^2 + c_{i7}\xi^2\eta + c_{i8}\xi\eta^2 + c_{i9}\xi^2\eta^2, \quad i = 1,2,\ldots,9 \tag{9.90}$$

so that they require nine nodes. A rectangular element satisfying this requirement is shown in Fig. 9.25. Hence, following the usual approach, the quadratic interpolation

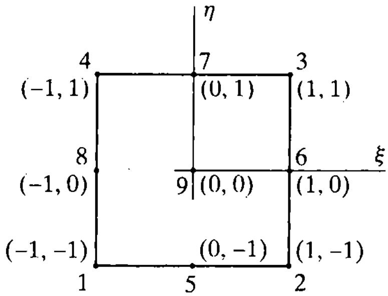


Figure 9.25 Rectangular element with nine nodes

functions can be shown to have the form

$$\begin{aligned}
\phi_1 &= \frac{1}{4}\xi\eta(1-\xi)(1-\eta), & \phi_2 &= -\frac{1}{4}\xi\eta(1+\xi)(1-\eta) \\
\phi_3 &= \frac{1}{4}\xi\eta(1+\xi)(1+\eta), & \phi_4 &= -\frac{1}{4}\xi\eta(1-\xi)(1+\eta) \\
\phi_5 &= -\frac{1}{2}\eta(1-\xi^2)(1-\eta), & \phi_6 &= \frac{1}{2}\xi(1+\xi)(1-\eta^2) \\
\phi_7 &= \frac{1}{2}\eta(1-\xi^2)(1+\eta), & \phi_8 &= -\frac{1}{2}\xi(1-\xi)(1-\eta^2) \\
\phi_9 &= (1-\xi^2)(1-\eta^2)
\end{aligned} \tag{9.91}$$

The quadratic interpolation functions ϕ_1, ϕ_5 and ϕ_9 are shown in Figs. 9.26a,b and c, respectively.

The quadratic interpolation functions given by Eqs. (9.91) are characterized by eight external nodes and one internal node. Because internal nodes do not contribute to the interelement connectivity, their usefulness has come into question. A family of rectangular elements known as *serendipity elements* contains only external nodes (Ref. 9). Of course, the four-node element was already examined earlier in this section. The eight-node element is obtained by simply omitting the ninth node from Fig. 9.25 and the last term in Eq. (9.90). Following the same procedure as above, we obtain the interpolation functions

$$\begin{aligned}
\phi_1 &= -\frac{1}{4}(1-\xi)(1-\eta)(1+\xi+\eta), & \phi_2 &= -\frac{1}{4}(1+\xi)(1-\eta)(1-\xi+\eta) \\
\phi_3 &= -\frac{1}{4}(1+\xi)(1+\eta)(1-\xi-\eta), & \phi_4 &= -\frac{1}{4}(1-\xi)(1+\eta)(1+\xi-\eta) \\
\phi_5 &= \frac{1}{2}(1-\xi^2)(1-\eta), & \phi_6 &= \frac{1}{2}(1+\xi)(1-\eta^2) \\
\phi_7 &= \frac{1}{2}(1-\xi^2)(1+\eta), & \phi_8 &= \frac{1}{2}(1-\xi)(1-\eta^2)
\end{aligned} \tag{9.92}$$

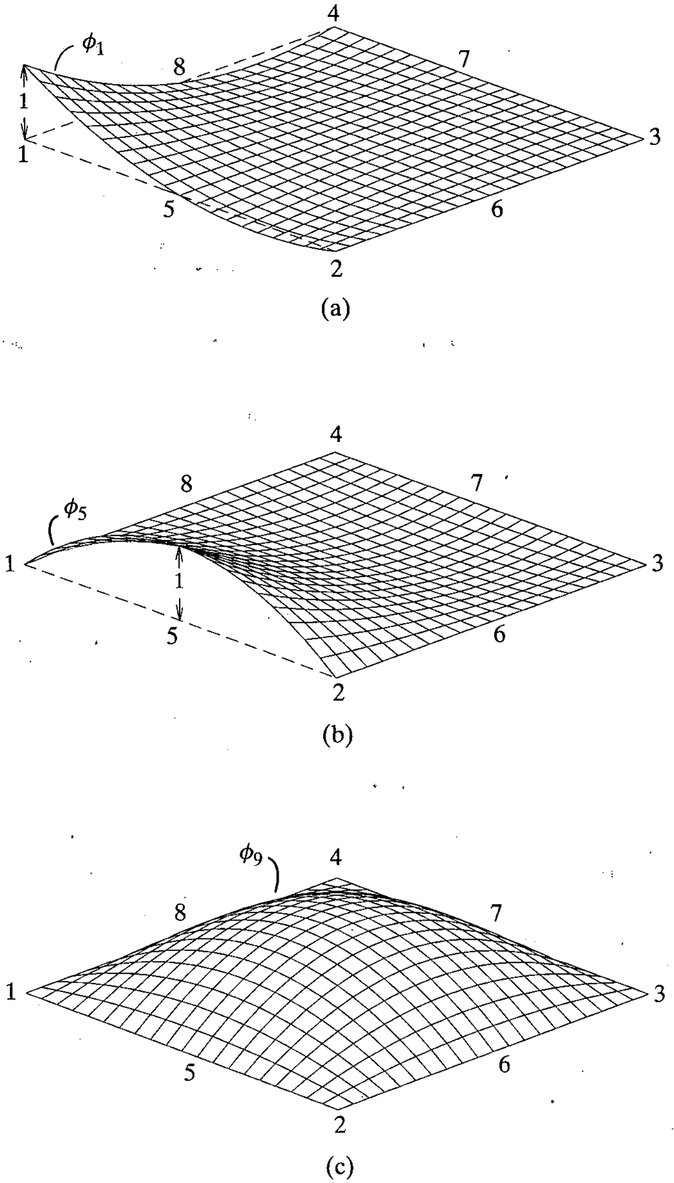


Figure 9.26 Quadratic interpolation functions for rectangular elements

Figures 9.27a and 9.27b show the eight-node interpolation functions ϕ_1 and ϕ_5, respectively.

Comparing Figs. 9.26 a and 9.27a, we conclude that there is not much difference between the functions ϕ_1 with eight nodes and nine nodes. On the other hand, from Figs. 9.26b and 9.27b, we see that there is significant difference between the functions ϕ_5 with eight nodes and nine nodes.

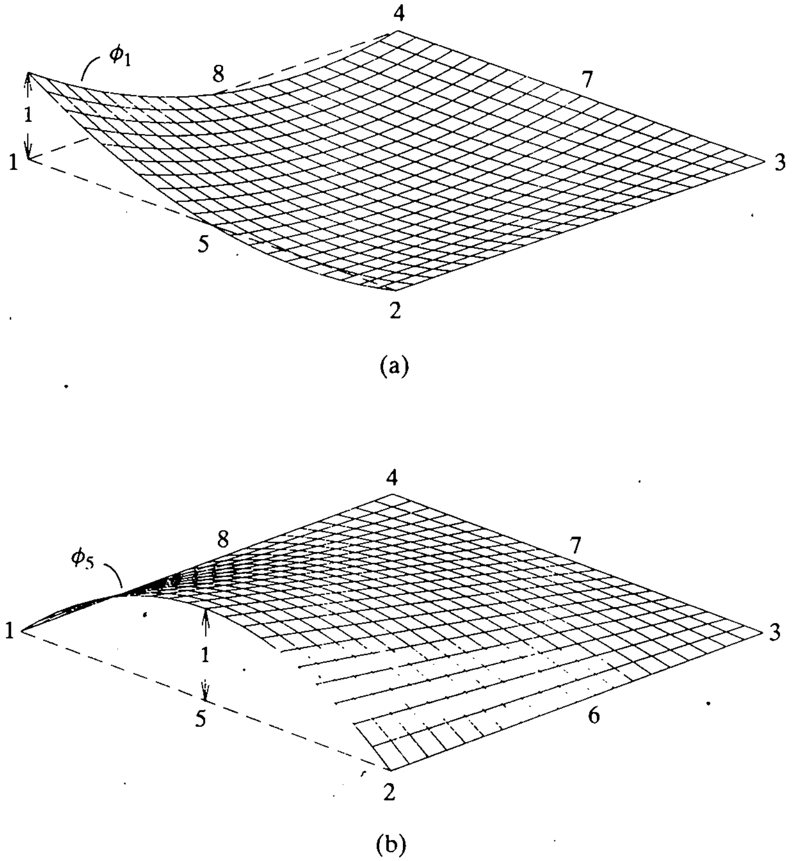


Figure 9.27 Eight-node interpolation functions for rectangular elements

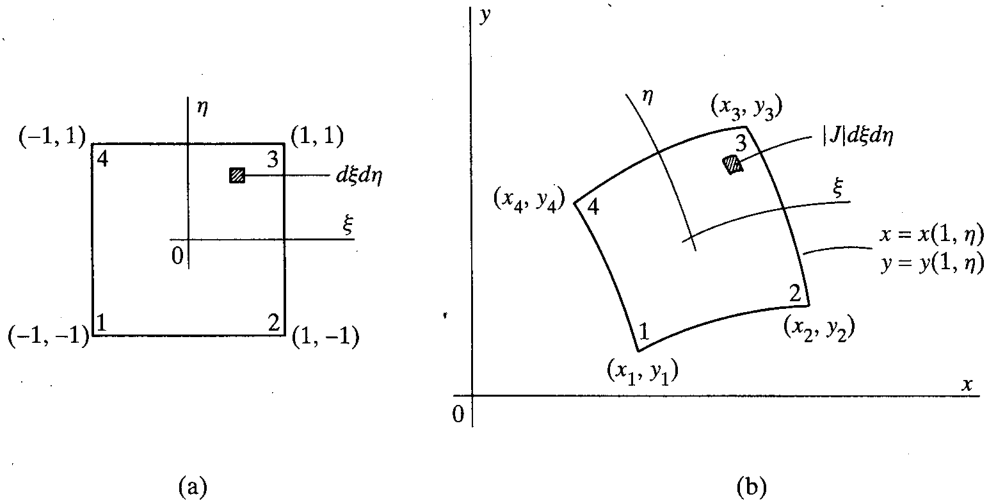
9.7 ISOPARAMETRIC ELEMENTS

At times, the shape of the membrane boundary requires more versatile elements, such as general quadrilateral elements and elements with curved boundaries. Yet, it is difficult to forsake triangular and rectangular elements with their many advantages. Fortunately, it is not necessary to abandon them due to some ingenuous coordinate transformations introduced by Taig (Ref. 24) and generalized by Irons and Ergatoudis et al. (Refs. 12 and 9, respectively).

The coordinate transformation

$$x = x(\xi, \eta), \qquad y = y(\xi, \eta) \tag{9.93}$$

represents a mapping of the points (ξ, η) in the ξ, η-plane onto points (x, y) in the x, y-plane. In particular, the objective is to use a coordinate transformation capable of straightening out curved elements. Figure 9.28a shows a rectangular element in the ξ, η-plane and Fig. 9.28b shows a curved element in the x, y-plane. We refer to the element in the ξ, η-plane as the *master element* and to the element in the x, y-plane as an *isoparametric element*. Because in the finite element method displacements are



(a) (b)

Figure 9.28 **(a)** Master element **(b)** Isoparametric element

approximated by piecewise polynomials, there are reasons to believe that piecewise polynomials can also be used for mapping one element onto another. We assume that the finite element approximation of the displacement w over a given element with n nodes has the form

$$w(\xi,\eta) = \sum_{i=1}^{n} \phi_i(\xi,\eta)w_i = \boldsymbol{\phi}^T(\xi,\eta)\mathbf{w} \tag{9.94}$$

where $\boldsymbol{\phi}$ is an n-vector of interpolation functions and $\mathbf{w}$ an n-vector of nodal displacements. Moreover, we assume that the mapping of (ξ, η) onto (x, y) is given by

$$x = \sum_{i=1}^{n} x_i\phi_i(x,\eta) = \boldsymbol{\phi}^T(\xi,\eta)\mathbf{x},\ y = \sum_{i=1}^{n} y_i\phi_i(\xi,\eta) = \boldsymbol{\phi}^T(\xi,\eta)\mathbf{y} \tag{9.95}$$

in which $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]^T$ and $\mathbf{y} = [y_1\ y_2\ \ldots\ y_n]^T$ are n-vectors with entries equal to the x-and y-components, respectively, of the nodal points (x_i, y_i). Equations (9.95) are said to represent an *isoparametric* mapping. If the dimension of $\boldsymbol{\phi}$ in Eq. (9.95) is lower, or higher, than the dimension of $\boldsymbol{\phi}$ in Eq. (9.94), the transformation is said to be *subparametric*, or *superparametric*, respectively. We are concerned only with isoparametric transformations.

Next, we consider the problem of deriving the stiffness and mass matrices for isoparametric elements. As in the case of rectangular elements, the element stiffness matrix requires the transformation from rectangular coordinates to natural coordinates. In particular, we recall from Sec. 9.6 that the stiffness matrix involves the partial derivatives $\partial w/\partial x$ and $\partial w/\partial y$, as well as the differential element of area

$dx\,dy$. Hence, for the general transformation described by Eqs. (9.93), we can write

$$dx = \frac{\partial x}{\partial \xi} d\xi + \frac{\partial x}{\partial \eta} d\eta, \qquad dy = \frac{\partial y}{\partial \xi} d\xi + \frac{\partial y}{\partial \eta} d\eta \tag{9.96}$$

Equations (9.96) can be rewritten in the matrix form

$$\begin{bmatrix} dx \\ dy \end{bmatrix} = J^T \begin{bmatrix} d\xi \\ d\eta \end{bmatrix} \tag{9.97}$$

where

$$J = \begin{bmatrix} \partial x/\partial \xi & \partial y/\partial \xi \\ \partial x/\partial \eta & \partial y/\partial \eta \end{bmatrix} \tag{9.98}$$

represents the *Jacobian matrix*. Then, the differential element of area can be shown to transform according to (Ref. 22)

$$dx\,dy = |J|\,d\xi\,d\eta \tag{9.99}$$

in which

$$|J| = \frac{\partial x}{\partial \xi}\frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta}\frac{\partial y}{\partial \xi} \tag{9.100}$$

is the *Jacobian determinant*, or simply the Jacobian of the transformation. The inverse of the transformation (9.93) has the form

$$\xi = \xi(x, y), \qquad \eta = \eta(x, y) \tag{9.101}$$

so that we can write

$$\begin{bmatrix} d\xi \\ d\eta \end{bmatrix} = \bar{J}^T \begin{bmatrix} dx \\ dy \end{bmatrix} \tag{9.102}$$

where

$$\bar{J} = \begin{bmatrix} \partial \xi/\partial x & \partial \eta/\partial x \\ \partial \xi/\partial y & \partial \eta/\partial y \end{bmatrix} \tag{9.103}$$

Comparing Eqs. (9.97) and (9.102), we conclude that

$$\begin{bmatrix} \partial \xi/\partial x & \partial \eta/\partial x \\ \partial \xi/\partial y & \partial \eta/\partial y \end{bmatrix} = J^{-1} = \frac{1}{|J|}\begin{bmatrix} \partial y/\partial \eta & -\partial y/\partial \xi \\ -\partial x/\partial \eta & \partial x/\partial \eta \end{bmatrix} \tag{9.104}$$

Next we consider

$$\frac{\partial w}{\partial x} = \frac{\partial \xi}{\partial x}\frac{\partial w}{\partial \xi} + \frac{\partial \eta}{\partial x}\frac{\partial w}{\partial \eta}, \qquad \frac{\partial w}{\partial y} = \frac{\partial \xi}{\partial y}\frac{\partial w}{\partial \xi} + \frac{\partial \eta}{\partial y}\frac{\partial w}{\partial y} \tag{9.105}$$

in which $\partial \xi/\partial x$, $\partial \xi/\partial y$, $\partial \eta/\partial x$ and $\partial \eta/\partial y$ can be obtained from Eq. (9.104). Equations (9.105) in conjunction with Eq. (9.104) can be arranged in the matrix form

$$\begin{aligned} \begin{bmatrix} \partial w/\partial x \\ \partial w/\partial y \end{bmatrix} &= \begin{bmatrix} \partial \xi/\partial x & \partial \eta/\partial x \\ \partial \xi/\partial y & \partial \eta/\partial y \end{bmatrix}\begin{bmatrix} \partial w/\partial \xi \\ \partial w/\partial \eta \end{bmatrix} \\ &= \frac{1}{|J|}\begin{bmatrix} \partial y/\partial \eta & -\partial y/\partial \xi \\ -\partial x/\partial \eta & \partial x/\partial \xi \end{bmatrix}\begin{bmatrix} \partial w/\partial \xi \\ \partial w/\partial \eta \end{bmatrix} \end{aligned} \tag{9.106}$$

At this point, we are ready to write expressions for the isoparameteric element stiffness and mass matrices. To this end, we assume that Fig. 9.28 represents the transformation between the master element and a typical isoparametric element j, insert Eqs. (9.99) and (9.106) into Eq. (9.60), consider Eq. (9.94) with $\mathbf{w}$ replaced by $\mathbf{w}_j$, as well as Eqs. (9.95) with $\mathbf{x}$ and $\mathbf{y}$ replaced by $\mathbf{x}_j$ and $\mathbf{y}_j$, respectively, and obtain

$$\begin{aligned}
[w, w]_j &= T \int_{A_j} \begin{bmatrix} \partial w/\partial x \\ \partial w/\partial y \end{bmatrix}^T \begin{bmatrix} \partial w/\partial x \\ \partial w/\partial y \end{bmatrix} dx\, dy \\
&= T \int_{-1}^{1} \int_{-1}^{1} \frac{1}{|J|_j} \begin{bmatrix} \dfrac{\partial w}{\partial \xi} \\[2ex] \dfrac{\partial w}{\partial \eta} \end{bmatrix}^T \begin{bmatrix} \left(\dfrac{\partial x}{\partial \eta}\right)^2 + \left(\dfrac{\partial y}{\partial \eta}\right)^2 & -\left(\dfrac{\partial x}{\partial \xi}\dfrac{\partial x}{\partial \eta} + \dfrac{\partial y}{\partial \xi}\dfrac{\partial y}{\partial \eta}\right) \\[2ex] -\left(\dfrac{\partial x}{\partial \xi}\dfrac{\partial x}{\partial \eta} + \dfrac{\partial y}{\partial \xi}\dfrac{\partial y}{\partial \eta}\right) & \left(\dfrac{\partial x}{\partial \xi}\right)^2 + \left(\dfrac{\partial y}{\partial \xi}\right)^2 \end{bmatrix} \begin{bmatrix} \dfrac{\partial w}{\partial \xi} \\[2ex] \dfrac{\partial w}{\partial \eta} \end{bmatrix} d\xi\, d\eta \\
&= \mathbf{w}_j^T K^{(j)} \mathbf{w}_j
\end{aligned} \tag{9.107}$$

where

$$K^{(j)} = T \int_{-1}^{1} \int_{-1}^{1} \frac{1}{|J|_j} \begin{bmatrix} \dfrac{\partial \boldsymbol{\phi}}{\partial \xi} & \dfrac{\partial \boldsymbol{\phi}}{\partial \eta} \end{bmatrix} B \begin{bmatrix} \partial \boldsymbol{\phi}^T/\partial \xi \\ \partial \boldsymbol{\phi}^T/\partial \eta \end{bmatrix} d\xi\, d\eta \tag{9.108}$$

is the isoparametric element stiffness matrix, in which, from Eq. (9.100),

$$|J|_j = \mathbf{x}_j^T \left[\frac{\partial \boldsymbol{\phi}}{\partial \xi}\frac{\partial \boldsymbol{\phi}^T}{\partial \eta} - \frac{\partial \boldsymbol{\phi}}{\partial \eta}\frac{\partial \boldsymbol{\phi}^T}{\partial \xi} \right] \mathbf{y}_j \tag{9.109}$$

and

$$B = \begin{bmatrix} \mathbf{x}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \eta}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \eta} \mathbf{x}_j + \mathbf{y}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \eta}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \eta} \mathbf{y}_j & -\left(\mathbf{x}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \xi}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \eta} \mathbf{x}_j + \mathbf{y}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \xi}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \eta} \mathbf{y}_j \right) \\[2ex] \text{symm} & \mathbf{x}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \xi}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \xi} \mathbf{x}_j + \mathbf{y}_j^T \dfrac{\partial \boldsymbol{\phi}}{\partial \xi}\dfrac{\partial \boldsymbol{\phi}^T}{\partial \xi} \mathbf{y}_j \end{bmatrix} \tag{9.110}$$

Moreover, introducing Eqs. (9.94) and (9.99) into Eq. (9.62), we can write

$$(\sqrt{m}w, \sqrt{m}w)_j = \rho \int_{A_j} w^2 dx\, dy = \rho \int_{-1}^{1} \int_{-1}^{1} |J|_j w^2\, d\xi\, d\eta = \mathbf{w}_j^T M^{(j)} \mathbf{w}_j \tag{9.111}$$

where

$$M^{(j)} = \rho \int_{-1}^{1} \int_{-1}^{1} |J|_j \boldsymbol{\phi} \boldsymbol{\phi}^T d\xi \, d\eta \tag{9.112}$$

is the isoparametric element mass matrix.

The simplest and most common isoparametric transformation is that in which the interpolation functions are *bilinear* and the rectangular element in the ξ, η-plane has four nodes. In this case, from Eqs. (9.81), the vector of interpolation functions is simply

$$\boldsymbol{\phi} = \frac{1}{4}[(1 - \xi)(1 - \eta) \;\; (1 + \xi)(1 - \eta) \;\; (1 + \xi)(1 + \eta) \;\; (1 - \xi)(1 + \eta)]^T \tag{9.113}$$

Of course, the vectors $\mathbf{x}_j$ and $\mathbf{y}_j$ in Eqs. (9.109) and (9.110) are four-dimensional. From Eq. (9.113), we write

$$\frac{\partial \boldsymbol{\phi}}{\partial \xi} = \frac{1}{4}\begin{bmatrix} -(1 - \eta) \\ 1 - \eta \\ 1 + \eta \\ -(1 + \eta) \end{bmatrix}, \qquad \frac{\partial \boldsymbol{\phi}}{\partial \eta} = \frac{1}{4}\begin{bmatrix} -(1 - \xi) \\ -(1 + \xi) \\ 1 + \xi \\ 1 - \xi \end{bmatrix}. \tag{9.114}$$

so that the Jacobian determinant, Eq. (9.109), can be written in the explicit form

$$|J|_j = \frac{1}{8}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}_j^T \begin{bmatrix} 0 & \xi(1 - \eta) & -(\xi - \eta) & -(1 - \xi) \\ & 0 & 1 + \xi & -(\xi + \eta) \\ & \text{symm} & 0 & 1 + \eta \\ & & & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}_j \tag{9.115}$$

Similarly, we can write

$$\frac{\partial \boldsymbol{\phi}}{\partial \xi}\frac{\partial \boldsymbol{\phi}^T}{\partial \xi} = \frac{1}{16}\begin{bmatrix} (1 - \eta)^2 & -(1 - \eta)^2 & -(1 - \eta^2) & 1 - \eta^2 \\ & (1 - \eta)^2 & 1 - \eta^2 & -(1 - \eta^2) \\ & \text{symm} & (1 + \eta)^2 & -(1 + \eta)^2 \\ & & & (1 + \eta)^2 \end{bmatrix} \tag{9.116a}$$

$$\frac{\partial \boldsymbol{\phi}}{\partial \xi}\frac{\partial \boldsymbol{\phi}^T}{\partial \eta} =$$

$$\frac{1}{16}\begin{bmatrix} (1 - \xi)(1 - \eta) & -(1 - \xi)(1 - \eta) & -(1 - \xi)(1 + \eta) & (1 - \xi)(1 + \eta) \\ (1 + \xi)(1 - \eta) & -(1 + \xi)(1 - \eta) & -(1 + \xi)(1 + \eta) & (1 + \xi)(1 + \eta) \\ -(1 + \xi)(1 - \eta) & (1 + \xi)(1 - \eta) & (1 + \xi)(1 + \eta) & -(1 + \xi)(1 + \eta) \\ -(1 - \xi)(1 - \eta) & (1 - \xi)(1 - \eta) & (1 - \xi)(1 + \eta) & -(1 - \xi)(1 + \eta) \end{bmatrix} \tag{9.116b}$$

$$\frac{\partial \boldsymbol{\phi}}{\partial \eta}\frac{\partial \boldsymbol{\phi}^T}{\partial \eta} = \frac{1}{16}\begin{bmatrix} (1-\xi)^2 & 1-\xi^2 & -(1-\xi^2) & -(1-\xi)^2 \\ & (1+\xi)^2 & -(1+\xi)^2 & -(1-\xi^2) \\ & \text{symm} & (1+\xi)^2 & 1-\xi^2 \\ & & & (1-\xi)^2 \end{bmatrix} \tag{9.116c}$$

The element stiffness matrix is obtained by inserting Eq. (9.110) in conjunction with Eqs. (9.116), as well as Eqs. (9.114) and (9.115), into Eq. (9.108) and carrying out the necessary integrations for given vectors $\mathbf{x}_j$ and $\mathbf{y}_j$ of global nodal positions. Similarly, the element mass matrix is obtained by inserting Eqs. (9.113) and (9.115) into Eq. (9.112) and performing the indicated integrations. The nature of the integrals in Eqs. (9.108) and (9.112) dictates that the integrations be carried out numerically, which is consistent with the idea of assigning as much of the routine work as possible to the computer.

The assembly process is similar to the one described in Sec. 9.5. The element stiffness and mass matrices are 4 × 4 and require the global nodal positions. Then, the global stiffness and mass matrices are assembled through the use of a connectivity array, as in Sec. 9.5. As an illustration, we consider the quadrilateral membrane shown in Fig. 9.29. The numbering scheme is as for the rectangular membrane of Fig. 9.24, so that the connectivity array remains in the form of Eq. (9.87). On the other hand, the nodal positions are different and are given in Table 9.3

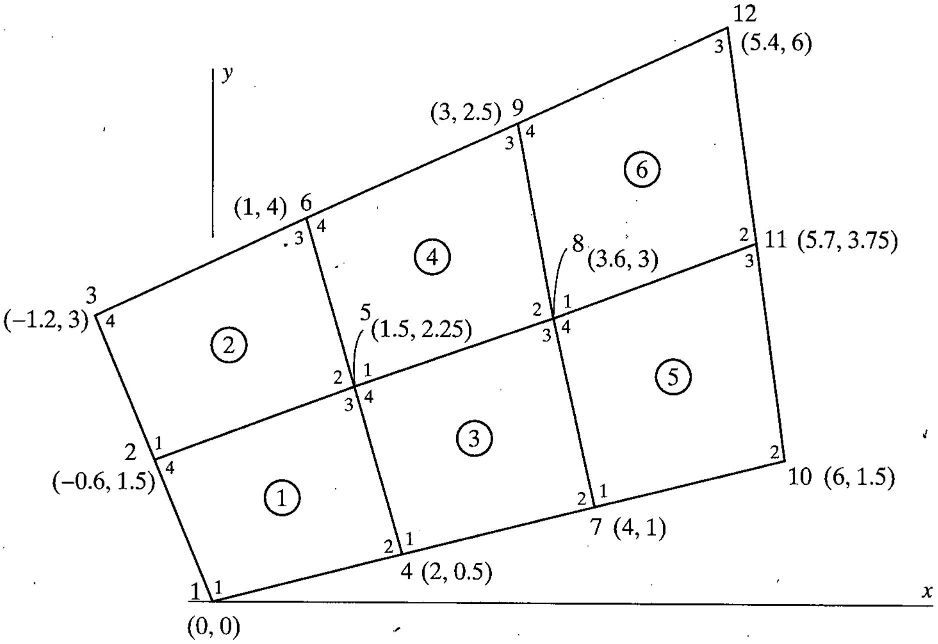


Figure 9.29 Quadrilateral membrane

If a given element has a curved edge, then the isoparameteric transformation must involve higher-degree interpolation functions, which complicates matters appreciably. For a discussion of this subject, see Ref. 22 (p. 158).

TABLE 9.3 Nodal Position Vectors

	$i \backslash j$	1	2	3	4	5	6
$\mathbf{x}_j$	1	0	−0.6	2	1.5	4	3.6
	2	2	1.5	4	3.6	6	5.7
	3	1.5	1	3.6	3.2	5.7	5.4
	4	−0.6	−1.2	1.5	1	3.6	3.2
$\mathbf{y}_j$	1	0	1.5	0.5	2.25	1	3
	2	0.5	2.25	1	3	1.5	3.75
	3	2.25	4	3	5	3.75	6
	4	1.5	3	2.25	4	3	5

9.8 VIBRATION OF PLATES

As can be concluded from Sec. 7.13, plates are significantly more complex than membranes, and the same can be said about finite element solutions to plate problems versus membrane problems. The difficulties can be traced to the fact that the stiffness operator L is the fourth-order biharmonic operator ∇^4 for plates and only the second-order Laplacian operator ∇^2 for membranes. Hence, whereas for membranes the continuity conditions between elements are imposed on the displacement w alone, for plates the continuity conditions are imposed on the displacement and its derivatives. Usually three continuity conditions, one concerning the displacement and two concerning rotations about two orthogonal axes, must be satisfied at each node. If complete interelement slope continuity is imposed, the derivation of interpolation functions becomes an intimidating task. The task is significantly simpler if one insists only on interelement displacement continuity and slope continuity at nodes alone, which gives rise to so-called *nonconforming* elements. Continuity at interface points can be restored by introducing new interpolation functions, thus obtaining *conforming* elements, but accuracy tends to suffer (Ref. 28, p. 227).

The process of deriving element stiffness and mass matrices involves the choice of finite elements. We begin with rectangular elements, which are the simplest for plate vibration. To this end, we consider the master element of Fig. 9.30, characterized by three degrees of freedom at each node, one translation and two rotations. Hence, the element nodal displacement vector is twelve-dimensional, having the form

$$\mathbf{w} = \begin{bmatrix} w_1 & \theta_{\xi 1} & \theta_{\eta 1} & w_2 & \theta_{\xi 2} & \dots & \theta_{\eta 4} \end{bmatrix}^T \tag{9.117}$$

The relation between the translation and rotations in terms of natural coordinates is simply

$$\theta_\xi = \partial w / \partial \eta, \qquad \theta_\eta = -\partial w / \partial \xi \tag{9.118}$$

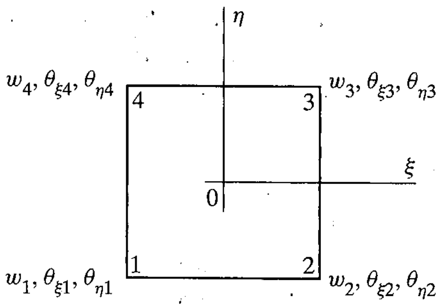

η
$w_4, \theta_{\xi 4}, \theta_{\eta 4}$
$w_3, \theta_{\xi 3}, \theta_{\eta 3}$
4
3
ξ
0
1
2
$w_1, \theta_{\xi 1}, \theta_{\eta 1}$
$w_2, \theta_{\xi 2}, \theta_{\eta 2}$

Figure 9.30 Rectangular element with three degrees of freedom at each node

The derivation of the interpolation functions follows the usual procedure, but the details are different. Because there are twelve degrees of freedom, the polynomials must contain twelve constants. A complete fourth-order polynomials has fourteen terms that together with a constant term makes fifteen terms. Hence, certain terms must be omitted. A good choice for the interpolation functions is (Ref. 28)

$$\begin{aligned}\phi_i(\xi,\eta) &= c_{i1} + c_{i2}\xi + c_{i3}\eta + c_{i4}\xi^2 + c_{i5}\xi\eta + c_{i6}\eta^2 + c_{i7}\xi^3 + c_{i8}\xi^2\eta \\ &\quad + c_{i9}\xi\eta^2 + c_{i10}\eta^3 + c_{i11}\xi^3\eta + c_{i12}\xi\eta^3, \qquad i = 1, 2, \ldots, 12\end{aligned} \tag{9.119}$$

This polynomial has the advantage that along any line ξ = constant or η = constant, such as the element boundaries, it reduces to a cubic containing four constants. But, a cubic is defined uniquely by four constants, so that the displacements and slopes at the two ends of an element boundary define the displacement along this boundary uniquely. Because the nodal displacements and slopes are shared by adjacent elements, displacement continuity is ensured at all interface points. This does not guarantee continuity of the slope in the direction normal to the boundary, so that this is a nonconforming element (Ref. 28). As usual, we assume that the displacement at any point of the master element is given by

$$w(\xi,\eta) = \boldsymbol{\phi}^T \mathbf{w} \tag{9.120}$$

where $\boldsymbol{\phi} = [\phi_1 \ \phi_2 \ \ldots \ \phi_{12}]^T$ is a vector of interpolation functions, whose components are given by Eqs. (9.119). The coefficients $c_{i1}, c_{i2}, \ldots, c_{i12}$ in Eqs. (9.119) can be obtained by modifying the approach used in Sec. 9.5 for membranes so as to take into account Eqs. (9.118). Consistent with this, we write

$$A = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix} \tag{9.121}$$

in which

$$A_k = \begin{bmatrix} 1 & \xi_k & \eta_k & \xi_k^2 & \xi_k\eta_k & \eta_k^2 & \xi_k^3 & \xi_k^2\eta_k & \xi_k\eta_k^2 & \eta_k^3 & \xi_k^3\eta_k & \xi_k\eta_k^3 \\ 0 & 0 & 1 & 0 & \xi_k & 2\eta_k & 0 & \xi_k^2 & 2\xi_k\eta_k & 3\eta_k^2 & \xi_k^3 & 3\xi_k\eta_k^2 \\ 0 & -1 & 0 & -2\xi_k & -\eta_k & 0 & -3\xi_k^2 & -2\xi_k\eta_k & -\eta_k^2 & 0 & -3\xi_k^2\eta_k & -\eta_k^3 \end{bmatrix}, \quad ; \; k = 1, 2, 3, 4 \tag{9.122}$$

The terms ξ_k and η_k $(k = 1, 2, 3, 4)$ in Eqs. (9.122) represent the natural coordinates of the nodal points, or

$$\xi_1 = \eta_1 = -1, \quad \xi_2 = 1; \quad \eta_2 = -1, \quad \xi_3 = \eta_3 = 1, \quad \xi_4 = -1, \quad \eta_4 = 1 \tag{9.123}$$

Hence, inserting Eqs. (9.123) into Eqs. (9.122) and the result into Eq. (9.121), we have

$$A = \begin{bmatrix} 1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 0 & 0 & 1 & 0 & -1 & -2 & 0 & 1 & 2 & 3 & -1 & -3 \\ 0 & -1 & 0 & 2 & 1 & 0 & -3 & -2 & -1 & 0 & 3 & 1 \\ 1 & 1 & -1 & 1 & -1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 1 & -2 & 0 & 1 & -2 & 3 & 1 & 3 \\ 0 & -1 & 0 & -2 & 1 & 0 & -3 & 2 & -1 & 0 & 3 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 2 & 0 & 1 & 2 & 3 & 1 & 3 \\ 0 & -1 & 0 & -2 & -1 & 0 & -3 & -2 & -1 & 0 & -3 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\ 0 & 0 & 1 & 0 & -1 & 2 & 0 & 1 & -2 & 3 & -1 & -3 \\ 0 & -1 & 0 & 2 & -1 & 0 & -3 & 2 & -1 & 0 & -3 & -1 \end{bmatrix} \tag{9.124}$$

which has the inverse

$$A^{-1} = \frac{1}{8} \begin{bmatrix} 2 & 1 & -1 & 2 & 1 & 1 & 2 & -1 & 1 & 2 & -1 & -1 \\ -3 & -1 & 1 & 3 & 1 & 1 & 3 & -1 & 1 & -3 & 1 & 1 \\ -3 & -1 & 1 & -3 & -1 & -1 & 3 & -1 & 1 & 3 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 \\ 4 & 1 & -1 & -4 & -1 & -1 & 4 & -1 & 1 & -4 & 1 & 1 \\ 0 & -1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & -1 & -1 & 0 & -1 & -1 & 0 & -1 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & -1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 & -1 & 1 & 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 1 & 0 & 1 & -1 & 0 & -1 & 1 & 0 & -1 \\ -1 & -1 & 0 & 1 & 1 & 0 & -1 & 1 & 0 & 1 & -1 & 0 \end{bmatrix} \tag{9.125}$$

Inserting the columns of A^{-1} into Eqs. (9.119), we obtain the interpolation functions

$$\begin{aligned}
\phi_1 &= \frac{1}{8}\left(2 - 3\xi - 3\eta + 4\xi\eta + \xi^3 + \eta^3 - \xi^3\eta - \xi\eta^3\right) \\
\phi_2 &= \frac{1}{8}\left(1 - \xi - \eta + \xi\eta - \eta^2 + \xi\eta^2 + \eta^3 - \xi\eta^3\right) \\
\phi_3 &= -\frac{1}{8}\left(1 - \xi - \eta - \xi^2 + \xi\eta + \xi^3 + \xi^2\eta - \xi^3\eta\right) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\phi_{12} &= -\frac{1}{8}\left(1 - \xi + \eta - \xi^2 - \xi\eta + \xi^3 - \xi^2\eta + \xi^3\eta\right)
\end{aligned} \tag{9.126}$$

Next we address the task of deriving the element stiffness and mass matrices. Tò this end, we consider Eqs. (7.316) and (7.318) and write the element energy inner product

$$[w, w]_j = \int_{A_j} D_E \left\{ (\nabla^2 w)^2 + (1-\nu)\left[2\left(\frac{\partial^2 w}{\partial x \partial y}\right)^2 - \frac{\partial^2 w}{\partial x^2}\frac{\partial^2 w}{\partial y^2} - \frac{\partial^2 w}{\partial y^2}\frac{\partial^2 w}{\partial x^2} \right]\right\} dA_j, \quad x, y \text{ in } D_j \tag{9.127}$$

where D_E is the plate flexural rigidity, Eq. (7.317), and the weighted inner product

$$\left(\sqrt{m}w, \sqrt{m}w\right)_j = \int_{A_j} m w^2 dx\, dy, \qquad x, y \text{ in } D_j \tag{9.128}$$

To transform the integrals to ones in terms of natural coordinates, we refer to Fig. 9.31 and write the relations

$$x = x_j + a\xi, \qquad y = y_j + b\eta \tag{9.129}$$

Then, following the usual procedure, we obtain the element stiffness matrix

$$\begin{aligned}
K^{(j)} = ab \int_{-1}^{1}\int_{-1}^{1} D_{Ej} \Bigg\{ &\left(\frac{1}{a^2}\frac{\partial^2 \boldsymbol{\phi}}{\partial \xi^2} + \frac{1}{b^2}\frac{\partial^2 \boldsymbol{\phi}}{\partial \eta^2}\right)\left(\frac{1}{a^2}\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi^2} + \frac{1}{b^2}\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \eta^2}\right) \\
&+ \frac{1-\nu}{a^2b^2}\left[2\frac{\partial^2 \boldsymbol{\phi}}{\partial \xi\, \partial \eta}\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi\, \partial \eta} - \left(\frac{\partial^2 \boldsymbol{\phi}}{\partial \xi^2}\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \eta^2} + \frac{\partial^2 \boldsymbol{\phi}}{\partial \eta^2}\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi^2}\right)\right]\Bigg\} d\xi\, d\eta
\end{aligned} \tag{9.130}$$

in which D_{Ej} is the flexural rigidity for element j, and the element mass matrix

$$M^{(j)} = ab \int_{-1}^{1}\int_{-1}^{1} m_j \boldsymbol{\phi}\boldsymbol{\phi}^T d\xi\, d\eta \tag{9.131}$$

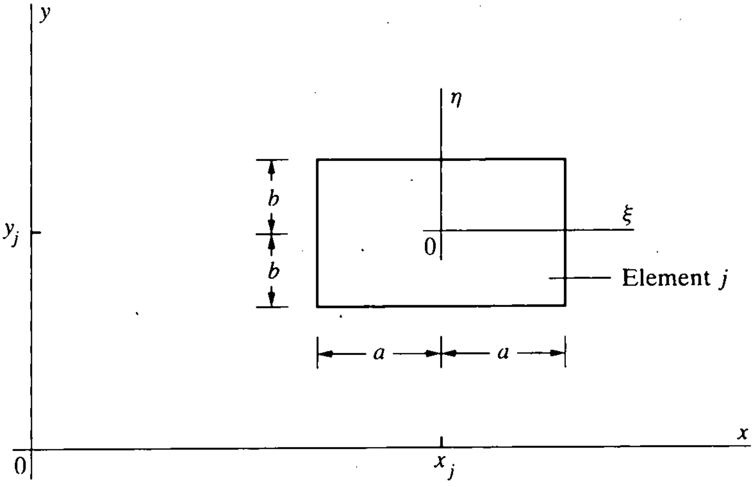


Figure 9.31 Global and local coordinates for a rectangular element

where m_j is the mass density for element j. The element stiffness and mass matrices are obtained by introducing Eqs. (9.126) into Eqs. (9.130) and (9.131), respectively, and performing the indicated integrations.

The assembly process for rectangular plate elements is similar to that for rectangular membrane elements, the main difference being that for plates there are three degrees of freedom per node. Keeping this in mind, the connectivity array for rectangular plate elements remains the same as for rectangular membrane elements.

The generalization of rectangular elements to quadrilateral elements for plate vibration is not trivial, and serious difficulties can be anticipated (Ref. 28, p. 240).

At this point, we turn our attention to triangular plate elements. In view of our discussion earlier in this section, we consider a triangular element with three nodes and three degrees of freedom per node, as shown in Fig. 9.32, for a total of nine degrees of freedom per element. Here we encounter immediately a problem, as the lowest-degree polynomial admissible is the cubic, which has nine terms. Adding the constant term, this would yield interpolation functions in terms of area coordinates

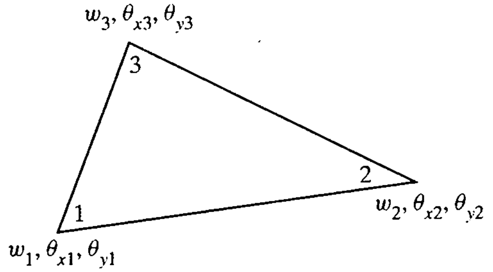

$w_3, \theta_{x3}, \theta_{y3}$
3
2
1
$w_2, \theta_{x2}, \theta_{y2}$
$w_1, \theta_{x1}, \theta_{y1}$

Figure 9.32 Triangular element with three degrees of freedom at each node

of the form

$$\phi_i = c_{i1} + c_{i2}\xi_1 + c_{i3}\xi_2 + c_{i4}\xi_1^2 + c_{i5}\xi_1\xi_2 + c_{i6}\xi_2^2 + c_{i7}\xi_1^3 + c_{i8}\xi_1^2\xi_2 + c_{i9}\xi_1\xi_2^2 + c_{i10}\xi_2^3 \tag{9.132}$$

where the dependence on the third area coordinate ξ_3 is implicit. Because there are ten constants and only nine degrees of freedom, there are several alternatives. One alternative is to add an internal node with only the transverse displacement as a degree of freedom (Ref. 22). Another alternative is to take arbitrarily $c_{i8} = c_{i9}$. An alternative with superior convergence characteristics consists of the interpolation functions (Ref. 28, p. 244)

$$\begin{aligned}
\phi_1 &= \xi_1 + \xi_1^2\xi_2 + \xi_1^2\xi_3 - \xi_1\xi_2^2 - \xi_1\xi_2^2 \\
\phi_2 &= (y_3 - y_1)(\xi_3\xi_1^2 + \frac{1}{2}\xi_1\xi_2\xi_3) - (y_1 - y_2)(\xi_1^2\xi_2 + \frac{1}{2}\xi_1\xi_2\xi_3) \\
\phi_3 &= (x_1 - x_3)(\xi_3\xi_1^2 + \frac{1}{2}\xi_1\xi_2\xi_3) - (x_2 - x_1)(\xi_1^2\xi_2 + \frac{1}{2}\xi_1\xi_2\xi_3) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\phi_9 &= (x_3 - x_2)(\xi_2\xi_3^2 + \frac{1}{2}\xi_1\xi_2\xi_3) - (x_1 - x_3)(\xi_3^2\xi_1 + \frac{1}{2}\xi_1\xi_2\xi_3)
\end{aligned} \tag{9.133}$$

and note that a difference in the sign of $\phi_2, \phi_3, \phi_5, \ldots, \phi_9$ is due to the fact that the rotations defined in Ref. 28 are the negative of those defined by Eqs. (9.118).

The element stiffness and mass matrices are as given by Eqs. (9.127) and (9.128), respectively, where

$$w = \boldsymbol{\phi}^T \mathbf{w} \tag{9.134}$$

in which $\boldsymbol{\phi}$ is a nine-dimensional vector with components given by Eqs. (9.133) and $\mathbf{w}$ is the nine-dimensional nodal vector

$$\mathbf{w} = [w_1\ \theta_{x1}\ \theta_{y1}\ w_2\ \theta_{x2}\ \theta_{y2}\ \ w_3\ \theta_{x3}\ \theta_{y3}]^T \tag{9.135}$$

and note that the rotations are defined in terms of rectangular coordinates as

$$\theta_{xi} = (\partial w/\partial y)_i, \qquad \theta_{yi} = -(\partial w/\partial x)_i, \qquad i = 1, 2, 3 \tag{9.136}$$

The relation between the rectangular coordinates and area coordinates are given by Eqs. (9.55). Hence, using Eqs. (9.55) with A replaced by A_j and Eq. (9.134) with $\mathbf{w}$

replaced by $\mathbf{w}_j$, we can write

$$\begin{aligned}
\frac{\partial w}{\partial x} &= \frac{\partial w}{\partial \xi_1}\frac{\partial \xi_1}{\partial x} + \frac{\partial w}{\partial \xi_2}\frac{\partial \xi_2}{\partial x} + \frac{\partial w}{\partial \xi_3}\frac{\partial \xi_3}{\partial x} \\
&= \frac{1}{2A_j}\left[\frac{\partial \boldsymbol{\phi}^T}{\partial \xi_1}(y_2 - y_3) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_2}(y_3 - y_1) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_3}(y_1 - y_2)\right]\mathbf{w}_j \\
\frac{\partial w}{\partial y} &= \frac{\partial w}{\partial \xi_1}\frac{\partial \xi_1}{\partial y} + \frac{\partial w}{\partial \xi_2}\frac{\partial \xi_2}{\partial y} + \frac{\partial w}{\partial \xi_3}\frac{\partial \xi_3}{\partial y} \\
&= \frac{1}{2A_j}\left[\frac{\partial \boldsymbol{\phi}^T}{\partial \xi_1}(x_3 - x_2) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_2}(x_1 - x_3) + \frac{\partial \boldsymbol{\phi}^T}{\partial \xi_3}(x_2 - x_1)\right]\mathbf{w}_j
\end{aligned} \tag{9.137}$$

Then, using the chain rule again, we obtain

$$\begin{aligned}
\frac{\partial^2 w}{\partial x^2} &= \frac{1}{4A_j^2}\left[\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1^2}(y_2 - y_3)^2 + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2^2}(y_3 - y_1)^2 + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_3^2}(y_1 - y_2)^2\right. \\
&\quad + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_2}(y_2 - y_3)(y_3 - y_1) + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_3}(y_2 - y_3)(y_1 - y_2) \\
&\quad \left. + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2\,\partial \xi_3}(y_3 - y_1)(y_1 - y_2)\right]\mathbf{w}_j \\
\frac{\partial^2 w}{\partial y^2} &= \frac{1}{4A_j^2}\left[\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1^2}(x_3 - x_2)^2 + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2^2}(x_1 - x_3)^2 + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_3^2}(x_2 - x_1)^2\right. \\
&\quad + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_2}(x_3 - x_2)(x_1 - x_3) + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_3}(x_3 - x_2)(x_2 - x_1) \\
&\quad \left. + 2\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2\,\partial \xi_3}(x_1 - x_3)(x_2 - x_1)\right]\mathbf{w}_j \\
\frac{\partial^2 w}{\partial x\,\partial y} &= \frac{1}{4A_j^2}\left\{\frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1^2}(x_3 - x_2)(y_2 - y_3) + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2^2}(x_1 - x_3)(y_3 - y_1)\right. \\
&\quad + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_3^2}(x_2 - x_1)(y_1 - y_2) \\
&\quad + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_2}[(x_3 - x_2)(y_3 - y_1) + (x_1 - x_3)(y_2 - y_3)] \\
&\quad + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_1\,\partial \xi_3}[(x_3 - x_2)(y_1 - y_2) + (x_2 - x_1)(y_2 - y_3)] \\
&\quad \left. + \frac{\partial^2 \boldsymbol{\phi}^T}{\partial \xi_2\,\partial \xi_3}[(x_1 - x_3)(y_1 - y_2) + (x_2 - x_1)(y_3 - y_1)]\right\}\mathbf{w}_j
\end{aligned} \tag{9.138}$$

The element stiffness matrix can be obtained by inserting Eqs. (9.138) in conjunction with Eqs. (9.133) into Eq. (9.127) and performing the indicated integrations. Moreover, the element mass matrix can be produced by introducing Eqs. (9.133) into Eq. (9.131) and carrying out the required integrations. In both regards, Eq. (9.66) should prove useful.

The assembly process for triangular plate elements, including the connectivity array, is similar to that for triangular membrane elements, with the exception of the fact that for plates there are three degrees of freedom at each node instead of one.

The generation of the element stiffness and mass matrices and the assembly process are routine and can be carried out conveniently on a computer.

9.9 ERRORS IN THE APPROXIMATE EIGENVALUES AND EIGENFUNCTIONS

In view of the fact that for self-adjoint systems the finite element method represents a Rayleigh-Ritz method, if the elements are conformable, then its convergence is assured, although convergence is possible even with nonconformable elements (Ref. 28, p. 244). Clearly, the problem of nonconformable elements arises in plates in bending. It appears that convergence can be assumed if the elements pass the *patch test* (see Ref. 28, Secs. 2.7 and 11.2). It has been shown by Irons (Ref. 4) in conjunction with the interpolation functions given by Eqs. (9.133) that a mesh generated by three sets of equally spaced parallel lines, such as that of Fig. 9.33a, passes the patch test and a mesh of the type shown in Fig. 9.33b does not. The issue of convergence is to a large extent only of academic interest, as for most systems the finite element method is known to converge. Indeed, more important is the rate of convergence. It is here that the finite element method pays the price for the simplicity arising from the use of low-degree polynomials as admissible functions.

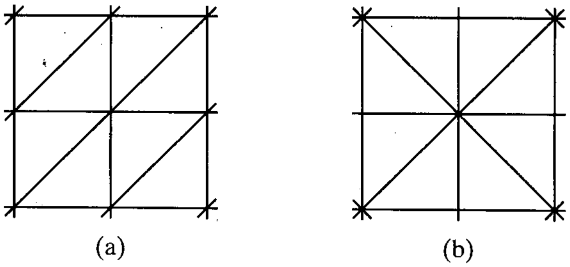


Figure 9.33 **(a)** Mesh passing the patch test **(b)** Mesh failing the patch test

Convergence is a qualitative concept. A more quantitative measure of the accuracy of the approximate solution can be obtained from error estimates. In the case of approximate solutions derived by the finite element method, the error estimates can be quantified to some extent by tying them not only to the dimension of the Ritz space but also to the degree of the elements. In this section, we give a summary of results presented in Ref. 22 (Sec. 6.3).

We are concerned with eigenvalue problems described by the differential equation

$$Lw = \lambda m w \quad \text{over } D \tag{9.139}$$

where L is a linear homogeneous self-adjoint differential operator of order $2p$, and the boundary conditions

$$B_i w = 0 \quad \text{on } S, \qquad i = 1, 2, \dots, p \tag{9.140}$$

in which B_i are boundary operators of maximum order $2p - 1$. Multiplying Eq. (9.139) through by w and integrating by parts with due consideration to boundary conditions (9.140), we obtain the weak form of the eigenvalue problem

$$[w, w] = \lambda(\sqrt{m}w, \sqrt{m}w) \tag{9.141}$$

where $[w, w]$ is an energy inner product, a measure of the potential energy, and $(\sqrt{m}w, \sqrt{m}w)$ is a weighted inner product, a measure of the kinetic energy. Then, dividing Eq. (9.141) by $(\sqrt{m}w, \sqrt{m}w)$, we obtain the Rayleigh's quotient

$$\lambda = R(w) = \frac{[w, w]}{(\sqrt{m}w, \sqrt{m}w)} \tag{9.142}$$

As shown in Sec. 7.14, rendering Rayleigh's quotient stationary is equivalent to solving the weak form of the eigenvalue problem.

Throughout this chapter, we have been concerned with finite element approximate solutions to the eigenvalue problem obtained by rendering Rayleigh's quotient stationary. Such solutions, denoted by $w^{(n)}$, have been assumed to have the form of linear combinations of admissible functions ϕ_i $(i = 1, 2, \dots, n)$ from the subspace $\mathcal{R}_n$ of the energy space $\mathcal{K}_G^p$, or

$$w^{(n)} = \boldsymbol{\phi}^T \mathbf{a} \tag{9.143}$$

where $\boldsymbol{\phi}$ is an n-vector of admissible functions and $\mathbf{a}$ an n-vector of undetermined coefficients. Inserting Eq. (9.143) into Eq. (9.142), we obtain the discretized version of Rayleigh's quotient

$$R(w^{(n)}) = \frac{[w^{(n)}, w^{(n)}]}{(\sqrt{m}w^{(n)}, \sqrt{m}w^{(n)})} = \frac{\mathbf{a}^T K \mathbf{a}}{\mathbf{a}^T M \mathbf{a}} \tag{9.144}$$

in which K and M are stiffness and mass matrices. Rayleigh's quotient has stationary values in the form of approximate eigenvalues $\lambda_r^{(n)}$ at the eigenvectors $\mathbf{a}_r$. The eigenvectors are orthogonal with respect to M and K and can be normalized so that

$$\mathbf{a}_s^T M \mathbf{a}_r = \delta_{rs}, \qquad \mathbf{a}_s^T K \mathbf{a}_r = \lambda_r^{(n)} \delta_{rs}, \qquad r, s = 1, 2, \dots, n \tag{9.145}$$

Introducing the eigenvectors $\mathbf{a}_r$ into Eq. (9.143), we obtain the approximate eigenfunctions

$$w_r^{(n)} = \boldsymbol{\phi}^T \mathbf{a}_r, \qquad r = 1, 2, \dots, n \tag{9.146}$$

which satisfy the orthonormality relations

$$(\sqrt{m}w_s^{(n)}, \sqrt{m}w_r^{(n)}) = \delta_{rs}, \qquad [w_s^{(n)}, w_r^{(n)}] = \lambda_r^{(n)} \delta_{rs}, \qquad r, s = 1, 2, \dots, n \tag{9.147}$$

Then, using the analogy with Eq. (7.404), we can write the maximin theorem for the approximate eigensolutions in the form

$$\lambda_{r+1}^{(n)} = \max_{v_i} \min_{w^{(n)}} R(w^{(n)}), \qquad (w^{(n)}, v_i) = 0, \qquad i = 1, 2, \dots, r \tag{9.148}$$

where v_i are r independent but otherwise arbitrary functions.

Using the maximin theorem, it is shown in Ref. 22 (Sec. 6.3) that the approximate eigenvalues are bounded for small h by

$$\lambda_r \leq \lambda_r^{(n)} \leq \lambda_r + 2\delta h^{2(k-p)}(\lambda_r^{(n)})^{k/p}, \qquad r = 1, 2, \dots, n \tag{9.149}$$

in which h is the longest side of the finite elements, $k-1$ is the degree of the elements and δ is some constant. The factor $h^{2(k-p)}$ in inequalities (9.149) implies that, for a given small h, the errors decrease as the degree of the elements increases. On the other hand, the factor $(\lambda_r^{(n)})^{k/p}$ indicates that the errors increase with the mode number, so that higher approximate eigenvalues tend to be inaccurate. As a rule of thumb, more than one half of the approximate eigenvalues must be regarded as unreliable (Ref. 22, p. 227). Hence, the dimension of the Ritz space $\mathcal{R}_n$ must be at least twice as large as the number of accurate eigenvalues desired.

As shown in Ref. 22 (Sec. 6.3) error estimates for approximate eigenfunctions are also possible. These are not pointwise error estimates, but estimates in an average sense. Indeed, it is demonstrated in Ref. 22 that for small h

$$\|\sqrt{m}(w_r - w_r^{(n)})\| \leq c[h^k + h^{2(k-p)}](\lambda_r^{(n)})^{k/2p}, \qquad r = 1, 2, \dots, n \tag{9.150a}$$

and

$$[w_r - w_r^{(n)}, w_r - w_r^{(n)}] \leq c' h^{2(k-p)}(\lambda_r^{(n)})^{k/p}, \qquad r = 1, 2, \dots, n \tag{9.150b}$$

Errors of a different type arise when the mass matrix is generated by a different process than the stiffness matrix, giving rise to so-called *inconsistent mass matrices*. The most common example of an inconsistent mass matrix is the *lumped mass matrix*. The use of a lumped matrix tends to raise the value of the denominator in the Rayleigh's quotient, thus lowering the approximate eigenvalues compared to the eigenvalues computed using a consistent mass matrix. Because the eigenvalues computed by means of a Rayleigh-Ritz process tend to be higher than the actual ones, this may appear as a good way of improving the estimates. This is not the case, however, as the use of inconsistent mass matrices violates the Rayleigh-Ritz code, so that the Rayleigh-Ritz theory can no longer be counted on to argue that the actual eigenvalues represent lower bounds for the approximate eigenvalues. As a result, the use of lumped masses can cause significant errors (Ref. 22, p. 228, and Ref. 25).

Before abandoning this section, it should be noted that, whereas the maximin theorem does apply to the finite element method, in general the separation theorem (Sec. 8.5) cannot be demonstrated to hold true. However, the fact that no proof exists should not be construed to imply that the separation theorem does not hold for the finite element method.

9.10 THE HIERARCHICAL FINITE ELEMENT METHOD

As demonstrated in Sec. 8.5, the accuracy of the approximate eigenvalues obtained by means of the Rayleigh-Ritz method can be improved by simply increasing the number of admissible functions in the linear combination representing the approximate solution. On the other hand, accuracy can be improved in the finite element method by refining the mesh, which implies a reduction in the width h of the element, or equivalently an increase in the number of elements. The procedure is characterized by the fact that the degree p of the elements is a fixed, generally low number. No confusion should arise from the fact that earlier in this text we used the symbol p in connection with the order of the stiffness operator L. Another way of improving the accuracy of the finite element approximation is to keep the width h constant and to increase the degree p of the polynomials. To distinguish between the two, the approach whereby the accuracy is improved by refining the finite element mesh is referred to as the *h-version* of the finite element method, and the approach whereby the degree of t he polynomials is increased is known as the *p -version* (Refs. 2, 3 and 23).

Because in the p-version of the finite element method accuracy is improved through an increase in the degree of the polynomials, which implies an increase in the number of admissible functions in the approximation, the p-version has something in common with the classical Rayleigh-Ritz method. Of course, one of the main differences remains, as the admissible functions used in the p-version of the finite element method are local functions and in the classical Rayleigh-Ritz method they are global functions. This gives the p-version greater versatility than the h-version. As a result, the rate of convergence of the p-version can be higher than that of the h-version.

In the p-version of the finite element method it is possible to choose from a variety of different sets of polynomials, provided the sets are complete. Particularly desirable are the so-called *hierarchical* polynomials, which have the property that the set of polynomials in the approximation of degree p represents a subset of the polynomials in the approximation of degree $p+1$. This version is referred to as the *hierarchical finite element method* (Refs. 18, 27 and 29) and is characterized by the fact that the mass and stiffness matrices possess the embedding property indicated by Eqs. (8.100). As a result, the separation theorem holds true for the hierarchical finite element method (Ref. 14).

Next, we demonstrate the hierarchical finite element method by applying it to the eigenvalue problem of a beam in bending. In Sec. 9.4, we have shown that the most common polynomials for beams in bending are the Hermite cubics

$$\phi_1 = 3\xi^2 - \xi^3, \quad \phi_2 = \xi^2 - \xi^3, \quad \phi_3 = 1 - 3\xi^2 + 2\xi^3, \quad \phi_4 = -\xi + 2\xi^2 - \xi^3 \tag{9.151}$$

A suitable set of hierarchical functions consists of the polynomials

$$\begin{aligned}\phi_5 &= \xi^2(1-\xi)^2 \\ \phi_6 &= \xi^2(1-\xi)^2(1-2\xi) \\ \phi_7 &= \xi^2(1-\xi)^2(1-3\xi)(2-3\xi) \\ &\cdots\cdots\cdots\cdots\cdots\cdots \\ \phi_{4+i} &= \xi^2(1-\xi)^2\prod_{j=2}^{i}(j-1-i\xi)\end{aligned} \tag{9.152}$$

and we note that all hierarchical functions and their slope have zero values at the nodes $\xi = 0$ and $\xi = 1$. As a result, when one hierarchical function is added to the approximating series for a given element, one row and one column are added to the element stiffness and mass matrices without disturbing the previously computed entries, so that they possess the *embedding property*. Hence, the separation theorem applies.

To develop an appreciation for the hierarchical finite element method, we consider the numerical example of Ref. 14, in which the five lowest eigenvalues of a uniform cantilever beam were computed by means of the h-version and p-version of the finite element method. Results are displayed in Tables 9.4 and 9.5, respectively. To place matters in proper perspective, we note that in the h-version 4, 6, 8 and 10 elements imply 8, 12, 16 and 20 degrees of freedom, respectively. On the other hand, the number of degrees of freedom for the p-version ranges from 9 to 12 in the upper columns in Table 9.5, from 10 to 16 in the middle columns and from 11 to 20 in the bottom columns. Hence, a direct comparison can be made between columns 2,3 and 4 in Table 9.4 and the extreme right upper, middle and bottom columns in Table 9.5, respectively. Clearly, the hierarchical finite element method gives significantly more accurate results than the ordinary finite element method, particularly for the higher eigenvalues. Examining results in the upper columns in Table 9.5, it is clear that the separation theorem holds true for the hierarchical finite element method. It turns out that the same is true for the eigenvalues in Table 9.4, the middle columns in Table 9.5 and the lower columns in Table 9.5. Not too much should be read into this, however. Indeed, the explanation is that for one-dimensional domains the eigenvalues tend to be well spaced, so it would be quite difficult to violate the separation theorem.

TABLE 9.4 The Five Lowest Eigenvalues Computed by the h-Version

Four Elements	Six Elements	Eight Elements	Ten Elements
0.14065	0.14064	0.14064	0.14064
0.88241	0.88160	0.88145	0.88141
2.48700	2.47240	2.46991	2.46852
4.90631	4.86724	4.84691	4.84068
9.12550	8.11453	8.04064	8.01453

TABLE 9.5 The Five Lowest Eigenvalues Computed by the p-Version Using Four Finite Elements

$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2, 3 $\phi_1, \phi_2, \ldots, \phi_5$ on 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2 $\phi_1, \phi_2, \ldots, \phi_5$ on 3, 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1 $\phi_1, \phi_2, \ldots, \phi_5$ on 2, 3, 4	$\phi_1, \phi_2, \ldots, \phi_5$ on all
0.14065	0.14064	0.14064	0.14064
0.88221	0.88190	0.88142	0.88140
2.48687	2.48008	2.47145	2.46875
4.89846	4.89464	4.89295	4.85038
8.92140	8.53673	8.26909	8.01927
$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2, 3 $\phi_1, \phi_2, \ldots, \phi_6$ on 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2 $\phi_1, \phi_2, \ldots, \phi_6$ on 3, 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1 $\phi_1, \phi_2, \ldots, \phi_6$ on 2, 3, 4	$\phi_1, \phi_2, \ldots, \phi_6$ on all
0.14064	0.14064	0.14064	0.14064
0.88220	0.88188	0.88141	0.88138
2.48633	2.47948	2.47074	2.46790
4.89359	4.88554	4.87950	4.83619
8.90123	8.51833	8.24920	7.99920
$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2, 3 $\phi_1, \phi_2, \ldots, \phi_7$ on 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1, 2 $\phi_1, \phi_2, \ldots, \phi_7$ on 3, 4	$\phi_1, \phi_2, \ldots, \phi_4$ on 1 $\phi_1, \phi_2, \ldots, \phi_7$ on 2, 3, 4	$\phi_1, \phi_2, \ldots, \phi_7$ on all
0.14064	0.14064	0.14064	0.14064
0.88220	0.88188	0.88141	0.88138
2.48633	2.47947	2.47073	2.46789
4.89351	4.88545	4.87941	4.83609
8.89793	8.51363	8.24405	7.99442

9.11 SYSTEM RESPONSE

The response of systems discretized by the finite element method can be obtained by an approach similar to that used in Sec. 8.10 in conjunction with the classical Rayleigh-Ritz method. However, in view of the fact that our treatment of the finite element method is based on the variational approach, we will modify the derivation accordingly. As a result, the approach adopted here resembles that used in Sec. 8.7 in conjunction with the assumed-modes method, which is really a variant of the Rayleigh-Ritz method.

The kinetic energy of a distributed-parameter system can be written as a weighted inner product in the form

$$T(t) = \frac{1}{2}\left(\sqrt{m(P)}\dot{w}(P,t), \sqrt{m(P)}\dot{w}(P,t)\right) \tag{9.153}$$

Similarly, the potential energy can be expressed as the energy inner product

$$V(t) = \frac{1}{2}[w(P,t),\ w(P,t)] \tag{9.154}$$

Moreover, the virtual work of the nonconservative forces has the form

$$\overline{\delta W_{nc}}(t) = \int_D f(P,t)\,\delta w(P,t)\,dD(P) \tag{9.155}$$

where $f(P,t)$ is the distributed force.

The response of a system discretized by the finite element method can be obtained by simply letting the nodal coordinates be functions of time. Hence, by analogy with Eq. (9.5), we write the response in the form

$$w(P,t) \cong w^{(n)}(P,t) = \boldsymbol{\phi}^T(P)\mathbf{q}_j(t) \text{ over } D_j, \qquad j = 1, 2, \ldots, n \tag{9.156}$$

where $\boldsymbol{\phi}$ is a vector of interpolation functions and $\mathbf{q}_j(t)$ a vector of time-dependent nodal displacements for element j. Introducing Eq. (9.156) into Eq. (9.153), we obtain the discretized kinetic energy

$$T(t) \cong \frac{1}{2}\sum_{j=1}^{n}\dot{\mathbf{q}}_j^T\left(\sqrt{m}\boldsymbol{\phi},\ \sqrt{m}\boldsymbol{\phi}^T\right)_j\dot{\mathbf{q}}_j = \frac{1}{2}\sum_{j=1}^{n}\dot{\mathbf{q}}_j^T M_j\dot{\mathbf{q}}_j \tag{9.157}$$

in which

$$M_j = \left(\sqrt{m}\boldsymbol{\phi}, \sqrt{m}\boldsymbol{\phi}^T\right)_j, j = 1, 2, \ldots, n \tag{9.158}$$

are recognized as the element mass matrices. In the same manner, the discretized potential energy can be written as

$$V(t) \cong \frac{1}{2}\sum_{j=1}^{n}\mathbf{q}_j^T\left[\boldsymbol{\phi}, \boldsymbol{\phi}^T\right]_j\mathbf{q}_j = \frac{1}{2}\sum_{j=1}^{n}\mathbf{q}_j^T K_j\mathbf{q}_j \tag{9.159}$$

where

$$K_j = \left[\boldsymbol{\phi}, \boldsymbol{\phi}^T\right]_j, \qquad j = 1, 2, \ldots, n \tag{9.160}$$

are the element stiffness matrices. Similarly, the discretized virtual work has the expression

$$\overline{\delta W_{nc}}(t) \cong \sum_{j=1}^{n}\int_{D_j} f\boldsymbol{\phi}^T\,\delta\mathbf{q}_j\,dD = \sum_{j=1}^{n}\mathbf{Q}_j^T\,\delta\mathbf{q}_j \tag{9.161}$$

in which

$$\mathbf{Q}_j(t) = \int_{D_j} f(P,t)\boldsymbol{\phi}(P)\,dD(P), \qquad j = 1, 2, \ldots, n \tag{9.162}$$

are the element nodal force vectors.

The assembly process amounts to using the connectivity array (Sec. 9.5) to eliminate redundant coordinates. It is not difficult to show that, following the assembly, the kinetic energy, potential energy and virtual work assume the usual form

$$T = \frac{1}{2}\dot{\mathbf{q}}^T M\dot{\mathbf{q}}, \qquad V = \frac{1}{2}\mathbf{q}^T K\mathbf{q} \tag{9.163a, b}$$

and

$$\overline{\delta W_{nc}} = \mathbf{Q}^T \delta\mathbf{q} \tag{9.163c}$$

where $\mathbf{q}$ is the global nodal displacement vector, M and K are the global mass and stiffness matrices and $\mathbf{Q}$ is the global nodal force vector.

Using Lagrange's equations, the nodal equations of motion for the system are simply

$$M\ddot{\mathbf{q}} + K\mathbf{q} = \mathbf{Q} \tag{9.164}$$

The solution of Eq. (9.164) was discussed in Sec. 4.10.

Example 9.5

Determine the global mass matrix, stiffness matrix and nodal force vector defining the response of a uniform simply supported beam to the distributed force $f(x,t) = (1 - x/2L)f(t)$ derived by the finite element method using Hermite cubics as interpolation functions.

We begin by determining the element mass matrix, stiffness matrix and nodal force vector. To this end, we recall from Eqs. (9.41) that the Hermite cubics have the form

$$\phi_1 = 3\xi^2 - 2\xi^3, \quad \phi_2 = \xi^2 - \xi^3, \quad \phi_3 = 1 - 3\xi^2 + 2\xi^3, \quad \phi_4 = -\xi + 2\xi^2 - \xi^3 \tag{a}$$

so that, using Eqs. (9.158) and recalling that $x = (j - \xi)h$, the element mass matrices are

$$M_j = (\sqrt{m}\boldsymbol{\phi}, \sqrt{m}\boldsymbol{\phi}^T)_j = \int_{(j-1)h}^{jh} m\boldsymbol{\phi}\boldsymbol{\phi}^T dx = mh\int_0^1 \boldsymbol{\phi}\boldsymbol{\phi}^T d\xi$$

$$= mh\int_0^1 \begin{bmatrix} 3\xi^2 - 2\xi^3 \\ \xi^2 - \xi^3 \\ 1 - 3\xi^2 + 2\xi^3 \\ -\xi + 2\xi^2 - \xi^3 \end{bmatrix} \begin{bmatrix} 3\xi^2 - 2\xi^3 \\ \xi^2 - \xi^3 \\ 1 - 3\xi^2 + 2\xi^3 \\ -\xi + 2\xi^2 - \xi^3 \end{bmatrix}^T d\xi$$

$$= \frac{mL}{420n}\begin{bmatrix} 156 & 22 & 54 & -13 \\ & 4 & 13 & -3 \\ \text{symm} & & 156 & -22 \\ & & & 4 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{b}$$

and, using Eqs. (9.160), the element stiffness matrices are

$$K_j = [\boldsymbol{\phi}, \boldsymbol{\phi}^T]_j = \int_{(j-1)h}^{jh} EI\frac{d^2\boldsymbol{\phi}}{dx^2}\frac{d^2\boldsymbol{\phi}^T}{dx^2}dx = \frac{EI}{h^3}\int_0^1 \boldsymbol{\phi}''\boldsymbol{\phi}''^T d\xi$$

$$= \frac{EI}{h^3}\int_0^1 \begin{bmatrix} 6 - 12\xi \\ 2 - 6\xi \\ -6 + 12\xi \\ 4 - 6\xi \end{bmatrix} \begin{bmatrix} 6 - 12\xi \\ 2 - 6\xi \\ -6 + 12\xi \\ 4 - 6\xi \end{bmatrix}^T d\xi$$

$$= \frac{EIn^3}{L^3}\begin{bmatrix} 12 & 6 & -12 & 6 \\ & 4 & -6 & 2 \\ \text{symm} & & 12 & -6 \\ & & & 4 \end{bmatrix}, \qquad j = 1, 2, \ldots, n \tag{c}$$

and we observe that M_j and K_j are consistent with results obtained in Example 9.3. Then, from Eqs. (9.162), the element nodal force vectors are

$$\mathbf{Q}_j(t) = \int_{(j-1)h}^{jh} f(x,t)\boldsymbol{\phi}\, dx$$

$$= hf(t)\int_0^1 \left[1 - \frac{(j-\xi)h}{2L}\right]\left[3\xi^2 - 2\xi^3 \;\; \xi^2 - \xi^3 \;\; 1 - 3\xi^2 + 2\xi^3 \;\; -\xi + 2\xi^2 - \xi^3\right]^T d\xi$$

$$= hf(t)\left\{\left(1 - \frac{jh}{2L}\right)\int_0^1 \begin{bmatrix} 3\xi^2 - 2\xi^3 \\ \xi^2 - \xi^3 \\ 1 - 3\xi^2 + 2\xi^3 \\ -\xi + 2\xi^2 - \xi^3 \end{bmatrix} d\xi + \frac{h}{2L}\int_0^1 \begin{bmatrix} 3\xi^3 - 2\xi^4 \\ \xi^3 - \xi^4 \\ \xi - 3\xi^3 + 2\xi^4 \\ -\xi^2 + 2\xi^3 - \xi^4 \end{bmatrix} d\xi\right\}$$

$$= \frac{Lf(t)}{120n^2}\left\{(2n - j)\begin{bmatrix} 30 \\ 5 \\ 30 \\ -5 \end{bmatrix} + \begin{bmatrix} 21 \\ 3 \\ 9 \\ -2 \end{bmatrix}\right\}, \qquad j = 1, 2, \ldots, n \tag{d}$$

At this point, we are ready for the assembly process. In view of the fact that for a simply supported beam the displacement is zero at both ends, $w_0 = w_n = 0$, the structure of the global mass and stiffness matrices is somewhat different from that indicated in Sec. 9.4. Indeed, in the case at hand, we must strike out the first and $(2n-1)$st rows and columns from the global mass and stiffness matrices. Hence, using Eqs. (b), we obtain the global mass matrix

$$M = \frac{mL}{420n}\begin{bmatrix} 4 & 13 & -3 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 312 & 0 & 54 & -13 & \cdots & 0 & 0 & 0 \\ & & 8 & 13 & -3 & \cdots & 0 & 0 & 0 \\ & & & 312 & 0 & \cdots & 0 & 0 & 0 \\ & & & & 8 & \cdots & 0 & 0 & 0 \\ \text{symm} & & & & & \cdots & \cdots & \cdots & \cdots \\ & & & & & & 312 & 0 & -13 \\ & & & & & & & 8 & -3 \\ & & & & & & & & 4 \end{bmatrix} \tag{e}$$

Similarly, using Eqs. (c), we can write the global stiffness matrix

$$K = \frac{EIn^3}{L^3}\begin{bmatrix} 4 & -6 & 2 & 0 & 0 & \cdots & 0 & 0 & 0 \\ & 24 & 0 & -12 & 6 & \cdots & 0 & 0 & 0 \\ & & 8 & -6 & 2 & \cdots & 0 & 0 & 0 \\ & & & 24 & 0 & \cdots & 0 & 0 & 0 \\ & & & & 8 & \cdots & 0 & 0 & 0 \\ \text{symm} & & & & & \cdots & \cdots & \cdots & \cdots \\ & & & & & & 24 & 0 & 6 \\ & & & & & & & 8 & 2 \\ & & & & & & & & 4 \end{bmatrix} \tag{f}$$

Finally, from Eqs. (d), striking out the first and $(2n-1)$st components and adding the contributions from $\mathbf{Q}_{j-1}$ and $\mathbf{Q}_j$ corresponding to shared nodal forces, we obtain the global nodal force vector

$$\mathbf{Q}(t) = \frac{Lf(t)}{120n^2}[10n-2 \quad 120n-60 \quad -4 \quad 120n-120 \quad -4 \quad \dots \quad 60n+60 \quad 5n+1 \quad -5n-2]^T \tag{g}$$

and we recognize that the global nodal displacement vector has the form

$$\mathbf{q}(t) = [h\theta_0(t) \; w_1(t) \; h\theta_1(t) \; w_2(t) \; h\theta_2(t) \; \dots \; w_{n-1}(t) \; h\theta_{n-1}(t) \; h\theta_n(t)]^T \tag{h}$$

9.12 SYNOPSIS

As far as vibration theory is concerned, the finite element method represents a Rayleigh-Ritz method. The main difference between the original Rayleigh-Ritz method, referred to as the classical Rayleigh-Ritz method, and the finite element method lies in the nature of the admissible functions. Indeed, the classical Rayleigh-Ritz method uses global admissible functions and the finite element method uses local admissible functions, generally called interpolation functions. This difference has profound implications and is responsible for the success of the finite element method. In the first place, because the method works with small subdomains, i.e., the finite elements over which the local functions are defined, systems with very complex parameter distributions and irregular geometries can be accommodated. We recall that one of the weaknesses of the classical Rayleigh-Ritz method and the weighted residuals methods, such as Galerkin's method and the collocation method, is the difficulty in handling systems with irregular boundaries. Moreover, because the finite elements are generally very small, good representation of the motion can be achieved with interpolation functions in the form of low-degree polynomials. An important aspect of this is that the interpolation functions can be prescribed for given classes of systems, a clear advantage over the classical Rayleigh-Ritz method and Galerkins' method, for which the generation of suitable admissible and comparison functions, respectively, requires experience and ingenuity. Another important aspect is that the whole finite element spatial discretization process lends itself to easy computer coding. This includes the generation of a finite element mesh, the computation of the element mass matrix, stiffness matrix and nodal force vector and the assembly of these element quantities into global quantities. With the power of computers to carry out the various steps increasing at a dizzying rate, finite element models involving thousands of degrees of freedom are not uncommon.

The enthusiasm for the finite element method should be tempered somewhat by other considerations. Some of the advantages cited above are more important to static stress and structural analyses than to vibrations. In particular, stress analysis problems are more typical of three-dimensional systems, which tend to be bulkier and hence less prone to vibrate than one-dimensional and two-dimensional systems. Hence complex geometries are more frequently encountered in stress analysis than in vibrations. Then, for problems for which the classical Rayleigh-Ritz method is able to produce a solution, it is reasonable to expect that, for a given accuracy, such a solution requires appreciably fewer degrees of freedom than a solution produced by the finite element method.

On balance, however, the advantages of the finite element method far outweigh any disadvantages, as the method is capable of producing solutions where other methods fail. Already very popular, the use of the finite element method can only increase as vibration problems become progressively more complex.

PROBLEMS

(to be solved by the finite element method)

9.1 Solve Problem 8.1 using linear interpolation functions and determine the order of the problem required to match the accuracy of the lowest natural frequency computed in Problem 8.1.

9.2 Solve Problem 8.16 using linear interpolation functions and determine the order of the problem required to match the accuracy of the lowest natural frequency computed in Problem 8.16 with $n = 6$.

9.3 Solve Problem 9.2 with Problem 8.17 replacing Problem 8.16.

9.4 Solve Problem 9.1 using quadratic interpolation functions. Discuss convergence characteristics compared with the solution obtained in Problem 9.1.

9.5 Solve Problem 9.4 with Problem 9.2 replacing Problem 9.1.

9.6 Solve Problem 9.4 with Problem 9.3 replacing Problem 9.1.

9.7 Use the approach of Sec. 9.3 to derive the cubic elements given by Eqs. (9.35).

9.8 Use Eqs. (9.17) to derive the element stiffness and mass matrices for a nonuniform string fixed at $x = 0$ and supported by a spring at $x = L$ using cubic interpolation functions. Then, indicate the structure of the global stiffness and mass matrices and give the matrices in explicit form.

9.9 Solve Problem 9.4 using cubic interpolation functions.

9.10 Solve Problem 9.5 using cubic interpolation functions.

9.11 Solve Problem 9.6 using cubic interpolation functions.

9.12 Solve Problem 8.4 using Hermite cubics and determine the order of the problem required to match the accuracy of the lowest natural frequency computed in Problem 8.4.

9.13 Solve Problem 9.12 with Problem 8.18 with $n = 6$ replacing Problem 8.4.

9.14 Solve Problem 9.12 with Problem 8.19 with $n = 6$ replacing Problem 8.4.

9.15 Solve Problem 7.39 with $a = 2b = 4h$ using sixteen triangular elements in conjunction with linear interpolation functions. Compare the three lowest natural frequencies with those obtained in Problem 7.39 and draw conclusions.

9.16 Solve Problem 9.15 with quadratic interpolation functions instead of linear.

9.17 Use the approach of Sec. 9.5 to derive the cubic interpolation functions given by Eqs. (9.79).

9.18 Determine the element stiffness and mass matrices for the cubic interpolation functions given by Eqs. (9.79).

9.19 Use the approach of Sec. 9.5 to derive the bilinear interpolation functions given by Eqs. (9.81).

9.20 Solve Problem 9.15 using eight rectangular elements in conjunction with bilinear interpolation functions. Compare results with those obtained in Problem 9.15 and draw conclusions.

9.21 Determine the element stiffness and mass matrices for the serendipity elements given by Eqs. (9.92).

9.22 Use Eqs. (9.108) and (9.112) to determine the element stiffness matrix and mass matrix, respectively, for the quadrilateral membrane of Fig. 9.29 using the bilinear interpolation functions given by Eq. (9.113).

9.23 Determine the global stiffness and mass matrices for the membrane of Problem 9.22.

9.24 Determine the element stiffness and mass matrices for triangular plate elements using the interpolation functions given by Eqs. (9.133).

9.25 Determine the global stiffness and mass matrices for the plate of Problem 8.23 using results from Problem 9.24.

9.26 Derive the response of the shaft of Problem 9.3 to the distributed torque $m(x, t) = m_0 x(L - x)u(t)$, where $u(t)$ is the unit step function.

9.27 Solve Problem 9.26 with Problem 9.6 replacing Problem 9.3.

9.28 Derive the response of the cantilever beam of Problem 9.12 to the concentrated force $F(t) = F_0[u(t) - u(t - T)]$ applied at $x = L$, where $u(t)$ is the unit step function.

9.29 Derive the response of the beam of Problem 9.14 to the distributed force $f(x, t) = f_0(1 - x/2L)\,\delta(t)$, where $\delta(t)$ is the unit impulse.

9.30 Derive the response of the membrane of Problem 9.15 to the force per unit area $f(x, y, t) = f_0 y(b - y)u(t)$, where $u(t)$ is the unit step function.

9.31 Solve Problem 9.30 with Problem 9.16 replacing Problem 9.15.

9.32 Derive the response of the plate of Problem 9.25 to the force $f(x, y, t) = f_0[r(t) - r(t - T)]$ distributed uniformly over the rectangular area defined by $a/2 < x < 3a/4$, $b/4 < y < 3b/4$, where $r(t)$ is the unit ramp function (Sec. 1.7).

BIBLIOGRAPHY

1. Argyris, J. H. and Kelsey, S., *Energy Theorems and Structural Analysis*, Butterworth, London, 1960.
2. Babuska, I., Szabo, B. A. and Katz, I. N., "The p-Version of the Finite Element Method," *SIAM Journal of Numerical Analysis*, Vol. 18, No. 3, 1981, pp. 515–545.
3. Babuska, I. and Dorr, B. M., "Error Estimates for the Combined h and p Versions of the Finite Element Method," *Numerical Mathematics*, Vol. 37, 1981, pp. 257–277.
4. Bazeley, G. P., Cheung, Y. K., Irons, B. M. and Zienkiewicz, O. C., "Triangular Elements in Bending-Conforming and Non-Conforming Solutions," *Proceedings of the Conference on Matrix Methods in Structural Mechanics*, Wright-Patterson Air Force Base, OH, 1965.
5. Clough, R. W., "The Finite Element in Plane Stress Analysis," *Proceedings of the 2nd ASCE Conference on Electronic Computation*, Pittsburgh, PA, September 1960.
6. Clough, R. W., "The Finite Element Method in Structural Mechanics," in *Stress Analysis*, Eds.: O. C. Zienkiewicz and G. S. Hollister, Wiley, New York, 1965.
7. Clough, K. W. and Tocher, J. L., "Finite Element Stiffness Matrices for Analysis of Plates in Bending," *Proceedings of the Conference on Matrix Methods in Structural Mechanics*, Wright-Patterson Air Force Base, OH, 1965.
8. Courant, R., "Variational Methods for the Solution of Problems of Equilibrium and Vibrations," *Bulletin of the American Mathematical Society*, Vol. 49, January 1943, pp. 1–23.
9. Ergatoudis, J., Irons, B. M. and Zienkiewicz, O. C., "Curved Isoparametric Quadrilateral Elements for Finite Element Analysis," *International Journal for Solids and Structures*, Vol. 4, 1968, pp. 31–42.

10. Gallagher, R. H., *Finite Element Analysis: Fundamentals*, Prentice Hall, Englewood Cliffs, NJ, 1975.

11. Huebner, K. H. and Thornton, E. A., *The Finite Element Method for Engineers*, 2nd ed., Wiley, New York, 1983.

12. Irons, B. M., "Engineering Application of Numerical Integration in Stiffness Method," *AIAA Journal*, Vol. 14, 1966, pp. 2035–2037.

13. Meirovitch, L., *Computational Methods in Structural Dynamics*, Sijthoff and Noordhoff, Alphen aan den Rijn, The Netherlands, 1980.

14. Meirovitch, L. and Baruh, H., "On the Inclusion Principle for the Hierarchical Finite Element Method," *International Journal for Numerical Methods in Engineering*, Vol. 19, 1983, pp. 281–291.

15. Meirovitch, L. and Kwak, M. K., "On the Convergence of the Classical Rayleigh-Ritz Method and the Finite Element Method," *AIAA Journal*, Vol. 28, No. 8, 1990, pp. 1509–1516.

16. Melosh, R. J., "Basis for Derivation of Matrices for the Direct Stiffness Method," *AIAA Journal*, Vol. 1, No. 7, 1963, pp. 1631–1636.

17. Oliveira, E. A. de, "Theoretical Foundations of the Finite Element Method," *International Journal for Solids and Structures*, Vol. 4, 1968, pp. 929–952.

18. Peano, A., "Hierarchies of Conforming Finite Elements for Plane Elasticity and Plate Bending," *Computers and Mathematics with Applications*, Vol. 2, 1976, pp. 211–224.

19. Pian, T.H.H. and Tong, P., "Basis of Finite Element Methods for Solid Coutinua," *International Journal for Numerical Methods in Engineering*, Vol. 1, 1969, pp. 3–28.

20. Reddy, J. N., *An Introduction to the Finite Element Method*, McGraw-Hill, New York, 1984.

21. Shames, I. H. and Dym, C. L., *Energy and Finite Element Methods in Structural Mechanics*, McGraw-Hill, New York, 1985.

22. Strang, G. and Fix, G.J., *An Analysis of the Finite Element Method*, Prentice-Hall, Englewood Cliffs, NJ, 1973.

23. Szabo, B. A., "Some Recent Developments in the Finite Element Method, *Computers and Mathematics with Applications*, Vol. 5, 1979, pp. 99–115.

24. Taig, I.C. "Structural Analysis by the Matrix Displacement Method," *English Electric Aviation Report SD-17*, 1961.

25. Tong, P., Pian, H. H. and Bucciarelli, L. L., "Mode Shapes and Frequencies of the Finite Element Method Using Consistent and Lumped Masses," *Journal of Composite Structures*, Vol. 1, 1977, pp. 623–638.

26. Turner, M. J., Clough, R. W., Martin, H. C. and Topp, L. J., "Stiffness and Deflection Analysis of Complex Structures," *Journal of Aeronautical Sciences*, Vol. 23, 1956, pp. 805–823.

27. Zienkiewicz, O. C., Irons, B. M., Scott, F. E. and Campbell, J. S., "Three Dimensional Stress Analysis," *Proceedings of the IUTAM Symposium on High Speed Computing of Elastic Structures*, Liége, Belgium, 1970.

28. Zienkiewicz, O. C. and Taylor, R. L., *The Finite Element Method*, 4th ed., McGraw-Hill, London, 1991.

29. Zienkiewicz, O. C., Kelly, D. W., Gago, J. and Babuska, I., "Hierarchical Finite Element Approaches, Error Estimates and Adaptive Refinement," *Report C/R/382/81*, Institute for Numerical Methods in Engineering, University College of Swansea, UK, 1981.

A

ELEMENTS OF LAPLACE TRANSFORMATION

The solution of difficult mathematical problems can often be simplified greatly by means of a suitable transformation. The process consists of transforming a difficult problem into a simpler problem, solving the simpler problem and inverse transforming the solution to the simpler problem to obtain the solution to the original problem. Transformations are widely used in vibrations. In particular, the Laplace transformation can be used to obtain efficient solutions to ordinary differential equations of motion for linear time-invariant systems, also known as systems with constant coefficients.

A.1 INTEGRAL TRANSFORMATIONS

We consider a function $f(t)$ defined by an ordinary differential equation and certain initial conditions and propose to obtain a solution by transforming the problem for $f(t)$ into a problem for $F(s)$ given by the *integral transformation*

$$F(s) = \int_a^b f(t)K(s,t)\,dt \tag{A.1}$$

where $K(s,t)$ is a given function of s and t called the *kernel* of the transformation. When the limits a and b are finite, $F(s)$ is a finite transformation. Such an integral transformation converts the differential equation into an algebraic equation in terms of the transformed function $F(s)$, where s is a parameter. In the process, initial conditions are accounted for automatically. The algebraic equation for $F(s)$ can in general be solved without much difficulty, and the function $f(t)$ is obtained by inverse transforming $F(s)$.

Integral transform methods can also be used to solve boundary-value problems defined by partial differential equations. In such a case one transformation reduces the number of independent variables by one. Hence, if two independent variables are involved, instead of solving a partial differential equation one must solve an ordinary differential equation, which is in general a considerably easier task.

The main difficulty in using integral transform methods lies in the inverse transformation, which for the most part involves evaluation of an integral. The transform and its inverse constitute a *transform pair*, and for many transformations in use there are tables of transform pairs available. More often than not, it is possible to find the inverse transformation in transform tables, thus eliminating the need to evaluate an inversion integral. One of the most widely used integral transformations is the Laplace transformation.

A.2 THE LAPLACE TRANSFORMATION

The Laplace transformation method provides a very convenient means of solving linear ordinary differential equations for linear time-invariant systems, or systems with constant coefficients. Such equations arise frequently in the study of vibrations of discrete linear systems. This type of problem was treated by Heaviside by means of his operational calculus. Much of Heaviside's work was based on intuition and the mathematical treatment was often obscure. The Laplace transformation method, although similar to Heaviside's operational calculus, is mathematically rigorous. The method accounts automatically for initial conditions and provides a great deal of insight into the physics of the system.

We consider a function $f(t)$ given for all values of time larger than zero, $t \geq 0$, and define the *unilateral Laplace transformation* of $f(t)$ as

$$\mathcal{L}f(t) = F(s) = \int_0^\infty e^{-st} f(t)\, dt \tag{A.2}$$

We note that the kernel of the transformation is $K(s, t) = e^{-st}$, where s is a *subsidiary variable*, which is in general a complex quantity. The complex plane defined by s is referred to as the *s-plane*, or the *Laplace plane*.

The function $f(t)$ is subject to certain restrictions, because the integral, Eq. (A.2), must converge. If $f(t)$ is such that

$$|e^{-st} f(t)| < Ce^{-(s-a)t}, \qquad \operatorname{Re} s > a \tag{A.3}$$

where C is a constant and Re s denotes the real part of s, then the Laplace transformation of $f(t)$ exists. Condition (A.3) implies that $f(t)$ does not increase more rapidly than Ce^{at} with increasing t. Such a function $f(t)$ is said to be of *exponential order* and is denoted $f(t) = O(e^{at})$. Another condition for the existence of the Laplace transformation is that $f(t)$ be piecewise continuous, which means that in a given interval it has a finite number of finite discontinuities and no infinite discontinuity. Most functions describing physical phenomena satisfy these conditions.

A.3 TRANSFORMATION OF DERIVATIVES

Our interest lies in solving ordinary differential equations by means of the Laplace transformation method, which requires the transformation of derivatives of $f(t)$. To this end, we assume that the Laplace transformation of $f(t)$ exists and consider first the transform of $df(t)/dt$. Integrating by parts, we can write

$$\begin{aligned}\mathcal{L}\frac{df(t)}{dt} &= \int_0^\infty e^{-st}\frac{df(t)}{dt}dt \\ &= e^{-st}f(t)\Big|_0^\infty - \int_0^\infty (-se^{-st})f(t)\,dt = -f(0) + sF(s)\end{aligned} \tag{A.4}$$

In general, for the nth derivative of $f(t)$, we obtain

$$\mathcal{L}\frac{d^n f(t)}{dt^n} = \mathcal{L}f^{(n)}(t) = -f^{(n-1)}(0) - sf^{(n-2)}(0) - \dots - s^{n-1}f(0) + s^n F(s) \tag{A.5}$$

where we adopted the notation

$$\left.\frac{d^{(n-j)}f(t)}{dt^{n-j}}\right|_{t=0} = f^{(n-j)}(0) \tag{A.6}$$

Equation (A.5) is valid only if $f(t)$ satisfies the conditions prescribed in Sec. A.3 and all its derivatives through the $(n-1)$st are continuous.

A.4 TRANSFORMATION OF ORDINARY DIFFERENTIAL EQUATIONS

The differential equation of motion of a mass-damper-spring system is (Sec. 2.3)

$$m\frac{d^2x(t)}{dt^2} + c\frac{dx(t)}{dt} + kx(t) = f(t) \tag{A.7}$$

where m, c and k are the mass, coefficient of viscous damping and spring constant, respectively, $x(t)$ is the displacement of m and $F(t)$ the force acting on m. Transforming both sides of Eq. (A.7) and using Eq. (A.5), we can write

$$m[s^2X(s) - sx(0) - \dot{x}(0)] + c[sX(s) - x(0)] + kX(s) = F(s) \tag{A.8}$$

where $x(0)$ and $\dot{x}(0)$ are the initial displacement and velocity, respectively. Solving Eq. (A.8) for $X(s)$, we obtain

$$X(s) = \frac{1}{ms^2 + cs + k}F(s) + \frac{ms + c}{ms^2 + cs + k}x(0) + \frac{m}{ms^2 + cs + k}\dot{x}(0) \tag{A.9}$$

Equation (A.9) is called the *subsidiary equation* of the differential equation, Eq. (A.7). To obtain the response $x(t)$, we must evaluate the inverse transformation of $X(s)$. Note that the Laplace transformation method provides automatically for initial conditions.

A.5 THE INVERSE TRANSFORMATION

Equation (A.9) gives the transform $X(s)$ of $x(t)$, which is a function of s. To obtain the actual solution $x(t)$ of Eq. (A.7), we must carry out an inverse transformation, which is denoted symbolically as

$$\mathcal{L}^{-1}X(s) = x(t) \tag{A.10}$$

meaning that the inverse transform of $X(s)$ is $x(t)$.

In general, the operation $\mathcal{L}^{-1}F(s)$ involves the evaluation of the integral

$$f(t) = \mathcal{L}^{-1}F(s) = \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{st}F(s)\,ds \tag{A.11}$$

where the path of integration is a line parallel to the imaginary axis crossing the real axis at Re $s = \gamma$ and extending from $-\infty$ to $+\infty$. In many cases, however, we can carry out the inverse Laplace transformation without having to recourse to line integrals. This is the case when Jordan's lemma can be used to replace the line integral by a closed contour integral, which in turn can be evaluated by means of the residue theorem (Ref. 1, Sec. 9-15). In most cases, however, it is possible to obtain the inverse transformation by means of tables of Laplace transform pairs (see Sec. A.9). Quite often, to increase the usefulness of the tables, we can use the method of partial fractions to reduce seemingly complicated functions to a form listed in tables.

A.6 SHIFTING THEOREMS

We consider the function

$$f_1(t) = f(t)e^{at} \tag{A.12}$$

where a is a real or complex number, and evaluate its Laplace transformation as follows:

$$\begin{aligned} F_1(s) &= \int_0^\infty [f(t)e^{at}]e^{-st}\,dt \\ &= \int_0^\infty f(t)e^{-(s-a)t}\,dt = F(s-a) \end{aligned} \tag{A.13}$$

It follows that

$$\mathcal{L}f(t)e^{at} = F(s-a) \tag{A.14}$$

Hence, the effect of multiplying $f(t)$ by e^{at} in the real domain is to shift the transform $F(s)$ of $f(t)$ by an amount a in the s-domain. Because the s-domain is a complex plane, this theorem is also called the *complex shifting theorem.*

Next, we consider the Laplace transformation

$$F(s) = \int_0^\infty e^{-s\lambda}f(\lambda)\,d\lambda \tag{A.15}$$

and let $\lambda = t - a$, so that

$$F(s) = \int_0^\infty e^{-s(t-a)} f(t-a)dt = e^{as} \int_0^\infty e^{-st}[f(t-a)u(t-a)]dt \quad \text{(A.16)}$$

where $u(t-a)$ is the unit step function initiated at $t = a$. Multiplying both sides of Eq. (A.16) by e^{-as}, we obtain

$$e^{-as}F(s) = \int_0^\infty e^{-st}[f(t-a)u(t-a)]dt \quad \text{(A.17)}$$

Hence,

$$\mathcal{L}^{-1}e^{-as}F(s) = f(t-a)u(t-a) \quad \text{(A.18)}$$

Equation (A.18) represents the *shifting theorem in the real domain.*

A.7 METHOD OF PARTIAL FRACTIONS

Under consideration is the inverse transformation of the function

$$F(s) = \frac{A(s)}{B(s)} \quad \text{(A.19)}$$

where both $A(s)$ and $B(s)$ are polynomials in s, and we recall that Eq. (A.9) is of this type. In general, $B(s)$ is a polynomial of higher degree than $A(s)$. We assume that $B(s)$ is a polynomial of nth degree and write it in the factored form

$$B(s) = (s - a_1)(s - a_2)\dots(s - a_n) \quad \text{(A.20)}$$

in which $a_1, a_2, \dots, a_n$ are the roots. We consider first the case in which all n roots are distinct, in which case Eq. (A.19) can be expressed as

$$F(s) = \frac{A(s)}{B(s)} = \frac{C_1}{s - a_1} + \frac{C_2}{s - a_2} + \dots + \frac{C_n}{s - a_n} \quad \text{(A.21)}$$

where the coefficients C_k are given by

$$C_k = \lim_{s \to a_k}[(s - a_k)F(s)] = \left.\frac{A(s)}{B'(s)}\right|_{s=a_k} \quad \text{(A.22)}$$

in which B' is the derivative of B with respect to s. But, using the table of Laplace transforms in Sec. A.9 and considering the complex shifting theorem, we have

$$\mathcal{L}^{-1}\frac{1}{s - a_k} = e^{a_k t} \quad \text{(A.23)}$$

so that the inverse transform of Eq. (A.21) is

$$\begin{aligned} f(t) &= \mathcal{L}^{-1}F(s) = C_1e^{a_1t} + C_2e^{a_2t} + \ldots + C_ne^{a_nt} \\ &= \sum_{k=1}^{n} C_ke^{a_kt} \\ &= \sum_{k=1}^{n} \lim_{s\to a_k}[(s-a_k)F(s)e^{st}] = \sum_{k=1}^{n} \frac{A(s)}{B'(s)}e^{st}\bigg|_{s=a_k} \end{aligned} \tag{A.24}$$

The roots $a_1, a_2, \ldots, a_n$ are called *simple poles* of $F(s)$. It should be noted that Eq. (A.24) can be used for functions other than ratios of two polynomials, provided the function has simple poles. Poles are points at which the function $F(s)$ becomes infinite.

Next, we consider the case in which $B(s)$ has a multiple root of order k, i.e., $F(s)$ has a pole of order k as opposed to poles of first order, or simple poles, examined previously. Hence, we consider

$$B(s) = (s-a_1)^k(s-a_2)(s-a_3)\ldots(s-a_n) \tag{A.25}$$

The partial-fractions expansion in this case is of the form

$$\begin{aligned} F(s) = \frac{A(s)}{B(s)} &= \frac{C_{11}}{(s-a_1)^k} + \frac{C_{12}}{(s-a_1)^{k-1}} + \ldots + \frac{C_{1k}}{s-a_1} \\ &\quad + \frac{C_2}{s-a_2} + \frac{C_3}{s-a_3} + \ldots + \frac{C_n}{s-a_n} \end{aligned} \tag{A.26}$$

It is not difficult to show that the coefficients corresponding to the repeated root a_1 are

$$C_{1r} = \frac{1}{(r-1)!}\frac{d^{r-1}}{ds^{r-1}}[(s-a_1)^kF(s)]_{s=a_1}, \qquad r = 1, 2, \ldots, k \tag{A.27}$$

The simple poles of $F(s)$ are treated by means of (A.22), as previously. For the higher-order pole, we use the table of Laplace transforms in conjunction with the complex shifting theorem and write

$$\mathcal{L}^{-1}\frac{1}{(s-a_1)^r} = \frac{t^{r-1}}{(r-1)!}e^{a_1t} \tag{A.28}$$

so that the inverse transform of Eq. (A.26) becomes

$$\begin{aligned} f(t) &= \left[C_{11}\frac{t^{k-1}}{(k-1)!} + C_{12}\frac{t^{k-2}}{(k-2)!} + \ldots + C_{1k}\right]e^{a_1t} \\ &\quad + C_2e^{a_2t} + C_3e^{a_3t} + \ldots + C_ne^{a_nt} \end{aligned} \tag{A.29}$$

Equation (A.29) can be shown to be equal to

$$f(t) = \frac{1}{(k-1)!}\frac{d^{k-1}}{ds^{k-1}}[(s-a_1)^kF(s)e^{st}]_{s=a_1} + \sum_{i=2}^{n}[(s-a_i)F(s)e^{st}]_{s=a_1} \tag{A.30}$$

A.8 THE CONVOLUTION INTEGRAL

We consider two functions $f_1(t)$ and $f_2(t)$, both defined for $t \geq 0$, and assume that $f_1(t)$ and $f_2(t)$ possess the Laplace transforms $F_1(s)$ and $F_2(s)$, respectively. Then, we consider the integral

$$x(t) = \int_0^t f_1(\tau) f_2(t - \tau)\, d\tau = \int_0^\infty f_1(\tau) f_2(t - \tau)\, d\tau \tag{A.31}$$

The function $x(t)$ given by these integrals, sometimes denoted by $x(t) = f_1(t) * f_2(t)$, is called the *convolution* of the functions f_1 and f_2 over the interval $(0, \infty)$. The validity of the change in the upper limit of the first integral can be explained by the fact $f_2(t - \tau) = 0$ for $\tau > t$, or $t - \tau < 0$. Transforming both sides of Eq. (A.31), we obtain

$$\begin{aligned} X(s) &= \int_0^\infty e^{-st} \left[\int_0^\infty f_1(\tau) f_2(t - \tau)\, d\tau \right] dt \\ &= \int_0^\infty f_1(\tau)\, d\tau \int_0^\infty e^{-st} f_2(t - \tau)\, dt \\ &= \int_0^\infty f_1(\tau)\, d\tau \int_\tau^\infty e^{-st} f_2(t - \tau)\, dt \end{aligned} \tag{A.32}$$

where the limit in the second integral was changed because $f_2(t - \tau) = 0$ for $t < \tau$.

Next, we let $t - \tau = \lambda$ in the second integral and note that $\lambda = 0$ for $t = \tau$, so that

$$\begin{aligned} X(s) &= \int_0^\infty f_1(\tau)\, d\tau \int_\tau^\infty e^{-st} f_2(t - \tau)\, dt \\ &= \int_0^\infty f_1(\tau)\, d\tau \int_0^\infty e^{-s(\tau+\lambda)} f_2(\lambda)\, d\lambda \\ &= \int_0^\infty e^{-s\tau} f_1(\tau)\, d\tau \int_0^\infty e^{-s\lambda} f_2(\lambda)\, d\lambda = F_1(s) F_2(s) \end{aligned} \tag{A.33}$$

From Eqs. (A.31) and (A.33), it follows that

$$\begin{aligned} x(t) &= \mathcal{L}^{-1} X(s) \\ &= \mathcal{L}^{-1} F_1(s) F_2(s) \\ &= \int_0^t f_1(\tau) f_2(t - \tau)\, d\tau = \int_0^t f_1(t - \tau) f_2(\tau)\, d\tau. \end{aligned} \tag{A.34}$$

This is true because it does not matter which of the functions $f_1(t)$ and $f_2(t)$ is shifted. Equation (A.34) can be stated in words as the following theorem: *The inverse transformation of the product of two transforms is equal to the convolution of their inverse transforms.* The integrals in (A.34) are called *convolution integrals.* A special case of Eq. (A.34) was encountered in Sec. 1.7 without any reference to Laplace transformations.

A.9 TABLE OF LAPLACE TRANSFORM PAIRS

$f(t)$	$F(s)$
$\delta(t)$(Dirac delta function)	1
$u(t)$(unit step function)	$\frac{1}{s}$
$t^n \quad n = 1, 2, \ldots$	$\frac{n!}{s^{n+1}}$
$e^{-\omega t}$	$\frac{1}{s+\omega}$
$te^{-\omega t}$	$\frac{1}{(s+\omega)^2}$
$\cos \omega t$	$\frac{s}{s^2+\omega^2}$
$\sin \omega t$	$\frac{\omega}{s^2+\omega^2}$
$\cosh \omega t$	$\frac{s}{s^2-\omega^2}$
$\sinh \omega t$	$\frac{\omega}{s^2-\omega^2}$
$1 - e^{-\omega t}$	$\frac{\omega}{s(s+\omega)}$
$1 - \cos \omega t$	$\frac{\omega^2}{s(s^2+\omega^2)}$
$\omega t - \sin \omega t$	$\frac{\omega^3}{s^2(s^2+\omega^2)}$
$\omega t \ \cos \omega t$	$\frac{\omega(s^2-\omega^2)}{(s^2+\omega^2)^2}$
$\omega t \ \sin \omega t$	$\frac{2\omega^2 s}{(s^2+\omega^2)^2}$
$\frac{1}{(1-\zeta^2)^{1/2}\omega} e^{-\zeta\omega t} \sin(1-\zeta^2)^{1/2}\omega t$	$\frac{1}{s^2+2\zeta\omega s+\omega^2}$
$e^{-\zeta\omega t}\left[\cos(1-\zeta^2)^{1/2}\omega t + \frac{\zeta}{(1-\zeta^2)^{1/2}} \sin(1-\zeta^2)^{1/2}\omega t\right]$	$\frac{s+2\zeta\omega}{s^2+2\zeta\omega s+\omega^2}$

BIBLIOGRAPHY

1. Kaplan, W., *Advanced Calculus*, 2nd ed., Addison-Wesley, Reading, MA, 1973.

B

ELEMENTS OF LINEAR ALGEBRA

The vibration of linear discrete systems is governed by sets of simultaneous linear ordinary differential equations. The solution of such sets of equations can be obtained most conveniently by means of a linear transformation rendering the set of equations independent. To find this linear transformation, it is necessary to solve a set of homogeneous algebraic equations containing a certain parameter. The problem of determining the values of the parameter such that the set of equations admits nontrivial solutions is known as the algebraic eigenvalue problem, a very important problem in linear algebra. This appendix is devoted to concepts from linear algebra of interest to the study of vibrations, such as vector spaces and matrices.

B.1 LINEAR VECTOR SPACES

In discussing vector spaces, it proves convenient to introduce the concept of a field. A *field* is defined as a set of scalars possessing certain algebraic properties. The real numbers constitute a field, and so do the complex numbers.

We consider a set of elements F such that for any two elements α and β in F it is possible to define another two unique elements belonging to F, the first denoted by $\alpha + \beta$ and called the *sum* of α and β, and the second denoted by $\alpha\beta$ and called the *product* of α and β. The set F is called a *field* if these two operations satisfy the five field postulates:

1. *Commutative laws.* For all α and β in F,
   (i) $\alpha + \beta = \beta + \alpha$, (ii) $\alpha\beta = \beta\alpha$.
2. *Associative laws.* For all α, β and γ in F,
   (i) $(\alpha + \beta) + \gamma = \alpha(\beta + \gamma)$, (ii) $(\alpha\beta)\gamma = \alpha(\beta\gamma)$.
3. *Distributive laws.* For all α, β and γ in F,
   $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
4. *Identity elements.* There exist in F elements 0 and 1 called the zero and the unity elements, respectively, such that $0 \neq 1$, and for all α in F,
   (i) $\alpha + 0 = \alpha$, (ii) $1\alpha = \alpha$.
5. *Inverse elements.*
   i. For every element α in F there exists a unique element $-\alpha$, called the additive inverse of α, such that $\alpha + (-\alpha) = 0$.

ii. For element $\alpha \neq 0$ in F there exists a unique element α^{-1}, called the multiplicative inverse of α, such that $\alpha\alpha^{-1} = 1$.

Next, we define the concept of *linear vector space*, also referred to as *linear space* and *vector space*. Let L be a set of elements called *vectors* and F a field of *scalars*. Then, if L and F are such that two operations, namely, *vector addition* and *scalar multiplication*, are defined for L and F, the set of vectors together with the two operations are called a *linear vector space L over a field F*. For every two elements $\mathbf{x}$ and $\mathbf{y}$ in L, it satisfies the postulates:

1. *Commutativity.* $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
2. *Associativity.* $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$.
3. There exists a unique vector $\mathbf{0}$ in L such that $\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$.
4. For every vector $\mathbf{x}$ in L there exists a unique vector $-\mathbf{x}$ such that $\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0}$.

Hence, the rules of vector addition are similar to those of ordinary algebra. Moreover, for any vector $\mathbf{x}$ in L and any scalar α in F, there is a unique *scalar product* $\alpha\mathbf{x}$ which is also an element of L. The scalar multiplication must be such that, for all α and β in F and all $\mathbf{x}$ and $\mathbf{y}$ in L, it satisfies the postulates:

5. *Associativity.* $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$.
6. *Distributivity.* (i) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$, (ii) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$.
7. $1\mathbf{x} = \mathbf{x}$, where 1 is the unit scalar, and $0\mathbf{x} = \mathbf{0}$.

We have considerable interest in a vector space L possessing n elements of the field F, i.e., in a vector space of n-tuples. We write any two such vectors in L as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{B.1}$$

and refer to them as *n-vectors*. The set of all n-vectors is called the *vector space* L^n. Then, the addition of these two vectors is defined as

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \cdots\cdots\cdot \\ x_n + y_n \end{bmatrix} \tag{B.2}$$

Moreover, if α is a scalar in F, then the product of a scalar and a vector is defined as

$$\alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \cdots \\ \alpha x_n \end{bmatrix} \tag{B.3}$$

Let S be a subset of the vector space L. Then, S is a *subspace* of L if the following statements are true:

1. If $\mathbf{x}$ and $\mathbf{y}$ are in S, then $\mathbf{x}+\mathbf{y}$ is in S.
2. If $\mathbf{x}$ is in S and α is in F, then $\alpha\mathbf{x}$ is in S.

B.2 LINEAR DEPENDENCE

We consider a set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ in a linear space L and a set of scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$ in F. Then, the vector $\mathbf{x}$ given by

$$\mathbf{x} = \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_n\mathbf{x}_n \tag{B.4}$$

is said to be a *linear combination* of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ with coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$. The vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be *linearly independent* if the relation

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_n\mathbf{x}_n = \mathbf{0} \tag{B.5}$$

can be satisfied only for the trivial case, i.e., only when all the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ are identically zero. If relation (B.5) is satisfied and at least one of the coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ is different from zero, then the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be *linearly dependent*, with the implication that one vector is a linear combination of the remaining $n-1$ vectors.

The subspace S of L consisting of all the linear combinations of the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ is called a subspace *spanned* by the vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$. If $S = L$, then $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to *span L*.

Example B.1

Consider the two independent vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix} \tag{a}$$

in a three-dimensional space. The set of all linear combinations of $\mathbf{x}_1$ and $\mathbf{x}_2$ span a plane passing through the origin and the tips of $\mathbf{x}_1$ and $\mathbf{x}_2$. The three vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 5 \\ 0 \\ 5 \end{bmatrix} \tag{b}$$

span the same plane, because $\mathbf{x}_3$ lies in the plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$. Hence, the three vectors are linearly dependent. Indeed, it can be easily verified that

$$\mathbf{x}_1 + 2\mathbf{x}_2 - \mathbf{x}_3 = \mathbf{0} \tag{c}$$

so that $\mathbf{x}_3$ is really a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$ (see Fig. B.1).

On the other hand, the three vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 5 \\ 1 \\ 5 \end{bmatrix} \tag{d}$$

are linearly independent because

$$\alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \alpha_4\mathbf{x}_4 \neq \mathbf{0} \tag{e}$$

for all cases other than the trivial one. The three vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_4$ span a three-dimensional space.

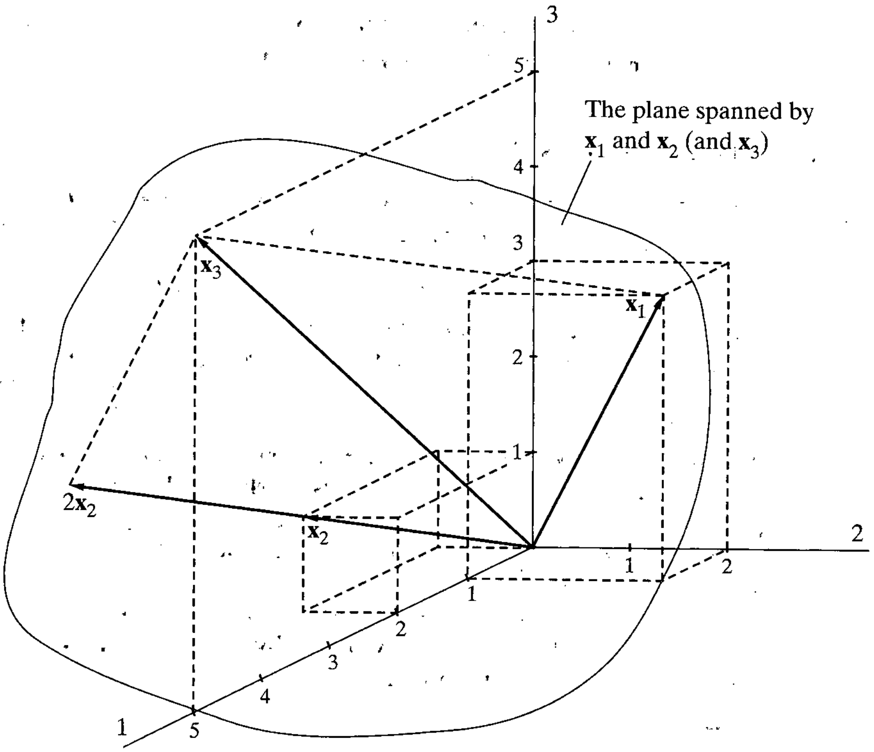

3
5
4
3
2
1
2
1
2
1
2
3
4
5
1
$\mathbf{x}_1$
$\mathbf{x}_2$
$\mathbf{x}_3$
$2\mathbf{x}_2$
The plane spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$ (and $\mathbf{x}_3$)

Figure B.1 Linearly dependent vectors

B.3 BASES AND DIMENSION OF A VECTOR SPACE

A vector space L over F is said to be finite-dimensional if there exists a finite set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ that span L, i.e., such that every vector in L is a linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

Let L be a vector space over F. A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ that span L is called a *generating system* of L. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent and span L, then the generating system is called a *basis* for L. If L is a finite-dimensional vector space, any two bases for L contain the same number of vectors. The basis can be regarded as the generalization of the concept of a coordinate system.

Let L be a finite-dimensional vector space over F. The *dimension* of L is defined as the number of vectors in any basis for L. This integer is denoted by dim L. The vector space L^n is spanned by n linearly independent vectors, so that dim $L^n = n$.

Consider an arbitrary n-vector $\mathbf{x}$ in L^n with components $x_1, x_2, \ldots, x_n$ and introduce a set of n-vectors given by

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{B.6}$$

The vector $\mathbf{x}$ can be written in terms of the vectors $\mathbf{e}_i$ $(i = 1, 2, \ldots n)$ as the linear combination

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n. = \sum_{i=1}^{n} x_i\mathbf{e}_i \tag{B.7}$$

It follows that L^n is spanned by the set of vectors $\mathbf{e}_i$ $(i = 1, 2, \ldots, n)$, so that the vectors $\mathbf{e}_i$ constitute a generating system of L^n. The set of vectors $\mathbf{e}_i$ can be verified as being linearly independent and they are generally referred to as the *standard basis for L^n*.

Example B.2

The vectors $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_4$ of Example B.1 form a basis for a three-dimensional vector space. Any vector $\mathbf{x}$ in L^3 can be written as a unique linear combination of $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_4$. For example, it can be verified that the vector

$$\mathbf{x} = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix} \tag{a}$$

can be represented in the form

$$\mathbf{x} = 2\mathbf{x}_1 + 3\mathbf{x}_2 - \mathbf{x}_4 \tag{b}$$

The same vector $\mathbf{x}$ can be also represented in the terms of the standard basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ for L^3. Indeed, it is easy to see that

$$\mathbf{x} = 3\mathbf{e}_1 + 0\mathbf{e}_2 + 4\mathbf{e}_3 \tag{c}$$

B.4 INNER PRODUCTS AND ORTHOGONAL VECTORS

Various concepts encountered in two- and three-dimensional spaces, such as the length of a vector and orthogonality, can be generalized to n-dimensional spaces. This requires the introduction of additional definitions.

Let L^n be an n-dimensional vector space defined over the field F of scalars. If to each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in L^n is assigned a unique scalar in F, called the *inner product* of $\mathbf{x}$ and $\mathbf{y}$, then L^n is said to be an *inner product space*. The vectors $\mathbf{x}$ and $\mathbf{y}$ can be complex, in which case $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ denote their complex conjugates. The inner product is denoted by $(\mathbf{x}, \mathbf{y})$ and must satisfy the following postulates:

1. $(\mathbf{x}, \mathbf{x}) \geq 0$ for all $\mathbf{x}$ in L^n and $(\mathbf{x}, \mathbf{x}) = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
2. $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$.
3. $(\lambda\mathbf{x}, \mathbf{y}) = \lambda(\mathbf{x}, \mathbf{y})$ and $(\mathbf{x}, \lambda\mathbf{y}) = \bar{\lambda}(\mathbf{x}, \mathbf{y})$ for all λ in F.
4. $(\mathbf{x}, \mathbf{y} + \mathbf{z}) = (\mathbf{x}, \mathbf{y}) + (\mathbf{x}, \mathbf{z})$ for all $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ in L^n.

The most common definition of the *complex inner product* is

$$(\mathbf{x}, \mathbf{y}) = x_1\bar{y}_1 + x_2\bar{y}_2 + \ldots + x_n\bar{y}_n \tag{B.8}$$

which represents a complex number. When $\mathbf{x}$ and $\mathbf{y}$ are real vectors, Eq. (B.8) reduces to

$$(\mathbf{x}, \mathbf{y}) = x_1y_1 + x_2y_2 + \ldots + x_ny_n \tag{B.9}$$

which defines the *real inner product*, a real number. A finite-dimensional inner product space defined over the real scalar field is called a *Euclidean space.*

It is often desirable to have a measure of the size of a vector. Such a measure is called the *norm.* It is designated by the symbol $\|\mathbf{x}\|$ and is required to possess the following properties:

1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$
2. $\|\lambda \mathbf{x}\| = |\lambda| \|\mathbf{x}\|$ for any scalar λ
3. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$

where property 3 is known as the triangle inequality. Note that $|\lambda|$ denotes the absolute value, or modulus of λ.

A commonly used norm is the *quadratic norm*

$$\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})^{1/2} \tag{B.10}$$

which defines the *length* of the vector $\mathbf{x}$. In the case of real vector spaces, Eq. (B.10) reduces to

$$\|\mathbf{x}\| = \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} \tag{B.11}$$

which defines the *Euclidean norm.* Equation (B.11) can be recognized as the extension to n dimensions of the ordinary concept of length of a vector in two and three dimensions.

A vector whose norm is equal to unity, $\|\mathbf{x}\| = (\mathbf{x}, \mathbf{x})^{1/2} = 1$, is called a *unit vector.* Any nonzero vector can be *normalized* so as to form a unit vector by simply dividing the vector by its norm

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|} \tag{B.12}$$

It is easy to verify that the vectors $\mathbf{e}_i$ defined by Eqs. (B.6) are unit vectors.

When the vectors $\mathbf{x}$ and $\mathbf{y}$ are real, the inner product is sometimes referred to as the *dot product.* We recall from ordinary vector analysis that the dot product of two vectors in the two- and three-dimensional space can be used to define the cosine of the angle between the two vectors. This concept can be generalized to the n-dimensional space by writing

$$\cos\theta = \frac{(\mathbf{x}, \mathbf{y})}{\|\mathbf{x}\| \|\mathbf{y}\|} = (\hat{\mathbf{x}}, \hat{\mathbf{y}}) \tag{B.13}$$

Any two vectors $\mathbf{x}$ and $\mathbf{y}$ in L^n are said to be *orthogonal* if and only if

$$(\mathbf{x}, \mathbf{y}) = 0 \tag{B.14}$$

which represents a generalization of the ordinary concept of perpendicularity. If each pair of vectors in a given set are mutually orthogonal, then the set is said to be an *orthogonal set.* If, in addition, the vectors have unit norms, the vectors are said to be *orthonormal. Any set of mutually orthogonal nonzero vectors in L^n is* *linearly independent.* To show this, we assume that the orthogonal set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ satisfies a relation of the type (B.5) and form the inner products

$$\begin{aligned} 0 = (\mathbf{x}_i, \mathbf{0}) &= (\mathbf{x}_i, \alpha_1\mathbf{x}_1 + \alpha_2\mathbf{x}_2 + \ldots + \alpha_n\mathbf{x}_n) \\ &= \alpha_1(\underline{f}x_i, \mathbf{x}_1) + \alpha_2(\mathbf{x}_i, \mathbf{x}_2) + \ldots + \alpha_n(\mathbf{x}_i, \mathbf{x}_n) \\ &= \alpha_i(\mathbf{x}_i, \mathbf{x}_i), \qquad i = 1, 2, \ldots, n \end{aligned} \tag{B.15}$$

Because $(\mathbf{x}_i, \mathbf{x}_i) \neq 0$, it follows that Eqs. (B.15) can be satisfied if and only if all the coefficients α_i are identically zero, so that the set of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ must be linearly independent. Owing to the independence property, orthogonal vectors, and in particular orthonormal vectors, are convenient choices for basis vectors. A classical example of an orthonormal set of vectors used as a basis are the unit vectors $\mathbf{e}_i$, which explains why these vectors are referred to as a standard basis for L^n.

B.5 THE GRAM-SCHMIDT ORTHOGONALIZATION PROCESS

Orthogonal vectors are by definition independent, but independent vectors are not necessarily orthogonal. A set of independent vectors, however, can be rendered orthogonal. In computational work, it is often desirable to work with a set of orthogonal vectors, so that the procedure for rendering independent vectors orthogonal is of special interest. The procedure is known as the *Gram-Schmidt orthogonalization process.*

We consider the set of independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ and denote the desired orthogonal vectors by $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$. These latter vectors can be normalized by dividing each of the vectors by its norm, so that the orthonormal vectors $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_n$ are given by

$$\hat{\mathbf{y}}_i = \mathbf{y}_i/\|\mathbf{y}_i\|, \qquad i = 1, 2, \ldots, n \tag{B.16}$$

The first vector of the desired orthonormal set is simply

$$\hat{\mathbf{y}}_1 = \hat{\mathbf{x}}_1 = \mathbf{x}_1/\|\mathbf{x}_1\| \tag{B.17}$$

The second vector, $\mathbf{y}_2$, must be orthogonal to $\hat{\mathbf{y}}_1$. A vector $\mathbf{y}_2$ satisfying this condition can be taken in the form

$$\mathbf{y}_2 = \mathbf{x}_2 - (\mathbf{x}_2, \hat{\mathbf{y}}_1)\hat{\mathbf{y}}_1 \tag{B.18}$$

Indeed, we have

$$(\mathbf{y}_2, \hat{\mathbf{y}}_1) = (\mathbf{x}_2, \hat{\mathbf{y}}_1) - (\mathbf{x}_2, \hat{\mathbf{y}}_1) = 0 \tag{B.19}$$

Of course, the vector $\mathbf{y}_2$ can be normalized by using the second of Eqs. (B.16) to obtain $\hat{\mathbf{y}}_2$. The third vector, $\mathbf{y}_3$, can be written in the form

$$\mathbf{y}_3 = \mathbf{x}_3 - (\mathbf{x}_3, \hat{\mathbf{y}}_1)\hat{\mathbf{y}}_1 - (\mathbf{x}_3, \hat{\mathbf{y}}_2)\hat{\mathbf{y}}_2 \tag{B.20}$$

which is orthonormal to $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$, as it satisfies

$$\begin{aligned} (\mathbf{y}_3, \hat{\mathbf{y}}_1) &= (\mathbf{x}_3, \hat{\mathbf{y}}_1) - (\mathbf{x}_3, \hat{\mathbf{y}}_1) = 0 \\ (\mathbf{y}_3, \hat{\mathbf{y}}_2) &= (\mathbf{x}_3, \hat{\mathbf{y}}_2) - (\mathbf{x}_3, \hat{\mathbf{y}}_2) = 0 \end{aligned} \tag{B.21}$$

The vector $\mathbf{y}_3$ can be normalized to obtain $\hat{\mathbf{y}}_3$. Generalizing, we can write

$$\mathbf{y}_i = \mathbf{x}_i - \sum_{j=1}^{i-1} (\mathbf{x}_i, \hat{\mathbf{y}}_j)\hat{\mathbf{y}}_j, \qquad i = 1, 2, \ldots, n \tag{B.22}$$

which can be used to compute $\hat{\mathbf{y}}_i$. Clearly, $\hat{\mathbf{y}}_i$ is orthonormal to $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_{i-1}$ $(i = 1, 2, \ldots, n)$. The process is concluded with the computation of $\hat{\mathbf{y}}_n$.

The Gram-Schmidt process described above can often yield computed vectors that are far from being orthogonal (Ref. 3, p. 148). An orthogonalization process mathematically equivalent but computationally superior to the Gram-Schmidt process is the *modified Gram-Schmidt process*. In the ordinary Gram-Schmidt process, an orthonormal basis $\hat{\mathbf{y}}_i$ $(i = 1, 2, \ldots, n)$ is computed in successive steps without altering the original vectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, n)$. In the modified Gram-Schmidt process, however, upon computing $\mathbf{y}_i$ the vectors $\mathbf{x}_{i+1}, \mathbf{x}_{i+2}, \ldots, \mathbf{x}_n$ are also changed by insisting that they be orthogonal to $\hat{\mathbf{y}}_i$, as well as to $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_{i-1}$. The first step is as given by Eq. (B.17), but in addition the vectors $\mathbf{x}_i$ $(i = 2, 3, \ldots, n)$ are modified by writing

$$\mathbf{x}_i^{(1)} = \mathbf{x}_i - (\hat{\mathbf{y}}_1, \mathbf{x}_i)\hat{\mathbf{y}}_1, \qquad i = 2, 3, \ldots, n \tag{B.23}$$

Forming the inner product $(\hat{\mathbf{y}}_1, \mathbf{x}_i^{(1)})$, we conclude that $\mathbf{x}_i^{(1)}$ $(i = 2, 3, \ldots, n)$ are all orthogonal to $\hat{\mathbf{y}}_1$. The next step consists of normalizing $\mathbf{x}_2^{(1)}$ to produce $\hat{\mathbf{y}}_2$ as well as of modifying $\mathbf{x}_i^{(1)}$ $(i = 3, 4, \ldots, n)$ by writing

$$\mathbf{x}_i^{(2)} = \mathbf{x}_i^{(1)} - (\hat{\mathbf{y}}_2, \mathbf{x}_i^{(1)})\hat{\mathbf{y}}_2, \qquad i = 3, 4, \ldots n \tag{B.24}$$

It is not difficult to verify that $\mathbf{x}_i^{(2)}$ $(i = 3, 4, \ldots, n)$ are all orthogonal to both $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$. Of course, $\hat{\mathbf{y}}_3$ is obtained by normalizing $\mathbf{x}_3^{(2)}$. Generalizing, the jth step consists of computing $\mathbf{x}_j^{(j-1)}$ and normalizing it to produce $\hat{\mathbf{y}}_j$, or

$$\hat{\mathbf{y}}_j = \mathbf{x}_j^{(j-1)} / \|\mathbf{x}_j^{(j-1)}\| \tag{B.25}$$

and then computing the modified vectors

$$\mathbf{x}_i^{(j)} = \mathbf{x}_i^{(j-1)} - (\hat{\mathbf{y}}_j, \mathbf{x}_i^{(j-1)})\hat{\mathbf{y}}_j, \qquad i = j+1, j+2, \ldots, n \tag{B.26}$$

which renders $\mathbf{x}_i^{(j)}$ orthogonal to $\hat{\mathbf{y}}_j$ as well as to $\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \ldots, \hat{\mathbf{y}}_{j-1}$, or

$$(\hat{\mathbf{y}}_k, \mathbf{x}_i^{(j)}) = 0, \qquad k = 1, 2, \ldots, j \tag{B.27}$$

The process is completed with the computation and normalization of $\mathbf{x}_n^{(n-1)}$, yielding $\hat{\mathbf{y}}_n$.

When the vectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, n)$ are independent, the ordinary and the modified Gram-Schmidt processes yield the same results. When the vectors $\mathbf{x}_i$ are nearly dependent, however, the ordinary Gram-Schmidt process fails to yield orthonormal vectors, but the modified Gram-Schmidt process does yield vectors that are nearly orthonormal (Ref. 3, p. 149).

Example B.3

Consider the vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \tag{a}$$

and obtain an orthonormal basis in terms of these vectors by the modified Gram-Schmidt process. Use Euclidean norms for the vectors.

From Eq. (B.17), we obtain the first normalized vector

$$\hat{\mathbf{y}}_1 = \hat{\mathbf{x}}_1 = \mathbf{x}_1/\|\mathbf{x}_1\| = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{b}$$

so that, from Eqs. (B.23), we can write

$$\begin{aligned} \mathbf{x}_2^{(1)} &= \mathbf{x}_2 - (\hat{\mathbf{y}}_1, \mathbf{x}_2)\hat{\mathbf{y}}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \left(\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \right) \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \\ \mathbf{x}_3^{(1)} &= \mathbf{x}_3 - (\hat{\mathbf{y}}_1, \mathbf{x}_3)\hat{\mathbf{y}}_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} - \left(\frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \frac{2}{3} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \end{aligned} \tag{c}$$

The second vector of the orthonormal set is obtained by simply normalizing $\mathbf{x}_2^{(1)}$, or

$$\hat{\mathbf{y}}_2 = \mathbf{x}_2^{(1)}/\|\mathbf{x}_2^{(1)}\| = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \tag{d}$$

Finally, from Eqs. (B.24), we can write

$$\begin{aligned} \mathbf{x}_3^{(2)} &= \mathbf{x}_3^{(1)} - (\hat{\mathbf{y}}_2, \mathbf{x}_3^{(1)})\hat{\mathbf{y}}_2 \\ &= \frac{2}{3} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} - \left(\frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}, \frac{2}{3} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right) \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \end{aligned} \tag{e}$$

so that

$$\hat{\mathbf{y}}_3 = \mathbf{x}_3^{(2)}/\|\mathbf{x}_3^{(2)}\| = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \tag{f}$$

It can be verified easily that the vectors $\hat{\mathbf{y}}_1$, $\hat{\mathbf{y}}_2$ and $\hat{\mathbf{y}}_3$ are orthonormal.

B.6 MATRICES

A *matrix* is a rectangular array of scalars of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \tag{B.28}$$

The scalars a_{ij} $(i = 1, 2, \dots, m;\ j = 1, 2, \dots, n)$, called the *elements of A*, belong to a given field F. The field is assumed to be either the real field R or the complex field C. Because the matrix A has m rows and n columns it is referred to as an $m \times n$ *matrix*. It is customary to say that the *dimensions* of A are $m \times n$. The position of the element a_{ij} in the matrix A is in the ith row and jth column, so that i is referred to as the row index and j as the column index.

If $m = n$, the matrix A reduces to a *square matrix of order* n. The elements a_{ii} in the square matrix are called the *main diagonal elements of* A. The remaining elements are referred to as the *off-diagonal elements of* A. In the special case in which all the off-diagonal elements of A are zero, A is said to be a *diagonal matrix*. If A is diagonal and if all its diagonal elements are unity, $a_{ii} = 1$, then the matrix is called *unit matrix*, or *identity matrix*, and denoted by I. Introducing the Kronecker delta symbol δ_{ij} defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{B.29}$$

the identity matrix can be regarded as a matrix with every element equal to the Kronecker delta and can be written in the form $I = [\delta_{ij}]$. Similarly, a diagonal matrix D can be written in terms of the Kronecker delta in the form $D = [a_{ij}\delta_{ij}]$.

A square matrix A is said to be *upper (lower) triangular* if $a_{ij} = 0$ for $i > j$ $(i < j)$. If the diagonal elements of an upper (lower) triangular matrix are unity, then the matrix is referred to as *unit upper (lower) triangular*. A square matrix A is said to be *upper (lower) Hessenberg* if $a_{ij} = 0$ for $i > j + 1$ $(i < j - 1)$. If A is upper and lower Hessenberg simultaneously, then it is said to be *tridiagonal*. Clearly, a tridiagonal matrix has nonzero elements only on the main diagonal and on the diagonals immediately above and below the main diagonal.

A matrix obtained from A by interchanging all its rows and columns is referred to as the *transpose* of A and is denoted by A^T. Hence,

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \multicolumn{4}{c}{\dotfill} \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \tag{B.30}$$

It is obvious that if A is an $m \times n$ matrix, then A^T is an $n \times m$ matrix.

Next, we consider a square matrix A. If the elements of A are such that $a_{ij} = a_{ji}$, then the matrix A is said to be *symmetric*. Otherwise, the matrix is *nonsymmetric*. Hence, a matrix is symmetric if $A = A^T$. On the other hand, if the elements of A are such that $a_{ij} = -a_{ji}$ for $i \neq j$ and $a_{ii} = 0$, then the matrix A is said to be *skew symmetric*. It follows that A is skew symmetric if $A = -A^T$.

A matrix consisting of one column and n rows is called a *column matrix* and denoted by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} \tag{B.31}$$

The column matrix $\mathbf{x}$ can clearly be identified with a vector in L^n and is also known as a *column vector*. The transpose of the column matrix $\mathbf{x}$ is the *row matrix*

$$\mathbf{x}^T = [x_1\ x_2\ \ldots\ x_n] \tag{B.32}$$

and is also called a *row vector*.

A matrix with all its elements equal to zero is called the *zero matrix* or the *null matrix* and is denoted by 0, $\mathbf{0}$, or $\mathbf{0}^T$, depending on whether it is a rectangular, a column, or a row matrix, respectively.

B.7 BASIC MATRIX OPERATIONS

Two matrices A and B are said to be equal if and only if they have the same number of rows and columns and $a_{ij} = b_{ij}$ for all pairs of subscripts i and j.

If A and B are two $m \times n$ matrices, then the sum of A and B is defined as a matrix C whose elements are

$$c_{ij} = a_{ij} + b_{ij}, \qquad i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n \tag{B.33}$$

Clearly, $C = A + B$ is also an $m \times n$ matrix. Matrix addition is *commutative* and *associative*. Indeed, if A, B and C are arbitrary $m \times n$ matrices, then

$$A + B = B + A, \qquad (A + B) + C = A + (B + C) \tag{B.34}$$

The *product of a matrix A and a scalar α* implies that every element of A is multiplied by α. Hence, if A is an $m \times n$ matrix, then the statement $C = \alpha A$ implies

$$c_{ij} = \alpha a_{ij}, \qquad i = 1, 2, \ldots, m;\ j = 1, 2, \ldots, n \tag{B.35}$$

Next we define the *product of two matrices*. If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then the product $C = AB$ of the two matrices is an $m \times p$ matrix with the elements

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} \tag{B.36}$$

It is clear from the above that the matrix product is defined only if the number of columns of A is equal to the number of rows of B. In this case, the matrices A and B are said to be *conformable* in the order stated. The matrix product AB can be described as B *premultiplied by* A or A *postmultiplied by* B. It can also be described as B *multiplied on the left by* A or A *multiplied on the right by* B. Matrix multiplication is in general *not commutative*

$$AB \neq BA \tag{B.37}$$

or, stated differently, matrices A and B *do not commute*. In fact, unless $m = p$, the matrix product BA is not even defined. One notable exception is the case in which one of the matrices is the identity matrix because then

$$AI = IA = A \tag{B.38}$$

Clearly, the order of I must be such that the product is defined.

The matrix product

$$AB = 0 \tag{B.39}$$

does not imply that either A or B is a null matrix, or both A and B are null matrices. Indeed, an illustration of this statement is provided by

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{B.40}$$

The matrix product satisfies *associative laws*. If A, B and C are $m \times n$, $n \times p$ and $p \times q$ matrices, respectively, then it can be verified that

$$D = (AB)C = A(BC) \tag{B.41}$$

is an $m \times q$ matrix whose elements are given by

$$d_{ij} = \sum_{l=1}^{p}\sum_{k=1}^{n} a_{ik}b_{kl}c_{lj} = \sum_{k=1}^{n}\sum_{l=1}^{p} a_{ik}b_{kl}c_{lj} \tag{B.42}$$

The matrix product satisfies *distributive laws*. If A and B are $m \times n$ matrices, C is an $n \times m$ matrix and D is an $p \times q$ matrix, then it can be shown that

$$C(A + B) = CA + CB, \qquad (A + B)D = AD + BD \tag{B.43}$$

If A is an $m \times n$ matrix and B is an $n \times p$ matrix, so that the product $C = AB$ is given by Eq. (B.36), then

$$C^T = (AB)^T = B^T A^T \tag{B.44}$$

To show this, we recognize that to any element a_{ik} in A corresponds the element a_{ki} in A^T, and to any element b_{kj} in B corresponds the element b_{jk} in B^T. Then the product

$$\sum_{k=1}^{n} b_{jk}a_{ki} = c_{ji} \tag{B.45}$$

establishes the validity of Eq. (B.44). In words, *the transpose of a product of two matrices is equal to the product of the transposed matrices in reversed order.* As a corollary, it can be verified that if

$$C = A_1 A_2 \dots A_{s-1} A_s \tag{B.46}$$

then

$$C^T = A_s^T A_{s-1}^T \dots A_2^T A_1^T \tag{B.47}$$

We have considerable interest in the concepts of inner product and orthogonality of vectors, so that we will find it convenient to recast some of the relations in Sec. B.4 in terms of matrix notation. The *inner product* of two n-vectors $\mathbf{x}$ and $\mathbf{y}$, Eq. (B.9), can be expressed as

$$\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = c \tag{B.48}$$

and *it represents a scalar*. When the inner product is zero, or

$$\mathbf{x}^T\mathbf{y} = \mathbf{y}^T\mathbf{x} = 0 \tag{B.49}$$

the vectors **x** and **y** are said to be *orthogonal.*,

Less frequently, we encounter the *outer product* of two vectors **x** and **y**, defined as

$$\mathbf{x}\mathbf{y}^T = C \tag{B.50}$$

and *it represents a matrix* C. If **x** is an m-vector and **y** an n-vector, then the matrix C is $m \times n$. Clearly, the outer product is not symmetric in **x** and **y**, because

$$\mathbf{y}\mathbf{x}^T = C^T \neq \mathbf{x}\mathbf{y}^T \tag{B.51}$$

Example B.4

Calculate the matrix product AB, where

$$A = \begin{bmatrix} 2 & -3 \\ 1 & 5 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 3 & 7 \\ -1 & 4 & 2 \end{bmatrix} \tag{a}$$

What can be said about the matrix product BA?

The matrix product AB is formed as follows:

$$AB = \begin{bmatrix} 2(1) - 3(-1) & 2(3) - 3(4) & 2(7) - 3(2) \\ 1(1) + 5(-1) & 1(3) + 5(4) & 1(7) + 5(2) \end{bmatrix} = \begin{bmatrix} 5 & -6 & 8 \\ -4 & 23 & 17 \end{bmatrix} \tag{b}$$

The matrix product BA is not defined, because B is a 2×3 matrix and A a 2×2 matrix and hence the matrices are not conformable in that order.

Example B.5

Calculate the matrix products AB and CA, where

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & 4 & 1 & -2 \\ -1 & 5 & 7 & 3 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 3 \\ 7 & 4 \\ -2 & 2 \end{bmatrix} \tag{a}$$

The matrix products AB and CA are as follows:

$$AB = \begin{bmatrix} -1 & 5 & 7 & 3 \\ 2 & 4 & 1 & -2 \end{bmatrix}, \qquad CA = \begin{bmatrix} 3 & 1 \\ 4 & 7 \\ 2 & -2 \end{bmatrix} \tag{b}$$

We observe that the matrix AB is obtained from B by interchanging the rows. Similarly, CA is obtained from C by interchanging the columns. Hence, the effect of premultiplying a matrix by A is to permute its rows and the effect of postmultiplying a matrix by A is to permute its columns. For this reason, A is called a *permutation matrix*. In general, a permutation matrix is a matrix obtained by interchanging rows or columns of the identity matrix.

B.8 DETERMINANTS

If A is any square matrix of order n with elements in the field F, then it is possible to associate with A a number in F called the *determinant of* A, and denoted by det A or $|A|$. The determinant of A is said to be of *order* n and can be exhibited in the form

$$\det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \vdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{B.52}$$

Determinants have many interesting and useful properties. We examine only those properties pertinent to our study.

Unlike the matrix A, which represents a given array of numbers, the determinant of A represents a single number with a unique value that can be calculated by following rules for the expansion of determinants. The expansion rules can be most conveniently discussed by introducing the concept of minor determinants. The *minor determinant* $|M_{rs}|$ corresponding to the element a_{rs} is the determinant obtained from $|A|$ by striking out the rth row and sth column. Clearly, the order of $|M_{rs}|$ is $n-1$ The signed minor determinant corresponding to the element a_{rs} is called the *cofactor* of a_{rs} and is given by

$$\det A_{rs} = |A_{rs}| = (-1)^{r+s}|M_{rs}| \tag{B.53}$$

The value of the determinant of A can be obtained by expanding in terms of cofactors by the rth row as follows:

$$|A| = \sum_{s=1}^{n} a_{rs}|A_{rs}| \tag{B.54}$$

The determinant can also be expanded by the sth column in the form

$$|A| = \sum_{r=1}^{n} a_{rs}|A_{rs}| \tag{B.55}$$

The value of the determinant is unique, regardless of whether it is expanded by a row or a column, and regardless of which row or column. The expansion by cofactors is known as a *Laplace expansion*. The cofactors $|A_{rs}|$ are determinants of order $n-1$. If $n = 2$, then these cofactors are simply scalars. If $n > 2$, then these cofactors can be expanded in terms of their own cofactors, and the process repeated until the minor determinants are of order 2. As an illustration, let $n = 3$ and expand det A by the first row as follows:

$$\begin{aligned}
\det A &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}|A_{11}| + a_{12}|A_{12}| + a_{13}|A_{13}| \\
&= a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&= a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) \\
&\quad + a_{13}(a_{21}a_{32} - a_{22}a_{31})
\end{aligned} \tag{B.56}$$

Because the value of det A is the same, regardless of whether the determinant is expanded by a row or a column, it follows that

$$\det A = \det A^T \tag{B.57}$$

or *the determinant of a matrix is equal to the determinant of the transposed matrix*. It is easy to verify that *the determinant of a triangular matrix is equal to the product of the main diagonal elements*. It follows immediately that *the determinant of a diagonal matrix is equal to the product of the diagonal elements*, and *the determinant of the identity matrix is equal to* 1.

If det $A = 0$, then the matrix A is said to be *singular*, and if det $A \neq 0$, the matrix is *nonsingular*. Clearly, a matrix with an entire row or an entire column equal to zero is singular. The evaluation of determinants can be greatly simplified by invoking certain properties of determinants. In fact, it is often possible to establish that a determinant is zero without actually expanding it. From Eq. (B.54), one can deduce the following properties:

1. If two rows (or two columns) are interchanged, then the determinant changes sign.
2. If all the elements of one row (or of one column) are multiplied by a scalar α, then the determinant is multiplied by α.
3. The value of a determinant does not change if one row (or one column) multiplied by a scalar α is added or subtracted from another row (or another column).
4. If every element in one row (or one column) is the sum of two terms, then the determinant is equal to the sum of two determinants, each of the two determinants being obtained by splitting every sum so that one term is in one determinant and the remaining term is in the other determinant.

The above properties permit us to make two observations. Property 2 implies that $\det(\alpha A) = \alpha^n \det A$, where α is a scalar and n is the order of the matrix. On the other hand, property 3 implies that a determinant with two proportional rows, or two proportional columns, is equal to zero. Property 3 can be used in general to simplify the evaluation of a determinant, and in particular to show that its value is zero, if indeed this is the case.

Example B.6

Calculate the value of the determinant

$$|A| = \begin{vmatrix} 3 & 2 & -1 \\ 1 & 5 & 2 \\ 3 & -1 & 2 \end{vmatrix} \tag{a}$$

Expanding by the first row, we obtain

$$\begin{aligned} A &= 3\begin{vmatrix} 5 & 2 \\ -1 & 2 \end{vmatrix} - 2\begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} - 1\begin{vmatrix} 1 & 5 \\ 3 & -1 \end{vmatrix} \\ &= 3[5(2) - 2(-1)] - 2[1(2) - 2(3)] - [1(-1) - 5(3)] = 60 \end{aligned} \tag{b}$$

On the other hand, subtracting three times the second row from the first and third rows, we can write

$$A = \begin{vmatrix} 0 & -13 & -7 \\ 1 & 5 & 2 \\ 0 & -16 & -4 \end{vmatrix} \tag{c}$$

Next, expanding by the first column, we obtain

$$A = -\begin{vmatrix} -13 & -7 \\ -16 & -4 \end{vmatrix} = [(-13)(-4) - (-7)(-16)] = 60 \tag{d}$$

which is the same value as that given by Eq. (b).

B.9 INVERSE OF A MATRIX

If A and B are two $n \times n$ matrices such that

$$AB = BA = I \tag{B.58}$$

then B is said to be the *inverse* of A and is denoted by

$$B = A^{-1} \tag{B.59}$$

Note that at the same time A is the inverse of B, $A = B^{-1}$.

Next, we derive a formula for the inverse of a matrix. To this end, we consider Eq. (B.53) and introduce the *adjugate* of A in the form

$$\text{adj } A = [(-1)^{r+s}|M_{rs}|]^T \tag{B.60}$$

where $(-1)^{r+s}|M_{rs}|$ is the cofactor corresponding to a_{rs}. Then we can write

$$A \text{ adj } A = \left[\sum_{s=1}^{n}(-1)^{r+s}a_{ps}|M_{rs}|\right] \tag{B.61}$$

Recalling Eq. (B.54), we conclude that every element of A adj A can be regarded as a determinantal expansion. When $p = r$, the element is simply equal to det A. On the other hand, when $p \neq r$ the result is zero. This can be explained by recognizing that the determinant corresponding to $p \neq r$ is obtained from the matrix A by replacing the rth row by the pth row and keeping the pth row intact. Because the corresponding determinant has two identical rows, its value is zero. In view of this, Eq. (B.61) can be rewritten as

$$A \text{ adj } A = (\det A)I \tag{B.62}$$

where I is the identity matrix of order N. Multiplying Eq. (B.62) on the left by A^{-1} and dividing through by det A, we obtain

$$A^{-1} = \frac{\text{adj } A}{\det A} \tag{B.63}$$

If det $A = 0$, then no matrix B exists such that Eq. (B.58) is satisfied. To show this, we invoke the following theorem (Ref. 2, p. 134): *If A and B are $n \times n$ matrices, then*

$$\det AB = \det A \det B \tag{B.64}$$

But, from Eq. (B.58), we conclude that det $AB = I$ if $B = A^{-1}$ exists, so that det $A \neq 0$. Hence, if det $A = 0$, Eq. (B.58) cannot be satisfied, so that $B = A^{-1}$ does not exist. Recalling that when det $A = 0$ the matrix is singular, it follows that *an $n \times n$ matrix A has an inverse if and only if A is nonsingular.*

To calculate the inverse of a matrix of large order by means of formula (B.63) it is necessary to evaluate a large number of determinants. For example, if A is of order n, then the calculation of det A requires the evaluation of $n!/2$ determinants of order 2. Hence, as n increases, it is necessary to carry out an increasingly large number of multiplications with a progressive loss of accuracy, so that the use of the formula (B.63) is not recommended. In this text we study more efficient and more accurate methods for the calculation of the inverse of a matrix (see Sec. 6.1).

Next we consider the product of matrices given by Eq. (B.46). Multiplying both sides of Eq. (B.46) on the right by $A_s^{-1}, A_{s-1}^{-1}, \ldots, A_1^{-1}$, in sequence, and then on the left by C^{-1}, we obtain

$$C^{-1} = A_s^{-1}A_{s-1}^{-1}\ldots A_2^{-1}A_1^{-1} \tag{B.65}$$

or *the inverse of a product of matrices is equal to the product of the inverse matrices in reversed order.* Of course, Eq. (B.65) implies that all the inverse matrices in question exist.

Example B.7

Calculate the inverse of the matrix

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 1 & 5 & 2 \\ 3 & -1 & 2 \end{bmatrix} \tag{a}$$

First, we evaluate the minor determinants

$$\begin{gathered} |M_{11}| = \begin{vmatrix} 5 & 2 \\ -1 & 2 \end{vmatrix} = 12, \quad |M_{12}| = \begin{vmatrix} 1 & 2 \\ 3 & 2 \end{vmatrix} = -4, \quad |M_{13}| = \begin{vmatrix} 1 & 5 \\ 3 & -1 \end{vmatrix} = -16 \\ |M_{21}| = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3, \quad |M_{22}| = \begin{vmatrix} 3 & -1 \\ 3 & 2 \end{vmatrix} = 9, \quad |M_{23}| = \begin{vmatrix} 3 & 2 \\ 3 & -1 \end{vmatrix} = -9 \\ |M_{31}| = \begin{vmatrix} 2 & -1 \\ 5 & 2 \end{vmatrix} = 9, \quad |M_{32}| = \begin{vmatrix} 3 & -1 \\ 1 & 2 \end{vmatrix} = 7, \quad |M_{33}| = \begin{vmatrix} 3 & 2 \\ 1 & 5 \end{vmatrix} = 13 \end{gathered} \tag{b}$$

Using Eq. (B.60), we obtain the adjugate matrix

$$\text{adj } A = \begin{bmatrix} 12 & -(3) & 9 \\ -(-4) & 9 & -(7) \\ (-16) & -(-9) & 13 \end{bmatrix} = \begin{bmatrix} 12 & -3 & 9 \\ 4 & 9 & -7 \\ -16 & 9 & 13 \end{bmatrix} \tag{c}$$

Recalling from Example B.6 that det $A = 60$ and using Eq. (B.63), we obtain

$$A^{-1} = \frac{1}{60}\begin{bmatrix} 12 & -3 & 9 \\ 4 & 9 & -7 \\ -16 & 9 & 13 \end{bmatrix} \tag{d}$$

B.10 PARTITIONED MATRICES

On occasion it is convenient to partition matrices into submatrices. Then, under proper circumstances, certain matrix operations can be performed by treating the submatrices as if they were single elements. As an example, we consider a 3 × 4 matrix A and partition it as follows:

$$A = \left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array}\right] = \left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \tag{B.66}$$

where

$$A_{11} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad A_{12} = \begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix},$$
$$A_{21} = [a_{31} \quad a_{32}], \qquad A_{22} = [a_{33} \quad a_{34}] \tag{B.67}$$

are the submatrices of A. Next, we consider a 4×4 matrix B partitioned in the form

$$B = \left[\begin{array}{cc|cc} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ \hline b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{array}\right] = \left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right] \tag{B.68}$$

where

$$B_{11} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \qquad B_{12} = \begin{bmatrix} b_{13} & b_{14} \\ b_{23} & b_{24} \end{bmatrix}$$
$$B_{21} = \begin{bmatrix} b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix}, \qquad B_{22} = \begin{bmatrix} b_{33} & b_{34} \\ b_{43} & b_{44} \end{bmatrix} \tag{B.69}$$

It is not difficult to verify that the matrix product AB can be obtained by treating the submatrices A_{ik} and B_{kj} as if they were ordinary matrices. Indeed, the elements of the product $C = AB$ are

$$C_{ij} = \sum_{k=1}^{2} A_{ik}B_{kj}, \qquad i, j = 1, 2 \tag{B.70}$$

It should be pointed out, however, that products such as (B.70) are possible only if the matrix A_{ik} has as many columns as the matrix B_{kj} has rows, which is clearly true in the particular case at hand.

If the off-diagonal submatrices of a square matrix are null matrices, then the matrix is said to be *block-diagonal.* For block-diagonal matrices, the determinant of the matrix is equal to the product of the determinants of the submatrices on the main diagonal. For example, if B_{12} and B_{21} in Eq. (B.68) are null matrices, then

$$\det B = \det B_{11} \det B_{22} \tag{B.71}$$

Actually the above statement is true even if the matrix is only *block-triangular*, i.e., if only the submatrices above (or below) the main diagonal are null matrices.

B.11 SYSTEMS OF LINEAR ALGEBRAIC EQUATIONS

We consider a system of m nonhomogeneous linear equations in n unknowns $x_1, x_2, \ldots, x_n$ of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= c_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= c_m \end{aligned} \tag{B.72}$$

The system of equations can be written as the compact matrix equation

$$A\mathbf{x} = \mathbf{c} \tag{B.73}$$

where $A = [a_{ij}]$ is an $m \times n$ matrix known as the matrix of the coefficients, $\mathbf{x} = [x_1\ x_2 \ldots x_n]^T$ is the n-vector of the unknowns and $\mathbf{c} = [c_1\ c_2 \ldots c_m]^T$ is the m-vector of "nonhomogeneous terms." Our interest is in the conditions under which Eq. (B.73) has a solution.

The matrix A can be partitioned into n column vectors of dimension m in the form

$$A = [\mathbf{a}_1\ \mathbf{a}_2 \ldots \mathbf{a}_n] \tag{B.74}$$

where $\mathbf{a}_1 = [a_{11}\ a_{21} \ldots a_{m1}]^T$, etc., are the column vectors. In view of this, the matrix product $A\mathbf{x}$ can be looked upon as a linear combination of the columns of A, so that Eq. (B.73) can be written as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \ldots + x_n\mathbf{a}_n = \mathbf{c} \tag{B.75}$$

Equation (B.75) implies that the set of all products $A\mathbf{x}$ is the same as the set of linear combinations of the columns of A. The subspace of L^m spanned by the columns of A is called the *column space* of A and denoted by $\mathcal{R}(A)$. If $\mathbf{y}$ is an m-vector, then $\mathbf{y}^T$ is a row vector with m components. Now, if A is partitioned into m row vectors, then the product $\mathbf{y}^TA$ is linear combination of the rows of A whose coefficients are the components of $\mathbf{y}$. Hence, the *row space* of A, written $\mathcal{R}(A^T)$, is the subspace of L^n spanned by the row vectors of A.

Next, we define the *rank* of a matrix A, denoted by rank A, as the dimension of the linear space spanned by its columns. Because the latter is simply the dimension of $\mathcal{R}(A)$, we have

$$\text{rank } A = \dim \mathcal{R}(A) \tag{B.76}$$

It would appear that rank A should have been more properly referred to as the *column rank* of A, which would have naturally called for the introduction of a *row rank* of A as the dimension of $\mathcal{R}(A^T)$. It turns out, however, that the column rank and row rank of any matrix A are equal (Ref. 4, p. 93), so that no such distinction is necessary. In view of the definition of the dimension of a linear space, it follows that *the rank of a matrix A is equal to the maximum number of linearly independent columns of A, and it is also equal to the maximum number of linearly independent rows of A, where the two numbers must be the same.*

There is one more vector space associated with any $m \times n$ matrix A, the *nullspace* of A. It is denoted by $\mathcal{N}(A)$ and defined as the space of all the solutions $\mathbf{x} \neq \mathbf{0}$ satisfying the homogeneous equation $A\mathbf{x} = \mathbf{0}$. The dimension of the nullspace $\mathcal{N}$ is called the *nullity* of A, $\dim \mathcal{N} = \text{null } A$.

Let us return now to Eq. (B.73) and introduce the *augmented matrix* of the system defined by

$$B = [A, \mathbf{c}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & c_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & c_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & c_m \end{bmatrix} \tag{B.77}$$

Then, Eq. (B.73) *has a solution* $\mathbf{x}$ *if and only if the rank of the augmented matrix* B *is equal to the rank of* A. If a solution $\mathbf{x}$ exists, then $\mathbf{c}$ is a linear combination of the columns of A and hence lies in $\mathcal{R}(A)$. It follows that $\mathcal{R}(B) = \mathcal{R}(A)$, and rank $B =$ rank A.

The rank of an arbitrary matrix can be connected with the order of its nonsingular square submatrices. Indeed, *the rank of any matrix* A *is equal to the order of the square submatrix of* A *of greatest order whose determinant does not vanish.* It follows that:

(a) If $m \geq n$, then the largest possible rank of A is n. If rank $A =$ rank $B = n$, then Eq. (B.73) has a unique solution.

(b) If $m < n$, then the largest possible rank of A is m. If rank $A =$ rank $B = m$, then Eq. (B.73) has an infinity of solutions. A unique solution can be chosen in the form of the solution with the minimum norm

$$\mathbf{x} = A^T(AA^T)^{-1}\mathbf{c} \tag{B.78}$$

where AA^T is an $m \times m$ matrix of rank m and is therefore nonsingular.

The case in which the number of equations is equal to the number of unknowns is of particular interest. If A is a square matrix of order n, then the following statements are equivalent:

1. The rank of A is n, rank $A = n$.
2. The system $A\mathbf{x} = \mathbf{c}$ has a unique solution for arbitrary vectors $\mathbf{c}$.
3. The system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$, which implies that null $A = 0$.

The implication of statements 1 and 2 is that the matrix A is nonsingular, so that A possesses an inverse. Considering the case in which the matrix A in Eq. (B.73) is square and premultiplying both sides of the equation by A^{-1}, we obtain

$$\mathbf{x} = A^{-1}\mathbf{c} \tag{B.79}$$

Hence, when A is nonsingular the solution of Eq. (B.73) can be produced by simply calculating the inverse of A. We have shown in Sec. B.9 that A^{-1} can be obtained by dividing the adjugate of A by the determinant of A; this method for solving sets of simultaneous equations is generally known as *Cramer's rule.* This approach is mainly of academic interest, and in computational work the procedure is seldom used, especially for large order matrices A. Indeed, the procedure involves the evaluation of a large number of determinants, which is time-consuming and leads to loss of accuracy. In Sec. 6.1, we discuss a more efficient method for deriving the solution of Eq. (B.73), namely, the Gaussian elimination.

Next, we turn our attention to the homogeneous system $A\mathbf{x} = \mathbf{0}$. As pointed out earlier, the matrix product $A\mathbf{x}$ represents a linear combination of the column vectors of A. Because this linear combination must be equal to zero, it follows from Sec. B.2 that the columns of A are not independent. Hence, the rank of A must be less than n, so that det $A = 0$. This conclusion can be stated in a more formal manner by means of the well-known theorem of linear algebra: *If* A *is an* $n \times n$ *matrix, then the equation* $A\mathbf{x} = \mathbf{0}$ *has a nontrivial solution* $\mathbf{x} \neq \mathbf{0}$ *if and only if* det $A = 0$.

As an application of the above theorem, let us devise a test for the dependence of a set of n-vectors $\mathbf{y}_1$ $\mathbf{y}_2, \ldots, \mathbf{y}_n$. If the vectors are to be linearly dependent, then they must satisfy a relation of the type

$$\alpha_1\mathbf{y}_1 + \alpha_2\mathbf{y}_2 + \ldots + \alpha_n\mathbf{y}_n = \mathbf{0} \tag{B.80}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are constant scalars. Next we form the inner products $(\mathbf{y}_i, \mathbf{y}_j)$. We have shown in Sec. B.6, however, that vectors can be represented by column matrices. In view of this, the inner product can be written in the matrix form

$$(\mathbf{y}_i, \mathbf{y}_j) = \bar{\mathbf{y}}_j^T \mathbf{y}_i \tag{B.81}$$

Hence, premultiplying Eq. (B.80) by $\bar{\mathbf{y}}_1^T, \bar{\mathbf{y}}_2^T, \ldots, \bar{\mathbf{y}}_n^T$, in sequence, we obtain

$$\begin{aligned} \alpha_1\bar{\mathbf{y}}_1^T\mathbf{y}_1 + \alpha_2\bar{\mathbf{y}}_1^T\mathbf{y}_2 + \ldots + \alpha_n\bar{\mathbf{y}}_1^T\mathbf{y}_n &= 0 \\ \alpha_1\bar{\mathbf{y}}_2^T\mathbf{y}_1 + \alpha_2\bar{\mathbf{y}}_2^T\mathbf{y}_2 + \ldots + \alpha_n\bar{\mathbf{y}}_2^T\mathbf{y}_n &= 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \alpha_1\bar{\mathbf{y}}_n^T\mathbf{y}_1 + \alpha_2\bar{\mathbf{y}}_n^T\mathbf{y}_2 + \ldots + \alpha_n\bar{\mathbf{y}}_n^T\mathbf{y}_n &= 0 \end{aligned} \tag{B.82}$$

Equations (B.82) represent a set of n homogeneous simultaneous equations in the unknowns $\alpha_1, \alpha_2, \ldots, \alpha_n$. By the theorem just presented, Eqs. (B.82) have a nontrivial solution if and only if the determinant of the coefficients vanishes, or

$$|G| = \begin{vmatrix} \bar{\mathbf{y}}_1^T\mathbf{y}_1 & \bar{\mathbf{y}}_1^T\mathbf{y}_2 & \ldots & \bar{\mathbf{y}}_1^T\mathbf{y}_n \\ \bar{\mathbf{y}}_2^T\mathbf{y}_1 & \bar{\mathbf{y}}_2^T\mathbf{y}_2 & \ldots & \bar{\mathbf{y}}_2^T\mathbf{y}_n \\ \cdots & \cdots & \cdots & \cdots \\ \bar{\mathbf{y}}_n^T\mathbf{y}_1 & \bar{\mathbf{y}}_n^T\mathbf{y}_2 & \ldots & \bar{\mathbf{y}}_n^T\mathbf{y}_n \end{vmatrix} = 0 \tag{B.83}$$

where $|G|$ is known as the *Gramian determinant*. Hence, a necessary and sufficient condition for the set of vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n$ to be linearly dependent is that the Gramian determinant be zero.

As a simple illustration, we consider the unit vectors $\mathbf{e}_i$ given by Eqs. (B.6). In this case, the *Gramian matrix* G is equal to the identity matrix, $G = I$, so that $|G| = 1$. Hence, the unit vectors $\mathbf{e}_i$ are linearly independent.

Example B.8

Determine the rank and nullity of the matrix

$$A = \begin{bmatrix} 2 & -1 & 4 & 3 \\ 1 & 5 & -2 & 4 \\ 5 & 3 & 6 & 10 \\ -1 & 6 & -6 & 1 \end{bmatrix} \tag{a}$$

It is not difficult to verify that $\det A = 0$, so that rank $A < 4$. Hence, at least one of the columns (rows) of A is a linear combination of the other. By inspection, we observe that adding twice the first row to the second we obtain the third row. Moreover, subtracting the first row from the second we obtain the fourth row. Further search will reveal no other combinations of rows, so that two rows of A are linearly independent. It follows that rank $A = 2$.

To determine the nullity of A, we determine first the null space of A by solving the equation

$$A\mathbf{x} = \mathbf{0} \tag{b}$$

which can be written in the explicit form

$$\begin{aligned} 2x_1 - x_2 + 4x_3 + 3x_4 &= 0 \\ x_1 + 5x_2 - 2x_3 + 4x_4 &= 0 \\ 5x_1 + 3x_2 + 6x_3 + 10x_4 &= 0 \\ -x_1 + 6x_2 - 6x_3 + x_4 &= 0 \end{aligned} \tag{c}$$

The above four equations can be reduced to the two equations

$$\begin{aligned} x_1 + (18/11)x_3 + (19/11)x_4 &= 0 \\ x_2 - (8/11)x_3 + (5/11)x_4 &= 0 \end{aligned} \tag{d}$$

while the remaining two equations are identically zero. It can be verified that every solution of Eqs. (d) can be written in the form

$$\mathbf{x} = \alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 \tag{e}$$

where

$$\mathbf{u}_1 = [-18\ 8\ 11\ 0]^T, \qquad \mathbf{u}_2 = [-19\ \ -5\ 0\ 11]^T \tag{f}$$

The vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ are clearly independent and they span the null space. Hence, they form a basis for the space. The dimension of the null space $\mathcal{N}(A)$ is two, dim $\mathcal{N} = 2$, so that the nullity of A is two, null $A = 2$.

B.12 LINEAR TRANSFORMATIONS

As shown in Sec. B.3, any n-vector $\mathbf{x}$ in L^n with components $x_1, x_2, \ldots, x_n$ can be expressed as the linear combination

$$\mathbf{x} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + \ldots + x_n\mathbf{e}_n = \sum_{j=1}^{n} x_i\mathbf{e}_i \tag{B.84}$$

where $\mathbf{e}_i$ $(i = 1, 2, \ldots, n)$ are the standard unit vectors. Moreover, the scalars $x_1, x_2, \ldots, x_n$ are called the *coordinates* of the vector $\mathbf{x}$ with respect to the basis $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ (Fig. B.2). Next, we consider an $n \times n$ matrix A and write

$$\mathbf{x}' = A\mathbf{x} \tag{B.85}$$

The resulting vector $\mathbf{x}'$ is another vector in L^n, so that Eq. (B.85) represents a *linear transformation* on the vector space L^n which maps the vector $\mathbf{x}$ into a vector $\mathbf{x}'$.

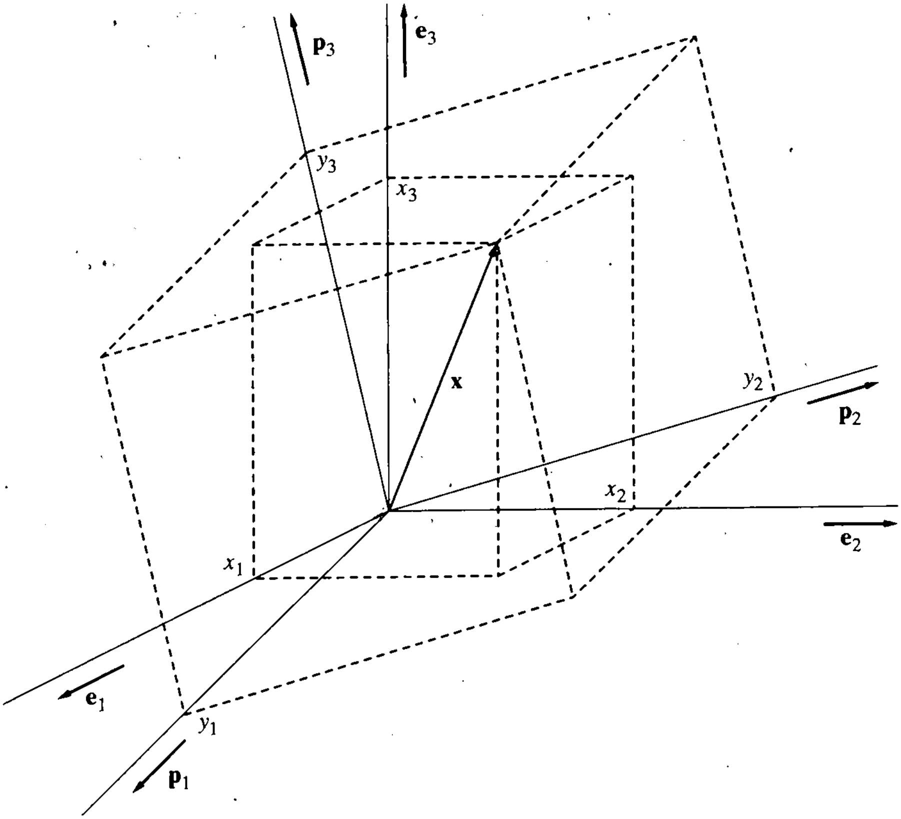


Figure B.2 Decomposition of a three-dimensional vector $\mathbf{x}$ in terms of the standard basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ and an arbitrary basis $\mathbf{p}_1$, $\mathbf{p}_2$, $\mathbf{p}_3$

Our interest lies in expressing the vector $\mathbf{x}$ in terms of any arbitrary basis $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ for L^n, rather than the standard basis, as follows:

$$\mathbf{x} = y_1\mathbf{p}_1 + y_2\mathbf{p}_2 + \ldots + y_n\mathbf{p}_n = \sum_{i=1}^{n} y_i\mathbf{p}_i = P\mathbf{y} \tag{B.86}$$

where

$$P = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_n] \tag{B.87}$$

is an $n \times n$ matrix of basis vectors and

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{B.88}$$

is an n-vector whose components $y_1, y_2, \ldots, y_n$ are the coordinates of $\mathbf{x}$ with respect to the basis $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ (Fig. B.2). By the definition of a basis, the vectors $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$ are linearly independent, so that the matrix P is nonsingular.

Similarly, denoting by $y_1', y_2', \ldots, y_n'$ the coordinates of $\mathbf{x}'$ with respect to the basis $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n$, we can use the analogy with Eq. (B.86) and write

$$\mathbf{x}' = P\mathbf{y}' \tag{B.89}$$

Then, inserting Eqs. (B.86) and (B.89) into Eq. (B.85), we have

$$P\mathbf{y}' = A(P\mathbf{y}) \tag{B.90}$$

so that, premultiplying both sides of Eq. (B.90) by P^{-1}, we obtain

$$\mathbf{y}' = B\mathbf{y} \tag{B.91}$$

where

$$B = P^{-1}AP \tag{B.92}$$

The matrix B represents the same linear transformation as A, but in a different coordinate system. Two square matrices A and B related by an equation of the type (B.92) are said to be *similar* and Eq. (B.92) itself represents a *similarity transformation.*

A similarity transformation of particular interest is the orthonormal transformation. A matrix P is said to be *orthonormal* if it satisfies

$$P^TP = I \tag{B.93}$$

from which it follows that an orthonormal matrix also satisfies

$$P^{-1} = P^T \tag{B.94}$$

Introducing Eq. (B.94) into Eq. (B.92), we obtain

$$B = P^TAP \tag{B.95}$$

Equation (B.95) represents an *orthonormal transformation*, a very important special type of similarity transformation, particularly when the matrix A is symmetric.

B.13 THE ALGEBRAIC EIGENVALUE PROBLEM

The equations for the free vibration of discrete systems can be written in the state form

$$\frac{d\mathbf{x}(t)}{dt} = A\mathbf{x}(t) \tag{B.96}$$

where $\mathbf{x}(t)$ is the n-dimensional state vector and A is an $n \times n$ matrix of coefficients. Equation (B.96) represents a set of homogeneous ordinary differential equations and has the solution

$$\mathbf{x}(t) = e^{\lambda t}\mathbf{x} \tag{B.97}$$

in which λ is a constant scalar and $\mathbf{x}$ a constant vector. Inserting Eq. (B.97) into Eq. (B.96) and dividing through by $e^{\lambda t}$, we obtain

$$A\mathbf{x} = \lambda\mathbf{x} \tag{B.98}$$

Equation (B.98) represents the *algebraic eigenvalue problem* and can be stated as follows: *Determine the values of the parameter λ so that Eq. (B.98) admits nontrivial solutions.* The values of λ are known as *eigenvalues* and are the roots of the *characteristic equation*

$$\det(A - \lambda I) = 0 \tag{B.99}$$

Similarity transformations are often used in numerical algorithms for the algebraic eigenvalue problem. To examine the reason why, we consider Eq. (B.92) and write the characteristic determinant

$$\begin{aligned}\det(B - \lambda I) = \det(P^{-1}AP - \lambda I) &= \det(P^{-1}(A - \lambda I)P) \\ &= \det P^{-1} \det(A - \lambda I) \det P \end{aligned} \tag{B.100}$$

But,

$$\det(P^{-1}P) = \det P^{-1} \det P = 1 \tag{B.101}$$

so that

$$\det(B - \lambda I) = \det(A - \lambda I) \tag{B.102}$$

Because matrices A and B possess the same characteristic determinant, they possess the same eigenvalues. It follows that *eigenvalues do not change under similarity transformations*. Of course the similarity transformations generally used in numerical algorithms for the solution of the eigenvalue problem are orthogonal transformations.

The characteristic determinant can be expressed in the form of the *characteristic polynomial*

$$\begin{aligned}\det(A - \lambda I) &= (-1)^n(\lambda - \lambda_1)(\lambda - \lambda_2)\dots(\lambda - \lambda_n) = (-1)^n \prod_{i=1}^{n}(\lambda - \lambda_i) \\ &= (-1)^n(\lambda^n + c_1\lambda^{n-1} + \dots + c_{n-1}\lambda + c_n)\end{aligned} \tag{B.103}$$

where $\prod$ is the product symbol. Because the eigenvalues do not change under similarity transformations, it follows that *the coefficients $c_i (i = 1, 2, \dots, n)$ of the polynomial are invariant.* Two of the coefficients have special significance, namely, c_1, and c_n. It can be verified that

$$c_1 = -\int_{i=1}^{n} \lambda_i = -\int_{i=1}^{n} a_{ii} = -\mathrm{tr} A \tag{B.104}$$

in which tr A denotes the *trace* of the matrix A, defined as the sum of the diagonal elements of A. Hence, *the trace of A is invariant under similarity transformations.* Similarly, it can be shown that

$$c_n = \prod_{i=1}^{n}(-\lambda_i) = (-1)^n \prod_{i=1}^{n} \lambda_i = (-1)^n \det A \tag{B.105}$$

from which we conclude that *the determinant of A is invariant under similarity transformations.*

B.14 MATRIX NORMS

As with vectors, it is useful to assign a single number to a matrix, thus providing a measure of the magnitude of the matrix in some sense. Such a measure is provided by the norm. The *norm of a square matrix* A is a nonnegative number $\|A\|$ satisfying the conditions

1. $\|A\| \geq 0,\ \ \|A\| = 0$ if and only if $A = 0$.
2. $\|kA\| = |k|\|A\|$ for any complex scalar k.
3. $\|A + B\| \leq \|A\| + \|B\|$.
4. $\|AB\| \leq \|A\| \cdot \|B\|$.

Corresponding to any vector norm, one can associate with any matrix A a nonnegative quantity defined by $\max\|A\mathbf{x}\|/\|\mathbf{x}\|,\ \|\mathbf{x}\| \neq 0$. This quantity is a function of the matrix A and it satisfies the conditions of a matrix norm. It is called the matrix norm *subordinate* to the vector norm. Because

$$\|A\| = \max\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}, \qquad \|\mathbf{x}\| \neq 0 \tag{B.106}$$

we have

$$\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\| \tag{B.107}$$

where inequality (B.104) is true for $\|\mathbf{x}\| \neq 0$ or for $\|\mathbf{x}\| = 0$. Matrix and vector norms satisfying an inequality of the type (B.104) for all A and $\mathbf{x}$ are said to be *compatible*. Hence, a vector norm and its subordinate matrix norm are always compatible.

A matrix norm of particular importance is the *Euclidean norm*, denoted by $\|A\|_E$ and defined as

$$\|A\|_E = \left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\right)^{1/2} \tag{B.108}$$

The Euclidean norm has the advantage that it is easy to compute. Moreover, it has the important property that its value is invariant under orthogonal transformations (Ref. 3, p. 287).

BIBLIOGRAPHY

1. Franklin, J. N., *Matrix Theory*, Prentice Hall, Englewood Cliffs, NJ, 1968.
2. Murdoch, D. C., *Linear Algebra*, Wiley, New York, 1970.
3. Noble, B. and Daniel, J.W., *Applied Linear Algebra*, 2nd ed., Prentice Hall, Englewood Cliffs, NJ, 1977.
4. Strang, G., *Linear Algebra and Its Applications*, Harcourt Brace Jovanovich, 3rd ed., San Diego, CA, 1988.

Author Index

Subject Index