620.00153
DIX
2009
C1

620.00153

ASTU022552EIT

0079089

Finite Element Methods For Engineers

U. S. Dixit

Indian Institute of Technology, Guwahati

Not to be taken out side
the Library

CENGAGE Learning™

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

CENGAGE
Learning™

Finite Element Methods for Engineers

U. S. Dixit

Development Editor:
Swati Meherishi

Assistant Development Editor:
Ligi John

Senior Product Manager:
Sunil Kumar Agarwal

Production Manager:
Ashok Asthaniya

Cover Designer:
iDesign

Compositor:
Anvi Composers

© 2009 Cengage Learning India Pvt. Ltd

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored, or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, without the prior written permission of the publisher.

For permission to use material from this text or product, email your request to **asia.publishing@cengage.com**

ISBN-13: **978-81-315-0987-6**

ISBN-10: **81-315-0987-7**

Cengage Learning India Pvt. Ltd
418, F.I.E., Patparganj
Delhi 110092

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at:
www.cengage.com/global

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

For product information, visit **www.cengage.co.in**

Adama University
Library

Printed in India
First Impression 2009

To All My Teachers

Contents

Preface *xi*

1. Finite Element Method: A Quick Introduction 1

1.1 Introduction 2
1.2 Axial Rod Problem: Direct FEM Formulation 4
1.3 Beam Problem: Direct FEM Formulation 9
1.4 Power of FEM 14
1.5 Difference Between FEM and FDM 17
1.6 Conclusion 18
Solved Problems 19
Exercises 27
References 33

2. Introduction to Calculus of Variations 35

2.1 Introduction 36
2.2 Functionals 36
2.3 Extremization of a Functional 38
2.4 Obtaining the Variational Form from a Differential Equation 50
2.5 Principle of Virtual Work 55
2.6 Principle of Minimum Potential Energy 56
2.7 Conclusion 59
Solved Problems 59
Exercises 61
References 66

3. Approximation Methods for Solving Differential Equations 67

3.1 Introduction 68
3.2 Ritz Method 68
3.3 Galerkin Method 71
3.4 Least Square Method 73
3.5 Collocation Method 73
3.6 Sub-domain Method 74
3.7 Conclusion 74
Solved Problems 74
Exercises 76
References 79

4. Ritz and Galerkin FEM Formulations 81
4.1 Introduction 82
4.2 Completeness and Compatibility 82
4.3 Concept of Shape Functions 84
4.4 Developing Elemental Equations by Ritz Method 87
4.5 Developing Elemental Equations by Galerkin Method 94
4.6 Conclusion 97
Solved Problems 97
Exercises 101
References 103

5. One-Dimensional FEM Formulations 105
5.1 Introduction 106
5.2 Steady-state Heat Conduction 106
5.3 Longitudinal Deformation of a Rod 117
5.4 Fluid Flow Problem 119
5.5 Conclusion 120
Solved Problems 120
Exercises 124

6. FEM Formulation for Bending of Beams 129
6.1 Introduction 130
6.2 Galerkin FEM Formulation 130
6.3 Ritz FEM Formulation 134
6.4 Conclusion 135
Solved Problems 135
Exercises 139

7. FEM Formulation for Trusses and Frames 143
7.1 Introduction 144
7.2 Formulation for Trusses 144
7.3 An Example for FEM Analysis of Plane Trusses 149
7.4 FEM Formulation for Frames 152
7.5 Assembly Procedure 155
7.6 Conclusion 157
Solved Problems 157
Exercises 159

8. Introduction to Two-Dimensional and Three-Dimensional FEM 163
8.1 Introduction 164
8.2 Triangular Elements 164
8.3 Tetrahedral Elements 169
8.4 Rectangular Elements 170
8.5 Brick Elements 171
8.6 Governing Differential Equation for Two-dimensional Heat Conduction 172
8.7 Weak Form and FEM Formulation 172
8.8 A Note on the Assembly in Two Dimensions 175
8.9 Poisson Equation in Three Dimensions 177
8.10 Fluid Flow Problem 178
8.11 Torsion of Circular and Non-circular Cross-section 179
8.12 A Note on Exploiting Symmetry in FEM 179
8.13 Summary 181
Solved Problems 181
Exercises 183
References 185

9. Numerical Integration 187
9.1 Introduction 188
9.2 One-dimensional Integration Formulae 188
9.3 Two-dimensional Integration Formulae 195
9.4 Conclusion 197
Solved Problems 197
Exercises 198
References 200

10. Two-Dimensional FEM: Detailed Discussion 203
10.1 Introduction 204
10.2 Natural Coordinates and Iso-parametric, Sub-parametric, and Super-parametric Elements 204
10.3 Four-noded Quadrilateral Elements 205
10.4 Serendipity Elements 206
10.5 Eight-noded Curvilinear Elements 208
10.6 Numerical Evaluation of Coefficient Matrices and Right-Hand-side Vectors for Two-dimensional Elements 209
10.7 Conclusion 209
Solved Problems 210
Exercises 212
References 213

11. FEM Formulation for Plane Stress and Plane Strain Problems 215
11.1 Introduction 216
11.2 Basic Equations 217
11.3 Boundary Conditions 218
11.4 Index Notation 218
11.5 FEM Formulation 220
11.6 Types of Elements 223
11.7 Numerical Evaluation of Element Matrices and Vectors 225
11.8 Assembly of Element Matrices 227
11.9 Boundary Conditions and Solutions 228
11.10 Gradient Estimates 228
11.11 Stress Analysis Using an FEM Package: An Example 229
11.12 Summary 231
Solved Problems 231
Exercises 234
References 236

12. FEM Formulation for Free Vibration Problems 237
12.1 Introduction 238
12.2 Vibrations of a Rod 238
12.3 Vibrations of a Beam 244
12.4 Finding out Natural Frequencies of a Damped System 248
12.5 Formulation of Two- and Three-dimensional Problems: A Note 252
12.6 Conclusion 253
Solved Problems 253
Exercises 256
References 259

13. FEM Formulation for Time-History Problems 261
13.1 Introduction 262
13.2 Classification of Partial Differential Equations 262
13.3 Time Response of a Parabolic Equation 264
13.4 Forced Vibration Problems 268
13.5 Conclusion 270
Solved Problems 270
Exercises 273
References 275

14. FEM Formulation for Plate Problem 277
14.1 Introduction 278
14.2 Thin-plate Formulation 278
14.3 Various Thin-plate Elements 281
14.4 Thick-plate Formulation 286
14.5 Locking Phenomenon 288
14.6 An Example 288
14.7 Conclusion 290
Solved Problems 290
Exercises 292
References 293

15. Introduction to Non-Linear FEM 295
15.1 Introduction 296
15.2 Direct Substitution Method 297
15.3 Newton–Raphson Method 301
15.4 Conclusion 302
Solved Problems 302
Exercises 303

16. FEM Formulation for Flow Problems 307
16.1 Introduction 308
16.2 Strip-drawing Process: Governing Equations and Boundary Conditions 309
16.3 Constitutive Relations 313
16.4 A Note on Different Types of FEM Formulations 315
16.5 Weak Formulation for Mixed Pressure–Velocity Formulation 315
16.6 Finite Element Approximation 318
16.7 Finite Element Equations 320
16.8 Discretization of Strip 324
16.9 Application of Boundary Conditions 325
16.10 Post-processing 326
16.11 FEM Formulation of Fluid Flow with Inertia Efect 327
16.12 Conclusion 328
Solved Problems 329
Exercises 330
References 333

17. Error Analysis in Finite Element Method 335
17.1 Introduction 336
17.2 Error Measures 336

17.3 Types of Error Estimates 338
17.4 Conclusion 349
Solved Problems 349
Exercises 351
References 352

18. Solving FEM Problems on a Computer 355
18.1 Introduction 356
18.2 Developing an FEM Code 356
18.3 Finite Element Packages 360
18.4 Conclusion 362
Solved Problems 362
Exercises 366
References 368

19. Miscellaneous Topics 369
19.1 Introduction 370
19.2 Finite Element Solutions Versus Exact Solutions 370
19.3 Accuracy of Derivatives of Primary Variables 373
19.4 Essential and Natural Boundary Conditions 374
19.5 Mesh Refinement 374
19.6 Effect of Geometry of an Element on FEM 375
19.7 Using FEM in Fracture Mechanics 377
19.8 Infinite Elements 381
19.9 Ill-conditioned Systems 381
19.10 Patch Test 383
19.11 Three-dimensional Finite Elements 384
19.12 Conclusion 387
Solved Problems 387
Exercises 390
References 391

Appendix A 393
A.1 Divergence Theorem 394
A.2 The Rule of Differentiation with Respect to a Vector 394
A.3 Solving the System of Equations $[A]\{x\} = \{b\}$ 395
A.4 Pseudocode for Solving a Two-dimensional Heat Conduction Problem 397
Reference 407

Answers to Multiple Choice Questions 409

Index 417

Preface

Finite element method (FEM) started as a method of stress analysis and was found to replace experimental techniques in many cases. It was soon realized that FEM can be used for solving problems relating to heat transfer, fluid mechanics, electromagnetism, and manufacturing process modelling, and a host of other areas. Many variants of the method have been developed to solve a number of engineering problems. Basically, FEM represents a class of numerical methods for solving integral and differential equations, in which the domain of the problem is discretized into a number of elements. Within each element, the solution is approximated by piecewise continuous functions.

About the Book

Finite Element Methods For Engineers is designed to serve as a textbook for an introductory course on FEM for undergraduate and postgraduate students of engineering. FEM is becoming increasingly popular as an undergraduate and postgraduate course, due to its increasing applications in industries and academic research. Despite this, there seems to be dearth of introductory textbooks on FEM, which are suitable for a single-semester course. Considering this difficulty, I prepared lecture notes on FEM with the support of the Quality Improvement Programme (QIP) at IIT Guwahati. Those lecture notes are the foundation of this textbook. This book has been designed in keeping with the requirements of students. It will serve as a simple introduction to FEM for undergraduate and postgraduate courses alike. It will also prove useful to practicing engineers and research scholars looking for a quick introduction to the method.

The book introduces FEM as a general tool for solving governing equations of engineering problems. The examples and exercise problems have been chosen from a number of different fields of engineering.

Content and Structure

The subject matter of this book has been organized into 19 chapters based on my experience of teaching this course at the undergraduate and postgraduate levels.

Students are generally eager to learn the subject in its entirety quickly, so that they may apply it to solve problems in their respective disciplines. Therefore, **Chapter 1** of the book provides a quick introduction to FEM with the help of some one-dimensional FEM problems. After reading this chapter, students can solve one-dimensional problems and write simple programs to solve these problems.

Chapter 2 provides an introduction to the calculus of variations. After reading this chapter, the reader will be able to convert differential equations into their variational form and vice versa.

Chapter 3 presents some classical numerical methods for solving differential equations that provide the basis for FEM.

Two popular FEM formulations techniques—Ritz FEM and Galerkin FEM are presented in **Chapter 4**.

Chapter 5 covers the application of FEM to a number of different one-dimensional engineering problems.

Chapter 6 discusses the FEM formulation of the beam bending problem.

FEM formulation of trusses and frames is presented in **Chapter 7**.

Following the comprehensive discussions on one-dimensional FEM formulations in Chapters 1 to 7, **Chapter 8** introduces FEM formulations of two- and three-dimensional problems.

Chapter 9 is devoted to numerical integration techniques used for generating the coefficient matrices and right-hand-side vectors.

Detailed discussions on FEM formulations, particularly iso-parametric, sub-parametric, and super-parametric formulations, can be found in **Chapter 10**.

Chapter 11 elucidates the concepts of two-dimensional elasticity problems.

Chapter 12 presents free vibration problems.

Forced vibration and other time-dependent problems are discussed in **Chapter 13**.

Chapter 14 presents an analysis of plate bending problems.

Chapters 15 and **16** describe FEM formulations for non-linear problems.

Error analysis is discussed in **Chapter 17**.

Chapter 18 provides guidelines for writing codes for FEM implementation and using commercially available FEM packages.

Chapter 19 covers a number of miscellaneous topics related to FEM to complete the treatment of FEM as a single-semester course.

Appendix A reviews mathematical concepts such as divergence theorem and differentiation with respect to vectors. Pseudocodes for solving matrix equations and two-dimensional heat conduction problems are also given in this appendix.

Pedagogy

The content of this book has been designed to appeal to first-time students of FEM. The book is written in a lucid manner with an ample number of examples that illustrate concepts. Each chapter begins with a mini table of contents that enables readers to view the list of inclusions within that chapter. An introduction at the beginning of every chapter serves to link the content within the chapter to preceding and subsequent discussions. Examples are interspersed through the text and solved

problems at the end of each chapter demonstrate problem solving methodologies. Every chapter includes a summary which provides a bird's eye view of the concepts discussed. The students' understanding of the concepts can be tested through end-of-chapter exercises which include problems and objective-type questions. It is essential for the reader to go through the examples and solved problems and then solve exercise problems for a complete understanding of FEM. Answers to the multiple choice questions have been included at the end of the book. A list of references has been provided at the end of each chapter for those interested in further reading.

Acknowledgements

A lot of people have helped in the preparation of this book directly or indirectly. It is not possible for me to name everyone here, but there are a few names that must be mentioned.

I am thankful to QIP, IIT Guwahati and QIP Coordinator Prof. Rajeev Tiwari for their support in converting my lecture notes into a textbook. I have learnt a lot from Prof. P.M. Dixit of IIT Kanpur. I am thankful to him for his teaching and valuable discussions on FEM. A number of teachers (current and erstwhile) at IIT Kanpur have been a source of inspiration for me. Of them, I'd especially like to acknowledge Dr B.P. Singh, Dr V. Sundararajan, Dr Amitabha Ghosh, Dr Gautam Biswas, Dr K. Muralidhar, Dr B.N. Banerjee, Dr P.K. Basudhar, and Dr N.N. Kishore.

My friends and colleagues have helped tremendously in the development of this text through both encouragement and criticism. For this, I am thankful to Dr U.K. Saha and Dr A.K. Singh of IIT Guwahati, Dr K.M. Pandey of NIT Silchar, and Dr Sudipto Ghosh of IIT Kharagpur. I am thankful to all students who have either taken a course on FEM and/or worked on their projects with me. My students Sandeep Somani, M. Tiwari, Srinivasa Rao Natra, and Karimulla Shaik deserve special mention.

I acknowledge Dr Sumit Basu of IIT Kanpur and Dr D. K. Sehgal of IIT Delhi for providing valuable review feedback on the manuscript. The editorial team at Cengage Learning has been very cooperative through the writing of this book. I must thank Ligi John for doing a commendable editing job. I cannot adequately acknowledge in words the support and encouragement extended by Swati Meherishi towards development and time-scheduling of the book.

I hope that the book will be well received by both teachers and students. The readers' suggestions about further improvement of the book may be sent to uday@iitg.ernet.in or usd1008@yahoo.com.in. All such feedback will be highly appreciated.

U. S. Dixit

1

CHAPTER

Finite Element Method: A Quick Introduction

1.1 Introduction

1.2 Axial Rod Problem: Direct FEM Formulation

1.3 Beam Problem: Direct FEM Formulation

1.4 Power of FEM

1.5 Difference Between FEM and FDM

1.6 Conclusion

1.1 INTRODUCTION

Finite element method (FEM) is a numerical method used to solve differential and integral equations. Since the behaviour of almost all physical systems can be represented with the help of differential equations, FEM can be used to analyse a number of physical problems. It was originally developed as a technique to analyse complex structural systems. The discovery of FEM is often attributed to Courant (1943), while its use in structural (aircraft) analysis was first reported by Turner, et al. (1956). FEM was named by Clough (1960). This method can be applied to various areas of study. With the help of this method, differential equations can be reduced to algebraic equations by using suitable approximations for the variables over the elements. This method also makes the application of complicated boundary conditions to physical problems easy. Complicated geometries and variations of material properties can also be handled using FEM. Due to these advantages, FEM has emerged as a versatile and powerful tool in computational engineering.

This chapter provides an introduction to FEM. For better understanding of the method, two one-dimensional problems have been considered in the chapter—the axial rod problem and the beam problem. In the *axial rod problem*, one is usually interested in finding the axial displacement of each point of the rod under the action of prescribed loads. However, in the *beam problem*, the vertical deflection and its slope need to be determined at each point. Thus, the axial rod problem is a one degree-of-freedom per node problem, while the beam problem is a two degrees-of-freedom per node problem. However, it will be seen that the finite element procedure is similar for both the problems. In fact, it is similar for any problem, irrespective of its dimension and degrees of freedom. The following steps constitute FEM: (i) pre-processing, (ii) obtaining elemental equations, (iii) assembly, (iv) application of boundary conditions, (v) solution and (vi) post-processing.

The individual steps are discussed below in detail.

Pre-processing In this step, the geometry is discretized into a number of small elements. The elements may be of different shapes. Each element is characterized by a number of points called *nodes*. The complete system of elements is called a *mesh* and the process of generating the elements is called *mesh generation*. Figure 1.1 shows finite element meshes with different types of elements:

- Figure 1.1a shows a mesh consisting of 3 two-noded linear elements. There are four nodes in the mesh. Node 2 is common between elements 1 and 2. Similarly, node 3 is common between elements 2 and 3.
- Figure 1.1b shows a mesh consisting of 3 three-noded elements. Element 2 has nodes 3, 4, and 5, in which node 4 is the internal node of the element, node 3 is shared by element 1, and node 5 is shared by element 3.
- Figure 1.1c shows a mesh consisting of 8 three-noded triangular elements. Note that in this case, node 5 is common to all the elements.
- Figure 1.1d shows a mesh consisting of 12 four-noded rectangular elements.
- Figure 1.1e shows a three-dimensional mesh consisting of several tetrahedral elements. This mesh has been generated using the pre-processor of a commercial FEM package.

All the meshes shown in Fig. 1.1 consist of elements of the same size. In general, a mesh may consist of elements of different sizes and may even contain different types of elements.

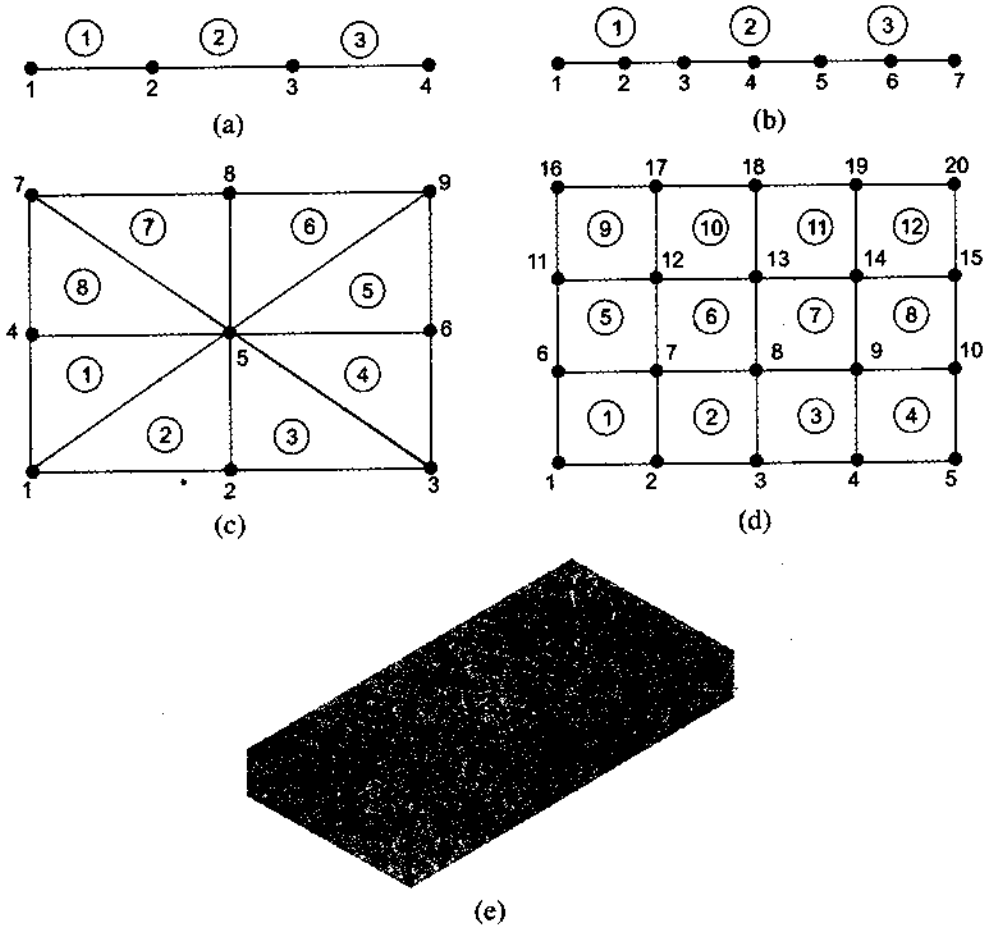


Figure 1.1 *Finite element mesh using (a) two-noded linear elements (b) three-noded linear elements (c) three-noded triangular elements (d) four-noded rectangular elements (e) four-noded tetrahedral elements*

Obtaining Elemental Equations In this step, algebraic equations are obtained for each element. Although the equations may be derived by a number of methods, they have been derived in this chapter using direct finite element formulation, in which algebraic equations are obtained directly from the physics of the problem.

Assembly In this step, the elemental stiffness equations are assembled to yield a global system of equations.

Application of Boundary Conditions In this step, the assembled system of equations is modified by applying the prescribed boundary conditions.

Solution In this step, the modified global system of equations is solved, and the solution obtained is in the form of values of primary variables at nodes, such as nodal displacements in axial rod problem and nodal deflections and slopes in beam problem.

Post-processing In this step, various secondary quantities are computed from the obtained solution. For example, stresses and strains are computed from the obtained nodal displacements in axial rod problem.

1.2 AXIAL ROD PROBLEM: DIRECT FEM FORMULATION

Consider an axial rod loaded at one end with a force P as shown in Fig. 1.2. In general, the rod may be made of a non-homogeneous material, have a variable cross-section, and be loaded with concentrated forces at different points or with distributed forces at different segments of the rod. However, to introduce FEM, a trivial problem of uniform axial rod loaded with force P at one end is chosen. Deflections, strains, and stresses at different points of the rod will be calculated using FEM.

A differential equation governing the problem can be obtained with axial deflection u as the independent variable and point coordinate x as the dependent variable. In FEM, the differential equation is converted to an algebraic equation. However, for this particular problem, the algebraic equation can be obtained directly from the physics of the problem. Hence, the methodology described here is called *direct finite element formulation*.

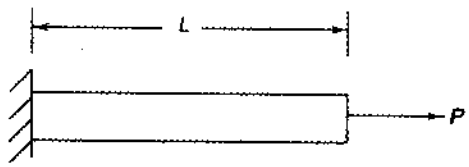


Figure 1.2 *An axial rod loaded at one end*

1.2.1 Pre-processing

The first step in this method involves the discretization of the rod into a number of small segments known as *elements*. For example, in Fig. 1.3, the rod has been divided into three elements. The end points of each element are called *nodes*. Thus, in this problem, there are a total of three elements and four nodes. Each element is designated by its two nodes and coordinates of each node are noted. This step is called *pre-processing or mesh generation*.

Figure 1.3 *Pre-processing of the axial rod problem: (a) discretization of the rod (b) a typical element*

Figure 1.3b shows a typical element, e, which is acted upon by two forces, F_i and F_j, respectively. These forces may be internal or external. For example, for the problem shown in Fig. 1.3, only node 4 experiences external force, while all other nodes are subjected to internal forces. In Fig. 1.3b, the force on node i acts in the left direction, whereas that on node j acts in the right direction. These directions have been chosen such that the element is subjected to tensile forces at both the nodes. Since the forces are acting away from the elements at both the nodes, they produce tensile stresses. One could follow a convention, in which both the forces are directed towards right. In that case, however, the force at node i will produce compressive stress, while that at node j will produce tensile stress.

1.2.2 Elemental Stiffness Matrix and Load Vector

In order to obtain governing algebraic equations, displacement in each element is assumed to be linear. This will indeed be so, if the element is composed of homogeneous material following Hooke's law, has uniform cross-sectional area, and loads are only point loads acting at the ends. Figure 1.3b shows a general element, with end nodes designated by i and j. The tensile strain, ε_t, in the element is given by

$$\varepsilon_t = \frac{u_j - u_i}{h} \tag{1.1}$$

where u_i and u_j are axial displacements at nodes i and j, respectively and h is the element length (equal to $L/3$ in this case). The corresponding tensile stress, σ_t, is given as follows

$$\sigma_t = E\,\frac{u_j - u_i}{h} \tag{1.2}$$

where E is Young's modulus of elasticity. The force F_j applied at the jth node is defined as the product of stress, σ_t, and the area of cross-section, A. Hence

$$AE\,\frac{u_j - u_i}{h} = F_j \tag{1.3}$$

Thus, a relationship between force F_j and nodal deflections is obtained. A similar relation between the force F_i and nodal deflections can be obtained in the following manner. The compressive strain, ε_c, in the element is given by

$$\varepsilon_c = \frac{u_i - u_j}{h} \tag{1.4}$$

and the corresponding compressive stress, σ_c, is given by

$$\sigma_c = E\,\frac{u_i - u_j}{h} \tag{1.5}$$

Hence, the relation for the force F_i can be written as

$$AE\,\frac{u_i - u_j}{h} = -F_i \tag{1.6}$$

Note the negative sign at the right hand side of Eq. (1.6). This is because force F_i is assumed to be tensile in Fig. 1.3b. Also note that Eqs (1.3) and (1.6) suggest that $F_i = F_j$. This indeed is the condition for the rod to be in equilibrium and hence Eqs (1.3) and (1.6) are the same. However, we retain both equations at this stage and write them in the matrix form as

$$\frac{AE}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_i \\ u_j \end{Bmatrix} = \begin{Bmatrix} -F_i \\ F_j \end{Bmatrix} \tag{1.7}$$

In the compact form, they can be written as

$$[k^e]\{u^e\} = \{F^e\} \tag{1.8}$$

where

$$[k^e] = \frac{AE}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{1.9}$$

$$\{u^e\} = \begin{Bmatrix} u_i \\ u_j \end{Bmatrix} \tag{1.10}$$

and

$$\{F^e\} = \begin{Bmatrix} -F_i \\ F_j \end{Bmatrix} \tag{1.11}$$

Compare Eq. (1.8) with the equation for a spring loaded with force F:

$$kx = F \tag{1.12}$$

In analogy with this, matrix $[k^e]$ is called the *elemental stiffness* matrix and its elements will have units N/m in SI system, $\{u^e\}$is called *elemental displacement* or *primary variable vector* and $\{F^e\}$ is called the *elemental load vector*.

Let us observe the elemental system of equations given by Eq. (1.7). This system cannot be solved to yield the values of u_i and u_j because of the following reasons:

1. In general, F_i and F_j are internal forces, which are unknown.
2. Even if the values of F_i and F_j are known, the elemental stiffness matrix cannot be inverted to yield solution, because this matrix is singular and its rank is 1. Physically, this means that just by prescribing the values of two end forces, one cannot predict the displacement because infinite numbers of rigid body modes are possible.

In order to overcome the first difficulty, i.e., to get rid of internal forces, the elemental stiffness matrices are assembled. The second difficulty is overcome by prescribing one geometric boundary condition (i.e., prescribing axial deflection at one node). The assembly procedure and the application of boundary conditions are illustrated in the following subsections.

1.2.3 Assembly Procedure

For the given problem, the elemental equations for three elements are written as follows:

$$\frac{AE}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ F_2 \end{Bmatrix} \tag{1.13}$$

$$\frac{AE}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -F_2 \\ F_3 \end{Bmatrix} \tag{1.14}$$

$$\frac{AE}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} -F_3 \\ F_4 \end{Bmatrix} \tag{1.15}$$

These elemental stiffness equations can be assembled to yield global stiffness equations, considering u_1, u_2, u_3, and u_4 to be unknowns. In the assembled system of equations, internal forces will be eliminated. There are a number of ways in which the assembly operation may be understood. We follow a simple approach in which elemental system of equations for each element is written in global form and then these global equations are algebraically added. Thus, the Eqs (1.13) to (1.15) are written as

$$\frac{AE}{h}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ F_2 \\ 0 \\ 0 \end{Bmatrix} \tag{1.16}$$

$$\frac{AE}{h}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} 0 \\ -F_2 \\ F_3 \\ 0 \end{Bmatrix} \tag{1.17}$$

$$\frac{AE}{h}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ -F_3 \\ F_4 \end{Bmatrix} \tag{1.18}$$

Adding Eqs (1.16) to (1.18), we obtain

$$\frac{AE}{h}\begin{bmatrix} 1+0+0 & -1+0+0 & 0+0+0 & 0+0+0 \\ -1+0+0 & 1+1+0 & 0-1+0 & 0+0+0 \\ 0+0+0 & 0-1+0 & 0+1+1 & 0+0-1 \\ 0+0+0 & 0+0+0 & 0+0-1 & 0+0+1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ 0 \\ 0 \\ F_4 \end{Bmatrix} \tag{1.19}$$

Note that in the system of equations given by Eq. (1.19), internal forces F_2 and F_3 are eliminated. However, F_1 and F_4 still remain. They can be eliminated only by applying boundary conditions. Also note that the assembled global stiffness matrix is singular, with rank 3. Thus, one nodal displacement needs to be prescribed.

1.2.4 Application of Boundary Conditions and Solution

For the problem discussed above, F_4 is equal to the externally applied load P. This is called force or natural boundary condition. However, F_1 is unknown. On the node at which F_1 acts, $u_1 = 0$. This is called essential or geometric boundary condition. There are various ways in which this boundary condition can be applied. A simple way is to replace the first equation by $u_1 = 0$, which is same as $(AE/h_1)u_1 = 0$. Thus, the assembled system of equations, after the application of boundary conditions, becomes

$$\frac{AE}{h}\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ P \end{Bmatrix} \tag{1.20}$$

Compare Eq. (1.19) and Eq. (1.20). These equations differ only in the first row. However, note that Eq. (1.20) has not been obtained by modifying the first row of Eq. (1.19). Instead, the first row of Eq. (1.19) has been replaced by the equation corresponding to the boundary conditions. This means that the first equation of system of equations given by Eq. (1.19) is still valid. However, we do not utilize it for solving for nodal displacements as its right hand side is unknown. Instead we use the simple equation given by $(AE/h_1)u_1 = 0$, which has been supplied by the geometric boundary condition.

There are various methods of solving the linear system of equations given by Eq. (1.20). The solution yields

$$u_1 = 0,\ u_2 = \frac{PL}{3AE},\ u_3 = \frac{2PL}{3AE},\ u_4 = \frac{PL}{AE} \tag{1.21}$$

Notice that these are exact displacements, obtainable from elementary strength of materials. This is no surprise, as the exact displacement function is linear and a linear displacement field (via constant strain) was assumed in each element.

1.2.5 Post-processing

After the nodal displacements have been obtained, strains and stresses in the elements can be computed. This is known as post-processing. In this example, strain in the element 2, $\varepsilon^{(2)}$, is given by

$$\varepsilon^{(2)} = \frac{u_3 - u_2}{h} = \frac{u_3 - u_2}{(L/3)} = \frac{P}{AE} \tag{1.22}$$

and the stress, $\sigma^{(2)}$, is given by

$$\sigma^{(2)} = E\varepsilon^{(2)} = \frac{P}{A} \tag{1.23}$$

In the same way, strains and stresses in other elements may also be computed. The displacement at any point inside the element can be found out by linear interpolation.

1.3 BEAM PROBLEM: DIRECT FEM FORMULATION

Consider a beam rigidly fixed at the ends and loaded in the centre by a load P (Fig. 1.4). In general, the beam can be of any arbitrary cross-section and can be loaded with any complex loading function. For the sake of simplicity, a beam of uniform cross-sectional area is considered and only the deflection caused by bending is considered. Deflection due to shear is not taken into consideration.

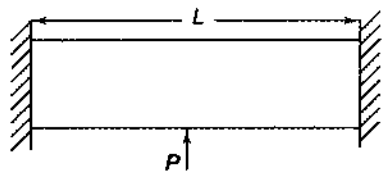


Figure 1.4 *Fixed–fixed beam with a central load*

1.3.1 Pre-processing

The entire beam can be divided into 2 two-noded equal elements (Fig. 1.5a) such that element 1 is composed of nodes 1 and 2 and element 2 is composed of nodes 2 and 3. It is important to discuss the concept of connectivity matrix here. *Connectivity matrix* is a simple representation of element–node relations, in which the row indicates element number, the column indicates local (elemental) node number, and the element of the matrix denotes the global node number. Thus, the connectivity matrix for the present mesh may be written as

$$\begin{array}{cc} \text{local node 1} & \text{local node 2} \\ \left[\begin{array}{cc} 1 & 2 \\ 2 & 3 \end{array}\right] & \begin{array}{l} \text{element 1} \\ \text{element 2} \end{array} \end{array} \tag{1.24}$$

The first row of the connectivity matrix given in Eq. (1.24) shows that global nodes for the element 1 are 1 and 2, and the second row shows that the global nodes for the element 2 are 2 and 3. Given the connectivity matrix and coordinates of the node, the mesh can be easily constructed.

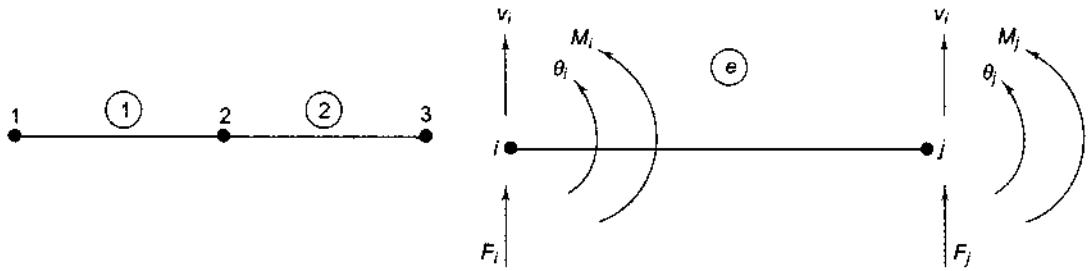


Figure 1.5 *Pre-processing of the beam problem: (a) discretization of the beam (b) a typical element*

1.3.2 Elemental Stiffness Equations

From elementary mechanics of materials, it is known that the deformation of axial rod is characterized by axial displacement of each point, whereas in the beam problem, both the vertical displacement and the slope at each point needs to be prescribed. Thus, a typical node in the element has two degrees of freedom: vertical deflection and slope. Figure 1.5b shows a typical two-noded element. The force acting on the two nodes, i and j, is denoted by F_i and F_j, respectively, while the corresponding moments are denoted by M_i and M_j. In general, F_i depends on the elastic property of the element and displacements at the two nodes. For a linear elastic material with small deflections, the force is given by

$$F_i = k_{11} v_i + k_{12} \theta_i + k_{13} v_j + k_{14} \theta_j \tag{1.25}$$

where ks are coefficients dependent on the geometry and material of the element. Mathematically, this means that the vertical force at a node is a linear function of the vertical deflections and slopes at the nodes. Similar equations can be written for M_i, F_j, and M_j. Thus, the elemental equations can be represented as

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix} \begin{bmatrix} v_i \\ \theta_i \\ v_j \\ \theta_j \end{bmatrix} = \begin{bmatrix} F_i \\ M_i \\ F_j \\ M_j \end{bmatrix} \tag{1.26}$$

In order to derive the values of coefficients, we proceed as follows. Assume that v_j and θ_j are both equal to 0 in Fig. 1.5b. The first two equations of the system of equations given by Eq. (1.26) reduce to

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \begin{Bmatrix} v_i \\ \theta_i \end{Bmatrix} = \begin{Bmatrix} F_i \\ M_i \end{Bmatrix} \tag{1.27}$$

The system then represents a cantilever beam loaded by a vertical force F_i and moment M_i at one end. The deflection and slope of that end can be obtained from elementary strength of materials using the following equations:

$$\frac{F_i h^3}{3EI} - \frac{M_i h^2}{2EI} = v_i \tag{1.28}$$

$$-\frac{F_i h^2}{2EI} - \frac{M_i h}{EI} = \theta_i \tag{1.29}$$

where h is the length of the element and is equal to $L/2$ and I is the second moment of inertia of the cross-section of the beam about the centroidal horizontal axis. (It is assumed that the bending is symmetric, i.e., the cross-section of the beam is symmetric and the loading is along the line of symmetry.) In the matrix form, the equations can be written as

$$\begin{bmatrix} \frac{h^3}{3EI} & \frac{-h^2}{2EI} \\ \frac{-h^2}{2EI} & \frac{h}{EI} \end{bmatrix} \begin{Bmatrix} F_i \\ M_i \end{Bmatrix} = \begin{Bmatrix} v_i \\ \theta_i \end{Bmatrix} \tag{1.30}$$

Inverting the above equations, we obtain

$$\frac{EI}{h^3} \begin{bmatrix} 12 & 6h \\ 6h & 4h^2 \end{bmatrix} \begin{Bmatrix} v_i \\ \theta_i \end{Bmatrix} = \begin{Bmatrix} F_i \\ M_i \end{Bmatrix} \tag{1.31}$$

Comparing Eq. (1.31) with Eq. (1.27), we obtain

$$k_{11} = \frac{12EI}{h^3},\ k_{12} = k_{21} = \frac{6EI}{h^2},\ k_{22} = \frac{4EI}{h} \tag{1.32}$$

In order to derive other terms of the first two columns of Eq. (1.26), we make use of following equations of equilibrium:

$$F_i + F_j = 0 \tag{1.33}$$

$$M_i + M_j - F_i h = 0 \tag{1.34}$$

The third equation of Eq. (1.26) gives

$$k_{31} v_i + k_{32} \theta_i = F_j = -F_i = -(k_{11} v_i + k_{12} \theta_i) \tag{1.35}$$

Hence

$$k_{31} = -k_{11} \quad k_{32} = -k_{12} \tag{1.36}$$

From the fourth equation of Eq. (1.26), we get

$$k_{41} v_i + k_{42} \theta_i = M_j = -M_i + F_i h = -(k_{21} v_i + k_{22} \theta_i) + (k_{11} v_i + k_{12} \theta_i) h \tag{1.37}$$

If we substitute the expressions for k_{11}, k_{12}, k_{21}, and k_{22} from Eq. (1.32) into Eq. (1.37) and compare the coefficients of v_i and θ_i on the left and right hand side of the equation, we obtain

$$k_{41} = 6EI/h^2 \quad \text{and} \quad k_{42} = 2EI/h^2 \tag{1.38}$$

The other elements of the matrix can be obtained by keeping node i fixed. If node i is fixed, the third and fourth equations of Eq. (1.26) reduce to

$$\begin{bmatrix} k_{33} & k_{34} \\ k_{43} & k_{44} \end{bmatrix} \begin{Bmatrix} v_j \\ \theta_j \end{Bmatrix} = \begin{Bmatrix} F_j \\ M_j \end{Bmatrix} \tag{1.39}$$

The system then represents a cantilever beam loaded by a vertical force F_j and moment M_j at one end. The deflection and slope of that end can be obtained from elementary strength of materials. They are given by the following equations:

$$\frac{F_j h^3}{3EI} + \frac{M_j h^2}{2EI} = v_j \tag{1.40}$$

$$\frac{F_j h^2}{2EI} + \frac{M_j h}{EI} = \theta_j \tag{1.41}$$

In the matrix form, the equations can be written as

$$\begin{bmatrix} \dfrac{h^3}{3EI} & \dfrac{h^2}{2EI} \\ \dfrac{h^2}{2EI} & \dfrac{h}{EI} \end{bmatrix} \begin{Bmatrix} F_j \\ M_j \end{Bmatrix} = \begin{Bmatrix} v_j \\ \theta_j \end{Bmatrix} \tag{1.42}$$

Inverting the above equation, we obtain

$$\frac{EI}{h^3} \begin{bmatrix} 12 & -6h \\ -6h & 4h^2 \end{bmatrix} \begin{Bmatrix} v_j \\ \theta_j \end{Bmatrix} = \begin{Bmatrix} F_j \\ M_j \end{Bmatrix} \tag{1.43}$$

Comparing Eq. (1.43) with Eq. (1.39), we obtain

$$k_{33} = \frac{12EI}{h^3}, \; k_{34} = k_{43} = \frac{6EI}{h^2}, \; k_{44} = \frac{4EI}{h} \tag{1.44}$$

Similarly, from equilibrium consideration, we can obtain

$$k_{13} = -12EI/h^3 \quad k_{14} = -k_{23} = 6EI/h^2 \quad k_{24} = 2EI/h^2 \tag{1.45}$$

Thus, the elemental stiffness matrix is given by

$$\frac{EI}{h^3} \begin{bmatrix} 12 & 6h & -12 & 6h \\ 6h & 4h^2 & -6h & 2h^2 \\ -12 & -6h & 12 & -6h \\ 6h & 2h^2 & -6h & 4h^2 \end{bmatrix} \tag{1.46}$$

The resulting stiffness matrix is exact, not approximate, for linear elastic bending of uniform beam element.

1.3.3 Assembly Procedure

In order to perform the assembly, elemental equations can be written in global form. The first elemental equation in global coordinate system is written as

$$\frac{EI}{h^3}\begin{bmatrix} 12 & 6h & -12 & 6h & 0 & 0 \\ 6h & 4h^2 & -6h & 2h^2 & 0 & 0 \\ -12 & -6h & 12 & -6h & 0 & 0 \\ 6h & 2h^2 & -6h & 4h^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}\begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \\ v_3 \\ \theta_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ M_1 \\ F_2^{(1)} \\ M_2^{(1)} \\ 0 \\ 0 \end{Bmatrix} \tag{1.47}$$

Here, the superscript (1) on forces and moments indicate contributions from element 1. The second elemental equation in global coordinates is given by

$$\frac{EI}{h^3}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 6h & -12 & 6h \\ 0 & 0 & 6h & 4h^2 & -6h & 2h^2 \\ 0 & 0 & -12 & -6h & 12 & -6h \\ 0 & 0 & 6h & 2h^2 & -6h & 4h^2 \end{bmatrix}\begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \\ v_3 \\ \theta_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ F_2^{(2)} \\ M_2^{(2)} \\ F_3^{(1)} \\ M_3^{(1)} \end{Bmatrix}. \tag{1.48}$$

Adding the system of equations given by Eq. (1.47) and Eq. (1.48), the following global system of equations is obtained:

$$\frac{EI}{h^3}\begin{bmatrix} 12 & 6h & -12 & 6h & 0 & 0 \\ 6h & 4h^2 & -6h & 2h^2 & 0 & 0 \\ -12 & -6h & 12+12 & -6h+6h & -12 & 6h \\ 6h & 2h^2 & -6h+6h & 4h^2+4h^2 & -6h & 2h^2 \\ 0 & 0 & -12 & -6h & 12 & -6h \\ 0 & 0 & 6h & 2h^2 & -6h & 4h^2 \end{bmatrix}\begin{Bmatrix} v_1 \\ \theta_1 \\ v_2 \\ \theta_2 \\ v_3 \\ \theta_3 \end{Bmatrix} = \begin{Bmatrix} F_1 \\ M_1 \\ P \\ 0 \\ F_3 \\ M_3 \end{Bmatrix} \tag{1.49}$$

Note that

$$F_2^{(1)} + F_2^{(2)} = P \quad \text{and} \quad M_2^{(1)} + M_2^{(2)} = 0 \tag{1.50}$$

At each node, the internal forces and moments of two elements sharing the node cancel each other. The external forces are shared by both the elements.

1.3.4 Boundary Conditions and Solution

It can be verified that the rank of assembled global stiffness matrix is 4. Hence, at least two essential boundary conditions are required to obtain the solution. However, in this case, we have four essential (geometric) boundary conditions:

$$v_1 = \theta_1 = v_3 = \theta_3 = 0 \tag{1.51}$$

From Eq. (1.51), it can be seen that the unknowns for the problem are v_2 and θ_2; only two equations are needed to determine these. We choose third and fourth equations of Eq. (1.49) for this purpose as the right hand side of these equations is known. After substituting the values of the prescribed degrees of freedom, these equations reduce to

$$\frac{EI}{h^3}\begin{bmatrix} 24 & 0 \\ 0 & 8h^2 \end{bmatrix}\begin{Bmatrix} v_2 \\ \theta_2 \end{Bmatrix} = \begin{Bmatrix} P \\ 0 \end{Bmatrix} \tag{1.52}$$

Solving Eq. (1.52), we obtain

$$v_2 = \frac{Ph^3}{24EI} = \frac{P(L/2)^3}{24EI} = \frac{PL^3}{192EI} \quad \text{and} \quad \theta_2 = 0 \tag{1.53}$$

It can be verified that the same values are obtained from elementary strength of materials.

1.3.5 Post-processing

By finite element analysis, we obtain nodal deflections and slope. The task of post-processing is to find out the slopes and deflection at any point of the beam, the shear force, and the bending moment. Knowing the shear force and bending moment at any section of the beam, the stresses can be calculated. In Section 1.2.5, it was suggested that the displacement at any point inside the element can be found by linear interpolation of the nodal displacements. Many times, students make the mistake of linearly interpolating the nodal deflections in a beam problem too; in doing so, the information of nodal slopes is not utilized. If the slopes and deflections at the nodes are known, displacement can be expressed as a cubic polynomial in an element. The deflection at a point can be found by evaluating the value of the cubic polynomial at that point. The slope at a point can be determined from the first derivative of the cubic polynomial, the bending moment from the second derivative, and the shear force from the third derivative.

1.4 POWER OF FEM

Two simple one-dimensional examples have been considered to illustrate the finite element methodology. However, these problems cannot be considered to be a part of the set of representative problems

that can demonstrate the power of FEM as a tool for solving the differential equations. For one-dimensional problems, there are many methods that can compete with FEM. The utility of FEM is realized while solving two-dimensional or three-dimensional problems, especially those involving complicated geometry and non-homogeneous and anisotropic material properties. Some examples of the application of FEM are discussed in this section.

Figure 1.6 shows an L-block with a groove, deformed by applied loads. The deformation of the block is studied using FEM, by discretizing the domain into a number of quadrilateral elements. Note the refined mesh near the groove. Because of the presence of a sharp stress gradient in the neighbourhood of the groove and the need for accurate determination of the stresses at the groove, this region consists of elements of small sizes.

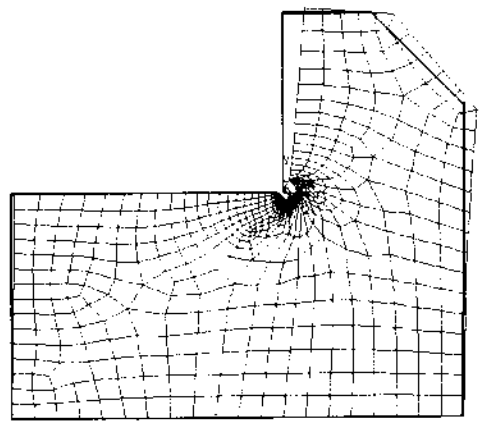


Figure 1.6 *Deformed (meshed) and undeformed (with thick boundaries) L-block with a groove*

Figure 1.7 shows uniform and non-uniform meshing of a plate that is bent with the help of a laser beam. The laser beam scans the plate across its width at the mid-length of the plate. One end of the plate is clamped. Due to temperature gradient, the plate gets bent. The bend angle may be obtained by thermo-mechanical analysis. In the decoupled scheme of the analysis by FEM, first the temperature field is determined and then an elasto-plastic structural analysis is carried out by inputting the temperature field. The plate is meshed using three-dimensional elements. In Fig. 1.7, one meshing scheme is the uniform meshing with brick elements, while the other is a non-uniform meshing with tetrahedral elements. In the non-uniform meshing scheme, the mesh is refined around the scanning path of the laser beam.

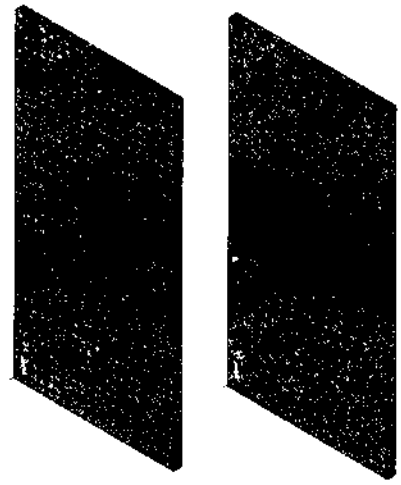
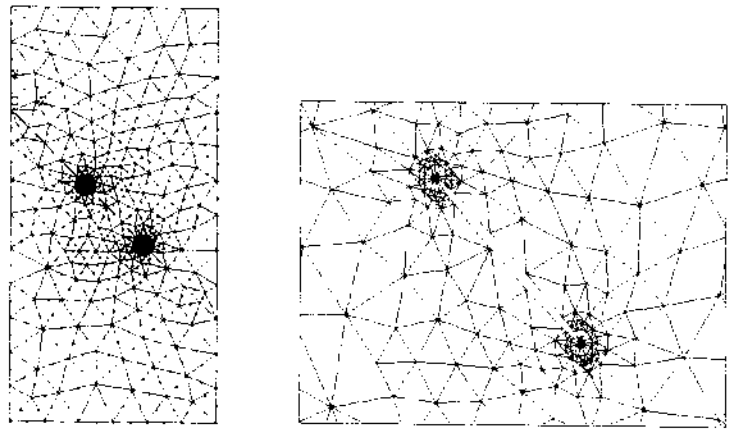


Figure 1.7 *Uniform and non-uniform three-dimensional meshing of a plate deformed using a laser beam*



(a) (b)

Figure 1.8 *Figure 1.8. A plate with cracks (a) non-uniform meshing with six-noded triangular nodes (b) magnified view of the region surrounding the crack*

Figure 1.8a shows a plate with a crack. The plate has been meshed with six-noded triangular elements. A magnified view containing the crack is shown in Fig. 1.8b. Note that the mesh is refined near the crack tips as very high stress gradients exist near them. Figure 1.8 presents an example of the problems of fracture mechanics.

These examples illustrate that FEM may be used to effectively solve a number of engineering problems. The application of FEM is not restricted to solid mechanics problems; this method may be

utilized for solving problems involving fluid mechanics, electromagnetism, solid–fluid interactions, etc. Since FEM can be used in solving any problem that can be modelled in the form of differential equation(s), it has an immense scope.

1.5 DIFFERENCE BETWEEN FEM AND FDM

Before the advent of FEM, *finite difference method* (FDM) was considered to be one of the most powerful techniques for numerically solving the differential equations. A beginner is often keen to know the difference between FEM and FDM. Therefore, in this section, the difference between the two techniques is briefly described.

The basic concepts of FEM may be described as follows:

- The solution is approximated in the form of piecewise continuous functions. These functions may be algebraic, trigonometric, or any other type. Note that in the problem discussed in Section 1.2, displacement inside an element is assumed to be a linear function of the coordinate in the length direction, while in the problem discussed in Section 1.3, beam deflection inside an element is a cubic function of the length–direction coordinate.
- Coefficients of these functions are obtained in such a manner that the errors are minimized. Note that in the problems studied in Section 1.2 and 1.3, we obtained the nodal values instead of the coefficients of the approximating polynomial functions. However, we obtained coefficients indirectly. For example, for a straight line $y = mx + c$, if you obtain the y-coordinates y_1 and y_2 at two x-coordinates x_1 and x_2, respectively, you indirectly obtain m and c as well. The slope, m, is given by $(y_2 - y_1)/(x_2 - x_1)$. Similarly, c can be obtained. Also, note that in Section 1.2 and 1.3 we have not explicitly mentioned about the error. We have in fact chosen the problems in which FEM provides zero errors in the solutions. In general, FEM can be based on minimizing one of the different error measures. These concepts will become clearer in due course of the book.

We have different FEM formulation techniques depending on the form of functions, definition of different errors, and procedure for minimizing errors.

In FDM, derivatives are replaced by discrete analogue of derivatives. For example

$$\frac{du}{dx} \approx \frac{u(x+h) - u(x)}{h} \tag{1.54}$$

where h is a small finite step length. Thus, the system of differential equations gets approximated to a system of linear simultaneous algebraic equations, just as in FEM. However, the basic concept here is different from FEM. In this method, the variable is not approximated by any piecewise continuous function.

In certain cases, the FEM and FDM may provide the same solution. For example, in the problem of a rod subjected to an axial load (as in Section 1.2), the axial displacement in an element may approximated as

$$u = a + bx = \left(1 - \frac{x}{h}\right)u_1 + \frac{x}{h}\, u_2 \tag{1.55}$$

Adama University

where u_1 and u_2 are axial displacements at the two end nodes and h is the length of the element. In this case

$$\frac{du}{dx} = \frac{u_2 - u_1}{h} \tag{1.56}$$

which is the same as finite difference approximation. On the other hand, if axial displacement is approximated by some function such as

$$u = a + b \sin x \tag{1.57}$$

then the derivative approximation in FEM will be different from the derivative approximation in FDM.

It is not convenient to apply FDM method to a problem involving a boundary of very complex geometry and involving non-homogeneous material. If a proper choice of piecewise continuous polynomials is made, FEM may take much less time than FDM. However, in certain problems, FDM has been found to provide better results due to better approximation of differential equations and boundary conditions.

1.6 CONCLUSION

In this chapter, finite element method has been introduced by taking one-dimensional problems as examples. The physics of a rod subjected to axial loads is modelled by a second-order ordinary differential equation. Similarly, the physics of the beam subjected to bending is modelled by a fourth-order differential equation. These concepts are discussed in greater detail in subsequent chapters. FEM involves converting the differential equations into algebraic equations. The procedure to do this will be explained in later chapters. In this chapter, we directly obtained the algebraic equations by making suitable assumptions. This type of procedure is generally known as direct FEM. In some cases, the governing differential equation may be of zero order, such as in Solved Problem 1.1, the problem of current flow in a circuit consisting of only resistive elements. Obtaining the algebraic equations for such problems is straightforward. For two and three-dimensional problems, the methodology is similar. In that case, we shall obtain the partial differential equations instead of ordinary differential equations.

In this chapter, we have taken up very simple problems to illustrate FEM. Therefore, many details have been omitted. FEM is a very effective technique for problems with complex geometry. Also, as the equations are developed element by element and then assembled, incorporation of non-homogeneous material properties becomes quite easy.

We are trying to understand FEM as a tool to solve differential equations. FEM can be applied to a large number of engineering problems. Although it originated as a technique of solving elastic structure problems, of late it has been applied to plastic deformation problems also. It is also widely used for solving heat transfer, fluid mechanics, and electromagnetism problems. In manufacturing area, FEM is being used to model metal forming, metal cutting, and non-traditional machining processes. The problems of dynamics and vibrations are also successfully solved using FEM.

SOLVED PROBLEMS

1.1 Use FEM to solve the electric circuit shown in Fig. 1.9.

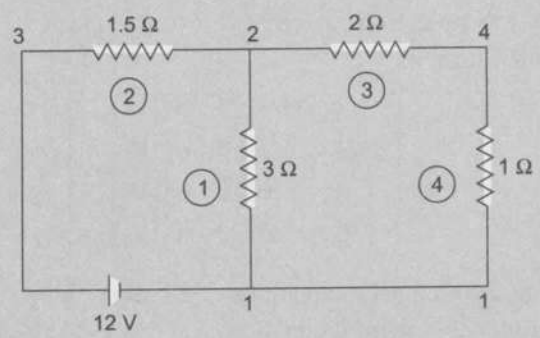


Figure 1.9 *Electric circuit for Problem 1.1*

Solution The electric circuit is discretized into four elements as shown in Fig. 1.9. Each register is considered as one element and voltage is the primary variable. Notice that node 1 is shown at two positions, as there is no voltage drop between these two positions. Now we develop the elemental equations:

For element 1, using Ohm's law, we can write

$$\frac{V_1 - V_2}{3} = I_1^{(1)} \tag{1.1.1}$$

where $I^{(1)}{}_1$ is the current entering into element (register) 1 at node 1. By convention, the current entering into an element is taking as positive. Similarly, for element 1, we can also write

$$\frac{V_2 - V_1}{3} = I_2^{(1)} \tag{1.1.2}$$

where $I_2^{(1)}$ is the current entering into element 1 at node 2. Combining these equations, we obtain

$$\frac{1}{3}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} V_1 \\ V_2 \end{Bmatrix} = \begin{Bmatrix} I_1^{(1)} \\ I_2^{(1)} \end{Bmatrix} \tag{1.1.3}$$

Similarly for element 2

$$\frac{1}{1.5}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} V_2 \\ V_3 \end{Bmatrix} = \begin{Bmatrix} I_2^{(2)} \\ I_3^{(2)} \end{Bmatrix} \tag{1.1.4}$$

For element 3

$$\frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} V_2 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} I_2^{(3)} \\ I_4^{(3)} \end{Bmatrix} \tag{1.1.5}$$

For element 4

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} V_4 \\ V_1 \end{Bmatrix} = \begin{Bmatrix} I_4^{(4)} \\ I_1^{(4)} \end{Bmatrix} \tag{1.1.6}$$

Equations (1.1.1) to (1.1.4) need to be assembled. For this purpose, the first elemental equation can be expressed in the global form as

$$\begin{bmatrix} 1/3 & -1/3 & 0 & 0 \\ -1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} I_1^{(1)} \\ I_2^{(1)} \\ 0 \\ 0 \end{Bmatrix} \tag{1.1.7}$$

Similarly, by expressing the elemental equations of all other elements in global form and adding them, we get the following global system of equations:

$$\begin{bmatrix} \frac{1}{3}+1 & \frac{-1}{3} & 0 & -1 \\ \frac{-1}{3} & \frac{1}{3}+\frac{1}{1.5}+\frac{1}{2} & \frac{-1}{1.5} & \frac{-1}{2} \\ 0 & \frac{-1}{1.5} & \frac{1}{1.5} & 0 \\ -1 & \frac{-1}{2} & 0 & \frac{3}{2} \end{bmatrix} \begin{Bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} I_1^{(1)} + I_1^{(4)} \\ I_2^{(1)} + I_2^{(2)} + I_2^{(3)} \\ I_3^{(2)} \\ I_4^{(3)} + I_4^{(4)} \end{Bmatrix} \tag{1.1.8}$$

In accordance with Kirchhoff's current law

$$I_2^{(1)} + I_2^{(2)} + I_2^{(3)} = 0\,, \quad I_4^{(3)} + I_4^{(4)} = 0 \tag{1.1.9}$$

Therefore, Eq. (1.1.8) becomes

$$\begin{bmatrix} \frac{4}{3} & \frac{-1}{3} & 0 & -1 \\ \frac{-1}{3} & 1.5 & \frac{-1}{1.5} & \frac{-1}{2} \\ 0 & \frac{-1}{1.5} & \frac{1}{1.5} & 0 \\ -1 & \frac{-1}{2} & 0 & \frac{3}{2} \end{bmatrix} \begin{Bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} I_1^{(1)} + I_1^{(4)} \\ 0 \\ I_3^{(2)} \\ 0 \end{Bmatrix} \tag{1.1.10}$$

Applying the boundary conditions, $V_1 = 0$ and $V_3 = 12$, we get

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1.5 & \frac{-1}{1.5} & \frac{1}{2} \\ 0 & 0 & 1 & 0 \\ 0 & \frac{-1}{2} & 0 & \frac{3}{2} \end{bmatrix} \begin{Bmatrix} V_1 \\ V_2 \\ V_3 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 12 \\ 0 \end{Bmatrix} \tag{1.1.11}$$

From Eq. (1.1.11), the following solution is obtained:

$$V_1 = 0,\ V_2 = 6\ \text{V},\ V_3 = 12\ \text{V},\ V_4 = 2\text{V}$$

Having obtained the voltages at each node, we can obtain current in any register. For example, the current passing through register 2 is

$$\frac{V_3 - V_2}{1.5} = \frac{12-6}{1.5} = 4\,\text{A}$$

1.2 The temperatures at the two ends of a 1-m long steel rod having a uniform cross-sectional area of 1 cm^2 are 100°C and 50°C, respectively, as shown in Fig. 1.10. The rate of internal heat generated in the steel rod is 100 W/m^3. Taking the thermal conductivity of steel as 50 W/m-°C, find out the temperature distribution in the rod using FEM with (a) two elements, (b) three elements, and (c) four elements. Assume that the heat transfer coefficient is very low and hence the heat transfer due to convection is negligible.

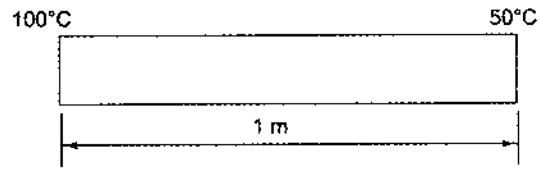


Figure 1.10 *A rod with end temperatures of* 100°C *and* 50°C

Solution The problem is of one-dimensional heat conduction. First, we develop the elemental equations for a typical line element of length h (Fig. 1.11). We follow the convention in which the heat entering into the element is positive and heat leaving the element is negative.

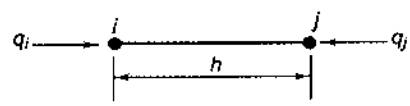


Figure 1.11 *A typical element*

We assume that the temperature gradient inside an element is constant. The rate of heat generation inside the element is $\dot{Q}AH$, where $\dot{Q}$ denotes the rate of heat generation per unit volume. This heat generation is modelled as if $\dot{Q}Ah/2$ amount of heat is entering from both ends per unit time. Let q_i and q_j be the additional (not due to heat generation) heat fluxes (rates of heat transfer per unit at cross-sectional area) at the nodes i and j, respectively. Fourier's law of heat conduction provides for a typical element with nodal temperatures T_i and T_j:

$$q_i A + \dot{Q}Ah = kA\frac{T_i - T_j}{h} \tag{1.2.1}$$

$$q_j A + \dot{Q}Ah = kA\frac{T_j - T_i}{h} \tag{1.2.2}$$

where k is the thermal conductivity and A is the cross-sectional area of the element. Thus, the elemental equations are

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_i \\ T_j \end{Bmatrix} = \begin{Bmatrix} q_{iA} + \dot{Q}Ah/2 \\ q_{jA} + \dot{Q}Ah/2 \end{Bmatrix} \tag{1.2.3}$$

The problem may be solved by taking two-, three-, and four-element cases as discussed below:

(a) Two-element case

The mesh is shown in Fig. 1.12.

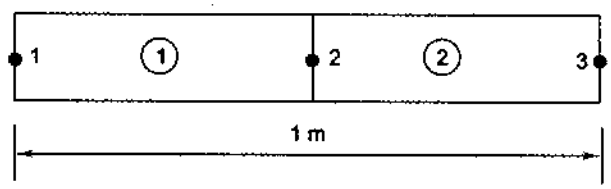


Figure 1.12 *A mesh with two elements*

The elemental equations for the first element may be written in the following form:

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} Aq_1^{(1)} + 50Ah \\ Aq_2^{(1)} + 50Ah \end{Bmatrix} \tag{1.2.4}$$

where the superscript (1) on q indicates the element number.

The elemental equations for the second element are represented as:

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} Aq_2^{(2)} + 50Ah \\ Aq_3^{(2)} + 50Ah \end{Bmatrix} \tag{1.2.5}$$

Assembling Eqs (1.2.4) and (1.2.5), we obtain:

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)} + 50Ah \\ q_2^{(1)}A + q_2^{(2)} + 100Ah \\ q_2^{(2)}A + q_2^{(2)} + 50Ah \end{Bmatrix} \tag{1.2.6}$$

At node 2: $$q_2^{(1)}A + q_2^{(2)}A = 0$$

Therefore, the Eq. (1.2.6) may be written as

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 0 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)} + 50Ah \\ 100Ah \\ q_3^{(2)}A + 50Ah \end{Bmatrix}$$

By applying the boundary conditions, $T_1 = 100°C$ and $T_3 = 50°C$, the equations reduce to

$$\begin{bmatrix} 1 & 0 & 0 \\ -\frac{kA}{h} & \frac{2kA}{h} & -\frac{kA}{h} \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 100Ah \\ 50 \end{Bmatrix} \tag{1.2.8}$$

Substituting $k = 50$ W/m-°C, $A = 1\text{cm}^2$, $h = 0.5$ m in Eq. (1.2.8) we obtain

$$T_1 = 100°C, \quad T_2 = 75.25°C, \quad \text{and} \quad T_3 = 50°C \tag{1.2.9}$$

The exact temperature distribution (obtained by solving the governing differential equation) for this problem is

$$T = 100 - 50x + (x - x^2) \tag{1.2.10}$$

where T is in °C and x is in m. Equation (1.2.10) provides the temperature at node 2, i.e., at $x = 0.5$ m, as 75.25°C, the same as that obtained by FEM. Although the nodal temperatures obtained by FEM are exact, the temperature distribution obtained by FEM will be different from the exact temperature distribution given by Eq. (1.2.10). In FEM, the temperature between the nodes, in case of two-noded elements, will be obtained by linear interpolation of the nodal temperatures. Thus, the temperature distribution obtained by FEM is

$$T = \begin{cases} 100 - 49.5x & \text{for } 0 \le x \le 0.5 \\ 100.5 - 50.5x & \text{for } 0.5 \le x \le 1 \end{cases} \tag{1.2.11}$$

which is obviously different from the exact temperature distribution.

(b) Three-element case

The mesh is shown in Figure 1.13.

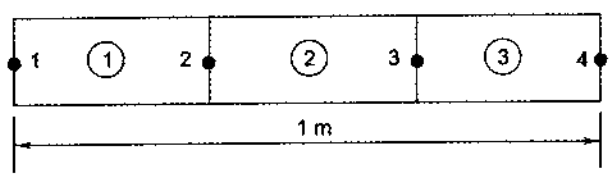


Figure 1.13 *A mesh with three elements*

The elemental equations for the first element are written as:

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)} + 50Ah \\ q_2^{(1)} + 50Ah \end{Bmatrix} \tag{1.2.12}$$

The elemental equations for the second element are written as

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} q_2^{(2)}A + 50Ah \\ q_3^{(2)}A + 50Ah \end{Bmatrix} \tag{1.2.13}$$

The elemental equations for the third element are written as

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} q_3^{(3)}A + 50Ah \\ q_4^{(3)}A + 50Ah \end{Bmatrix} \tag{1.2.14}$$

By assembling the above elemental equations, we obtain

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)}A + 50Ah \\ q_2^{(1)}A + q_2^{(2)}A + 100Ah \\ q_3^{(2)}A + q_3^{(3)}A + 100Ah \\ q_4^{(3)}A + 50Ah \end{Bmatrix} \tag{1.2.15}$$

For nodes 2 and 3

$$q_2^{(1)}A + q_2^{(2)}A = 0 \text{ and } q_3^{(2)}A + q_3^{(3)}A = 0 \tag{1.2.16}$$

Thus

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)}A + 50Ah \\ 100Ah \\ 100Ah \\ q_4^{(3)}A + 50Ah \end{Bmatrix} \tag{1.2.17}$$

By applying the boundary conditions, T_1 = 100°C and T_4 = 50°C, Eq. (1.2.17) becomes

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} & 0 \\ 0 & \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 100Ah \\ 100Ah \\ 50 \end{Bmatrix} \tag{1.2.18}$$

Substituting k = 50 W/m-°C, A = 1cm^2, and h = 1/3 m in Eq. (1.2.18), we obtain:

$$T_1 = 100°\text{C}, \quad T_2 = 83\frac{5}{9}°\text{C}, \quad T_3 = 66\frac{8}{9}°\text{C}, \quad T_4 = 50° \tag{1.2.19}$$

It can be easily verified that the nodal temperatures are exact. The temperatures between the two nodes can be obtained by linear interpolation between the nodes.

(c) Four-element case

The mesh is shown in Fig. 1.14.

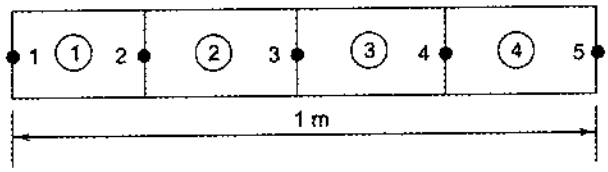


Figure 1.14 *A mesh with four elements*

The elemental equations for the first element are given by

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)}A + 50Ah \\ q_2^{(1)}A + 50Ah \end{Bmatrix} \tag{1.2.20}$$

The elemental equations for the second element are written as

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} q_2^{(2)}A + 50Ah \\ q_3^{(2)}A + 50Ah \end{Bmatrix} \tag{1.2.21}$$

The elemental equations for the third element are given as

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_4 \\ T_5 \end{Bmatrix} \begin{Bmatrix} q_4^{(4)}A + 50Ah \\ q_5^{(4)}A + 50Ah \end{Bmatrix} \tag{1.2.22}$$

The elemental equations for the fourth element are given by

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_4 \\ T_5 \end{Bmatrix} \begin{Bmatrix} q_4^{(4)}A + 50Ah \\ q_5^{(4)}A + 50Ah \end{Bmatrix} \tag{1.2.23}$$

By assembling the elemental matrices, we obtain

$$\frac{kA}{h}\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} q_1^{(1)}A + 50Ah \\ q_2^{(1)}A + q_2^{(2)}A + 100Ah \\ q_3^{(2)}A + q_3^{(3)}A + 100Ah \\ q_4^{(3)}A + q_4^{(4)}A + 100Ah \\ q_5^{(4)}A + 50Ah \end{Bmatrix} \tag{1.2.24}$$

For nodes 2, 3, and 4

$$q_2^{(1)}A + q_2^{(2)}A = 0\,,\ q_3^{(2)}A + q_3^{(3)}A = A = 0, \text{ and } q_4^{(3)}A + q_4^{(4)}A = 0 \tag{1.2.25}$$

Thus

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} & 0 & 0 \\ 0 & \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} & 0 \\ 0 & 0 & \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 100Ah \\ 100Ah \\ 100Ah \\ 50 \end{Bmatrix} \tag{1.2.26}$$

By applying the boundary conditions, Eq. (1.2.26) becomes

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} & 0 & 0 \\ 0 & \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} & 0 \\ 0 & 0 & \frac{-kA}{h} & \frac{2kA}{h} & \frac{-kA}{h} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ T_5 \end{Bmatrix} = \begin{Bmatrix} 100 \\ 100Ah \\ 100Ah \\ 100Ah \\ 50 \end{Bmatrix} T_1$$

Substituting $k = 50$ W/m-°C, $A = 1\text{cm}^2$, and $h = 1/4$ m in Eq. (1.2.27), we obtain

$$T_1 = 100°\text{C},\quad T_2 = 87\frac{11}{16}°\text{C},\quad T_3 = 75.25°\text{C},\quad T_4 = 62\frac{11}{16}°\text{C},\quad T_5 = 50°\text{C}$$

It can be verified that these temperatures are exact. The temperature at any point can be obtained by linear interpolation of the temperatures of two adjacent nodes.

It is observed that irrespective of the number of elements used, FEM provides exact nodal temperature in this case. However, the temperature distribution will be always different from the exact temperature distribution. In case of two-noded elements, FEM assumes piecewise linear temperature distribution, while the exact temperature distribution is quadratic. As the number of elements is increased, the approximate temperature distribution approaches the exact temperature distribution. Had we taken a three-noded element, in which the temperature distribution inside an element is assumed to be quadratic, we would have obtained exact temperature distribution by FEM even with one element.

EXERCISES

A. Multiple Choice Questions

1.1 Consider the following three global stiffness matrices:

$$\text{(i)}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1.5 & -1 \\ -1 & 1.5 & 0.5 \end{bmatrix} \quad \text{(ii)}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1.5 & -0.5 \\ 0 & -0.5 & 0 \end{bmatrix} \quad \text{(iii)}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1.5 & -0.5 \\ 0 & -0.5 & 0.5 \end{bmatrix}$$

For a non-uniform rod subjected to axial loads, which of the following is true?

(a) Only matrix (i) is a valid global stiffness matrix

(b) All matrices are valid global stiffness matrices

(c) Only matrix (iii) is a valid global stiffness matrix

(d) Matrices (ii) and (iii) are valid global stiffness matrices

1.2 Consider the following system of equations obtained after the assembly step of FEM:

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ 0 \\ 10 \end{Bmatrix}$$

After applying the boundary condition $u_3 = 1$, the system of equation ready for solution is

(a) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ 0 \\ 10 \end{Bmatrix}$

(b) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 1 & 0 & 0 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix}$

(c) $\begin{bmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 1 \end{Bmatrix}$

(d) $\begin{bmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 10 \end{Bmatrix}$

1.3 Consider the following system of equations obtained after the assembly step of FEM:

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -F_1 \\ 0 \\ 10 \end{Bmatrix}$$

After applying the boundary condition, the system of equation ready for solution is

(a) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 10 \end{Bmatrix}$

(b) $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 1 & 0 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}$

(d) $\begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 10 \end{Bmatrix}$

1.4 In a one-dimensional finite element, the physical displacement variable v and its derivative with respect to x, v', are the nodal variables. Given that at the first node ($x = 0$), $v = v' = 0$, and at the second node ($x = 1$), $v = 1$m and $v' = 3$, the displacement variable at ($x = 0.5$) is

(a) 1/8 (b) 1/4 (c) 1/2 (d) 1

1.5 The stiffness matrix for a beam problem is given by

$$\begin{bmatrix} k_{11} & k_{12} & k_{13} & k_{14} \\ k_{21} & k_{22} & k_{23} & k_{24} \\ k_{31} & k_{32} & k_{33} & k_{34} \\ k_{41} & k_{42} & k_{43} & k_{44} \end{bmatrix}$$

For this matrix, which of the following relation is valid?

(a) $k_{11} + k_{12} + k_{13} + k_{14} = 0$

(b) $k_{11} + k_{13} = 0$ and $k_{21} + k_{23} = 0$

(c) $k_{11} + k_{13} = 0$ and $k_{21} + k_{23} \neq 0$

(d) $k_{11} + k_{13} = 0$ and $k_{12} + k_{14} = 0$

B. Problems

1.6 A short rod of length l is rigidly supported at both ends and an axial load P is applied at the mid-length (Fig. 1.15). Taking two equal finite elements, find out the displacement at the point of application of load. Also find out the support reactions. The area of cross-section of the rod is A and Young's modulus of elasticity is E.

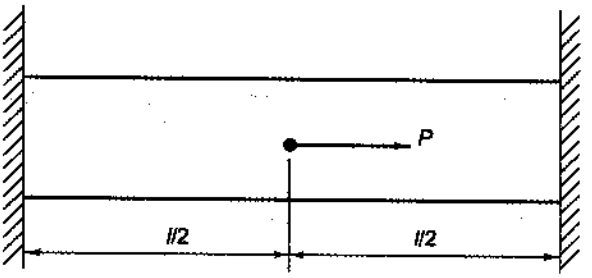


Figure 1.15 *A fixed–fixed rod of length l*

1.7 The short rod (area of cross-section A, Young's modulus of elasticity E) shown in Fig. 1.16 is fixed at one end, the other end being held by a spring of spring constant k. A load P is applied at the mid length. Using direct finite element formulation, find out the spring compression.

(Solve by two methods. In the first method, take two elements in the rod and put spring force as the natural boundary condition. In the second method, take two elements in the rod and treat spring as the third element; apply essential boundary conditions at the both ends.)

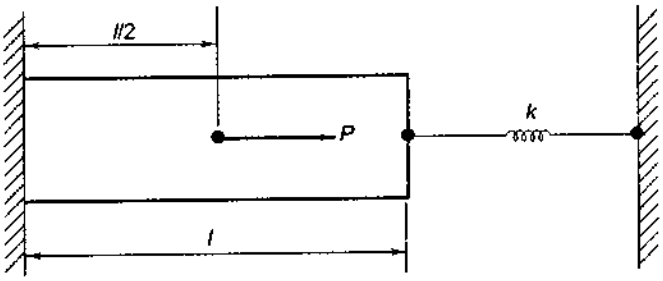


Figure 1.16 *Fixed rod supported by spring*

1.8 A cantilever beam of length l, second moment of inertia I, and Young's modulus of elasticity E is loaded by a load P (Fig. 1.17). Take only one finite element and find out the deflection and slope at the free end. Compare it with the solution obtained using the strength of the material approach. Using the fact that deflection of any point of this beam is a cubic polynomial in x (the distance of the point from the fixed end), find out the deflection at a distance of $l/2$ from the fixed end.

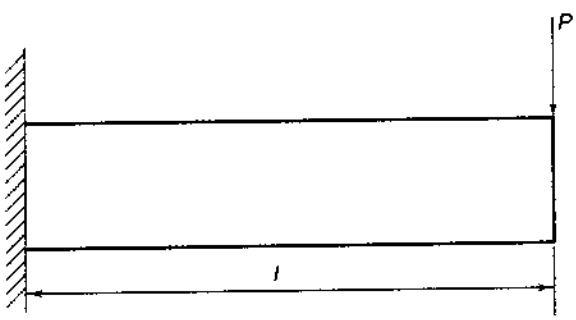


Figure 1.17 *A cantilever beam of length l*

1.9 Fourier's law of heat conduction in a rod gives

$$q = -kA\frac{\mathrm{d}T}{\mathrm{d}x}$$

where k is the thermal conductivity, A is the area of the rod, and T is the temperature. Using direct FEM approach, obtain the elemental stiffness and right hand side (load) vector for solving one-dimensional heat conduction problem. For the rod shown in Fig. 1.18, find out the temperature at node 2 by taking two elements. The rod is made of steel having the thermal conductivity 50 W/m-K.

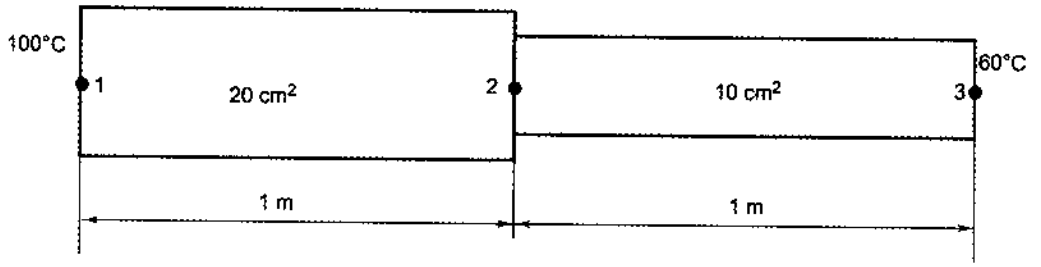


Figure 1.18 *Heat conduction in a rod*

1.10 One end of a steel rod is fixed and other end is pulled by an unknown force F as shown in Fig. 1.19. Due to the application of the load, the mid-length point of the rod moves by a distance of 0.2 mm. Using FEM with two equal elements, find out the value of the unknown force. The length of the rod is 1 m, area is 1 cm^2, and Young's modulus of elasticity is 200 GN/m^2. Note that in this problem only one boundary condition is known and the information regarding the mid-point displacement is given instead of the other boundary condition.

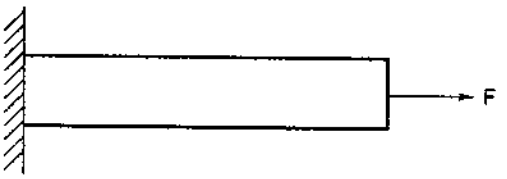


Figure 1.19 *A fixed rod being pulled by force F*

1.11 Using direct FEM approach, find out the currents in all resistors of the circuit shown in Fig. 1.20. Treat each resistor as one element and potentials at the two ends as primary variables. Elemental equations can be derived using Ohm's law and assembly can be carried out using Kirchhoff's current law.

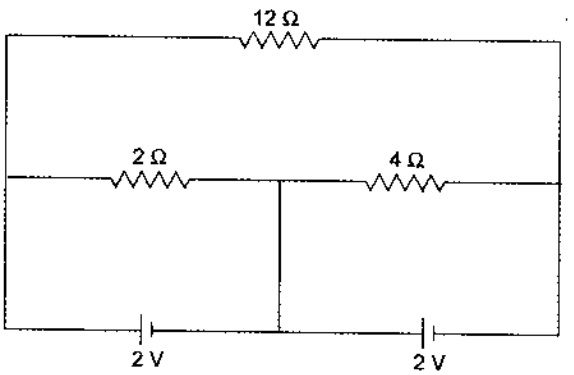


Figure 1.20 *An electric circuit*

1.12 Two beams of length l each are joined by a pin joint and then the combined beam is subjected to a load P (Fig. 1.21). Find the deflection at the load using FEM. Do you have to make any special comment about this problem?

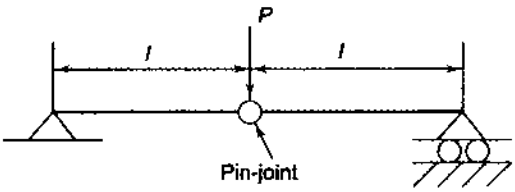


Figure 1.21 *Two beams joined by a pin joint*

1.13 A cantilever beam of length l is supported on a spring of spring constant k at its free end (Fig. 1.22). Using FEM, find out the deflection of the beam if a load P is applied at the mid-length of the beam.

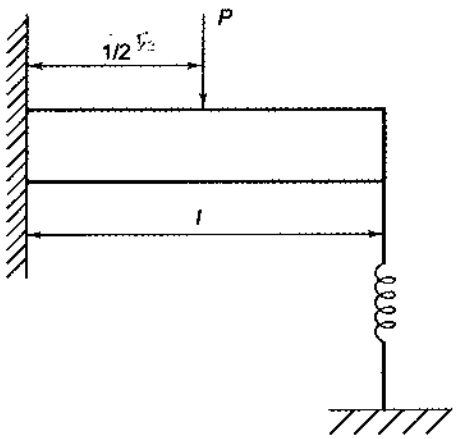


Figure 1.22 *A cantilever beam supported on a spring*

C. Computer-based Problems

1.14 The cross-sectional area of a rod varies as

$$A(x) = A_0\left(2 - \frac{x}{l}\right)$$

where A_0 is the area of cross-section at the free end, x is the longitudinal distance of a cross-section of area A from the fixed end, and l is the length of the rod. A load P is applied at the free end (Fig. 1.23). Find the displacement at the free end using FEM. Solve this problem 10 times by discretizing the rod into elements from 1 to 10. For each element, the average area of

the element should be taken. Plot the obtained displacement versus number of elements and comment on convergence. Take suitable numerical values for plotting the graph. Else, express the displacement in some non-dimensional form.

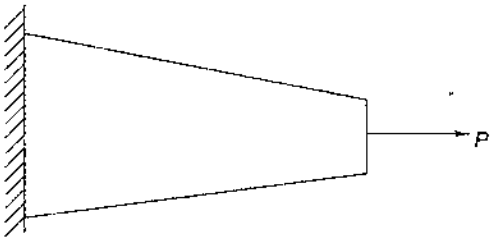


Figure 1.23 *A fixed rod with load P*

1.15 One-dimensional seepage through a porous medium is governed by Darcy's law which gives the flow in terms of the gradient of the total potential f. The law is similar to Fourier's law of heat conduction:

$$q = -kA\frac{\partial \phi}{\partial x}$$

where q is the flow, k is the permeability coefficient, and A is the area of cross-section of the porous medium. In the problem shown in Fig. 1.24, potentials on the two sides of the porous medium are h_1 and h_2. The thickness of the porous medium is t and permeability coefficients on the left and right sides are k_1 and k_r, respectively, the variation being linear across thickness. Solve this problem using FEM. Study the convergence by taking different numbers of elements.

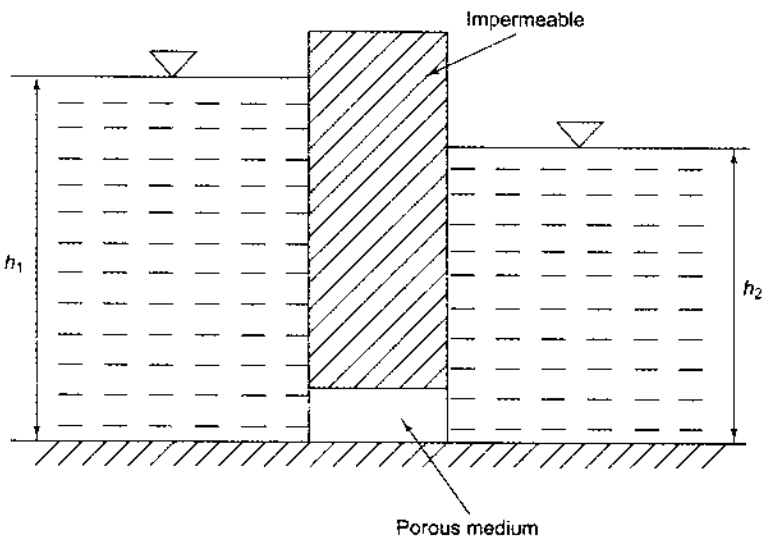


Figure 1.24 *One-dimensional seepage through a porous medium*

REFERENCES

Clough, RW, 1960, 'The finite element in plane stress analysis', *Proc.* 2nd *ASCE conference on Electronic Computation*, Pittsburgh, PA, USA.

Courant, R, 1943, 'Variational methods for the solution of problems of equilibrium and Vibrations', *Bulletin of the American Mathematical Society*, Vol. 49, pp 1–23.

Turner, MJ, Clough RW, Martin, HC, and Topp, LJ, 1956, 'Stiffness and deflection analysis of complex structures', *J Aero Sci*, Vol. 23, pp 805-823.

2

CHAPTER

Introduction to Calculus of Variations

2.1 Introduction

2.2 Functionals

2.3 Extremization of a Functional

2.4 Obtaining the Variational Form from a Differential Equation

2.5 Principle of Virtual Work

2.6 Principle of Minimum Potential Energy

2.7 Conclusion

2.1 INTRODUCTION

The concept of finite element method (FEM) was introduced in Chapter 1; this is an excellent method of solving differential equations. FEM is considered to be an extremely useful method as the behaviour of a physical system is often described by its governing differential equations. However, sometimes, it is convenient to derive an integral expression (called *variational form*), minimization or maximization of which leads to the same solution as obtained by solving the governing differential equations. For example, the governing differential equation for an Euler–Bernoulli beam (in which only the deflection caused by bending is considered) is given by

$$EI_{zz}\frac{d^4v}{dx^4} = q \tag{2.1}$$

where EI_{zz} is the flexural rigidity, v is the transverse deflection, a function of the length coordinate x, and q is the load intensity (load per unit length) function. Note that in Chapter 1, the notation I was used for the second moment of inertia of the cross-section of the beam. However, in this chapter, I will be used to indicate the variational form, and hence the second moment of inertia (about transverse z-axis) has been denoted by I_{zz}. To solve this differential equation, four boundary conditions are needed. Assume that we are solving the built-in beam problem for which the boundary conditions at both the ends are

$$v = 0$$

$$dv/dx = v' = 0 \tag{2.2}$$

It can be shown that the deflection of this beam of length L may also be obtained by minimizing the following integral:

$$\int_0^L \left[\frac{1}{2}EI_{zz}\left(\frac{d^2v}{dx^2}\right) - qv\right]dx \tag{2.3}$$

ensuring that the boundary conditions given by Eq. (2.2) are satisfied. The integral expression of Eq. (2.3) is called the *variational form*. This directly follows from the principle of minimum potential energy which will be discussed in Section 2.6.

Given a variational form, one can obtain the governing differential equations and solve them by using a suitable numerical method such as FEM. Differential equations can also be transformed into variational form. The branch of mathematics that deals with transforming a variational form to differential equation form and vice versa is called *calculus of variations*. In calculus, we are often interested in finding out a point at which a function attains minimum/maximum value; however, in calculus of variation, we find out the function that provides minimum/maximum value of the variational form. We will study the necessary techniques of calculus of variation in the following sections.

2.2 FUNCTIONALS

In calculus, we have to deal with different kinds of functions. A *function* consists of a dependent variable, whose value depends on one or many independent variables. For example, $y = f(x) = x^3$ is a function, in

which, for each value of the independent variable x, there exists a scalar value of the dependent variable y. Similarly, in the function $z = x^2 + y^2$, for each value of x and y, there exists a value of the dependent variable z.

Now consider the definite integral

$$I = \int_0^5 y\mathrm{d}x = \int_0^5 f(x)\mathrm{d}x \tag{2.4}$$

Here, for each particular function, there exists a scalar value of I. For example, when $f(x) = 1$, the value of I is 5. When $f(x) = x$, the value of I becomes 12.5. In literature, I is called a *functional* (function of a function), whose value depends on the independent function $f(x)$. The following integral expression is also an example of a functional:

$$I(y) = \int_a^b F(x, y, y')\mathrm{d}x \tag{2.5}$$

where F depends on x, $y \equiv y(x)$, and $y' \equiv \dfrac{\mathrm{d}y}{\mathrm{d}x}$. Mathematically, a functional is an operator I, which maps y into a scalar value.

A functional $l(y)$ is said to be linear in y, if and only if, it satisfies the following relation:

$$l(\alpha y + \beta z) = \alpha l(y) + \beta l(z) \tag{2.6}$$

for any scalars, α and β, and independent functions y and z. A functional $B(y, z)$ is said to be bilinear, if it is linear in each of its argument y and z, i.e.,

$$B(\alpha y_1 + \beta y_2, z) = \alpha B(y_1, z) + \beta B(y_2, z) \tag{2.7}$$

$$B(y, \alpha z_1 + \beta z_2) = \alpha B(y, z_1) + \beta B(y, z_2) \tag{2.8}$$

The functional is said to be symmetric if

$$B(y, z) = B(z, y) \tag{2.9}$$

An example of a linear functional is

$$I(u) = \int_0^L u\, f\, \mathrm{d}x \tag{2.10}$$

where f is a constant function and u is the independent variable function of x. An example of a symmetric bilinear functional is as follows:

$$I(u, v) = \int_0^L E\,A \frac{\mathrm{d}u}{\mathrm{d}x}\frac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x \tag{2.11}$$

where E and A are constant functions and u and v are independent variable functions of x.

In this chapter, the symbol I has often been used to denote a functional. It should not be confused with the second moment of inertia. The distinction will be clear from the context. At some places, the moment of inertia has been denoted by I_{zz} to avoid confusion.

Example 2.1 To clarify the concept of functional, consider the following expressions:

(a) $\int_0^3 \left(x + \sin x + 5x^2 e^x\right) dx$

(b) $\int_0^3 \left(x + \sin x + 5x^2 e^x + y\right) dx$

(c) $\int_0^3 \left(y^2 + y'^2\right) dx$

(d) $\int_0^x \left(x + \sin x + 5x^2 e^x + y\right) dx$

(e) $\int_0^x \left(y^2 + y'^2\right) dx$

For each of them, analyze whether it is function, functional, or neither a function nor a functional.

Solution Expression (a) can be integrated and will give a definite value. Thus, it represents just a constant value. (It can of course be considered to be a constant function.) Expression (b) can be evaluated only if y as a function of x is known. Thus, the expression is dependent on the function y and hence, is a functional. Expression (c) is also dependent on the function y and hence, is a functional. Expressions (d) and (e) will not provide the real value even if the expression of y in terms of x is known, hence, they represent neither functions nor functionals.

2.3 EXTREMIZATION OF A FUNCTIONAL

Consider a functional I, such that

$$I = \int_a^b F\left(x, y, \frac{dy}{dx}\right) dx$$

The relation between y and x is not known and the problem consists of finding this relation so that I becomes either a maximum or a minimum. Assume that $y_0(x)$ is some known relation between y and x, which extremizes I. Let another function in the neighbourhood of $y_0(x)$ be denoted by $y_0(x) + \varepsilon\eta(x)$, where $\eta(x)$ is an arbitrary (but sufficiently differentiable) function of x and ε is an arbitrary small quantity. The function $\eta(x)$ does not violate the geometric boundary conditions of the problem. Hence, wherever y is prescribed, h is zero. The function $\varepsilon\eta(x)$ is called δy, the variation of y at a given x, δ being a variational operator. Variational operator δ is in many ways similar to differential operator d and has similar mathematical properties. However, they are conceptually different. Differential of a function dy is a first-order approximation to the change in function along a particular curve, while the variational of a function δy is a first-order approximation to the change from curve to curve.

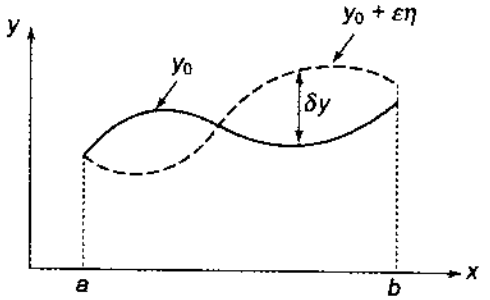


Figure 2.1 *Variation of a function*

Figure 2.1 shows the plot of function y_0 with a solid line. Assume that the value of function at one end point is prescribed, then a general function $y_0(x) + \varepsilon\eta(x)$ is shown by a dotted curve. If we put the general function in the functional, I will be obtained as a function of ε. The necessary condition for extremum of this function is

$$\frac{\mathrm{d}I}{\mathrm{d}\varepsilon} = 0 \tag{2.12}$$

Strictly speaking, the condition given in Eq. 2.12 is valid for maxima, minima, or saddle points. In most of the practical problems, a saddle point is not observed. If y_0 itself is extremum, then Eq. (2.12) should hold good at $\varepsilon = 0$, such that,

$$\left.\frac{\mathrm{d}I}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0 \tag{2.13}$$

Replacing y by $y_0(x) + \varepsilon\eta(x)$ in functional I, we obtain

$$I = \int_a^b F(x,\ y_0 + \varepsilon\eta,\ y_0' + \varepsilon\eta')\,\mathrm{d}x \tag{2.14}$$

where a dash in the superscript indicates differentiation with respect to x. At a fixed value of x, one can use Taylor's theorem such that

$$F(x, y, y') = F(x, y_0, y_0') + \left(\frac{\partial F}{\partial y}\varepsilon\eta + \frac{\partial F}{\partial y'}\varepsilon\eta'\right) + \frac{\partial^2 F}{\partial y^2}\frac{(\varepsilon\eta)^2}{2!} + \frac{\partial^2 F}{\partial y \partial y'}(\varepsilon\eta\varepsilon\eta') + \frac{\partial^2 F}{\partial y'^2}\frac{(\varepsilon\eta')^2}{2!} + \cdots \tag{2.15}$$

where all the derivatives are evaluated at y_0 and y_0'. Note that while expanding F by Taylor series, we treat x as fixed and y and y' as two independent variables. Once x is fixed, y and y' become variables instead of functions and Eq. (2.15) becomes possible. Integrating Eq. (2.15) between a and b and taking the derivative with respect to ε, we obtain

$$\frac{\mathrm{d}I}{\mathrm{d}\varepsilon} = \int_a^b \left\{0 + \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right) + \left(\varepsilon\frac{\partial^2 F}{\partial y^2}\eta^2 + 2\varepsilon\frac{\partial^2 F}{\partial y' \partial y}\eta\eta' + \varepsilon\frac{\partial^2 F}{\partial y'^2}\right) + \cdots\right\}\mathrm{d}x \tag{2.16}$$

Applying the condition for extremum, we get

$$\left.\frac{\mathrm{d}I}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0 = \int_a^b \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta'\right)\mathrm{d}x \tag{2.17}$$

The $\left.\frac{\mathrm{d}I}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0}\varepsilon$ is also called the first variation of I, δI. Thus, the condition for extremizing a functional is $\delta I = 0$. Similarly, $\left.\frac{\mathrm{d}^2 I}{\mathrm{d}\varepsilon^2}\right|_{\varepsilon=0}\varepsilon^2$ is called the second variation of I, $\delta^2 I$, which can tell if the function is minimized, maximized, or neither minimized nor maximized. When the right hand side of the above equation is integrated by parts, so as to reduce the order of derivative of η, Eq. (2.17) gets transformed to

$$\int_a^b \left\{ \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \right\} \eta \mathrm{d}x + \frac{\partial F}{\partial y'} \eta \Big|_a^b \tag{2.18}$$

Thus

$$\delta I = \int_a^b \left\{ \frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \right\} \eta \mathrm{d}x + \frac{\partial F}{\partial y'} \eta \Big|_b - \frac{\partial F}{\partial y'} \eta \Big|_a = 0 \tag{2.19}$$

If the value of y is prescribed at a boundary, η at that boundary will be zero as there is no variation at that point. At other places η is arbitrary. However, we can put in Eq. (2.19), an arbitrary η, which is 0 at both the boundaries. In that case, we obtain

$$\int_{x_1}^{x_2} \left(\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \right) \eta \mathrm{d}x = 0 \tag{2.20}$$

In view of the η being arbitrary, Eq. (2.20) implies that

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0 \tag{2.21}$$

Thus, extremization of the functional I requires the satisfaction of Eq. (2.21). Substituting Eq. (2.21), in Eq. (2.19), we have

$$\frac{\partial F}{\partial y'} \eta \Big|_b - \frac{\partial F}{\partial y'} \eta \Big|_a = 0 \tag{2.22}$$

At a particular boundary, say at point a, η is either zero or arbitrary, and hence, can be made zero. Thus

$$\frac{\partial F}{\partial y'} \eta \Big|_b = 0 \tag{2.23}$$

In the same way, it can be shown that

$$\frac{\partial F}{\partial y'} \eta \Big|_a = 0 \tag{2.24}$$

Equations (2.23) and (2.24) are the two boundary conditions, which must be satisfied along with the differential equation given by Eq. (2.21) for the extremization of the functional. Boundary conditions imply that at the boundaries either η should be 0, i.e., the value of y should be prescribed or $\frac{\partial F}{\partial y'}$ should be zero. The first type of boundary condition is called *geometric or essential boundary condition*, while the second type of boundary condition is called *natural boundary condition*. Equation (2.21) is called Euler–Lagrange equation.

Example 2.2 Find out the Euler–Lagrange equation, which extremizes the following functional:

$$I = \int_0^1 (y'^2 + y^2 + 2xy)\ \mathrm{d}x \tag{2.25}$$

Solution Using Eq. (2.21), we have

$$2y + 2x - \frac{d}{dx}(2y') = 0$$

that is

$$y + x - y'' = 0 \tag{2.26}$$

as

$$\frac{\partial F}{\partial y'} = 2y' \tag{2.27}$$

The boundary conditions are as follows:

$$\text{either } y' = 0 \text{ or } y \text{ is prescribed at } x = 0 \text{ and } 1 \tag{2.28}$$

Note that the variational form of the differential equation given in Eq. (2.26) does not depend on the prescribed value of y at the boundaries.

Example 2.3 Find out the governing differential equation that minimizes the following functional:

$$\int_{x_1}^{x_2} \frac{\sqrt{1 + (dy/dx)^2}}{x}\, dx \tag{2.29}$$

Solution Here

$$F = \frac{\sqrt{1 + y'^2}}{x} \tag{2.30}$$

Using Euler–Lagrange equation given by Eq. (2.21):

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0 \tag{2.31}$$

As F does not contain y, the first term in Eq. (2.31) is zero. Thus

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = \frac{d}{dx}\left(\frac{y'}{x\sqrt{1 + y'^2}}\right) = 0 \tag{2.32}$$

Note that while partially differentiating F with respect to y', x has been treated as a constant. Equation (2.32) implies

$$\frac{y'}{x\sqrt{1 + y'^2}} = C \quad \text{(a constant)} \tag{2.33}$$

This is the required differential equation for minimization of Eq. (2.29). Although we have not proved the sufficiency condition for minimization, it can be observed that Eq. (2.29) is unbounded for maximization, because by choosing a suitable function, dy/dx can be made as large as required, making the functional as large as desired. Thus, any function y providing the finite value of the functional of Eq. (2.29) is not a maximizing solution. Thus, the solution obtained by solving Eq. (2.33) provides either a minimum or a saddle point solution. To confirm that it is a minimum solution we need to consider the second variation, which is beyond the scope of this textbook.

Example 2.4 Find out the equation of the curve whose end points are (x_1, y_1) and (x_2, y_2) and which generates a surface with minimum area after it completes one revolution about the x-axis.

Solution An arbitrary curve passing through points (x_1, y_1) and (x_2, y_2) is shown in Fig. 2.2. When this curve is revolved around x-axis, a surface of revolution is obtained. We have to find out the equation of the curve that generates the minimum surface area after revolution around x-axis

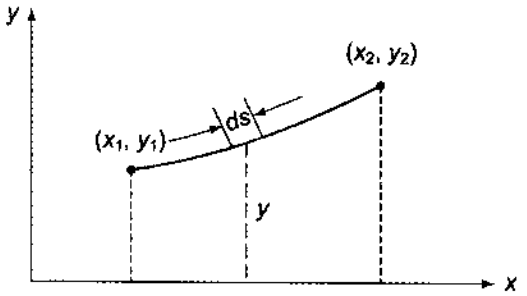


Figure 2.2 *A curve passing through two points*

A small length ds of the curve produces an area 2π yds after one complete revolution about x-axis, such that

$$dA = 2\pi y ds = 2\pi y\left(1+\left(\frac{dy}{dx}\right)^2\right)^{1/2} dx \tag{2.34a}$$

Thus, the total area of surface of revolution is

$$S = \int_{x_1}^{x_2} 2\pi y\left(1+\left(\frac{dy}{dx}\right)^2\right)^{1/2} dx \tag{2.34b}$$

As 2π is a constant, minimum surface area of surface of revolution will be obtained by minimizing the following functional:

$$\int_{x_1}^{x_2} y\sqrt{1+y'^2}\,dx \tag{2.35}$$

Here

$$F = y\sqrt{1+y'^2} \tag{2.36}$$

Using Euler–Lagrange equation given in Eq. (2.21), we obtain

$$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = \sqrt{1+y'^2} - \frac{d}{dx}\left(\frac{yy'}{\sqrt{1+y'^2}}\right) = 0 \tag{2.37}$$

Thus, the equation of the required curve may be obtained by solving the following differential equation:

$$\sqrt{1+y'^2} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{yy'}{\sqrt{1+y'^2}}\right) \tag{2.38}$$

A somewhat simple form of the differential equation may be obtained by using the following alternate form of the Euler–Lagrange equation:

$$\frac{\partial F}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}x}\left(F - y'\frac{\partial F}{\partial y'}\right) = 0 \tag{2.39}$$

It can be easily demonstrated that this form is the same as Eq. (2.21). For this, we rewrite Eq. (2.39) as

$$\frac{\partial F}{\partial x} - \frac{\mathrm{d}F}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial F}{\partial y'}\right) = 0 \tag{2.40}$$

As F is the function of x, y, and y', using the chain rule of differentiation, the total derivative of F with respect to x can be written as follows:

$$\frac{\mathrm{d}F}{\mathrm{d}x} = \frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{\partial F}{\partial y'}\frac{\mathrm{d}y'}{\mathrm{d}x} \tag{2.41}$$

Substituting Eq. (2.41) in Eq. (2.40), we get

$$-\frac{\partial F}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} - \frac{\partial F}{\partial y'}\frac{\mathrm{d}y'}{\mathrm{d}x} + \frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial F}{\partial y'}\right) = 0 \tag{2.42}$$

Noting that

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y'\frac{\partial F}{\partial y'}\right) = \frac{\partial F}{\partial y'}\frac{\mathrm{d}y'}{\mathrm{d}x} + y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) \tag{2.43}$$

Equation (2.42) can be written as

$$-y'\frac{\partial F}{\partial y} + y'\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0 \tag{2.44}$$

Assuming that y' is not equal to 0, we can divide Eq. (2.44) by $-y'$ to obtain

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) = 0 \tag{2.45}$$

which is the same as the Euler–Lagrange equation given in Eq. (2.21).

The form of Eq. (2.39) is very useful when F is not an explicit function of x, like in the present problem. Using Eq. (2.39), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(y\sqrt{1+y'^2}-\frac{yy'^2}{\left(1+y'^2\right)^{1/2}}\right)=0 \tag{2.46}$$

or

$$\left(y\sqrt{1+y'^2}-\frac{yy'^2}{\left(1+y'^2\right)^{1/2}}\right)=\frac{y}{\left(1+y'^2\right)^{1/2}}=c \tag{2.47}$$

where c is a constant. Thus, the differential equation providing the equation of the required curve is given by

$$\frac{y^2}{\left(1+y'^2\right)}=c^2 \tag{2.48}$$

Equation (2.48) gives

$$y'=\frac{\mathrm{d}y}{\mathrm{d}x}=\frac{\sqrt{y^2-c^2}}{c} \tag{2.49}$$

Thus

$$\int\frac{c\mathrm{d}y}{\sqrt{y^2-c^2}}=\int\mathrm{d}x \tag{2.50}$$

or

$$x=c\cosh^{-1}\left(y/c\right)+d \tag{2.51}$$

where d is a constant. The curve given by Eq. (2.51) provides the minimum surface of revolution. The constants c and d may be obtained by solving the following non-linear equations:

$$x_1=c\cosh^{-1}\left(y_1/c\right)+d \tag{2.52}$$

$$x_2=c\cosh^{-1}\left(y_2/c\right)+d \tag{2.53}$$

as the curve passes through two end points. We have not verified the sufficiency condition for the functional to be minimum. However, intuitively, one may join the two end points by a curve having very large values of y-coordinates and obtain a surface of revolution whose area may be made as large as desired. Thus, the problem is unbounded for the maximization case. Therefore, the curve given by Eq. (2.51) either minimizes the functional or provides saddle points. By considering the second variation, it can be confirmed that this curve actually minimizes the functional.

If F depends on several dependent variables, that is $F=F(y_1,y_1',y_2',y_2',\cdots,y_n,y_n',x)$ where each $y_i=y_i(x)$, the analysis proceeds as before, leading to n separate but simultaneous equations for the $y_i(x)$:

$$\frac{\partial F}{\partial y}-\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y_i'}\right)=0,\ i=1,\ldots,n \tag{2.54}$$

with corresponding boundary conditions. With n independent variables, we need to extremize multiple integrals of the form:

$$I = \int \ldots \int F\left(y, \frac{\partial y}{\partial x_1}, \ldots, \frac{\partial y}{\partial x_n}, x_1, \ldots, x_n\right) \mathrm{d}x_1 \ldots x_n \tag{2.55}$$

Using the same kind of analysis as before, we find that the extremizing function $y = y(x_1, x_2, \ldots, x_n)$ must satisfy

$$\frac{\partial F}{\partial y} - \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(\frac{\partial F}{\partial y_{x_i}}\right) = 0 \tag{2.56}$$

where y_{x_i} stands for $\dfrac{\partial y}{\partial x_i}$.

We will now derive the necessary differential equation for extremizing $I = \int_a^b F(x, y, y', y'')\mathrm{d}x$. We can apply the procedure adopted for extremizing $I = \int_a^b F\left(x, y, \dfrac{\mathrm{d}y}{\mathrm{d}x}\right)\mathrm{d}x$; however, we will now follow a short but not so rigorous approach. We will make use of the variational operator δ, and the fact that this operator behaves in the same way as a differential operator. In particular, the following identities are useful:

$$\delta(\alpha u \pm \beta v) = \alpha \delta u \pm \beta \delta v \tag{2.57}$$

$$\delta(uv) = u\delta v + v\delta u \tag{2.58}$$

$$\delta\left(\frac{u}{v}\right) = \frac{v\delta u - u\delta v}{v^2} \tag{2.59}$$

where u and v are the variable functions, and α and β are the known constant functions. In addition, we shall also make use of the following two properties:

Property 1 Differentiation and variation commute, i.e., $(\delta y)' = \delta y'$.

Proof We can write

$$(\delta y)' = \frac{\mathrm{d}}{\mathrm{d}x}(\delta y) = \frac{\mathrm{d}}{\mathrm{d}x}(\varepsilon\eta) = \varepsilon\frac{\mathrm{d}\eta}{\mathrm{d}x} \tag{2.60}$$

and

$$\delta(y') = \delta\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right) = \frac{\mathrm{d}}{\mathrm{d}x}(y + \varepsilon\eta) - \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y}{\mathrm{d}x} + \varepsilon\frac{\mathrm{d}\eta}{\mathrm{d}x} - \frac{\mathrm{d}y}{\mathrm{d}x} = \varepsilon\frac{\mathrm{d}\eta}{\mathrm{d}x} \tag{2.61}$$

The expressions in Eq. (2.60) and Eq. (2.61) are equal.

Hence proved.

Property 2 Integration and variation commute, that is, $\delta\int_a^b y(x)\,\mathrm{d}x = \int_a^b (\delta y)\,\mathrm{d}x$

Proof

$$\delta\int_a^b y(x)\mathrm{d}x = \int_a^b (y+\varepsilon\eta)\mathrm{d}x - \int_a^b (y)\mathrm{d}x$$

$$= \int_a^b y\mathrm{d}x + \int_a^b (\varepsilon\eta)\mathrm{d}x - \int_a^b y\mathrm{d}x$$

$$= \int_a^b (\varepsilon\eta)\mathrm{d}x = \int_a^b \delta y\mathrm{d}x$$

Hence proved.

Using Property 1, we can write

$$\delta I = \int_a^b \delta F(x, y, y', y'')\mathrm{d}x \tag{2.62}$$

The first variation of F is given by

$$\delta F(x, y, y', y'') = F(x, y+\varepsilon\eta, y'+\varepsilon\eta', y''+\varepsilon\eta'') - F(x, y, y', y'') \tag{2.63}$$

where ε is a very small quantity. The expansion using Taylor series leads to

$$\delta F = \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y'' + O\left(\varepsilon^2\right) \tag{2.64}$$

Note that $\delta y \equiv \varepsilon\eta$ and so on. As $\varepsilon \to 0$, we can write

$$\delta F = \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y'' \tag{2.65}$$

Thus

$$\delta I = \int_a^b \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y''\right)\mathrm{d}x \tag{2.66}$$

Making use of Property 1, we can write

$$\delta I = \int_a^b \left(\frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\frac{\mathrm{d}}{\mathrm{d}x}(\delta y) + \frac{\partial F}{\partial y''}\frac{\mathrm{d}^2}{\mathrm{d}x^2}(\delta y)\right)\mathrm{d}x \tag{2.67}$$

Integrating the second and third terms by parts, we obtain

$$\delta I = \int_a^b \left(\frac{\partial F}{\partial y}\delta y - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\delta y - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y''}\right)\frac{\mathrm{d}}{\mathrm{d}x}(\delta y)\right)\mathrm{d}x + \frac{\partial F}{\partial y'}\delta y\Big|_a^b + \frac{\partial F}{\partial y''}\frac{\mathrm{d}}{\mathrm{d}x}(\delta y)\Big|_a^b \tag{2.68}$$

Integrating the third term by parts, we get

$$\delta I = \int_a^b \left\{\frac{\partial F}{\partial y}\delta y - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right)\delta y + \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\partial F}{\partial y''}\right)\delta y\right\}\mathrm{d}x + \frac{\partial F}{\partial y'}\delta y\Big|_a^b + \frac{\partial F}{\partial y''}\delta y'\Big|_a^b - \frac{\mathrm{d}}{\mathrm{d}x}\frac{\partial F}{\partial y''}\delta y\Big|_a^b \tag{2.69}$$

At this point, note that if you had adopted the procedure of putting $y = y_0(x) + \varepsilon\eta(x)$ in the functional, then the necessary condition for extremum would have been

$$\left.\frac{\mathrm{d}I}{\mathrm{d}\varepsilon}\right|_{\varepsilon=0} = 0 = \int_a^b \left(\frac{\partial F}{\partial y}\eta + \frac{\partial F}{\partial y'}\eta' + \frac{\partial F}{\partial y''}\eta'' \right) \mathrm{d}x \tag{2.70}$$

Multiplying the integral in Eq. (2.70) by ε, we can see that this is the same as Eq. (2.66). Thus, it is seen that the necessary condition for extremization is $\delta I = 0$. We can argue it in a different way also. Assume that $\delta I \neq 0$. In that case it is possible to increase and decrease the functional by applying a small variation. The extremum is reached when it is not possible to increase or decrease the functional by application of a small variation. Note also that if a function can be increased by applying a small variation, it can be decreased in the same manner by applying a variation in opposite direction. Thus, the first variation must vanish for extremization.

Applying the necessary condition for extremization, we get

$$\delta I = \int_a^b \left[\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) + \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\partial F}{\partial y''}\right) \right] \delta y \,\mathrm{d}x + \left\{ \frac{\partial F}{\partial y'} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y''}\right) \right\} (\delta y)\Big|_a^b + \frac{\partial F}{\partial y''}\delta y'\Big|_a^b = 0 \tag{2.71}$$

Thus, the differential equation is

$$\frac{\partial F}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y'}\right) + \frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(\frac{\partial F}{\partial y''}\right) = 0 \tag{2.72}$$

The boundary conditions are

$$\left\{ \frac{\partial F}{\partial y'} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y''}\right) \right\} \delta y = 0 \text{ at } x = a \text{ and at } x = b \tag{2.73}$$

and

$$\frac{\partial F}{\partial y''}\delta y' = 0 \text{ at } x = a \text{ and at } x = b \tag{2.74}$$

Equation (2.73) suggests that at a boundary point, we can have either

$$\delta y = 0$$

or

$$\frac{\partial F}{\partial y'} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial F}{\partial y''}\right) = 0$$

The former is called the *essential (or geometric) boundary condition*, while the latter is called the *natural boundary condition*. Similarly in Eq. (2.74), the boundary condition $\delta y' = 0$ is called essential boundary condition and $\partial F/\partial y'' = 0$ is referred to as the natural boundary condition. By convention, the boundary conditions associated with the variational operator are always called essential boundary

condition and other boundary conditions are called natural boundary conditions. Note that $\delta y' = 0$ does not imply that the slope is zero. It means that the slope has been prescribed.

There can be various explanations as to why the boundary conditions are named as essential and natural. One possible explanation is as follows. The essential boundary conditions are the set of boundary conditions that are sufficient for solving the problem completely. The natural boundary conditions are the boundary conditions involving higher order derivative terms and are not sufficient for solving the differential equation completely. This means that the representation of essential boundary condition(s) is essential in the set of boundary conditions of the problem and hence the name essential boundary condition. In most of the physical problems, the natural boundary conditions represent the interaction with the outside world (nature). For example, applied forces and moments in structural problems, and heat flux going to or coming from the surroundings in heat transfer problems form a part of natural boundary conditions and show the influence of outside world on the domain of the problem. Also, the natural boundary conditions affect the nature of variational form, thus justifying the name.

Example 2.5 Total potential energy of a beam of length l loaded by a load of intensity $q(x)$ is given by

$$I = \int_0^l \left(\frac{1}{2} EI_{zz} w''^2 - qw \right) \mathrm{d}x \tag{2.75}$$

where E and I_{zz} are the Young's modulus of elasticity and second moment of area about the axis perpendicular to the plane of bending, respectively, and are known functions of x, and w is the unknown beam deflection function. Find out the governing differential equation and boundary conditions.

Solution Minimization of the total potential energy will lead to finding out the beam deflection under load intensity q. Application of Eq. (2.72) gives

$$-q - 0 + \frac{\mathrm{d}^2}{\mathrm{d}x^2}(EI_{zz} w'') = 0$$

or

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}(EI_{zz} w'') = q \tag{2.76}$$

which is the well known Euler–Bernoulli beam equation. The boundary conditions are

$$\left[0 - \frac{\mathrm{d}}{\mathrm{d}x}(EI_{zz} w'') \right] \delta w = 0 \text{ at } x = 0, l \tag{2.77}$$

and

$$EI_{zz} w'' \delta w' = 0 \text{ at } x = 0, l \tag{2.78}$$

The term $\frac{\mathrm{d}}{\mathrm{d}x}(EI_{zz} w'')$ represents the shear force, while $EI_{zz} w''$ represents the bending moment. We will not talk about the sign convention at this stage. Thus, at both ends of the beam, two sets of the boundary conditions need to be satisfied:

Set 1 either the shear force is zero or the deflection is prescribed.

Set 2 either the bending moment is zero or the slope is prescribed.

The boundary conditions concerning the slope and deflection are called 'geometric' or 'essential' boundary conditions, while those involving shear and bending moment are called 'natural' or 'force' boundary conditions. Boundary conditions for three differently supported beams are shown in Fig. 2.3.

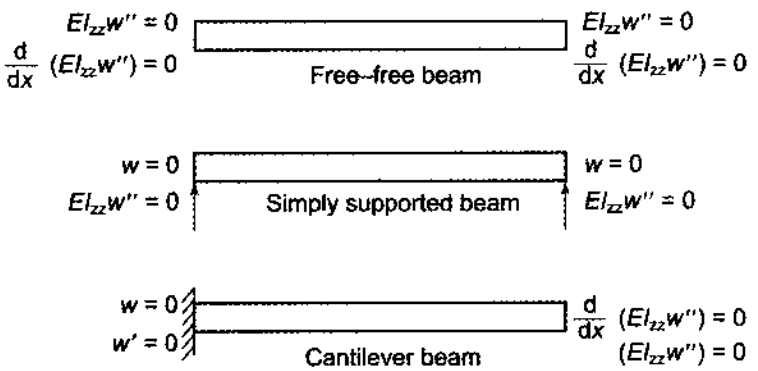


Figure 2.3 *Boundary conditions for three different types of beams*

The problem can also be directly solved by using the variational operator. Equating the first variation of the total potential energy to zero, we obtain

$$\delta I = \delta\int_0^l \left(\frac{1}{2}EI_{zz}w''^2 - qw\right)\mathrm{d}x = 0 \tag{2.79}$$

Using Property 2 of the variational operator, we can write

$$\delta I = \int_0^l \delta\left(\frac{1}{2}EI_{zz}w''^2 - qw\right)\mathrm{d}x = 0 \tag{2.80}$$

Making use of Eqs. (2.57) and (2.58), Eq. (2.80) can be written as

$$\int_0^l \left(EI_{zz}w''\delta w'' - q\delta w\right)\mathrm{d}x = 0 \tag{2.81}$$

Integrating the first part of the integrand by parts once to reduce the order of derivative of $\delta w''$, we get

$$EI_{zz}w''\delta w'\Big|_0^l - \int_0^l \left(\frac{\mathrm{d}}{\mathrm{d}x}\left(EI_{zz}w''\right)\delta w' - q\delta w\right)\mathrm{d}x = 0 \tag{2.82}$$

Integrating by parts the first part of integrand once more to reduce the order of derivative of $\delta w'$, we obtain

$$EI_{zz}w''\delta w'\Big|_0^l - \frac{\mathrm{d}}{\mathrm{d}x}\left(EI_{zz}w''\right)\delta w\Big|_0^l + \int_0^l\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(EI_{zz}w''\right)\delta w - q\delta w\right)\mathrm{d}x = 0 \tag{2.83}$$

or

$$EI_{zz}w''\delta w'\Big|_0^l - \frac{\mathrm{d}}{\mathrm{d}x}\left(EI_{zz}w''\right)\delta w\Big|_0^l + \int_0^l\left(\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(EI_{zz}w''\right) - q\right)\delta w\,\mathrm{d}x = 0 \tag{2.84}$$

As δw and $\delta w'$ are arbitrary except at the boundary points where their values are prescribed, all the terms in Eq. (2.84) are separately zero. This provides

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(EI_{zz}w''\right) - q = 0 \tag{2.85}$$

with the boundary conditions

$$EI_{zz}w''\delta w' = 0 \quad \text{at } x = 0 \text{ and } l \tag{2.86}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}x}(EI_{zz}w'')\delta w = 0 \text{ at } x = 0 \text{ and } l \tag{2.87}$$

2.4 OBTAINING THE VARIATIONAL FORM FROM A DIFFERENTIAL EQUATION

In this section, we will discuss in detail the procedure for converting a differential equation into a variational form. Consider the differential equation

$$L(\phi) - f = 0 \tag{2.88}$$

where L is linear or non-linear differential operator, ϕ is a scalar function defined over the domain D, and f is a known scalar function. Note that many books use symbol L only for linear differential operators. Here, it has been used to represent both. For example, consider the linear ordinary differential equation

$$\frac{\mathrm{d}^4\phi}{\mathrm{d}x^4} + 5\frac{\mathrm{d}^2\phi}{\mathrm{d}x^2} - 10\phi - 5x = 0 \tag{2.89}$$

where ϕ is a function of x. Here

$$L \equiv \frac{\mathrm{d}^4}{\mathrm{d}x^4} + 5\frac{\mathrm{d}^2}{\mathrm{d}x^2} - 10 \tag{2.90}$$

is a linear operator and $f \equiv 5x$.

Consider the non-linear ordinary differential equation

$$\frac{d^2\phi}{dx^2} + \phi^2 = 0 \tag{2.91}$$

Here

$$L \equiv \frac{d^2}{dx^2} + \phi \tag{2.92}$$

is a non-linear operator and $f \equiv 0$.

Consider the differential equation

$$\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + 5 = 0 \tag{2.93}$$

where ϕ is a function of x and y. In this case,

$$L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2.94}$$

is a linear operator and $f \equiv -5$.

To obtain the variational form of Eq. (2.88), we can use the variational operator. Multiplying the equation by a variation of ϕ and integrating it over the domain, D, we obtain

$$\int_D [L(\phi) - f]\delta\phi \, dD = 0 \tag{2.95}$$

We keep on manipulating Eq. (2.95) using integration by parts till we are able to put it in the variational form given by

$$\delta\int_D [L^*(\phi) - f\phi] dD = 0 \tag{2.96}$$

where L^* is suitable operator obtained at the end of the procedure. Finally, we obtain the following variational form of the given differential equation:

$$I(\phi) = \int_D [L^*(\phi) - f\phi] dD \tag{2.97}$$

In the process, the order of derivatives gets reduced. Generally, in a particular term, we attempt to keep the order of derivative on $\delta\phi$ and on the associated expression same. The procedure will become clearer from Examples 2.6 to 2.10.

Example 2.6 The governing differential equation for a rod loaded with axial force is

$$\frac{d}{dx}\left(EA\frac{du}{dx}\right) + q = 0 \tag{2.98}$$

where E is Young's modulus of elasticity, A is the area of cross-section, q is the load intensity (load per unit length in axial direction), and u is the axial displacement (function of axial coordinate x). Obtain the variational form of this equation. Assume that the boundary conditions are

$$\text{at } x=0, u=0$$

$$\text{at } x=l,\ EA\frac{\mathrm{d}u}{\mathrm{d}x}=P \tag{2.99}$$

where l is the length of the rod.

Solution As a first step, multiply the governing differential equation given in Eq. (2.98) by δu and integrate between 0 to l. Thus

$$\int_0^l\left[\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q\right]\delta u\,\mathrm{d}x=0 \tag{2.100}$$

Integrating Eq. (2.100) by parts, we obtain

$$\delta u\,EA\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_0^l+\int_0^l\delta u\,q\mathrm{d}x-\int_0^l\frac{\mathrm{d}}{\mathrm{d}x}(\delta u)\,EA\frac{\mathrm{d}u}{\mathrm{d}x}\mathrm{d}x=0 \tag{2.101}$$

As the variational operator behaves like a differential operator and $\delta u=0$ at $x=0$, we can write

$$\int_0^l\delta\left[\frac{1}{2}EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2\right]\mathrm{d}x-\int_0^l\delta(qu)\mathrm{d}x-\delta\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}u\right)_l=0 \tag{2.102}$$

Therefore

$$\delta\left[\int_0^l\left[\frac{1}{2}EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2-qu\right]\mathrm{d}x-Pu(l)\right]=0 \tag{2.103}$$

Hence, the variational form is given by

$$I=\int_0^l\left[\frac{1}{2}EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2-qu\right]\mathrm{d}x-Pu(l) \tag{2.104}$$

The reader may observe that this is the total potential energy of the rod. Thus, the displacement function of the rod will be one which extremizes the total potential energy amongst the functions that satisfy the essential boundary condition, that is, $u=0$ at $x=0$.

Example 2.7 The heat conduction in a rod without heat generation is governed by the equation

$$k\frac{\mathrm{d}^2T}{\mathrm{d}x^2}=0 \tag{2.105}$$

where k is the thermal conductivity and T is the temperature. Assume that the temperatures at the two ends of the rod are prescribed. Obtain the variational form of the problem.

Solution Multiplying the differential equation given in Eq. (2.105) by δT and integrating between 0 and l (length of the rod), we obtain

$$\int_0^l k\frac{\mathrm{d}^2T}{\mathrm{d}x^2}\delta T\,\mathrm{d}x=0 \tag{2.106}$$

Integrating by parts, we get

$$k\frac{\mathrm{d}T}{\mathrm{d}x}\delta T\bigg|_0^l - \int_0^l \frac{\mathrm{d}}{\mathrm{d}x}(\delta T)k\frac{\mathrm{d}T}{\mathrm{d}x}\mathrm{d}x = 0 \tag{2.107}$$

Making use of the fact that variation of the temperature at the ends is 0 and using the properties of the variational operator, we can write

$$\int_0^l \delta\left(\frac{1}{2}k\frac{\mathrm{d}T}{\mathrm{d}x}\frac{\mathrm{d}T}{\mathrm{d}x}\right)\mathrm{d}x = 0 \tag{2.108}$$

or

$$\delta\int_0^l \left(\frac{1}{2}k\frac{\mathrm{d}T}{\mathrm{d}x}\frac{\mathrm{d}T}{\mathrm{d}x}\right)\mathrm{d}x = 0 \tag{2.109}$$

Hence, the variational form is given by

$$I = \int_0^l \frac{1}{2}k\left(\frac{\mathrm{d}T}{\mathrm{d}x}\right)^2 \mathrm{d}x \tag{2.110}$$

Thus, for finding out the temperature distribution, extremize Eq. (2.110) while satisfying the essential boundary condition at the ends. Unlike in the case of rod subjected to axial load, we cannot assign any name to the functional I in Eq. (2.110).

Example 2.8 Steady-state heat conduction in an isotropic and homogeneous plate, in which the temperature in the direction of thickness remains constant, is governed by the following differential equation:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0 \tag{2.111}$$

Assume that the temperatures at the edges have been prescribed. Obtain the variational form of the problem.

Solution We first multiply the differential equation by δT and integrate it over the domain. Thus

$$\int_A \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right)\delta T\,\mathrm{d}A = 0 \tag{2.112}$$

Now, we can integrate this equation by parts to reduce the order of derivative; however, it is better to make use of the divergence theorem (Appendix A.1). Thus, Eq. (2.112) can be written as

$$\delta I = \int_A \nabla^2 T\,\delta T\,\mathrm{d}A = \int_A \nabla\cdot(\nabla T\,\delta T)\mathrm{d}A - \int_A \nabla T\cdot\nabla(\delta T)\mathrm{d}A = 0 \tag{2.113}$$

Applying the divergence theorem (Appendix A.1), we obtain

$$\delta I = \int_{\Gamma} (\nabla T.\hat{n})\delta T \mathrm{d}\Gamma - \int_{A} \nabla T.\nabla(\delta T)\mathrm{d}A = 0 \tag{2.114}$$

where Γ is the boundary of the domain. As the temperature is prescribed on the boundary, δT becomes 0 there. Thus, Eq. (2.114) becomes

$$\delta I = -\int_{A} \nabla T \cdot \nabla(\delta T)\mathrm{d}A = -\int_{A}\left(\frac{\partial T}{\partial x}\frac{\partial \delta T}{\partial x} + \frac{\partial T}{\partial y}\frac{\partial \delta T}{\partial y}\right)\mathrm{d}A = 0 \tag{2.115}$$

Using the properties of the variational operator, we can write

$$\delta I = -\delta\int_{A}\frac{1}{2}\left\{\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2\right\}\mathrm{d}A = 0 \tag{2.116}$$

Thus, the variational form is given by

$$I = -\int_{A}\frac{1}{2}\left\{\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial y}\right)^2\right\}\mathrm{d}A \tag{2.117}$$

Example 2.9 Obtain the variational form of the following differential equations:

$$\frac{\partial^2 u}{\partial x^2} + v = 0 \tag{2.118}$$

$$\frac{\partial^2 v}{\partial y^2} + u = 0 \tag{2.119}$$

where u and v are functions of x and y. Assume that u and v are prescribed at the boundaries of the domain.

Solution Considering that the main parameter in Eq. (2.118) is u and in Eq. (2.119) is v, we write

$$\delta I = \int_{A}\left(\frac{\partial^2 u}{\partial x^2} + v\right)\delta u\,\mathrm{d}x\mathrm{d}y + \int_{A}\left(\frac{\partial^2 v}{\partial y^2} + v\right)\delta v\,\mathrm{d}x\mathrm{d}y = 0 \tag{2.120}$$

Applying the divergence theorem (Appendix A.1) in component form, we obtain

$$\delta I = \oint_{\Gamma}\frac{\partial u}{\partial x}n_x\delta u + \oint_{\Gamma}\frac{\partial v}{\partial y}n_y\delta v - \int_{A}\left(\frac{\partial u}{\partial x}\frac{\partial \delta u}{\partial x} + \frac{\partial v}{\partial y}\frac{\partial \delta v}{\partial y} - v\delta u - u\delta v\right)\mathrm{d}x\mathrm{d}y = 0 \tag{2.121}$$

The first two terms vanish because u and v are prescribed at the boundary Γ. Using this fact and the properties of the variational operator, we get

$$\delta I = -\int_{A}\left(\frac{1}{2}\delta\left(\frac{\partial u}{\partial x}\right)^2 + \frac{1}{2}\delta\left(\frac{\partial v}{\partial y}\right)^2 - \delta(vu)\right)\mathrm{d}x\mathrm{d}y = 0 \tag{2.122}$$

Thus, the variational form is

$$I = -\int_A \left(\frac{1}{2}\left(\frac{\partial u}{\partial x}\right)^2 + \frac{1}{2}\left(\frac{\partial v}{\partial y}\right)^2 - vu \right) \mathrm{d}x\mathrm{d}y = 0 \tag{2.123}$$

Example 2.10 The static equilibrium equation for a stretched thin membrane is given by

$$T\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) + q(x, y) = 0 \tag{2.124}$$

where T is the uniform tension per unit length in the membrane, w is the vertical (transverse to x-y plane) deflection and $q(x, y)$ is the load per unit area. Obtain the variational form of Eq. (2.124).

Solution Multiplying the differential equation by δw and integrating it over the domain, we obtain

$$\delta I = \int_A \left[T\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) + q(x, y) \right] \delta w \mathrm{d}A = 0 \tag{2.125}$$

Using divergence theorem (Appendix A.1) the order of derivative can be reduced as

$$\delta I = \int_\Gamma (\nabla w \cdot \hat{n}) \delta T \mathrm{d}\Gamma - \int_A \nabla w \cdot \nabla(\delta w) \mathrm{d}A + \int_A q \delta w \mathrm{d}A = 0 \tag{2.126}$$

As the deflections at the edges are zero, the first term of Eq. (2.126) vanishes. In view of the properties of the variational operator δ, Eq. (2.126) reduces to

$$\delta I = -\delta \int_A \frac{1}{2}\left[\left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial y}\right)^2\right] \mathrm{d}A + \delta \int_A q w \mathrm{d}A = 0 \tag{2.127}$$

Thus, the variational form is

$$I = -\int_A \frac{1}{2}\left[\left(\frac{\partial w}{\partial x}\right)^2 + \left(\frac{\partial w}{\partial y}\right)^2 - qw\right] \mathrm{d}A = 0 \tag{2.128}$$

2.5 PRINCIPLE OF VIRTUAL WORK

We obtained the variational form of the rod subjected to axial loading by converting the differential equation into the variational form. We could have obtained it using the principle of virtual work. According to it, when we subject a loaded body in equilibrium to small, compatible virtual displacements (which do not violate the essential boundary conditions), the total internal virtual work is equal to the total external virtual work. The internal virtual work per unit volume will be equal to the product of real stresses and virtual strain, which can be integrated to yield the total internal virtual work. Note that virtual displacements are small, imaginary displacements, which when applied do not cause

any change in the forces or stresses. During the application of virtual displacements, the stresses can be considered to be constant. That is why we are able to take internal virtual work per unit volume equal to the product of real stresses and virtual strains. The virtual strains are found by taking the derivatives of the virtual displacement function. Therefore, the virtual displacement function must be differentiable. The external work will be the summation of work done by all forces acting on the body when subjected to virtual displacements.

If the rod of Example (2.6) is subjected to virtual displacement δu, the internal work will be given by

$$W_{\text{int}} = \int_0^l \frac{\mathrm{d}}{\mathrm{d}x}(\delta u)\, EA \frac{\mathrm{d}u}{\mathrm{d}x}\mathrm{d}x \tag{2.129}$$

and the external work will be given by

$$W_{\text{ext}} = \int_0^l \delta u\, q\mathrm{d}x + P\,\delta u(l) \tag{2.130}$$

Applying the principle of virtual work, we obtain

$$W_{\text{int}} - W_{\text{ext}} = \int_0^l \frac{\mathrm{d}}{\mathrm{d}x}(\delta u)\, EA \frac{\mathrm{d}u}{\mathrm{d}x}\mathrm{d}x - \int_0^l \delta u\, q\mathrm{d}x - P\,\delta u(l) = 0 \tag{2.131}$$

The Eq. (2.131) is called the integral form of the problem. If we treat δ as the variational operator, we can easily obtain Eq. (2.104), which represents the variational form of the problem. We shall be discussing about the principle of virtual work for some other problems later.

2.6 PRINCIPLE OF MINIMUM POTENTIAL ENERGY

In the process of obtaining the variational form of the rod problem, we have obtained a functional whose extremization along with the satisfaction of essential boundary condition will provide us the solution. We indicated that this functional is actually the total potential energy of the rod composed of (a) the strain energy of elastic distortion and (b) the potential possessed by applied loads. We could have written this functional using the principle of minimum potential energy for stable conservative mechanical system. For a conservative mechanical systems, one can express the energy content of the system in terms of its configuration, without reference to whatever deformation history or path may have led to the configuration.

The statement of principle of minimum potential energy is as follows:

> Among all possible configurations of a conservative system satisfying internal compatibility and essential boundary conditions, those that keep the body in stable equilibrium make the potential energy minimum with respect to small admissible variations of displacement. The admissible variations are those that preserve the constraints of the system.

This principle is applicable even if the material behaviour is non-linear.

As a simple example of principle of minimum potential energy, consider the case of a linear spring of stiffness k stretched by a force F. When the spring stretches by an amount x, the strain energy U stored in the spring is given by

$$U = \frac{1}{2}kx^2 \tag{2.132}$$

When the spring moves by an amount x in the direction of the force, its work potential becomes $-Fx$, assuming that in non-stretched configuration, the work potential is zero. This is because the spring has lost its capacity to do work by an amount Fx. We can confirm that negative of gradient of the work potential is F, i.e., if we differentiate the work potential with respect to x and multiply by -1, we get F. This is due to the fact that for a conservative system, the gradient of work potential is equal to the negative of the force. The total potential energy of the system is the sum of strain energy and work potential, that is

$$\Pi = \frac{1}{2}kx^2 - Fx \tag{2.133}$$

The spring will stretch by the amount that minimizes the total potential energy, the necessary condition for which is

$$\frac{\mathrm{d}\Pi}{\mathrm{d}x} = kx - F = 0 \tag{2.134}$$

This provides

$$x = F/k \tag{2.135}$$

which is the condition for static equilibrium. We have obtained it from the principle of minimum potential energy. (As the second derivative of Π with respect to x is negative, x obtained from Eq. (2.135) minimizes the total potential energy.)

As another example of finding total potential energy of a system, consider a rod subjected to axial loads (Fig. 2.4). It is known from strength of materials that strain energy density (strain energy per unit volume) of the linear elastic body with stress, σ, and strain, ε, is given as

$$\frac{\mathrm{d}U}{\mathrm{d}V} = \frac{1}{2}\sigma\varepsilon \tag{2.136}$$

For the rod subjected to axial displacement, u, the strain, ε, is given by

$$\varepsilon = \frac{\mathrm{d}u}{\mathrm{d}x} \tag{2.137}$$

and the stress, σ, by

$$\sigma = E\frac{\mathrm{d}u}{\mathrm{d}x} \tag{2.138}$$

where E is Young's modulus of elasticity. Thus, the strain energy density (dU/dV) is

$$\frac{\mathrm{d}U}{\mathrm{d}V} = \frac{1}{2}E\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 \tag{2.139}$$

Therefore, the strain energy of the rod (U) is

$$U = \int_V \frac{1}{2} E \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 \mathrm{d}V = \int_0^l \frac{1}{2} EA \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 \mathrm{d}x \tag{2.140}$$

where A is the area of cross-section and l is the length of the rod. Now, at location x, the rod may be considered to be subjected to a concentrated load $q\mathrm{d}x$, where q is the load intensity (load per unit length); the corresponding displacement is u and the work potential due to the load is $qu\mathrm{d}x$. Therefore, total work potential due to q is given by

$$V_q = -\int_0^l qu\mathrm{d}x \tag{2.141}$$

The work potential due to concentrated load P is

$$V_P = -Pu(l) \tag{2.142}$$

where u is (l) the displacement at $x = l$. Thus, the total potential energy of the rod is given by

$$\Pi = U + V_q + V_P = \int_0^l \frac{1}{2} EA \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 \mathrm{d}x - \int_0^l qu\mathrm{d}x - Pu(l) \tag{2.143}$$

The reader will come across the expressions for potential energy of other systems as we move along.

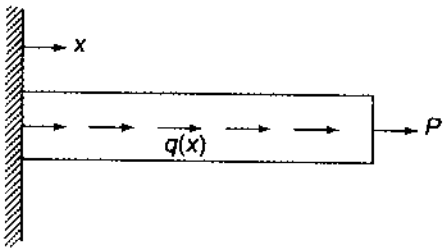


Figure 2.4 *A rod subjected to axial loads*

The reader may have a doubt regarding how one can know that the functional in Eq. (2.104) has to be minimized and not maximized as the functional was obtained by putting the necessary condition for extremization. For ascertaining that the functional has to be minimized, one needs to calculate the second variation $\delta^2 I$, which we have not done in this chapter. However, often the form of the functional and the physics of the problem can provide the answer. For example, we can see that Eq. (2.56) is unbounded for maximization problem. We can choose the function $u = -Mx$, where M is a large positive number, and hence make the functional as large as desired. Thus, there is no physically realistic solution for a maximization problem. Therefore, physically realistic solution should correspond to a minimization problem.

2.7 CONCLUSION

In this chapter, we have briefly introduced calculus of variations. We can model a physical problem in the form of differential equations or in the form of integral. The differential form is called *strong form* because it contains higher order derivatives, while the integral form containing lower order derivatives is called the *weak form*. In most of the physical problems, it is possible to convert one form into the other. The integral form can also be obtained using principle of virtual work or principle of minimum potential energy. These principles have been developed for mechanical systems. Thus, calculus of variations enables us to obtain the variational form for many other physical problems. For further details, the reader may refer to Reddy (1993), Cook (1989), Bathe (1982), Riley (1998), and Arfken (2005).

SOLVED PROBLEMS

2.1 The governing equations for free vibrations of a rod of length l can be obtained from Hamilton's principle. According to it, the motion of the particles of the rod from time t_1 to t_2 is such that the following line integral has a stationary value for the actual path of motion:

$$H = \int_{t_1}^{t_2}\int_0^l \left(\frac{1}{2}\rho A\left(\frac{\partial u}{\partial t}\right)^2 - \frac{1}{2}EA\left(\frac{\partial u}{\partial x}\right)^2\right)\mathrm{d}x\,\mathrm{d}t$$

where A is the area of cross-section of the rod, ρ is the density, E is Young's modulus of elasticity, and u is the displacement of a particle. The area of cross-section, the density, and Young's modulus of elasticity may be functions of x, while u is a function of both x and t. It is assumed that the position of particles is prescribed at time t_1 and t_2. Obtain the governing differential equation for the free vibration of the rod.

Solution For stationary value of H

$$\delta H = 0 \tag{2.1.1}$$

Thus

$$\int_{t_1}^{t_2}\int_0^l \left(\frac{1}{2}\rho A\delta\left(\frac{\partial u}{\partial t}\right)^2 - \frac{1}{2}EA\delta\left(\frac{\partial u}{\partial x}\right)^2\right)\mathrm{d}x\,\mathrm{d}t = 0 \tag{2.1.2}$$

Consider that the variation and integration commute and there is no variation on ρ, E, and A. Using Eq. (2.33), Eq. (2.1.2) becomes

$$\int_{t_1}^{t_2}\int_0^l \left(\rho A\left(\frac{\partial u}{\partial t}\right)\delta\left(\frac{\partial u}{\partial t}\right) - EA\left(\frac{\partial u}{\partial x}\right)\delta\left(\frac{\partial u}{\partial x}\right)\right)\mathrm{d}x\,\mathrm{d}t = 0 \tag{2.1.3}$$

The integration by parts to obtain the integrand in the form of a function multiplied by δu provides

$$\int_0^l \rho A\left(\frac{\partial u}{\partial t}\right)\delta u\Big|_{t_1}^{t_2}\mathrm{d}x - \int_t^{t_2}\int_0^l \rho A\left(\frac{\partial^2 u}{\partial t^2}\right)\delta u\mathrm{d}x\,\mathrm{d}t - \int_{t_1}^{t_2} EA\left(\frac{\partial u}{\partial x}\right)\delta u\Big|_0^l \mathrm{d}t + \int_{t_1}^{t_2}\int_0^l \frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right)\delta u\mathrm{d}x\,\mathrm{d}t = 0 \tag{2.1.4}$$

In view of the prescribed u, the first term in Eq. (2.1.4) becomes 0. Thus

$$-\int_t^{t_2}\int_0^l \rho A\left(\frac{\partial^2 u}{\partial t^2}\right)\delta u\mathrm{d}x\,\mathrm{d}t + \int_{t_1}^{t_2}\int_0^l \frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right)\delta u\mathrm{d}x\,\mathrm{d}t - \int_{t_1}^{t_2} EA\left(\frac{\partial u}{\partial x}\right)\delta u\Big|_0^l = 0 \tag{2.1.5}$$

or

$$\int_{t_1}^{t_2}\int_0^l\left[\frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right) - \rho A\left(\frac{\partial^2 u}{\partial t^2}\right)\right]\delta u\mathrm{d}x\,\mathrm{d}t - \int_{t_1}^{t_2} EA\left(\frac{\partial u}{\partial x}\right)\delta u\mathrm{d}t\Big|_0^l = 0 \tag{2.1.6}$$

As δu is arbitrary and Eq. (2.1.6) is valid for any time interval $t_2 - t_1$, we obtain

$$\frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right) - \rho A\left(\frac{\partial^2 u}{\partial t^2}\right) = 0 \tag{2.1.7}$$

and

$$EA\left(\frac{\partial u}{\partial x}\right)\delta u\Big|_0^l = 0 \tag{2.1.8}$$

Equation (2.1.7) is the governing differential equation and Eq. (2.1.8) provides the boundary conditions. Equation (2.1.8) implies that either the displacement is prescribed at one end or $EA\,\partial u/\partial x$ is zero.

2.2 Consider the following differential equation:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + 5 = 0 \qquad \text{for } 0 < x < 1$$

with boundary conditions:

$$u = 0 \quad \text{at } x = 0$$

$$\frac{\mathrm{d}u}{\mathrm{d}x} + u = 0 \quad \text{at } x = 1$$

Obtain the variational form of the differential equation.

Solution Multiplying the differential equation by δu and integrating between 0 and 1, we obtain

$$\delta I = \int_0^1\left(\frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + 5\right)\delta u\mathrm{d}x = 0 \tag{2.2.1}$$

Integrating the first term of the integrand by parts to evenly distribute the order of derivative on u and δu, we get

$$\delta I = \frac{\mathrm{d}u}{\mathrm{d}x}\delta u\Big|_0^1 - \int_0^1\left(\frac{\mathrm{d}u}{\mathrm{d}x}\frac{\mathrm{d}\delta u}{\mathrm{d}x} - 5\delta u\right)\mathrm{d}x = 0 \tag{2.2.2}$$

As the variational and differential operator commute, Eq. (2.2.2) can be written as

$$\delta I = \frac{\mathrm{d}u}{\mathrm{d}x}\delta u\Big|_0^1 - \int_0^1\left(\frac{\mathrm{d}u}{\mathrm{d}x}\delta\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right) - 5\delta u\right)\mathrm{d}x = 0 \tag{2.2.3}$$

Using Eqs. (2.32) and (2.33) and the fact that integration and variation commute, Eq. (2.2.3) becomes

$$\delta I = \frac{\mathrm{d}u}{\mathrm{d}x}\delta u\Big|_0^1 - \delta\int_0^1\left(\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 - 5u\right)\mathrm{d}x = 0 \tag{2.2.4}$$

The variation δu is 0 at $x = 0$ and δu (1) at $x = 1$, where $u(1)$ is the value of u at $x = 1$. At $x = 1$, du/dx is $-u(1)$. Therefore, Eq. (2.2.4) becomes

$$\delta I = -u(1)\delta u(1) - \delta\int_0^1\left(\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 - 5u\right)\mathrm{d}x = 0 \tag{2.2.5}$$

which in view of Eqs (2.32) and (2.33) becomes

$$\delta I = -\delta\left(\frac{u^2(1)}{2}\right) - \delta\int_0^1\left(\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 - 5u\right)\mathrm{d}x = 0 \tag{2.2.6}$$

Thus, the variational functional I is

$$I = -\left(\frac{u^2(1)}{2}\right) - \int_0^1\left(\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 - 5u\right)\mathrm{d}x \tag{2.2.7}$$

Multiplying Eq. (2.2.7) by -1, the variational function can be expressed as

$$I = \left(\frac{u^2(1)}{2}\right) + \int_0^1\left(\frac{1}{2}\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)^2 - 5u\right)\mathrm{d}x$$

EXERCISES

A. Multiple Choice Questions

2.1 Consider the functional

$$\Pi = \int_0^{10}\frac{1}{2}\left(\frac{\mathrm{d}u(x)}{\mathrm{d}x}\right)^2\mathrm{d}x - 5u(5)$$

where $u(5)$ is the value of function $u(x)$ at $x = 5$. The corresponding differential equation is

(a) $\dfrac{d^2u}{dx^2} + 5\delta(x) = 0$

(b) $\dfrac{d^2u}{dx^2} + 5u = 0$

(c) $\dfrac{d^2u}{dx^2} - 5\delta(x-5) = 0$

(d) $\dfrac{d^2u}{dx^2} + 5\delta(x-5) = 0$

where δ is Dirac-delta function.

2.2 Consider the following variational form:

$$\int_0^5 \left(\frac{1}{y} + y'^2 \right) dx$$

The corresponding differential equation is

(a) $2\dfrac{d^2y}{dx^2} = \dfrac{1}{y^2}$

(b) $2\dfrac{d^2y}{dx^2} + \dfrac{1}{y^2} = 0$

(c) $2\dfrac{d^2y}{dx^2} + \dfrac{1}{x^2} = 0$

(d) $\dfrac{d^2y}{dx^2} = 0$

2.3 For the cantilever beam shown in Fig. 2.5, w is the deflection function and w_L is the deflection at $x = L$.

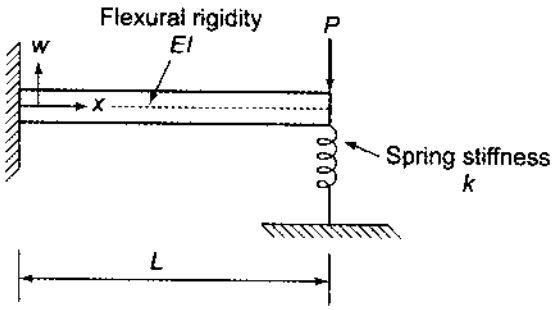


Figure 2.5 *A cantilever beam*

The total potential energy of the beam–spring system is

(a) $\int_0^L \frac{1}{2} EI \left(\frac{dw}{dx} \right)^2 dx + \frac{1}{2} k w_L^2 - P w_L$

(b) $\int_0^L \frac{1}{2} EI \left(\frac{dw}{dx} \right)^2 dx - \frac{1}{2} k w_L^2 - P w_L$

(c) $\int_0^L \left\{ EI \left(\frac{dw}{dx} \right)^2 - k w^2 \right\} dx - P w_L$

(d) $\int_0^L \left\{ EI \left(\frac{dw}{dx} \right)^2 + k w^2 \right\} dx - P w_L$

2.4 The spring constant of a spring is (4+1000x) N/m, where the spring deflection x is in m. The expression for the strain energy (in N-m) of the spring is

(a) $2x^2 + \frac{10^3 x^3}{3}$

(b) $2\times 10^{-6} x^2 + \frac{10^{-9} x^3}{3}$

(c) $\frac{10^{-3}}{2}(4+1000x)x$

(d) $\frac{10^{-6}}{2}(4+1000x)x^2$

2.5 Consider two linear springs of spring constant k whose one ends are fixed and the other ends are stretched by a displacement x. One spring is virtually displaced by δx and the other spring is provided an actual infinitesimal displacement dx. Which of the following is the correct statement?

(a) The work done by the virtual displacement is $kx\delta x$ and that by actual displacement is $\frac{1}{2}kx\,dx$.

(b) The force after the application of the virtual displacement is $kx\delta x$, while it is $k(x+\delta x)$ after the application of the actual displacement.

(c) The virtual displacement is always positive, whereas the actual displacement can be positive or negative.

(d) The work done by the applied force due to virtual displacement is zero.

B. Problems

2.6 The functional governing static buckling of the column in Fig. 2.6 is given as

$$\Pi = \frac{1}{2}\int_0^L EI\left(\frac{d^2w}{dx^2}\right)^2 dx - \frac{P}{2}\int_0^L\left(\frac{dw}{dx}\right)^2 dx + \frac{1}{2}kw_L^2 \qquad \text{(a)}$$

$$\text{where } w\big|_{x=0} = 0 \quad \text{and} \quad \frac{dw}{dx}\bigg|_{x=0} = 0 \qquad \text{(b)}$$

Invoke the stationary $\delta\Pi = 0$ conditions to derive the governing differential equation of the problem and the natural boundary conditions.

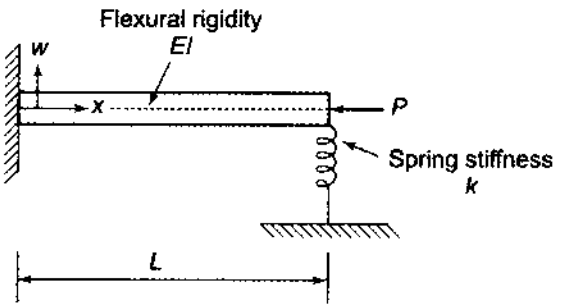


Figure 2.6 *Static buckling of a column*

2.7 Obtain the variational form of heat transfer problem in a rod (Fig. 2.7). Assume A, k, and r are constants.

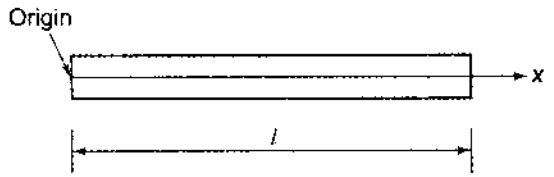


Figure 2.7 *A rod of length l*

Given, differential equation:

$$\frac{d^2T}{dx^2} + \frac{r}{kA} = 0$$

and boundary conditions:

(a) Essential boundary conditions: $T = 0$ at $x = 0$

(b) Natural boundary conditions: $\frac{dT}{dx} = \frac{-T}{2}$ at $x = l$

2.8 A certain physical problem has the functional

$$\Pi = \int_0^L \left(\frac{1}{2}\phi_{,x}^2 - 50\phi \right) dx$$

Essential boundary conditions are $\phi = 0$ at $x = 0$ and $x = L$. Express ϕ as a function of x, which extremizes the functional.

2.9 The potential energy of an isotropic plate that carries lateral pressure q is

$$\Pi_p = \frac{D}{2}\iint \left\{ (w_{,xx} + w_{,yy})^2 - 2(1-\nu)(w_{,xx}w_{,yy} - w_{,xy}^2) - 2\frac{qw}{D} \right\} dx\,dy$$

where D is a constant called flexural rigidity. Show that the governing differential equation is

$$\nabla^4 w = q/D$$

where ∇^4 is the bi-harmonic operator.

2.10 Steady-state heat conduction without heat generation is governed by the following differential equation (for domain with constant thermal conductivity):

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$$

If the temperatures at the boundary are specified, obtain the variational form of the above equation. How will the variational form change if the temperature at two sides is prescribed, while the other two sides are insulated?

2.11 Consider the following boundary value problem

$$\frac{d^2u}{dx^2} + u = 1 \qquad 0 \le x \le 1$$

with $u(0) = 0$ and $\frac{du}{dx} = 0$ at $x = 1$. Convert this problem into variational form.

2.12 The bending of a beam is governed by the following differential equation

$$\frac{d^2}{dx^2}\left(EI_{zz}\frac{d^2v}{dx^2}\right) + \frac{dm_z}{dx} - p_y = 0$$

with the following essential boundary conditions:

at $x = 0$, $v = v^*$ and $\frac{dv}{dx} = \theta_z^*$

and the following natural boundary conditions:

at $x = l$, $EI_{zz}\frac{d^2v}{dx^2} = M_z^*$ and $\frac{d}{dx}\left(EI_{zz}\frac{d^2v}{dx^2}\right) + m_z = -V_y^*$

where m_z is the distributed moment per unit length about z-axis and p_y is the distributed force per unit length in y-direction. Obtain the variational functional.

2.13 Using calculus of variations, show that the shortest curve joining two points is a straight line.

2.14 A frictionless wire in a vertical plane connects two points A and B, A being higher than B. Let the position of A be fixed at the origin of the xy-coordinate system, but allow B to lie anywhere on the vertical line $x = x_0$. Find the shape of the wire such that a bead of mass m placed on it at A will slide under gravity to B in the shortest possible time.

2.15 Consider a mechanical system whose configuration can be uniquely defined by a number of coordinates (usually distances and angles) together with time t, and which only experiences forces derivable from a potential. Hamilton's principle states that in moving from one configuration at time t_1 to another configuration at time t_2, the motion of the system is such as to make the following functional stationary:

$$H = \int_{t_0}^{t_1} L(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n, t)\,dt$$

The Lagrangian L is defined in terms of the kinetic energy T and the potential energy V (with respect to some reference situation) by $L = T - V$. Here V is a function of the q_i only, not of the $\dot{q}_i$. Applying the Euler–Lagrange equation to H, we obtain Lagrange's equations, such that

$$\frac{\partial L}{\partial q_i} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_i}\right) \qquad i = 1, \ldots, n$$

Using Hamilton's principle, derive the wave equation for small transverse oscillations of a taut string.

REFERENCES

Arfken, GB, and Weber, HJ, 2005, *Mathematical Methods for Physicists, Sixth edition*, Academic Press, Burlington.

Bathe, KJ, 1982, *Finite Element Procedures in Engineering Analysis*, Prentice-Hall, Englewood Cliffs, NJ, USA.

Cook, RD, Malkus, DS, and Plesha, ME, 1989, *Concepts and Applications of Finite Element Analysis, Third edition*, John Wiley, New York.

Reddy, JN, 1993, *An Introduction to the Finite Element Method*, McGraw-Hill, New York.

Riley, KF, Hobson, MP, and Bence, SJ, 1998, *Mathematical Methods for Physics and Engineering*, Cambridge University Press, Cambridge.

3

CHAPTER

Approximation Methods For Solving Differential Equations

3.1 Introduction

3.2 Ritz Method

3.3 Galerkin Method

3.4 Least Square Method

3.5 Collocation Method

3.6 Sub-domain Method

3.7 Conclusion

3.1 INTRODUCTION

In this chapter, we will study some classical function approximation methods that have been in use before the advent of finite element method (FEM) for solving differential equations. Methods discussed in this chapter approximate the function globally as a weighted summation of several linearly independent functions. The weights of various functions are found so that error can be minimized. No discretization of the domain is carried out. The procedure may be considered as FEM with just one element.

3.2 RITZ METHOD

The Ritz method was introduced by a German physicist and mathematician, W. Ritz (1908). It operates directly on the variational form of the differential equation. Therefore, it is also called *direct method* of solving variational problems. If the variational form is given, we can solve it using Ritz method without bothering about the Euler–Lagarange equation. If a differential equation is provided, we have to convert it to a variational form. The method is sometimes associated with Lord Rayleigh and hence it is also known as *Rayleigh–Ritz* method. However, based on the study of history, Leissa (2005) has observed that Rayleigh method is quite different from the Ritz method and therefore Rayleigh's name should not be associated with the method. The steps that need to be carried out in the method are as follows:

- Approximate the function as a weighted sum of linearly independent functions. The approximated function must satisfy the essential boundary conditions of the problem, while the natural boundary conditions need not be satisfied. The chosen functions must be complete such that starting from a constant term, the successively higher degree terms should be taken (without missing the terms in between) from a series of complete functions. In the context of polynomials, the series $1, x, x^2, x^3$ is complete but $1, x^2, x^3, x^4$ is incomplete. To give an example, if u is a function of x, it can be represented as

$$u = \sum_{i=1}^{n} a_i \phi_i \tag{3.1}$$

  where ϕ_i is the ith basis function and a_i is the corresponding weight or coefficient.
- In Eq. (3.1), the functions must be chosen in such a way that the essential boundary conditions are satisfied. We can obtain the linear equations corresponding to each essential boundary condition by putting the boundary conditions in Eq. (3.1). With the help of these equations, some coefficients can be eliminated from Eq. (3.1).
- Put the approximation u that satisfies the essential boundary conditions into variational form. The variational form will now be a function of the coefficients. Extremize the variational form. For this purpose, we take partial derivatives of the variational form with respect to the coefficients and make them 0. The number of equations thus obtained will be equal to the number of coefficients. These equations can be solved to yield the unknown coefficients in u. Having known the coefficients, approximate solution can be constructed.

Example 3.1 Consider the following boundary value problem:

$$\frac{d^2u}{dx^2} + u = 1 \quad 0 \le x \le 1 \quad \text{with } u(0) = 0 \quad \text{and} \quad \frac{du}{dx} = 0 \text{ at } x = 1 \tag{3.2}$$

Solve this equation using Ritz's method.

Solution First we have to convert it into variational form. For this, consider

$$\delta I = \int_0^1 \left(\frac{d^2u}{dx^2} + u - 1 \right) \delta u \, dx \tag{3.3}$$

After integrating the first term of Eq. (3.3) by parts, the expression can be written as

$$\delta I = \frac{du}{dx} \delta u \Big|_0^1 - \int_0^1 \left(\frac{du}{dx} \right) \frac{d}{dx} (\delta u) dx + \int_0^1 u \delta u dx - \int_0^1 \delta u dx \tag{3.4}$$

In view of the boundary conditions and the fact that variational and differential operators are commutative, Eq. (3.4) becomes

$$\delta I = -\int_0^1 \frac{du}{dx} \delta \left(\frac{du}{dx} \right) dx + \int_0^1 u \delta u dx - \int_0^1 \delta u dx \tag{3.5}$$

As the variational operator behaves like a differential operator, we can write

$$\delta I = -\int_0^1 \frac{1}{2} \delta \left(\frac{du}{dx} \right)^2 dx + \int_0^1 \frac{1}{2} \delta \left(u^2 \right) dx - \int_0^1 \delta u dx \tag{3.6}$$

or

$$\delta I = \delta \left[-\int_0^1 \frac{1}{2} \left(\frac{du}{dx} \right)^2 dx + \int_0^1 \frac{1}{2} \left(u^2 \right) dx - \int_0^1 u dx \right] \tag{3.7}$$

Therefore

$$I = \int_0^1 \left\{ -\frac{1}{2} \left(\frac{du}{dx} \right)^2 + \frac{1}{2} u^2 - u \right\} dx \tag{3.8}$$

Now let us solve this problem using Ritz's method. If we approximate u by $u = a + bx$, then considering the first essential boundary conditions, i.e., $u(0) = 0$, we obtain $a = 0$. Hence, $u = bx$ and $\frac{du}{dx} = b$. Putting the value of u and $\frac{du}{dx}$ in the variational form, we obtain

$$I = \int_0^1 \left\{ -\frac{1}{2} b^2 + \frac{1}{2} b^2 x^2 - bx \right\} dx \tag{3.9}$$

Extremizing Eq. (3.9) with respect to b, we obtain

$$\frac{dI}{db} = \int_0^1 \left\{ -b + bx^2 - x \right\} dx = \left[-b + \frac{b}{3} - \frac{1}{2} \right] = 0 \tag{3.10}$$

This gives $b = -\frac{3}{4}$. Thus, $u = -\frac{3}{4}$.

Therefore, the exact solution of this differential equation is

$$u = A \sin x + B \cos x + 1 \tag{3.11}$$

where constants A and B are to be determined from the boundary conditions. Since $u(0) = 0$, $B = -1$, Also

$$\frac{du}{dx} = A \cos x - B \sin x \tag{3.12}$$

At $x = 1$, $\frac{du}{dx} = 0$. Hence, $A \cos(1) = B \sin(1)$, or $A = B \tan(1) = -\tan(1)$. Thus

$$u = -\tan(1) \sin x - \cos x + 1 \tag{3.13}$$

The exact value of u at $x = 0.5$ and 1 is given as follows:

$$u(0.5) = -0.6242 \qquad u(1) = -0.8508 \tag{3.14}$$

The approximate solutions at these points are

$$u(0.5) = -0.375 \qquad u(1) = -0.75 \tag{3.15}$$

Now let us add one more term in the approximating function and take $u = a + bx + cx^2$. Essential boundary condition, $u(0) = 0$, gives $a = 0$. Thus

$$u = bx + cx^2 \tag{3.16}$$

$$\frac{du}{dx} = b + 2cx \tag{3.17}$$

$$I = \int_0^1 \left\{ -\frac{1}{2}(b + 2cx)^2 + \frac{1}{2}(bx + cx^2)^2 - (bx + cx^2)^2 \right\} dx \tag{3.18}$$

Now, we have to extremize this with respect to b and c. Thus

$$\frac{\partial I}{\partial b} = \int_0^1 \{-(b + 2cx) + (bx + cx^2)x - x\} dx = 0 \tag{3.19}$$

and

$$\frac{\partial I}{\partial b} = \int_0^1 \{-(b + 2cx)2x + (bx + cx^2)x^2 - x^2\} dx = 0 \tag{3.20}$$

After integration and simplification of Eqs (3.18) and (3.19), we obtain the following expressions

$$8b + 9c = -6 \tag{3.21}$$

$$45b + 68c = -20 \tag{3.22}$$

Solving Eqs (3.21) and (3.22), we obtain, $b = -1.6402$ and $c = 0.7913$. Thus, the approximate solution is

$$u = -1.6402x + 0.7913x^2 \tag{3.23}$$

It gives $u(0.5) = -0.6223$ and $u(1) = -0.8489$, very near to the exact solution [Eq (3.14)]. Increasing the terms in the approximation polynomial will keep on increasing the accuracy. Now, let us see how the approximate solutions satisfy the natural boundary conditions. When we take linear approximation, du/dx is constant and equal to -0.75 everywhere. This obviously gives a large error in the natural boundary conditions. However, when we take quadratic approximation, we obtain

$$\frac{du}{dx} = -1.6402 + 1.5826\,x \tag{3.24}$$

giving the value -0.0576 at $x = 1$, much nearer to exact value of 0. Thus, we see that with Ritz's procedure, the natural boundary conditions also get satisfied. The accuracy of the solution obtained by this method is dependent on the number of terms used for approximating the solution.

3.3 GALERKIN METHOD

Boundary value problems may be solved without converting into variational form by the method proposed by Soviet scientist B.G. Galerkin (1871–1945). Galerkin published the method in 1915 [Galerkin (1905)]. Before this, Bubnov (1913) had applied this method to some specific problems, but did not give the method in a general form. To recognize the work of Bubnov, the method is also known as Bubnov–Galerkin method. Galerkin's method works directly on the differential equation, which is called the *strong form*. The variational form is the *weak form* because the highest order of derivative gets reduced, thus weakening the continuity requirement on primary variables.

Consider a differential equation

$$L(u) + f = 0 \tag{3.25}$$

where L is the differential operator and f is the known function. The boundary conditions may be

$$B_i u = q_i \text{ on } S_i \qquad i = 1, 2, \ldots, n \tag{3.26}$$

The following steps need to be followed in the method:

- Approximate the function as a weighted sum of linearly independent functions as in Ritz's method. However, here the approximated function must satisfy the essential as well as natural boundary conditions of the problem. The chosen functions must be complete. To give an example, if u is a function of x satisfying all the boundary conditions, it can be represented as

$$u = \psi_0 + \sum_{i=1}^{n} a_i \psi_i \tag{3.27}$$

  where ψ_i is the ith basis function, and a_i is the corresponding coefficient. Note that no coefficient is associated with ψ_0. It is the necessary function in the approximation in order to satisfy the boundary conditions.

- Substitute the approximating function in Eq. (3.25) and obtain the residual as follows:

$$R = L\left(\sum_{i=1}^{n} a_i \psi_i + \psi_0\right) + f \tag{3.28}$$

If you are very lucky, the residual R will be zero and thus you obtain the exact solution accidentally. In general, the residual is non-zero and has to be minimized by adjusting the coefficients a_i.

- In Galerkin method, coefficients are found by making the weighted integrals of residual zero. The n linearly independent basis functions, $\psi_i s$, act as weights. Thus

$$\int_D R\psi_i \, \mathrm{d}D = 0 \qquad i = 1, 2, \ldots, n \tag{3.29}$$

where D is the domain and integral is a definite integral over the domain. Equation (3.29) provides us n simultaneous equations and we can determine n unknown coefficients, thus obtaining the approximating function.

Example 3.2 **Solve the differential equation given in Example (3.1) using Galerkin method.**

Solution Let us consider the following quadratic approximation:

$$u = a + bx + cx^2 \tag{3.30}$$

In view of the essential boundary conditions, $a = 0$. The natural boundary condition gives

$$b + 2c = 0 \tag{3.31}$$

Thus, the approximating function becomes

$$u = c(x^2 - 2x) \tag{3.32}$$

The residual is given by

$$R = 2c + c(x^2 - 2x) - 1 \tag{3.33}$$

The coefficient c is found by minimizing the weighted integral of R, the weight being $(x^2 - 2x)$ as it is associated with the coefficient c. Thus

$$\int_0^1 (x^2 - 2x)[2c + c(x^2 - 2x) - 1]\mathrm{d}x = 0 \tag{3.34}$$

giving $c = 5/6$. Hence, the approximate solution is

$$u = \frac{5}{6}(x^2 - 2x) \tag{3.35}$$

This solution satisfies both the boundary conditions. Its values at $x = 0.5$ and $x = 1$ are -0.625 and -0.8333, respectively. Let us compare the values at these points with the values obtained using Ritz method and the exact solution. Table 3.1 shows this comparison. At one point, Galerkin method is closer to the exact solution, while at the other point Ritz method is closer.

Table 3.1 *Comparison of the Values of u at Two Points Obtained by Galerkin and Ritz Methods with the Exact Solution*

Point	*u*		
	Galerkin method	*Ritz method*	*Exact solution*
$x = 0.5$	-0.6250	-0.6223	-0.6242
$x = 1.0$	-0.8333	-0.8489	-0.8508

3.4 LEAST SQUARE METHOD

In Galerkin method, the error is minimized by taking the weighted integral of residual. The residual can also be minimized in a least square sense giving rise to the *least square method*. The approximating function should satisfy both the natural and essential boundary conditions. When the residual given by Eq. (3.28) is minimized in a least square sense, we get n simultaneous equations as follows:

$$\frac{\partial}{\partial a_i}\int_D R^2 \mathrm{d}D = 0 \quad i = 1, 2, \ldots, n \tag{3.36}$$

Example 3.3 Solve the differential equation given in Example (3.1) by using the least square method.

Solution Let us consider the approximating function of Example (3.2). The residual is given by Eq. (3.33). There is only one unknown c. Therefore

$$\frac{\partial}{\partial c}\int_0^1 \left(2c + c(x^2 - 2x) - 1\right)^2 \mathrm{d}x = 0 \tag{3.37}$$

or

$$\int_0^1 2\left(2 + (x^2 - 2x)\right)\left(2c + c(x^2 - 2x) - 1\right)\mathrm{d}x = 0 \tag{3.38}$$

Solving Eq. (3.38) we obtain $c = 5/7$. Therefore, the solution is

$$u = \frac{5}{7}(x^2 - 2x) \tag{3.39}$$

3.5 COLLOCATION METHOD

In this method, the residual is made equal to 0 at n points called *collocation points*. This gives n simultaneous equations, which can be solved for n unknown coefficients. The method is very simple. However, here the solution depends on the chosen collocation points. For example, if the residual in Eq. (3.33) is made equal to zero at $x = 0.5$, we get $c = 4/5$, giving a solution

$$u = \frac{4}{5}(x^2 - 2x) \tag{3.40}$$

If the residual is made equal to zero at $x = 1/3$, we get $c = 9/13$ giving

$$u = \frac{9}{13}(x^2 - 2x) \tag{3.41}$$

Of course, if we take many collocation points and approximate the function to a higher degree of polynomial, we may expect to get better solution even by this method.

3.6 SUB-DOMAIN METHOD

In this method, the domain is divided into n sub-domains and the integration of the residual is made zero over each sub-domain to generate n equations. Consider that the approximating function is given by Eq. (3.27). To find out the unknown coefficients, we can divide the whole domain into n sub-domains, integrate the residual over each sub-domain, and force it to zero for each sub-domain. Thus, we make the residual zero in an average sense over each sub-domain.

If we take the approximating function of Example (3.2) and solved it by sub-domain method, we would get the solution as

$$u = \frac{3}{4}\left(x^2 - 2x\right) \tag{3.42}$$

This can be verified easily. As there is only one unknown in the approximating function, the entire domain is taken as a sub-domain.

3.7 CONCLUSION

In this chapter, a number of classical methods have been introduced for solving the differential equations. These methods can form the basis of FEM, each method giving rise to one type of formulation. The main difference between these methods and FEM is that here a continuous function is approximated for the whole domain, whereas in FEM a number of locally continuous functions are chosen.

SOLVED PROBLEMS

3.1 The total potential energy of a simply supported beam of length l loaded by a concentrated force P at the middle is

$$\Pi = \int_0^l \frac{1}{2} EI \left(\frac{d^2 v}{dx^2}\right)^2 dx - Pv(x = l/2)$$

where EI is the modulus of flexural rigidity and v is the transverse deflection of the beam. Assuming $v = A \sin(\pi x/l)$, find out the maximum deflection of the beam by Ritz method. Assume EI to be constant.

Solution For a simply supported beam, the deflection is 0 at both the ends. The function $A \sin(\pi x/l)$ is 0 at $x = 0, l$. Thus, the essential boundary conditions are satisfied (the natural boundary conditions need not be satisfied in Ritz method). Substituting the approximating function in the expression for total potential energy, we obtain

$$\Pi = \int_0^l \frac{1}{2} EI \left(A \frac{p^2}{l^2} \right)^2 \sin^2 \frac{px}{l} \, dx - PA \tag{3.1.1}$$

For the extremization of Eq. (3.1.1), we consider

$$\frac{d\Pi}{dA} = 0$$

Thus

$$AEI \int_0^l \sin^2 \frac{\pi x}{l} \, dx - P = AEI \int_0^l \frac{\left(1 - \cos \frac{2\pi x}{l}\right)}{2} dx - P = 0$$

or

$$AEI \frac{\pi^4}{2l^4} l = P$$

which provides

$$A = \frac{2Pl^3}{\pi^4 EI} = \frac{Pl^3}{48.7EI}$$

This is the required maximum deflection of the beam. It is interesting to mention here that the maximum deflection obtained by exactly solving the governing differential equation of the beam is given as

$$v_{max} = \frac{PL^3}{48EI}$$

which is very close to the approximate solution obtained using Ritz method.

3.2 Consider the differential equation

$$\frac{d^2u}{dx^2} + x^2 u = 0, \quad 0 < x < 1$$

with boundary conditions:

$$\text{at} \quad x = 0, u = 0 \quad \text{and} \quad \text{at} \quad x = 1, du/dx = 0$$

Taking the approximation $u = a + bx + cx^2 + dx^2$, solve the differential equation by using Galerkin method.

Solution In Galerkin method, the approximation must satisfy the essential as well as natural boundary conditions. Putting $u = 0$ at $x = 0$ in the approximation, we obtain $a = 0$. Similarly, as du/dx is equal to $b + 2cx + 3dx^2$, the second boundary condition provides

$$b + 2c + 3d = 0 \quad \text{or} \quad b = -2c - 3d$$

Thus, substituting the values of a and b, the approximation becomes

$$u = c(x^2 - 2x) + d(x^3 - 3x) \tag{3.2.1}$$

which provides

$$\frac{d^2u}{dx^2} = 2c + 6dx \tag{3.2.2}$$

Substituting these approximate values in the differential equation, we get the residual R as

$$R = 2c + 6dx + cx^2(x^2 - 2x) + dx^2(x^3 - 3x) \tag{3.2.3}$$

As per Galerkin method, we now make the weighted residual with respect to weight functions $(x^2 - 2x)$ and $(x^3 - 3x)$ zero:

$$\int_0^1 \left(2c + 6dx + cx^2\left(x^2 - 2x\right) + dx^2(x^3 - 3x)\right)\left(x^2 - 2x\right)dx = 0 \tag{3.2.4}$$

$$\int_0^1 \left(2c + 6dx + cx^2\left(x^2 - 2x\right) + dx^2(x^3 - 3x)\right)\left(x^3 - 3x\right)dx = 0 \tag{3.2.5}$$

Integrating Eqs (3.2.4) and (3.2.5), the following linear equations are obtained:

$$1.0571c + 1.9607d = 0 \quad \text{and} \quad 1.9607c + 3.746\,d = 0 \tag{3.2.6}$$

The solution of Eq. (3.2.6) is $c = d = 0$. Thus, Galerkin method provides $u = 0$. This solution satisfies the differential equation and boundary conditions exactly and would have been obtained by inspection. Therefore, this solution is a trivial solution. In this case, there is no non-trivial solution. For getting a non-trivial solution, you may solve the following differential equation:

$$\frac{d^2u}{dx^2} + \frac{\pi^2}{4}u = 0, \quad 0 < x < 1$$

with the same boundary conditions used in Solved Problem 3.2.

EXERCISES

A. Multiple Choice Questions

3.1 In which of the following methods, the approximating function need not satisfy the natural boundary conditions?

(a) Collocation method
(b) Galerkin method
(c) Ritz method
(d) Least square method

3.2 The boundary conditions for a second-order ordinary differential equation are $u = 0$ at $x = 0$ and $x = 1$. A valid approximation of u satisfying essential boundary conditions and completeness is

(a) $a_1 x(x-1) + a_2 x^2(x-1) + a_3 x^4(x-1)$ (b) $a_1 + a_2 x + a_3 x^2 + a_4 x^2$

(c) $a_1 + a_2 \sin \pi x + a_3 \cos \pi x$ (d) $a_1 x(x-1) + a_2 x(x-1)^2 + a_3 x(x-1)^3$

3.3 For a second-order differential equation, the variable u is approximated as $u = a + bx + cx^2$. If the boundary conditions are $u = 5$ at $x = 0$ and $du/dx = 0$ at $x = 1$, the appropriate approximation for using it in a Galerkin method is

(a) $5 + b(2x - x^2)$ (b) $5 + bx + cx^2$

(c) $5 + x + cx^2$ (d) $b(2x - x^2)$

3.4 For the differential equation

$$\frac{d^2u}{dx^2} = 2$$

with boundary conditions

$$u = 0 \quad \text{at} \quad x = 0 \quad \text{and} \quad \frac{du}{dx} = 1 \ \text{at} \ x = 1$$

the approximation $u = bx^2$ will provide exact solution by

(a) Ritz, Galerkin, and collocation methods

(b) Ritz method but not Galerkin method

(c) collocation method but not Ritz method

(d) Galerkin method but not Ritz method

3.5 In the weighted residual method, when the weight functions are taken equal to Dirac-delta function, the method is called

(a) Bubnov–Galerkin method (b) collocation method

(c) least square method (d) Ritz method

B. Problems

3.6 A certain problem of one-dimensional heat transfer is governed by the equation $\frac{d^2\phi}{dx^2} + \phi + 1 = 0$ and boundary conditions $\phi = 1$ at $x = 0$ and $\frac{d\phi}{dx} = 1$ at $x = 1$. Solve this problem by using Galerkin method.

3.7 Given a differential equation:

$$\frac{d^2}{dx^2}\left(EI_{zz}\frac{d^2v_0}{dx^2}\right) - \rho\omega^2 v_0 = 0$$

with boundary conditions:

$$v_0 = \frac{dv_0}{dx} = 0 \quad \text{at} \quad x = 0 \text{ and } l \quad \text{(both essential).}$$

EI_{zz} is a function of x and ω is a constant unknown angular velocity.

(a) Derive the variational functional associated with this problem. State explicitly the conditions to be satisfied by the δv as well as the properties of the operator used in the derivation.

(b) Using the approximation

$$v_0 = \sum_{i=1}^{n} a_i \phi_i(x) \text{ (a_i are unknown constants) and the Ritz method, derive}$$

the algebraic equations satisfied by the unknown constants in the following form:

$$\sum_{j=1}^{n} \left(k_{ij} - \omega^2 M_{ij}\right) a_j = 0 \quad \text{for} \quad i = 1, 2, \ldots n$$

State explicitly the condition to be satisfied by ϕ_i.

3.8 The variational functional of the problem shown in Fig. 3.1 is

$$I = \int_0^l \left(-\frac{1}{2} AE \left(\frac{du}{dx} \right)^2 + pu \right) dx + (Pu)\Big|_{x=l/2}$$

(a) Using the approximation $u = a_1 \dfrac{x(l-x)}{l^2}$ (a_1 is unknown constant), express I in terms of a_1

(Note: A, E, p are constants.).

(b) Find a_1 using Ritz method.

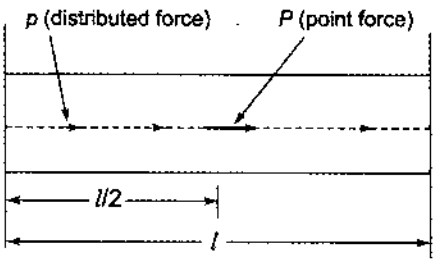


Figure 3.1 *A fixed–fixed rod of length l*

3.9 Solve the following differential equation using Galerkin method:

$$-\frac{d^2u}{dx^2} - cu + x^2 = 0 \quad \text{for} \quad 0 \le x \le 1$$

the boundary conditions being

$$\frac{du}{dx}(x=0) = 1 \quad \frac{du}{dx}(x=1) = \frac{4}{3}$$

3.10 Solve the following equation using Galerkin method:

$$\frac{d}{dx}\left(u\frac{du}{dx}\right) - f = 0 \quad \text{for} \quad 0 < x < 1$$

Boundary conditions: (i) Essential: $u = \sqrt{2}$ at $x = 0$

(ii) Natural : $\frac{du}{dx} = 0$ at $x = 1$

Take f to be a linear function of x : $b_1 + b_2 x$ (b_1, b_2 are constants).

Take $u = \sqrt{2} + a_1 x + a_2 x^2$ $(a_1, a_2 = \text{constants})$ as the approximation for the Galerkin method.

3.11 A certain problem of one-dimensional heat transfer is governed by the equation $\frac{d^2\phi}{dx^2} + \phi + 1 = 0$ and boundary conditions $\phi = 1$ at $x = 0$ and $\frac{d\phi}{dx} = 1$ at $x = 1$.

Use Ritz method to solve this problem. Approximate the function by a quadratic polynomial and compare with the exact solution.

3.12 Solve the following differential equation by Ritz method:

$-\frac{d^2u}{dx^2} - cu + x^2 = 0$ for $0 \le x \le 1$ for the boundary conditions: $\frac{du}{dx}(x = 0) = 1$, $\frac{du}{dx}(x = 1) = \frac{4}{3}$.

Solve it by all other methods which you have studied in this chapter.

REFERENCES

Bubnov, IG, 1913, 'Report on the works of Professor Timoshenko which were awarded the Zhuranski Prize'. *Symposium of the Institute of Communication Engineers*, No. 81, All union Special Planing office (SPB) (in Russian).

Galerkin, BG, 1915, 'Rods and plates: Series occurring in various questions concerning the elastic equilibrium of rods and plates', *Engineers Bulletin (Vestnik Inzhenerov)*, Vol. 19, pp. 897–908 (in Russian).

Leissa, AW, 2005, 'The historical bases of the Rayleigh and Ritz methods', *Journal of Sound and Vibration*, Vol. 287, pp. 961–978.

Ritz, W, 1908, 'Über eine neue methode zur Lösung gewisser Variations problem der mathematischen Physik', *Journal für Reine und Angewande Mathematik*, Vol. 135, pp. 1–61.

4
CHAPTER

Ritz and Galerkin FEM Formulations

4.1 Introduction

4.2 Completeness and Compatibility

4.3 Concept of Shape Functions

4.4 Developing Elemental Equations by Ritz Method

4.5 Developing Elemental Equations by Galerkin Method

4.6 Conclusion

4.1 INTRODUCTION

In Chapter 3, a number of classical methods for solving differential equations have been described, which approximate the solutions in the form of continuous functions. The adjustable parameters of the functions are found to minimize the error in the solution. In the finite element method (FEM), a similar technique is adopted, except that the approximating functions are only piecewise continuous. In FEM, a domain is discretized into a number of small elements. In each element, the solution is approximated by simple functions, usually but not necessarily, polynomial functions. Inside an element, the function is continuous. Continuity requirement at the interface between the two elements is somewhat weaker, as discussed later in this chapter. Each method described in Chapter 3 can form the basis of a finite element formulation. However, in this chapter, we shall restrict our discussion to two types of FEM formulations: Ritz FEM and Galerkin FEM.

In Ritz FEM, the finite element equations are obtained by following a procedure similar to the classical Ritz method. The difference is that in FEM, piecewise continuous functions are chosen instead of a globally continuous function for the whole domain. The elemental equations can be obtained by writing the variational expression for an element, putting a continuous interpolation function in that expression, and obtaining the coefficients of the interpolation function by extremizing the variational expression. The interpolation function, which approximates the actual solution should be complete and should satisfy the compatibility conditions. The concept of completeness and compatibility is discussed in detail in the next section.

Galerkin FEM follows a procedure similar to the classical Galerkin method. The piecewise continuous approximation is employed here too. However, before applying the method, equations are converted into the weak form. This is achieved by carrying out integration by parts.

4.2 COMPLETENESS AND COMPATIBILITY

In the context of classical Ritz method, completeness implies that basis functions of the approximating functions are such that if enough terms are taken, the primary variables and their derivatives can be approximated as accurately as we wish. In the context of finite elements, the set of basis functions are said to be complete if they can approximate the primary variables, and their derivatives appearing in the variational form, as accurately as desired by reducing the size of element. Thus, leaving aside the computational difficulties, if the size of the elements approach 0, the exact values of the primary variables and their derivatives (upto the order appearing in variational form) should be obtained.

A polynomial series is complete if it is of high enough degree and no terms are omitted. Fourier series are also complete. By 'high enough degree' we mean that the highest order derivative appearing in the variational form can be represented. For example, if

$$I = \int_0^h \left\{ \left(\frac{\mathrm{d}u}{\mathrm{d}x} \right)^2 - u \right\} \mathrm{d}x \tag{4.1}$$

where h is the length of the element, then $u = ax + b$ is a complete approximation (or interpolation) function, as it can represent du/dx by b. However, $u = a$ is not complete because it makes du/dx zero. The interpolation function $u = a + bx + cx^2$ is complete. It represents $\frac{du}{dx}$ by $b + 2cx$. However, $u = a + cx^2$ is not complete. Why? After all it can represent $\frac{du}{dx}$ by $2cx$. The reason is that it cannot approximate a constant derivative term. The approximate $\frac{du}{dx}$ will always be zero at $x = 0$. In practice, one may have a non-zero du/dx at $x = 0$.

Now, consider the following variational expression:

$$I = \int_0^h \left(\left(\frac{d^2u}{dx^2} \right)^2 - qu \right) dx - \left(\frac{d^3u}{dx^3} \right)_{x=h} \tag{4.2}$$

For this, a complete approximate function will be $u = a + bx + cx^2 + dx^3$. We have considered cubic polynomial because we have to represent $\frac{d^3u}{dx^3}$ at $x = h$, even though the highest order of derivative inside the integrand is 2. Thus, for completeness, the degree of approximating polynomial should be equal to the highest order of derivative in the variational expression of the elements.

We should also ensure that whatever be the approximation, I should not become infinite. If u and its second derivative are finite inside the element and the third derivative of u is also finite at $x = h$, then I of Eq. (4.2) will also be finite. In order that d^2u/dx^2 is finite inside the element, du/dx should be continuous. This is ensured as we usually select a smooth interpolation function for the element. For example, all the derivatives of $u = a + bx + cx^2 + dx^3$ exist. Now, let us pay attention to boundaries. If at $x = h$, du/dx is not continuous and changes abruptly, d^2u/dx^2 will not be finite there. What happens if d^2u/dx^2 is finite everywhere except at the boundaries? Well, in Eq. (4.2), if boundaries are considered to be encompassed between the limits of integration, the value of I becomes undefined. Considering the boundaries to be outside the integration limits, of course makes I finite. It appears perfect if, for the variational form of Eq. (4.2), we impose a compatibility condition between the elements that at the common nodes of the two elements, du/dx for both the elements should be same. In general, the approximating (interpolation) functions of the elements should have continuity equal to at least one order less than the highest order derivative in the integral. In Eq. (4.1), the approximation should be at least C^0 continuous through the domain. This will ensure that $\frac{du}{dx}$ is always finite inside the whole domain. The piecewise continuous polynomial may be different in two neighbouring elements. However, the values of primary variables should be the same for both the polynomials at the common node. Equation (4.2) requires C^1 continuity, i.e., u and $\frac{du}{dx}$ should be the same for both the elements at common nodes. Thus, for compatibility, one should check the highest order of the derivative inside the

integral expression. If the highest order of derivative is m, the function should be continuous up to order $m - 1$ throughout the domain, i.e., the function should be C^{m-1} continuous. In that case, the elements are said to be *compatible*. Note that for compatibility, we check the highest order of the derivative in the integral expression of the full variational form.

It is worth mentioning that in his book, Zienkiewicz (1979) has mentioned one of the criterion of convergence as follows:

> The displacement functions should be so chosen that the strains at the interface between elements are finite (even though indeterminate).

He has further explained that in the case of plate and shell problems, strains are defined by the second derivatives of deflection, whereas in the plane problems, they are defined by the first derivatives of the deflections. Zienkiewicz has not provided a mathematical proof for the criterion. The mathematical discussion on this criterion was carried out by Pian and Tong (1967) and Oliviera (1968). It is to be mentioned that compatibility requirement is not a strong requirement. While discussing the plate problems, we shall see that some non-compatible elements also provide good results. Of late, there is a growing interest in using discontinuous FEM in which the compatibility condition is not fulfilled, i.e., there is discontinuity at the interface [Li (2006) and Zienkiewicz, et al. (2003)]. However, this book will not deal with discontinuous FEM.

4.3 CONCEPT OF SHAPE FUNCTIONS

Consider the problem of minimizing I in Eq. (4.1). For this problem, the lowest degree of incomplete polynomial is $u = a + bx$. At the same time, there should be at least C^0 continuity. If we take a two-noded element of length h, we can express the constants a and b in terms of the nodal values of the primary variables. This will automatically ensure C^0 continuity. If the local coordinate of node 1 is 0 and that of node 2 is h and the values of the primary variable at these nodes are u_1 and u_2, respectively, then

$$u_1 = a + b \times 0 = a \tag{4.3}$$

and

$$u_2 = a + bh \times u_1 = bh \tag{4.4}$$

Substituting the value of a from Eq. (4.3) into Eq. (4.4), we get

$$b = \frac{u_2 - u_1}{h} \tag{4.5}$$

Hence, by expressing the constants a and b in terms of the nodal values of the primary variable, the piecewise continuous polynomial can be expressed as

$$u = u_1 + \frac{u_2 - u_1}{h} x = \left(1 - \frac{x}{h}\right) u_1 + \frac{x}{h} u_2 = N_1 u_1 + N_2 u_2 \tag{4.6}$$

where N_1 and N_2 are called *shape functions*, because they provide information regarding the shape of piecewise continuous approximating polynomial. In this case, the shape is linear.

We can also have three-noded elements. In such a case, a quadratic interpolation function $u = a + bx + cx^2$ may be considered. Let the coordinates of the three nodes be x_1, x_2, and x_3, respectively. Then

$$u_1 = a + bx_1 + cx_1^2 \tag{4.7}$$

$$u_2 = a + bx_2 + cx_2^2 \tag{4.8}$$

$$u_3 = a + bx_3 + cx_3^2 \tag{4.9}$$

Solving this system of equations, we can express the coefficients a, b, and c in terms of nodal values of the primary variables and arrange the expression for u to get the following form:

$$u = N_1u_1 + N_2u_2 + N_3u_3 \tag{4.10}$$

where N_1, N_2, and N_3 are the shape functions corresponding to nodes 1, 2, and 3, respectively. The same procedure can be extended to n-noded element. In that case, one has to solve n simultaneous equations for obtaining the coefficients of polynomial interpolation function in terms of the nodal values of the primary variables. These coefficients are substituted back in the polynomial expression and a rearrangement provides the following form:

$$u = N_1u_1 + N_2u_2 + \ldots + N_nu_n = \sum_{i=1}^{n} N_iu_i \tag{4.11}$$

where N_i is the ith shape function and u_i is the value of primary variable at the ith node. However, this procedure of obtaining the shape functions is tedious. A somewhat simpler way is to obtain the shape function on the basis of the three properties discussed below. One-dimensional C^0 continuous polynomial shape functions satisfy the following three properties:

Property 1 All shape functions of an n-noded element are polynomials of $n - 1$ degree.

This is because for an n-noded element, interpolation function will be of $n - 1$ degree. It is possible to represent a polynomial function of $n - 1$ degree in such a manner that its value is zero at all nodes except one node which can be any one out of the n nodes. Hence, the shape functions associated with all nodes should be of the same degree.

Property 2 For any shape function

$$N_i(x_j) = \delta_{ij} \tag{4.12}$$

where $N_i(x_j)$ is the value of the ith shape function at the jth node and δ_{ij} is the Kronecker delta (for $i = j$, $\delta_{ij} = 1$ and for $i \neq j$, $\delta_{ij} = 0$).

This property can be justified as follows: Let us assume a variable field such that it is non-zero at i^{th} node and all other nodal values of the variable are zero. In that case

$$u = N_iu_i \tag{4.13}$$

Now, at $x = x_i$, $u = u_i$. Hence, we have from Eq. (4.13)

$$u_i = N_i(x_i)u_i \quad \text{or} \quad N_i(x_i) = 1 \tag{4.14}$$

At $x = x_j, j \neq i$, the value of the $u = 0$. Hence, from Eq. (4.13), we can write

$$0 = N_i(x_j)\, u_i \quad \text{or} \quad N_i(x_j) = 0 \tag{4.15}$$

This proves Property 2.

Property 3 The sum of all shape functions of an element is unity.

This can be proved as follows: Assume that u is constant and equal to c throughout. Thus, the nodal values of the variable will also be c. From Eq. (4.11), we can write

$$c = \sum_{i=1}^{n} cN_i \quad \text{or} \quad \sum_{i=1}^{n} N_i = 1 \tag{4.16}$$

Lagrange's interpolation functions satisfy these properties. For n-noded element, Lagrange shape functions are given by

$$\begin{aligned}
N_1 &= \frac{(x_2 - x)(x_3 - x)(x_4 - x)\ldots(x_n - x)}{(x_2 - x_1)(x_3 - x_1)(x_4 - x_1)\ldots(x_n - x_1)} \\
N_2 &= \frac{(x_1 - x)(x_3 - x)(x_4 - x)\ldots(x_n - x)}{(x_1 - x_2)(x_3 - x_2)(x_4 - x_2)\ldots(x_n - x_2)} \\
&\vdots \\
N_n &= \frac{(x_1 - x)(x_3 - x)(x_4 - x)\ldots(x_{n-1} - x)}{(x_1 - x_n)(x_3 - x_n)(x_4 - x_n)\ldots(x_{n-1} - x_n)}
\end{aligned} \tag{4.17}$$

It can be easily seen that for these shape functions, the first two properties are satisfied. It is also easy to show that $\sum_{i=1}^{n} N_i = 1$ at all nodes, in view of Property 2. However, we have to prove that the sum of shape functions is 1 everywhere. We argue it as follows: It is clear that $\sum_{i=1}^{n} N_i$ can be at most an $n - 1$ degree polynomial. Thus

$$\sum_{i=1}^{n} N_i = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^{n-1} \tag{4.18}$$

Its value at all the nodes is 1. Hence

$$\begin{aligned}
1 &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_n x^{n-1} \\
1 &= a_0 + a_1 x_1 + a_2 x_2^2 \cdots + a_n x^{n-1} \\
&\vdots \\
1 &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_n x_n^{n-1}
\end{aligned} \tag{4.19}$$

This is a system of n equations in n unknowns and should give a unique solution. One solution is $a_0 = 1$ with all other coefficients equal to 0 (by inspection). Hence, from Eq. (4.18)

$$\sum_{i=1}^{n} N_i = 1 \tag{4.20}$$

throughout the element.

4.4 DEVELOPING ELEMENTAL EQUATIONS BY RITZ METHOD

Consider the problem

$$\frac{d^2u}{dx^2} + u - 1 = 0 \tag{4.21}$$

with boundary conditions

$$u(0) = 0 \quad \text{and} \quad \frac{du}{dx} = 0 \text{ at } x = 1 \tag{4.22}$$

Let us develop the elemental equations for this problem. Consider an element whose length is h. Adopting the local coordinate system, the coordinate of the first node is 0 and that of second is h. For the present problem, the form of the differential equation remains unchanged in the local coordinate system. In general, if X is the global coordinate, it is related to the local coordinate x of the element with first node as i by $X = X_i + x$, where X_i is the global coordinate of node i. It is clear that $dX = dx$ and thus Eq. (4.21) is the same in local and global coordinates. For obtaining the variational form of the differential equation we write

$$\delta I = \int_0^h \left[\frac{d^2u}{dx^2} + u - 1\right] \delta u \, dx = 0 \tag{4.23}$$

Integrating by parts and using the properties of the variational operator, we obtain

$$dI = \left(\frac{du}{dx}\right) du\Big|_0^h - \int_0^h \frac{1}{2}\, \delta\left(\frac{du}{dx}\right)^2 dx + \int_0^h \frac{1}{2}\delta(u^2)\, dx - \int_0^h du dx = 0 \tag{4.24}$$

Hence,

$$I = -\int_0^h \frac{1}{2}\left(\frac{du}{dx}\right)^2 dx + \int_0^h \frac{1}{2}u^2 dx - \int_0^h u dx + \left(\frac{du}{dx}\right) u\Big|_0^h \tag{4.25}$$

We observe that completeness up to first degree polynomial and C^0 continuity throughout the domain is enough. Thus, two-noded elements are enough for solving the problem. However, we can also take higher-noded C^0 continuity elements for better accuracy. Let u^e be the approximation inside an n-noded element:

$$u^e = \sum_{i=1}^{n} N_i u_i = \lfloor N_1\ N_2 \ldots N_n \rfloor \begin{Bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{Bmatrix} = \lfloor N \rfloor \{u^{ne}\} \tag{4.26}$$

Note that we are using 'half-bracket' symbol to denote a row vector, which is the transpose of a column vector with the same entries written in a column. By vector here we mean a one-dimensional array, not necessarily a physical quantity such as velocity. We are writing u^e in the form of matrix multiplication, as we are interested in getting the final finite element equations containing the matrices. A *column vector* is a special matrix with one column and a *row vector* is a special matrix with one row. In that way, u^e can be considered as a vector having only one element (entry), although the term scalar is more appropriate for it. Now, we know that if we pre-multiply a column vector with its transpose, we get sum of the squares of each entry, i.e.,

$$\lfloor v_1\ v_2\ .\ .\ v_n \rfloor \begin{Bmatrix} v_1 \\ v_2 \\ . \\ . \\ v_n \end{Bmatrix} = v_1^2 + v_2^2 + \cdots + v_n^2 \tag{4.27}$$

Thus, as u^e is also a vector with one element

$$(u^e)^2 = u^{e\mathrm{T}} u^e = \left(\lfloor N \rfloor \{u^{ne}\}\right)^{\mathrm{T}} \lfloor N \rfloor \{u^{ne}\} \tag{4.28}$$

Using the property of the transpose of the matrix product of two matrices

$$([A][B])^{\mathrm{T}} = [B]^{\mathrm{T}}[A]^{\mathrm{T}} \tag{4.29}$$

and noting that the transpose of a column vector is the corresponding row vector and *vice versa*, Eq. (4.28) can be written as

$$(u^e)^2 = \lfloor u^{ne} \rfloor \{N\} \lfloor N \rfloor \{u^{ne}\} \tag{4.30}$$

Some students may wonder why $(u^e)^2$ has not been written in the following way:

$$(u^e)^2 = \lfloor N \rfloor \{u^{ne}\} \lfloor N \rfloor \{u^{ne}\} \tag{4.31}$$

There is nothing wrong with Eq. (4.31). However, the form given in Eq. (4.30) is more convenient for differentiating the expressions with respect to nodal variables. This will be clear by referring to Appendix A.2. In fact, usually, in a variational expression, we shall substitute u^e in the following form:

$$u^e = \lfloor u^{ne} \rfloor \{N\} \tag{4.32}$$

instead of the form given in Eq. (4.26).

Differentiating u^e with respect to x we obtain, from Eq. (4.26)

$$\frac{\mathrm{d}u^e}{\mathrm{d}x} = \sum_{i=1}^{n} \frac{\mathrm{d}N_i}{\mathrm{d}x} u_i = \left\lfloor \frac{\mathrm{d}N_1}{\mathrm{d}x} \ \frac{\mathrm{d}N_2}{\mathrm{d}x} \ \cdots \ \frac{\mathrm{d}N_n}{\mathrm{d}x} \right\rfloor \begin{Bmatrix} u_1 \\ u_2 \\ . \\ . \\ u_n \end{Bmatrix} = \lfloor N_{,x} \rfloor \{u^{ne}\} \tag{4.33}$$

Note the use of comma notation in $\lfloor N_{,x} \rfloor$ to indicate the derivative of each entry of the row with respect to x. Similar to Eq. (4.30), we can write

$$\left(\frac{\mathrm{d}u^e}{\mathrm{d}x}\right)^2 = \lfloor u^{ne} \rfloor \{N_{,x}\} \lfloor N_{,x} \rfloor \{u^{ne}\} \tag{4.34}$$

Substituting Eqs (4.30), (4.32), and (4.34) in Eq. (4.25), we obtain

$$I = -\int_0^h \frac{1}{2} \lfloor u^{ne} \rfloor \{N_{,x}\} \lfloor N_{,x} \rfloor \{u^{ne}\} \mathrm{d}x + \int_0^h \frac{1}{2} \lfloor u^{ne} \rfloor \{N\} \lfloor N \rfloor \{u^{ne}\} \mathrm{d}x$$

$$-\int_0^h \lfloor u^{ne} \rfloor \{N\} \, \mathrm{d}x + \lfloor u_1 \ u_2 \ \ldots \ u_n \rfloor \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=0} \\ 0 \\ 0 \\ . \\ . \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=h} \end{Bmatrix} \tag{4.35}$$

In Eq. (4.25), I is a functional of u and $\mathrm{d}u/\mathrm{d}x$, whereas in Eq. (4.35) I is function of nodal values $u_1, u_2, \ldots, u_n$. Note how the last term of Eq. (4.25) has been written as the multiplication of a row vector and a column vector. Also, in the last term, we have made no approximation for the derivatives. Thus, $(\mathrm{d}u/\mathrm{d}x)_{x=0}$ means the exact derivative of u with respect to x at node 1, which we shall denote by $(\mathrm{d}u/\mathrm{d}x)_1$. Similarly, $(\mathrm{d}u/\mathrm{d}x)_{x=h}$ will be denoted by $(\mathrm{d}u/\mathrm{d}x)_n$, the exact derivative at node n. In a more compact form, Eq. (4.35) can be written as

$$I = -\int_0^h \frac{1}{2} \lfloor u^{ne} \rfloor \{N_{,x}\} \lfloor N_{,x} \rfloor \{u^{ne}\} \mathrm{d}x + \int_0^h \frac{1}{2} \lfloor u^{ne} \rfloor \{N\} \lfloor N \rfloor \{u^{ne}\} \mathrm{d}x$$

$$-\int_0^h \lfloor u^{ne} \rfloor \{N\} \, \mathrm{d}x + \lfloor u^{ne} \rfloor \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ 0 \\ 0 \\ . \\ . \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_n \end{Bmatrix} \tag{4.36}$$

For obtaining the solution, the variational expression given in Eq. (4.36) has to be extremized (or more precisely, we make the variational expression stationary). This means that the partial derivatives with respect to each nodal value have to be equated to zero. We write this in a concise form as

$$\frac{\partial I}{\partial \{u^{ne}\}} = \{0\} \tag{4.37}$$

Equation (4.37) actually means

$$\frac{\partial I}{\partial u_1} = 0, \quad \frac{\partial I}{\partial u_2} = 0, \quad \ldots, \quad \frac{\partial I}{\partial u_n} = 0 \tag{4.38}$$

Equation (4.37) is the concise representation of Eq. (4.38), in which the left hand side may be interpreted as the differentiations of I with respect to a vector, which is actually the gradient of I with respect to components of vector. The right hand side is a column vector of size n with each element (entry) equal to zero. Using the rule of differentiation with respect to a vector (Appendix A.2), we get

$$-\int_0^h \{N_{,x}\}\lfloor N_{,x}\rfloor\{u^{ne}\}\mathrm{d}x + \int_0^h \{N\}\lfloor N\rfloor\{u^{ne}\}\mathrm{d}x - \int_0^h \{N\}\mathrm{d}x + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_n \end{Bmatrix} = \{0\} \tag{4.39}$$

Hence

$$\int_0^h \left[\{N_{,x}\}\lfloor N_{,x}\rfloor\mathrm{d}x - \int_0^h \{N\}\lfloor N\rfloor\mathrm{d}x\right]\{u^{ne}\} = -\int_0^h \{N\}\mathrm{d}x + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_n \end{Bmatrix} \tag{4.40}$$

This is the elemental equation for an n-noded element.

For a typical two-noded element

$$N_1 = 1 - \frac{x}{h} \quad \text{and} \quad N_2 = \frac{x}{h} \tag{4.41}$$

Therefore

$$\{N_{,x}\} = \begin{Bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x} \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}x} \end{Bmatrix} = \begin{Bmatrix} \dfrac{-1}{h} \\ \dfrac{1}{h} \end{Bmatrix} \tag{4.42}$$

Hence

$$\int_0^h \{N_{,x}\}\lfloor N_{,x} \rfloor \mathrm{d}x = \int_0^h \begin{Bmatrix} \dfrac{-1}{h} \\ \dfrac{1}{h} \end{Bmatrix} \begin{bmatrix} \dfrac{-1}{h} & \dfrac{1}{h} \end{bmatrix} \mathrm{d}x = \int_0^h \frac{1}{h^2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \mathrm{d}x = \frac{1}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4.43}$$

Note that integration of the matrix has been carried out by integrating each of its elements (entries). Similarly

$$\int_0^h \{N\}\lfloor N \rfloor \mathrm{d}x = \int_0^h \begin{Bmatrix} 1-\dfrac{x}{h} \\ \dfrac{x}{h} \end{Bmatrix} \begin{bmatrix} 1-\dfrac{x}{h} & \dfrac{x}{h} \end{bmatrix} \mathrm{d}x$$

$$= \begin{bmatrix} \displaystyle\int_0^h \left(1-\frac{x}{h}\right)^2 \mathrm{d}x & \displaystyle\int_0^h \left(1-\frac{x}{h}\right)\frac{x}{h}\mathrm{d}x \\ \displaystyle\int_0^h \left(1-\frac{x}{h}\right)\frac{x}{h}\mathrm{d}x & \displaystyle\int_0^h \left(\frac{x}{h}\right)^2 \mathrm{d}x \end{bmatrix} = \begin{bmatrix} \dfrac{h}{3} & \dfrac{h}{6} \\ \dfrac{h}{6} & \dfrac{h}{3} \end{bmatrix} = \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{4.44}$$

and

$$\int_0^h \{N\}\mathrm{d}x = \int_0^h \begin{Bmatrix} 1-\dfrac{x}{h} \\ \dfrac{x}{h} \end{Bmatrix} \mathrm{d}x = \begin{Bmatrix} \dfrac{h}{2} \\ \dfrac{h}{2} \end{Bmatrix} \tag{4.45}$$

Substituting Eqs (4.43) to (4.45) for a two-noded element ($n = 2$) in Eq. (4.40):

$$\left[\frac{1}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} - \frac{h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right]\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} -\dfrac{h}{2} \\ -\dfrac{h}{2} \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{4.46}$$

Equation (4.46) is typical elemental equation for a two-noded element. It is to be understood that u_1 and u_2 refer to local nodal values. For example, if we discretize our domain (from $x = 0$ to $x = 1$) into two equal size elements as shown in Fig. 4.1, for the second element, the vector of nodal values will be $\{u_1 \;\; u_3\}^{\mathrm{T}}$ instead of $\{u_1 \;\; u_2\}^{\mathrm{T}}$. The same thing applies to the derivative at the nodes. Putting $h = 0.5$, the elemental equations for the first element are

$$\begin{bmatrix} 1.833 & -2.083 \\ -2.083 & 1.833 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{4.47}$$

Similarly, the elemental equations for the second element are

$$\begin{bmatrix} 1.833 & -2.083 \\ -2.083 & 1.833 \end{bmatrix} \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{4.48}$$

Figure 4.1 *Discretization of the domain into two elements*

We can assemble Eqs (4.47) and (4.48) using the procedure described in Chapter 1, which requires writing the elemental equations in the global form and simply adding them. A quicker way to accomplish the same task is as follows. First, make a format of final global equations with empty entries. Then, keep on putting the entries corresponding to elemental equations in corresponding places of global system of equations. If two or more entries are kept at the same place, they simply get added. For example, for the two-element problem, the format of global equations will be

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} + \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix} \tag{4.49}$$

We now place Eq. (4.47) in the space provided by the first two rows and first two columns, which gives us

$$\begin{bmatrix} 1.833 & -2.083 & 0 \\ -2.083 & 1.833 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.25 \\ 0 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \\ 0 \end{Bmatrix} \tag{4.50}$$

This is because the rows 1 and 2 and columns 1 and 2 of the elemental equations given in Eq. (4.33) correspond to the rows 1 and 2 and columns 1 and 2 of the elemental equations written in global form, respectively. Next, we place Eq. (4.48) in the space provided by the rows 2 and 3 and columns 2 and 3, because the first row and first column of elemental equations corresponds to the second row and second column, respectively, of the global version of these equations, and the second row and second column correspond to the third row and third column, respectively, of the global version. After appending Eq. (4.48) in Eq. (4.50), we obtain

$$\begin{bmatrix} 1.833 & -2.083 & 0 \\ -2.083 & 1.833+1.833 & -2.083 \\ 0 & -2.083 & 1.833 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.25-0.25 \\ -0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 - \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{4.51}$$

Note that in the locations of non-zero entries, new entries are added to the previous entries. Equation (4.51) is simplified to provide the following assembled (global) finite element equations:

$$\begin{bmatrix} 1.833 & -2.083 & 0 \\ -2.083 & 3.666 & -2.083 \\ 0 & -2.083 & 1.833 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.5 \\ -0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ 0 \\ -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{4.52}$$

Now, putting the boundary condition that the first derivative of u vanishes at $x = 1$, corresponding to node 3, we obtain

$$\begin{bmatrix} 1.833 & -2.083 & 0 \\ -2.083 & 3.666 & -2.083 \\ 0 & -2.083 & 1.833 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.25 \\ -0.5 \\ -0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ 0 \\ 0 \end{Bmatrix} \tag{4.53}$$

We still cannot solve the system of equations given by Eq. (4.53) as $(\mathrm{d}u/\mathrm{d}x)_1$ is unknown. We get one more equation $u_1 = 0$ from the essential boundary condition $u = 0$ at $x = 0$. In Eq. (4.53), the first equation may be replaced by this equation, as we did in Chapter 1 for applying the essential geometric boundary conditions. However, an efficient computational strategy is to eliminate the first row and first column from the system of equations. This reduces the number of equations to be solved. Removing one row from a system of linear equations does not require any modification in the remaining set of equations. However, when a column is removed, some modification is required in the right hand side vector. Suppose the system of equations is

$$[K]\{u\} = \{R\} \tag{4.54}$$

where $[K]$ is a coefficient matrix of size $n \times n$, $\{u\}$ is a $n \times 1$ vector of variables and $\{R\}$ is the right hand side vector of size $n \times 1$. If it is known that i^{th} variable u_i is equal to u_i^*, then i^{th} row and i^{th} columns can be eliminated. When ith column is removed, the following modification is needed in the right hand side vector

$$R_j^{\text{new}} = R_j - K_{ji} u_i^* \qquad j = 1 \text{ to } n \tag{4.55}$$

where K_{ji} indicates the entry (element) at j^{th} row and i^{th} column of the matrix $[K]$. In the present case, the prescribed variable u_1 is 0 and no modification is needed in the right hand side vector. The reduced system of equations is

$$\begin{bmatrix} 3.666 & -2.083 \\ -2.083 & 1.833 \end{bmatrix} \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} -0.5 \\ -0.25 \end{Bmatrix} \tag{4.56}$$

Solving it we get

$$u_2 = -0.604 \quad \text{and} \quad u_3 = -0.822 \tag{4.57}$$

Thus, $u(1) = -0.822$. (We have been retaining only three digits after decimal in calculations. The answer may differ slightly depending on the precision used.) It can be easily verified that FEM with three 2–noded elements gives $u(1) = -0.834$ and with six 2–noded elements, it gives $u(1) = -0.848$. The exact solution of the differential equation is

$$u = -\tan 1 \sin x - \cos x + 1 \tag{4.58}$$

(It is suggested that the reader finds out the exact solution of the differential equation or at least verify that Eq. (4.58) satisfies the differential equation and boundary conditions.) The exact values of u at $x = 0.5$ and 1 are

$$u(0.5) = -0.6242 \quad \text{and} \quad u(1) = -0.8508 \tag{4.59}$$

which are close to the FEM solution even with two elements.

We have illustrated the Ritz FEM by a simple one-dimensional problem. The procedure is similar for other problems. For any element, we substitute the approximating function in terms of the unknown nodal displacement and extremize the function with respect to the displacement vector. While developing the elemental equations for a differential equation containing the independent variable x, one should not forget to express the differential equation in the form of local variable, if the integration limits are in terms of local variables.

4.5 DEVELOPING ELEMENTAL EQUATIONS BY GALERKIN METHOD

For obtaining the finite element equations by Galerkin method, one has to obtain the weak form of the differential equation. Galerkin method is a special form of weighted residual method. In the weighted residuals method, we integrate the weighted residuals (obtained after putting the approximation for dependent variable in the differential equation) over the domain and make them equal to zero. Before actually putting the approximation for the dependent variable in the differential equation for obtaining the residual, often integration by parts is carried out to reduce the differentiability requirement of the approximation. We illustrate this method by considering the problem discussed in Section 4.4.

Let u^e be the approximation for u for the differential equation given in Eq. (4.21), in the eth element. Then the residual for the element is given by

$$R^e = \frac{d^2 u^e}{dx^2} + u^e - 1 \tag{4.60}$$

Let w be a suitable weight function. Making the integral of weighted residual equal to zero, we write

$$\int_0^h wR^e \mathrm{d}x = \int_0^h w\left[\frac{\mathrm{d}^2u^e}{\mathrm{d}x^2} + u^e - 1\right]\mathrm{d}x = 0 \tag{4.61}$$

In the first term of Eq. (4.61), u^e is differentiated twice. We can integrate this term by parts such that the differentiability requirement in w and u becomes the same. This provides

$$w\frac{\mathrm{d}u^e}{\mathrm{d}x}\bigg|_0^h - \int_0^h \frac{\mathrm{d}w}{\mathrm{d}x}\frac{\mathrm{d}u^e}{\mathrm{d}x}\mathrm{d}x + \int_0^h wu^e\mathrm{d}x - \int_0^h w\mathrm{d}x = 0 \tag{4.62}$$

Note that Eq. (4.62) has been obtained by only integrating the first term by parts. The other two terms have been left as they are.

In Eq. (4.62), the highest order of derivative in the entire expression is one. Therefore, a single degree polynomial satisfies the completeness requirement. Also, we observe that the highest order of derivative in the integrand is one. Thus, C^0 continuity is good enough for the approximating function. A two-noded element with one degree of freedom at each node fulfils this requirement. Thus, we can use the approximation

$$u^e = \lfloor N_1 \;\; N_2 \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \tag{4.63}$$

However, we shall develop the equations for n-noded element with the approximation given by Eq. (4.26). The first derivative of u^e is given by Eq. (4.33). We put this approximation everywhere inside the element, but not at the boundaries of the element. Like in the Ritz method, we put the exact derivatives at the boundaries, that is, we assume that

$$\left(\frac{\mathrm{d}u^e}{\mathrm{d}x}\right)_{x=0} = \left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=0} \quad \text{and} \quad \left(\frac{\mathrm{d}u^e}{\mathrm{d}x}\right)_{x=h} = \left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=h} \tag{4.64}$$

In Galerkin method, w is approximated in the same way as the shape function

$$w = \lfloor w^{ne} \rfloor \{N\} \tag{4.65}$$

where $\lfloor w^{ne} \rfloor$ is the row vector of the arbitrary nodal weights w_1 and w_2. (There are methods in which the form of the weight function can be quite different from the approximating function.) The basis of Galerkin method can be explained as follows. When an approximating function is represented by Eq. (4.11), the complete set of linearly independent function is given by the shape functions. That means, it is assumed that the exact function in the element can be written as the linear combination of the shape functions to a fair degree of accuracy. Now, in Eq. (4.61), if w is arbitrary, it will be ensured that residual R^e is zero everywhere. Just as we assumed that the exact function may be approximated as a linear combination of shape functions, we assume that the arbitrary weight function may also be approximated as a linear combination of shape functions

$$w = N_1 w_1 + N_2 w_2 + \cdots + N_n w_n = \lfloor w_1 \;\; w_2 \ldots w_n \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ . \\ . \\ N_n \end{Bmatrix} = \lfloor w^{ne} \rfloor \{N\} \tag{4.66}$$

Therefore

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \lfloor w^{ne} \rfloor \{N_{,x}\} \tag{4.67}$$

Substituting Eqs (4.26), (4.33), (4.66), and (4.67) in Eq. (4.62), we obtain

$$\lfloor w^{ne} \rfloor \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=0} \\ 0 \\ 0 \\ . \\ . \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=h} \end{Bmatrix} - \int_0^h \lfloor w^{ne} \rfloor \{N, x\} \lfloor N, x \rfloor \{u^{ne}\} \mathrm{d}x + \int_0^h \lfloor w^{ne} \rfloor \{N\} \lfloor N \rfloor \{u^{ne}\} \mathrm{d}x$$

$$- \int_0^h \lfloor w^{ne} \rfloor \{N\} \mathrm{d}x = 0 \tag{4.68}$$

As the nodal values of the weights are arbitrary and may be non-zero, Eq. (4.68) implies

$$\left[\int_0^h \{N_{,x}\} \lfloor N_{,x} \rfloor \mathrm{d}x - \int_0^h \{N\} \lfloor N \rfloor \mathrm{d}x \right] \{u^{ne}\} = - \int_0^h \{N\} \, \mathrm{d}x + \begin{Bmatrix} -\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=0} \\ 0 \\ 0 \\ . \\ . \\ \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_{x=h} \end{Bmatrix} \tag{4.69}$$

This is the elemental equation, which is the same as Eq. (4.40) obtained by Ritz method. (We again draw attention of the reader to the fact that derivatives at $x = 0$ and $x = h$ are the derivatives at node 1 and node n, respectively. The rest of the procedure is similar to that described in the previous section.)

If the variational form of a differential equation exists, then both the methods provide the same results for the same approximation. However, for many differential equations, the variational form does not exist. For example, the differential equation

$$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 7y = 0 \tag{4.70}$$

does not have a variational form. For solving this equation by FEM, we can use Galerkin procedure but not Ritz procedure.

4.6 CONCLUSION

In this chapter, two approaches of finite element formulation have been described—Ritz FEM and Galerkin FEM. If the variational form is given, it is convenient to apply Ritz FEM. On the other hand, if the differential equation is given, it is convenient to apply Galerkin FEM. One can convert the problem from a variational form to differential form and vice versa and use any of the two methods. However, for certain differential equations, the variational form does not exist at all. Applying Galerkin method to these problems is more convenient.

SOLVED PROBLEMS

4.1 Solve the following differential equation using FEM:

$$\frac{d^2y}{dx^2} + x = 0, \quad 0 \le x \le 3$$

with the boundary conditions: $y = 0$ at $x = 0$ and $x = 3$. Discretize the domain into three equal elements.

Solution We shall solve it by Galerkin FEM. We shall use the x symbol for local coordinates and X for global coordinate. The differential equation in global coordinates is given as

$$\frac{d^2y}{dX^2} + X = 0, \quad 0 \le X \le 3 \tag{4.1.1}$$

We discretize the domain into three elements as shown in Fig. 4.2. The global coordinates of the nodes are

$$X_1 = 0, \quad X_2 = 1, \quad X_3 = 2, \quad X_4 = 3$$

For a typical element with first node as i, the relation between the local and global coordinates is

$$X = X_i + x, \text{ and therefore, } dX = dx \tag{4.1.2}$$

which provides the local coordinate of ith node as 0. Substituting Eq. (4.1.2) in Eq. (4.1.1), we obtain the differential equation in local coordinates

$$\frac{d^2 y}{dx^2} + (X_i + x) = 0, \quad 0 \le X \le h$$

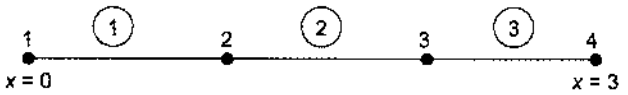


Figure 4.2 *Discretization of a domain into three elements*

Now, let the approximate solution in the element be given by function y^e. The residual inside the element is given by

$$R^e = \frac{d^2 y^e}{dx^2} + (X_i + x)$$

We have to make the weighted integral of the residual zero:

$$\int_0^h wR^e dx = \int_0^h w\left(\frac{d^2 y^e}{dx^2} + X_i + x\right) dx = 0$$

We integrate the first term by parts in the above integrand in order to evenly distribute the order of derivative on y^e and w. Thus

$$w\frac{dy^e}{dx}\Bigg|_0^h - \int_0^h \frac{dw}{dx}\frac{dy^e}{dx} dx + \int_0^h wX_i\, dx + \int_0^h wx\, dx = 0 \tag{4.1.3}$$

Observing the weak form given in Eq. (4.1.3), it is concluded that a two-noded element is good enough. For an element

$$y^e = \begin{bmatrix} 1 - \frac{x}{h} & \frac{x}{h} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix}$$

Thus

$$\frac{dy^e}{dx} = \begin{bmatrix} -\frac{1}{h} & \frac{1}{h} \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix}$$

In Galerkin method, the weight function is approximated in the same manner as the primary variable. Thus

$$w = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{Bmatrix} 1 - \frac{x}{h} \\ \frac{x}{h} \end{Bmatrix} \quad \text{and} \quad \frac{dw}{dx} = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{Bmatrix} -\frac{1}{h} \\ \frac{1}{h} \end{Bmatrix}$$

We substitute these approximations for y^e and w in Eq. (4.1.3); however, no approximation for the derivative of y^e is made at the boundaries. As a result, we obtain

$$\lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -\left(\dfrac{dy}{dx}\right)_1 \\ \left(\dfrac{dy}{dx}\right)_2 \end{Bmatrix} - \int_0^h \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -\dfrac{1}{h} \\ \dfrac{1}{h} \end{Bmatrix} \begin{bmatrix} -\dfrac{1}{h} & \dfrac{1}{h} \end{bmatrix} dx +$$

$$\int_0^h \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -\dfrac{1}{h} \\ \dfrac{1}{h} \end{Bmatrix} X_i \, dx + \int_0^h \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -\dfrac{1}{h} \\ \dfrac{1}{h} \end{Bmatrix} x \, dx = 0 \tag{4.1.4}$$

As the nodal weights w_1 and w_2 are arbitrary, Eq. (4.1.4) implies that

$$\begin{Bmatrix} -\left(\dfrac{dy}{dx}\right)_1 \\ \left(\dfrac{dy}{dx}\right)_2 \end{Bmatrix} - \int_0^h \frac{1}{h^2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} dx + \int_0^h X_i \begin{Bmatrix} 1 - \dfrac{x}{h} \\ \dfrac{x}{h} \end{Bmatrix} dx + \int_0^h \begin{Bmatrix} x - \dfrac{x^2}{h} \\ \dfrac{x^2}{h} \end{Bmatrix} dx = 0 \tag{4.1.5}$$

Simplifying Eq. (4.1.5), we obtain the following elemental equations:

$$\frac{1}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = \begin{Bmatrix} \dfrac{X_i h}{2} \\ \dfrac{X_i h}{2} \end{Bmatrix} + \begin{Bmatrix} \dfrac{h^2}{6} \\ \dfrac{h^2}{3} \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{dy}{dx}\right)_1 \\ \left(\dfrac{dy}{dx}\right)_2 \end{Bmatrix} \tag{4.1.6}$$

We now discretize the domain into three equal elements as shown in Fig. 4.2. For each element, $h = 1$. For element 1: $i = 1$, $X_1 = 0$. Hence

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = \begin{Bmatrix} \dfrac{1}{6} \\ \dfrac{1}{3} \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{dy}{dx}\right)_1 \\ \left(\dfrac{dy}{dx}\right)_2 \end{Bmatrix} \tag{4.1.7}$$

For element 2: $i = 2$, $X_2 = 0$. Hence

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} y_2 \\ y_3 \end{Bmatrix} = \begin{Bmatrix} \dfrac{1}{2} \\ \dfrac{1}{2} \end{Bmatrix} + \begin{Bmatrix} \dfrac{1}{6} \\ \dfrac{1}{3} \end{Bmatrix} + \begin{Bmatrix} -\left(\dfrac{dy}{dx}\right)_2 \\ \left(\dfrac{dy}{dx}\right)_3 \end{Bmatrix} \tag{4.1.8}$$

For element 3: $i = 3$, $X_3 = 0$. Hence

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} y_3 \\ y_4 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} + \begin{Bmatrix} \frac{1}{6} \\ \frac{1}{3} \end{Bmatrix} + \begin{Bmatrix} -\left(\frac{dy}{dx}\right)_3 \\ \left(\frac{dy}{dx}\right)_4 \end{Bmatrix} \tag{4.1.9}$$

Assembling the elemental equations [Eqs (4.1.7) to (4.1.9)], we obtain

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{Bmatrix} = \begin{Bmatrix} \frac{1}{6} + \frac{1}{2} + \frac{1}{6} \\ \frac{1}{3} + \frac{1}{2} + \frac{1}{6} \\ \frac{1}{2} + \frac{1}{3} + 1 + \frac{1}{6} \\ 1 + \frac{1}{3} \end{Bmatrix} + \begin{Bmatrix} -\left(\frac{dy}{dx}\right)_1 \\ 0 \\ 0 \\ \left(\frac{dy}{dx}\right)_4 \end{Bmatrix} \tag{4.1.10}$$

Now, we apply the boundary conditions: $y = 0$ at $x = 0$ and $x = 3$, which implies that $y_1 = y_4 = 0$. Hence, from the system of equations given by Eq. (4.1.10), we eliminate the first and fourth rows and columns to get the following system of equations:

$$\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} \tag{4.1.11}$$

The solution of the system of equations given by Eq. (4.1.11) is $y_2 = \frac{4}{3}$ and $y_3 = \frac{5}{3}$. Thus, all the nodal values are obtained. Then, the value of y at any point between the nodes may be found by the linear interpolation of nodal values. Thus, the complete solution is

$$y = \begin{cases} \frac{4x}{3} & \text{for } 0 \le x \le 1 \\ \frac{x}{3} + 1 & \text{for } 1 \le x \le 2 \\ \frac{-5x}{3} + 5 & \text{for } 2 \le x \le 3 \end{cases}$$

The exact solution of the differential equation is

$$y = \frac{3x}{2} - \frac{x^3}{6}$$

The readers may verify that in this problem FEM provides exact solution at the nodes, although both solutions are different. In general, it is expected that FEM solution is more accurate at nodes than at other points.

4.2 The x-coordinates of a four-noded element are 1, 2, 3, and 4, respectively. Find out the shape function N_3 corresponding to the third node with coordinates $x = 3$.

Solution Given $x_1 = 1, x_2 = 2, x_3 = 3, x_4 = 4$.

Using Eq. (4.17), we obtain

$$N_3 = \frac{(x_1 - x)(x_2 - x)(x_4 - x)}{(x_1 - x_3)(x_2 - x_3)(x_4 - x_3)} = \frac{(1-x)(2-x)(4-x)}{(1-3)(2-3)(4-3)} = \frac{1}{2}(1-x)(2-x)(4-x)$$

We can verify that this shape function is 0 at nodes 1, 2, and 4. Its value is 1 at node 4.

EXERCISES

A. Multiple Choice Questions

4.1 Consider the function shown in Fig. 4.3.

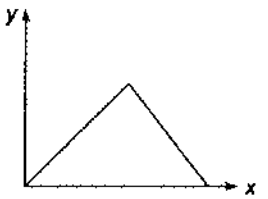


Figure 4.3 *Function for Exercise 4.1*

This function is

(a) C^0 continuous
(b) C^1 continuous
(c) C^2 continuous
(d) C^3 continuous

4.2 A C^2 continuous function ensures

(a) Continuity of slope but not of displacement
(b) Continuity of slope but not of curvature
(c) Continuity of displacement but not of slope
(d) Continuity of slope as well as curvature

4.3 For the differential equation, $\frac{d^4 y}{dx^4} + 1 = 0$, a complete polynomial that can be used as an interpolation function is

(a) $a + bx$
(b) $a + bx + cx^2$
(c) $a + bx + cx^2 + dx^3$
(d) $a + bx + cx^2 + dx^4$

4.4 Consider the following system of equations:

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} -\left(\dfrac{du}{dx}\right)_0 \\ 6 \\ 6 \\ 3 \end{Bmatrix}$$

If $u_1 = 1$ and it is decided to eliminate row and columns, the right hand side vector will be

(a) $\{7 \;\; 6 \;\; 3\}^T$ (b) $\{7 \;\; 7 \;\; 4\}^T$

(c) $\{7 \;\; 7 \;\; 3\}^T$ (d) $\{6 \;\; 6 \;\; 3\}^T$

4.5 For the differential equation

$$\frac{dy}{dx} = 1$$

The Galerkin FEM yields the following equations for an element of length h:

(a) $\int_0^h \lfloor N_{,x} \rfloor \{N_{,x}\} dx \{u^{ne}\} = \int_0^h \{N\} dx$ (b) $\int_0^h \{N_{,x}\} \lfloor N_{,x} \rfloor dx \{u^{ne}\} = \int_0^h \{N\} dx$

(c) $\int_0^h \{N\} \lfloor N_{,x} \rfloor dx \{u^{ne}\} = \int_0^h \{N\} dx$ (d) $\int_0^h \{N_{,x}\} \lfloor N \rfloor dx \{u^{ne}\} = \int_0^h \{N\} dx$

B. Problems

4.6 Solve the following boundary value problem using Ritz and Galerkin FEM:

Differential equation: $\dfrac{d^2y}{dx^2} + 5xy = 7$

Boundary conditions: at $x = 0$, $y = 0$ and at $x = 5$, $y = 6$.

Solve first by taking two elements and then three elements and compare the results.

4.7 Solve the following boundary value problem using Galerkin FEM:

$$\frac{d^2y}{dx^2} + 5\frac{dy}{dx} + 7y = 0$$

The boundary conditions are at $x = 0$, $y = 0$ and at $x = 1$, $dy/dx = 0$. Solve by taking two and three elements.

4.8 Solve the following problem using Ritz FEM:

$$\text{Minimize } I = \int_0^1 5\left(\frac{du}{dx}\right)^2 dx - 10u(1); \quad u(0) = 0$$

Start by taking one element and keep on increasing the elements till convergence is achieved.

4.9 Solve the following problem using Ritz FEM:

$$\text{Minimize } I = \int_0^1 \left(\left(\frac{du}{dx}\right)^2 - 5u\right) dx; \quad u(0) = 0$$

Start by taking one element and keep on increasing the elements till the convergence is achieved.

4.10 At a point, the values of the shape functions of a four-noded C^0 continuity element are: $N_1 = 0.1$, $N_2 = N_3 = 0.2$. Find N_4.

4.11 Solve Problem 4.9 by considering a three-noded element.

4.12 Solve the following problem using Ritz and Galerkin FEM:

$$\text{Differential equation: } \frac{d^2y}{dx^2} + 5\delta(x - 0.5) = 0 \qquad 0 \le x \le 1$$

$$\text{Boundary conditions: At } x = 0,\ y = 0 \text{ and at } x = 1,\ \frac{dy}{dx} = 0$$

Take 2 two-noded elements. Note that $\delta(x - 0.5)$ is a Dirac-delta function, which is 0 everywhere except at $x = 0.5$.

REFERENCES

Li, BQ, 2006, *Discontinuous Finite Element in Fluid Dynamics and Heat Transfer*, Springer-Verlag, London.

Oliveira, ERDA, 1968, 'Theoretical foundations of the finite element method', *Int. J. Solids Struct.*, Vol. 4, pp. 929–952.

Pian, THH, and Tong, P, 1967, 'The convergence of finite element method in solving linear elastic problems', *Int. J. Solids Struct.*, Vol. 3, pp. 865–880.

Zienkiewicz, OC, 1979, *The Finite Element Method, Third edition*, Tata McGraw-Hill, New Delhi.

Zienkiewicz, OC, Taylor, RL, Sherwin, SJ, and Piero, J, 2003, 'On discontinuous Galerkin methods', *Int. J. Numer. Meth. Engg*, Vol. 58, pp. 1119–1148.

5

CHAPTER

One-Dimensional FEM Formulations

5.1 Introduction

5.2 Steady-State Heat Conduction

5.3 Longitudinal Deformation of a Rod

5.4 Fluid Flow Problem

5.5 Conclusion

5.1 INTRODUCTION

In Chapter 4, the techniques of FEM formulation using Ritz and Galerkin methods were discussed in detail. In this chapter, we will consider some one-dimensional C^0 continuity problems that are of particular interest to engineers and will obtain the FEM formulations for them. The problems discussed in this chapter include specific problems from areas of heat transfer, solid mechanics, and fluid mechanics. Before applying the FEM procedure, the governing equations for these problems have to be developed. Each problem has been described in the following sections.

5.2 STEADY-STATE HEAT CONDUCTION

Suppose we have to find out the temperature distribution of the rod shown in Fig. 5.1. If the area of cross-section of the rod is small compared to its length, then the temperature across a particular cross-section can be taken as constant. Thus, temperature can be considered to be a function of the longitudinal coordinate x. The area of cross-section and thermal conductivity of the rod may also be considered as functions of x.

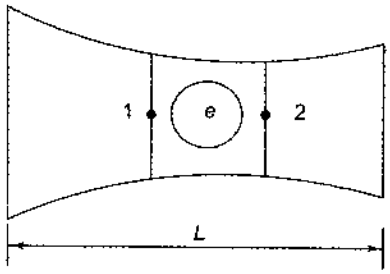


Figure 5.1 *A heat-conducting rod showing a typical finite element*

The governing law for heat-conduction problems is the Fourier heat-conduction equation, which is given as follows

$$q = -k\frac{\partial T}{\partial x} \tag{5.1}$$

where q is the heat flow per unit area (heat flux) in the direction of x, T is the temperature, and k is the thermal conductivity. In general, T may be a function of the length coordinate x as well as of time t; therefore, in Eq. (5.1), a partial derivative of T with respect to x has been considered. There can be different types of the boundary conditions for heat-conduction problems, which will be discussed later. For obtaining the differential equation governing this problem, we consider an infinitesimally small element of length 'dx' as shown in Fig. 5.2 and study the heat balance in the element. In steady-state, the heat generated will be equal to the net heat coming out of the rod. If $\dot{Q}$ denotes the rate of heat generation per unit volume, then the rate of heat generation in the element dx is given by $\dot{Q}A\,dx$,

where A is the average area of cross-section of the rod. The heat entering the rod per unit time from the left is equal to Aq. We can also say that the rate of heat coming out of the rod through the left hand side cross-section of the element is $-Aq$. The rate of heat coming out of the element from the right-hand side cross-section is equal to $Aq + \mathrm{d}(Aq)$, where $\mathrm{d}(Aq)$ denotes the differential increase in the heat flow. The convective heat transfer from the surface is given by $h_c(T - T_f)p\mathrm{d}x$, where h_c is the convective heat transfer coefficient, T_f is the temperature of the surrounding fluid, and p is the average perimeter of the cross-section of the infinitesimal element.

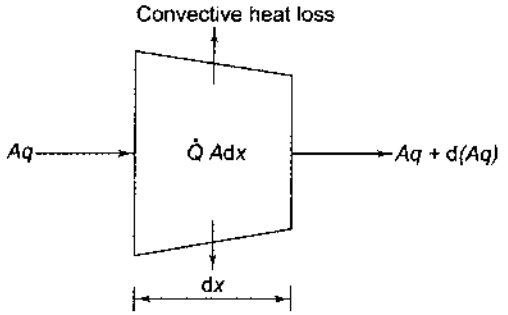


Figure 5.2 *Heat transfer from an infinitesimal element*

Thus, the heat balance in the element would be given by

$$Aq + \mathrm{d}(Aq) - Aq - h_c(T_f - T)p\mathrm{d}x = \dot{Q}A\mathrm{d}x \tag{5.2}$$

From Eq. (5.2), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}x}(Aq) - \dot{Q}A - h_c(T_f - T)p = 0 \tag{5.3}$$

Substituting the value of q from Eq. (5.1), we can write Eq. (5.3) as

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(-kA\frac{\mathrm{d}T}{\mathrm{d}x}\right) - \dot{Q}A - h_c(T_f - T)p = 0 \tag{5.4}$$

Thus, the governing differential equation of the system becomes

$$\frac{\mathrm{d}}{\mathrm{d}x}(AkT_{,x}) - \dot{Q}A - h_c(T_f - T)p = 0 \tag{5.5}$$

where the subscript x in T denotes differentiation with respect to x.

Now, we can proceed easily towards Galerkin FEM formulation. However, we shall first carry out Ritz FEM formulation. For this, we first need to obtain the variational form. The variational form of this problem, considering an element of length h, is given as

$$\mathrm{P} = \int_0^h \left\{\frac{1}{2}AkT_{,x}^2 + \frac{1}{2}h_c pT^2 - (\dot{Q}AT + h_c pT_f)\right\}\mathrm{d}x - kA\frac{\mathrm{d}T}{\mathrm{d}x}T\Big|_0^h \tag{5.6}$$

The reader can easily obtain this form by following the procedure described in Chapter 2. Taking into account both completeness and compatibility, a two-noded element with Lagrange shape functions would have sufficed; however, we shall carry out the formulation for a general n-noded element. Substituting the following equations:

$$T = \lfloor N \rfloor \{T^{ne}\} \tag{5.7}$$

$$T_{,x} = \left[N_{,x}\right]\{T^{ne}\} \tag{5.8}$$

in the variational form, we obtain

$$\Pi^e = \frac{1}{2}\{T^{ne}\}^{\mathrm{T}}\int_0^h Ak\left\{N_{,x}\right\}\left\lfloor N_{,x}\right\rfloor \mathrm{d}x\{T^{ne}\} + \frac{1}{2}\{T_c\}^{\mathrm{T}}\int_0^h h_c p\left\{N\right\}\left\lfloor N\right\rfloor \mathrm{d}x\{T^{ne}\}$$

$$-\{T^{ne}\}^{\mathrm{T}}\int_0^h \dot{Q}A\left\{N\right\}\mathrm{d}x - \{T^{ne}\}^{\mathrm{T}}\int_0^h h_c pT_f\{N\}\mathrm{d}x + \lfloor T_1\ T_2 \ldots T_n \rfloor \begin{Bmatrix} \left(kA\dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_{x=0} \\ 0 \\ \vdots \\ 0 \\ -\left(kA\dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_{x=h} \end{Bmatrix} \tag{5.9}$$

Note that in Eq. (5.9) we have expressed the last term of Eq. (5.6) in terms of the product of two vectors: a row vector of order $1 \times n$ and a column vector of order $n \times 1$. All the entries (or elements) in the column vector, except the first and the last term, are 0. Minimizing Eq. (5.9) with respect to $\{T^{ne}\}$, we get

$$0 = \int_0^h Ak\left\{N_{,x}\right\}\left\lfloor N_{,x}\right\rfloor \mathrm{d}x\{T^{ne}\} + \int_0^h h_c p\left\{N\right\}\left\lfloor N\right\rfloor \mathrm{d}x\{T^{ne}\}$$

$$-\int_0^h \dot{Q}A\{N\}\mathrm{d}x - \int_0^h h_c pT_f\{N\}\mathrm{d}x + \begin{Bmatrix} \left(kA\dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ 0 \\ \vdots \\ -\left(kA\dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{5.10}$$

Equation (5.10) provides the elemental equation. For ease of representation, we define the following matrix and column vectors:

$$[k^e] = \int_0^h Ak\{N_{,x}\}\left\lfloor N_{,x}\right\rfloor \mathrm{d}x + \int_0^h h_c p\{N\}\left\lfloor N\right\rfloor \mathrm{d}x \tag{5.11a}$$

$$[R^e] = \int_0^h \dot{Q}A\{N\}\mathrm{d}x + \int_0^h h_c pT_f\{N\}\mathrm{d}x \tag{5.11b}$$

$$[R_i^e] = \left\lfloor \left(kA\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1\ 0\ \ 0 \ldots 0\ \left(kA\frac{\mathrm{d}T}{\mathrm{d}x}\right)_n \right\rfloor^{\mathrm{T}} \tag{5.11c}$$

where $[k^e]$ is the elemental matrix, $[R^e]$ is the right hand side vector representing the heat generation and convection through the lateral surface, and $[R_i^e]$ is the right hand side vector representing internal flux. Thus, the elemental equations given by Eq. (5.10) get simplified to

$$[k^e]\{T^{ne}\} = [R^e] + [R_i^e] \tag{5.12}$$

The rest of the FEM procedure is explained in the following subsections. The elemental equations given by Eq. (5.12) are valid for a general n-noded element. In the following subsections, we will

- develop expressions for a general two-noded element,
- illustrate the assembly procedure for a specific domain discretized with two elements,
- describe the procedure for applying different types of boundary conditions,
- briefly discuss the method used for solving a large system of equations.

5.2.1 Elemental Equations for a Two-noded Element

For a two-noded element with the first node at $x = 0$ and the second node at $x = h$, we have

$$N_1 = 1 - \frac{x}{h},\ N_2 = \frac{x}{h},\ N_{1,x} = -\frac{1}{h},\ N_{2,x} = \frac{1}{h} \tag{5.13}$$

Therefore, the elemental matrix is written as

$$[k^e] = \int_0^h Ak \begin{Bmatrix} -\frac{1}{h} \\ \frac{1}{h} \end{Bmatrix} \begin{bmatrix} -\frac{1}{h} & \frac{1}{h} \end{bmatrix} \mathrm{d}x + \int_0^h h_c p \begin{Bmatrix} 1 - \frac{x}{h} \\ \frac{x}{h} \end{Bmatrix} \mathrm{d}x$$

$$= \frac{Ak}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \frac{h_c p}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{5.14}$$

We assume that A, k, h_c, and p are constant. The elemental right hand side vector representing the heat generation and convection through lateral surface is given by

$$[R^e] = \int_0^h \dot{Q}A \begin{Bmatrix} 1 - \frac{x}{h} \\ \frac{x}{h} \end{Bmatrix} \mathrm{d}x + \int_0^h h_c p T_f \begin{Bmatrix} 1 - \frac{x}{h} \\ \frac{x}{h} \end{Bmatrix} \mathrm{d}x$$

$$= \begin{Bmatrix} \frac{\dot{Q}Ah}{2} \\ \frac{\dot{Q}Ah}{2} \end{Bmatrix} + \begin{Bmatrix} \frac{h_c p T_f}{2} \\ \frac{h_c p T_f}{2} \end{Bmatrix} \tag{5.15}$$

The elemental right hand side vector representing the internal heat flux is given as

$$[R_i^e] = \begin{Bmatrix} -\left(kA\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ \left(kA\frac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{5.16}$$

5.2.2 Assembly Procedure for Two Elements

For better understanding of the assembly procedure, let us consider a stepped rod that has been discretized into two elements (Fig. 5.3). Element 1 corresponds to the larger area of cross-section, while element 2 corresponds to the smaller area of cross-section. In general, at a point where the cross-section changes abruptly, the element must also change. If there is heat generation in the rod and/or thermal conductivity is changing along the length, then to obtain reasonably accurate results, we can adopt one of the two ways: either take more number of elements or take higher noded elements. However, if there is no heat generation and the thermal conductivity does not change within a step of the rod, the discretization with two 2-noded elements will give an accurate solution. This is because the actual solution is piecewise linear for this case.

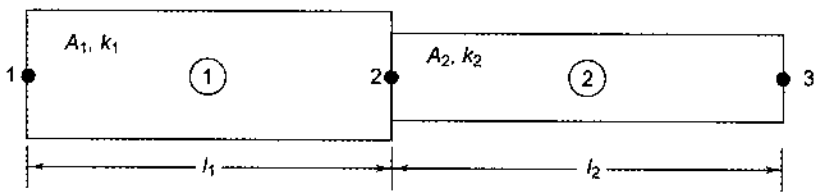


Figure 5.3 *A stepped rod discretized into two elements*

Here, we assume that the thermal conductivity is denoted by k, the cross-sectional areas of the two steps are denoted by A_1 and A_2, and there is no heat generation. We also assume that there is no heat transfer from the lateral surface. The elemental equations for the two elements are then written as

$$\frac{A_1 k}{l_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1 \dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ \left(kA_1 \dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{5.17a}$$

$$\frac{A_2 k}{l_1}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_2 \dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \\ \left(kA_2 \dfrac{\mathrm{d}T}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{5.17b}$$

Next, we need to assemble the two systems of equations. For this purpose, we make a format for a system of three equations and place Eq. (5.17a) in the block created by rows 1 and 2 and columns 1 and 2. Then, we append Eq. (5.17b) in the block created by rows 2 and 3 and columns 2 and 3. As a result, we obtain the following system of equations:

$$\begin{bmatrix} \frac{A_1k}{l_1} & -\frac{A_1k}{l_1} & 0 \\ -\frac{A_1k}{l_1} & \frac{A_1k}{l_1}+\frac{A_2k}{l_1} & -\frac{A_2k}{l_2} \\ 0 & -\frac{A_2k}{l_2} & \frac{A_2k}{l_2} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ \left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_2 - \left(kA_2\frac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \\ \left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{5.18}$$

We make use of the fact that at the interface of the two elements, heat coming out of one element is equal to heat going into the other element. This will hold true irrespective of the rate of heat generation and changes in thermal conductivities. In this problem, the first derivatives of temperature and heat flux are discontinuous at the interface, while the heat flow is continuous. Thus,

$$\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_2 = \left(kA_2\frac{\mathrm{d}T}{\mathrm{d}x}\right)_2 \tag{5.19}$$

Therefore, the assembled system of equations now becomes

$$\begin{bmatrix} \frac{A_1k}{l_1} & -\frac{A_1k}{l_1} & 0 \\ -\frac{A_1k}{l_1} & \frac{A_1k}{l_1}+\frac{A_2k}{l_1} & -\frac{A_2k}{l_2} \\ 0 & -\frac{A_2k}{l_2} & \frac{A_2k}{l_2} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ 0 \\ \left(kA_2\frac{\mathrm{d}T}{\mathrm{d}x}\right)_3 \end{Bmatrix} \tag{5.20}$$

5.2.3 Application of Boundary Conditions

In heat transfer problems, we usually encounter the following boundary conditions:

(i) **Dirichlet boundary condition** In this boundary condition, the temperature is prescribed on the boundary, that is, $T = T^*$. In one-dimensional problems, T^* will be the prescribed value of temperature, while in multi-dimensional problems, it will be a prescribed function. This boundary condition is an essential boundary condition.

(ii) **Neumann boundary condition** In this boundary condition, the heat flux is prescribed on the boundary, that is, $q_n = q^*$. In one-dimensional problems, the q^* will be the prescribed value of heat flux, while in multi-dimensional problems, it will be a prescribed function. This boundary condition is a natural boundary condition.

(iii) **Robin boundary condition** In this boundary condition, heat flux at the boundary is equated to the convective heat flux, that is, $q_n = h_c\,(T - T_f)$, where h_c is the surface convective heat transfer coefficient and T_f is the temperature of the surrounding medium. If we develop the variational form of the problem, this boundary condition gets incorporated naturally into it. Hence, it can be treated as a natural boundary condition. Note that the presence of these types of boundary

conditions is sufficient for finding out the temperature distribution. Therefore, in that sense, it can be called an essential boundary condition.

It is convenient to apply the natural boundary conditions first and the essential boundary conditions later, because natural boundary conditions require modifications only in the right hand side vector, whereas essential boundary conditions require modifications of the elemental matrix as well as the right hand side vector. Suppose, the heat flux (heat flow per unit area out) through the right hand side is given by q^*: it implies that

$$-\left(kA_2\frac{\mathrm{d}T}{\mathrm{d}x}\right)_3 = q^* A_2 \tag{5.21}$$

Therefore, Eq. (5.20) gets modified as

$$\begin{bmatrix} \frac{A_1k}{l_1} & -\frac{A_1k}{l_1} & 0 \\ -\frac{A_1k}{l_1} & \frac{A_1k}{l_1}+\frac{A_2k}{l_1} & -\frac{A_2k}{l_2} \\ 0 & -\frac{A_2k}{l_2} & \frac{A_2k}{l_2} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ 0 \\ -q^*\ A_2 \end{Bmatrix} \tag{5.22}$$

If we apply Robin boundary condition to the right hand side, we obtain the following equation

$$q_3A_2 = -kA_2\left(\frac{\mathrm{d}T}{\mathrm{d}x}\right)_3 = h_cA_2\left(T_3 - T_f\right) \tag{5.23}$$

Thus, Eq. (5.20) gets modified as

$$\begin{bmatrix} \frac{A_1k}{l_1} & -\frac{A_1k}{l_1} & 0 \\ -\frac{A_1k}{l_1} & \frac{A_1k}{l_1}+\frac{A_2k}{l_1} & -\frac{A_2k}{l_2} \\ 0 & -\frac{A_2k}{l_2} & \frac{A_2k}{l_2} \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ 0 \\ -h_cA_2\left(T_3 - T_f\right) \end{Bmatrix} \tag{5.24}$$

Note that Eq. (5.24) has a term containing the unknown, T_3, on the right hand side. It can be taken to the left hand side to modify Eq. (5.24) which provides the following system of equations:

$$\begin{bmatrix} \frac{A_1k}{l_1} & -\frac{A_1k}{l_1} & 0 \\ -\frac{A_1k}{l_1} & \frac{A_1k}{l_1}+\frac{A_2k}{l_1} & -\frac{A_2k}{l_2} \\ 0 & -\frac{A_2k}{l_2} & \frac{A_2k}{l_2}+h_cA_2 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} -\left(kA_1\frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 \\ 0 \\ h_cA_2T_f \end{Bmatrix} \tag{5.25}$$

Now, suppose there is an essential boundary condition, such that, $T_1 = T_1^*$. This boundary condition can be applied in one of the following ways:

(i) We can replace the first equation in the system of equations given by Eq. (5.25) by the equation describing the essential boundary condition. Thus, Eq. (5.25) gets modified to

$$\begin{bmatrix} 1 & 0 & 0 \\ -\dfrac{A_1 k}{l_1} & \dfrac{A_1 k}{l_1} + \dfrac{A_2 k}{l_1} & -\dfrac{A_2 k}{l_2} \\ 0 & -\dfrac{A_2 k}{l_2} & \dfrac{A_2 k}{l_2} + h_c A_2 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} T_1^* \\ 0 \\ h_c A_2 T_f \end{Bmatrix} \tag{5.26}$$

Although the system of equations given by Eq. (5.26) can be solved, the coefficient matrix has become unsymmetrical. The disadvantage in such a situation is that the special solvers for equations with symmetric coefficient matrix that take lesser computational time compared to a system of equations with a non-symmetric coefficient matrix cannot be used. The coefficient matrix given in Eq. (5.26) can be made symmetric by transferring the product of coefficients and the known temperature in each row from the left hand side to the right hand side. Thus, Eq. (5.26) can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \dfrac{A_1 k}{l_1} + \dfrac{A_2 k}{l_1} & -\dfrac{A_2 k}{l_2} \\ 0 & -\dfrac{A_2 k}{l_2} & \dfrac{A_2 k}{l_2} + h_c A_2 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} T_1^* \\ \dfrac{A_1 k}{l_1} T_1^* \\ h_c A_2 T_f \end{Bmatrix} \tag{5.27}$$

(ii) One can eliminate the first row and first column and obtain the following system of two equations:

$$\begin{bmatrix} \dfrac{A_1 k}{l_1} + \dfrac{A_2 k}{l_1} & -\dfrac{A_2 k}{l_2} \\ -\dfrac{A_2 k}{l_2} & \dfrac{A_2 k}{l_2} + h_c A_2 \end{bmatrix} \begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} \dfrac{A_1 k}{l_1} T_1^* \\ h_c A_2 T_f \end{Bmatrix} \tag{5.28}$$

Note that the right hand side vector has been modified by the method discussed in Chapter 4.

(iii) We can also apply the *penalty method* to the given problem. In this method, the equation describing the essential boundary condition is multiplied by a large penalty number M and is then added to the equation containing the unknown right hand side. In the present problem, after multiplying the essential boundary condition with M, we obtain

$$MT_1 = MT_1^* \tag{5.29}$$

which is then added to the first equation of Eq. (5.25) because it contains the unknown right hand side term. The first equation then becomes

$$\left(\frac{A_1 k}{l_1} + M\right) T_1 - \frac{A_1 k}{l_1} T_2 = -\left(kA_1 \frac{\mathrm{d}T}{\mathrm{d}x}\right)_1 + MT_1^* \tag{5.30}$$

As M is a very large number (say 10^6 times the maximum element of the first row), Eq. (5.30) can be written as

$$MT_1 - \frac{A_1 k}{l_1} T_2 \approx MT_1^* \tag{5.31}$$

Thus, the system of equation becomes

$$\begin{bmatrix} M & -\dfrac{A_1 k}{l_1} & 0 \\ -\dfrac{A_1 k}{l_1} & \dfrac{A_1 k}{l_1} + \dfrac{A_2 k}{l_1} & -\dfrac{A_2 k}{l_2} \\ 0 & -\dfrac{A_2 k}{l_2} & \dfrac{A_2 k}{l_2} + h_c A_2 \end{bmatrix} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} MT_1^* \\ 0 \\ h_c A_2 T_f \end{Bmatrix} \tag{5.32}$$

This system retains the symmetry of the coefficient matrix. The implementation of this method is very simple. Corresponding to the coefficient of prescribed variable, a large number is put in the coefficient matrix and the product of the large number and the prescribed value is put in the right hand side.

5.2.4 Solving the System of Equations

Equation (5.32) can be solved manually. However, a larger system of equations will have to be solved using a computer. There are many methods of solving a system of equations. One of the most popular methods is the *Gauss elimination method* with *pivoting*. In pivoting, the equations are arranged in such a manner that they have a larger value of coefficient matrix element at the diagonal. The Gauss elimination method then converts the coefficient matrix into an upper triangular form by using suitable row operations. The values of the variables are then found by back substitution. For highly ill-conditioned matrices, Gauss elimination method is not suitable. In such cases, the Householder method may be used.

5.2.5 Heat-Conduction Problem in Polar Coordinates

Consider heat flow in the radial direction in a hollow cylinder having inner radius equal to a and outer radius equal to b (Fig. 5.4). As a practical example we can think of a hot fluid flowing inside a thick pipe. The temperature of the fluid as well as that of the wall can be measured and the temperature distribution in the wall of the pipe can be obtained using FEM.

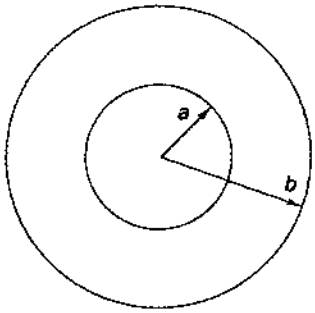


Figure 5.4 *A hollow cylinder*

In polar coordinates, the steady-state heat conduction without heat generation is given by the following equation:

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(kr\frac{\mathrm{d}T}{\mathrm{d}r}\right)=0 \tag{5.33}$$

If we reserve 'r' for local coordinates, Eq. (5.33) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}R}\left(kr\frac{\mathrm{d}T}{\mathrm{d}R}\right)=0 \tag{5.34}$$

We can solve this problem using Galerkin FEM. For this purpose, we make the integral of the weighted residual equal to zero for the jth element (with nodes j and k). Thus

$$\int_0^h w\frac{\mathrm{d}}{\mathrm{d}r}\left(\left(R_j+r\right)k\frac{\mathrm{d}T}{\mathrm{d}r}\right)\mathrm{d}r=0 \tag{5.35}$$

where R_j is the radius of jth node and r is the local coordinate. The integration by parts of Eq. (5.35) yields

$$w(R_j+r)\left(k\frac{\mathrm{d}T}{\mathrm{d}r}\right)\Bigg|_0^h-\int_0^h\frac{\mathrm{d}w}{\mathrm{d}r}(R_j+r)k\frac{\mathrm{d}T}{\mathrm{d}r}\,\mathrm{d}r=0 \tag{5.36}$$

Observing the weak form, we assess that completeness in T, $T_{,r}$ and continuity of T is adequate. Thus, we can approximate the temperature as

$$T=a+br=\begin{bmatrix}1-\dfrac{r}{h} & \dfrac{r}{h}\end{bmatrix}\begin{Bmatrix}T_j\\T_k\end{Bmatrix} \tag{5.37}$$

Therefore

$$\frac{\mathrm{d}T}{\mathrm{d}r}=\begin{bmatrix}-\dfrac{1}{h} & \dfrac{1}{h}\end{bmatrix}\begin{Bmatrix}T_j\\T_k\end{Bmatrix} \tag{5.38}$$

Putting this approximation in the weak form, we obtain

$$\int_0^h k(R_j + r)\ \{N_{,r}\}\left[N_{,r}\right] \mathrm{d}r\ \{T\}^{ne} = \begin{Bmatrix} -R_1 k \left.\dfrac{\mathrm{d}T}{\mathrm{d}r}\right|_1 \\ R_2 k \left.\dfrac{\mathrm{d}T}{\mathrm{d}r}\right|_2 \end{Bmatrix} \tag{5.39}$$

or

$$k\begin{bmatrix} \dfrac{1}{h^2} & -\dfrac{1}{h^2} \\ -\dfrac{1}{h^2} & \dfrac{1}{h^2} \end{bmatrix}\left(R_j + \frac{h}{2}\right)h\begin{Bmatrix} T_j \\ T_k \end{Bmatrix} = \begin{Bmatrix} -R_1 k\left(\dfrac{\mathrm{d}T}{\mathrm{d}r}\right)_1 \\ R_2 k\left(\dfrac{\mathrm{d}T}{\mathrm{d}r}\right)_2 \end{Bmatrix} \tag{5.40}$$

or

$$k\left(\frac{R_j}{h} + \frac{1}{2}\right)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} T_j \\ T_k \end{Bmatrix} = \frac{1}{2\pi}\begin{Bmatrix} H_j \\ -H_k \end{Bmatrix} \tag{5.41}$$

where H_j and H_k denote the heat flow rate at the two nodes, if the length of the cylinder is taken to be unity. Now, assume that there is a cylinder with an inner radius of 0.5 m, outer radius of 1.5 m, and wall thickness of 1 m. If we take two elements of equal lengths, the assembled matrix is written as

$$\begin{bmatrix} 1.5 & -1.5 & 0 \\ -1.5 & 1.5 + 2.5 & -2.5 \\ 0 & -2.5 & 2.5 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \frac{1}{2p}\begin{Bmatrix} H_1 \\ 0 \\ -H_3 \end{Bmatrix} \tag{5.42}$$

Let us impose the boundary conditions:

$$T_1 = T_a^* \quad T_3 = T_b^* \tag{5.43}$$

Application of these boundary conditions eliminates the first and third equations in Eq. (5.42). From the second equation, we obtain

$$-1.57\,T_a^* + 4T_2 - 2.5T_b^* = 0 \tag{5.44}$$

$$T_2 = 0.375T_a^* + 0.625T_b^* \tag{5.45}$$

The exact solution is given by

$$T_{\text{exact}} = T_a^* + \frac{T_b^* - T_a^*}{\ln\left(\dfrac{R_b}{R_a}\right)}\ln\left(\frac{R}{R_a}\right) = T_a^* + \frac{T_b^* - T_a^*}{\ln(3)}\ln\left(\frac{R}{0.5}\right) \tag{5.46}$$

Putting $R = 1.0$ m, we obtain the exact value of temperature at node 2 as

$$(T_2)_{\text{exact}} = 0.369T_a^* + 0.631T_b^* \tag{5.47}$$

Subtracting Eq. (5.47) from Eq. (5.45), the error e between FEM and exact solution is obtained as

$$e = 0.006\,(T_a^* + T_b^*) \tag{5.48}$$

Note that the error is just 0.006 times the sum of inner and outer surface temperatures. The error would have been more if the thermal conductivity were not uniform throughout the cylinder thickness.

5.3 LONGITUDINAL DEFORMATION OF A ROD

Consider a rod loaded with axial force of intensity q as shown in Fig. 5.5. Note that q is the force per unit length and has the unit N/m in SI system. However, q may be a function of x. The area of cross-section A and material properties are also functions of x. The governing differential equation is given by

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right) + q = 0 \tag{5.49}$$

The boundary conditions on the system are as follows

(i) $$\text{At } x = 0, \quad u = 0 \tag{5.50}$$

(ii) $$\text{At } x = L, \ \frac{\mathrm{d}u}{\mathrm{d}x} = 0 \tag{5.51}$$

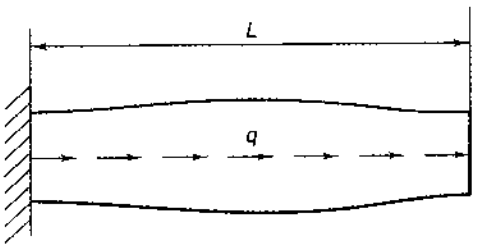


Figure 5.5 *A rod subjected to axial loads*

You may easily derive this differential equation by writing the expression for the total potential energy and applying Euler–Lagrange equations.

Let u^e be the approximate solution of the displacement u for an element e whose end coordinates are given by x_e and x_{e+1}. Taking the weighted residual with respect to the weight function w and obtaining the weak form, we get

$$\int_{x_e}^{x_e+1} w\left(\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u^e}{\mathrm{d}x}\right) + q\right)\mathrm{d}x = \int_{x_e}^{x_e+1}\left[-\frac{\mathrm{d}w}{\mathrm{d}x}\ EA\frac{\mathrm{d}u^e}{\mathrm{d}x} + qw\right]\mathrm{d}x + wEA\frac{\mathrm{d}u^e}{\mathrm{d}x}\bigg|_{x_e}^{x_e+1} = 0 \tag{5.52}$$

The highest order of the derivative appearing in the weak form is 1. Thus, a complete function of the form $a + bx$ is enough. Since the highest order of derivative inside the integral is 1, the function needs to be C^0 continuous everywhere in the domain $(0 < x < L)$. Thus, the value of y at any node

should be the same for both the elements to which the node belongs. A suitable linear approximation for the element is

$$u^e = \lfloor N_1 \quad N_2 \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \tag{5.53}$$

where N_1 and N_2 are called shape functions and u_1 and u_2 are the nodal values of u. Their values can be found from the following expressions:

$$N_1 = \frac{(x - x_{e+1})}{(-x_{e+1} + x_e)} \qquad N_2 = \frac{(x - x_e)}{(x_{e+1} - x_e)} \tag{5.54}$$

Differentiating Eq. (5.54) with respect to x, we obtain

$$\frac{\mathrm{d}u^e}{\mathrm{d}x} = \left\lfloor \frac{\mathrm{d}N_1}{\mathrm{d}x} \quad \frac{\mathrm{d}N_2}{\mathrm{d}x} \right\rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \lfloor N_{1,x} \quad N_{2,x} \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \tag{5.55}$$

A linear weight function is given by

$$w = \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \tag{5.56}$$

where w_1 and w_2 are the weights at the nodes. With these approximations, the weak form given by Eq. (5.52) is approximated as

$$-\int_{x_e}^{x_e+1} \lfloor w_1 \quad w_2 \rfloor EA \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \quad N_{2,x} \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \mathrm{d}x + \int_{x^e}^{x_e+1} q \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x$$

$$+ \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -E_1 A_1 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ E_2 A_2 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} = 0 \tag{5.57}$$

Without the loss of generality, we can take $x^e = 0$ and $x^{e+1} = h^e$, where h^e is the length of the element e. Then Eq. (5.57) becomes

$$\int_0^{h^e} \lfloor w_1 \quad w_2 \rfloor EA \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \quad N_{2,x} \rfloor \begin{Bmatrix} u_1^e \\ u_2^e \end{Bmatrix} \mathrm{d}x - \int_0^{h^e} \lfloor w_1 \quad w_2 \rfloor q \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x = \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -E_1 A_1 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ E_2 A_2 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{5.58}$$

As the weights are arbitrary, Eq. (5.58) becomes

$$\int_0^{h^e} EA \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \quad N_{2,x} \rfloor \mathrm{d}x \begin{Bmatrix} u_1^e \\ u_2^e \end{Bmatrix} = \int_0^{h^e} q \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} -E_1 A_1 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ E_2 A_2 \left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{5.59}$$

This is the finite element elemental equation. For uniform load intensity in an element, Eq. (5.59) is written as

$$\frac{EA}{h^e}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = \begin{Bmatrix} \dfrac{qh^e}{2} \\ \dfrac{qh^e}{2} \end{Bmatrix} + \begin{Bmatrix} -E_1A_1(y_1)' \\ E_2A_2(y_2)' \end{Bmatrix} \tag{5.60}$$

In Eq. (5.60), vectors on the right hand side are called *load vectors* and *internal load vectors*, respectively. It is interesting to note that the total load of qh^e in an element gets uniformly divided between the two nodes.

If there is a concentrated load Q at the second node, the load vector becomes

$$\int_0^{h^e} Q\,d(x - h^e)\begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\mathrm{d}x \tag{5.61}$$

where the concentrated load has been converted into distributed load by using Dirac-delta function. The value of the integral given in Eq. (5.61) is the value of integrand at the second node. Thus, the load vector is

$$Q\begin{Bmatrix} 0 \\ 1 \end{Bmatrix} = \begin{Bmatrix} 0 \\ Q \end{Bmatrix} \tag{5.62}$$

This means that the concentrated loads can be accounted for by adding these loads at the position of node.

5.4 FLUID FLOW PROBLEM

Consider the flow of a viscous and incompressible fluid flowing between two parallel plates (Fig. 5.6). Assume that the flow is fully developed. Because of the symmetry about the mid-plane parallel to the plates, we can consider only the portion between the mid-plane and one plate. The flow is parallel to the plates (along x-direction). The differential equation governing this problem is given by

$$\mu\frac{\mathrm{d}^2u}{\mathrm{d}y^2} = \frac{\mathrm{d}p}{\mathrm{d}x} \tag{5.63}$$

where u is the velocity, μ is the viscosity, and p is the pressure. If at a particular cross-section, the pressure gradient $\mathrm{d}p/\mathrm{d}x$ is specified, we can find out the velocity distribution along the y direction.

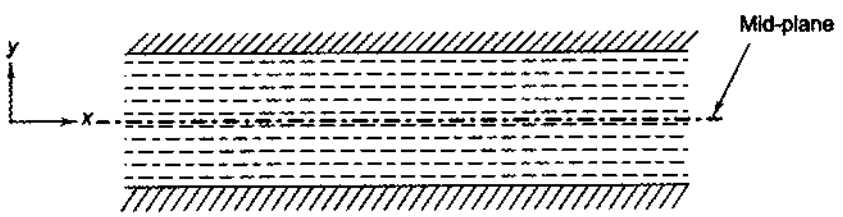


Figure 5.6 *Fluid flow between two parallel plates*

The boundary conditions for the problem are as follows:

(i) No-slip condition: At the plate surface, $u = 0$ (5.64)

(ii) Symmetry condition: $m\dfrac{\mathrm{d}u}{\mathrm{d}y} = 0$ (5.65)

The symmetry condition arises because of the fact that at the line of symmetry there cannot be a shear force. Equation (5.64) is an essential boundary condition, whereas Eq. (5.65) is a natural boundary condition.

The formulation of this problem is similar to the heat conduction problem or that of a rod loaded with axial loads. You may solve this problem and compare it with the exact solution. The exact solution is quadratic in y. If we take a three-noded element, we will obtain a result that perfectly matches the exact solution with a mesh consisting of one element only , whereas if we take a two-noded element, then there will be a non-zero error involved no matter how many elements are taken. The error will decrease with increase in the number of elements in the mesh, but will never be zero.

What is the point in solving this equation by using FEM? We already know the exact solution. The importance of FEM will become evident when we encounter a situation in which the viscosity varies as a function of y. In that case, it may be difficult to find out the exact solution. However, FEM can provide approximate solutions easily.

5.5 CONCLUSION

In this chapter, we have studied the problems of heat conduction, solid mechanics, and fluid flow. The problems from these three different areas have been chosen to emphasize the applicability of FEM as a general tool. By now it must have become clear that although FEM was initially developed to serve as a tool of structural mechanics, it has the potential of application to a wide variety of areas.

SOLVED PROBLEMS

5.1 For a two-noded finite element of length l, obtain the nodal contributions of the loads in the following situations:

(i) A concentrated axial load of magnitude P acts at a distance of $l/3$ from the first node.

(ii) The element is subjected to linearly varying load intensity. The load intensity at node 1 is q_1, and at node 2, it is q_2.

Solution (i) As seen from Eq. (5.59), the load vector for a rod subjected to an axial load of intensity q is given by

$$\text{load vector} = \int_0^l q\{N\}\mathrm{d}x \tag{5.1.1}$$

Here

$$q = P\delta\left(x - \frac{l}{3}\right) \tag{5.1.2}$$

Thus, substituting in Eq. (5.1.1), we obtain

$$\text{load vector} = \int_0^l P \begin{Bmatrix} 1 - \dfrac{x}{l} \\ \dfrac{x}{l} \end{Bmatrix} d\left(x - \frac{l}{3}\right) dx = P \begin{Bmatrix} 1 - \dfrac{x}{l} \\ \dfrac{x}{l} \end{Bmatrix}_{x=\frac{l}{3}} = P \begin{Bmatrix} \dfrac{2}{3} \\ \dfrac{1}{3} \end{Bmatrix} \tag{5.1.3}$$

Therefore, the contribution at the first node is 2/3 P and at second node 1/3 P. Sum of these contributions is P, which is equal to the applied load.

(ii) For this case:

$$q = q_1 + \frac{q_2 - q_1}{l} x = \begin{bmatrix} 1 - \dfrac{x}{l} & \dfrac{x}{l} \end{bmatrix} \begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} \tag{5.1.4}$$

Therefore, the load vector of Eq. (5.1.1) becomes

$$\text{load vector} = \int_0^l \begin{bmatrix} 1 - \dfrac{x}{l} & \dfrac{x}{l} \end{bmatrix} \begin{Bmatrix} q_1 \\ q_2 \end{Bmatrix} \begin{Bmatrix} 1 - \dfrac{x}{l} \\ \dfrac{x}{l} \end{Bmatrix} dx = \begin{Bmatrix} \dfrac{q_1 l}{3} + \dfrac{q_2 l}{6} \\ \dfrac{q_1 l}{6} + \dfrac{q_2 l}{3} \end{Bmatrix} \tag{5.1.5}$$

It can be verified that the sum of nodal loads is

$$\frac{(q_1 + q_2) l}{2}$$

This is equal to the total load in the element.

5.2 The thermal conductivity of a stainless steel rod of 0.1 m length and area of cross-section of 1 cm^2 is 20 W/m-°C. The rate of heat generation in the rod is 10^5 W/m^3. One end of the rod is kept at 0°C and the other end at 100°C. The rod is insulated except at the ends. Using finite element with two elements, find out the temperature at the mid-point of the rod. Also, find out the heat flow at the ends of the rod using FEM. Compare the results with the exact solution.

Solution

Figure 5.7 *Discretization with two elements*

Figure 5.7 shows a rod discretized with two elements. The length of each element, $h = 0.05$ m. As the lateral surface is insulated we take the connective heat transfer coefficient for this surface as 0, that is, $h_c = 0$. Using Eq. (5.14), we obtain

$$[k^e] = \frac{Ak}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \frac{10^{-4} \times 20}{0.05}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 0.04 & -0.04 \\ -0.04 & 0.04 \end{bmatrix}$$

In this problem, the rate of heat generation, $\dot{Q} = 10^5$ W/m^3. Using Eq. (5.15), we obtain

$$[R^e] = \begin{Bmatrix} \dfrac{\dot{Q}Ah}{2} \\ \dfrac{\dot{Q}Ah}{2} \end{Bmatrix} = \begin{Bmatrix} \dfrac{10^5 \times 10^{-4} \times 0.05}{2} \\ \dfrac{10^5 \times 10^{-4} \times 0.05}{2} \end{Bmatrix} = \begin{Bmatrix} 0.25 \\ 0.25 \end{Bmatrix}$$

Using Eq. (5.12), the elemental equations for element 1 can be written as

$$\begin{bmatrix} 0.04 & -0.04 \\ -0.04 & 0.04 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = \begin{Bmatrix} 0.25 \\ 0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(kA\dfrac{dT}{dx}\right)_1 \\ \left(kA\dfrac{dT}{dx}\right)_2 \end{Bmatrix} \tag{5.2.1}$$

Similarly, the elemental equations for element 2 are given by

$$\begin{bmatrix} 0.04 & -0.04 \\ -0.04 & 0.04 \end{bmatrix}\begin{Bmatrix} T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} 0.25 \\ 0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(kA\dfrac{dT}{dx}\right)_2 \\ \left(kA\dfrac{dT}{dx}\right)_3 \end{Bmatrix} \tag{5.2.2}$$

Assembling Eqs (5.2.1) and (5.2.2), we obtain

$$\begin{bmatrix} 0.04 & -0.04 & 0 \\ -0.04 & 0.04+0.04 & -0.04 \\ 0 & -0.04 & 0.04 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} 0.25 \\ 0.25+0.25 \\ 0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(kA\dfrac{dT}{dx}\right)_1 \\ \left(kA\dfrac{dT}{dx}\right)_2 - \left(kA\dfrac{dT}{dx}\right)_2 \\ \left(kA\dfrac{dT}{dx}\right)_3 \end{Bmatrix}$$

Thus, the assembled finite element equations are

$$\begin{bmatrix} 0.04 & -0.04 & 0 \\ -0.04 & 0.08 & -0.04 \\ 0 & -0.04 & 0.04 \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = \begin{Bmatrix} 0.25 \\ 0.5 \\ 0.25 \end{Bmatrix} + \begin{Bmatrix} -\left(kA\dfrac{dT}{dx}\right)_1 \\ 0 \\ \left(kA\dfrac{dT}{dx}\right)_3 \end{Bmatrix} \tag{5.2.3}$$

Now, we put the boundary conditions $T_1 = 0°C$ and $T_3 = 100°C$. For this purpose, we eliminate the first and third row and write the equation corresponding to the middle row. Thus

$$-0.04T_1 + 0.08T_2 - 0.04T_3 = 0.5 \tag{5.2.4}$$

Substituting the values of T_1 and T_3 and simplifying Eq. (5.2.4) we obtain $T_2 = 56.25°C$.

Having found the temperature at the mid-point, we can carry out post-processing to find out the heat flow through the ends. According to Eq. (5.1), the steady-state heat flux is given by

$$q = -k\frac{dT}{dx}$$

Using this heat flow for element 1 is

$$qA = -kA\lfloor N_{1,x} \quad N_{2,x} \rfloor \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} = kA\frac{T_1 - T_2}{h} \tag{5.2.5}$$

Note that a two-noded finite element provides a constant heat flow throughout the element. Substituting the numerical values, we obtain the heat flow for element 1 as –2.25 W. The minus sign indicates that heat is flowing in the negative x-direction, that is, the heat is coming out of the element through the left end. The heat flow for element 2 is determined using

$$qA = kA\frac{T_2 - T_3}{h} \tag{5.2.6}$$

On substituting the numerical values, we obtain –1.75 W. Again, the heat flow is in negative x-direction. This implies that the heat is entering the rod from the right end.

The governing differential equation of this problem is

$$\frac{d^2T}{dx^2} + \dot{Q} = 0 \tag{5.2.7}$$

with $T = 0°C$ at $x = 0$ and $T = 100°C$ at $x = 0.1$ m. Solving Eq. (5.27), we get

$$T = 1250x - 2500x^2$$

This gives the exact temperature at the mid-point as 56.25°C, same as that obtained by FEM. However, the temperature distribution inside the elements will differ. The exact temperature distribution is quadratic in an element, but FEM gives a linear distribution. Using Fourier's law, the exact heat flow is obtained as

$$\text{Exact heat flow} = -kA\frac{dT}{dx} = kA(5000x - 1250x)$$

Substituting the values of k and A, we get

$$\text{Exact heat flow} = kA(5000x - 1250x) = 10x - 2.5$$

Thus, the heat flow at the left end of the rod (at $x = 0$) is –2.5 W and at right end ($x = 0.1$), it is –1.5 W.

Interestingly, if we recover the heat flow from Eq. (5.2.4), we get the exact value in this case. However, in general, we are not interested only in heat flow at the ends, but also inside the elements. Using equations similar to Eqs (5.2.5) and (5.2.6) is the standard practice in FEM.

EXERCISES

A. Multiple Choice Questions

5.1 The heat generation in a two-noded element is 500 W/m. If the length of the element is 1 cm, the right hand side vector due to this heat generation is

(a) $\lfloor 2.5\text{W} \quad 2.5\text{W} \rfloor^T$ (b) $\lfloor 5\text{W} \quad 5\text{W} \rfloor^T$

(c) $\lfloor 250\text{W} \quad 250\text{W} \rfloor^T$ (d) $\lfloor 50\text{W} \quad 50\text{W} \rfloor^T$

5.2 In a three-noded element for a rod subjected to an axial load, the uniform load intensity inside the element is q. The coordinates of the nodes 1, 2, and 3 are given as 0, $h/2$, and h, respectively. The right hand side vector due to this load intensity is

(a) $\left\lfloor \frac{qh}{3} \quad \frac{qh}{3} \quad \frac{qh}{3} \right\rfloor^T$ (b) $\left\lfloor \frac{qh}{6} \quad \frac{2qh}{3} \quad \frac{qh}{6} \right\rfloor^T$

(c) $\left\lfloor \frac{qh}{4} \quad \frac{qh}{2} \quad \frac{qh}{4} \right\rfloor^T$ (d) $\left\lfloor \frac{qh}{6} \quad \frac{qh}{3} \quad \frac{qh}{6} \right\rfloor^T$

5.3 A viscous fluid is flowing between two parallel plates. The flow is steady, laminar, two-dimensional, and incompressible. In a plane parallel to the plates and equidistant from them,

(a) shear stress and shear strain are 0

(b) shear stress is 0, but not shear strain

(c) shear strain is 0, but not shear stress

(d) shear stress and shear strain are non-zero

5.4 For a fully developed laminar in a pipe, the volume flow rate Q is related to the pressure gradient by

$$Q = -\frac{\pi D^4}{128\mu}\frac{dP}{dx}$$

where μ is the dynamic viscosity of the fluid, and D *is* the pipe diameter. If the a pipe of length l is taken as a finite element, the element equations are given by

$$\begin{bmatrix} C & -C \\ -C & C \end{bmatrix} \begin{Bmatrix} P_1 \\ P_2 \end{Bmatrix} = \begin{Bmatrix} Q_1 \\ Q_2 \end{Bmatrix}$$

where C is equal to

(a) $\frac{\pi D^4}{l\mu}$ (b) $\frac{\pi D^4}{128 l\mu}$ (c) $\frac{\pi D^4 l}{128\mu}$ (d) $\frac{\pi D^4}{128\mu}$

5.5 In a three-noded (one node exactly at the middle of the end nodes) finite element for a rod subjected to axial loads, there are zero displacements at the left-end node and middle node and non-zero displacement at the right-end node. For this element, the FEM will predict zero strain at a distance of

(a) one half of the element length from the right-end node

(b) one fourth of the element length from the right-end node

(c) one fourth of the element length from left-end node

(d) zero from the left-end node.

B. Problems

5.6 An elastic rod of area of cross-section A, Young's modulus of elasticity E, and length l^e is divided into two elements of equal length with three nodes per element (two end points and the mid point) and one degree of freedom (i.e., u) per node (Fig. 5.8). Prove that the elemental stiffness and load vector are given by

$$[k^e] = \frac{AE}{3l^e}\begin{bmatrix} 7 & -8 & 1 \\ -8 & 16 & -8 \\ 1 & -8 & 7 \end{bmatrix} \qquad [f^e] = \frac{pl^e}{6}\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix}$$

Find the global coefficient matrix and the right side vector. Apply the essential boundary conditions. Assuming $u_2 = u_4$, determine the nodal displacement u_2 and u_3.
Prove that the exact solution for this problem is

$$u = \frac{px(l-x)}{2AE}$$

Compare the exact solution at the nodes 2 and 3 with their FEM values.

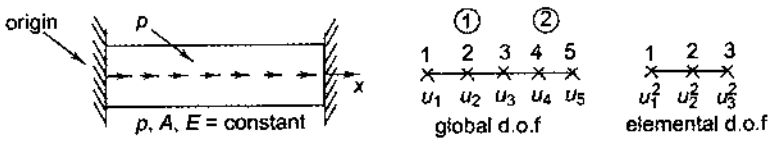


Figure 5.8 *An electric rod and its degrees of freedom*

5.7 In a one-dimensional stress analysis of an axial rod, one end is fixed and the other end just touches a wall at room temperature (Fig. 5.9). The rod is heated to 100°C. Model the rod by two linear elements. Give all the finite element matrix equations to evaluate the reactions at the ends. Also find the thermal stresses at point P at the centre of the element (2).
(Make use of initial strain approach).

Use the following values for calculations:

$A_1 = 2\ \text{cm}^2$ $\quad A_2 = 1\ \text{cm}^2$

$l_1 = 50\ \text{cm}$ $\quad l_2 = 50\ \text{cm}$

$E = 200\ \text{GPa}$ $\quad \Delta T = 100°\text{C}$ $\quad \alpha = 17.3 \times 10^{-6}$

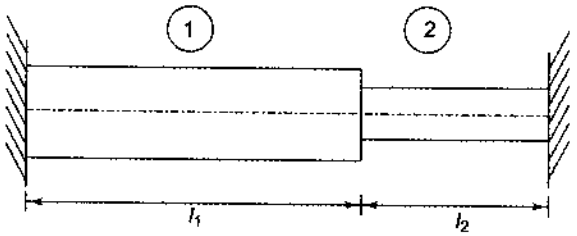


Figure 5.9 *An axial rod*

5.8 Consider the following problem, where

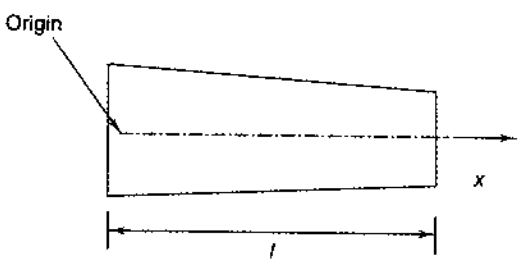
Differential equation: $\dfrac{d}{dx}\left(kA\dfrac{dT}{dx}\right) + r = 0$

Boundary conditions:

(i) Essential boundary condition: $T = T^*$ at $x = 0$

(ii) Natural (convection) boundary condition: $\left(kA\dfrac{dT}{dx}\right) = -hA(T - T_0)$ at $x = l$

$$\left(kA\frac{dT}{dx}\right) = -hA(T - T_0)$$



T = temperature difference
A = area of cross-section
r = heat generated per unit length
k = thermal conductivity
h = heat transfer coefficient
T_0 = ambient temperature
T^* = specified temperature at $x = 0$

Figure 5.10 *The system for Problem 5.8*

Obtain FEM formulation using Ritz FEM and solve by taking two elements (Fig. 510).

5.9 The equation for one-dimensional steady-state heat conduction in absence of heat generation is as follows:

$$\frac{d}{dx}\left(kA\frac{dT}{dx}\right) = 0$$

Let us take a rod with end-point coordinates of $x = 1$ and $x = 3$, respectively. Assume that $k = 1$ throughout the domain and non-uniform cross-sectional area is given by $A = x$. Boundary conditions are

$$T\,(\text{at } x = 1) = 300 \text{ K},\ T\,(\text{at } x = 3) = 400 \text{ K}$$

The stiffness matrix of an element for this problem is given by

$$\int_0^h \frac{kA}{h^2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} dx$$

where h is the length of the element.

Now the following two approaches can be used for computing the stiffness matrix:

(i) A is treated as constant and equal to the average area of the element

(ii) A is treated as varying

Solve this problem by both the approaches and compare them by studying their performance with different number of elements. (You can easily find out the exact solution also.) You may use MATLAB or write a program in C/C++.

6
CHAPTER

FEM Formulation for Bending of Beams

6.1 Introduction

6.2 Galerkin FEM Formulation

6.3 Ritz FEM Formulation

6.4 Conclusion

6.1 INTRODUCTION

By now we are familiar with finite element procedure for one-dimensional problems. We have taken several examples to illustrate the Galerkin and Ritz FEM formulations. However, so far we have encountered the problems requiring only C^0 continuity for the approximating functions. In this chapter, we will learn the formulation of the problems requiring C^1 continuity, that is, the value of the function as well as its first derivative should be continuous. To understand the finite element formulation of such problems, we choose the bending of an Euler–Bernoulli beam. The differential equation governing this problem is a fourth-order differential equation, requiring C^1 continuity. The procedure developed in this chapter will be useful for any fourth-order ordinary differential equation.

6.2 GALERKIN FEM FORMULATION

The differential equation that represents the bending of an Euler–Bernoulli beam is given by

$$EI_{zz}\frac{\mathrm{d}^4 v}{\mathrm{d}x^4} = q \tag{6.1}$$

where EI_{zz} is called the flexural rigidity, which is the product of Young's modulus of elasticity and second moment of area of the cross-section (with respect to a centroidal axis perpendicular to the plane of bending), v is the vertical deflection as a function of longitudinal coordinate x, and q is the load intensity (load per unit length) function. To completely solve this equation, we need four boundary conditions. However, out of these four boundary conditions, at least one boundary condition should be in the form of prescribed deflection and, in addition, one other boundary condition should be in the form of either a prescribed deflection or a prescribed slope. It is not essential to prescribe the second or third derivative of the deflection if the slope and deflections are prescribed instead. Thus, the boundary conditions on the slope and deflection are called *essential boundary conditions*, while the other boundary conditions involving the derivatives of deflection and slope are called *natural boundary conditions*. In the following subsections, we explain the steps of Galerkin FEM formulation.

6.2.1 Weak Form

The first step in Galerkin FEM is to obtain the weak form of a differential equation. For this purpose, we multiply the residual of the differential equation by a weight function w and integrate it by parts so as to evenly distribute the order of differentiation on v and w. In the weak form, both v and w will have the derivatives up to second order only. The weighted integral of the residual of Eq. (6.1) is given by

$$\int_0^l \left(EI\frac{\mathrm{d}^4 v}{\mathrm{d}x^4} - q \right) w\,\mathrm{d}x = \int_0^l EI\frac{\mathrm{d}^4 v}{\mathrm{d}x^4} w\,\mathrm{d}x - \int_0^l q\,\mathrm{d}x = 0 \tag{6.2}$$

where l is the length of the beam. (In case of an element of length h, the weak form can be obtained by simply replacing l with h.) We integrate the first integral by parts to reduce the order of derivative on v and obtain

$$EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3}w\bigg|_0^l - \int_0^l EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3}\frac{\mathrm{d}w}{\mathrm{d}x}\mathrm{d}x - \int_0^l q\mathrm{d}x = 0 \tag{6.3}$$

By doing this, we have reduced the order of the derivative on v while simultaneously causing an increase in the order of derivative on w. We further integrate the first integral in Eq. (6.3) to obtain

$$EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3}w\bigg|_0^l - EI\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\frac{\mathrm{d}w}{\mathrm{d}x}\bigg|_0^l + \int_0^l EI\frac{\mathrm{d}^2w}{\mathrm{d}x^2}\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\mathrm{d}x - \int_0^l qw\mathrm{d}x = 0 \tag{6.4}$$

We shall not integrate the first integral in Eq. (6.4) further, because by doing so the order of derivative on w will become three and that of v will reduce to one. At the present stage, the derivative on both w and v is of the order two.

In the weak form [represented by Eq. (6.4)], if we substitute w with δv, we get the variational form. Replacing w by δv in Eq. (6.4), we obtain

$$EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3}\delta v\bigg|_0^l - EI\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\frac{\mathrm{d}\delta v}{\mathrm{d}x}\bigg|_0^l + \int_0^l EI\frac{\mathrm{d}^2\delta v}{\mathrm{d}x^2}\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\mathrm{d}x - \int_0^l q\delta v\mathrm{d}x = 0 \tag{6.5}$$

which can be written as

$$\delta I = EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3}\delta v\bigg|_0^l - EI\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\frac{\mathrm{d}\delta v}{\mathrm{d}x}\bigg|_0^l + \delta\int_0^l EI\left(\frac{\mathrm{d}^2v}{\mathrm{d}x^2}\right)^2\mathrm{d}x - \delta\int_0^l qv\mathrm{d}x = 0 \tag{6.6}$$

Then, observing the terms on the boundary, we can recognize that at $x = 0$ and $x = l$

$$\text{either } EI\frac{\mathrm{d}^3v}{\mathrm{d}x^3} = 0 \text{ or } \delta v = 0 \tag{6.7}$$

and

$$\text{either } EI\frac{\mathrm{d}^2v}{\mathrm{d}x^2} = 0 \text{ or } \delta\frac{\mathrm{d}v}{\mathrm{d}x} = 0 \tag{6.8}$$

The first boundary condition in each of the Eqs (6.3) and (6.4) is called the natural boundary condition, while the boundary conditions of having zero variations in the deflection v and slope $\mathrm{d}v/\mathrm{d}x$ are called the essential (or geometric) boundary conditions. From the study of strength of materials, recall that $EI\,\mathrm{d}^2v/\mathrm{d}x^2$ is the bending moment and $EI\,\mathrm{d}^3v/\mathrm{d}x^3$ is the negative of the shear force. (The sign convention for the bending moment and shear force may differ from book to book.)

6.2.2 Choosing Suitable Approximating Functions

After obtaining the weak form, we have to choose the suitable approximating functions within an element. Observing the weak form given in Eq. (6.2), we see that the highest order of the derivative on

v is three; therefore, the approximating function should be thrice differentiable. A third-degree polynomial is a function of that kind. Thus, one may approximate v within an element as

$$v = a + bx + cx^2 + dx^3 \tag{6.9}$$

Inside the integral, the highest order of derivative is two; therefore, the overall approximation should be C^1 continuous. If we take a four-noded Lagrange element, it will not guarantee that the slope obtained at the end points will be same for two adjacent elements. However, if we obtain the constants a, b, c, and d in Eq. (6.5) by expressing them in terms of the slopes and deflections at the ends of the element, the continuity of slope is ensured. In Galerkin FEM, w is approximated in the same way as v. In the next subsection, we shall express the approximating functions in terms of nodal values of slope and deflection.

6.2.3 Hermitian Shape Functions

We denote the end points (nodes) of a beam element by 1 and 2 and use them as subscripts for specifying the value at a specified point. The coordinate of point 1 is 0 and that of 2 is h. Then

$$\begin{aligned} v_1 &= a \\ \left(\frac{dv}{dx}\right)_1 &= v_{1,x} = b \\ v_2 &= a + bh + ch^2 + dh^3 \\ \left(\frac{dv}{dx}\right)_2 &= v_{2,x} = b + 2ch + 3dh^2 \end{aligned} \tag{6.10}$$

With the help of these set of equations, we can express the constants in terms of the nodal values of slopes and deflection. Substituting them in Eq. (6.9) and rearranging, we obtain

$$v = N_1 v_1 + N_2 v_{1,x} + N_3 v_2 + N_4 v_{2,x} \tag{6.11}$$

where N_1, N_2, N_3, and N_4 are called *Hermitian shape* functions. Their values are given by

$$N_1 = 1 - 3\left(\frac{x}{h}\right)^2 + 2\left(\frac{x}{h}\right)^3 \tag{6.12}$$

$$N_2 = x\left(1 - \frac{x}{h}\right)^2 \tag{6.13}$$

$$N_3 = 3\left(\frac{x}{h}\right)^2 - 2\left(\frac{x}{h}\right)^3 \tag{6.14}$$

$$N_4 = x\left[\left(\frac{x}{h}\right)^2 - \frac{x}{h}\right] \tag{6.15}$$

If the coordinates of the first node is not 0, but x_i, the shape functions are given by

$$N_1 = 1 - 3\left(\frac{x - x_i}{h}\right)^2 + 2\left(\frac{x - x_i}{h}\right)^3 \tag{6.16}$$

$$N_2 = (x - x_i)\left(1 - \frac{x - x_i}{h}\right)^2 \tag{6.17}$$

$$N_3 = 3\left(\frac{x - x_i}{h}\right)^2 - 2\left(\frac{x - x_i}{h}\right)^3 \tag{6.18}$$

$$N_4 = (x - x_i)\left[\left(\frac{x - x_i}{h}\right)^2 - \frac{x - x_i}{h}\right] \tag{6.19}$$

6.2.4 Elemental Equations

For obtaining the elemental stiffness matrix for a beam element e of length h, we change l to h in Eq. (6.2) and substitute the approximations

$$v^e = \lfloor N_1 \;\; N_2 \;\; N_3 \;\; N_4 \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} \tag{6.20}$$

and

$$w = \lfloor w_1 \;\; w_{1,x} \;\; w_2 \;\; w_{2,x} \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} = \lfloor w^{ne} \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \tag{6.21}$$

Thus

$$\int_0^h \lfloor w^{ne} \rfloor EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} \mathrm{d}x = \int_0^h \lfloor w^{ne} \rfloor q \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x + \lfloor w^{ne} \rfloor \begin{Bmatrix} -V_1 \\ -M_1 \\ V_2 \\ M_2 \end{Bmatrix} \tag{6.22}$$

As the nodal weights are arbitrary, Eq. (6.22) gives the following elemental equations:

$$\int_0^h EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \mathrm{d}x \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} = \int_0^h q \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} -V_1 \\ -M_1 \\ V_2 \\ M_2 \end{Bmatrix} \tag{6.23}$$

Here, the integral on the left hand side of the equality sign is called the *stiffness matrix*, the last term on the right hand side is called the *internal load vector*, and the middle term is called the *load vector*.

6.2.5 Assembly, Boundary Conditions, and Solution

After obtaining the elemental equations, we need to assemble them as usual. During assembly, internal loads—shear force and bending moments—get cancelled everywhere except at the boundaries. These are tackled by means of the boundary conditions. To obtain a solution either the values of the loads or their slopes and/or deflections need to be prescribed. If these are provided, we can solve the equations.

6.3 RITZ FEM FORMULATION

It is usually possible to find a functional, which on minimization along with the satisfaction of geometric boundary conditions will lead to the solution of the differential equation with prescribed boundary conditions. Ritz method can be used in such situations. The method is illustrated by considering the example of a beam. In a typical beam element of length h, the total potential energy may be written as

$$\Pi = \int_0^h \frac{1}{2} EI \left(\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right)^2 \mathrm{d}x - \int_0^h qv \, \mathrm{d}x + V_1 v_1 - V_2 v_2 + M_1 v_{1,x} - M_2 v_{2,x} \tag{6.24}$$

where the first term denotes the *strain energy* and all the other terms represent the *potential due to loads*. Here, v is the beam deflection, V is the shear force, and M is the moment. The subscripts 1 and 2 indicate *nodes*. The highest order of derivative in this expression is 3, since the shear force V contains the third derivative of v. Hence, v should be of the form $a + bx + cx^2 + dx^3$. Since the highest order of derivative inside the integral is 2, the approximating function should be C^1 continuous. One suitable approximation is given by Eq. (6.20). Substituting this approximation and the second derivatives of the approximation in Eq. (6.24), we obtain

$$\begin{aligned} \Pi = & \int_0^h \frac{1}{2} \lfloor v_1 \quad v_{1,x} \quad v_2 \quad v_{2,x} \rfloor EI \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} \mathrm{d}x \\ & - \int_0^h q \lfloor v_1 \quad v_{1,x} \quad v_2 \quad v_{2,x} \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x + \lfloor v_1 \quad v_{1,x} \quad v_2 \quad v_{2,x} \rfloor \begin{Bmatrix} V_1 \\ M_1 \\ -V_2 \\ -M_2 \end{Bmatrix} \end{aligned} \tag{6.25}$$

For minimizing the expression given in Eq. (6.25), we differentiate it with respect to $\lfloor v_1 \quad v_{1,x} \quad v_2 \quad v_{2,x} \rfloor^T$ and set it equal to zero. Thus

$$\frac{\partial \Pi}{\partial \lfloor v_1 \quad v_{1,x} \quad v_2 \quad v_{2,x} \rfloor^T} = 0 \tag{6.26}$$

which provides

$$\int_0^h EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \mathrm{d}x \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} = \int_0^h q \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} -V_1 \\ -M_1 \\ V_2 \\ M_2 \end{Bmatrix} \tag{6.27}$$

This is the same equation as obtained by Galerkin formulation [given in Eq. (6.23)]. Thus, for the Euler–Bernoulli beam problem, Galerkin and Ritz formulations provide the same finite element equations. There is no difference between both the methods regarding the procedures of assembly, application of boundary conditions, and solution of simultaneous linear equations.

6.4 CONCLUSION

In this chapter, FEM formulation of beam problem is carried out by using Galerkin and Ritz FEM formulations. It is observed that both the formulations yield the same results. The procedure can be employed for any fourth-order differential equation.

SOLVED PROBLEMS

6.1 Find the load vector due to external loads for a beam element of length h under the following loadings:

(i) Uniformly distributed load of intensity q

(ii) A clockwise moment M acting at a distance of $h/4$ from the left end

Solution The load vector is given by

$$\text{Load vector} = \int_0^h q \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x = \int_0^h q \begin{Bmatrix} 1 - 3\dfrac{x^2}{h^2} + 2\dfrac{x^3}{h^3} \\ x\left(1 + \dfrac{x^2}{h^2} - 2\dfrac{x}{h}\right) \\ 3\dfrac{x^2}{h^2} - 2\dfrac{x^3}{h^3} \\ x\left(\dfrac{x^2}{h^2} - \dfrac{x}{h}\right) \end{Bmatrix} \mathrm{d}x \tag{6.1.1}$$

Using Eq. (6.1.1) we can find out the load vector for the two cases as follows:

(i) Here q is constant and can be taken outside the integral. The definite integration in Eq. (6.1.1) yields

$$\text{Load vector} = q\left\{\begin{matrix} x-\dfrac{x^3}{h^2}+\dfrac{x^4}{2h^3} \\ \dfrac{x^2}{2}+\dfrac{x^4}{4h^2}-\dfrac{2}{3}\dfrac{x^3}{h} \\ \dfrac{x^3}{h^2}-\dfrac{x^4}{2h^3} \\ \dfrac{x^4}{4h^2}-\dfrac{x^3}{3h} \end{matrix}\right\}\Bigg|_0^h = \left\{\begin{matrix} \dfrac{qh}{2} \\ \dfrac{qh^2}{12} \\ \dfrac{qh}{2} \\ -\dfrac{qh^2}{12} \end{matrix}\right\} \tag{6.1.2}$$

We observe that the total load of qh in the element is equally shared between the two nodes. In addition, there are moments of the magnitude $qh^2/12$, although the net moment is zero.

(ii) Just as we can convert a concentrated load into a distributed load using a Dirac-delta function, we can also convert the concentrated moment using a *doublet* function. Thus, the given load is written as

$$q = M\eta\left(x-\frac{h}{4}\right)$$

It is important to note the sign convention that while transforming a concentrated moment into a uniformly distributed load, the clockwise moment has been considered to be positive. Using Eq. (6.1.1):

$$\text{Load vector} = \int_0^h M\eta\left(x-\frac{h}{4}\right)\begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix}\mathrm{d}x \tag{6.1.3}$$

In Eq. (6.1.3), M is a constant. We carry out the integration by parts, by treating $M\eta\left(x-\frac{h}{4}\right)$ as the first function and the vector of shape functions as the second function. Thus

$$\text{Load vector} = \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix}\Bigg|_0^h \int_0^h M\eta\left(x-\frac{h}{4}\right)\mathrm{d}x - \int_0^h \begin{Bmatrix} \dfrac{\mathrm{d}N_1}{\mathrm{d}x} \\ \dfrac{\mathrm{d}N_2}{\mathrm{d}x} \\ \dfrac{\mathrm{d}N_3}{\mathrm{d}x} \\ \dfrac{\mathrm{d}N_4}{\mathrm{d}x} \end{Bmatrix}\left(\int_0^x M\eta\left(x-\frac{h}{4}\right)\mathrm{d}x\right)\mathrm{d}x$$

$$= \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix}\Bigg|_0^h \left\{ M\delta\left(x - \frac{h}{4}\right)\right\}\Bigg|_0^h - \int_0^h M \begin{Bmatrix} \dfrac{dN_1}{dx} \\ \dfrac{dN_2}{dx} \\ \dfrac{dN_3}{dx} \\ \dfrac{dN_4}{dx} \end{Bmatrix} \delta\left(x - \frac{h}{4}\right) dx$$

$$= 0 - M \begin{Bmatrix} \dfrac{dN_1}{dx} \\ \dfrac{dN_2}{dx} \\ \dfrac{dN_3}{dx} \\ \dfrac{dN_4}{dx} \end{Bmatrix}_{x=\frac{h}{4}} = -M \begin{Bmatrix} -\dfrac{6x}{h^2} + \dfrac{6x^2}{h^3} \\ 1 + \dfrac{3x^2}{h^2} - \dfrac{4x}{h} \\ \dfrac{6x}{h^2} - \dfrac{6x^2}{h^3} \\ \dfrac{3x^2}{h^2} - \dfrac{2x}{h} \end{Bmatrix}_{x=\frac{h}{4}} = \begin{Bmatrix} \dfrac{9M}{8h} \\ -\dfrac{3M}{16} \\ -\dfrac{9M}{8h} \\ \dfrac{5M}{16} \end{Bmatrix}$$

We observe the following points:

Net force in the element = $\dfrac{9M}{8h} - \dfrac{9M}{8h} = 0$ and

Net moment (taking the moment about the left node) = $-\dfrac{3M}{16h} + \dfrac{5M}{16h} - \dfrac{9M}{8h} h = -M$

Thus, the finite element procedure has produced a statically equivalent system.

6.2 How will you obtain the finite element model of a beam if it is supported on flexible supports at both ends?

Solution Assume that the beam is supported on linear and torsion springs at both the ends. The spring constant of the left-end linear spring is k_l and that of the left-end torsion spring is k_{tl}. For the right end, the spring constants of linear and torsion springs are k_r and k_{tr}, respectively. The expression for the potential energy of a typical element (excluding the elements at the extreme left and extreme right) is given by Eq. (6.24):

$$\Pi = \int_0^h \frac{1}{2} EI \left(\frac{d^2v}{dx^2}\right)^2 dx - \int_0^h qv\, dx + V_1 v_1 - V_2 v_2 + M_1 v_{1,x} - M_2 v_{2,x}$$

In the above expression, it is assumed that the shear forces and bending moments do not depend on the deflections and slope and are constant for an applied loading.

For the element at the extreme left

$$V_1 = k_l v_1 \quad \text{and} \quad M_1 = k_{tl} v_{1,x}$$

Note that shear force is dependent on the deflection and moment on slope. The total potential energy for this element is given by

$$\Pi = \int_0^h \frac{1}{2} EI \left(\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right)^2 \mathrm{d}x - \int_0^h qv \, \mathrm{d}x + \frac{1}{2} k_l v_1^2 - V_2 v_2 + \frac{1}{2} k_{tl} v_{1,x}^2 - M_2 v_{2,x}$$

Substituting the approximation

$$v^e = \lfloor N_1 \;\; N_2 \;\; N_3 \;\; N_4 \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix}$$

for the element, we obtain

$$\Pi = \int_0^h \frac{1}{2} \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor EI \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \lfloor N_{1,xx} \;\; N_{2,xx} \;\; N_{3,xx} \;\; N_{4,xx} \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} \mathrm{d}x$$

$$+ \frac{1}{2} \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor \begin{bmatrix} k_l & 0 & 0 & 0 \\ 0 & k_{tl} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} \tag{6.2.1}$$

$$- \int_0^h q \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x - \frac{1}{2} \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor \begin{Bmatrix} 0 \\ 0 \\ -V_2 \\ -M_2 \end{Bmatrix}$$

Minimizing the expression in Eq. (6.2.1) with respect to nodal deflections and slopes, we get

$$\frac{\partial \Pi}{\partial \lfloor v_1 \;\; v_{1,x} \;\; v_2 \;\; v_{2,x} \rfloor^{\mathrm{T}}} = \left(\int_0^h EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \lfloor N_{1,xx} \;\; N_{2,xx} \;\; N_{3,xx} \;\; N_{4,xx} \rfloor \mathrm{d}x + \begin{bmatrix} k_l & 0 & 0 & 0 \\ 0 & k_{tl} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \right) \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix}$$

$$= \int_0^h q \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} 0 \\ 0 \\ V_2 \\ M_2 \end{Bmatrix}$$

Therefore, the effect of flexible support on the left hand side can be accounted for by modifying the stiffness matrix of the element at the extreme left and treating the left end as free. The modification in the stiffness matrix is as follows: We add k_l to the first element on the leading diagonal and k_{tl} to the second element on the leading diagonal.

A similar procedure has to be carried out for the element at the extreme right. Here, we add k_r to the third element on the leading diagonal and k_{tr} to the fourth element on the leading diagonal.

EXERCISES

A. Multiple Choice Questions

6.1 In a beam element of length h, a transverse load P is applied at a distance of $h/4$ from the left hand side. The load vector due to this load will be

(a) $\left\lfloor \frac{3P}{4} \;\; 0 \;\; \frac{P}{4} \;\; 0 \right\rfloor^T$ (b) $\left\lfloor \frac{27P}{32} \;\; 0 \;\; \frac{5P}{32} \;\; 0 \right\rfloor^T$

(c) $\left\lfloor \frac{27P}{32} \;\; \frac{9Ph}{64} \;\; \frac{5P}{32} \;\; \frac{-3Ph}{64} \right\rfloor^T$ (d) $\left\lfloor \frac{27P}{32} \;\; \frac{9Ph}{64} \;\; \frac{5P}{32} \;\; \frac{-6Ph}{64} \right\rfloor^T$

6.2 An element has a concentrated moment M at the node 2. The load vector due to it will be

(a) $\lfloor 0 \;\; 0 \;\; 0 \;\; M \rfloor^T$ (b) $\lfloor 0 \;\; 0 \;\; 0 \;\; -M \rfloor^T$

(c) $\left\lfloor -\frac{M}{h} \;\; 0 \;\; \frac{M}{h} \;\; M \right\rfloor^T$ (d) $\lfloor 0 \;\; 0 \;\; 0 \;\; 0 \rfloor^T$

6.3 Which of the following is a correct statement regarding the units of shape functions N_1 and N_2 in SI system?

(a) Both N_1 and N_2 have units as meters.

(b) Both N_1 and N_2 are dimensionless.

(c) N_1 is dimensionless, N_2 is expressed in m^{-1}.

(d) N_1 is dimensionless, N_2 is expressed in metres.

6.4 For solving a fourth-order ordinary differential equation to obtain y as a function of x, a beam element of length 1 m is used. On solving the equation, y is obtained as 0 at the first node and 1/3 m at the second node. The first derivative dy/dx is obtained as 0 at the first node and 1/2 at the second node. The third derivative, d^3y/dx^3, in the element is

(a) $-1/6$ (b) $1/6$

(c) -1 (d) $-1/2$

6.5 For a beam element, which of the following relations is true?

(a) $N_1 + N_2 + N_3 + N_4 = 1$ (b) $N_1 + N_3 = 1$

(c) $N_1 + N_2 = 1$ (d) $N_2 + = N_4$

B. Problems

6.6 (i) A beam is modelled by one element (as shown in Fig. 6.1). From finite element equation evaluate the deflection at free end due to concentrated load, P.

(ii) If the beam is acted upon by a uniformly distributed load of intensity q per unit length find the deflection at the free end. Here, first take one element and then keep on increasing the number of elements. Plot the error in end-point deflection versus number of elements.

(iii) Give reasons why the deflection result in the first case is exact but not in the second case.

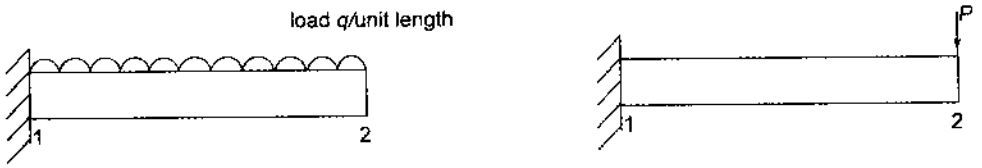


Figure 6.1 *The beam of Problem 6.6*

Given A, I, E, and l are the area of cross-section, second moment of inertia, Young's modulus of elasticity, and length, respectively.

6.7 A load P is applied at the end of the cantilever which is supported on a spring of stiffness k (Fig. 6.2). The length of the cantilever is L and flexural rigidity is given as EI.

(i) Write down the essential and natural boundary conditions of the problem.

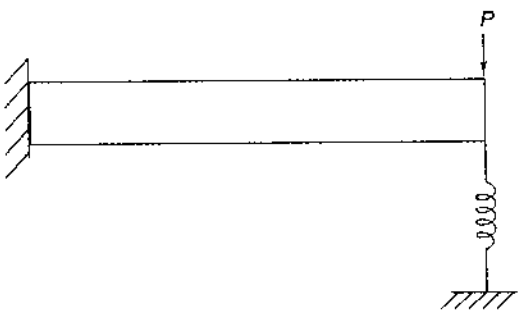


Figure 6.2 *A cantilever with a load P at one end*

(ii) Take one element to solve this problem by FEM. The stiffness matrix for that element is given by

$$\frac{EI}{L}\begin{bmatrix} \frac{12}{L^2} & \frac{6}{L} & \frac{-12}{L^2} & \frac{6}{L} \\ & 4 & \frac{-6}{L} & 2 \\ SYM & & \frac{12}{L^2} & \frac{-6}{L} \\ & & & 4 \end{bmatrix}$$

Generate the load matrix, apply boundary conditions, and obtain the slope and deflection at the ends.

6.8 The functional governing the static buckling of the column [shown in Fig. (6.3)] is given as

$$\Pi = \frac{1}{2}\int_0^L EI\left(\frac{d^2w}{dx^2}\right)^2 dx - \frac{P}{2}\int_0^L\left(\frac{dw}{dx}\right)^2 dx + \frac{1}{2}kw_L^2 \tag{6.10.1}$$

where $w_L = w|_{x=L}$ and the essential boundary conditions are

$$w|_{x=0} = 0, \qquad \left.\frac{dw}{dx}\right|_{x=0} = 0 \tag{6.10.2}$$

Invoke the stationary condition $\delta\Pi = 0$ to derive the differential equation governing the problem and the natural boundary conditions.

For the case, when load P is zero, carry out the FEM formulation by Ritz FEM.

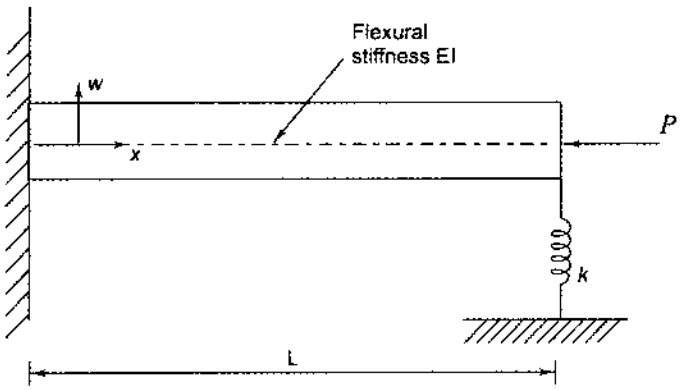


Figure 6.3 *A cantilever beam loaded at one end with P*

6.9 The distribution of the bending moment M in a beam subjected to a loading by a distributed load $w(x)$ per unit length satisfies the equation $d^2M/dx^2 = w(x)$. A beam of unit length is simply supported (i.e., $M = 0$) at both ends and carries a load $w(x) = \sin \pi x$ per unit length.

Carry out FEM formulation for obtaining the bending moment distribution. Obtain the bending moment distributions by considering two and three elements.

7

CHAPTER

FEM Formulation for Trusses and Frames

7.1 Introduction

7.2 Formulation for Trusses

7.3 An Example for FEM Analysis of Plane Trusses

7.4 FEM Formulation for Frames

7.5 Assembly Procedure

7.6 Conclusion

7.1 INTRODUCTION

In this chapter, we will deal with the finite element formulations of two important engineering structures: trusses and frames. A brief description of trusses and frames is as follows:

Truss A truss is a structure constructed to support a load such that when a load is applied on a truss, neither the load nor any part of the structure experiences a displacement. However, in practice, the structure undergoes elastic deformation. The members of the truss are straight slender bars or rods and, in an ideal case, are connected to each other by pin joints. In real situations, the members may be riveted, bolted, or welded at the ends. However, as long as they are slender and cannot carry bending load, they may be assumed to be connected through pin joints. The loads in a truss are applied only at the joints. In cases where the weight of the member is not negligible, it is assumed that half of the weight of each member acts as an applied force at each of their joints.

There are two types of trusses: plane trusses and space trusses. In *plane trusses*, both the truss structure and the applied loads lie in the same plane. In *space trusses*, either the structure or the applied loads, or both, lie in different planes. Each member of the truss can be considered as a rod or bar subjected to an axial load. Therefore, its stiffness can be easily calculated by the method introduced in Chapter 1. However, one has to transform the stiffness matrix into a global form so that the assembly can be easily done.

Frame A frame is a structure in which at least one member carries the transverse load. Members of a frame resist axial as well as bending loads. They may be connected with pin or fixed joints. If two members are connected by a fixed joint, the angle between them is assumed to remain unaltered during deformation. In the case of frames, the local stiffness matrix can be easily written as the combination of the stiffness matrices of a beam and a rod subjected to axial load. The local stiffness matrix has to be converted to global form for assembly.

In this chapter, we shall discuss the formulation for trusses and frames based on the finite element formulation for rods (subjected to axial loads) and beams. It is interesting to see that based on the one-dimensional formulation for rods and beams, we are able to solve two- and three-dimensional problems of trusses and frames. We will also introduce a general assembly procedure that can be easily implemented in a computer code.

7.2 FORMULATION FOR TRUSSES

Figure 7.1 shows two Cartesian coordinate systems, which have the same origin but are rotated with respect to each other. A vector having the components in the coordinate system $x'-y'-z'$ has components u', v', and w' in the $x-y-z$ system. These components are related as

$$\begin{Bmatrix} u' \\ v' \\ w' \end{Bmatrix} = [R] \begin{Bmatrix} u \\ v \\ w \end{Bmatrix} \tag{7.1}$$

where $[R]$, given by

$$[R] = \begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix} \tag{7.2}$$

is called the *rotation matrix*. The following points need to be clearly understood:

- The direction cosines of x' with respect to the x–y–z system are l_1, m_1, and n_1, respectively which means that l_1 is the cosine of angle between x' and x, m_1 is the cosine of angle between x' and y, and n_1 is the cosine of angle between x' and z.
- The direction cosines of y' with respect to the x–y–z system are l_2, m_2, and n_2, respectively which means that l_2 is the cosine of angle between y' and x, m_2 is the cosine of angle between $y'$and y, and n_2 is the cosine of angle between y' and z.
- The direction cosines of z' with respect to the x–y–z system are l_3, m_3, and n_3, respectively.

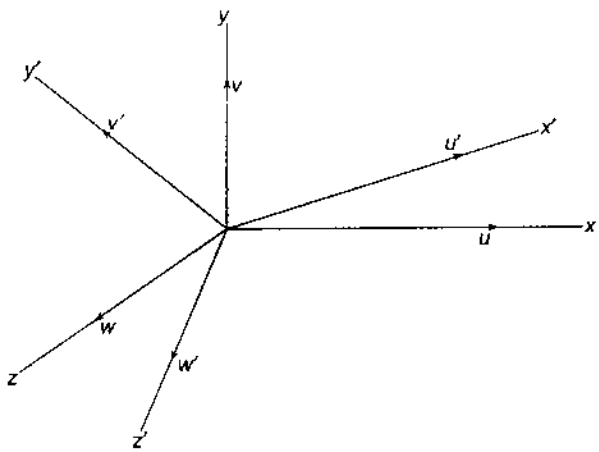


Figure 7.1 *Rotated coordinate system*

The rotation matrix is orthogonal, that is, its inverse is equal to its transpose. Thus, from Eq. (7.1), we get

$$\begin{Bmatrix} u \\ v \\ w \end{Bmatrix} = [R]^{-1} \begin{Bmatrix} u' \\ v' \\ w' \end{Bmatrix} = [R]^{\mathrm{T}} \begin{Bmatrix} u' \\ v' \\ w' \end{Bmatrix} \tag{7.3}$$

If $\{d\} = [u \;\; v \;\; w]^{\mathrm{T}}$ is a displacement vector expressed in terms of its components along x, y, and z axes and $\{d'\} = [u' \;\; v' \;\; w']$ is the same displacement vector expressed in terms of its components along x', y', and z' axes, then

$$\{d'\} = [R]\{d\} \quad \text{and} \quad \{d\} = [R]^{\mathrm{T}}\{d'\} \tag{7.4}$$

If $\{r\} = [F_x \ F_y \ F_z]^T$ is a force vector expressed in terms of its components along x, y, and z axis and $[r']$ is the same force vector expressed in terms of its components along x', y', and z', then

$$\{r'\} = [R]\{r\} \quad \text{and} \quad \{r\} = [R]^T\{r'\} \tag{7.5}$$

Note that due to rotation of the coordinate system, the force components also transform in the same way as the displacement components. In fact, rotation of the coordinate system causes the components of all the vectors to transform in the same way.

Now, let us consider a general linear transformation of the components of displacement vector $\{d\}$, not necessarily due to the rotation of coordinate system. The vector $\{d\}$ may, in fact, contain the components of more than one vector. For example, for a three-noded planar element, it may mean $\{u_1 \ v_1 \ u_2 \ v_2 \ u_3 \ v_3\}^T$, representing the two components of displacement at the three nodes in the sequence. Also the components need not represent only linear displacement; they may represent rotations also. In that sense, the displacement vector $\{d\}$ and force vector $\{r\}$ in a coordinate system should be interpreted as generalized displacements and forces. Their inner (dot) product, of course, should give work as result. Let the vector of transformed components of displacement be $\{d'\}$ and let the transformation matrix be $[T]$, i.e., $\{d'\} = [T]\{d\}$, where $[T]$ is not necessarily orthogonal and may not even be square. It can be shown that components of force vector in the original and the new coordinate systems are related as

$$\{r\} = [T]^T \{r'\} \tag{7.6}$$

This can be proved as follows. It is known that the dot product of a force vector and a displacement vector is work, which is a scalar quantity. A scalar quantity remains invariant under the transformation of the coordinate system. If a force whose components are given by the vector $\{r\}$ in a coordinate system is displaced and the displacement components are given by $\{d\}$, then

$$\text{Work done by the force} = \{d\}^T \{r\}$$

If we calculate the work done in the transformed coordinate system, then

$$\text{Work done by the force } \{d'\}^T \{r'\}$$

This implies that

$$\{d\}^T \{r\} = \{d'\}^T \{r'\}$$

Substituting the first relation of Eq. (7.4), in the above inequality, and using $\{d'\} = \{T\}\{d\}$, we get

$$\{d\}^T \{r\} = \{d\}^T [T]^T \{r'\}$$

from which we can write $\{d\}^T\left[\{r\} - [T]^T\{r'\}\right] = 0$. As $\{d\}$ is arbitrary, it implies that the expression $\{r\} - \{T\}^T\{r'\}$ is zero. Therefore, $\{r\} = \{T\}^T\{r'\}$. Thus, we have proved the relation given by Eq. (7.6).

Now, consider the elemental equations in the global x–y–z system and the local x'–y'–z' system

$$[k]\{d\} = r \quad \text{or} \quad [k']\{d'\} = \{r'\} \tag{7.7}$$

As already discussed, if $\{d'\} = [T]\{d\}$, then $\{r\} = [T]^{\mathrm{T}}\{r'\}$. Therefore, we can write

$$[k]\{d\} = \{r\} = [T]^{\mathrm{T}}\{r'\}$$

Replacing $\{r'\}$ in the above expression by the second relation of Eq. (7.7), we obtain

$$[k]\{d\} = [T]^{\mathrm{T}}[k']\{d'\}$$

Replacing $\{d'\}$ by $[T]\{d\}$, we get

$$[k]\{d\} = [T]^{\mathrm{T}}[k'][T]\{d\}$$

Therefore

$$[k] = [T]^{\mathrm{T}}[k'][T] \tag{7.8}$$

This is the equation for transforming the local stiffness equation into the global stiffness equation for any structure, not just for the member of a truss.

For a plane truss, the local stiffness matrix for a member of the truss is given as (refer Chapter 1)

$$[k'] = \frac{AE}{L}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{7.9}$$

where A is the area of cross-section, E is Young's modulus of elasticity, and L is the length of the truss. In the local coordinates, a member of the truss has two degrees of freedom, the nodal axial displacements u_1' and u_2'. In the global coordinate system there are four degrees of freedom u_1, v_1, u_2, and v_2, where u_1 and u_2 denote the displacements of nodes 1 and 2, respectively, along the global x-axis, and v_1 and v_2 denote the displacements of nodes 1 and 2, respectively, along the global y-axis. The local degrees of freedom are related to the global degrees of freedom as

$$\begin{Bmatrix} u_1' \\ u_2' \end{Bmatrix} = [T]_{2\times 4}\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} \tag{7.10}$$

where

$$[T]_{2\times 4} = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ 0 & 0 & \cos\theta & \sin\theta \end{bmatrix} \tag{7.11}$$

In Eq. (7.11), θ is the counterclockwise angle made by the member of the truss with the global x-axis. The formula for transforming the local stiffness matrix into the global stiffness matrix is given as shown in Eq. (7.8).

$$[k] = [T]_{4\times 2}^{\mathrm{T}}[k']_{2\times 2}[T]_{2\times 4}$$

which leads to

$$[k] = \frac{AE}{L}\begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta & -\cos^2\theta & -\cos\theta\sin\theta \\ & \sin^2\theta & -\cos\theta\sin\theta & -\sin^2\theta \\ & & \cos^2\theta & \cos\theta\sin\theta \\ & & & \sin^2\theta \end{bmatrix} \tag{7.12}$$

For a member of space truss also, the local stiffness matrix is given by Eq. (7.9). Here also, in the local coordinates, a member of the truss has two degrees of freedom, the nodal axial displacements u_1' and u_2'. In the global coordinate system, there are six degrees of freedom u_1, v_1, w_1, u_2, v_2, and w_2, where w_1 and w_2 denote the displacements of nodes 1 and 2, respectively, along the global z-axis. The local degrees of freedom are related to the global degrees of freedom as

$$\begin{Bmatrix} u_1' \\ u_2' \end{Bmatrix} = [T]_{2\times 6} \begin{Bmatrix} u_1 \\ v_1 \\ w_1 \\ u_2 \\ v_2 \\ w_2 \end{Bmatrix} \tag{7.13}$$

where

$$[T]_{2\times 6} = \begin{bmatrix} l_1 & m_1 & n_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l_1 & m_1 & n_1 \end{bmatrix} \tag{7.14}$$

Note that l_1, m_1, and n_1 are the direction cosines relating the local x' with the global x-axis. As the local axis is taken along the member, l_1, m_1, and n_1 are in fact the direction cosines of the member itself with respect to global coordinate system. Using the formula for transforming the local stiffness matrix into a global form given in Eq. (7.8), we get

$$[k] = \begin{bmatrix} l_1 & 0 \\ m_1 & 0 \\ n_1 & 0 \\ 0 & l_1 \\ 0 & m_1 \\ 0 & n_1 \end{bmatrix} \frac{AE}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} l_1 & m_1 & n_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & l_1 & m_1 & n_1 \end{bmatrix} \tag{7.15}$$

Equation (7.15) provides a 6×6 global stiffness matrix.

We have described the procedure to obtain the stiffness matrix for a member of the truss in the global form. Let us now consider the force vector. In Chapter 1, we had taken the forces as $-F_i$ and F_j, respectively, for a rod element with nodes i and j. We had adopted the convention of taking tensile forces as positive, which means the force at node i is actually directed from i to j and, thus, produces a negative tensile stress. This convention was helpful in explaining the concept of assembly to a beginner. However, here we adopt the convention of taking positive forces in positive x' direction. Thus, the nodal force vector in local coordinates for an element with nodes 1 and 2 may be written as

$$r' = \begin{Bmatrix} F_{x'1} \\ F_{x'2} \end{Bmatrix} \tag{7.16}$$

We emphasize here that this force vector actually represents two forces: the force at node 1 and that at node 2. For a plane truss, the force components in a local coordinate system can be transformed to a global coordinate system as follows:

$$\{r\} = \begin{bmatrix} \cos\theta & 0 \\ \sin\theta & 0 \\ 0 & \cos\theta \\ 0 & \sin\theta \end{bmatrix} \begin{Bmatrix} F_{x'1} \\ F_{x'2} \end{Bmatrix} = \begin{Bmatrix} (F_x)_1^1 \\ (F_y)_1^1 \\ (F_x)_2^1 \\ (F_y)_2^1 \end{Bmatrix} \tag{7.17}$$

In Eq. (7.17), the number in the superscript indicates the element number. Thus, $(F_x)_1^1$ indicates the force component in the global x-direction at node 1 of element 1.

In the same way, the local and global components of force for a three-dimensional truss member are related in the following way:

$$\{r\} = \begin{bmatrix} l_1 & 0 \\ m_1 & 0 \\ n_1 & 0 \\ 0 & l_1 \\ 0 & m_1 \\ 0 & n_1 \end{bmatrix} \begin{Bmatrix} F_{x'1} \\ F_{x'2} \end{Bmatrix} = \begin{Bmatrix} (F_x)_1^1 \\ (F_y)_1^1 \\ (F_z)_1^1 \\ (F_x)_2^1 \\ (F_y)_2^1 \\ (F_z)_2^1 \end{Bmatrix} \tag{7.18}$$

Here, $(F_z)_1^1$ represents the z-component of force at node 1 and $(F_z)_2^1$ represents the z-component of force at node 2.

7.3 AN EXAMPLE FOR FEM ANALYSIS OF PLANE TRUSSES

Figure 7.2 shows a plane truss. The point 1 is pinned and point 3 is supported on a roller support. The roller can roll in a plane inclined at 45° from the horizontal. We will solve this problem using FEM. We shall assume that appropriate units have been used for all physical quantities and the applied force causes elastic deformation only.

Let for element 1–2, $AE = 3\sqrt{2}$ (where A = area of cross-section and E = Young's modulus), for element 2–3, $AE = 3$, and for element 3–1, $AE = 3$. For element 1–2, $\theta = 135°$, using Eq. (7.12), finite element equations for this element may be written as

$$\frac{3\sqrt{2}}{3\sqrt{2}}\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ & & \frac{1}{2} & -\frac{1}{2} \\ & & & \frac{1}{2} \end{bmatrix}\begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{Bmatrix} = \begin{Bmatrix} (F_x)_1^1 \\ (F_y)_1^1 \\ (F_x)_2^1 \\ (F_y)_2^1 \end{Bmatrix}$$

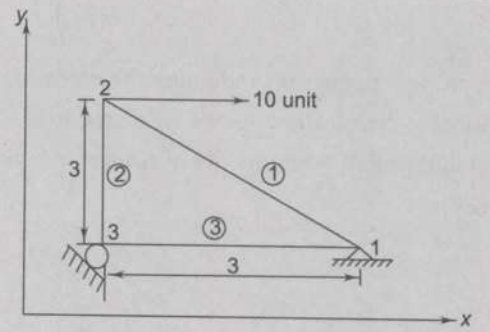


Figure 7.2 *A typical plane truss*

The displacement and force components in the above system of equations are in global coordinate system. For element 2–3, $\theta = -90°$, giving rise to the following system of equations:

$$\frac{3}{3}\begin{bmatrix} 0 & 0 & 0 & 0 \\ & 1 & 0 & -1 \\ & & 0 & 0 \\ & & & 1 \end{bmatrix}\begin{Bmatrix} u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} (F_x)_2^2 \\ (F_y)_2^2 \\ (F_x)_3^2 \\ (F_y)_3^2 \end{Bmatrix}$$

Note that here the global node 2 has been considered as the first local node, while the global node 3 has been considered as the second local node. That is why we have taken $\theta = -90°$. If we choose global node 3 as the first local node and the global node 2 as the second local node, then θ will be 90°. For element 1–3, we choose global node 1 as the first node and global node 3 as the second node. Hence, $\theta = 180°$, giving rise to the following system of equations:

$$\frac{3}{3}\begin{bmatrix} 1 & 0 & -1 & 0 \\ & 0 & 0 & 0 \\ & & 1 & 0 \\ & & & 0 \end{bmatrix}\begin{Bmatrix} u_1 \\ v_1 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} (F_x)_1^3 \\ (F_y)_1^3 \\ (F_x)_3^3 \\ (F_y)_3^3 \end{Bmatrix}$$

Assembling the elemental equations for the three elements of the truss, we obtain

$$\begin{bmatrix} 0.5+1 & -0.5+0 & -0.5 & 0.5 & -1 & 0 \\ -0.5+0 & 0.5+0 & 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0.5+0 & -0.5+0 & 0 & 0 \\ 0.5 & -0.5 & -0.5+0 & 0.5+1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0+0 & 1+0 \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} (F_x)_1^1 + (F_x)_1^3 \\ (F_y)_1^1 + (F_y)_1^3 \\ (F_x)_2^1 + (F_x)_2^2 \\ (F_y)_2^1 + (F_y)_2^2 \\ (F_x)_3^2 + (F_x)_3^3 \\ (F_y)_3^2 + (F_y)_3^3 \end{Bmatrix}$$

For assembly, we first prepare a system of equations for six global degrees of freedom with all the elements in the coefficient matrix and take the right hand side vector as 0. Then, we place the stiffness matrices and right hand side vectors of the three elements in the appropriate locations. The right hand side of the above system of equations is modified by considering the equilibrium of the forces acting at a point. At node 1, $(F_x)_1^1$ denotes the x-component of force acting on element 1 and $(F_x)_1^3$ denotes the x-component of force acting on element 2. The sum of these forces will be equal to the external force acting on the truss at node 1 and will be denoted by $(F_x)_1$. Actually, this force is the horizontal component of the reaction. Similarly, $(F_y)_1^1 + (F_y)_1^3$ will be denoted by $(F_y)_1$. At node 2, there is an external force of magnitude 10 in the horizontal direction. Therefore

$$(F_x)_2^1 + (F_x)_2^2 = 10;\ (F_y)_2^1 + (F_y)_2^2 = 0$$

We shall denote $(F_x)_3^2 + (F_x)_3^3$ by $(F_x)_3$ and $(F_y)_3^2 + (F_y)_3^3$ by $(F_y)_3$, where $(F_x)_3$ and $(F_y)_3$ are the two components of reaction at node 3. Thus, the assembled system of equation can be written as

$$\begin{bmatrix} 0.5+1 & -0.5+0 & -0.5 & 0.5 & -1 & 0 \\ -0.5+0 & 0.5+0 & 0.5 & -0.5 & 0 & 0 \\ -0.5 & 0.5 & 0.5+0 & -0.5+0 & 0 & 0 \\ 0.5 & -0.5 & -0.5+0 & 0.5+1 & 0 & -1 \\ -1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0+0 & 1+0 \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} (F_x)_1 \\ (F_y)_1 \\ 10 \\ 0 \\ (F_x)_3 \\ (F_y)_3 \end{Bmatrix}$$

Now we apply the boundary conditions. As $u_1 = 0$, $v_1 = 0$, we can eliminate the first two rows and columns to get

$$\begin{bmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 1.5 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \begin{Bmatrix} u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} 10 \\ 0 \\ (F_x)_3 \\ (F_y)_3 \end{Bmatrix}$$

At node 3, due to roller support, force along the inclined plane is zero. Hence

$$(F_x)_3 \cos 45° - (F_y)_3 \sin 45° = 0$$

or

$$(F_x)_3 - (F_y)_3 = 0$$

To enforce the above condition, we subtract the fourth row from the third row to get

$$\begin{bmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 1.5 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix} \begin{Bmatrix} u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} 10 \\ 0 \\ 0 \\ (F_y)_3 \end{Bmatrix}$$

We still cannot solve the above system of equation, because $(F_y)_3$ is unknown. However, we can obtain additional equation from the condition that the normal displacement is zero at node 3:

$$u_3 \cos 45° + v_3 \cos 45° = 0$$

that is, $u_3 + v_3 = 0$. We replace the fourth equation by this equation to get

$$\begin{bmatrix} 0.5 & -0.5 & 0 & 0 \\ -0.5 & 1.5 & 0 & -1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{Bmatrix} u_2 \\ v_2 \\ u_3 \\ v_3 \end{Bmatrix} = \begin{Bmatrix} 10 \\ 0 \\ 0 \\ 0 \end{Bmatrix}$$

Solving the above equation, we obtain

$$u_2 = 40, \quad v_2 = 20, \quad u_3 = -10, \quad v_3 = 10$$

Once we have got the nodal displacements in global form we can find out the axial displacements using Eq. (7.10). The difference of axial displacement divided by the length of the member of truss provides the longitudinal strain. Multiplying the longitudinal strain by Young's modulus of elasticity provides the longitudinal stress.

7.4 FEM FORMULATION FOR FRAMES

Elemental equation for a member of a frame is written as

$$\begin{bmatrix} \frac{AE}{L} & 0 & 0 & -\frac{AE}{L} & 0 & 0 \\ 0 & 12\frac{EI}{L^3} & 6\frac{EI}{L^2} & 0 & -12\frac{EI}{L^3} & 6\frac{EI}{L^2} \\ 0 & 6\frac{EI}{L^2} & 4\frac{EI}{L} & 0 & -6\frac{EI}{L^2} & 2\frac{EI}{L} \\ -\frac{AE}{L} & 0 & 0 & \frac{AE}{L} & 0 & 0 \\ 0 & -12\frac{EI}{L^3} & -6\frac{EI}{L^2} & 0 & 12\frac{EI}{L^3} & -6\frac{EI}{L^2} \\ 0 & 6\frac{EI}{L^2} & 2\frac{EI}{L} & 0 & -6\frac{EI}{L^2} & 4\frac{EI}{L} \end{bmatrix} \begin{Bmatrix} u_1' \\ v_1' \\ v_{1,x'}' \\ u_2' \\ v_2' \\ v_{2,x'}' \end{Bmatrix} = \begin{Bmatrix} F_{x'1} \\ -V_1 \\ -M_1 \\ F_{x'2} \\ V_2 \\ M_2 \end{Bmatrix} \tag{7.19}$$

or
$$[k']\{d'\} = \{r'\} \tag{7.20}$$

where $\{r'\}$ is the generalized force vector in local coordinates. Note that Eq. (7.19) has been obtained by combining the equations of rod and beam assuming that there is no coupling between the axial and transverse displacement. As per our sign convention, the positive vertical force in a plane whose normal is along the positive x-direction produces a positive shear force. Similarly, a counter-clockwise moment on that plane produces a positive moment. Similarly, on a plane whose normal is along negative x-direction, a positive vertical force produces a negative shear force and a counter-clockwise moment produces a negative bending moment. Keeping this sign convention in mind, we can represent the generalized force vector as

$$\{r'\} = \begin{Bmatrix} F_{x'1} \\ F_{y'1} \\ M_{z'1} \\ F_{x'2} \\ F_{y'2} \\ M_{z'2} \end{Bmatrix} \tag{7.21}$$

where the forces in the direction of positive axes and counterclockwise moments are considered positive.

We denote the displacement vector in the rotated coordinate system by

$$\{d'\} = \begin{bmatrix} u_1' & v_1' & v_{1,x'}' & u_2' & v_2' & v_{2,x'}' \end{bmatrix}^{\mathrm{T}} \tag{7.22}$$

Note that sometimes, a dash (′) is used to represent differentiation. However, here it refers to rotated coordinates and a comma in the subscript denotes differentiation. In the global coordinate system, the displacement vector is given by

$$\{d\} = \begin{bmatrix} u_1 & v_1 & v_{1,x} & u_2 & v_2 & v_{2,x} \end{bmatrix}^{\mathrm{T}} \tag{7.23}$$

Both the displacement vectors are related to each other in the following manner:

$$\begin{Bmatrix} u_1' \\ v_1' \\ v_{1,x'}' \\ u_2' \\ v_2' \\ v_{2,x'}' \end{Bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \cos\theta & \sin\theta & 0 \\ 0 & 0 & 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ v_{1,x} \\ u_2 \\ v_2 \\ v_{2,x} \end{Bmatrix} \tag{7.24}$$

where θ is the angle which the frame element makes from the global x-axis. Note that rotation of axis system does not change slopes. Thus

$$v_{1,x'}' = v_{1,x} \tag{7.25}$$

Actually

$$v'_{1,x'} = \lim_{\Delta x \to 0} \frac{\Delta v'_1}{\Delta x'} = \lim_{\Delta x \to 0} \frac{\Delta v_1 \left(-\sin\theta + \cos\theta\right)}{\Delta x\left(-\sin\theta + \cos\theta\right)} = v_{1,x} \tag{7.26}$$

Equation (7.24) can be written as

$$\{d'\} = [T]\{d\} \tag{7.27}$$

Then, the stiffness matrix in global system is obtained as

$$[k] = [T]^{\mathrm{T}}\,[k'][T] \tag{7.28}$$

The remaining procedure is same as that described in the case of a truss. At a joint, the sum of all internal forces is equal to the external force. This fact is used during assembly. In case of rigid joints, displacements as well as rotations at a joint are same for all the members. If two members are connected by pin or spherical joints, the displacements are common, but rotations differ. In such cases, separate additional degrees of freedom for rotations have to be assigned to the members during assembly procedure. There may be situations in which a load acts on a point other than a joint. In that case, the loads can be easily transferred to joints as is done in beams and rods. For example, if an axial load of intensity $q_{x'}$ acts on a frame member, it provides the following load vector:

$$\text{load vector} = \int_0^L q_{x'} \begin{Bmatrix} 1 - \dfrac{x'}{L} \\ \dfrac{x'}{L} \end{Bmatrix} \mathrm{d}x \tag{7.29}$$

The first element of the load vector corresponds to the transferred load at joint at $x' = 0$ and the second element corresponds to the transferred load at joint at $x' = L$. The loads will be along x' direction. Similarly, if a transverse load of intensity $q_{y'}$ acts on the frame member, it provides the following load vector:

$$\text{load vector} = \int_0^L q_{y'} \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \mathrm{d}x \tag{7.30}$$

where N_1, N_2, N_3, and N_4 are the shape functions of a beam element. The first and third elements of the load vector correspond to forces at joints $x' = 0$ and $x' = L$ respectively. The direction of the forces is along y'. The second and fourth elements of the load vectors correspond to counterclockwise moments at the joints. The concentrated force and moments can be converted to load intensity by using Dirac-delta and doublet functions, respectively. It was explained in detail in the Solved Problem 6.1 of Chapter 6.

7.5 ASSEMBLY PROCEDURE

Usually, the truss and frame problems are solved on computer. Therefore, one needs an automatic assembly procedure. This can be done using the concept of connectivity matrix introduced in Chapter 1. The connectivity matrix provides the global number of each local node of an element. The row of the connectivity matrix indicates a row number and column the local node number. If C is a connectivity matrix, the element $C(i, j)$ indicates the global node number of the jth node of the ith element. The connectivity matrix for the structure in Fig. 7.2 is written as

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 1 & 3 \end{bmatrix} \tag{7.31}$$

If we provide the global coordinates of all nodes along with the connectivity matrix, one can graphically construct the structure. The connectivity matrix is not unique. For example, for the structure shown in Fig. 7.2, one can write the connectivity matrix as

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix} \tag{7.32}$$

Here, there is a slight change in the last row. The global number of first local node in this case is 3 instead of 1. This does not pose any problem as long as we recognize the node of a member that is being called the first node. Assembled equations in both cases will remain the same. For the connectivity matrix of Eq. (7.31), the third member makes an angle of 180° from the x-axis, while it makes an angle of 0° for the connectivity matrix of Eq. (7.32).

The assembly procedure involves picking the stiffness matrix and right hand side vector of each element (member) one by one and placing the entries in the appropriate position in the global stiffness matrix and right hand side vector. To understand how the procedure can be implemented in a computer code, one should understand the concept of degree of freedom of a node. In a plane truss, each node has two degrees of freedom (the displacements in x and y directions). In a three-dimensional truss, each node has three degrees of freedom (the displacements in three directions). In a plane frame, each member has three degrees of freedom (two displacements and one slope). A space frame has six degrees of freedom (three displacements and three slopes). If the number of degree of freedom per node is d_f, then in the elemental equations of element i, the rows and columns will correspond to jth node as per the following equation:

$$\text{Rows and column corresponding to } j^{\text{th}} \text{ node} = (j-1)d_f + 1, (j-1)d_f + 2, \ldots, jd_f \tag{7.33}$$

If j^{th} node of element i corresponds to p^{th} global node, then corresponding global rows and columns will be as per the following equation:

$$\text{Global rows and column correspondings} = (p-1)d_f + 1, (p-1)d_f + 2, \ldots, pd_f \tag{7.34}$$

where

$$p = C(i, j) \tag{7.35}$$

Now we can develop an algorithm to carry out the assembly as per the following steps:

Step 1 Initialize the global stiffness and global right hand side vector to 0. In the C language, the following statements will do it:

```
for (i=1; i<=n; i=i+1) /*outer loop*/
{
for (j=1; j<=n; j+1) /*inner loop*/
{
gstiff [i][j] = 0.0;
}
grhs[i]=0.0;
}
```

Note that in C language, i and j can start from 0, but we have started them from 1 in order to avoid confusion. Similarly, instead of writing $i = i + 1$, we can write i++. However, our strategy in this book will be to use a simplified form of programming, so that those having only preliminary knowledge of C language can also understand the basic concepts.

Step 2 Start from the first element. Place each element (entry) of elemental stiffness matrix and elemental right hand side vector into the appropriate location in the global stiffness matrix with the help of the connectivity matrix. If there is already some non-zero quantity in a location of global stiffness matrix, the new quantity is added to it. This has to be done for all the elements. (Assume that there are I elements.) In C language, the following statements will implement step 2:

```
for (i=1; i<=I; i=i+1) /*outer loop*/
{
stiffness(i); /*calling stiffness matrix of element i*/
righthand(i); /*calling right hand side vector of element i*/

for (j=1; j<=2; j=j+1) /*middle loop, df is the degree of freedom
per node*/
{
p=C[i][j];
for (j1=1; j1<=df; j1=j1+1)
{
grhs[(p-1)*df+j1]=grhs[(p-1)*df+j1]+erhs[(j-1)*df+j1)]; /*placing
elemental entries into global location*/
}
for (k=1; k<=2; k=k+1) /*inner loop */
```

```
{
q=C[i][k];
for (k1=1; k1<=df;  k1=k1+1)
{
for(k2=1; k2<=df; k2=k2+1)
{
gstiff[(p-1)*df+k1, (q-1)*df+k2]= gstiff[(p-1)*df+k1,
(q-1)*df+k2]+estiff[(j-1)*df+k1, (k-1)*df+k2]; /*placing
elemental entries into global location*/
}}}}}
```

This procedure is valid when at a particular node the values for all degrees of freedom are same for all elements. If some values differ (for example, at a pin joint, slopes for the two elements may be different), then a similar procedure may be used. One just has to find the location in the global system for a particular elemental entry.

7.6 CONCLUSION

In this chapter, the FEM formulations for trusses and frames are discussed in detail. We have solved one example for the truss problem. For the frames, we have not solved any problem till now. We shall solve one sample problem. As the size of the assembled stiffness matrix in frame problems is usually large, it would be better if these problems are solved using a computer.

SOLVED PROBLEMS

7.1 To reduce the deflection of a cantilever beam of diameter 5 cm and length 1 m, a support is provided by a rod of the same diameter, which joins the tip of the cantilever beam with the wall as shown in Fig. 7.3. Using FEM, obtain the deflections of the tip of a beam loaded with a 1 kN applied load, with and without support. Take the material of the beam and rod as aluminum having $E = 69$ GPa.

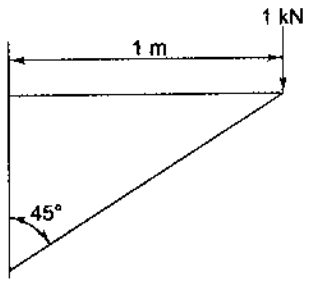


Figure 7.3 *A beam fixed at one end and supported by a rod at the other end*

Solution First consider that there is no support. In this case, one can solve it as a beam problem. However, we take the beam as a frame element. Using Eq. (7.19), elemental equations are written as

$$10^4\begin{bmatrix} 13548 & 0 & 0 & -13548 & 0 & 0 \\ 0 & 25.4028 & 12.7014 & 0 & -25.408 & 12.7041 \\ 0 & 12.7041 & 8.4676 & 0 & -12.7014 & 4.2338 \\ -13548 & 0 & 0 & 13548 & 0 & 0 \\ 0 & -25.4028 & -12.7014 & 0 & 25.4028 & -12.7014 \\ 0 & 12.704 & 4.2338 & 0 & -12.7014 & 8.4676 \end{bmatrix}\begin{Bmatrix} u_1' \\ v_1' \\ v_{1,x'}' \\ u_2' \\ v_2' \\ v_{2,x'}' \end{Bmatrix}=\begin{Bmatrix} F_{x'1} \\ -V_1 \\ -M_1 \\ 0 \\ 1000 \\ 0 \end{Bmatrix} \tag{7.1.1}$$

Invoking the boundary condition at the fixed end, the first three rows and columns can be eliminated providing

$$10^4\begin{bmatrix} 13548 & 0 & 0 \\ 0 & 25.4028 & -12.7014 \\ 0 & -12.7014 & 8.4676 \end{bmatrix}\begin{Bmatrix} u_2' \\ v_2' \\ v_{2,x'}' \end{Bmatrix}=\begin{Bmatrix} 0 \\ 1000 \\ 0 \end{Bmatrix}$$

Solving Eq. (7.1.2), we get $u_2' = 0$ and $v_2' = 15.76 \times 10^{-3}$ m. Thus, the vertical deflection at the end is 15.76 mm.

In the presence of a support, we can treat the system as a problem of a frame with two members. In the global form, the stiffness matrix of the first element is given by Eq. (7.1.1). For the second element, we consider the fixed node as the first node and obtain the local stiffness matrix as

$$10^4\begin{bmatrix} 9580 & 0 & 0 & -9580 & 0 & 0 \\ 0 & 8.9808 & 6.3507 & 0 & -8.9808 & 6.3507 \\ 0 & 6.3507 & 5.9876 & 0 & -6.3507 & 2.9938 \\ -9580 & 0 & 0 & 9580 & 0 & 0 \\ 0 & -8.9808 & -6.3507 & 0 & 8.9808 & -6.3507 \\ 0 & 6.3507 & 2.9938 & 0 & -6.3507 & 5.9876 \end{bmatrix} \tag{7.1.3}$$

This has to be transformed to global system, the transformation matrix for which is

$$\begin{bmatrix} 0.7071 & 0.7071 & 0 & 0 & 0 & 0 \\ -0.7071 & 0.7071 & 0 & 0 & -0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.7071 & 0.7071 & 0 \\ 0 & 0 & 0 & -0.7071 & 0.7071 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{7.1.4}$$

Using Eq. (7.28), the stiffness matrix in the global system is

$$10^4\begin{bmatrix} 4794 & 4785 & -4.5 & 4794 & 4785 & 4.5 \\ 4785 & 4794 & 4.5 & -4785 & -4794 & 4.5 \\ -4.5 & 4.5 & 6 & 4.5 & -4.5 & 3 \\ -4794 & -4785 & 4.5 & 4794 & 4785 & 4.5 \\ -4785 & -4794 & -4.5 & 4785 & 4794 & -4.5 \\ -4.5 & 4.5 & 3 & 4.5 & -4.5 & 6 \end{bmatrix} \tag{7.1.5}$$

We can assemble the stiffness matrices of the two elements to obtain a 9×9 global matrix. However, nodes 1 and 3 are fixed and, so, we can eliminate the corresponding rows and columns. Finally, we get the following set of equations

$$10^8\begin{bmatrix} 1.4374 & 0.4785 & 0.0004 \\ 0.4785 & 0.4803 & -0.0011 \\ 0.0004 & -0.0011 & 0.0012 \end{bmatrix}\begin{Bmatrix} u_2 \\ v_2 \\ v_{2,x} \end{Bmatrix} = \begin{Bmatrix} 0 \\ 1000 \\ 0 \end{Bmatrix} \tag{7.1.6}$$

This leads to $v_2 = 0.03128$ mm. Thus, we see that the transverse displacement gets drastically reduced.

EXERCISES

A. Multiple Choice Questions

7.1 For a particle in plane motion, the x and y displacement components are 3 and 4, respectively, in the axis system x–y. Another axis system, x'–y', is obtained by rotating the system x–y by 30° in the counterclockwise direction. The x' and y' components of the displacement of the particle are

(a) $\frac{3\sqrt{3}}{2}+2, -2+\frac{3\sqrt{3}}{2}$

(b) $-2+\frac{3\sqrt{2}}{2}, \frac{3\sqrt{3}}{2}+2$

(c) $\frac{3}{2}+2\sqrt{3}, -2\sqrt{3}+\frac{3}{2}$

(d) $\frac{3\sqrt{3}}{2}+2, 2\sqrt{3}-\frac{3}{2}$

7.2 Consider the following rotation matrix:

$$\begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix}$$

where l_i, m_i, and n_i represent the direction cosines of ith axis with respect to a global axis system. The inverse of the rotation matrix is given by

(a) $\begin{bmatrix} l_1 & m_1 & n_1 \\ l_2 & m_2 & n_2 \\ l_3 & m_3 & n_3 \end{bmatrix}$ (b) $\begin{bmatrix} l_1 & l_2 & l_3 \\ m_1 & m_2 & m_3 \\ n_1 & n_2 & n_3 \end{bmatrix}$

(c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} l_3 & m_3 & n_3 \\ l_2 & m_2 & n_2 \\ l_1 & m_1 & n_1 \end{bmatrix}$

7.3 Choose the incorrect statement from the following:

(a) The members of an ideal truss carry out only axial loads.

(b) Frames never have pin joints.

(c) Statically indeterminate frames can be analyzed by finite element method.

(d) Stiffness matrix for a truss is never negative definite.

7.4 The connectivity matrix for a truss is given as follows:

$$C = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{bmatrix}$$

The global stiffness matrix is $[K]$ and the elemental stiffness matrices are $[k^1]$, $[k^2]$, and $[k^3]$, respectively. The element (entry) at fifth row and sixth column of the global stiffness matrix is given by

(a) $K(5, 6) = k^2(3, 4) + k^3(1, 2)$ (b) $K(5, 6) = k^2(3, 4) + k^3(3, 4)$

(c) $K(5, 6) = k^2(1, 2) + k^3(3, 4)$ (d) $K(5, 6) = k^2(3, 4) + k^3(3, 4)$

7.5 The angle between two lines is 30° in a coordinate system. With respect to another coordinate system that is obtained by rotating the present coordinate system 20° in the counterclockwise direction, the angle between the lines will be

(a) 10° (b) 20°

(c) 30° (d) 50°

B. Problems

7.6 Solve the truss problem discussed in the chapter again and find the axial forces in all the elements. Also find the support reactions.

7.7 Figure 7.4 represents a five-member truss made of steel. The area of cross-section each member is 6 cm^2. It is loaded by a vertical force of 100 N at the top point at extreme right. Find the deflection of that point.

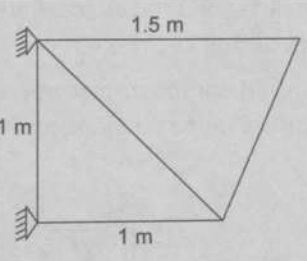


Figure 7.4 *A five-member truss*

7.8 All members of the frame shown in Fig. 7.5 have equal length. Assume the length, A, and EI to be unity. If a unit load is applied as shown, find out the deflection at the point of application of the load and maximum stress in the frame.

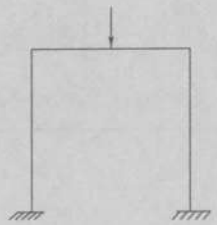


Figure 7.5 *A frame with members of equal length*

7.9 In Fig. 7.6, a cantilever beam is supported on a stepped rod. Consider one beam element and two rod elements to solve the problem. The stiffness matrix for beam element is given by

$$\frac{EI}{h^3}\begin{bmatrix} 12 & 6h & -12 & 6h \\ 6h & 4h^2 & -6h & 2h^2 \\ -12 & -6h & 12 & -6h \\ 6h & 2h^2 & -6h & 4h^2 \end{bmatrix}$$

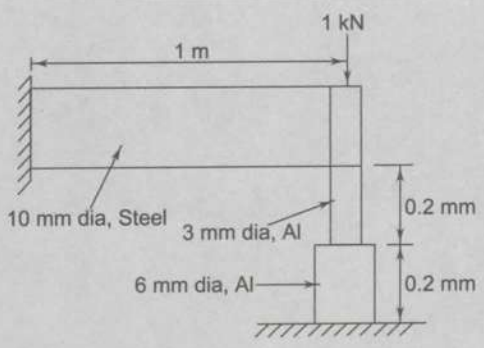


Figure 7.6 *A cantilever beam supported on a stepped rod*

If you remember the stiffness matrix for a rod element, use it. Else, derive the matrix. Solve this problem to find out the deflections and stresses.

7.10 For the truss shown in Fig. 7.7, find out the displacements at 2, 3, and 4 when a vertical load P is applied at 4. Find out the stresses in the all elements.

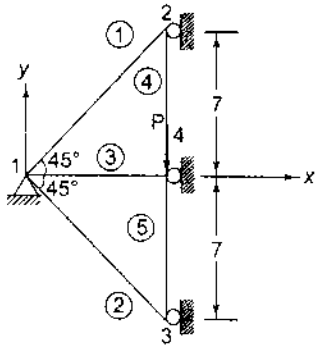


Figure 7.7 *The truss discussed in Problem 7.10*

8

CHAPTER

Introduction to Two-Dimensional and Three-Dimensional FEM

8.1 Introduction
8.2 Triangular Elements
8.3 Tetrahedral Elements
8.4 Rectangular Elements
8.5 Brick Elements
8.6 Governing Differential Equation for Two-Dimensional Heat Conduction
8.7 Weak Form and FEM Formulation
8.8 A Note on the Assembly in Two Dimensions
8.9 Poisson Equation in Three Dimensions
8.10 Fluid Flow Problems
8.11 Torsion of Circular and Non-circular Cross-section
8.12 A Note on Exploiting Symmetry in FEM
8.13 Summary

8.1 INTRODUCTION

So far we have discussed problems that involved the solution of one-dimensional differential equations only. Now, we move towards solving physical problems involving two-dimensional and three-dimensional differential equations. In this chapter, we shall study the FEM formulation for Laplace and Poisson equations. We shall start this chapter with a discussion on different types of elements and then develop the finite element equations. The detailed derivation of two-dimensional equations will also be provided. The reader can easily extend the procedure to obtain FEM equations for problems involving three-dimensional equations.

8.2 TRIANGULAR ELEMENTS

In one dimension, a finite element can be described by two nodes and the line passing through these nodes is said to form a *line element*. On the other hand, in two dimensions, a minimum of three nodes are required to make an element. These nodes should be able to form a triangle.

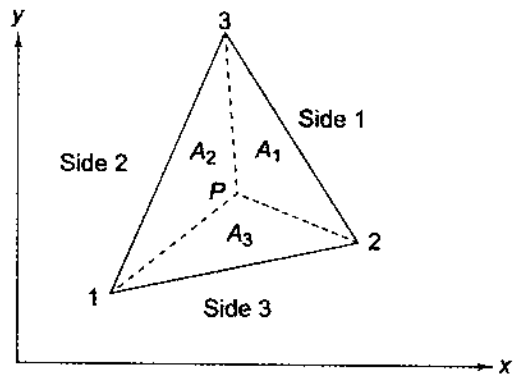


Figure 8.1 *A triangular element*

Figure 8.1 shows a three-noded triangular element in two dimensions. It has three nodes 1, 2, and 3. P is any internal point within the element. Point P and any two nodes of the element make a triangle such that triangle $P23$ has area A_1, triangle $P13$ has area A_2, and triangle $P12$ has area A_3. Let us consider that we have to interpolate the temperature T within the element. We need to express the constants in terms of nodal temperatures. As there are three nodes, we can determine three constants. So what type of function should be taken? For better visualization, one can construct a triangle similar to Pascal's triangle, as shown in Fig. 8.2.

$$
\begin{array}{ccccc}
 & & 1 & & \\
 & x & & y & \\
 & x^2 & xy & y^2 & \\
x^3 & x^2y & xy^2 & y^3 & \\
x^4 & x^3y & x^2y^2 & xy^3 & y^4 \\
x^5 & x^4y & x^3y^2 & x^2y^3 & xy^4 \quad y^5
\end{array}
$$

Figure 8.2 *Triangle used for choosing an appropriate interpolation function*

In Fig. 8.2, the first row contains only one term, i.e., 1, in which the sum of the powers of x and y is 0. The second row contains two terms, i.e., x and y, in which the sum of the powers of x and y is 1. The third row contains the terms in which the powers of x and y add to give 2 and so on. For three-noded elements, we choose the first two rows of the triangle. Thus

$$T = a + bx + cy \tag{8.1}$$

This linear equation must be satisfied at all the nodes. Therefore

$$T_1 = a + bx_1 + cy_1, \quad T_2 = a + bx_2 + cy_2, \quad T_3 = a + bx_3 + cy_3 \tag{8.2}$$

Equation (8.2) can be written in the matrix form as follows:

$$\begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix} \begin{Bmatrix} a \\ b \\ c \end{Bmatrix} = \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} \tag{8.3}$$

From Eq. (8.1), we get

$$T = [1 \ x \ y] \begin{Bmatrix} a \\ b \\ c \end{Bmatrix} = [1 \ x \ y] \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{bmatrix}^{-1} \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} = [N_1 \ N_2 \ N_3] \begin{Bmatrix} T_1 \\ T_2 \\ T_3 \end{Bmatrix} \tag{8.4}$$

where N_1, N_2, and N_3 are *shape functions* given as follows:

$$
\begin{aligned}
N_1 &= \frac{(xy_2 - x_2y) + (x_2y_3 - x_3y_2) + (x_3y - y_3x)}{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - y_3x_1)} = \frac{\text{area}(P23)}{\text{area}(123)} \\
N_2 &= \frac{(x_1y - xy_1) + (xy_3 - x_3y) + (x_3y_1 - y_3x_1)}{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - y_3x_1)} = \frac{\text{area}(P31)}{\text{area}(123)} \\
N_3 &= \frac{(x_1y_2 - x_2y_1) + (x_2y - xy_2) + (xy_1 - yx_1)}{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - y_3x_1)} = \frac{\text{area}(P12)}{\text{area}(123)}
\end{aligned}
\tag{8.5}
$$

where P is the point having coordinates (x, y). It is clearly seen that the sum of the shape functions is 1. We can define the natural coordinates of a point (x, y) as

$$\xi_1 = \frac{(xy_2 - x_2y) + (x_2y_3 - x_3y_2) + (x_3y - y_3x)}{(x_1y_2 - x_2y_1) + (x_2y_3 - x_3y_2) + (x_3y_1 - y_3x_1)} = \frac{\text{area}(P23)}{\text{area}(123)}$$

$$\xi_2 = \frac{(x_1 y - x y_1) + (x y_3 - x_3 y) + (x_3 y_1 - y_3 x_1)}{(x_1 y_2 - x_2 y_1) + (x_2 y_3 - x_3 y_2) + (x_3 y_1 - y_3 x_1)} = \frac{\text{area}(P31)}{\text{area}(123)} \tag{8.6}$$

$$\xi_3 = \frac{(x_1 y_2 - x_2 y_1) + (x_2 y - x y_2) + (x y_1 - y x_1)}{(x_1 y_2 - x_2 y_1) + (x_2 y_3 - x_3 y_2) + (x_3 y_1 - y_3 x_1)} = \frac{\text{area}(P12)}{\text{area}(123)}$$

The natural coordinates of every point will be bounded between 0 and 1 and

$$\xi_1 + \xi_2 + \xi_3 = 1 \tag{8.7}$$

Thus, there are only two independent natural coordinates. Note that the equations given in Eq. (8.6) imply

$$x = \xi_1 x_1 + \xi_2 x_2 + \xi_3 x_3 \tag{8.8a}$$

$$y = \xi_1 y_1 + \xi_2 y_2 + \xi_3 y_3 \tag{8.8b}$$

It is seen that the physical coordinates are linear functions of the natural coordinates and vice versa. For a three-noded triangular element

$$N_1 = \xi_1; \quad N_2 = \xi_2; \quad N_3 = \xi_3 \tag{8.9}$$

The three-noded linear triangular element is also called a *constant strain triangle* (CST) element, because if we approximate the displacement components using this element, the strains will be constant.

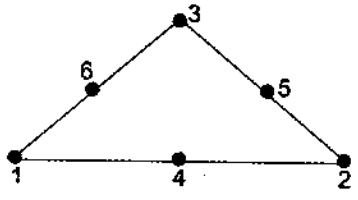


Figure 8.3 *A six-noded triangular element*

A six-noded triangular element is shown in Fig. 8.3 in which the nodes 1, 2, and 3 form the vertices of a triangle (such nodes are called *corner nodes*) and nodes 4, 5, and 6 lie between nodes 1 and 2, 2 and 3, and 3 and 1, respectively (called *mid-side nodes*). Quadratic interpolation can be carried out for this element. Taking the first three rows from the triangle shown in Fig. 8.2, T can be approximated as

$$T = a + bx + cy + dx^2 + exy + fy^2 \tag{8.10}$$

We can find out the shape functions following the same procedure as we used for a three-noded element. However, here we derive the shape functions in a different way using natural coordinates, realizing that shape functions will be quadratic functions in physical as well as natural coordinate systems. In natural coordinates, the shape functions corresponding to i^{th} node are written as

$$N_i = a_i \xi_1^2 + b_i \xi_2^2 + c_i \xi_3^2 + d_i \xi_1 \xi_2 + e_i \xi_2 \xi_3 + f_i \xi_1 \xi_3 \tag{8.11}$$

Note that in Eq. (8.11) the sum of the powers of natural coordinates is 2 in each term. One may wonder why a constant term or a term containing natural coordinates of degree 1 is not present. This is

because even if a constant term is present, it can be multiplied by $(\xi_1 + \xi_2 + \xi_3)^2$, which is equal to 1. Similarly, a term containing a single-degree natural coordinate can be multiplied by $(\xi_1 + \xi_2 + \xi_3)$. Thus, ultimately, we can obtain the form given in Eq. (8.11).

Let us first derive the expression for N_4, which can be written as

$$N_4 = a_4\xi_1^2 + b_4\xi_2^2 + c_4\xi_3^2 + d_4\xi_1\xi_2 + e_4\xi_2\xi_3 + f_4\xi_1\xi_3 \tag{8.12}$$

It is '1' at node 4 and '0' at all other nodes. At node 1, $\xi_1 = 1$, $\xi_2 = 0$, and $\xi_3 = 0$. Thus,

$$a_4 = 0 \tag{8.13}$$

Similarly, it can be shown that

$$b_4 = c_4 = 0 \tag{8.14}$$

At node 5, $\xi_2 = \xi_3 = \frac{1}{2}$, thus

$$\frac{1}{4}e_4 = 0 \quad \text{or} \quad e_4 = 0 \tag{8.15}$$

At node 6, $\xi_1 = \xi_3 = \frac{1}{2}$, thus

$$\frac{1}{4}f_4 = 0 \quad \text{or} \quad f_4 = 0 \tag{8.16}$$

Finally, at node 4, $\xi_1 = \xi_2 = \frac{1}{2}$, thus

$$\frac{1}{4}d_4 = 1 \quad \text{or} \quad d_4 = 4 \tag{8.17}$$

Therefore

$$N_4 = 4\xi_1\xi_2 \tag{8.18}$$

Similarly, it can be shown (rather it can be directly written by noting the similarity) that

$$N_5 = 4\xi_2\xi_3 \quad \text{and} \quad N_6 = 4\xi_1\xi_3 \tag{8.19}$$

Let us now derive the expression for N_1, which can be written as

$$N_1 = a_1\xi_1^2 + b_1\xi_2^2 + c_1\xi_3^2 + d_1\xi_1\xi_2 + e_1\xi_2\xi_3 + f_1\xi_1\xi_3 \tag{8.20}$$

This shape function is 0 at all nodes expect at node 1, where its value is 1.

At node 2, $\xi_1 = \xi_3 = 0$ and $\xi_2 = 1$, which gives

$$b_1 = 0 \tag{8.21}$$

At node 3, $\xi_1 = \xi_2 = 0$ and $\xi_3 = 1$, which gives

$$c_1 = 0 \tag{8.22}$$

At node 5, $\xi_1 = 0$, $\xi_2 = \xi_3 = 1/2$, which gives

$$e_1 = 0 \tag{8.23}$$

At node 1, $\xi_2 = \xi_3 = 0$ and $\xi_1 = 1$, which gives

$$a_1 = 0 \tag{8.24}$$

At node 4, $\xi_1 = \xi_2 = 1/2$ and $\xi_3 = 0$, which gives

$$\frac{a_1}{4}+\frac{d_1}{4}=0 \text{ or } d_1=-1 \tag{8.25}$$

Lastly, at node 6, $\xi_1 = \xi_3 = 1/2$ and $\xi_2 = 0$, thus giving

$$f_1=-1 \tag{8.26}$$

Therefore, the expression for N_1 is obtained as

$$N_1=\xi_1^2-\xi_1\xi_2-\xi_1\xi_3 \tag{8.27}$$

Equation (8.27) can be simplified as

$$N_1=\xi_1^2-\xi_1(\xi_2+\xi_3)=\xi_1^2-\xi_1(1-\xi_1)=\xi_1(2\xi_1-1) \tag{8.28}$$

By analogy, we can write

$$N_2=\xi_2(2\xi_2-1) \text{ and } N_3=\xi_3(2\xi_3-1) \tag{8.29}$$

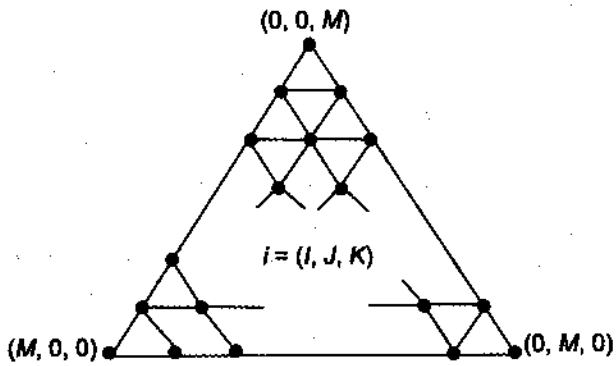


Figure 8.4 *A triangular element interpolating a polynomial of degree M*

Thus, we have derived the shape functions for a linear (degree 1) and a quadratic (degree 2) element. Now, let us derive the expression for an interpolation polynomial of degree M. A similar procedure needs to be followed. However, Argyris, et al. (1968) and some others, such as Silvester (1969) and Taylor (1972), have developed a simpler procedure in which each node i is represented by the triplet (I, J, K). For example, the corner nodes in the baseline are written as $(M, 0, 0)$ and $(0, M, 0)$. On the baseline, the first coordinate keeps decreasing by one from left to right as we move from one node to the other, while the second coordinate keeps increasing by 1. For any node

$$I+J+K=M \tag{8.30}$$

A simple way of obtaining a triplet corresponding to a node is to assume that the nodes are equispaced on sides (they need not be actually in physical coordinates), and then multiplying the natural coordinates of the nodes by M. For a node having designation (I, J, K), the shape function is given by

$$N_i=l_I^I(\xi_1)l_J^J(\xi_2)l_K^K(\xi_3) \tag{8.31}$$

where

$$l_n^n(x)=\frac{(x-x_0)(x-x_1)\ldots(x-x_{n-1})}{(x_n-x_0)(x_n-x_1)\ldots(x_n-x_{n-1})} \tag{8.32}$$

where starting from $x_0 = 0$, x_is denote the one-dimensional natural coordinates. (You may divide a line with end coordinates 0 and 1 into M equal parts to obtain the coordinates $x_0, x_1, \ldots, x_M$. The coordinate of x_M will be 1.) Take $l_0^0 = 1$. Using this method for the cubic triangle shown in Fig. 8.5, we get

$$N_1 = \frac{1}{2}(3\xi_1 - 1)(3\xi_1 - 2)\xi_1, \quad N_4 = \frac{9}{2}\xi_1\xi_2(3\xi_1 - 1), \quad N_{10} = 27\xi_1\xi_2\xi_3 \tag{8.33}$$

where N_1 is the corner node, N_4 is the node adjacent to node 1 on the line connecting nodes 1 and 2, and N_{10} is the internal node. For a detailed derivation of Eq. (8.33), see Solved Problem 8.1.

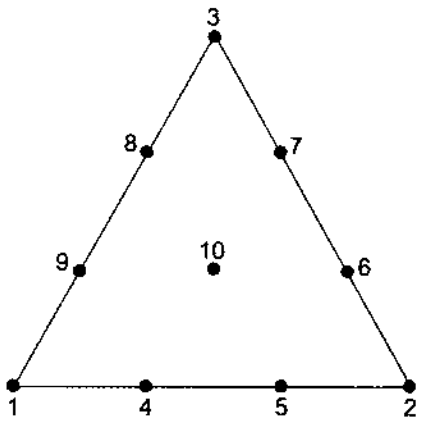


Figure 8.5 *A cubic triangular element*

8.3 TETRAHEDRAL ELEMENTS

Tetrahedral elements are three-dimensional elements. The formulae for these elements may be derived in the same manner as discussed in the previous section. A linear tetrahedral element has 4 nodes, a quadratic element has 10 nodes, and a cubic element has 20 nodes. Figures 8.6(a) and (b) shows two tetrahedral elements having 4 nodes and 10 nodes, respectively. We can define the coordinates of a point P in the element by

$$\xi_1 = \frac{\text{Volume P234}}{\text{Volume 1234}}, \quad \xi_2 = \frac{\text{Volume P134}}{\text{Volume 1234}}, \quad \xi_3 = \frac{\text{Volume P124}}{\text{Volume 1234}}, \quad \xi_4 = \frac{\text{Volume P123}}{\text{Volume 1234}} \tag{8.34}$$

The shape functions for linear elements are given by

$$N_i = \xi_i \tag{8.35}$$

For a quadratic tetrahedron, we have the following relation for the typical nodes:

For the corner node 1, the shape function can be written as

$$N_1 = (2\xi_1 - 1)\xi_1 \tag{8.36}$$

For the mid-side node 5, the shape function is given by

$$N_5 = 4\xi_1\xi_2 \tag{8.37}$$

You can easily write the shape functions for the other nodes.

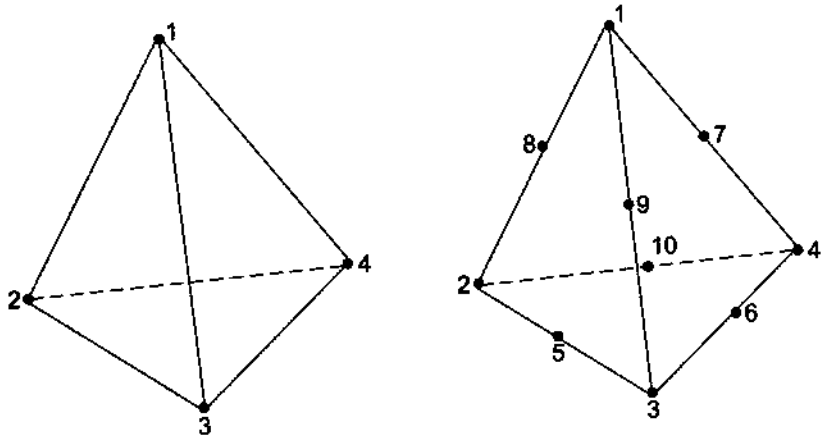


Figure 8.6 *Tetrahedral elements having (a) 4 nodes (b) 10 nodes*

8.4 RECTANGULAR ELEMENTS

Figure 8.7 shows two rectangular elements with 4 and 9 nodes. With the help of a four-noded element, a variable, say temperature T, may be interpolated as

$$T = a + bx + cy + dxy = (A + Bx)(C + Dy) \tag{8.38}$$

Thus, the interpolation function is the product of two linear functions in x and y. We can obtain the constants a, b, c, and d (or A, B, C, and D) in terms of the nodal values T_1, T_2, T_3, and T_4 and express the variable T in the following form:

$$T = N_1T_1 + N_2T_2 + N_3T_3 + N_4T_4 \tag{8.39}$$

However, the shape functions can be easily obtained if we consider the shape function at a node to be the product of one-dimensional shape functions along x and y. Thus, if $x_1 = x_4 = p$, $x_2 = x_3 = q$ and $y_1 = y_2 = r$, $y_3 = y_4 = s$, then the shape functions would be written as

$$\begin{aligned} N_1 &= \frac{(x-q)}{(p-q)}\frac{(y-s)}{(r-s)}; \quad N_2 = \frac{(x-p)}{(q-p)}\frac{(y-s)}{(r-s)} \\ N_3 &= \frac{(x-p)}{(q-p)}\frac{(y-r)}{(s-r)}; \quad N_4 = \frac{(x-q)}{(p-q)}\frac{(y-r)}{(s-r)} \end{aligned} \tag{8.40}$$

The shape functions given in Eq. (8.40) satisfy the three properties mentioned in Chapter 4. These shape functions are called *Lagrangian shape functions* and the corresponding elements shown in Fig. 8.7 are called *Lagrangian elements*. You can easily derive the shape function for nine-noded elements, which will be products of one-dimensional quadratic shape functions.

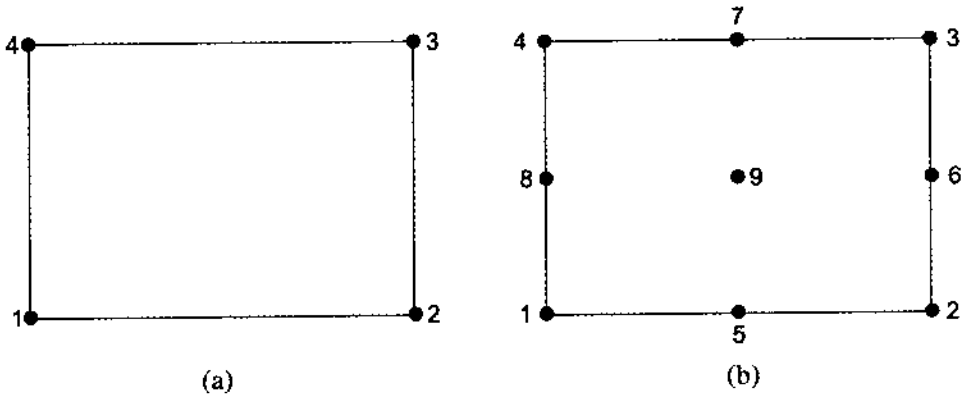


Figure 8.7 *Rectangular elements having (a) 4 nodes (b) 9 nodes*

8.5 BRICK ELEMENTS

On extending the rectangular elements to three dimensions, we obtain *brick* elements. Figure 8.8 shows 8-noded and 20-noded brick elements. The element shown in Fig. 8.8(a) is called a Lagrangian 8-noded brick element. The shape functions of this element are tri-linear. The three-dimensional element having tri-quadratic shape functions is an element having 27 nodes as shown in Fig. 8.7(b). The 20-noded element is actually a 'serendipity' element, which we shall discuss later. The shape functions for 8-noded elements can be obtained as the product of shape functions along x, y, and z. However, the shape functions for a 20-noded element cannot be obtained by simple multiplication of one-dimensional shape functions.

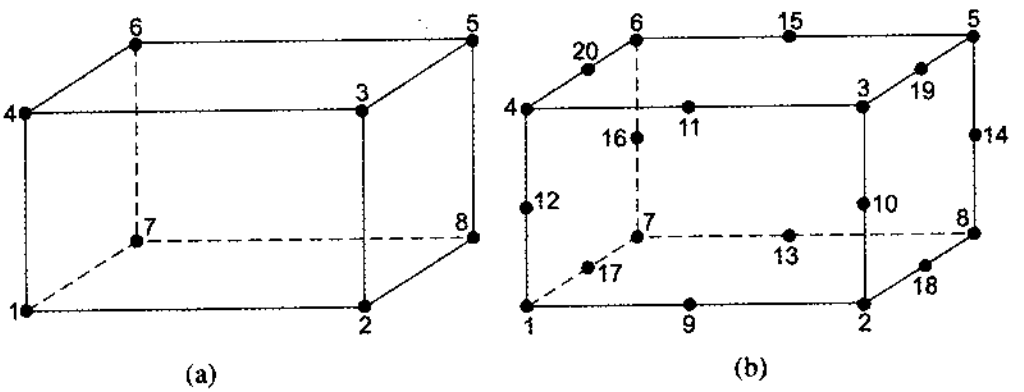


Figure 8.8 *Brick elements having (a) 8 nodes (b) 20 nodes*

8.6 GOVERNING DIFFERENTIAL EQUATION FOR TWO-DIMENSIONAL HEAT CONDUCTION

A finite element model is developed and implemented for solving a two-dimensional, steady-state heat-conduction problem for an orthotropic, non-homogeneous material. The governing equation is written as follows:

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right)+\dot{Q}=0 \tag{8.41}$$

where T is the temperature field, k_x is the thermal conductivity in x direction, k_y is the thermal conductivity in y direction, and $\dot{Q}$ is the rate of heat generation per unit volume. As the material is non-homogenous, both k_x and k_y are functions of x and y. The boundary conditions are as follows:

(i) Essential boundary condition: $T = T^*(x, y)$, prescribed temperature field on the boundary Γ_1.

(ii) Natural boundary conditions: Boundary heat flux conditions on boundary Γ_2

$q_n = \alpha^*(x, y)$ is the prescribed boundary heat flux or

$q_n = \beta(T - T_\infty)$ is the convective heat flux

where

$$q_n = -k_x \frac{\partial T}{\partial x}\hat{n}_x - k_y \frac{\partial T}{\partial x}\hat{n}_y \tag{8.42}$$

is the normal heat flux on the boundary Γ_2, β is the convective heat transfer coefficient, and T_∞ is the ambient temperature. In Eq. (8.42), $\hat{n}_x$ and $\hat{n}_y$ are the components of the unit outward normal on the surface. The convention is that the heat coming out of a surface is positive, while the heat going into it is negative. We shall carry out the FEM formulation using Galerkin procedure.

8.7 WEAK FORM AND FEM FORMULATION

Once the discretization of the domain is done, the weak form is developed on the arbitrary typical element. The element is denoted by Ω^e with boundary Γ^e. Let T^e be the finite element temperature field within an element. As it is not the exact temperature field, on substituting in the left-hand side of Eq. (8.41), in general, a non-zero residue, R^e will be obtained, i.e.,

$$R^e = \frac{\partial}{\partial x}\left(k_x \frac{\partial T^e}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T^e}{\partial y}\right)+\dot{Q} \tag{8.43}$$

The weighted integral statement of the governing equation is then given by

$$\iint_{\Omega^e} wR^e \mathrm{d}\Omega = 0 \tag{8.44}$$

where w is the *weighting function.*

Thus

$$\int_{\Omega^e} w\left[-\frac{\partial}{\partial x}\left(k_x \frac{\partial T^e}{\partial x}\right)-\frac{\partial}{\partial y}\left(k_y \frac{\partial T^e}{\partial y}\right)-\dot{Q}\right]\mathrm{d}x\mathrm{d}y = 0 \tag{8.45}$$

Using the component form of gradient theorem, given as

$$\int_{\Omega^e} \frac{\partial}{\partial x}\left(w\frac{\partial F}{\partial x}\right)\mathrm{d}\Omega = \oint_{\Gamma^e} wF\hat{n}_x \mathrm{d}s \tag{8.46}$$

and

$$-w\frac{\partial F}{\partial x} = \frac{\partial w}{\partial x}F - \frac{\partial}{\partial x}(wF) \tag{8.47}$$

Eq. (8.45) becomes

$$\int_{\Omega^e}\left(k_x\frac{\partial w}{\partial x}\frac{\partial T}{\partial x} + k_y\frac{\partial w}{\partial y}\frac{\partial T}{\partial y} - w\dot{Q}\right)\mathrm{d}x\mathrm{d}y - \oint_{\Gamma_e} w\left(k_x\frac{\partial T}{\partial x}\hat{n}_x + k_y\frac{\partial T}{\partial x}\hat{n}_y\right)\mathrm{d}s = 0 \tag{8.48}$$

or

$$\int_{\Omega^e}\left(k_x\frac{\partial w}{\partial x}\frac{\partial T^e}{\partial x} + k_x\frac{\partial w}{\partial y}\frac{\partial T^e}{\partial y}\right)\mathrm{d}x\mathrm{d}y = \int_{\Omega^e} w\dot{Q}\mathrm{d}x\mathrm{d}y - \oint_{\Gamma^e} wq_n\mathrm{d}s = 0 \tag{8.49}$$

where q_n is the heat flux in the direction of unit normal **n**.

For a convective boundary, the natural boundary condition is a balance of energy transfer across the boundary due to conduction and/or convection (i.e., Newton's law of cooling):

$$k_x\frac{\partial T}{\partial x}\hat{n}_x + k_y\frac{\partial T}{\partial x}\hat{n}_y = -\beta(T - T_\infty) \tag{8.50}$$

where β is the convective conductance (or convective heat transfer coefficient), T_∞ is the ambient temperature of the surrounding fluid medium. It is the presence of the term $\beta(T - T_\infty)$ that requires some modification of Eq. (8.49). In the weak form given in Eq. (8.41), the boundary integral is modified to account for the convective heat transfer term in Eq. (8.48) as follows:

$$\int_{\Omega^e}\left(k_x\frac{\partial w}{\partial x}\frac{\partial T}{\partial x} + k_y\frac{\partial w}{\partial y}\frac{\partial T}{\partial y} - w\dot{Q}\right)\mathrm{d}x\mathrm{d}y + \int_{\Gamma_{e1}} wq_n\,\mathrm{d}s + \int_{\Gamma_{e2}} \beta w(T - T_\infty)\mathrm{d}s = 0 \tag{8.51}$$

or

$$\mathrm{B}(w,T) - \mathrm{l}(w) = 0 \tag{8.52}$$

where w is the weight function, and B(.,.) and l(.) are the bilinear and linear forms, respectively.

$$\mathrm{B}(w,T) = \int_{\Omega^e}\left(k_x\frac{\partial w}{\partial x}\frac{\partial T}{\partial x} + k_y\frac{\partial w}{\partial y}\frac{\partial T}{\partial y}\right)\mathrm{d}x\mathrm{d}y + \int_{\Gamma_{e2}} \beta wT\mathrm{d}s \tag{8.53}$$

$$\mathrm{l}(w) = \int_{\Omega^e} w\dot{Q}\mathrm{d}x\mathrm{d}y + \oint_{\Gamma_{e2}} \beta wT_\infty\mathrm{d}s - \int_{\Gamma_{e1}} wq_n\mathrm{d}s \tag{8.54}$$

Note that Γ_{e1} is the boundary on which heat flux is prescribed and Γ_{e2} is the boundary on which convective heat transfer takes place. The finite element model is obtained by substituting the finite element approximation of the form

$$T = \sum_{j=1}^{n} T_j^e N_j^e(x,y) \tag{8.55}$$

and

$$w = \sum_{j=1}^{n} w_j^e N_j^e(x,y) \tag{8.56}$$

for T and w, respectively, into Eq. (8.52). In Eqs (8.55) and (8.56), n is the number of the nodes in the elements. Superscript e in these equations indicates the element number, but will be omitted later in some cases for the sake of brevity. We shall use the notation

$$\left\{T^{ne}\right\}=\begin{Bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{Bmatrix} \tag{8.57}$$

$$\left\lfloor w^{ne} \right\rfloor=\left\lfloor w_1\ w_2 \ldots w_n \right\rfloor \tag{8.58}$$

and

$$\left\lfloor N \right\rfloor=\left\lfloor N_1\ N_2 \ldots N_n \right\rfloor=\left\{N\right\}^{\mathrm{T}} \tag{8.59}$$

In Eq. (8.57), the nodal temperatures of an element have been arranged in the form of a column vector. In Eq. (8.58), the nodal weights of an element are arranged in the form of a row vector. In Eq. (8.59), the elemental shape functions are arranged in the form of a row vector. With these notations, Eqs (8.55) and (8.56) can be written as

$$T=\left\lfloor N \right\rfloor\left\{T^{ne}\right\} \tag{8.60}$$

$$w=\left\lfloor w^{ne} \right\rfloor\left\{N\right\} \tag{8.61}$$

The partial differentiation of Eqs (8.60) and (8.61) with respect to x and y provides

$$\frac{\partial T}{\partial x}=\left\lfloor N_{,x} \right\rfloor\left\{T^{ne}\right\} \tag{8.62}$$

$$\frac{\partial T}{\partial y}=\left\lfloor N_{,y} \right\rfloor\left\{T^{ne}\right\} \tag{8.63}$$

$$\frac{\partial w}{\partial x}=\left\lfloor w^{ne} \right\rfloor\left\{N_{,x}\right\} \tag{8.64}$$

$$\frac{\partial w}{\partial y}=\left\lfloor w^{ne} \right\rfloor\left\{N_{,x}\right\} \tag{8.65}$$

Substituting Eqs (8.60) to (8.65) into Eq. (8.52), we obtain

$$\begin{aligned}&\int_{\Omega^e}\left\lfloor w^{ne} \right\rfloor\left(k_x\left\{N_{,x}\right\}\left\lfloor N_{,x} \right\rfloor\left\{T^{ne}\right\}+k_y\left\{N_{,y}\right\}\left\lfloor N_{,y} \right\rfloor\left\{T^{ne}\right\}-\left\{N\right\}\dot{Q}\right)\mathrm{d}x\mathrm{d}y\\&+\int_{\Gamma_{e1}}\left\lfloor w^{ne} \right\rfloor\left\{N\right\}q_n\,\mathrm{d}s+\int_{\Gamma_{e2}}\beta\left\lfloor w^{ne} \right\rfloor\left\{N\right\}\left(\left\lfloor N \right\rfloor\left\{T^{ne}\right\}-T_\infty\right)\mathrm{d}s=0\end{aligned} \tag{8.66}$$

Note that on a boundary only the shape functions corresponding to nodes on the boundary may be non-zero. Thus, the shape function vectors within the boundary integral will have many entries as 0. In Galerkin method, nodal weights are arbitrary. If the dot product of a vector with an arbitrary vector is 0, the vector must be a zero vector. This implies

$$\begin{aligned}&\int_{\Omega^e}\left(k_x\left\{N_{,x}\right\}\left\lfloor N_{,x} \right\rfloor+k_y\left\{N_{,y}\right\}\left\lfloor N_{,y} \right\rfloor\right)\left\{T^{ne}\right\}\mathrm{d}x\mathrm{d}y+\int_{\Gamma_{e2}}\beta\left\{N\right\}\left\lfloor N \right\rfloor\left\{T^{ne}\right\}\mathrm{d}s\\&=\int_{\Omega^e}\dot{Q}\left\{N\right\}\mathrm{d}x\mathrm{d}y-\int_{\Gamma_{e1}}\left\{N\right\}q_n\,\mathrm{d}s+\int_{\Gamma_{e2}}\beta T_\infty\left\{N\right\}\mathrm{d}s\end{aligned} \tag{8.67}$$

Equation (8.67) can be written in the following form:

$$[k^e]\left\{T^{ne}\right\} = [R^e] + [R_i^e] \tag{8.68}$$

where $[k^e]$ is the elemental matrix, $[R^e]$ is the right-hand side matrix representing heat generation and convection through the surfaces, and $[R_i^e]$ is the right-hand side vector representing internal flux. Here

$$\left[k^e\right] = \int_{\Omega^e} \left(k_x \left\{N_{,x}\right\}\left\lfloor N_{,x}\right\rfloor + k_y \left\{N_{,y}\right\}\left\lfloor N_{,y}\right\rfloor\right) \mathrm{d}x\mathrm{d}y + \int_{\Gamma_{e2}} \beta\{N\}\lfloor N\rfloor \,\mathrm{d}s \tag{8.69}$$

$$[R^e] = \int_{\Omega^e} \dot{Q}\{N\}\mathrm{d}x\mathrm{d}y + \int_{\Gamma_{e2}} \beta T_\infty \{N\}\,\mathrm{d}s \tag{8.70}$$

$$[R_i^e] = -\int_{\Gamma_{e1}} \{N\}\, q_n \,\mathrm{d}s \tag{8.71}$$

During assembly, the internal heat flux vectors will cancel each other everywhere except at the boundaries. Equation (8.68) actually provides the following n equations for an n-noded element:

$$\sum_{j=1}^{n} k_{ij}^e T_j^e = F_i^e + P_i^e, \quad i = 1 \text{ to } n \tag{8.72}$$

where

$$k_{ij}^e = \int_{\Omega^e} \left(k_x N_{i,x} N_{j,x} + k_y N_{i,y} N_{j,y}\right) \mathrm{d}x\mathrm{d}y + \int_{\Gamma_{e2}} \beta N_i N_j \mathrm{d}s \tag{8.73}$$

$$F_i^e = \int_{\Omega^e} \dot{Q} N_i \mathrm{d}x\mathrm{d}y + \beta^e \int_{\Gamma_{e2}} N_i T_\infty \mathrm{d}s \tag{8.74}$$

$$P_i^e = -\int_{\Gamma_{e1}} q_n^e N_i \mathrm{d}s \tag{8.75}$$

Once the elemental equations are obtained, they are assembled to form a set of linear simultaneous equations, the solution of which yields the temperature field. The assembly is based on the principle of maintaining the continuity of primary variables, i.e., temperatures, and noting that at the interface of an element, the heat coming out of one element is equal to the heat going into the other elements.

The variational form of the differential equation can be obtained easily by putting δT in place of w in the weak form. It is given by

$$\Pi = \int_{\Omega^e} \left[\frac{1}{2} k_x (T,x)^2 + \frac{1}{2} k_y (T,y)^2 - \dot{Q}T\right] \mathrm{d}\Omega^e + \int_{\Gamma_{e1}} q_n T \mathrm{d}s + \int_{\Gamma_{e2}} \left(\frac{1}{2} hT^2 - hTT_\infty\right) \mathrm{d}s \tag{8.76}$$

Now, Ritz FEM can be employed to obtain the FEM equations.

8.8 A NOTE ON THE ASSEMBLY IN TWO DIMENSIONS

The concept of a connectivity matrix was introduced in Chapter 1. It was further discussed in Chapter 7. In two-dimensional problems, it gains major importance. In a connectivity matrix, a row

denotes the element number and a column denotes the local node number. An element of the connectivity matrix indicates the global node number. For the mesh of rectangular elements shown in Fig. 8.9(a), the connectivity matrix is given by

$$\begin{bmatrix} 1 & 2 & 6 & 5 \\ 2 & 3 & 7 & 6 \\ 3 & 4 & 8 & 7 \\ 5 & 6 & 10 & 9 \\ 6 & 7 & 11 & 10 \\ 7 & 8 & 12 & 11 \\ 9 & 10 & 14 & 13 \\ 10 & 11 & 15 & 14 \\ 11 & 12 & 16 & 15 \end{bmatrix}$$

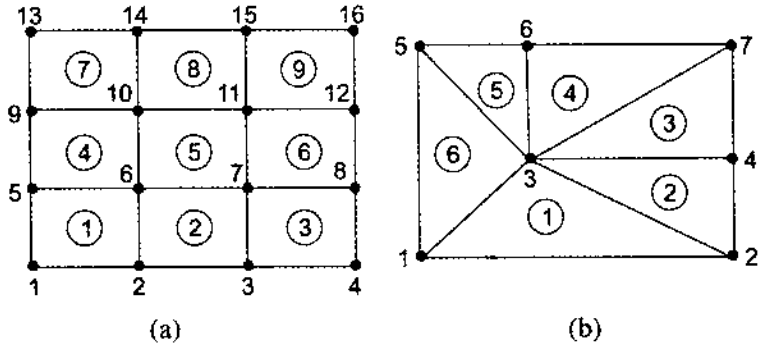


Figure 8.9 *Mesh of (a) rectangular elements (b) triangular elements*

For the mesh of triangular elements shown in Fig. 8.9(b), the connectivity matrix is given by

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 4 \\ 3 & 4 & 7 \\ 3 & 7 & 6 \\ 3 & 6 & 5 \\ 1 & 3 & 5 \end{bmatrix}$$

The task of assembly is to identify the global location of each element in the local matrix and load vector and put the element in that place. The assembly procedure as discussed in Chapter 7 can be implied. A pseudocode (akin to C language) for the assembly of four-noded rectangular elements is given below

```
/*Pseudocode for assembly of four-noded rectangular elements*/
for (i=1; i<=number_of_nodes; i=i+1) /*outer loop*/
{
for (j=1; j<=number_of_nodes; j+1) /*inner loop*/
{
gstiff [i][j] = 0.0;   /*initializing global stiffness matrix*/
}
grhs[i]=0.0;     /*initializing global right hand side vector*/
}
for (i=1; i<= number_of_elements; i=i+1) /*outer loop*/
{
stiffness(i); /*calling stiffness matrix of element i*/
righthand(i); /*calling right hand side vector of element i*/
for (j=1; j<=4; j=j+1) /* inner loop 1 */
{
p=C[i][j];
grhs[p]=grhs[p]+erhs[j)]; /*placing elemental entries into global
location*/
for (k=1; k<=4; k=k+1) /*inner loop 2*/
{
q=C[i][k];
gstiff[p, q]= gstiff [p, q]+estiff[(j, k]; /*placing elemental
entries into global location*/
}}}
```

8.9 POISSON EQUATION IN THREE DIMENSIONS

The steady-state heat-conduction equation for an orthotropic material in three dimensions is given by

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right)+\dot{Q}=0 \tag{8.77}$$

Proceeding as before, the FEM formulation for this problem will provide the elemental coefficient matrix as

$$\left[k^e\right]=\int_{V^e}\left(k_x\left\lfloor N_{,x}\right\rfloor\left\{N_{,x}\right\}+k_y\left\lfloor N_{,y}\right\rfloor\left\{N_{,y}\right\}+k_z\left\lfloor N_{,z}\right\rfloor\left\{N_{,z}\right\}\right)\mathrm{d}x\mathrm{d}y\mathrm{d}z \tag{8.78}$$

The reader can derive this as well as the load vector. Similarly, the right-hand-side vector representing heat generation is obtained as

$$\{R^e\} = \int_{V^e} \dot{Q}\{N\}\mathrm{dxdydz} \tag{8.79}$$

and the right-hand-side vector representing internal heat flux is given as

$$\{R_t^e\} = -\int_{\Gamma_e} \{N\}\, q_n\, \mathrm{d}s = \int_{\Gamma_e} \{N\}\left(k_x \frac{\partial T}{\partial x}\hat{n}_x + k_y \frac{\partial T}{\partial x}\hat{n}_y + k_z \frac{\partial T}{\partial x}\hat{n}_z\right)\mathrm{d}s \tag{8.80}$$

The assembly procedure is similar to that discussed in the previous section.

8.10 FLUID FLOW PROBLEM

In fluid dynamics, *vorticity* is a mathematical concept that represents the amount of circulation or rotation in a fluid. Vorticity at any point is defined as the curl of velocity and hence is a vector. For an irrotational, two-dimensional flow, the vorticity is 0. Therefore, we can write

$$\frac{\partial u}{\partial y} - \frac{\partial v}{\partial x} = 0 \tag{8.81}$$

where u and v are the fluid velocities in the x and y directions, respectively. For an incompressible flow, the continuity equation is given by

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{8.82}$$

For solving Eqs (8.81) and (8.82), we define a potential function $\phi(x, y)$ such that

$$u = -\frac{\partial \phi}{\partial x} \quad \text{and} \quad v = -\frac{\partial \phi}{\partial y} \tag{8.83}$$

Equation (8.81) is identically satisfied. Substitution of Eq. (8.83) into Eq. (8.82) provides

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} = 0 \tag{8.84}$$

This is a Laplace equation and the FEM formulation can be obtained similar to the case of steady-state heat conduction. At any boundary

$$\frac{\partial \phi}{\partial n} = \frac{\partial \phi}{\partial x}\hat{n}_x + \frac{\partial \phi}{\partial y}\hat{n}_y = u\hat{n}_x + v\hat{n}_y = v_n \tag{8.85}$$

where $\frac{\partial \phi}{\partial n}$ denotes the derivative of the velocity potential in the direction of the unit normal vector to the boundary. At a solid boundary, $v_n = 0$ which is a natural boundary condition. The prescribed velocity boundary condition can also form the natural boundary conditions by expressing the velocity in terms of derivatives of the velocity potential. In addition to natural boundary conditions, at least on one point we need to provide an arbitrary value of velocity potential that forms an essential boundary condition.

8.11 TORSION OF CIRCULAR AND NON-CIRCULAR CROSS-SECTION

The torsional behaviour of a shaft of a circular and non-circular cross-section is governed by the following equation

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + 2G\theta = 0 \tag{8.86}$$

where θ is the angle of twist per unit length, G is the shear modulus of the shaft material, and ϕ is the stress function. The shear stress components at any point can be calculated as

$$\tau_{zx} = \frac{\partial \phi}{\partial y} \qquad \tau_{zy} = -\frac{\partial \phi}{\partial x} \tag{8.87}$$

The boundary condition on ϕ is that it is zero at the boundary of the shaft. Thus, the problem is similar to the heat-conduction problem with heat generation and prescribed temperature on the boundary.

8.12 A NOTE ON EXPLOITING SYMMETRY IN FEM

Consider the problem of heat conduction in a square plate as shown in Fig. 8.10(a). Three sides of the plate are kept at 400 K and the fourth side is kept at 600 K. We can discretize the domain as shown in Fig. 8.10(b). The present mesh consists of 25 nodes. It is clear that this problem possesses symmetry. We can exploit the symmetry to reduce the burden of computation. The problem is symmetric about the middle vertical line connecting nodes 3 and 23. Thus, we can divide the domain in two parts by this line and analyse only one part shown in Fig. 8.10(c). The boundary condition on the middle line is given by

$$\frac{\partial T}{\partial x} = 0 \tag{8.88}$$

This can be justified physically and mathematically as follows:

Physical justification

At the middle vertical line, the heat coming out from one half of the domain will be same as the heat going into the other half as there is no heat generation or absorption at the middle line. Now, if we assume that the heat is coming from the left half of the domain, then in the right half of the domain, it must enter through the middle line. However, due to symmetry, there is no difference between the right- and left-hand-side domains. A clockwise rotation of 180° about the middle line will transform the left-hand-side domain into the right-hand-side domain. It is not possible that a mere rotation of the domain can change the heat flux direction. Therefore, for the sake of consistency, we have to accept that the heat flux through the middle line is 0. Thus

$$-k_x \frac{\partial T}{\partial x} = 0 \tag{8.89}$$

proving the validity of Eq. (8.88).

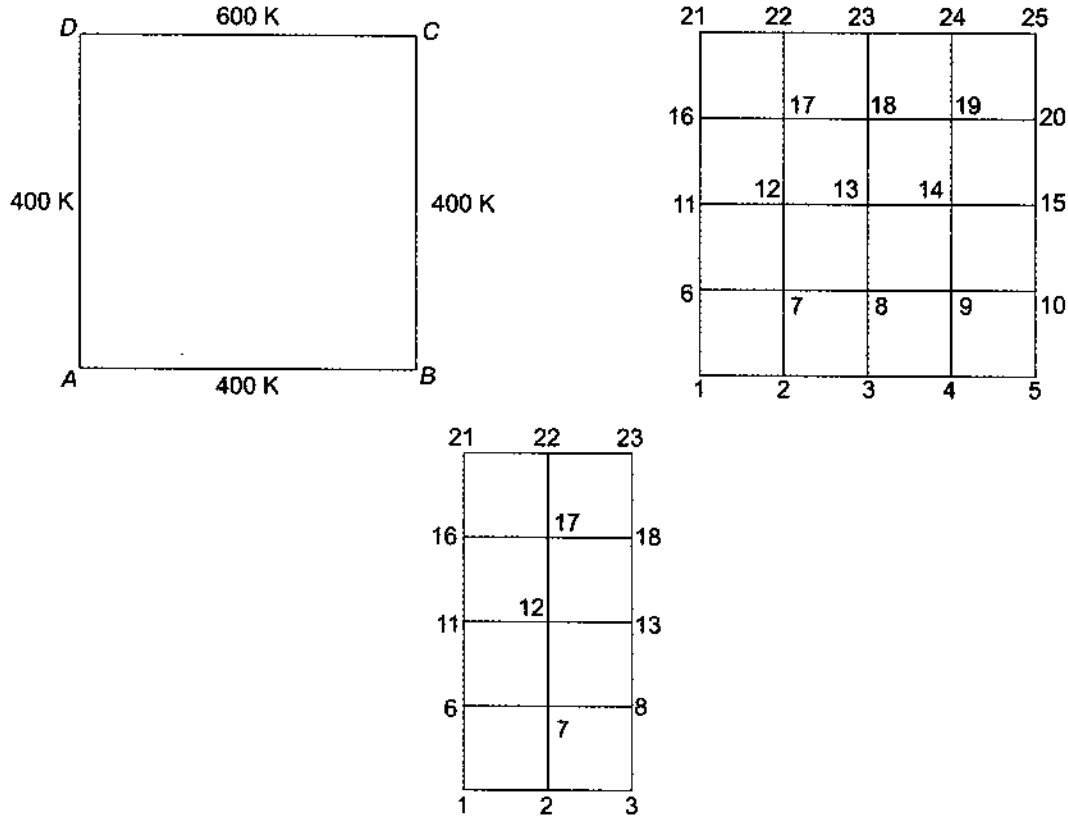


Figure 8.10 *An application of symmetry condition: (a) Domain of the problem (b) Meshing of the full domain (c) Meshing of half the domain*

Mathematical justification

Since there is no concentrated heat source and the thermal conductivities are constant functions of the coordinates, the temperature field is continuous. A discontinuous temperature field will not satisfy Laplace equation. Therefore, at the middle vertical line, $\partial T/\partial x$ will be either positive or negative. Assume that it is positive. In that case, for the left side, it indicates that the temperature at a point at the left of the middle vertical line will be lower with respect to a point on the middle line. Also, the positive value of $\partial T/\partial x$ indicates that for the right-hand-side domain, the temperature at an adjacent point towards the right of the middle-line point will be higher than the temperature on the middle-line point. Thus, there is no symmetry. For symmetry, Eq. (8.88) must hold good.

It is clear that the above arguments will not hold good if there is a concentrated heat source on the middle edge. In that case, the temperature field will have only C^0 continuity. In fact, the derivative at the middle line is not defined. For solving such problems, it is better to consider the entire domain, although an experienced FEM user can still exploit symmetry!

8.13 SUMMARY

In this chapter, an introduction to two-dimensional and three-dimensional FEM is provided. We have limited our discussion to Laplace and Poisson equations. We have considered simple types of elements. More details of FEM formulation with isoparametric elements will be provided in the next chapter.

SOLVED PROBLEMS

8.1 Find out the shape functions N_1, N_4, and N_{10} for the cubic triangle shown in Fig. 8.5.

Solution The shape functions will be cubic polynomial, i.e., $M = 3$. The natural coordinate of node 1 is (1, 0, 0). Node 1 will be designated by (3, 0, 0), which is obtained by multiplying the natural coordinate with 3. Noting that $I = 3$, $J = K = 0$, we use Eq. (8.31)

$$N_1 = l_3^3(\xi_1)l_0^0(\xi_2)l_0^0(\xi_3) = l_3^3(\xi_1)$$

$$= \frac{(\xi_1 - 0)\left(\xi_1 - \frac{1}{3}\right)\left(\xi_1 - \frac{2}{3}\right)}{(1-0)\left(1-\frac{1}{3}\right)\left(1-\frac{2}{3}\right)}$$

$$= \frac{1}{2}(3\xi_1 - 1)(3\xi_1 - 2)\xi_1$$

Natural coordinates of node 4 are (2/3, 1/3, 0). We multiply these by 3 and obtain the designation of node 4 as (2, 1, 0). Thus, $I = 2$, $J = 1$, and $K = 0$. Using Eq. (8.28), we obtain

$$N_4 = l_2^2(\xi_1)l_1^1(\xi_2)l_0^0(\xi_3) = l_2^2(\xi_1)l_1^1(\xi_2)$$

$$= \frac{(\xi_1 - 0)\left(\xi_1 - \frac{1}{3}\right)}{\left(\frac{2}{3}-0\right)\left(\frac{2}{3}-\frac{1}{3}\right)}\frac{(\xi_2 - 0)}{\left(\frac{1}{3}-0\right)}$$

$$= \frac{9}{2}\xi_1\xi_2(3\xi_1 - 1)$$

Natural coordinates of node 10 are (1/3, 1/3, 1/3). Multiplying by 3, we get the designation of node 10 as (1, 1, 1). Thus, $I = J = K = 1$. Thus

$$N_{10} = l_1^1(\xi_1)l_1^1(\xi_2)l_1^1(\xi_3)$$

$$= \frac{(\xi_1 - 0)}{\left(\frac{1}{3}-0\right)}\frac{(\xi_2 - 0)}{\left(\frac{1}{3}-0\right)}\frac{(\xi_3 - 0)}{\left(\frac{1}{3}-0\right)}$$

$$= 27\xi_1\xi_2\xi_3$$

8.2 Derive all shape functions of an eight-noded brick element shown in Fig. 8.11.

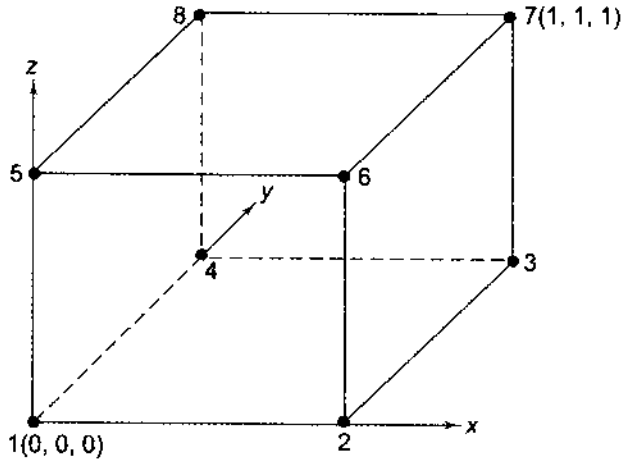


Figure 8.11 *An eight-noded brick element*

Solution The coordinates of any node are (0 or 1, 0 or 1). In one dimension, along *x-axis*, the shape function corresponding to node 0 is

$$N_1^x = (1 - x)$$

Similarly, the shape function corresponding to node 1 is

$$N_2^x = x$$

In the same way, the one-dimensional shape functions along y and z axes are given by

$$N_1^y = (1 - y), \quad N_2^y = y; \quad N_1^z = (1 - z), \quad N_2^z = z$$

Thus, shape functions corresponding to node 1 with coordinates (0, 0, 0) are given as follows

$$N_1 = N_1^x N_1^y N_1^z = (1 - x)(1 - y)(1 - z)$$

The shape function corresponding to node 2 with coordinates (1, 0, 0) is written as

$$N_2 = N_2^x N_1^y N_1^z = x(1 - y)(1 - z)$$

The shape function corresponding to node 3 with coordinates (1, 1, 0) is given by

$$N_3 = N_2^x N_2^y N_1^z = xy(1 - z)$$

The shape function corresponding to node 4 with coordinates (0, 1, 0) is written as

$$N_4 = N_1^x N_2^y N_1^z = (1 - x)\, y(1 - z)$$

The shape function corresponding to node 5 with coordinates (0, 0, 1) is given as

$$N_5 = N_1^x N_1^y N_2^z = (1 - x)(1 - y)z$$

The shape function corresponding to node 6 with coordinates (1, 0, 1) is given by

$$N_6 = N_2^x N_1^y N_2^z = x(1-y)z$$

The shape function corresponding to node 7 with coordinates (1, 1, 1) is written as

$$N_7 = N_2^x N_2^y N_2^z = xyz$$

The shape function corresponding to node 8 with coordinates (0, 1, 1) is given as

$$N_8 = N_1^x N_2^y N_2^z = (1-x)\,y\,z$$

EXERCISES

A. Multiple Choice Questions

8.1 For a two-dimensional heat-conduction problem in a plate (having the temperature boundary conditions shown in Fig. 8.12), at the line of symmetry (x-axis):

(a) $\dfrac{\partial T}{\partial x} = \dfrac{\partial T}{\partial y} = 0$

(b) $\dfrac{\partial T}{\partial y} = 0$

(c) $\dfrac{\partial T}{\partial x} = 0$

(d) $\dfrac{\partial T}{\partial x} + \dfrac{\partial T}{\partial y} = 0$

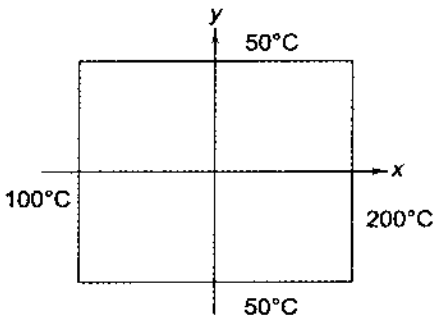


Figure 8.12 *Two-dimensional heat-conduction in a plate*

8.2 For the triangular element shown in Fig. 8.13

$$\frac{\partial T}{\partial x} = 1 \qquad \frac{\partial T}{\partial y} = 2$$

Taking thermal conductivity to be 1, the magnitude of heat flux through BC is

(a) $\dfrac{\sqrt{3}}{2} + 1$

(b) $\dfrac{1}{2} + \sqrt{3}$

(c) 3

(d) $\sqrt{5}$

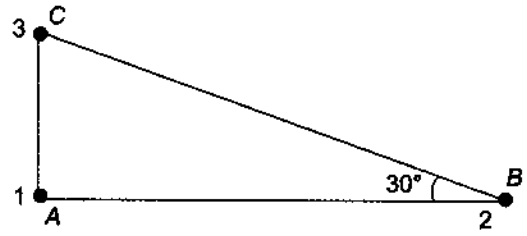


Figure 8.13 *Triangular element of Question 8.2*

8.3 A triangular element is shown in Fig. 8.14 with physical coordinates of the nodes as indicated in the figure. The physical coordinates of a point with natural coordinates as $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right)$ are

(a) (3/4, 1) (b) (3/2, 2) (c) (1/2, 1/4) (d) (3, 7)

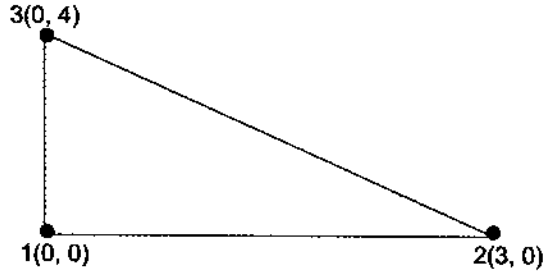


Figure 8.14 *Triangular element of Question 8.3*

8.4 The temperatures at the four nodes of a four-noded rectangular element are 300 K, 350 K, 450 K, and 400 K. The temperature at the centre of the element is

(a) 400 K (b) 350 K (c) 300 K (d) 375 K

8.5 The temperatures at the four nodes of a four-noded rectangular element are 300 K, 350 K, 400K, and 450 K. The line joining the nodes 1 and 2 is 1-cm long and that joining the nodes 2 and 3 is 2-cm long. The x-axis is along the line joining nodes 1 and 2. The value of $\partial T/\partial x$ at the centre of the element is

(a) 50 K/cm (b) 25 K/cm (c) 12.5 K/cm (d) 100 K/cm

B. Problems

8.6 Derive all the shape functions for a cubic triangular element using the simplified formula described in this chapter.

8.7 Sketch a 15-noded triangular element. Find out the expression for the shape functions of this element.

8.8 Figure 8.15 shows a mesh consisting of linear triangular elements. Write down the connectivity matrix of the mesh.

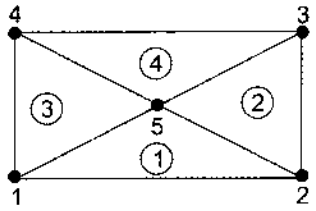


Figure 8.15 *A mesh of linear triangular elements*

8.9 In the triangle shown in Fig. 8.16, a point has natural coordinates (1/3, 1/3, 1/3). Find out its x and y coordinates.

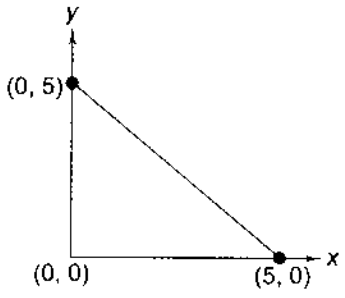


Figure 8.16 *Triangle corresponding to Problem 8.9*

8.10 A square element has a point heat source of 100 W at the center. Find out its contribution at all the corner nodes.

8.11 Develop an expression for the stiffness matrix for a triangular element for solving steady-state heat-conduction problem. Treat the material as orthotropic.

8.12 Develop an expression for the stiffness matrix for a rectangular element of sides a and b for solving steady-state heat-conduction problem. Treat the material as orthotropic.

REFERENCES

Argyris, JH, Fried, I, and Scharpf, DW, 1968, 'The TET 20 and the TEA 8 elements for the matrix displacement method,' *Aero. J.*, 72, 618–25.

Silvester, P, 1969, 'Higher order polynomial triangular elements for potential problems', *Int. J. Eng. Sci.*, 7, 849–61.

Taylor, RL, 1972, 'On completeness of shape functions for finite element *analysis*', *Int. J. Num. Meth. Eng.*, 4, 17–22.

9

CHAPTER

Numerical Integration

9.1 Introduction

9.2 One-Dimensional Integration Formulae

9.3 Two-Dimensional Integration Formulae

9.4 Conclusion

9.1 INTRODUCTION

In finite element method, the integrations arising in the computation of coefficient matrix and right-hand-side vector are often carried out numerically. The process of numerical integration is known as *quadrature*. In this chapter, we shall describe the commonly used quadrature formulae. The accuracy of numerical integration affects the accuracy of the solution to a great extent. Therefore, we have devoted a complete chapter for a detailed study of this topic.

9.2 ONE-DIMENSIONAL INTEGRATION FORMULAE

Integration basically represents the process of summation. Figure 9.1 shows a graphical representation of a function $f(x)$ in the form of a curve with points 1, 2, 3, and 4, where points 1 and 4 are the end points of the curve. The area under the curve can be calculated by dividing it into three equal intervals where Δx is the interval size of each part and $f(x_1)$, $f(x_2)$, and $f(x_3)$ are the heights. The area can be given by the sum

$$S = f(x_1)\Delta x + f(x_2)\Delta x + f(x_3)\Delta x \tag{9.1}$$

This sum denotes the shaded area in the figure. The sum need not be exactly equal to the area under the curve $f(x)$ in the first quadrant. However, if we divide the curve into a large number of parts, say n parts, instead of just three parts as in Fig. 9.1, then the summation

$$S = \sum_{i=1}^{n} f_i(x)\Delta x \tag{9.2}$$

will be very close to the area under the curve in the first quadrant. As $n \to \infty$ and $\Delta x \to 0$, the summation S in Eq. (9.2) tends to the area under the curve $f(x)$ on the positive quadrant. The summation in Eq. (9.2) when $n \to \infty$ and $\Delta x \to 0$ is denoted as $\int f(x)\,dx$. The symbol '$\int$' is a distorted form of S which implies summation. As integration basically represents summation, all numerical integration formulae involve summation. We cannot, of course, take $n = \infty$ when summing Eq. (9.2) manually or even by using the fastest computer. Therefore, in the numerical integration procedure, approximations have to be incorporated. There are always finite terms in the summation. However, depending on the function and the type of quadrature formula, sometimes the error in the evaluation of integral can be made zero even with finite number of terms in the formula. In this section, we describe two commonly used numerical integration procedures: Newton–Cotes quadrature and Gauss quadrature.

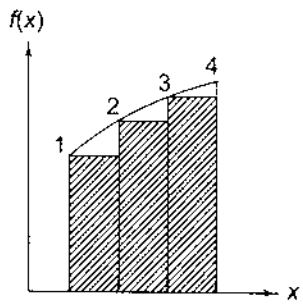


Figure 9.1 *The plot of function f(x) and illustration of the concept of integration*

9.2.1 Newton–Cotes Quadrature

In Eq. (9.2), if we treat Δx as a weight then the integral $\int f(x)\,\mathrm{d}x$ can be considered to be approximately the weighted sum of function values at discrete sampling points. In general, the weights or interval sizes Δx need not be equal. Similarly, the discrete sampling points need not be equally spaced. However, in developing formulae for Newton–Cotes quadrature, the points are usually taken to be equally spaced and are always specified before deciding the weights. This implies that for developing the formulae, the points at which the function is to be evaluated are determined a *priori*, usually at equal intervals. A polynomial is passed through these points and is exactly integrated to obtain a closed-form expression. The coefficients associated with the function values are taken as weights in quadrature formula.

The n values of the function define a polynomial of degree $n-1$. It can be shown [Mathews (2001)] that the errors will be of the order $O(\Delta^3)$ in a two-point formula and of the order $O(\Delta 5)$ in three-point and four-point formulae, where Δ is defined as the point spacing. For a general n-point Newton–Cotes quadrature formula, the integral can be written as

$$I = \int_{-1}^{1} f(\xi)\,\mathrm{d}\xi = \sum_{i=1}^{n} w_i f(\xi_i)\,\mathrm{d}t \tag{9.3}$$

where w_i are the weights. We have taken the limits from -1 to $+1$. If the limits are different, they can be transformed to lie from -1 to $+1$ with the help of linear transformation of the variables. For example, if x varies from a to b, the relation

$$x = \frac{a+b}{2} + \frac{(b-a)}{2}\xi \tag{9.4}$$

can be used to ensure that ξ varies from -1 to $+1$. Note that Eq. (9.4) provides

$$\mathrm{d}x = \frac{(b-a)}{2}\mathrm{d}\xi \tag{9.5}$$

We will now develop a two point Newton–Cotes formula as discussed below.

For two sampling points, i.e., for $n = 2$, a straight line can be easily fitted. Thus, the given function is approximated as

$$f(\xi) = a + b\xi \tag{9.6}$$

We choose two sampling points as $\xi = -1$ and $\xi = 1$. At $\xi = -1$

$$f(-1) = a - b \tag{9.7}$$

At $\xi = 1$

$$f(1) = a + b \tag{9.8}$$

From Eqs (9.7) and (9.8), we obtain

$$a = \frac{1}{2}\left[f(-1) + f(1)\right] \tag{9.9}$$

$$b = \frac{1}{2}\left[f(+1) - f(-1)\right] \tag{9.10}$$

Thus

$$f(\xi)=\frac{1}{2}\left[f(-1)+f(1)\right]+\frac{1}{2}\left[f(+1)-f(-1)\right]\xi \tag{9.11}$$

Therefore

$$I=\int_{-1}^{1} f(\xi)\mathrm{d}\xi = f(-1)+f(1) \tag{9.12}$$

Equation (9.12) is called *trapezoidal rule.*

Now, we develop a three-point Newton–Cotes quadrature formula. For $n = 3$, we can approximate the function as

$$f(\xi)=a+b\xi+c\xi^2 \tag{9.13}$$

The three sampling points are chosen as $\xi=-1$, $\xi=0$, and $\xi=1$. Integrating the function between the limits –1 and +1, we obtain

$$\int_{-1}^{+1} f(\xi)\mathrm{d}\xi = 2a+\frac{2}{3}c \tag{9.14}$$

Evaluating the function given in Eq. (9.13) at three sampling points, we get

$$f(-1)=a-b+c \tag{9.15}$$

$$f(0)=a \tag{9.16}$$

$$f(1)=a+b+c \tag{9.17}$$

Substituting Eq. (9.16) in Eqs (9.15) and (9.17), we obtain

$$c-b=f(-1)-f(0) \tag{9.18}$$

$$c+b=f(1)-f(0) \tag{9.19}$$

Solving Eqs (9.18) and (9.19), we get

$$c=\frac{f(-1)+f(1)-2f(0)}{2} \tag{9.20}$$

Now, substituting the values of a and c in Eq. (9.14), we obtain

$$\begin{aligned}\int_{-1}^{+1} f(\xi)d\xi &= 2f(0)+\frac{1}{3}f(-1)+\frac{1}{3}f(1)-\frac{2}{3}f(0)\\ &=\frac{1}{3}\left[f(-1)+4f(0)+f(1)\right]\end{aligned} \tag{9.21}$$

This is called *Simpson's one-third rule.*

In the same way, we can obtain a four-point formula. For $n = 4$, we approximate the function as

$$f(\xi)=a+b\xi+c\xi^2+d\xi^3 \tag{9.22}$$

The four sampling points required are chosen as $\xi=-1, \xi=-1/3, \xi=1/3$, and $\xi = 1$. Integrating the function between the limits –1 and +1, we get

$$\int_{-1}^{+1} f(\xi)\mathrm{d}\xi = 2a+\frac{2}{3}c \tag{9.23}$$

Evaluating the function given in Eq. (9.22) at four sampling points, we obtain the following values:

$$f(-1) = a - b + c - d \tag{9.24}$$

$$f\left(-\frac{1}{3}\right) = a - \frac{1}{3}b + \frac{1}{9}c - \frac{1}{27}d \tag{9.25}$$

$$f\left(\frac{1}{3}\right) = a + \frac{1}{3}b + \frac{1}{9}c + \frac{1}{27}d \tag{9.26}$$

$$f(1) = a + b + c + d \tag{9.27}$$

Equations (9.24) to (9.27) can be solved to provide the values of a, b, c, and d. However, Eq. (9.23) shows that for evaluating the integral, we need the values of a and c only. Thus, we can eliminate b and d by adding Eqs (9.24) and (9.27) together and Eqs (9.25) and (9.26) together to obtain

$$2a + 2c = f(-1) + f(1) \tag{9.28}$$

$$2a + \frac{2}{9}c = f\left(-\frac{1}{3}\right) + f\left(\frac{1}{3}\right) \tag{9.29}$$

Solving Eqs (9.28) and (9.29), we get

$$c = \frac{9}{16}\left[f(-1) + f(1) - f\left(-\frac{1}{3}\right) - f\left(\frac{1}{3}\right)\right] \tag{9.30}$$

$$a = \frac{9}{16}f\left(-\frac{1}{3}\right) + \frac{9}{16}f\left(\frac{1}{3}\right) - \frac{1}{16}f(-1) - \frac{1}{16}f(1) \tag{9.31}$$

Substituting the values of a and c in Eq. (9.23), we obtain

$$\int_{-1}^{+1} f(\xi)d\xi = \frac{1}{4}\left[f(-1) + 3f\left(-\frac{1}{3}\right) + 3f\left(\frac{1}{3}\right) + f(1)\right] \tag{9.32}$$

The above formula is also called *Simpson's 3/8 rule*. Here, the spacing h between the sampling points is 2/3. Thus, the formula can be written as

$$\int_{-1}^{+1} f(\xi)d\xi = \frac{3}{8}h\left[f(-1) + 3f\left(-\frac{1}{3}\right) + 3f\left(\frac{1}{3}\right) + f(1)\right] \tag{9.33}$$

In the above formula, the weight factor is 3/8 times the spacing between the points. Owing to this, the formula is called *Simpson's 3/8 rule*.

Example 9.1 Evaluate the following integral

$$I = \int_0^4 x^3\,dx \tag{9.34}$$

using (a) trapezoidal rule (b) Simpson's one-third rule (c) Simpson's 3/8 rule. Compare the solutions with the exact solution.

Solution The exact integration provides $I = 64$. To carry out numerical integration, we first transform the integral such that the limits lie between –1 and +1. Using Eqs (9.4) and (9.5), we obtain

$$x = \frac{0+4}{2} + \frac{(4-0)}{2}\xi = 2 + 2\xi \tag{9.35}$$

and

$$dx = 2\,d\xi \tag{9.36}$$

Thus, the integral can be written as

$$I = \int_{-1}^{1} (2+2\xi)^3\, 2\,d\xi \tag{9.37}$$

(a) Using *trapezoidal rule*, given by Eq. (9.12), we get

$$I = 2\times(2-2)^3 + 2\times(2+2)^3 = 128 \tag{9.38}$$

The exact value is 64. Therefore, the error in the evaluation of integral is 100%.

(b) Using *Simpson's one-third rule* [Eq. (9.21)], we obtain

$$I = \frac{1}{3}\left[f(-1) + 4f(0) + f(1)\right] = \frac{1}{3}\left[2\times(2-2)^3 + 4\times 2^4 + 2\times(2+2)^3\right] = 64 \tag{9.39}$$

This is equal to the exact solution because Simpson's one-third rule can carry out the exact integration for a cubic polynomial. One may wonder how Simpson's one-third rule can carry out exact integration of a cubic polynomial when a quadratic interpolation function given by Eq. (9.13) was used for developing the formula. To understand the real cause let us approximate the function by a cubic polynomial [Eq. (9.22)]. Now, at the sampling points of Simpson's one-third rule, the function values are given by

$$f(-1) = a - b - d + c \tag{9.40}$$

$$f(0) = a \tag{9.41}$$

$$f(1) = a + b + d + c \tag{9.42}$$

Equations (9.40) to (9.42) are similar to Eqs (9.15) to (9.17) except that b gets replaced by $b + d$. Both the systems of equations will give the same values for a and c. Integrating the cubic function between the limits –1 and +1, we get

$$\int_{-1}^{1} \left(a + b\xi + c\xi^2 + d\xi^3\right) d\xi = 2a + \frac{2}{3}c \tag{9.43}$$

Compare it with Eq. (9.14). Regardless of whether we use quadratic or cubic interpolation polynomial, we obtain the same formula. Hence, Simpson's one-third rule gives the exact integration up to a cubic polynomial.

(c) Using *Simpson's 3/8 rule* [Eq. (9.32)], we obtain

$$I = \frac{1}{4}\left[f(-1) + 3f\left(\frac{-1}{3}\right) + 3f\left(\frac{1}{3}\right) + f(1)\right]$$

$$= \frac{1}{4}\left[0 + 3\times 2\times\left(2 - \frac{2}{3}\right)^3 + 3\times 2\times\left(2 + \frac{2}{3}\right)^3 + 2\times 4^3\right] = 64 \tag{9.44}$$

This is also exact. Both Simpson's one-third rule and 3/8 rule provide the exact integration for a cubic polynomial. However, contrary to intuition, Simpson's one-third rule is more accurate [Mathews (2001)].

9.2.2 Gauss Quadrature

In this method, the integral can be written as

$$\int_{-1}^{+1} f(\xi)\mathrm{d}\xi = \sum_{i=1}^{n} w_i f(\xi_i) \tag{9.45}$$

where ξ_i is the sampling point and w_i is the associated weight. Here, the sampling points and associated weights are optimally chosen, which leads to the improvement of accuracy. We will develop one and two Gauss-point formulae as follows:

One-Gauss-point formula

Consider a function $f(\xi) = a + b\xi$, which can be evaluated exactly by a one-Gauss-point formula. Then

$$\int_{-1}^{+1} (a + b\xi)\mathrm{d}\xi = w_1 f(\xi_1) = w_1(a + b\xi_1) \tag{9.46}$$

or

$$2a = w_1 a + w_1 b\xi_1 \tag{9.47}$$

The above relation is true for any arbitrary a and b. Hence, comparing the coefficient of a and b on both sides we get $w_1 = 2$ and $\xi_1 = 0$ Thus, for one-Gauss-point formula, we have to evaluate the function at $\xi = 0$ and multiply it by 2.

Two-Gauss-point formula

Consider a function $f(\xi) = a + b\xi + c\xi^2 + d\xi^3$, which can be evaluated exactly by a two-Gauss-point formula. Then

$$\int_{-1}^{+1} f(\xi)\mathrm{d}\xi = w_1 f(\xi_1) + w_2 f(\xi_2) \tag{9.48}$$

or

$$\int_{-1}^{+1} \left(a + b\xi + c\xi^2 + d\xi^3\right)d\xi = w_1\left(a + b\xi_1 + c\xi_1^2 + d\xi_1^3\right) + w_2\left(a + b\xi_2 + c\xi_2^2 + d\xi_2^3\right) \tag{9.49}$$

or

$$2a + \frac{2c}{3} = (w_1 + w_2)a + (w_1\xi_1 + w_2\xi_2)b + \left(w_1\xi_1^2 + w_2\xi_2^2\right)c + \left(w_1\xi_1^3 + w_2\xi_2^3\right)d \tag{9.50}$$

As a, b, c, and d are arbitrary, we can equate the coefficients on both sides to obtain the following:

$$w_1 + w_2 = 2 \tag{9.51}$$

$$w_1\xi_1 + w_2\xi_2 = 0 \tag{9.52}$$

$$w_1\xi_1^2 + w_2\xi_2^2 = \frac{2}{3} \tag{9.53}$$

$$w_1\xi_1^3 + w_2\xi_2^3 = 0 \tag{9.54}$$

Solving Eqs (9.51) to (9.54), we obtain

$$\xi_1 = -\frac{1}{\sqrt{3}}, \quad \xi_2 = \frac{1}{\sqrt{3}}, \quad w_1 = 1, w_2 = 2$$

Similarly, for a three-Gauss-point formula

$$\xi_1 = -0.77459667, \quad w_1 = \frac{5}{9}$$
$$\xi_2 = 0, \quad w_2 = \frac{8}{9}$$
$$\xi_3 = +0.77459667, \quad w_3 = \frac{5}{9}$$

The following important points need to be noted in case of a one-dimensional Gauss-quadrature:

(i) As any Gauss-quadrature formula must evaluate a constant function accurately, we can write

$$\int_{-1}^{+1} a \mathrm{d}\xi = 2a = \sum_{i=1}^{n} w_i a \tag{9.55}$$

Thus

$$\sum w_i = 2$$

(ii) With n-Gauss-point formula, a polynomial of degree $2n - 1$ can be evaluated exactly.

Table 9.1 lists the weights and Gauss points up to eight-Gauss-point formulae. Using eight Gauss points, polynomials of degree up to 15 can be evaluated exactly.

Table 9.1 *Sampling Points and Weights for Gauss Quadrature*

Number of Gauss Points	*Sampling Points*	*Weights*
1	0.0	2
2	±0.577350269189626	1
3	±0.774596669241483	5/9
	0.0	8/9
4	±0.861136311594053	0.347854845137454
	±0.339981043584856	0.652145154862546
5	0.0	0.568888888888889
	±0.538469310105683	0.478628670499365
	±0.906179845938684	0.236926885056189
6	±0.238619186083197	0.467913934572691
	±0.661209386466265	0.661209386466265
	±0.932469514203152	0.171324472379170
7	0.0	0.417959183673469
	±0.405845151377397	0.381830050505119
	±0.741531185597374	0.217705171887777
	±0.949101717342159	0.177484766168810
8	±0.183434642495650	0.362683783378362
	±0.525532409916329	0.313706645877887
	±0.796666477413627	0.222381034453374
	±0.960289856497536	0.101228536290376

9.3 TWO-DIMENSIONAL INTEGRATION FORMULAE

One-dimensional Gauss quadrature can be extended to two and three dimensions as well. The method by which the integration can be carried out in square and triangular domains is described in the following subsections:

9.3.1 Integration Over a Square Region

A square domain is shown in Fig. 9.2. The numerical integration over this domain can be carried out as follows:

$$I = \int_{-1}^{1}\int_{-1}^{-1} f(\xi,\eta)\mathrm{d}\xi\mathrm{d}\eta$$

$$= \sum_{i=1}^{n_1} w_i \int_{-1}^{+1} f(\xi,\eta_i)\mathrm{d}\xi$$

where n_1 is the number of Gauss points in the direction η. The application of one-dimensional Gauss quadrature again leads to

$$= \sum_{i=1}^{n_1} w_i \sum_{j=1}^{n_2} w_j f(\xi_j,\eta_i)$$

$$= \sum_{i=1}^{n_1}\sum_{j=1}^{n_2} w_i w_j f(\xi_j,\eta_i)$$

where n_2 is the number of Gauss points in the direction ξ.

In order to apply this formula, the four corners of the region must be (–1, –1), (1, –1), (1, 1), and (–1, 1). If a rectangular region is provided, it can be transformed to a square region in natural coordinates $\xi - \eta$ by linear transformation. In fact, any quadrilateral region can be transformed to a square region in natural coordinates. We shall discuss this transformation in the next chapter.

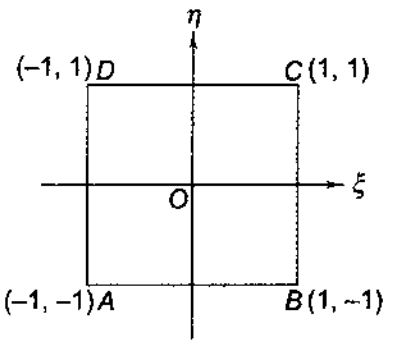


Figure 9.2 *A square domain*

9.3.2 Integration Over a Triangular Region

Figure 9.3 shows a triangular element in natural coordinates. The coordinates of the corners of the triangle are given by (0, 0), (1, 0), and (0, 1). If f is a function of three natural coordinates, it can be integrated over the triangular domain as given below

$$I = \int_0^1 \int_0^{1-\xi_1} f(\xi_1, \xi_2, \xi_3) \mathrm{d}\xi_2 \, \mathrm{d}\xi_1 \tag{9.56}$$

Note that the sum of the three natural coordinates is 1. Therefore, ξ_3 can be eliminated to express f as another function F, which is a function of the first two natural coordinates, ξ_1 and ξ_2.

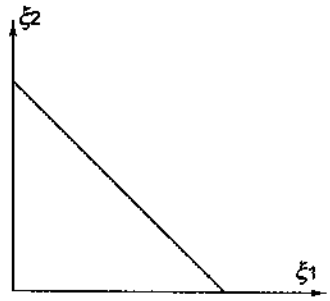


Figure 9.3 *A triangular element*

In terms of the first two natural coordinates, the integral is written as

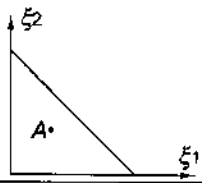
$$I = \int_0^1 \int_0^{1-\xi_2} F(\xi_1, \xi_2) \mathrm{d}\xi_1 \, \mathrm{d}\xi_2 \tag{9.57}$$

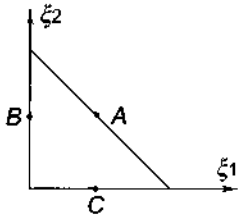
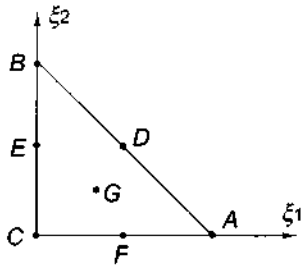
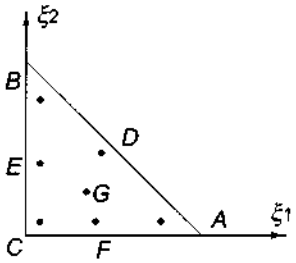
This can be integrated by applying the Gauss-point formula twice, first along ξ_1 with limit 0 to $1 - \xi_2$ and then along ξ_2 with limits 0 to 1. Note that the limits have to be transformed to lie between −1 and +1. The calculations involved in solving this integral are tedious; therefore simplified expressions have been developed for this purpose. One of the methods is as follows:

$$\int_0^1 \int_0^{1-\xi_2} F(\xi_1, \xi_2) \mathrm{d}\xi_1 \, \mathrm{d}\xi_2 = \sum_{i=1}^{n_G} w_i F\left(\xi_1^i, \xi_2^i\right) \tag{9.58}$$

The weights w_i and the coordinates $\left(\xi_1^i, \xi_2^i\right)$ of the Gauss points corresponding to various values of n_G are provided in Table 9.2. Further details on the methods of numerical integration have been provided by Carnaham, et al. (1969), Froberg (1969), Hammer, et al. (1956), and Cowper (1973).

Table 9.2 *Gauss Points and Weights for a Triangular Domain*

n_G	*Coordinates* $\left(\xi_1^i, \xi_2^i\right)$ *of the Gauss Points*	*Weights* (w_i)
1	 $A = (1/3, 1/3)$	$w_A = 1/2$ Formula exact for complete first degree polynomial

n_G	Coordinates $\left(\xi_1^i, \xi_2^i\right)$ of the Gauss points	Weights (w_i)
3	 $A = (1/2, 1/2)$ $B = (0, 1/2)$ $C = (1/2, 0)$	$w_A = w_B = w_C = 1/6$ Formula exact for complete second degree polynomial
7	 $A = (1, 0)$ $B = (0, 1)$ $C = (0, 0)$ $D = (1/2, 1/2)$ $E = (0, 1/2)$ $F = (1/2, 0)$ $G = (1/3, 1/3)$	$w_A = w_B = w_C = 3/120$ $w_D = w_E = w_F = 8/120$ $w_G = 27/120$ Formula exact for complete third degree polynomial
7	 $A = (\alpha_1, \beta_1)$, $D = (\alpha_2, \alpha_2)$, $B = (\beta_1, \alpha_1)$, $E = (\beta_2, \alpha_2)$, $C = (\beta_1, \beta_1)$, $F = (\alpha_2, \beta_2)$, $G = (1/3, 1/3)$ where $\alpha_1 = \frac{9+2\sqrt{15}}{21}$, $\alpha_2 = \frac{6+\sqrt{15}}{21}$ $\beta_1 = \frac{6-\sqrt{15}}{21}$, $\beta_2 = \frac{9-2\sqrt{15}}{21}$.	$w_A = w_B = w_C = \frac{155-\sqrt{15}}{2400}$ $w_D = w_E = w_F = \frac{155+\sqrt{15}}{2400}$ $w_G = 9/80$ Formula exact for complete fifth degree polynomial

9.4 CONCLUSION

In this chapter, various numerical integration formulae have been discussed. First, the numerical integration in one dimension using Newton-Cotes and Gauss quadrature has been described. The advantage of Gauss-quadrature scheme has been highlighted. Then, the Gauss-quadrature formula has been extended to two dimensions. The reader can also extend the formula to three dimensions.

SOLVED PROBLEMS

9.1. Using one and two Gauss-point numerical integration formulae, evaluate the following integral:

$$I = \int_3^5 \frac{1}{x} dx$$

Compare the results with the exact integration and trapezoidal rule.

Solution The exact integration is given by

$$\int_3^5 \frac{1}{x}\mathrm{d}x = \ln\left(\frac{5}{3}\right) = 0.510826 \tag{9.1.1}$$

For solving the integral by Gauss-quadrature, we have to first transform the limits from −1 to 1. Using Eq. (9.4), we obtain

$$x = \frac{5+3}{2} + \frac{(5-3)}{2}\xi = 4 + \xi \tag{9.1.2}$$

and by using Eq. (9.5), we obtain

$$\mathrm{d}x = \frac{(5-3)}{2}\mathrm{d}\xi = \mathrm{d}\xi \tag{9.1.3}$$

Thus

$$I = \int_{-1}^{1} \frac{1}{(4+\xi)}\mathrm{d}\xi \tag{9.1.4}$$

Now, we can evaluate the integral given in Eq. (9.1.4) by the following three methods:

(i) **One-Gauss-point formula**

The Gauss point is $\xi = 0$ and the weight is 2. Thus

$$I = 2 \times \frac{1}{4} = 0.5$$

(ii) **Two-Gauss-point formula**

The Gauss points are $\xi = -1/\sqrt{3}$ and $\xi = 1/\sqrt{3}$ and the weight for both points is 1. Thus

$$I = \frac{1}{\left(4 - 1/\sqrt{3}\right)} + \frac{1}{\left(4 + 1/\sqrt{3}\right)} = \frac{24}{47} = 0.510638$$

(iii) **Trapezoidal rule**

The sampling points are $\xi = -1$ and $\xi = 1$. Using Eq. (9.12), we obtain

$$I = \frac{1}{3} + \frac{1}{5} = \frac{8}{15} = 0.53333333$$

It is observed that even the two-Gauss-point formula gives a fairly accurate result; the results match up to three decimal places. As the integrand is not a finite degree polynomial, there will always be some errors on using Gauss–Legendre and Newton–Cotes quadratures. It is also seen that the trapezoidal rule, which also uses two points, provides a lesser accuracy compared to the two-point Gauss quadrature.

EXERCISES

A. Multiple Choice Questions

9.1 Simpson's one-third rule is an example of

(a) Newton–Cotes quadrature
(b) Gauss–Legendre quadrature
(c) trapezoidal rule
(d) Newton–Raphson method

9.2 Which is the most accurate quadrature among the following methods?
(a) Trapezoidal rule
(b) Simpson's one-third rule
(c) Simpson's 3/8 rule
(d) Newton–Cotes quadrature with two sampling points

9.3 For exactly integrating a polynomial of degree 10, the number of Gauss points required in Gauss quadrature is
(a) 3 (b) 4 (c) 5 (d) 6

9.4 The value obtained after the integration of $\left(1-\frac{1}{2}\xi_1-\frac{1}{2}\xi_2\right)$ in the triangular domain (unit intercepts on the axes) shown in Fig. 9.4 is
(a) 1/3 (b) 1/8 (c) 3/8 (d) 3/4

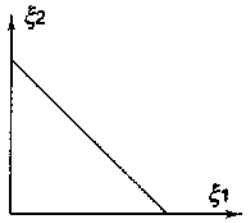


Figure 9.4 *A triangular domain*

9.5 For the exact integration of function $x^2 y$ in a square domain, the minimum number of sampling points needed in Gauss quadrature is
(a) 2 (b) 3 (c) 4 (d) 5

B. Problems

9.6 Integrate the following function:

$$f(x)=10+20x-\frac{3x^2}{10}+\frac{4x^3}{100}-\frac{5x^4}{1000}+\frac{6x^5}{10000}$$

between 8 and 12 using three-point Gauss formulation. The three points and their corresponding weights are as follows:

Points	Weights
0	8/9
$\pm\sqrt{0.6}$	5/9

9.7 Consider the element shown in Fig. 9.5, which is used to solve heat-conduction problem. The length of each side is 4 mm.

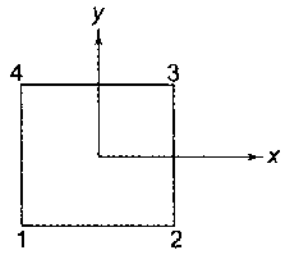


Figure 9.5 *Square element used in heat-conduction problem*

The uniform heat generation in the element is

$$Q = (2.0 - |x|)(2.0 - |y|)\,\mathrm{W/m^3}$$

where x and y are in mm. Obtain the load vector due to heat generation using the appropriate Gauss-point formula. Use the following table

Number of Gauss points	*Location*	*Weight*
1	0.0	2.0
2	± 0.5773502692	1.0
3	± 0.7745966692	0.55555555556
	0.0	0.88888888889

9.8 Integrate the expression $x^2 + y^2$ over the triangle shown in Fig. 9.6.

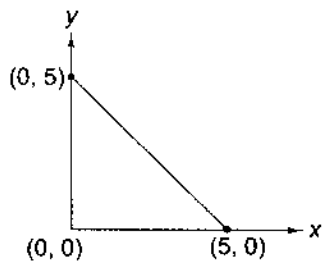


Figure 9.6 *Triangular domain for Problem 9.8*

9.9 Consider the evaluation of $\int_{-1}^{+1}\int_{-1}^{+1} f(x, y)\,dx\,dy$. How many total Gauss points should be used for the evaluation of it, if the $f(x, y)$ is

(a) $x^2 + y^2$ (b) $3x^2y^3 + y^4$ (c) $\dfrac{x+y}{x-y}$

REFERENCES

Carnahan, B, Luther, HA, and Wilkes, JO, 1969, *Applied Numerical Methods*, John Wiley, New York.

Cowper, GR, 1973, 'Gaussian quadrature formulas for triangles', *International Journal for Numerical Methods in Engineering*, Vol. 7, pp. 405-408.

Froberg, CE, 1969, *Introduction to Numerical Analysis*, Addison-Wesley, Reading, MA.

Hammer, PC, Marlowe, OJ, and Stroud, AH, 1956, 'Numerical integration over simplexes and cones', *Mathematical Tables and other Aids to Computations*, Vol. 10, pp. 130–137, The National Research Council, Washington, DC.

Mathews, JH, 2001, *Numerical Methods for Mathematics, Science and Engineering, Second edition*, Prentice-Hall of India, New Delhi.

10
CHAPTER

Two-Dimensional FEM: Detailed Discussion

10.1 Introduction

10.2 Natural Coordinates and Iso-Parametric, Sub-Parametric, and Super-Parametric Elements

10.3 Four-Noded Quadrilateral Elements

10.4 Serendipity Elements

10.5 Eight-Noded Curvilinear Elements

10.6 Numerical Evaluation of Coefficient Matrices and Right-Hand-Side Vectors for Two-Dimensional Elements

10.7 Conclusion

10.1 INTRODUCTION

In this chapter, we shall provide more details regarding two-dimensional FEM. First, it will be emphasized that the use of natural coordinates is very convenient in FEM formulation since they allow us to write the formulae for the shape functions of a particular type of element. When the functions are expressed in the form of natural coordinates, it becomes very convenient to carry out numerical integration. The type of mapping between physical and natural coordinates influences the behaviour of the element. In this context, a description of iso-parametric, sub-parametric, and super-parametric elements have been provided. Certain other types of element have also been introduced in the chapter.

10.2 NATURAL COORDINATES AND ISO-PARAMETRIC, SUB-PARAMETRIC, AND SUPER-PARAMETRIC ELEMENTS

You have already become familiar with natural coordinates and have seen how convenient it becomes to perform numerical integration in natural coordinates. This is not the only advantage of using natural coordinates, as will be clear after studying this chapter. Let us begin this discussion by learning how to deal with natural coordinates for one-dimensional elements.

Consider a two-noded rod element, where the coordinates of the two nodes are given by x_1 and x_2. We need to transform the coordinates from physical to natural coordinate system, such that x_1 transforms to -1 and x_2 transforms to $+1$. A linear transformation of the form $x = a + b\xi$ is appropriate for this purpose. The constants a and b can be easily determined. We already have carried out linear interpolation of the primary variable of the problem using C^0 continuity shape functions. The physical variable can be interpolated in the same way. Thus

$$x = N_1 x_1 + N_2 x_2 \tag{10.1}$$

where $N_1 = (1 - \xi)/2$ and $N_2 = (1 + \xi)/2$.

The elements in which the field variables and physical variables are approximated in the same way are called *iso-parametric elements*. If we use three nodes (instead of two) in an element and approximate a primary variable such as temperature in the following way:

$$T = N_1 T_1 + N_2 T_2 + N_3 T_3 \tag{10.2}$$

and transform the physical coordinate by Eq. (10.1), it will be called *sub-parametric element* and the corresponding formulation will be known as *sub-parametric formulation*. In sub-parametric formulation, the primary variable is approximated by a higher degree of polynomial than the geometry. In super-parametric formulation, the reverse case holds true.

Let the stiffness matrix be written as

$$[K^e] = \int_{x_1}^{x_2} \begin{Bmatrix} \mathrm{d}N_1/\mathrm{d}x \\ \mathrm{d}N_2/\mathrm{d}x \end{Bmatrix} \begin{bmatrix} \mathrm{d}N_1/\mathrm{d}x & \mathrm{d}N_2/\mathrm{d}x \end{bmatrix} \mathrm{d}x \tag{10.3}$$

In Eq. (10.3), we have to transform every variable to natural coordinates. From Eq. (10.1), we have

$$\mathrm{d}x = \frac{x_2 - x_1}{2}\mathrm{d}\xi = J\,\mathrm{d}\xi \tag{10.4}$$

where J is called the *Jacobian* and is defined as the ratio of the sizes of the element in physical and natural coordinate systems. Basically, in one dimension, the Jacobian is the ratio of infinitesimal length in the physical coordinates to the corresponding mapped length in natural coordinates. If the transformation is linear, the value of the Jacobian is same everywhere. In two dimensions, the determinant of the Jacobian matrix is the ratio of infinitesimal area in physical coordinate system to that in natural coordinate system. In three dimensions, the determinant of the Jacobian matrix is the ratio of infinitesimal volume in physical coordinate system to that in natural coordinate system.

To determine the derivative of the shape function with respect to x, we employ chain rule. Thus

$$\frac{\mathrm{d}N_1}{\mathrm{d}x} = \frac{\mathrm{d}N_1}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\mathrm{d}x} = J^{-1}\frac{\mathrm{d}N_1}{\mathrm{d}\xi} \tag{10.5}$$

Similarly

$$\frac{\mathrm{d}N_2}{\mathrm{d}x} = \frac{\mathrm{d}N_2}{\mathrm{d}\xi}\frac{\mathrm{d}\xi}{\mathrm{d}x} = J^{-1}\frac{\mathrm{d}N_2}{\mathrm{d}\xi} \tag{10.6}$$

By substituting Eqs (10.4) to (10.6) in Eq. (10.3) and integrating, it can be verified that

$$[K^e] = \frac{1}{(x_2 - x_1)}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{10.7}$$

For a beam element, the field variable (deflection) is approximated as a cubic polynomial, but the geometry is approximated by linear functions only. Therefore, it is a sub-parametric formulation. The stiffness matrix for a beam element is written as

$$\left[K^e\right] = EI\int_{x_1}^{x_2}\{N''\}\lfloor N''\rfloor \mathrm{d}x \tag{10.8}$$

where E is the Young's modulus of elasticity and I is the second moment of inertia of the cross-section about the axis perpendicular to neutral axis and load. We can transform dx as discussed earlier. It can be easily determined that

$$\frac{\mathrm{d}^2N_i}{\mathrm{d}x^2} = (J^{-1})^2\frac{\mathrm{d}^2N_i}{\mathrm{d}\xi^2} \tag{10.9}$$

10.3 FOUR-NODED QUADRILATERAL ELEMENTS

In quadrilateral elements, we can transform the geometry in the following manner:

$$x = \sum_{i=1}^{4} N_i x_i, \qquad y = \sum_{i=1}^{4} N_i y_i \tag{10.10}$$

Cook, et al. (1989) have provided an interesting exercise problem for a quadrilateral element. The problem is discussed as follows:

Let us use the quadrilateral element shown in Fig. 10.1 in the finite element model. Imagine that its x-direction displacement field is $u = a_1 + a_2x + a_3y + a_4xy$, where the a_i are constants. How does u vary with x or y along each side? Do you think the element will be compatible with its neighbours?

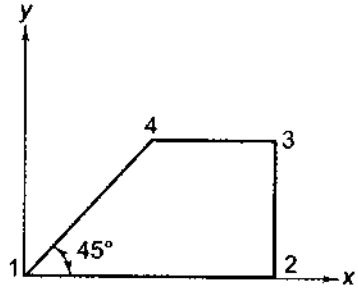


Figure 10.1 *A quadrilateral element*

Let us solve this problem. Along the line 1–2, $y = 0$, i.e., $u = a_1 + a_2x$. Thus, the displacement field is linear. Along the line 2–3, $x = c$ (a constant), $u = a_1 + a_2c + a_3y + a_4cy$, which is linear in y. Similarly, it can be shown that the approximated displacement field is linear along the line 3–4, while along 1–4, $y = x$, and therefore, $u = a_1 + a_2y + a_3y + a_4y^2$. Thus, the displacement varies in a quadratic manner along 1–4. A quadratic function cannot be fitted just by the nodal values of 1 and 4, the other nodes will influence it. Therefore, this element is not compatible. If there is another element whose one edge coincides with 1–4, then the approximate displacement field need not be the same for both the elements. The reason is that only nodes 1 and 4 are common between the two elements, not the other nodes.

Now, if the element is transformed into an square element using Eq. (10.10), then the displacement field is approximated as

$$u = N_1u_1 + N_2u_2 + N_3u_3 + N_4u_4 \tag{10.11}$$

where N_i are the Lagrangian shape functions. Now, along the line 1–4, N_3 and N_2 are 0, while the value of the natural coordinate ξ is equal to –1 along the line 1–4. Therefore

$$u = \frac{(1-\eta)}{2}u_1 + \frac{(1+\eta)}{2}u_4 \tag{10.12}$$

This is a linear equation. The displacement field along the line 1–4 depends on the nodal values of the nodes 1 and 4. The same thing will hold true for the element that interfaces with this element along 1–4. Therefore, this element will be compatible.

10.4 SERENDIPITY ELEMENTS

The term 'serendipity' literally means 'the faculty of making fortunate discoveries by accident'. The *serendipity elements* of FEM are called so because they were probably discovered as a result of numerical experimentation and were found to be useful for FEM formulations. You will see that the derivation of shape functions is carried out in an unusual manner. In this section, the derivation of shape functions for an eight-noded rectangular serendipity element will be carried out to clarify the concepts to the readers.

Figure 10.2 shows an eight-noded rectangular serendipity element. It is similar to a nine-noded element, but there is no internal node. Let us derive the shape functions for this element.

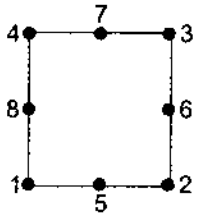


Figure 10.2 *An eight-noded serendipity element*

For C^0 element, the shape functions should be such that the value of the shape function should be 1 at the corresponding node and 0 at the other nodes. Also, the sum of the shape functions should be 1. Moreover, each shape function should be a quadratic function. We shall derive the expressions in natural coordinates, in which node 1 has the coordinates (–1, 1) and node 2 has the coordinates (1, 1). Let us choose N_1 corresponding to a four-noded square element, such that

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta) \tag{10.13}$$

This shape function is 1 at node 1 and 0 at the nodes 2, 3, 4, 6, and 7. However, what about its values at nodes 5 and 8? At node 5, $\xi = 0$ and $\eta = -1$ and therefore, $N_1 = ½$. Similarly, at node 8, $x = -1$ and $\eta = 0$, thus giving $N_1 = ½$. How can we modify the shape function in Eq. (10.13) so that it becomes 0 at nodes 5 and 8 without affecting its value at other nodes? A little thought will show that the following shape function can achieve this purpose:

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta) - \frac{1}{2}N_5 - \frac{1}{2}N_8 \tag{10.14}$$

At node 5, N_5 is 1 and therefore the term $-\frac{1}{2}N_5$ is equal to –1/2, thus making the overall expression 0. The last term does the same thing at node 8. At other nodes neither the last term nor the second last term has any effect, as these terms are 0 by the very characteristic of the shape functions.

Similarly, we can write the other shape functions as

$$N_2 = \frac{1}{4}(1+\xi)(1-\eta) - \frac{1}{2}N_5 - \frac{1}{2}N_6 \tag{10.15}$$

$$N_3 = \frac{1}{4}(1+\xi)(1+\eta) - \frac{1}{2}N_7 - \frac{1}{2}N_6 \tag{10.16}$$

$$N_4 = \frac{1}{4}(1-\xi)(1+\eta) - \frac{1}{2}N_7 - \frac{1}{2}N_8 \tag{10.17}$$

Note that

$$\begin{aligned}\sum_{i=1}^{8} N_i &= \frac{1}{4}(1-\xi)(1-\eta) - \frac{N_5}{2} - \frac{N_8}{2} + \frac{1}{4}(1+\xi)(1-\eta) - \frac{N_5}{2} - \frac{N_6}{2} \\ &+ \frac{1}{4}(1+\xi)(1+\eta) - \frac{N_6}{2} - \frac{N_7}{2} + \frac{1}{4}(1-\xi)(1+\eta) - \frac{N_8}{2} - \frac{N_7}{2} \\ &+ N_5 + N_6 + N_7 + N_8\end{aligned} \tag{10.18}$$

This simplifies to

$$\sum_{i=1}^{8} N_i = \frac{1}{4}(1-\xi)(1-\eta) + \frac{1}{4}(1+\xi)(1-\eta) + \frac{1}{4}(1+\xi)(1+\eta) + \frac{1}{4}(1-\xi)(1+\eta) \tag{10.19}$$

The expression given in Eq. (10.19) is the sum of the Lagrangian functions of a four-noded element, which is 1. Thus, the shape functions given in Eqs (10.14) to (10.17) can serve our purpose. However, we have to find the expressions for N_5, N_6, N_7, and N_8. If we consider N_5 to be the product of a quadratic shape function in ξ and a linear shape function in η, our problem will be solved, for then the shape function will be 0 at $\xi = -1$ and $+1$ and $\eta = 1$, and it will be non-zero and equal to 1 only at $\xi = 0$ and $\eta = -1$, which are the conditions we needed to satisfy. Thus

$$N_5 = \frac{1}{2}(1-\eta)(1-\xi^2) \tag{10.20}$$

Similarly

$$N_6 = \frac{1}{2}(1+\xi)(1-\eta^2) \tag{10.21}$$

$$N_7 = \frac{1}{2}(1+\eta)(1-\xi^2) \tag{10.22}$$

$$N_8 = \frac{1}{2}(1-\xi)(1-\eta^2) \tag{10.23}$$

Thus, we have derived all the shape functions for an eight-noded serendipity element.

10.5 EIGHT-NODED CURVILINEAR ELEMENTS

An eight-noded curvilinear element can be transformed to a square element by the following transformations:

$$x = \sum_{i=1}^{8} N_i x_i, \qquad y = \sum_{i=1}^{8} N_i y_i \tag{10.24}$$

where the shape functions correspond to eight-noded square serendipity element and they are functions of the natural coordinates. For this element

$$\frac{\partial N_i}{\partial \xi} = \frac{\partial N_i}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial N_i}{\partial y}\frac{\partial y}{\partial \xi} \tag{10.25}$$

and

$$\frac{\partial N_i}{\partial \eta} = \frac{\partial N_i}{\partial x}\frac{\partial x}{\partial \eta} + \frac{\partial N_i}{\partial y}\frac{\partial y}{\partial \eta} \tag{10.26}$$

Transforming Eqs (10.25) and (10.26) in the matrix form, we obtain

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} \\ \dfrac{\partial N_i}{\partial \eta} \end{Bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial x}{\partial \eta} & \dfrac{\partial y}{\partial \eta} \end{bmatrix} \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix} = [J] \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix} \tag{10.27}$$

where $[J]$ is the Jacobian and its determinant is equal to $(dx\, dy)/(d\xi\, d\eta)$. Note that in most of the mathematics books, the Jacobian is defined as $[J]^T$. The determinants of $[J]$ and $[J]^T$ are of course equal.

From Eq. (10.27), it is seen that

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \end{Bmatrix} = [J]^{-1} \begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} \\ \dfrac{\partial N_i}{\partial \eta} \end{Bmatrix} \tag{10.28}$$

The elements of Jacobian matrix can be computed as

$$\frac{\partial x}{\partial \xi} = \sum_{i=1}^{8} \frac{\partial N_i}{\partial \xi} x_i \text{ and so on} \tag{10.29}$$

10.6 NUMERICAL EVALUATION OF COEFFICIENT MATRICES AND RIGHT-HAND-SIDE VECTORS FOR TWO-DIMENSIONAL ELEMENTS

Coefficient matrices and right-hand-side vectors of two-dimensional elements are evaluated using Gauss quadrature. For this purpose, the elements are transformed to natural coordinates; the integrand is expressed in terms of the natural coordinates. For double integration, $dxdy$ is replaced by $|J|d\xi d\eta$, where

$$|J| = \begin{vmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial x}{\partial \eta} & \dfrac{\partial y}{\partial \eta} \end{vmatrix} \tag{10.30}$$

The boundary integrals are transformed with the help of the relation

$$\int_{\Gamma} (\cdots) d\Gamma = \int_{-1}^{+1} (\cdots) |J_b| d\zeta \tag{10.31}$$

where

$$|J_b| = \sqrt{\left(\frac{dx}{d\zeta}\right)^2 + \left(\frac{dy}{d\zeta}\right)^2} \tag{10.32}$$

Along the boundary, the coordinates (x, y) are approximated using one-dimensional shape functions.

10.7 CONCLUSION

In this chapter, the importance of natural coordinates has been highlighted. The method of obtaining the derivatives in physical coordinates from the known derivatives in the natural coordinates has been explained in this chapter. Four-noded quadrilateral elements and eight-noded serendipity elements have been introduced. As the concepts of numerical integration and Jacobian have been clearly defined and explained, the readers should now be able to write their own FEM codes for solving two-dimensional problems. Some related problems have been included in the exercises.

SOLVED PROBLEMS

10.1 The element shown in Fig. 10.3 is a transition element used to connect a four-noded element with a nine-noded element. Find out all the shape functions for this element.

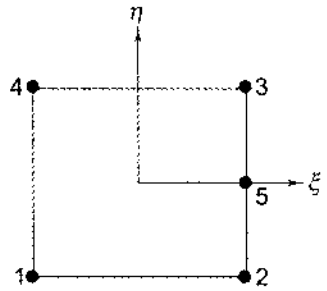


Figure 10.3 *A transition element*

Solution We derive the shape functions of this element following the methodology described in Section 10.4. The shape functions of a four-noded Lagrangian element are given by

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta), \qquad N_2 = \frac{1}{4}(1+\xi)(1-\eta)$$

$$N_3 = \frac{1}{4}(1+\xi)(1+\eta), \qquad N_4 = \frac{1}{4}(1-\xi)(1+\eta)$$

We observe that N_1 is 1 at node 1 and 0 at all other nodes, and N_4 is 1 at node 4 and 0 at all other nodes. Hence, they can be suitable shape functions for nodes 1 and 4, respectively. The shape function N_2 is 1 at node 2 and 0 at other nodes except at node 5. At node 5, its value is ½. We can replace it by the following shape function:

$$N_2 = \frac{1}{4}(1+\xi)(1-\eta) - \frac{N_5}{2}$$

The shape function N_5 will be a suitable shape function that will be 1 at node 5 and 0 at the other nodes. This will ensure that N_2 is 1 at node 2 and 0 at other nodes. In the same way, we can write

$$N_3 = \frac{1}{4}(1+\xi)(1-\eta) - \frac{N_5}{2}$$

We observe that

$$N_1 + N_2 + N_3 + N_4 + N_5 = \frac{1}{4}(1-\xi)(1-\eta) + \frac{1}{4}(1+\xi)(1-\eta) + \frac{1}{4}(1+\xi)(1+\eta) - \frac{N_5}{2}$$

$$+ \frac{1}{4}(1-\xi)(1+\eta) - \frac{N_5}{2} + N_5 = 1$$

Thus, the required property of shape functions of C^0 continuity elements is satisfied. Now, we obtain the expression for N_5. The line 2–5–3 is a three-noded linear element. If we consider the shape function $(1-\eta^2)$, it will be 0 at nodes 2 and 3 and 1 at node 5. Interestingly, it will be 0 at nodes 1 and 4 also. Therefore, such a function may be chosen. However, this function is not 0 on the entire line 1–4. Thus, the primary variable will not be a linear interpolation function on line 1–4. If we want to

ensure that primary variable is interpolated linearly on 1– 4, it is better to use the following shape function for N_5:

$$N_5 = \frac{(1+\xi)}{2}\left(1-\eta^2\right)$$

This function makes N_5 zero at line 1–4 and $(1-\eta^2)$ at line 2–3.

Thus, the five shape functions are given by

$$N_1 = \frac{1}{4}(1-\xi)(1-\eta), \quad N_2 = \frac{1}{4}(1+\xi)(1-\eta) - \frac{(1+\xi)}{4}\left(1-\eta^2\right)$$

$$N_3 = \frac{1}{4}(1+\xi)(1+\eta) - \frac{(1+\xi)}{4}\left(1-\eta^2\right), \quad N_4 = \frac{1}{4}(1-\xi)(1+\eta),$$

$$N_5 = \frac{(1+\xi)}{2}\left(1-\eta^2\right)$$

10.2 Transform the quadrilateral element shown in Fig. 10.4 into a square element in natural coordinates. Find out the Jacobian and its determinant.

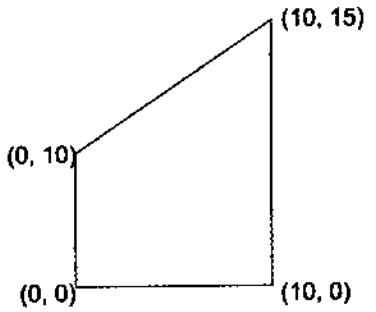


Figure 10.4 *A quadrilateral element*

Solution Using Eq. (10.10), we can write the variables as

$$x = \frac{10}{4}(1+\xi)(1-\eta) + \frac{10}{4}(1+\xi)(1+\eta) = 5(1+\xi)$$

and

$$y = \frac{15}{4}(1+\xi)(1+\eta) + \frac{10}{4}(1-\xi)(1+\eta) = (1+\eta)(6.25+1.25\xi)$$

Using Eq. (10.27), the Jacobian is written as

$$[J] = \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial x}{\partial \eta} & \dfrac{\partial y}{\partial \eta} \end{bmatrix} = \begin{bmatrix} 5 & 1.25(1+\eta) \\ 0 & (6.25+1.25) \end{bmatrix}$$

The determinant of the Jacobian is given by

$$|J| = 31.25 + 6.25\xi$$

Observe that the determinant is not constant.

EXERCISES

A. Multiple Choice Questions

10.1 In a two-dimensional finite element procedure for solving heat-conduction problems, the geometry is approximated by nine-noded Lagrangian elements and temperature by four-noded Lagrangian elements. The type of formulation used in this procedure is called

(a) sub-parametric (b) iso-parametric (c) super-parametric (d) consistent

10.2 The four nodes of a quadrilateral element for solving heat-conduction problem are (0, 0), (10, 0), (10, 20), and (0, 10). At which of the following points will the temperature be the average of the four nodal temperatures in an iso-parametric formulation?

(a) (0, 5) (b) (5, 7.5) (c) (5, 10) (d) (10, 15)

10.3 In an eight-noded serendipity element, the value of each shape function associated with the corner nodes at the center point, i.e., at $\xi = \eta = 0$, is

(a) 0.5 (b) 0.25 (c) -0.25 (d) 0

10.4 A rectangular element of size 50×100 is mapped into a square domain of size 2×2. The determinant of Jacobian is

(a) 1250 (b) 2500 (c) 5000 (d) 4

10.5 A cubic element of size $8 \times 8 \times 8$ is mapped into a cubic element of size $2 \times 2 \times 2$. The determinant of Jacobian is

(a) 4 (b) 8 (c) 16 (d) 64

B. Problems

10.6 In Fig. 10.5, the temperatures at the coordinates (0, 0), (10, 0), (10, 15), and (0,10) are 30 K, 60 K, 60 K, and 20 K, respectively. Find out the partial derivative of the temperature with respect to horizontal x coordinate at the point (5, 6.25). Consider iso-parametric formulation.

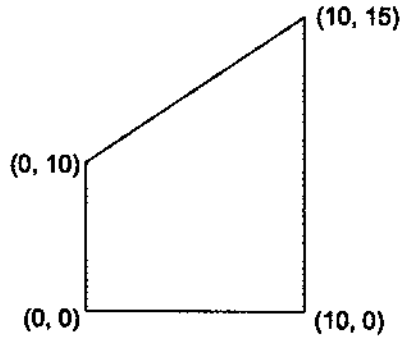


Figure 10.5 *A quadrilateral element for Problem 10.6*

10.7 Write down the expression for all the nine shape functions of a nine-noded Lagrangian element (Fig. 10.6). The physical coordinates of the element are shown in the figure. Find out the physical coordinate at $\xi = 0.5$ and $\eta = 0.5$. The natural coordinates of the nodes are (– 1, – 1), (– 1, 1), (1, 1), (– 1, 1), (0, – 1), (1, 0), (0, 1), (– 1, 0), and (0, 0).

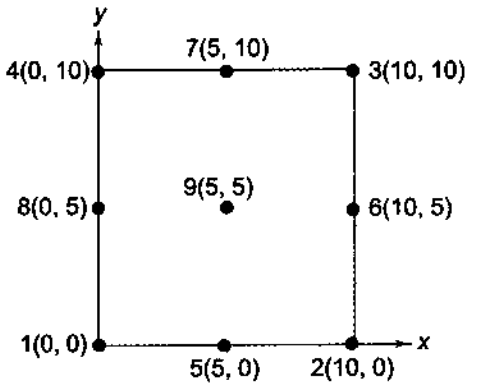


Figure 10.6 *A nine-noded Lagrangian element*

10.8 For a quadrilateral element shown in Fig. 10.7, the temperatures at nodes 1, 2, 3, and 4 are 100°C, 150°C, 50°C, and 70°C, respectively. This element is to be used in an iso-parametric FEM formulation. The shape functions are chosen such that the element is compatible with the adjoining elements at all the other edges. Find the temperatures at points (7.5, 5) and (5, 5).

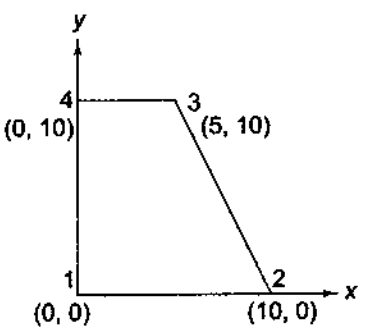


Figure 10.7 *A quadrilateral element*

REFERENCES

Cook, RD, Malkus, DS, and Plesha, ME, 1989, *Concepts and Applications of Finite Element Analysis, Third edition*, John Wiley, New York.

11

CHAPTER

FEM Formulation for Plane Stress and Plane Strain Problems

11.1 Introduction
11.2 Basic Equations
11.3 Boundary Conditions
11.4 Index Notation
11.5 FEM Formulation
11.6 Types of Elements
11.7 Numerical Evaluation of Element Matrices and Vectors
11.8 Assembly of Element Matrices
11.9 Boundary Conditions and Solutions
11.10 Gradient Estimates
11.11 Stress Analysis Using an FEM Package: An Example
11.12 Summary

11.1 INTRODUCTION

Consider a linear elastic solid of domain Ω having a uniform thickness bounded by two parallel planes. On any of these parallel planes, let the closed boundary of the solid be denoted by Γ. The top view (in a plane perpendicular to the thickness of the solid) is shown in Fig. 11.1 along with the imposed geometric constraints and a distributed load on the solid. The meaning of these constraints is explained in Fig. 11.2, in which u denotes the displacement in x direction and v denotes the displacement in y direction. If the thickness in z direction is small compared to the size Ω of the domain, the problem may be approximated as a *plane stress problem*. In case of plane stress problems, the following assumptions are made. The body forces, if they exist, cannot vary in the direction of the thickness and cannot have components in the z direction; the applied boundary forces must be uniformly distributed across the thickness (i.e., they must be constant in the z direction); and no loads can be applied on the parallel planes bounding the bottom and top surfaces. The assumption that the forces are zero on the parallel planes implies that for plane stress problems the stresses in the z-plane (the plane whose normal is along z direction) are zero. The statement can be written as

$$\sigma_{zx} = 0, \quad \sigma_{zy} = 0, \quad \sigma_{zz} = 0 \tag{11.1}$$

Owing to the symmetry of stress components, we can write

$$\sigma_{xz} = \sigma_{zx} = 0, \quad \sigma_{yz} = \sigma_{zy} = 0 \tag{11.2}$$

Thus, in plane stress problems, the possible non-zero stress components are σ_{xx}, σ_{yy}, σ_{xy}, and σ_{yx}. Because of symmetry, σ_{xy} is equal to σ_{yx}. Thus, one needs to find three components of stress at each point in a plane stress problem.

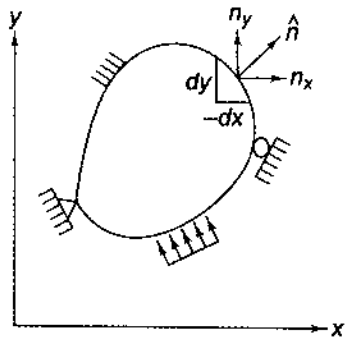


Figure 11.1 *A solid of domain Ω alongwith the geometric constraints and distributed load*

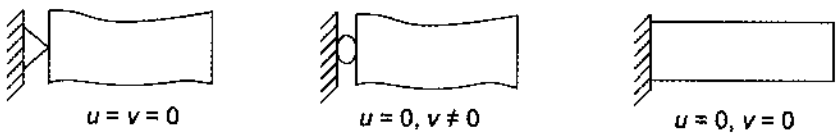


Figure 11.2 *Graphical symbols for geometric constraints: (a) pin-joint (b) roller-support (c) fixed*

Plane strain is defined as a deformation state in which there is no deformation in one direction, say z direction, and the deformations in other directions are functions of x and y but not of z. Thus, there is no stretching along the z direction and a line parallel to z direction always remains orthogonal to x and y direction. This implies that strain components involving z-axis are zero, i.e., $\varepsilon_{zz} = \varepsilon_{yz} = \varepsilon_{zx} = 0$. The stresses and strains in different materials are related in different ways. Materials can be classified into three types on the basis of their properties: materials that have different properties in different directions are called *anisotropic materials*; materials whose properties are not dependent on the direction are called *isotropic materials*; and materials that have properties that are different in three mutually perpendicular directions at a point and have three mutually perpendicular planes of material symmetry are called *orthotropic materials*. For isotropic and orthotropic materials in a plane strain state, the non-zero stress components are $\sigma_{xx}, \sigma_{yy}, \sigma_{xy}$ and σ_{zz}. However, σ_{zz} is not an independent component and can be obtained if σ_{xx} and σ_{yy} are known. For example, for an isotropic material in a plane strain

$$\varepsilon_{zz} = \frac{1}{E}\left\{\sigma_{zz} - \nu\left(\sigma_{xx} + \sigma_{yy}\right)\right\} = 0 \tag{11.3}$$

where E is Young's modulus of elasticity and ν is Poisson's ratio. Thus, for an isotropic material in plane strain

$$\sigma_{zz} = \nu\left(\sigma_{xx} + \sigma_{yy}\right) \tag{11.4}$$

Similarly, one can relate σ_{zz} to σ_{xx} and σ_{yy} for an orthotropic material.

As there are only three independent components of stress in plane stress as well as plane strain problems for orthotropic and isotropic materials, the corresponding FEM formulations for both types of problems are similar. The only difference lies in the constitutive matrices for both problems. In this chapter, FEM formulation for plane stress and plane strain problems will be discussed.

11.2 BASIC EQUATIONS

The governing equations of motion for the plane elasticity problems are given by

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + f_x = \rho \frac{\partial^2 u}{\partial t^2} \tag{11.5}$$

$$\frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + f_y = \rho \frac{\partial^2 v}{\partial t^2} \tag{11.6}$$

where f_x and f_y denote the body forces per unit volume along the x and y directions, respectively; ρ is the density of the material; σ_{xx} and σ_{yy} are the normal stresses; u and v are the displacements along x and y directions, respectively; and σ_{xy} is the shear stress. Strain–displacement relations are given by

$$\varepsilon_{xx} = \frac{\partial u}{\partial x}, \quad \varepsilon_{yy} = \frac{\partial v}{\partial y}, \quad 2\varepsilon_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \tag{11.7}$$

For plane stress problems, stress and strain are related by the constitutive matrix $[D]$ in the following manner:

$$\{\sigma\} = \begin{Bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{Bmatrix} = \begin{bmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ 2\varepsilon_{xy} \end{Bmatrix} = [D]\{\varepsilon\} \tag{11.8}$$

where d_{ij} ($d_{ij} = d_{ji}$) are the elasticity (material) constants for an orthotropic material with the material principal directions coinciding with the coordinate axes (x, y) used to describe the problem. For an isotropic material in plane stress, d_{ij} are given by

$$d_{11} = d_{22} = \frac{E}{1-\nu^2}, \quad d_{12} = d_{21} = \frac{E\nu}{1-\nu^2}, \quad d_{33} = \frac{E}{2(1+\nu)} \tag{11.9}$$

For plane strain problems

$$d_{11} = d_{22} = \frac{E(1-\nu)}{(1+\nu)(1-2\nu)}, \quad d_{12} = d_{21} = \frac{E\nu}{(1+\nu)(1-2\nu)}, \quad d_{33} = \frac{E}{2(1+\nu)} \tag{11.10}$$

If we substitute constitutive and strain–displacement relations in Eqs (11.5) and (11.6), we get two second-order differential equations for the unknown displacements u and v. Solving the differential equations for the displacements, strains can be found by using Eq. (11.7) and stresses by Eq. (11.8). However, for solving the differential equations, we need boundary conditions, which are described in the following section.

11.3 BOUNDARY CONDITIONS

For the given problem, the essential or geometric boundary conditions are written as

$$u = u^*, \quad v = v^* \quad \text{on } \Gamma_u \tag{11.11}$$

where u^* and v^* are the prescribed displacement components, while the natural boundary conditions are given by

$$t_x = \sigma_x n_x + \sigma_{xy} n_y = t_x^*, \quad t_y = \sigma_{xy} n_x + \sigma_y n_y = t_y^* \quad \text{on } \Gamma_t \tag{11.12}$$

where n_x and n_y are the components of the unit normal vector on the boundary, Γ_u and Γ_t are portions of the boundary Γ ($\Gamma = \Gamma_u \cdot \Gamma_t$), t_x^* and t_y^* are the specified boundary tractions, and u^* and v^* are the specified displacements. The traction is also called *stress vector* and is equal to the force per unit area on a prescribed surface. Only one element of each pair, (u, t_x) and (v, t_y) may be specified at a boundary point. This is because one can control either the displacement at a point or the force.

11.4 INDEX NOTATION

Before discussing the finite element formulation, we shall describe index notation for representing the components of vectors and tensors. A *tensor* of order 2 is a physical quantity that has nine components in a three-dimensional space and transforms according to a definite rule under the rotation of the reference frame. Physical laws can be expressed using tensors in a form that applies to any coordinate system. The examples of the second-order tensors are *stress* and *strain*. In the index notation, stress

components are represented as σ_{ij}, where i indicates the plane and j denotes the direction. In three-dimensional space, the indices i and j take on the values 1, 2, and 3; thus, σ_{ij} represents nine components of the stress tensor. Similarly, C_{ijkl} represents 81 components of a fourth-order tensor. A tensor of order N has 3^N components in three-dimensional space. Index notations are useful in expressing the mathematical expressions containing scalars, vectors, and tensors in a concise way. In this book, we shall use indices only in the form of subscripts.

A subscript, which is not repeated in a term is called a *free index*, e.g., σ_{ij} is a component of tensor having two free indices. The nine components of the tensor can be represented over the range of subscripts 1 to 3 as

$$\left[\sigma_{ij}\right] = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{11.13}$$

The repeated subscript in a term is called a *dummy index*. Einstein provided a summation convention according to which when an index is repeated in a term, it implies that we are summing over all possible terms that are obtained by assigning different values to repeated index in the range. For example, consider the following expression, relating the traction vector, stress tensor, and direction cosines of the normal to the plane

$$\sigma_{ij} n_j = t_i \tag{11.14}$$

The relation given in Eq. (11.14) is a Cauchy's relation, considering the stress tensor as a symmetric tensor. In this equation, j is repeated twice on the left-hand side; therefore, it is a dummy index. On the other hand, i is a free index that is not repeated. Considering the three-dimensional Cartesian coordinate system in which subscripts i and j vary from 1 to 3, Eq. (11.14) can be expanded as follows:

$$\sigma_{i1} n_1 + \sigma_{i2} n_2 + \sigma_{i3} n_3 = t_i \tag{11.15}$$

In Eq. (11.15), summation convention is applied in which the dummy index j takes all values from 1 to 3 in the same equation. Equation (11.15) itself represents three equations corresponding to three values of the free index. Thus, Eq. (11.15) can be further written in an unabridged notation as

$$\begin{aligned} \sigma_{11} n_1 + \sigma_{12} n_2 + \sigma_{13} n_3 &= t_1, \\ \sigma_{21} n_1 + \sigma_{22} n_2 + \sigma_{23} n_3 &= t_2, \\ \sigma_{31} n_1 + \sigma_{32} n_2 + \sigma_{33} n_3 &= t_3 \end{aligned} \tag{11.16}$$

The rank or order of a tensor is dependent on the number of free indices attached while representing its components. For example, σ_{ij} represents the components of a second-order tensor because it has two free indices, i and j. Similarly, η_{ijk} represents the components of a third-order tensor. However, λ_{iijk} does not represent the components of a fourth-order tensor; it represents the components of a second-order tensor, as there are only two free indices, j and k. More clearly, if the range of index i is 1 to 3, then

$$\lambda_{iijk} = \lambda_{11jk} + \lambda_{22jk} + \lambda_{33jk} \tag{11.17}$$

Equation (11.17) represents the components of a second-order tensor, provided it obeys the transformation rule. Thus, the requirement is that λ_{iijk} must transform like a stress tensor under the rotation of coordinate axes.

We now introduce two important symbols that will help in reducing the complexity of equations involving tensors. The δ_{ij} is a Kronecker delta symbol; it is represented as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{11.18}$$

The symbol ε_{ijk} is known as the permutation symbol, the alternating symbol, Levy-Civita symbol, or the alternator; it represents the following conditions:

$$\varepsilon_{ijk} = \begin{cases} 1 & \text{if } i, j, k \text{ take values in the cyclic order,} \\ -1 & \text{if } i, j, k \text{ take values in the acyclic order,} \\ 0 & \text{if two or all of } i, j, k \text{ take the same value} \end{cases} \tag{11.19}$$

Thus, ε_{123} is 1, ε_{321} is -1, and ε_{113} is 0. Another important notation is the *comma notation*. If f represents the function $f(x_1, x_2, x_3,)$, then the partial derivative of f with respect to x_i may be represented using the comma notation as follows:

$$f_{,i} = \frac{\partial f}{\partial x_i}, \quad i = 1, 2, 3 \tag{11.20}$$

In the expanded form, the Eq. (11.20) denotes

$$f_{,1} = \frac{\partial f}{\partial x_1}, \quad f_{,2} = \frac{\partial f}{\partial x_2}, \quad f_{,3} = \frac{\partial f}{\partial x_3}$$

Thus, the comma notation $()_{,i}$ denotes the partial differential operator $(\partial / \partial x_i)$.

11.5 FEM FORMULATION

There are a number of techniques of FEM formulation, the most popular being Ritz and Galerkin techniques. Some other methods are collocation method, sub-domain method, least-square method, over-determined collocation method, etc. In this chapter, we will discuss FEM formulation using the Ritz method.

In the Ritz FEM method, the variables whose values are to be determined are approximated by piecewise continuous polynomials. The coefficients of these polynomials are obtained by minimizing the total potential energy of the system. In FEM, usually, these coefficients are expressed in terms of unknown values of primary variables. Thus, if an element has four nodes, the displacement field u can be approximated as

$$u = \sum_{i=1}^{4} N_i u_i \tag{11.21}$$

where u_i are the nodal displacements in x direction and N_i are the shape functions, which are functions of coordinates.

For plane elastic body, the total potential energy of an element is given by

$$\Pi^e = \int_{V^e} \frac{1}{2}\sigma_{ij}\varepsilon_{ij}\mathrm{d}V - \int_{V^e} f_i u_i \mathrm{d}V - \int_{S_t^e} t_i^* u_i \mathrm{d}S \tag{11.22}$$

where V^e denotes the volume of element e, S_t^e is the bounding surface of the domain V^e on which tractions are prescribed, σ_{ij} and ε_{ij} are the components of stress and strain tensors, respectively, and f_i and t_i^* are the components of body force (per unit volume) and boundary stress vectors (tractions), respectively. Note that the index notation with summation convention has been used for writing Eq. (11.22). In index notation, the stress components are written as

$$\sigma_{11} = \sigma_x, \quad \sigma_{12} = \sigma_{xy}, \quad \sigma_{22} = \sigma_y \tag{11.23}$$

Similarly, the body force and traction components are given by

$$f_1 = f_x, \quad f_2 = f_y, \quad t_1^* = t_x^*, \quad t_2^* = t_y^* \tag{11.24}$$

The first term in Eq. (11.22) corresponds to strain energy stored in the element, the second term represents the potential due to the body force, and the third term represent the potential due to surface forces. For plane stress problems with thickness h^e, it is assumed that all functions are independent of the thickness coordinate z. Hence

$$\begin{aligned}\Pi^e = h^e \int_{\Omega^e} \frac{1}{2}(\sigma_{xx}\varepsilon_{xx} + \sigma_{yy}\varepsilon_{yy} + 2\sigma_{xy}\varepsilon_{xy})\mathrm{d}x\mathrm{d}y \\ -h^e \int_{\Omega^e} (f_x u + f_y v)\mathrm{d}x\mathrm{d}y - h^e \int_{\Gamma_t^e} (t_x^* u + t_y^* v)\mathrm{d}\Gamma\end{aligned} \tag{11.25}$$

where f_x and f_y are the body forces per unit volume and t_x^* and t_y^* are boundary forces per unit area, Ω^e is the area of the element in $x-y$ plane, and Γ_t^e is the boundary of Ω^e on which traction is specified. Equation (11.25) can be rewritten as

$$\Pi^e = h^e \int_{\Omega^e} \frac{1}{2}\left(\begin{Bmatrix}\varepsilon_{xx}\\ \varepsilon_{yy}\\ 2\varepsilon_{xy}\end{Bmatrix}^{\mathrm{T}}\begin{Bmatrix}\sigma_{xx}\\ \sigma_{yy}\\ \sigma_{xy}\end{Bmatrix}\right)\mathrm{d}x\mathrm{d}y - h^e \int_{\Omega^e}\begin{Bmatrix}u\\ v\end{Bmatrix}^{\mathrm{T}}\begin{Bmatrix}f_x\\ f_y\end{Bmatrix}\mathrm{d}x\mathrm{d}y - h^e \int_{\Gamma_t^e}\begin{Bmatrix}u\\ v\end{Bmatrix}^{\mathrm{T}}\begin{Bmatrix}t_x^*\\ t_y^*\end{Bmatrix}\mathrm{d}\Gamma \tag{11.26}$$

Note that it is possible that on a certain boundary only one traction component is specified. In such cases, the last integral would consist of one term instead of two terms.

The finite element model of the plane elasticity equations is developed using the matrix form given in Eq. (11.26). The displacements u and v are approximated by the Lagrange family of interpolation functions (shape functions). Let u and v be approximated over Ω^e by the finite element interpolations

$$u \approx \sum_{i=1}^{n} u_i^e N_i^e(x,y), \quad v \approx \sum_{i=1}^{n} v_i^e N_i^e(x,y) \tag{11.27}$$

where n is the number of nodes representing the element e, N_i^e are the displacement shape functions, and u_i^e and v_i^e are the nodal displacements in x and y directions, respectively. The displacements and strains over element e are given by

$$\begin{Bmatrix} u^e \\ v^e \end{Bmatrix} = \begin{Bmatrix} \sum_{i=1}^{n} u_i^e N_i^e \\ \sum_{i=1}^{n} v_i^e N_i^e \end{Bmatrix} = \begin{bmatrix} N_1 & N_2 & \dots & N_n & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & N_1 & N_2 & \dots & N_n \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \\ v_1 \\ v_2 \\ \vdots \\ v_n \end{Bmatrix} \tag{11.28}$$

or

$$\begin{Bmatrix} u^e \\ v^e \end{Bmatrix} = \begin{bmatrix} N_1 & 0 & N_2 & 0 & \dots & N_n & 0 \\ 0 & N_1 & 0 & N_2 & \dots & 0 & N_n \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ \vdots \\ u_n \\ v_n \end{Bmatrix} \equiv \left[N^e\right]\left\{\Delta^e\right\} \tag{11.29}$$

The vector of strain components may be written as

$$\left\{\varepsilon^e\right\} = \begin{Bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{xy} \end{Bmatrix} = \begin{bmatrix} N_{1,x} & 0 & N_{2,x} & 0 & \dots & N_{n,x} & 0 \\ 0 & N_{1,y} & 0 & N_{2,y} & \dots & 0 & N_{n,y} \\ N_{1,y} & N_{1,x} & N_{2,y} & N_{2,x} & \dots & N_{n,y} & N_{2,x} \end{bmatrix} \begin{Bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \vdots \\ \vdots \\ u_n \\ v_n \end{Bmatrix} = \left[B^e\right]\left\{\Delta^e\right\} \tag{11.30}$$

The matrix $[B^e]$ is also called *strain gradient matrix* and can be represented as

$$\left[B^e\right] = \begin{bmatrix} \dfrac{\partial}{\partial x} & 0 \\ 0 & \dfrac{\partial}{\partial y} \\ \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial x} \end{bmatrix} \left\{N^e\right\} = [T^e]\left\{N^e\right\} \tag{11.31}$$

where $[T^e]$ is the matrix of differential operators. Using Eq. (11.8), we obtain

$$\{\sigma^e\} = [D^e]\left\{\varepsilon^e\right\} = [D^e][B^e][\Delta^e] \tag{11.32}$$

Substituting the expressions for displacements, strains, and stresses, given in Eqs (11.29), (11.30), and (11.32), in Eq. (11.26), we get

$$\Pi^e = \frac{1}{2} h^e \int_{\Omega^e} \{\Delta^e\}^T ([B^e]^T [D^e][B^e]\{\Delta^e\}) \mathrm{d}x\mathrm{d}y - h^e \int_{\Omega^e} \{\Delta^e\}^T [N^e]^T \begin{Bmatrix} f_x \\ f_y \end{Bmatrix} \mathrm{d}x\mathrm{d}y$$

$$-h^e \int_{\Gamma^e} \{\Delta^e\}^T [N^e]^T \begin{Bmatrix} t_x^* \\ t_y^* \end{Bmatrix} \mathrm{d}\Gamma = \{\Delta^e\}^T \left(\frac{1}{2} [k^e]^T \{\Delta^e\} - \{f^e\} - \{Q^e\} \right) \tag{11.33}$$

For minimizing the potential energy of the element we differentiate the expression given in Eq. (11.33) with respect to $\{\Delta^e\}$ and get

$$[k^e]\ \{\Delta^e\} = \{f^e\} + \{Q^e\} \tag{11.34}$$

where

$$[k^e] = h^e \int_{\Omega^e} [B^e]^T [D^e][B^e]\, \mathrm{d}x\mathrm{d}y \tag{11.35}$$

$$\{f^e\} = h^e \int_{\Omega^e} [N^e]^T \begin{Bmatrix} f_x \\ f_y \end{Bmatrix} \mathrm{d}x\mathrm{d}y \tag{11.36}$$

$$\{Q^e\} = h^e \int_{\Gamma^e} [N^e]^T \begin{Bmatrix} t_x^* \\ t_y^* \end{Bmatrix} \mathrm{d}\Gamma \tag{11.37}$$

The element stiffness matrix $[k^e]$ is of order $2n \times 2n$ and the elemental load vector

$$[F^e] = \{f^e\} + \{Q^e\} \tag{11.38}$$

is of the order $2n \times 1$, where n is the number of nodes of the element.

11.6 TYPES OF ELEMENTS

From the total potential energy expression for an element, given in Eq. (11.33), it becomes clear that the highest order of derivative in the integrand is 1. Hence, elements with C^0 continuity can be used. The displacement function in an element should be at least linear. A linear displacement function can provide non-zero strain in an element. A constant displacement function provides zero strain and hence is not suitable for solving elasticity problems. Shape functions or interpolation functions N_i are used in the finite element analysis to interpolate the nodal displacements of any element to any point within each element. In natural coordinates, the shape functions for the four-noded quadrilateral elements shown in Fig. 11.3 are written as

$$N_1 = \frac{(1-\xi)(1-\eta)}{4}, \quad N_2 = \frac{(1+\xi)(1-\eta)}{4}, \quad N_3 = \frac{(1+\xi)(1+\eta)}{4}, \quad N_4 = \frac{(1-\xi)(1+\eta)}{4} \tag{11.39}$$

where ξ and η are the natural coordinates and are functions of the physical coordinates x and y, respectively. In the natural coordinate system, the coordinates of the four nodes are (– 1, – 1), (1, 21), (1, 1), and (–1, 1).

One of the earliest finite elements used for solving the elasticity problem is a three-noded triangular element shown in Fig. 11.4. This element was introduced in Chapter 8. In this element, an arbitrarily located point P (shown by a small unfilled circle) divides a triangle 1–2–3 into three sub-areas A_1, A_2, and A_3. Then, the natural coordinates of the point P are defined as ratios of areas, and are written as

$$\xi_1 = \frac{A_1}{A}, \quad \xi_2 = \frac{A_2}{A}, \quad \xi_3 = \frac{A_3}{A} \tag{11.40}$$

where A is the area of triangle 1–2–3. Since $A = A_1 + A_2 + A_3$, the ξ_i are not independent. They satisfy the constraint equation

$$\xi_1 + \xi_2 + \xi_3 = 1 \tag{11.41}$$

For this triangle, shape functions in terms of natural coordinates are given as

$$N_1 = \xi_1, \quad N_2 = \xi_2, \quad N_3 = \xi_3 \tag{11.42}$$

The displacement function in this element is linear in natural as well as physical coordinates. Thus, the displacement field obtained using these shape functions provides constant strain inside the triangular element. This element is therefore called a *constant strain triangle* (CST).

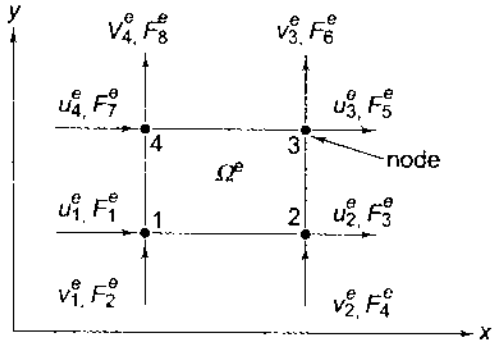


Figure 11.3 *A quadrilateral element*

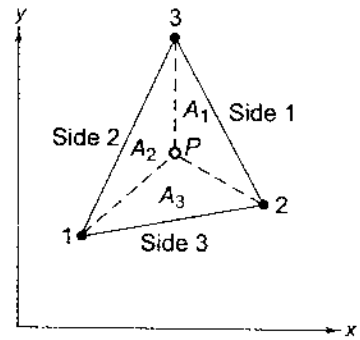


Figure 11.4 *A triangular element*

One can use the higher degree elements in quadrilateral and triangular form. The elements can be transformed in natural coordinates. A six-noded element that could be used for solving the elasticity problems was shown in Fig. 8.3. At each node, there will be two degrees of freedom. Thus, the element has 12 degrees of freedom. Since the displacement field in this element is quadratic, the strain field will be linear. Compared to CST element, this element is expected to provide greater accuracy. Thus, using this element, a coarse mesh may provide the same accuracy as obtained by using a fine mesh in case of a CST element. In quadrilateral elements, eight- or nine-noded elements are commonly used.

11.7 NUMERICAL EVALUATION OF ELEMENT MATRICES AND VECTORS

The evaluation of the element matrices in Eqs (11.35) to (11.37) is done by using numerical integration techniques. For all area and line integrals, Gauss quadrature rule is used. For quadrilateral elements, all physical domain integrations are transformed to the (ξ, η) plane as shown in Fig. 11.5. As a result, the integral can be written as

$$\int_{\Omega} f(x, y)\,\mathrm{d}x\mathrm{d}y = \int_{-1}^{1}\int_{-1}^{1} f(\xi, \eta)|J|\,\mathrm{d}\xi\mathrm{d}\eta \tag{11.43}$$

where $|J|$ is the determinant of the Jacobian matrix of the transformation and is given by

$$|J| = \begin{vmatrix} J_1 & J_2 \\ J_3 & J_4 \end{vmatrix} = \begin{vmatrix} \sum_{i=1}^{n} N_{i,\xi}^{e} x_i & \sum_{i=1}^{n} N_{i,\xi}^{e} y_i \\ \sum_{i=1}^{n} N_{i,\eta}^{e} x_i & \sum_{i=1}^{n} N_{i,\eta}^{e} y_i \end{vmatrix} \tag{11.44}$$

where n is the number of nodes of the element, $N_i^e(\xi, \eta)$ is the shape function corresponding to node i, and x_i and y_i are the physical coordinates of node i.

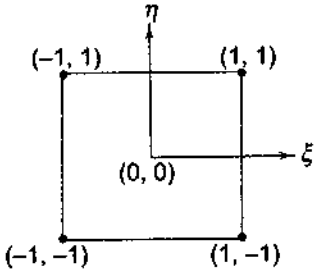


Figure 11.5 *Natural coordinate system*

In writing Eq. (11.44), the iso-parametric formulation has been used, i.e., the interpolation functions used for the geometry variables (x, y) and field variables (u, v) are the same. Also, the spatial derivatives are transformed to the (ξ, η) plane using

$$\begin{Bmatrix} N_{i,x} \\ N_{i,y} \end{Bmatrix} = [J]^{-1} \begin{Bmatrix} N_{i,\xi} \\ N_{i,\eta} \end{Bmatrix} \tag{11.45}$$

where

$$[J]^{-1} = \begin{bmatrix} \dfrac{J_4}{|J|} & \dfrac{-J_2}{|J|} \\ \dfrac{-J_3}{|J|} & \dfrac{J_1}{|J|} \end{bmatrix} \tag{11.46}$$

Finally, the Gauss-quadrature scheme gives

$$\int_{-1}^{1}\int_{-1}^{1} f(\xi, \eta)\,\mathrm{d}\xi\mathrm{d}\eta = \sum_{r=1}^{n1}\sum_{s=1}^{n2} w_r w_s f(\xi_r, \eta_s) \tag{11.47}$$

where n_1 is the number of Gauss points in ξ direction, n_2 is number of Gauss points in η direction, and w_r and w_s are the weights of the corresponding Gauss points. If the displacement field within each

element is assumed to be bilinear, then 2 × 2 Gauss quadrature exactly integrates all terms of the elemental stiffness matrix.

Now, consider the evaluation of a boundary integral of the type

$$Q_i^e = \int_{\Gamma^e} q_n^e N_i(s)\mathrm{d}s \tag{11.48}$$

where q_n^e is a known function (here boundary stress) of the distance s along the boundary Γ^e. It is not necessary to compute such integrals when no portion of Γ^e coincides with the boundary Γ of the total Ω. This is because for any interior boundary the tractions from adjacent elements cancel each other. For evaluating the integral on the external boundary, the two-dimensional line integral in (x, y) plane is transformed to a one-dimensional line integral in the natural coordinate plane by using the fact that along any element side, one of the natural coordinates is constant. Thus

$$\int_0^s f(x, y)\mathrm{d}s = \int_{-1}^{1} f(\varsigma)J_b\mathrm{d}\varsigma \tag{11.49}$$

where the boundary Jacobian J_b is given by

$$J_b = \sqrt{\left(\frac{\mathrm{d}x}{\mathrm{d}\varsigma}\right)^2 + \left(\frac{\mathrm{d}y}{\mathrm{d}\varsigma}\right)^2} \tag{11.50}$$

For integration in a triangular element, we transform the element into an isosceles right-angled triangle, whose two equal sides are of unit length. Such a triangle is shown in Fig. 11.6 in $\xi_1 - \xi_2$ natural coordinate system. Now, the integration over a domain can be carried out using Gauss quadrature. For this purpose, first the integrand and the domain are transformed from the physical to natural coordinates, and then the weighted sum of the integrand is obtained by taking n_G sampling points. Thus

$$\int_\Omega f(x, y)\mathrm{d}x\mathrm{d}y = \int_0^1\int_0^{1-\xi_2} f(\xi_1, \xi_2)|\mathrm{J}|\mathrm{d}\xi_1\mathrm{d}\xi_2 = \sum_{i=1}^{n_G} w_i f\left(\xi_1^i, \xi_2^i\right)|J|^i \tag{11.51}$$

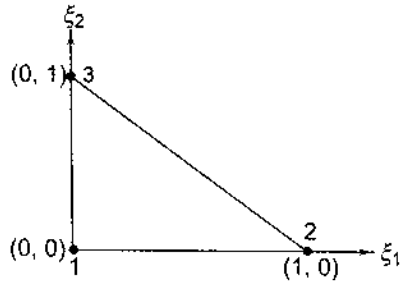


Figure 11.6 *A typical triangular element in natural coordinate system*

where $|J|^i$ indicates the determinant of the Jacobian at the sampling points. The determinant of the Jacobian is given by Eq. (11.44). It can be shown that in the case of CST element with straight sides, the determinant of the Jacobian is twice the area of the triangle in physical coordinates. Line integral is performed as discussed earlier.

11.8 ASSEMBLY OF ELEMENT MATRICES

Once the element matrices are obtained, they are assembled to form a set of linear simultaneous equations, the solution of which yields the displacement field. The assembly is based on the principle of maintaining the continuity of the primary variables (displacements in this case) and the equilibrium of the secondary variables (tractions).

A pseudocode for the assembly in plane elasticity problems using CST elements is given below. The total number of elements in the mesh is denoted by 'number_of_elements' and total number of nodes by 'number_of_nodes'. In an element, there are three nodes; however, there are two degrees of freedom at each node. Thus, corresponding to a node i, there are two degrees of freedom $2i-1$ and $2i$. For an elemental matrix, corresponding to node pair (i, j), the entries will be at locations $(2i-1, 2j-1)$, $(2i-1, 2j)$, $(2i, 2j-1)$, and $(2i, 2j)$. The entries at these locations have to be placed at appropriate location in the global stiffness matrix, which is decided by the connectivity matrix $[C]$.

```
/*Pseudocode for assembly with CST elements*/
for (i=1; i<=2*number_of_nodes; i=i+1)  /*outer loop*/
{
for (j=1; j<= 2*number_of_nodes; j+1)  /*inner loop*/
{
gstiff [i][j] = 0.0;  /*initializing global stiffness matrix*/
}
grhs[i]=0.0;  /*initializing global right hand side vector*/
}
for (i=1; i<= number_of_elements; i=i+1)  /*outer loop*/
{
stiffness(i); /*calling stiffness matrix of element i*/
righthand(i); /*calling right hand side vector of element i*/
p=C[i][j];
for (j=1; j<=3; j=j+1)
{
grhs[2*p-1]=grhs[2*p-1]+erhs[2*j-1)]; /*placing elemental entries
into global location*/
grhs[2*p]=grhs[2*p]+erhs[2*p)]; /*placing elemental entries into
global location*/
for (k=1; k<=4; k=k+1)   /*inner loop */
{
q=C[i][k];
gstiff[2*p-1, 2*q-1]= gstiff[2*p-1, 2*q-1]+estiff[2*j-1, 2*k-1];
/*placing elemental entries into global location*/
gstiff[2*p-1, 2*q]= gstiff[2*p-1, 2*q]+estiff[2*j-1, 2*k];
/*placing elemental entries into global location*/
gstiff[2*p, 2*q-1]= gstiff[2*p, 2*q-1]+estiff[2*j, 2*k-1];
```

```
/*placing elemental entries into global location*/
gstiff[2*p, 2*q]= gstiff[2*p, 2*q]+estiff[2*j, 2*k];
/*placing elemental entries into global location*/
}}}
```

11.9 BOUNDARY CONDITIONS AND SOLUTIONS

As already discussed in the preceding chapters, boundary conditions are of the following two types:

(i) *Essential or geometric boundary conditions*, which are imposed on primary variables such as displacements and

(ii) *Natural or force boundary conditions*, which are imposed on secondary variables such as force and traction.

The force boundary conditions may be imposed during the evaluation of the elemental right-hand side itself, while the prescribed displacement boundary conditions are imposed after the assembly of the element matrices. Then, the global system of linear equations is solved by any numerical technique to get the displacements at global nodes.

11.10 GRADIENT ESTIMATES

In finite element calculations, one often requires accurate estimates of the derivatives of the primary variable. For example, in plane stress or plane strain analysis, the primary unknowns to be computed are the displacement components of the nodes. However, in many cases, the strains and stresses are of prime importance, which are computed from the derivatives of the displacements. Since finite element solutions represent only interpolate solutions, they are expected to be exact at the nodes and approximate elsewhere. Such accuracy is rare, but in general, one observes that the computed values of the primary variables are most accurate at the nodes than at other points of the element. Thus, for the sake of simplicity it may be assumed that the nodal values of an element are exact in a sufficiently refined mesh. Based on this assumption, Barlow (1976, 1989) observed that estimates of the derivatives are the least accurate at the nodes and the most accurate at certain Gauss points for many FEM problems. These points of higher accuracy are called *Barlow points* or *optimal points*. In one-dimensional linear elements, the center of the element is taken as the optimal position for calculating the first derivative.

For four-noded C^0 continuity quadrilateral elements, the Barlow point corresponds to the point of one-Gauss-point quadrature. For eight- and nine-noded C^0 continuity quadrilateral elements, Barlow points correspond to sampling points of two-Gauss-point quadrature. There is a considerable lack of consensus regarding the optimal points of the triangular elements [Rajendran and Liew (2003)]. According to Strange and Fix (1973), the midpoints of the edges may be the optimal points for linear and quadratic triangular elements. Lazarov and Andreev (1988) observed that the errors in some but not all the components of strain are minimum at the Gauss points of the edges. Rajendran and Liew (2003) observed that for CST element, the optimal point is at the centroid of the element. This point was called the *best-fit point*. Barlow points (corresponding to the most accurate, direct strain

component) for this element fall on the edges. The authors observed that the best-fit point is frame invariant with respect to a rotation of the element, which is not the case with Barlow point. The best-fit point corresponds to the sampling point of one-Gauss-point quadrature. Also, in a CST element, for stress and strain, the same best-fit point is obtained. In this element, Barlow point for stress does not exist. For a six-noded triangular element, Barlow points for direct strain exist for the orientations in which one natural axis makes an angle of 0° or 90° with the physical axis. In natural coordinates, these are (0.211, 0, 0.789) and (0.789, 0, 0.211) for ε_x in 0° orientation of the element. These points coincide with the Barlow points for stress when the shear strain does not exist. If shear strain is present, Barlow points for stress do not exist. The approximate best-fit points for stress as well strain are (1/6, 1/6, 2/3), (2/3, 1/6, 1/6), and (1/6, 2/3, 1/6).

The above discussion brings out the fact that the gradients should not be calculated at the nodes. They can be calculated at Barlow points or the best-fit points mentioned above. However, these points are only the expected points of high accuracy and need not be the exact points of high accuracy. As the refinement level of the mesh increases, the possibility of these points being the real high-accuracy points increases. Hence, for accurate estimate of gradients, one should use a refined element and in addition estimate the values at Barlow points or best-fit points, but certainly not at the nodes.

11.11 STRESS ANALYSIS USING AN FEM PACKAGE: AN EXAMPLE

As an example of finding out the stress in the whole domain, consider the problem of a plate with the hole loaded by uniformly distributed tensile load (Fig. 11.7). This problem was solved on ANSYS, commercial FEM software. Because of the symmetry of the problem, only a quarter of the plate needs to be analyzed. On the line of symmetries, tangential traction and normal displacement components will be zero. Figure11.8 shows the finite element mesh and boundary conditions. Here, quadrilateral elements have been used. Figure 11.9 shows the contours of longitudinal stresses; the applied tensile stress was 100 MPa. Note that in the domain far away from the hole, stresses are close to this value. There is a region of high stress concentration near the highest vertical point of the circle, the stresses there being around 336 MPa. One can obtain the contours of other stress components and equivalent stresses also.

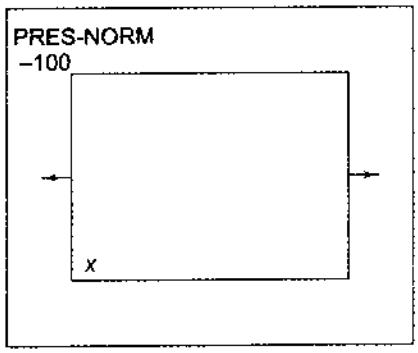


Figure 11.7 *A plate with a hole loaded by uniformly distributed tensile load on both sid es*

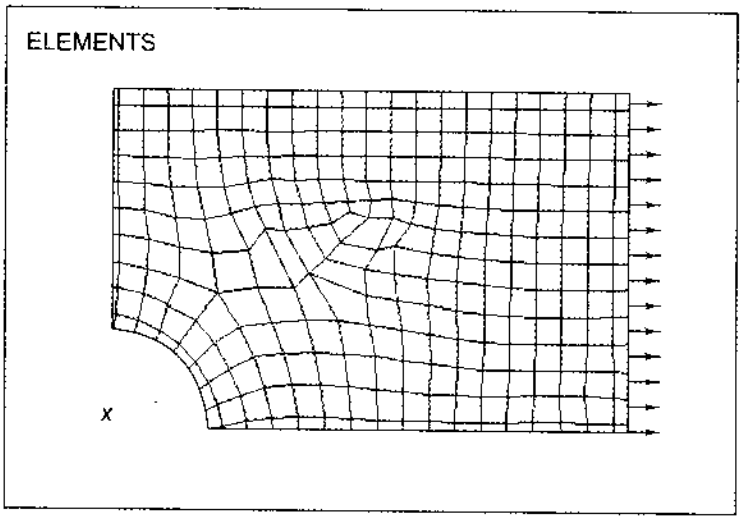


Figure 11.8 *Finite element mesh for solving quarter plate problem*‡

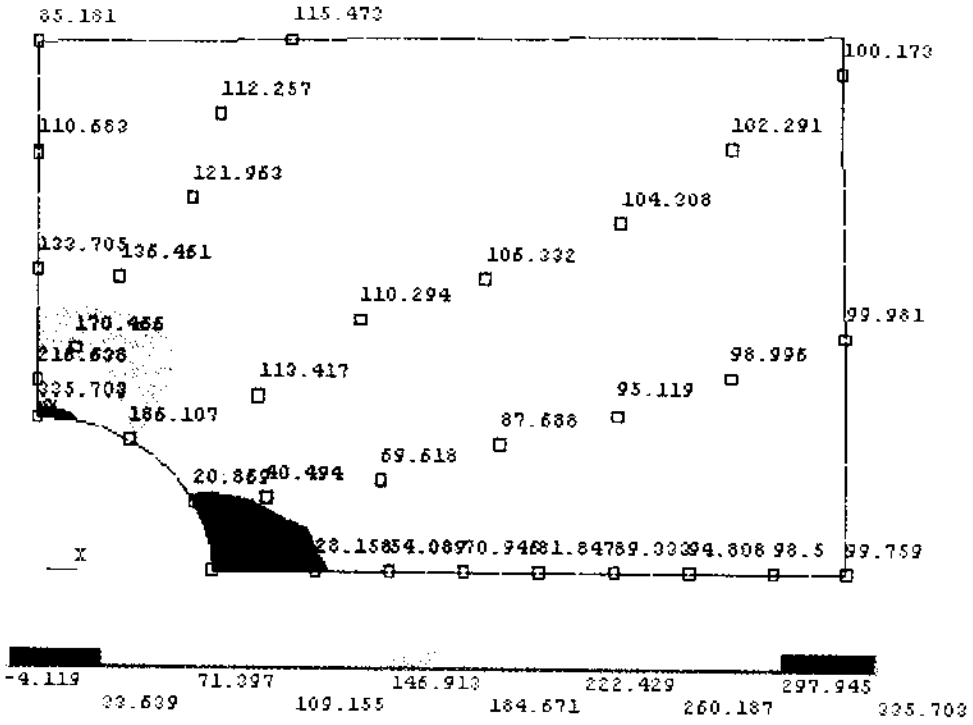


Figure 11.9 *Contours of longitudinal stresses*‡

‡ The symbol indicates that the coloured image is provided at the centre of the book.

11.12 SUMMARY

In this chapter, the finite element formulation of plane stress and plane strain problems is discussed. The only difference between plane strain and plane stress formulations is that both have different constitutive matrices. After obtaining the finite element solution in the form of nodal displacements, stresses and strains can be found by post-processing the results. As an example, a problem was solved in ANSYS and has been presented here to demonstrate the effectiveness of FEM for finding the stresses in the whole domain.

SOLVED PROBLEMS

11.1 A concentrated in-plane load is applied normal to the free side surface of a plate fixed at one edge. The plate can be considered in plane stress. Prove that the FEM predicted displacement in the direction of force at the point will be less or equal to the actual displacement.

Solution Using Eq. (11.33), the potential energy of a finite element is given by

$$\Pi^e = \frac{1}{2}\lfloor \Delta^e \rfloor [k^e]\{\Delta^e\} - \lfloor \Delta \rfloor \{f^e\} - \lfloor \Delta^e \rfloor \{Q^e\} \tag{11.1.1}$$

where the symbol $\lfloor \Delta^e \rfloor$ has been used, instead of $\{\Delta^e\}^T$, indicate row vector.

As there is no body force, the potential energy of the element becomes

$$\Pi^e = \frac{1}{2}\lfloor \Delta^e \rfloor [k^e]\{\Delta^e\} - \lfloor \Delta^e \rfloor \{Q^e\} \tag{11.1.2}$$

The total potential energy of the plate as modelled by FEM with a mesh consisting of N elements is given by

$$\Pi_{\text{FEM}} = \sum_{e=1}^{N} \Pi^e = \frac{1}{2}\sum_{e=1}^{N} \lfloor \Delta^e \rfloor [k^e]\{\Delta^e\} - \sum_{e=1}^{N} \lfloor \Delta^e \rfloor \{Q^e\} \tag{11.1.3}$$

We can express each elemental matrix in the global form as we did in Chapter 1. In the global form of the elemental matrix all the entries except those corresponding to $[k^e]$ will be zero. Similarly, the elemental force vector can also be written in the global form. Denoting the global form of stiffness matrices by $[k^e]_g$, the global form of the elemental load vector by $[Q^e]_g$, and the vector containing all degrees of freedom of the mesh by $\{\Delta\}$, the potential energy of the mesh can be written as

$$\Pi_{\text{FEM}} = \sum_{e=1}^{N} \Pi^e = \frac{1}{2}\sum_{e=1}^{N} \left(\lfloor \Delta \rfloor [k^e]_g \{\Delta\}\right) - \sum_{e=1}^{N} \left(\lfloor \Delta \rfloor \{Q^e\}_g\right) \tag{11.1.4}$$

The expression given in Eq. (11.1.4) can also be written as

$$\Pi_{\text{FEM}} = \sum_{e=1}^{N} \Pi^e = \frac{1}{2}\lfloor \Delta \rfloor \sum_{e=1}^{N} [k^e]_g \{\Delta\} - \lfloor \Delta \rfloor \sum_{e=1}^{N} \{Q^e\}_g \tag{11.1.5}$$

which gives

$$\Pi_{FEM} = \frac{1}{2}\lfloor \Delta \rfloor [K]\{\Delta\} - \lfloor \Delta \rfloor \{Q\} \tag{11.1.6}$$

where $[K]$ is the global stiffness matrix and $\{Q\}$is the global load vector. In the present case, the internal loads will balance each other and the global load vector will consist of one externally applied load. Minimizing Eq. (11.1.6) with respect to $\{\Delta\}$, we obtain

$$[K]\{\Delta\} = \{Q\} \tag{11.1.7}$$

Substituting Eq. (11.1.7) in Eq. (11.1.6), we get

$$\Pi_{FEM} = \frac{1}{2}\lfloor \Delta \rfloor \{Q\} - \lfloor \Delta \rfloor \{Q\} = -\frac{1}{2}\lfloor \Delta \rfloor \{Q\} \tag{11.1.8}$$

We also know that actual potential energy at equilibrium of the structure is given by

$$\Pi_{act} = \text{Strain energy} + \text{Potential due to the external loads} \tag{11.1.9}$$

In the present case, let the external force be F_1 and the corresponding displacement as u_1. (For convenience you may assume that the x-axis is aligned along F_1 and node 1 of the mesh is at the point of application of F_1.) In that case, the actual potential energy at the equilibrium is

$$\Pi_{act} = \text{Strain energy} - F_1 u_1 \tag{11.1.10}$$

However, the stored strain energy is equal to the work done by the external forces. In the present case

$$\text{Strain energy} = \frac{1}{2} F_1 u_1 \tag{11.1.11}$$

Thus

$$\Pi_{act} = \frac{1}{2} F_1 u_1 - F_1 u_1 = -\frac{1}{2} F_1 u_1 \tag{11.1.12}$$

From the principle of minimum potential energy we know that at equilibrium, the actual potential energy is the minimum. Hence, the potential energy of the FEM mesh will be greater or equal to the actual potential energy, that is

$$-\frac{1}{2}\lfloor \Delta \rfloor \{Q\} \geq -\frac{1}{2} F_1 u_1 \tag{11.1.13}$$

As discussed earlier, the external load vector $\{Q\}$ consists of just one load and is assumed to act at node 1 in x-direction, then

$$Q = \begin{Bmatrix} F_1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{Bmatrix} \tag{11.1.14}$$

The row vector of nodal displacements may be written as

$$\lfloor \Delta \rfloor = \lfloor \hat{u}_1 \ \hat{u}_2 \ \hat{u}_3 \ldots \rfloor \tag{11.1.15}$$

where the hat symbol is used to indicate FEM nodal displacement. Putting Eqs (11.1.14) and (11.1.15) in Eq. (11.1.13), we obtain

$$-\frac{1}{2}F_1\hat{u}_1 \geq -\frac{1}{2}F_1u_1 \tag{11.1.16}$$

which implies

$$\hat{u}_1 \leq u_1 \tag{11.1.17}$$

Thus, it has been proved that FEM predicted displacement is less than or equal to the actual displacement.

11.2 The body force per unit volume in x-direction in a triangular CST element of 1-mm thickness is 30 N/mm^3. The area of the element is 30 mm^2. Find out the load vector due to the body force.

Solution The load vector is calculated using Eq. (11.36) as

$$\{f^e\} = h^e \int_{\Omega^e} [N^e]^{\mathrm{T}} \begin{Bmatrix} f_x \\ f_y \end{Bmatrix} \mathrm{d}x\mathrm{d}y \tag{11.2.1}$$

For our problem, $h^e = 1$ mm, $f_x = 30$ N/mm^3, and $f_y = 0$. For the CST element, $[N^e]$ is given by

$$\left[N^e\right] = \begin{bmatrix} N_1 & 0 & N_2 & 0 & N_3 & 0 \\ 0 & N_1 & 0 & N_2 & 0 & N_3 \end{bmatrix} = \begin{bmatrix} \xi_1 & 0 & \xi_2 & 0 & \xi_3 & 0 \\ 0 & \xi_1 & 0 & \xi_2 & 0 & \xi_3 \end{bmatrix} \tag{11.2.2}$$

where the shape functions have been expressed in natural coordinates. As mentioned in Eq. (8.7)

$$\xi_1 + \xi_2 + \xi_3 = 1 \tag{11.2.3}$$

Substituting the values in Eq. (11.2.1) and transforming the integrand and domain into natural coordinates, we get

$$\{f^e\} = \int_{\Omega_n^e} \begin{bmatrix} \xi_1 & 0 \\ 0 & \xi_1 \\ \xi_2 & 0 \\ 0 & \xi_2 \\ \xi_3 & 0 \\ 0 & \xi_3 \end{bmatrix} \begin{Bmatrix} 30 \\ 0 \end{Bmatrix} |J| \mathrm{d}\xi_1 \mathrm{d}\xi_2 = \int_{\Omega_n^e} \begin{bmatrix} 30\xi_1 \\ 0 \\ 30\xi_2 \\ 0 \\ 30\xi_3 \\ 0 \end{bmatrix} |J| \mathrm{d}\xi_1 \mathrm{d}\xi_2 \tag{11.2.4}$$

where symbol Ω_n^e is used to indicate the area in natural coordinates,

For CST element, determinant of Jacobian matrix is equal to twice the area in physical coordinates. In the present case it is 60 mm^2. Thus

$$\left\{f^e\right\} = \int_{\Omega_n^e} 60 \begin{bmatrix} 30\xi_1 \\ 0 \\ 30\xi_2 \\ 0 \\ 30\xi_3 \\ 0 \end{bmatrix} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \tag{11.2.5}$$

The integral given in Eq. (11.2.5) can be evaluated by Gauss quadrature. From Table 9.2, we see that one-Gauss-point formula is appropriate. The Gauss point is (1/3, 1/3) and the corresponding weight is 1/2. This means $\xi_1 = \xi_2 = \xi_3 = 1/3$. Thus, the integration in Eq. (11.2.5) yields

$$\{f^e\} = \frac{1}{2} \times 60 \begin{bmatrix} 30 \times \frac{1}{3} \\ 0 \\ 30 \times \frac{1}{3} \\ 0 \\ 30 \times \frac{1}{3} \\ 0 \end{bmatrix} = \begin{bmatrix} 300 \\ 0 \\ 300 \\ 0 \\ 300 \\ 0 \end{bmatrix} \text{N} \tag{11.2.6}$$

Thus, the load vector is obtained. We observe that the total horizontal load in the element has been equally divided between the nodes.

EXERCISES

A. Multiple Choice Questions

11.1 The body force in x-direction in a four-noded Lagrangian element of area 10 mm^2 and thickness 2 mm is 40 N/mm^3. The contribution of the body force at each node will be

(a) 100 N (b) 200 N (c) 400 N (d) 800 N

11.2 The body force in x-direction in a nine-noded Lagrangian element of area 10 mm^2 and thickness 2 mm is 40 N/mm^3. The contribution of this load to the center node of the element is

(a) 800/9 N (b) 1600/9 N (c) 2400/9 N (d) 12800/9 N

11.3 The x-displacement at four nodes of a square element is u, $-u$, u, and $-u$ and the y-displacement is zero. For this element,

(a) All the components of the strain are zero (b) $\varepsilon_{xx} = \varepsilon_{xy} = 0$

(c) $\varepsilon_{xx} = 0, \varepsilon_{xy} \neq 0$ (d) $\varepsilon_{yy} = 0, \varepsilon_{xy} \neq 0$

11.4 In plane stress FEM formulation, the vector containing the components of strain is related to the displacement vector by the relation $\{\varepsilon\} = [B]\{u\}$. The matrix $[B]$ is constant for

(a) a CST element (b) any triangular element

(c) any rectangular element (d) a four-noded quadrilateral element

11.5 At one edge of a CST element the traction is 50 N/mm^2. The traction vector is along x-axis. The thickness of the element is 1 mm and the length of the edge is 5 mm. Contribution of this traction to one node on the edge will be

(a) 25 N (b) 50 N (c) 125 N (d) 250 N

A

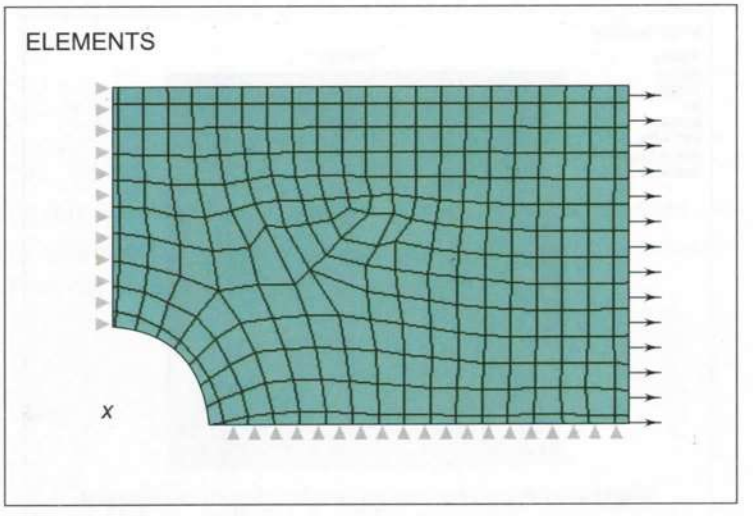


Figure 11.8 *Finite element mesh for solving quarter plate problem*

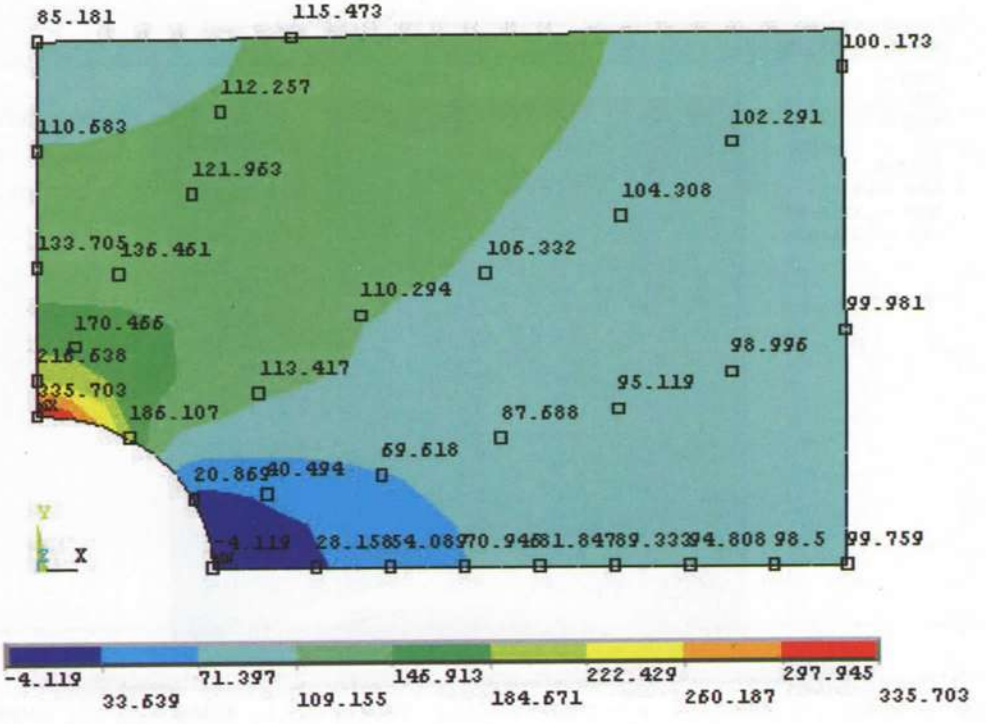


Figure 11.9 *Contours of longitudinal stresses*

12.1 INTRODUCTION

In the preceding chapters, we have discussed FEM formulations for some two-dimensional problems. In the present chapter, we will extend our discussions to time-dependent problems. In solid mechanics, two types of time-dependent problems are solved; in which their governing differential equation(s) contain time derivatives. In one of the problems, dynamic response, i.e., the change in the displacement of the particle with time, is studied, while in the other type, we need to find out the natural frequencies of vibrations. In this chapter, we shall find out the natural frequencies of vibration for one-dimensional problems. We have chosen one-dimensional problems to make the treatment simple. On the completion of this chapter, the reader will be able to carry out finite element formulations for two-dimensional and three-dimensional problems as well.

12.2 VIBRATIONS OF A ROD

The governing differential equation for the undamped free vibration of a rod is written as

$$\frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right)-\rho A\frac{\partial^2 u}{\partial t^2}=0 \tag{12.1}$$

where u is the axial displacement function, E is Young's modulus of elasticity of the rod, A is the area of cross-section of the rod, and ρ is the density of the rod. If the parameters E, A, and ρ are constants, Eq. (12.1) can be solved analytically. However, if E, A, and ρ vary along the rod, the numerical methods may be needed for solving it. Let us carry out FEM formulation for Eq. (12.1).

Consider the approximating function in an element of length h to be u^e. In that case, the residual is given by

$$R^e=\frac{\partial}{\partial x}\left(EA\frac{\partial u^e}{\partial x}\right)-\rho A\frac{\partial^2 u^e}{\partial t^2} \tag{12.2}$$

Making the weighted integral of the residual equal to 0, we obtain

$$\int_0^h wR^e\,\mathrm{d}x=0 \tag{12.3}$$

where we assume that x is a local coordinate and the differential equation [Eq. (12.1)] has been expressed in local coordinates. Thus, we can write Eq. (12.3) as

$$\int_0^h w\left(\frac{\partial}{\partial x}\left(EA\frac{\partial u^e}{\partial x}\right)-\rho A\frac{\partial^2 u^e}{\partial t^2}\right)\mathrm{d}x=0 \tag{12.4}$$

Integrating this equation by parts to reduce the order of the derivative with respect to x, we obtain

$$wEA\frac{\partial u^e}{\partial x}\bigg|_0^h-\int_0^h\frac{\partial w}{\partial x}EA\frac{\partial u^e}{\partial x}\mathrm{d}x-\int_0^h\rho Aw\frac{\partial^2 u^e}{\partial t^2}\mathrm{d}x=0 \tag{12.5}$$

The approximating function u^e is a function of x and t. It can be expressed in terms of the nodal displacements as in the case of static problem, where the nodal displacements will be functions

of t. Taking into account the factors of completeness and compatibility, a two-noded C^0 continuity element is good enough to serve our purpose. Thus

$$u^e = \lfloor N_1 \quad N_2 \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix}, \quad \ddot{u}^e = \lfloor N_1 \quad N_2 \rfloor \begin{Bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{Bmatrix} = \lfloor N \rfloor \left\{ \ddot{u}^{ne} \right\} \tag{12.6}$$

where double dots indicate second derivative with respect to time. We already know the expressions for the shape functions in local coordinate system can be written as

$$N_1 = 1 - \frac{x}{h}; \quad N_2 = \frac{x}{h} \tag{12.7}$$

In Galerkin FEM, the weight functions are approximated in the same way as the primary variable. Thus

$$w = \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \tag{12.8}$$

Substituting the expressions u^e and w in the weak form, we obtain

$$\begin{aligned} & w_2 EA \frac{\partial u^e}{\partial x}\bigg|_h - w_1 EA \frac{\partial u^e}{\partial x}\bigg|_0 - \int_0^h \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} EA \lfloor N_{1,x} \quad N_{2,x} \rfloor \mathrm{d}x \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \\ & - \int_0^h \lfloor w_1 \quad w_2 \rfloor \rho A \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \quad N_2 \rfloor \mathrm{d}x \begin{Bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{Bmatrix} = 0 \end{aligned} \tag{12.9}$$

where '$,x$' in the subscript indicates differentiation with respect to x. Noting that node 1 corresponds to $x = 0$ and node 2 corresponds to $x = h$ and taking the factor $\lfloor w_1 \quad w_2 \rfloor$ common, Eq. (12.9) can be written as

$$\lfloor w_1 \quad w_2 \rfloor \left[\begin{Bmatrix} -EA \dfrac{\partial u}{\partial x}\bigg|_2 \\ EA \dfrac{\partial u}{\partial x}\bigg|_1 \end{Bmatrix} - \int_0^h \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} EA \lfloor N_{1,x} \quad N_{2,x} \rfloor \mathrm{d}x \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} - \int_0^h \rho A \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \quad N_2 \rfloor \mathrm{d}x \begin{Bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{Bmatrix} \right] = 0 \tag{12.10}$$

As the nodal weights are arbitrary, we get the following elemental equations:

$$\int_0^h \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} EA \lfloor N_{1,x} \quad N_{2,x} \rfloor \mathrm{d}x \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} + \int_0^h \rho A \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \quad N_2 \rfloor \mathrm{d}x \begin{Bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{Bmatrix} = \begin{Bmatrix} -EA \dfrac{\partial u}{\partial x}\bigg|_2 \\ EA \dfrac{\partial u}{\partial x}\bigg|_1 \end{Bmatrix} \tag{12.11}$$

If ρ, E, and A are constant inside an element, the system of equations given in Eq. (12.11) may be written as

$$\frac{\rho A h}{6} \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \begin{Bmatrix} \ddot{u}_1 \\ \ddot{u}_2 \end{Bmatrix} + \frac{EA}{h} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} -EA \dfrac{\partial u}{\partial x}\bigg|_1 \\ EA \dfrac{\partial u}{\partial x}\bigg|_2 \end{Bmatrix} \tag{12.12}$$

We observe that the elemental equations are not algebraic equations, but ordinary differential equations in time. Writing Eq. (12.12) in notational form, we obtain

$$[m^e]\{\ddot{u}^{ne}\}+[k^e]\{u^{ne}\}=\{F_i^e\} \tag{12.13}$$

where $[m^e]$ is the elemental mass matrix, $[k^e]$ is the stiffness matrix, and the right-hand side of the equation represents the internal load vector.

Let us study the elemental mass matrix. If we sum all the elements of the mass matrix, we get ρAh, which is the total mass of the element. Thus, as a result of FEM formulation, the total mass of the elements has been distributed in the elements of the stiffness matrix. The mass matrix obtained in this way is called *consistent mass matrix* because its derivation is consistent with the derivation of stiffness matrix. A simpler way would have been to distribute the total mass at the two nodes. Then the mass matrix would be written as

$$\left[m^e\right]=\frac{\rho Ah}{2}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \tag{12.14}$$

This type of mass matrix is called a *lumped mass matrix* and has the advantage of being strongly diagonally dominant.

Assembling the equations we get the following system of equations:

$$\left[M_g\right]\left\{\ddot{U}_g\right\}+[K_g]\left\{U_g\right\}=\begin{Bmatrix}-EA\left.\dfrac{\partial u}{\partial x}\right|_1\\ 0\\ 0\\ \vdots\\ 0\\ EA\left.\dfrac{\partial u}{\partial x}\right|_N\end{Bmatrix} \tag{12.15}$$

where $[M_g]$ is the global mass matrix, $[K_g]$ is the global stiffness matrix, and $\{U_g\}$ is the global displacement vector. The nodes 1 and N are the boundary nodes. It is clear that during assembly, the internal loads cancel each other everywhere except at the boundaries. Equation (12.5) can be solved only after the boundary conditions are applied. In free-vibration problems, no forces are applied at the boundary. Therefore

$$EA\left.\frac{\partial u}{\partial x}\right|_1=EA\left.\frac{\partial u}{\partial x}\right|_N=0 \tag{12.16}$$

and Eq. (12.15) becomes

$$\left[M_g\right]\left\{\ddot{U}_g\right\}+[K_g]\left\{U_g\right\}=\{0\} \tag{12.17}$$

where $\{0\}$ indicates a vector, in which all entries are zero and the size of the vector is the same as that of $\{U_g\}$. Now, three different types of situations can arise. First, both ends are free; second, one end is fixed, while the other end is free; and third, both ends are fixed. In the first situation, there are no

essential boundary conditions. In the second situation, one essential boundary condition is present. The row and column corresponding to that boundary condition can be eliminated from the matrices of Eq. (12.17) and the corresponding row can be deleted from the vectors. For example, if the first node is fixed, the first row and column will be eliminated from $[M_g]$ and $[K_g]$ and the first row will be eliminated from $\{U_g\}$, $\{\ddot{U}_g\}$, and $\{0\}$. In the third situation, rows and columns corresponding to two boundary conditions will be eliminated from $[M_g]$ and $[K_g]$ and corresponding rows will be eliminated from $\{U_g\}$, $\{\ddot{U}_g\}$, and $\{0\}$. After the application of essential boundary conditions, i.e., by eliminating the required rows and columns, the system of equations given by Eq. (12.7) becomes

$$[M]\{\ddot{U}\}+[K]\{U\}=\{0\} \tag{12.18}$$

where $\{0\}$ is the zero vector and has the same size as $\{U\}$. In Eq. (12.18), rows and columns corresponding to the fixed boundary conditions have been eliminated.

To solve Eq. (12.18), we assume

$$\{U\}=\begin{Bmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \end{Bmatrix} e^{\lambda t} = \{X\}e^{\lambda t} \tag{12.19}$$

Then, the system of equations becomes

$$\lambda^2[M]\{X\}+[K]\{X\}=0 \tag{12.20}$$

To solve the system of equations given in Eq. (12.20), we can substitute $\lambda = i\omega$ and obtain the equation

$$-\omega^2[M]\{X\}+[K]\{X\}=0 \tag{12.21}$$

Note that substituting $\lambda = -i\omega$ will also lead to the same equation. Equation (12.21) is an eigenvalue problem, ω^2 corresponding to eigenvalue and $\{X\}$ corresponding to eigenvectors. If the order of $[M]$ and $[K]$ is $n \times n$, there will be a total of n eigenvalues; some of these may be repeated eigenvalues, in general. Displacement at the ith node can be obtained by a linear combination of the solutions corresponding to $\lambda = i\omega$ and $\lambda = -i\omega$. Thus, we can write the displacement as

$$U_i = X_i e^{i\omega t} + Y_i e^{-i\omega t} = C_i \cos\omega t + D_i \sin\omega t \tag{12.22}$$

which is a periodic function of time with angular frequency ω. The constants C_i and D_i can be obtained from the initial conditions. For example, if the displacement U_i is zero at time $t = 0$, then C_i will be equal to zero and the displacement will be given by

$$U_i = D_i \sin\omega t \tag{12.23}$$

The constant D_i can be obtained if the velocity of the node i is known at $t = 0$. It is seen that all the nodes vibrate with the same frequency. In fact, in Eq. (12.19), we assumed the solution in which the nodal displacement of the ith node is written as $X_i e^{\lambda t}$. It can be verified that if we assume the solution of the form $X_i e^{\lambda_i t}$, in which λ_i are not equal, the solution will be trivial with all nodal displacements being zero.

In general, the frequencies predicted by FEM are expected to be higher than the actual frequencies because FEM makes the structure stiffer by constraining the motion. The static deflections predicted

by FEM are expected to be lower than the actual deflection. This was already discussed in the previous chapter. Refining the mesh improves the accuracy and also provides the frequencies of higher modes of vibrations. Higher the mode, lesser is the accuracy in the estimation of natural frequency.

Example 12.1 Obtain the expression for the first non-zero natural frequency of vibration for a uniform free–free (both ends free) rod by FEM with two and three elements.

Solution From Eqs (12.12) and (12.13), it is seen that

$$[m^e] = \frac{\rho A h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad [k^e] = \frac{EA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{12.24}$$

We divide the rod of length l in two elements. For both the elements, E, A, and ρ are the same and $h = l/2$. The assembled mass and stiffness matrices are given by

$$[M_g] = \frac{\rho A h}{6}\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2+2 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad [K_g] = \frac{EA}{h}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1+1 & -1 \\ 0 & -1 & 1 \end{bmatrix} \tag{12.25}$$

Since the ends of the rod are free, we can use Eq. (12.16) and write

$$EA\left.\frac{\partial u}{\partial x}\right|_1 = EA\left.\frac{\partial u}{\partial x}\right|_3 = 0 \tag{12.26}$$

Thus, right-hand side vector in Eq. (12.15) will be a zero vector and no rows and columns will be eliminated, i.e., after applying the boundary conditions, the assembled mass and stiffness matrices remain unchanged as shown

$$[M] = [M_g], \quad [K] = [K_g] \tag{12.27}$$

Using Eq. (12.21), we can write

$$-\omega^2 \frac{\rho A h}{6}\begin{bmatrix} 2 & 1 & 6 \\ 1 & 4 & 1 \\ 0 & 1 & 2 \end{bmatrix}\{X\} + \frac{EA}{h}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}\{X\} = 0 \tag{12.28}$$

This is a generalized eigenvalue problem and can be solved by standard methods coded in the packages such as MATLAB®. Here, we solve it manually. For convenience, we can write

$$\alpha = \omega^2 \frac{\rho h^2}{6E} \tag{12.29}$$

Then, Eq. (12.29) can be written as

$$\begin{bmatrix} 1-2\alpha & -1-\alpha & 0 \\ -1-\alpha & 2-4\alpha & -1-\alpha \\ 0 & -1-\alpha & 1-2\alpha \end{bmatrix}\{X\} = 0 \tag{12.30}$$

For non-trivial solution, i.e., for the solution in which $\{X\}$ is not a zero vector, the determinant of the coefficient matrix is zero. Thus

$$\begin{bmatrix} 1-2\alpha & -1-\alpha & 0 \\ -1-\alpha & 2-4\alpha & -1-\alpha \\ 0 & -1-\alpha & 1-2\alpha \end{bmatrix} = 0 \tag{12.31}$$

Expanding the determinant, we obtain

$$12\alpha^3 - 30\alpha^2 + 12\alpha = 0 \tag{12.32}$$

The solution of Eq. (12.32) is $\alpha = 0, 0.5, 2$. Corresponding to $\alpha = 0$, the natural frequency is $\omega = 0$. The corresponding eigenvector can be found by solving Eq. (12.30) after substituting $\alpha = 0$ and writing $\{X\}^T$ as $\{X_1, X_2, X_3\}^T$. Thus

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} X_1 \\ X_2 \\ X_3 \end{Bmatrix} = 0 \tag{12.33}$$

The three equations given in Eq. (12.33) can be written as

$$X_1 - X_2 = 0 \tag{12.34a}$$

$$-X_1 + 2X_2 - X_3 = 0 \tag{12.34b}$$

$$-X_2 + X_3 = 0 \tag{12.34c}$$

From Eqs (12.34a) and (12.34c), we obtain

$$X_1 = X_2 = X_3 \tag{12.35}$$

Since Eq. (12.35) satisfies Eq. (12.34b) no unique solution is obtained. One solution can be obtained by putting $X_1 = 1$, which gives the eigenvector as $\{1, 1, 1\}^T$. The normalized eigenvector is $\{1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}\}^T$, whose magnitude is 1. It is seen that this eigenvector provides the same displacement at all nodes. Thus, the mode corresponding to zero frequency is a rigid-body mode.

Corresponding to $\alpha = 0.5$, frequency ω is obtained using Eq. (12.29) as

$$\omega = \frac{\sqrt{3}}{h}\sqrt{\frac{E}{\rho}} = \frac{2\sqrt{3}}{l}\sqrt{\frac{E}{\rho}} = \frac{3.4641}{l}\sqrt{\frac{E}{\rho}} \tag{12.36}$$

This is the first non-zero natural frequency. The mode shape corresponding to this frequency can be obtained by finding out the eigenvector as in the case of zero frequency.

If we discretize the rod into three elements of equal length with $h = l/3$, the system of equations to be solved becomes

$$-\omega^2 \frac{\rho A h}{6} \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 0 \\ 0 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} \{X\} + \frac{EA}{h} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \{X\} = 0 \tag{12.37}$$

Using Eq. (12.29), the system of equations given in Eq. (12.37) can be written as

$$-\alpha\begin{bmatrix}2 & 1 & 0 & 0\\ 1 & 4 & 1 & 0\\ 0 & 1 & 4 & 1\\ 0 & 0 & 1 & 2\end{bmatrix}\{X\}+\begin{bmatrix}1 & -1 & 0 & 0\\ -1 & 2 & -1 & 0\\ 0 & -1 & 2 & -1\\ 0 & 0 & -1 & 1\end{bmatrix}\{X\}=0 \tag{12.38}$$

Solving this generalized eigenvalue problem with the help of EIG function of MATLAB, we get

$$\alpha = 0,\ 0.2,\ 1,\ 2 \tag{12.39}$$

Using Eq. (12.29), the first non-zero frequency is obtained as

$$\omega = \frac{3.2863}{l}\sqrt{\frac{E}{\rho}} \tag{12.40}$$

The exact solution provides the first natural frequency as [Bathe (1996)]

$$\omega = \frac{\pi}{l}\sqrt{\frac{E}{\rho}} \approx \frac{3.1416}{l}\sqrt{\frac{E}{\rho}} \tag{12.41}$$

We observe that as a result of mesh refinement, the first natural frequency moves towards exact solution.

12.3 VIBRATIONS OF A BEAM

In the absence of damping, the governing differential equation of the motion of the free vibration of a beam is given as

$$\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 v}{\partial x^2}\right)+\rho A\frac{\partial^2 v}{\partial t^2}=0 \tag{12.42}$$

where ρA is the mass per unit length of the beam and it may be a function of x. Here, the vertical deflection of the beam is a function of both x and t. In Eq. (12.42), shear deformation and rotary inertia have been neglected. Assuming that the vertical deflection in an element is similar to that in the case of a static beam problem, we can write

$$v^e = \lfloor N_1 \quad N_2 \quad N_3 \quad N_4 \rfloor \begin{Bmatrix} v_1(t) \\ v_{1,x}(t) \\ v_2(t) \\ v_{2,x}(t) \end{Bmatrix} \tag{12.43}$$

Note that Eq. (12.43) is same as Eq. (6.11) except that nodal deflections/slopes are now functions of time. For the assumed deflection v^e, the residual is obtained as

$$R^e = \frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 v^e}{\partial x^2}\right)+\rho A\frac{\partial^2 v^e}{\partial t^2} \tag{12.44}$$

We make the weighted residual inside the element equal to zero; the equation then becomes

$$\int_0^h wR^e \mathrm{d}x = \int_0^h w\left(\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 v^e}{\partial x^2}\right) + \rho A\frac{\partial^2 v^e}{\partial t^2}\right)\mathrm{d}x = 0 \tag{12.45}$$

The first part of the expression is integrated by parts twice, so that the order of differentiation of v^e becomes two. The second part is left as it is. Equation (12.45) then becomes

$$w\frac{\partial}{\partial x}\left(EI\frac{\partial^2 v^e}{\partial x^2}\right)\Bigg|_0^h - \frac{\partial w}{\partial x}\left(EI\frac{\partial^2 v^e}{\partial x^2}\right)\Bigg|_0^h + \int_0^h \frac{\partial^2 w}{\partial x^2}\left(EI\frac{\partial^2 v^e}{\partial x^2}\right)\mathrm{d}x + \int_0^h \rho A\frac{\partial^2 v^e}{\partial t^2}\mathrm{d}x = 0 \tag{12.46}$$

In Galerkin FEM, the basis functions for approximating the deflections and weight functions are same. Hence

$$w = \begin{Bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{Bmatrix}^{\mathrm{T}} \lfloor N_1 \quad N_2 \quad N_3 \quad N_4 \rfloor^{\mathrm{T}} = \lfloor w_1 \quad w_2 \quad w_3 \quad w_4 \rfloor \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \tag{12.47}$$

Putting the approximation given in Eq. (12.47) in Eq. (12.46), we obtain

$$\lfloor w_1 \quad w_2 \quad w_3 \quad w_4 \rfloor \left[\int_0^h EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \mathrm{d}x \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix}\right.$$

$$\left. + \int_0^h \rho A \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \lfloor N_1 \quad N_2 \quad N_3 \quad N_4 \rfloor \mathrm{d}x \begin{Bmatrix} \ddot{v}_1 \\ \ddot{v}_{1,x} \\ \ddot{v}_2 \\ \ddot{v}_{2,x} \end{Bmatrix}\right]$$

$$= \lfloor w_1 \quad w_2 \quad w_3 \quad w_4 \rfloor \begin{Bmatrix} \left(-\dfrac{\partial}{\partial x}\left(EI\dfrac{\partial^2 v}{\partial x^2}\right)\right)_0 \\ \left(+EI\dfrac{\partial^2 v}{\partial x^2}\right)_0 \\ \left(+\dfrac{\partial}{\partial x}\left(EI\dfrac{\partial^2 v}{\partial x^2}\right)\right)_h \\ \left(-EI\dfrac{\partial^2 v}{\partial x^2}\right)_h \end{Bmatrix} \tag{12.48}$$

where the double dot on a variable denotes double partial derivative with respect to time.

Since the nodal weights w_1 and w_2 are arbitrary, the elemental finite element equations become

$$\int_0^h EI \begin{Bmatrix} N_{1,xx} \\ N_{2,xx} \\ N_{3,xx} \\ N_{4,xx} \end{Bmatrix} \begin{bmatrix} N_{1,xx} & N_{2,xx} & N_{3,xx} & N_{4,xx} \end{bmatrix} \mathrm{d}x \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} + \int_0^h \rho A \begin{Bmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Bmatrix} \begin{bmatrix} N_1 & N_2 & N_3 & N_4 \end{bmatrix} \mathrm{d}x \begin{Bmatrix} \ddot{v}_1 \\ \ddot{v}_{1,x} \\ \ddot{v}_2 \\ \ddot{v}_{2,x} \end{Bmatrix}$$

$$= \text{Internal force vector} \tag{12.49}$$

The first integral on the left-hand side is same as that obtained in the left-hand side of Eq. (6.23) and is called *elemental stiffness matrix*, whereas the second integral is called *elemental mass matrix*. The elemental stiffness matrix for a beam is given by Eq. (1.46). The elemental mass matrix for the beam element is written as

$$[m^e] = \frac{\rho A h}{420} \begin{bmatrix} 156 & 22h & 54 & -13h \\ 22h & 4h^2 & 13h & -3h^2 \\ 54 & 13h & 156 & -22h \\ -13h & -3h^2 & -22h & 4h^2 \end{bmatrix} \tag{12.50}$$

This is called *consistent mass matrix*. The reader can verify that the entries associated with translational degrees of freedom add to the total mass of the element. One possible lumped mass matrix can be written as

$$\left[m^e\right] = \frac{\rho A h}{2} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{12.51}$$

in which one-half of the total mass has been lumped at each node, corresponding to the translational degree of freedom. It is also possible to lump the effect of the mass moment of inertia at the degrees of freedom corresponding to rotation. The elemental matrices can be assembled in the usual manner. The internal force vectors will cancel each other, except at the nodes where essential boundary conditions are prescribed. However, the rows and columns corresponding to essential boundary conditions may be eliminated from the global matrices. In view of this, the global finite element equations are obtained as

$$[K]\{V\} + [M]\{\ddot{V}\} = \{0\} \tag{12.52}$$

where $[K]$ is the global stiffness matrix, $[M]$ is the global mass matrix, and $\{V\}$ is the vector containing unknown displacements and slopes. In free vibrations, the structure undergoes harmonic motion. Thus

$$\{V\} = \{X\}\sin\omega t + \{Y\}\cos\omega t \quad \text{and} \quad \{\ddot{V}\} = -\omega^2\{X\}\sin\omega t - \omega^2\{Y\}\cos\omega t \tag{12.53}$$

where $\{X\}$ and $\{Y\}$ are the vectors which calculate amplitudes of vibrations at nodes and ω is the circular frequency. Substituting Eq. (12.53) in Eq. (12.52), we obtain the following eigenvalue problem for free vibration:

$$\left[[K]-\omega^2[M]\right]\{X\}\sin\omega t+\left[[K]-\omega^2[M]\right]\{Y\}\cos\omega t=\{0\} \tag{12.54}$$

Equation (12.54) holds true for all values of t. At $\omega t = 0$, Eq. (12.54) reduces to

$$\left[[K]-\omega^2[M]\right]\{Y\}=\{0\} \tag{12.55}$$

At $\omega t = \pi/2$, Eq. (12.54) reduces to

$$\left[[K]-\omega^2[M]\right]\{X\}=\{0\} \tag{12.56}$$

Equations (12.55) and (12.56) are actually the same eigenvalue problem. From any of these equations, the natural frequencies and mode shapes may be found out. The accuracy of the solution increases with finer discretization.

Note that instead of the assumed solution given in Eq. (12.51), we could have assumed solution of the form given in Eq. (12.19) and obtained the eigenvalue problem given by Eq. (12.55). Here, a slightly different form of the assumed solution has been used to provide confidence to readers and also to help them understand the concept better.

Example 12.2 Using only one finite element, obtain an expression for the natural frequency of a uniform cantilever Euler–Bernoulli beam. Do not consider rotary inertia of the cross-section.

Solution Let the length of the beam be l, flexural rigidity be EI, density of beam material be ρ, and area of cross-section be A. The stiffness matrix of the beam element of length l is given as

$$[k]=\frac{EI}{l^3}\begin{bmatrix} 12 & 6l & -12 & 6l \\ 6l & 4l^2 & -6l & 2l^2 \\ -12 & -6l & 12 & -6l \\ 6l & 2l^2 & -6l & 4l^2 \end{bmatrix}$$

and mass matrix is given by

$$[m^e]=\frac{\rho Al}{420}\begin{bmatrix} 156 & 22l & 54 & -13l \\ 22l & 4l^2 & 13l & -3l^2 \\ 54 & 13l & 156 & -22l \\ -13l & -3l^2 & -22l & 4l^2 \end{bmatrix} \tag{12.57}$$

In a cantilever beam, one end is fixed. Considering that the displacement and slope at the first node is zero, the first two rows and columns of the stiffness matrix and mass matrix can be deleted. Thus, the stiffness and mass matrices after the application of boundary conditions are given by

$$[K]=\frac{EI}{l^3}\begin{bmatrix} 12 & -6l \\ -6l & 4l^2 \end{bmatrix} \tag{12.58}$$

and

$$[M]=\frac{\rho Al}{420}\begin{bmatrix} 156 & -22l \\ -22l & 4l^2 \end{bmatrix} \tag{12.59}$$

Using Eq. (12.56), we obtain

$$\frac{EI}{l^3}\begin{bmatrix} 12 & -6l \\ -6l & 4l^2 \end{bmatrix}\{X\} - \frac{\omega^2 \rho A l}{420}\begin{bmatrix} 156 & -22l \\ -22l & 4l^2 \end{bmatrix}\{X\} = 0 \tag{12.60}$$

Substituting

$$\alpha = \frac{\omega^2 \rho A l^4}{420EI} \tag{12.61}$$

Equation (12.60) may be written as

$$\begin{bmatrix} 12 - 156\alpha & -6l + 22\alpha \\ -6l + 22\alpha & 4l^2 - 4l^2\alpha \end{bmatrix}\{X\} = 0 \tag{12.62}$$

For the non-trivial solution of Eq. (12.62), the determinant of the coefficient matrix should be zero. Thus

$$140\alpha^2 - 408\alpha + 12 = 0 \tag{12.63}$$

providing us with two roots: 0.0297 and 2.8846. Using $\alpha = 0.0297$ in Eq. (12.61), the first natural frequency is obtained as

$$\omega_1 = \frac{3.5319}{l^2}\sqrt{\frac{EI}{\rho A}} \tag{12.64}$$

and with $\alpha = 2.8846$, the second natural frequency is obtained as

$$\omega_2 = \frac{34.8071}{l^2}\sqrt{\frac{EI}{\rho A}} \tag{12.65}$$

The exact values of the two natural frequencies are [Rao and Gupta (1984)]

$$\omega_1 = \frac{3.5156}{l^2}\sqrt{\frac{EI}{\rho A}}, \quad \omega_2 = \frac{22.0336}{l^2}\sqrt{\frac{EI}{\rho A}} \tag{12.66}$$

It is seen that FEM solution with one element provides a very accurate solution for the first natural frequency (error being less than 0.5%). However, there is a significant error in the second natural frequency, which implies that a more refined mesh is needed to evaluate the second natural frequency with accuracy. It is worth noting that the frequencies are more than the corresponding exact values.

12.4 FINDING OUT NATURAL FREQUENCIES OF A DAMPED SYSTEM

In Sections 12.2 and 12.3, we have studied the vibrations of undamped systems. In case of undamped systems, the vibrations, once started, continue forever due to the absence of damping. However, in real systems, damping is always present, which causes the amplitude of the vibrations to decrease with time until it becomes zero; at this point, the vibration stops.

There are different ways in which damping can be handled depending on the type of damping and ease of computation. We shall discuss viscous damping in a rod in detail. The reader will then be able to obtain finite element formulation of beam and other structures in a similar manner. The governing differential equation for a beam with viscous damping is written as

$$\frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right)-c\frac{\partial u}{\partial t}-\rho A\frac{\partial^2 u}{\partial t^2}=0 \tag{12.67}$$

where c is a damping parameter analogous to viscosity. The parameter c has a unit of Ns/m^2. Equating the weighted integral over the element to zero, we obtain

$$\int_0^h w\left[\frac{\partial}{\partial x}\left(EA\frac{\partial u}{\partial x}\right)-c\frac{\partial u}{\partial t}-\rho A\frac{\partial^2 u}{\partial t^2}\right]\mathrm{d}x=0 \tag{12.68}$$

The first term in the integrand is integrated by parts once to reduce the order of derivative in x. The other two terms are retained as given. Considering completeness and compatibility, we understand that a two-noded, C^0 continuity element is good enough to serve our purpose. The function and its second derivative are approximated using Eq. (12.6). The first derivative with respect to time is approximated as

$$\dot{u}^e=\lfloor N_1\ \ N_2\rfloor\begin{Bmatrix}\dot{u}_1\\ \dot{u}_2\end{Bmatrix}=\lfloor N\rfloor\{\dot{u}^{ne}\} \tag{12.69}$$

Similar to Eq. (12.10), after substituting the approximation for displacement and weight, we obtain the following weak form:

$$\lfloor w_1\ \ w_2\rfloor\left[\begin{Bmatrix}-EA\left.\dfrac{\partial u}{\partial x}\right|_2\\ EA\left.\dfrac{\partial u}{\partial x}\right|_1\end{Bmatrix}-\int_0^h\begin{Bmatrix}N_{1,x}\\ N_{2,x}\end{Bmatrix}EA\lfloor N_{1,x}\ \ N_{2,x}\rfloor\mathrm{d}x\begin{Bmatrix}u_1\\ u_2\end{Bmatrix}-\int_0^h\rho A\begin{Bmatrix}N_1\\ N_2\end{Bmatrix}\lfloor N_1\ \ N_2\rfloor\mathrm{d}x\begin{Bmatrix}\ddot{u}_1\\ \ddot{u}_2\end{Bmatrix}\right]$$

$$-\lfloor w_1\ \ w_2\rfloor\int_0^h c\begin{Bmatrix}N_1\\ N_2\end{Bmatrix}\lfloor N_1\ \ N_2\rfloor\mathrm{d}x\begin{Bmatrix}\dot{u}_1\\ \dot{u}_2\end{Bmatrix}=0 \tag{12.70}$$

Note the presence of an extra term due to damping. As the nodal weights are arbitrary, we get the following elemental equations:

$$\int_0^h\begin{Bmatrix}N_{1,x}\\ N_{2,x}\end{Bmatrix}EA\lfloor N_{1,x}\ \ N_{2,x}\rfloor\mathrm{d}x\begin{Bmatrix}u_1\\ u_2\end{Bmatrix}+\int_0^h\rho A\begin{Bmatrix}N_1\\ N_2\end{Bmatrix}\lfloor N_1\ \ N_2\rfloor\mathrm{d}x\begin{Bmatrix}\ddot{u}_1\\ \ddot{u}_2\end{Bmatrix}$$

$$+\int_0^h c\begin{Bmatrix}N_1\\ N_2\end{Bmatrix}\lfloor N_1\ \ N_2\rfloor\mathrm{d}x\begin{Bmatrix}\dot{u}_1\\ \dot{u}_2\end{Bmatrix}=\begin{Bmatrix}-EA\left.\dfrac{\partial u}{\partial x}\right|_2\\ EA\left.\dfrac{\partial u}{\partial x}\right|_1\end{Bmatrix} \tag{12.71}$$

Writing Eq. (12.71) in notational form, we obtain

$$[m^e]\{\ddot{u}^{ne}\}+[c^e]\{\dot{u}^{ne}\}+[k^e]\{u^{ne}\}=\{F_i^e\} \tag{12.72}$$

where $[c^e]$ is the elemental damping matrix. For a constant damping parameter c, it is written as

$$[c^e]=ch\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \tag{12.73}$$

After the assembly and application of boundary conditions, we get the following system of equations:

$$[M]\{\ddot{U}\}+[C]\{\dot{U}\}+[K]\{U\}=\{0\} \tag{12.74}$$

where $[C]$ is the damping matrix. We shall get a similar equation in case of damped vibrations of a beam.

To solve Eq. (12.74), we assume the solution of the form given in Eq. (12.19). Substituting this form, we get

$$\lambda^2[M]\{X\}+\lambda[C]\{X\}+[K]\{X\}=0 \tag{12.75}$$

One way to solve the above equation is as follows [Lee, et al. (1999)]. We make use of the following identity:

$$[M]\{\dot{U}\}-[M]\{\dot{U}\}=\{0\} \tag{12.76}$$

In Eq. (12.76), substituting the approximation

$$\{U\}=\{X\}e^{\lambda t};\quad \{\dot{U}\}=\lambda\{X\}e^{\lambda t} \tag{12.77}$$

we get

$$[M]\{\lambda X\}e^{\lambda t}-[M]\lambda\{X\}e^{\lambda t}=0 \tag{12.78}$$

As $e^{\lambda t}$ is not equal to zero, Eq. (12.78) becomes

$$[M]\{\lambda X\}-[M]\lambda\{X\}=0 \tag{12.79}$$

or

$$[M]\{\lambda X\}=\lambda[M]\{X\} \tag{12.80}$$

Rewriting Eq. (12.75) as

$$-[K]\{X\}=\lambda[C]\{X\}+\lambda[M]\{\lambda X\} \tag{12.81}$$

and combining it with Eq. (12.80), we get

$$\begin{bmatrix}[-K] & [0]\\ [0] & [M]\end{bmatrix}\begin{Bmatrix}\{X\}\\ \{\lambda X\}\end{Bmatrix}=\lambda\begin{bmatrix}[C] & [M]\\ [M] & [0]\end{bmatrix}\begin{Bmatrix}\{X\}\\ \{\lambda X\}\end{Bmatrix} \tag{12.82}$$

Note that in Eq. (12.82), matrices and vectors are written in the partitioned form. The matrix $[0]$ is of the same size as the matrices $[M]$ and $[K]$. Thus, if $[M]$ and $[K]$ are $n\times n$ matrices, the matrices in Eq. (12.82) will be of the order $2n\times 2n$. Denoting

$$[A]=\begin{bmatrix}-[K] & [0]\\ [0] & [M]\end{bmatrix}\quad [B]=\begin{bmatrix}[C] & [M]\\ [M] & [0]\end{bmatrix}\quad \text{and}\quad \{Z\}=\begin{Bmatrix}\{X\}\\ \{\lambda X\}\end{Bmatrix} \tag{12.83}$$

we can write Eq. (12.82) as

$$[A]\{Z\}=\lambda[B]\{Z\} \tag{12.84}$$

This generalized eigenvalue problem can be solved by using standard methods. A similar procedure can be adopted for solving other problems with damping. In general, eigenvalues will be a complex numbers that can be represented as

$$\lambda=a+i\omega \tag{12.85}$$

where a and ω are real numbers and i is the square root of -1. Thus

$$\{U\}=\{X\}e^{at+i\omega t}=\{X\}e^{at}(\cos\omega t+i\sin\omega t) \tag{12.86}$$

This is a periodic function of frequency ω (time period $2\pi/\omega$). Thus, the imaginary part of the eigenvalue gives the frequency of the vibration. If the eigenvalue obtained is real, the system will not vibrate. For damped systems, a will be negative, hence reducing the amplitude of vibration with time due to the presence of the factor e^{at} in Eq. (12.86). However, if a is found to be positive, it indicates that the system is unstable and that the displacements will keep on increasing with time (till the physics of the problem changes the governing equation itself!).

Example 12.3 A rod with an area of cross-section 1 cm^2 and length 1 m is fixed at one end, while the other end is kept free. The rod is made of a steel-like material and possesses a density of 7.8 $\times 10^3$ kg/m^3 and Young's modulus of elasticity 200 GPa. Assume a damping parameter of 200 Ns/m^2. Also assume that the area of cross-section of the beam is such that it is stiffer in the transverse direction. Find out the fundamental frequency of the longitudinal vibration using FEM with one element and compare it with undamped fundamental frequency.

Solution The first natural frequency is called *fundamental frequency*. The governing differential equation for this problem is given by Eq. (12.67). In case of one element, we use the following values: $h = 1$ m, $\rho = 7.8 \times 10^3$ kg/m^3, $A = 1 \times 10^{-4}$ cm^2, $E = 200 \times 10^9$ N/m^2, and $c = 200$ Ns/m^2. We use standard SI units in all calculations. From Eq. (12.24), for one element, we obtain

$$[m^1] = \frac{\rho A h}{6}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 0.26 & 0.13 \\ 0.13 & 0.26 \end{bmatrix}, \quad [k^1] = \frac{EA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = 200\times 10^5 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and from Eq. (12.73), we get

$$[c^e] = \begin{bmatrix} 400 & 200 \\ 200 & 400 \end{bmatrix}$$

As we have taken only one element, there is no need to carry out the assembly procedure. Since one end is fixed, the boundary condition is applied by eliminating the first row and first column. Thus

$$[M] = [0.26]\text{kg}, \quad [K] = [200\times 10^5]\text{N/m}, \quad [C] = [400]\text{Ns/m} \tag{12.87}$$

From Eq. (12.83), we obtain

$$[A] = \begin{bmatrix} -200\times 10^5 & 0 \\ 0 & 0.26 \end{bmatrix}, \quad [B] = \begin{bmatrix} 400 & 0.26 \\ 0.26 & 0 \end{bmatrix} \tag{12.88}$$

Solving the generalized eigenvalue problem

$$[A]\{Z\} = \lambda[B]\{Z\}$$

by MATLAB®, we get two eigenvalues as

$$\lambda_1 = -769.2 + 8736.8i, \quad \lambda_2 = -769.2 - 8736.8i \tag{12.89}$$

The complex part of the eigenvalue is 8736.8. Thus, fundamental frequency is 8736.8 rad/s. For undamped system, matrix $[A]$ remains same, while $[B]$ is given by

$$[B] = \begin{bmatrix} 0 & 0.26 \\ 0.26 & 0 \end{bmatrix} \tag{12.90}$$

With this value of $[B]$, eigenvalues are obtained to be

$$\lambda_1 = 8770.6i, \quad \lambda_2 = -8770.6i \tag{12.91}$$

Thus, the fundamental frequency of vibration for the undamped system is 8770.6 rad/s. We observe that damping reduces the fundamental frequency.

In the above example, we notice that damping parameter is not dependent on frequency. In many structural materials, damping is quite small but is frequency-dependent. Such damping is often modelled by Rayleigh (or proportional) damping. In this model, the damping matrix is taken as a linear combination of stiffness and mass matrices. Thus

$$[C] = \alpha[K] + \beta[M] \tag{12.92}$$

where α is a stiffness proportional damping constant and β is a mass proportional damping constant. It is usual practice to characterize damping by a parameter called *damping ratio*, which is denoted by the symbol ξ. Damping ratio is defined as the ratio of the damping in the system to critical damping. At critical damping, the damping ratio ξ becomes equal to 1. If ξ is less than 1, the system is oscillatory and is called *under-damped*. If ξ is greater than 1, the system is non-oscillatory and is called *over-damped*. It can be shown [Bathe (1996)] that

$$\xi = \frac{1}{2}\left(\alpha\omega + \frac{\beta}{\omega}\right) \tag{12.93}$$

If damping ratios (ξ_1 and ξ_2) are determined experimentally at two different frequencies (ω_1 and ω_2), then the parameters α and β can be found from the following relations:

$$\alpha = 2(\xi_2\omega_2 - \xi_1\omega_1)/(\omega_2^2 - \omega_1^2) \tag{12.94}$$

$$\beta = 2\omega_1\omega_2(\xi_1\omega_2 - \xi_2\omega_1)/(\omega_2^2 - \omega_1^2) \tag{12.95}$$

After that, the damping matrix can be constructed using Eq. (12.92).

12.5 FORMULATION OF TWO- AND THREE-DIMENSIONAL PROBLEMS: A NOTE

The methodology discussed in the previous sections can be extended to two- and three-dimensional problems. In general, the elemental finite element equations are of the form

$$[m^e]\{\ddot{\Delta}^e\} + [c^e]\{\dot{\Delta}^e\} + [k^e]\{\Delta^e\} = \{F_i^e\} \tag{12.96}$$

where $\{\Delta^e\}$ is a vector containing nodal degrees of freedom. For plane strain and plane stress problems, the mass matrix is obtained as

$$[m^e] = \int_{\Omega^e} \rho t[N]^{\mathrm{T}}[N]\,\mathrm{d}x\mathrm{d}y \tag{12.97}$$

where ρ is the density of the material, t is the thickness, and $[N]$ is defined by Eq. (11.29). For three-dimensional problems, the mass matrix is given by

$$[m^e] = \int_{V^e} \rho[N]^{\mathrm{T}}[N]\,\mathrm{d}V \tag{12.98}$$

The damping matrix for plane strain and plane stress problems is given by

$$[c^e] = \int_{\Omega^e} c_A [N]^{\mathrm{T}} [N] \, \mathrm{d}x\mathrm{d}y \tag{12.99}$$

where c_A is the per unit area damping with units of Ns/m^3 in SI system. For three-dimensional problems, the damping matrix is given by

$$[c^e] = \int_{V^e} c_V [N]^{\mathrm{T}} [N] \, \mathrm{d}V \tag{12.100}$$

where c_V is the per unit volume damping with units of Ns/m^4 in SI system. Alternatively, Rayleigh damping scheme may be employed. The procedure of assembly, application of boundary conditions, and solution are similar and will not be repeated here.

12.6 CONCLUSION

In this chapter, finite element formulations of the free vibration of rod and beam have been carried out. The formulations provide eigenvalue problems. Usually, in these problems, the accuracy of the fundamental frequency is more than the accuracy of other natural frequencies. The accuracy keeps on decreasing for higher modes. The discretization should be carried out depending on how many modes are important for the problem. The degrees of freedom of FEM system should be significantly more than the highest mode for which the natural frequency is desired. Section 12.5 outlines the procedure for formulating the free-vibration problems in two and three dimensions.

SOLVED PROBLEMS

12.1 **Problems of free vibrations are essentially eigenvalue problems; the problems of structural instability also fall in this category. Obtain the finite element formulation for buckling of a beam by following the procedure similar to free-vibration problems.**

Solution The governing differential equation for the buckling of the beam is written as

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(EI\frac{\mathrm{d}^2 v}{\mathrm{d}x}\right) + P\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} = 0 \tag{12.1.1}$$

For an element of length h, making the weighted integration of this differential equation equal to zero will lead to

$$\int_0^h w\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left(EI\frac{\mathrm{d}^2 v}{\mathrm{d}x}\right)\mathrm{d}x + \int_0^h wP\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\mathrm{d}x = 0 \tag{12.1.2}$$

where w is the weight function. We integrate the first integral by parts twice and second integral once in order to distribute the order of derivative evenly between v and w. Thus

$$\begin{aligned} &w\frac{\mathrm{d}}{\mathrm{d}x}\left(EI\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\right)\Bigg|_0^h - \frac{\mathrm{d}w}{\mathrm{d}x}EI\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\Bigg|_0^h + \int_0^h EI\frac{\mathrm{d}^2 w}{\mathrm{d}x^2}\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\mathrm{d}x \\ &+ Pw\frac{\mathrm{d}v}{\mathrm{d}x}\Bigg|_0^h - \int_0^h P\frac{\mathrm{d}v}{\mathrm{d}x}\frac{\mathrm{d}w}{\mathrm{d}x}\mathrm{d}x = 0 \end{aligned} \tag{12.1.3}$$

The highest order of the derivative in Eq. (12.1.3) is 3 and the highest order of derivative in the integral is 2. Therefore, the completeness upto third degree and C^1 compatibility is required. Hermitian shape functions used to approximate a slender beam subjected to bending will be suitable. The approximation can now be written as

$$v = \lfloor N_1 \quad N_2 \quad N_3 \quad N_4 \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} = \lfloor N \rfloor \{v^{ne}\} \tag{12.1.4}$$

$$v_{,xx} = \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \begin{Bmatrix} v_1 \\ v_{1,x} \\ v_2 \\ v_{2,x} \end{Bmatrix} = \lfloor N_{,xx} \rfloor \{v^{ne}\} \tag{12.1.5}$$

In Galerkin FEM, the shape functions are approximated in the same way as weight functions. Therefore

$$w = \lfloor N_1 \quad N_2 \quad N_3 \quad N_4 \rfloor \begin{Bmatrix} w_1 \\ w_{1,x} \\ w_2 \\ w_{2,x} \end{Bmatrix} = \lfloor N \rfloor \{w^{ne}\} \tag{12.1.6}$$

$$w_{,xx} = \lfloor N_{1,xx} \quad N_{2,xx} \quad N_{3,xx} \quad N_{4,xx} \rfloor \begin{Bmatrix} w_1 \\ w_{1,x} \\ w_2 \\ w_{2,x} \end{Bmatrix} = \lfloor N_{,xx} \rfloor \{w^{ne}\} \tag{12.1.7}$$

Substituting the approximation in Eq. (12.1.3) and rearranging the terms, we obtain

$$\lfloor w^{ne} \rfloor \int_0^h EI \{N_{,xx}\} \lfloor N_{,xx} \rfloor \mathrm{d}x \{v^{ne}\} - P \lfloor w^{ne} \rfloor \int_0^h \{N_{,x}\} \lfloor N_{,x} \rfloor \mathrm{d}x \{v^{ne}\} =$$

$$\lfloor w^{ne} \rfloor \begin{Bmatrix} \left(\dfrac{\mathrm{d}}{\mathrm{d}x} \left(EI \dfrac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right) + P \dfrac{\mathrm{d}v}{\mathrm{d}x} \right)_1 \\ -\left(EI \dfrac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right)_1 \\ -\left(\dfrac{\mathrm{d}}{\mathrm{d}x} \left(EI \dfrac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right) + P \dfrac{\mathrm{d}v}{\mathrm{d}x} \right)_2 \\ \left(EI \dfrac{\mathrm{d}^2 v}{\mathrm{d}x^2} \right)_2 \end{Bmatrix} \tag{12.1.8}$$

Note that for a beam subjected to an axial load P, the shear force and bending moment are given by

$$V = -\frac{\mathrm{d}}{\mathrm{d}x}\left(EI\frac{\mathrm{d}^2 v}{\mathrm{d}x^2}\right) - P\frac{\mathrm{d}v}{\mathrm{d}x}, \quad M = EI\frac{\mathrm{d}^2 v}{\mathrm{d}x^2} \tag{12.1.9}$$

As $\lfloor w^{ne} \rfloor$ is arbitrary, Eq. (12.1.9) provides

$$\int_0^h EI\{N_{,xx}\}\lfloor N_{,xx}\rfloor \mathrm{d}x\{v^{ne}\} - P\int_0^h \{N_{,x}\}\lfloor N_{,x}\rfloor \mathrm{d}x\{v^{ne}\} = \begin{Bmatrix} -V_1 \\ -M_1 \\ V_2 \\ M_2 \end{Bmatrix} \tag{12.1.10}$$

Equation (12.1.10) is an elemental equation. During assembly the elemental shear forces and bending moments vanish everywhere except at the end nodes. At an end node, if a degree of freedom is not constrained, the generalized force (shear force for displacement and bending moment for slope) corresponding to that degree of freedom is zero and if a degree of freedom is constrained, the row and column corresponding to that degree of freedom can be eliminated. Thus, after assembly and application of the boundary conditions, we shall get the system of equations of the following form:

$$[K]\{V\} - P[F]\{V\} = \{0\} \tag{12.1.11}$$

Equation (12.1.11) represents an eigenvalue problem with axial force P as the eigenvalue. It can be solved in the usual manner.

12.2 Obtain the mass matrix for a three-noded element of length h and per unit length mass of m that undergoes longitudinal oscillations.

Solution The mass matrix of a three-noded element is given by

$$[m^e] = \int_0^h m\begin{Bmatrix} N_1 \\ N_2 \\ N_3 \end{Bmatrix}\lfloor N_1 \quad N_2 \quad N_3 \rfloor \mathrm{d}x \tag{12.2.1}$$

and the shape functions are written as

$$N_1 = \frac{(2x-h)(x-h)}{h^2}, \quad N_2 = \frac{4x(h-x)}{h^2}, \quad N_3 = \frac{x(2x-h)}{h^2} \tag{12.2.2}$$

Substituting the values given in Eq. (12.2.2) in Eq. (12.2.1) and integrating, we obtain

$$[m^e] = \frac{mh}{30}\begin{bmatrix} 4 & 2 & -1 \\ 2 & 16 & 2 \\ -1 & 2 & 4 \end{bmatrix} \tag{12.2.3}$$

EXERCISES

A. Multiple Choice Questions

12.1 The consistent mass matrix is always a

(a) positive definite matrix (b) positive semi-definite matrix

(c) negative definite matrix (d) negative semi-definite matrix

12.2 In Rayleigh damping, the damping matrix $[C]$ is a linear combination of the stiffness and mass matrices such that

$$[C] = \alpha[K] + \beta[M]$$

In this case, the minimum damping ratio is obtained at an angular frequency equal to

(a) $\sqrt{\dfrac{\alpha}{\beta}}$ (b) $\sqrt{\alpha\beta}$ (c) $\sqrt{\dfrac{\beta}{\alpha}}$ (d) 0

12.3 In SI system, the ratio α/β has the units

(a) s^{-1} (b) s^{-2} (c) s (d) s^2

12.4 A light rod has a mass M at its center. The consistent mass matrix of the rod for longitudinal vibration is

(a) $\begin{bmatrix} \frac{M}{2} & 0 \\ 0 & \frac{M}{2} \end{bmatrix}$ (b) $\begin{bmatrix} \frac{M}{4} & \frac{M}{4} \\ \frac{M}{4} & \frac{M}{4} \end{bmatrix}$ (c) $\begin{bmatrix} \frac{M}{3} & \frac{M}{6} \\ \frac{M}{6} & \frac{M}{3} \end{bmatrix}$ (d) $\begin{bmatrix} \frac{M}{2} & \frac{M}{4} \\ \frac{M}{4} & \frac{M}{2} \end{bmatrix}$

12.5 Arrange the following beams possessing the same geometry and made of the same material in increasing order of frequency: (i) simply supported (ii) fixed-fixed (iii) cantilever (iv) free-free.

(a) i, ii, iii, iv (b) iv, iii, i, ii (c) iv, iii, ii, i (d) ii, iii, i, iv

B. Problems

12.6 Figure 12.1(a) is a schematic diagram of a single-link flexible manipulator with length L, Young's modulus E, transverse area moment of inertia I, mass per unit length m, and hub inertia J_h; the element is subjected to a torque T. In the figure, XOY and POQ represent the fixed and rotating coordinate frames, respectively. The manipulator is assumed to vibrate dominantly in the XOY plane only. Considering the manipulator to be long and slender, transverse shear and rotary inertia effects are neglected and hence the manipulator is modelled by Euler–Bernoulli beam theory.

For small values of angular displacement θ, angular velocity $\dot{\theta}$, and deflection w, the governing linear partial differential equations for undamped motion are obtained as follows:

$$\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 w}{\partial x^2}\right) + m\frac{\partial^2 w}{\partial t^2} + m\,x\frac{d^2\theta}{dt^2} = q \tag{12.6.1}$$

$$J_h \frac{d^2\theta}{dt^2} + \int_0^L mx\left(\frac{\partial^2 w}{\partial t^2} + x\frac{d^2\theta}{dt^2}\right)dx = T \tag{12.6.2}$$

Note that m and I are functions of x—the distance of the point on the manipulator from the fixed point O—and transverse load q is a function of both x and time t. Tip mass M_p can be handled by using Dirac-delta function, i.e., by adding a term $M_p\delta(x - L)$ in m. The boundary conditions at the torque end are given as

$$w(0,t) = 0 \tag{12.6.3}$$

$$\left.\frac{\partial w}{\partial x}\right|_{x=0} = 0 \tag{12.6.4}$$

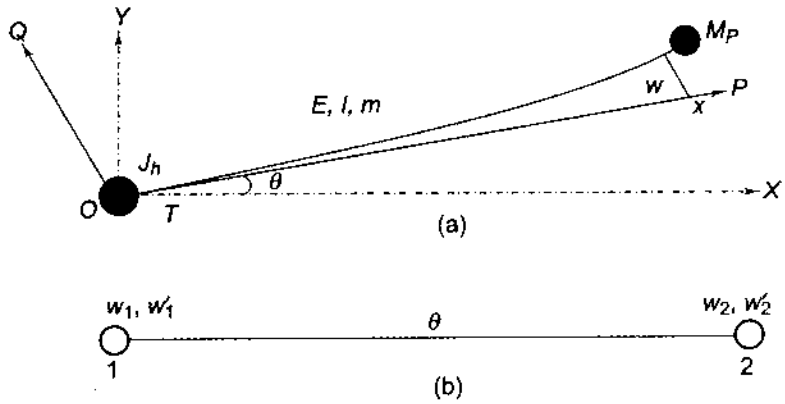


Figure 12.1 *(a) A single-link flexible manipulator (b) An element having five degrees of freedom*

The natural boundary conditions at the free end are written as

$$\left.\frac{\partial^2 w}{\partial x^2}\right|_{x=L} = 0 \tag{12.6.5}$$

$$\left.\frac{\partial}{\partial x}\left(EI\frac{\partial^2 w}{\partial x^2}\right)\right|_{x=L} = 0 \tag{12.6.6}$$

For FEM formulation, the manipulator is divided into n elements. A typical element having five degrees of freedom is shown in Fig. 12.1(b). Here, w_1, w_1', w_2, w_2' are the transverse deflections and slopes at the first and second nodes of the element.

Now, answer the following questions:

(a) Prove that the finite element formulation by Galerkin's approach yields the following elemental mass matrix for the ith element:

$$M_i^e = \frac{m_i h}{420}\begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} & m_{15} \\ m_{21} & 156 & 22h & 54 & -13h \\ m_{31} & 22h & 4h^2 & 13h & -3h^2 \\ m_{41} & 54 & 13h & 156 & -22h \\ m_{51} & -13h & -3h^2 & -22h & 4h^2 \end{bmatrix} \tag{12.6.7}$$

where m_i is the mass per unit length of the element obtained by multiplying the average area of the element with mass density, h is the element length, and

$$m_{11} = 140h^2(3i^2 - 3i + 1) \qquad m_{12} = m_{21} = 21h(10i - 7) \qquad m_{13} = m_{31} = 7h^2(5i - 3)$$

$$m_{14} = m_{41} = 21h(10i - 3) \qquad m_{15} = m_{51} = -7h^2(5i - 2)$$

(b) Prove that to take into account the hub inertia and tip mass, the following terms are to be added in the assembled mass matrix of size $(2n + 3) \times (2n + 3)$

(i) $J_h + M_p L^2$ in the first element of the principal diagonal

(ii) $M_p L$ at the element $(1, 2n + 2)$ and $(2n + 2, 1)$

(iii) M_p at the element $(2n + 2, 2n + 2)$.

(c) Prove that the element stiffness matrix is obtained as

$$K_i^e = \frac{EI_i}{h^3}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 12 & 6h & -12 & 6h \\ 0 & 6h & 4h^2 & -6h & 2h^2 \\ 0 & -12 & -6h & 12 & -6h \\ 0 & 6h & 2h^2 & -6h & 4h^2 \end{bmatrix} \tag{12.6.8}$$

where I_i is the average moment of inertia of the ith element. The size of an element being small, taking the average value of the area of cross-section and area moment of inertia yields sufficiently accurate results.

12.7 Carry out the FEM formulation of a rod vibrating freely with damping. For a uniform rod, taking two elements, find out the fundamental frequencies with and without damping.

12.8 Find out the fundamental frequency and the corresponding mode shape for a uniform rod subjected to the following boundary conditions:

(a) free–free (b) fixed–free (c) fixed–fixed

In each case, first find out the exact solution, then study the convergence of FEM solution starting from the one-element model to ten-element model.

12.9 A rod is pushed by a force F on a smooth surface. The rigid body displacement of the rod is denoted by a, which is the displacement of the node 1. The displacement of the particles relative to node 1 is denoted by u, which is a function of x. The governing equations of motion are written as

$$EA\frac{\partial^2 u}{\partial x^2} - m\frac{\partial^2 (a+u)}{\partial t^2} = 0$$

$$F = \int_0^l m\frac{\partial^2 (a+u)}{\partial t^2}\,\mathrm{d}x$$

where E is the Young modulus of elasticity, A is the area of cross-section, and m is the mass per unit length.

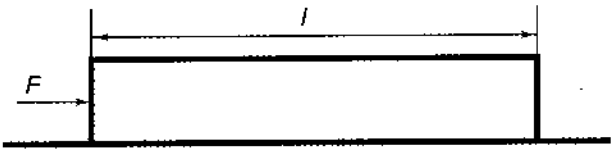


Figure 12.2 *A rod of length l with a force F acting on it*

The one-element finite element formulation of this problem leads to

$$[M]\begin{Bmatrix}\ddot{a}\\ \ddot{u}_1\\ \ddot{u}_2\end{Bmatrix} + [K]\begin{Bmatrix}a\\ u_1\\ u_2\end{Bmatrix} = \begin{Bmatrix}F\\ -EA\left.\dfrac{\mathrm{d}u}{\mathrm{d}x}\right|_1\\ 0\end{Bmatrix}$$

Find out the expression for mass matrix $[M]$ and stiffness matrix $[K]$.

REFERENCES

Bathe, KJ, 1996, *Finite Element Procedures*, Prentice-Hall, Englewood Cliffs, NJ.

Lee, IW, Kim, DO, and Jung, JH, 1999, 'Natural frequencies and mode shape sensitivities of damped system: Part 1—Distinct natural frequencies', *Journal of Sound and Vibration*, Vol. 222, pp. 399–412.

Rao, JS, and Gupta, K, 1984, *Introductory Course on Theory and Practice of Mechanical Vibrations*, New Age International (P) Limited, New Delhi.

13

CHAPTER

FEM Formulation for Time-History Problems

13.1 Introduction

13.2 Classification of Partial Differential Equations

13.3 Time Response of a Parabolic Equation

13.4 Forced Vibration Problems

13.5 Conclusion

13.1 INTRODUCTION

In this chapter we will study problems where we are interested in finding out the time response of the system. By *time response*, we mean the value of a variable as a function of time. In the previous chapter, we obtained the frequencies of free vibrations and mode shapes. If the mode shape, frequencies, and initial conditions are known, then the periodic motion is known completely. In general, the motion of an oscillating body is a linear combination of motions corresponding to different modes. In forced vibrations, we have a forcing function that may depend on time. Therefore, the motion at each time may have to be found incrementally. Usually, a finite difference approximation is made for the time derivative. Thus, these problems involve FEM and finite difference schemes in combination. Another way of obtaining the time-history in forced-vibration problems is *mode summation method*. In this method, the displacement of the oscillating body is approximated by the sum of a limited number of modal motions. This method is quite useful for finding out the time-history over a large time interval and can be used when some knowledge about dominant modes already exists. In this book, however, we shall not describe this method. Interested readers may consult reference books such as Bathe (1996) and Cook, et al. (1989) for a detailed discussion on this method.

Apart from the time-history problems in dynamics, we shall also describe transient heat-conduction problems. In fact, the methods described in this chapter may be used for solving a variety of problems. To present the treatment in this chapter in a somewhat general form we start our discussion with the classification of partial differential equations. In Section 13.3, we will present the scheme for obtaining the time response of a parabolic equation. The equation governing one-dimensional heat conduction is an example of a parabolic equation. We will also discuss the formulation of forced vibration problems in later sections.

13.2 CLASSIFICATION OF PARTIAL DIFFERENTIAL EQUATIONS

Consider a differential equation of the form

$$a\frac{\partial^2 z}{\partial x^2}+2h\frac{\partial^2 z}{\partial x\partial y}+b\frac{\partial^2 z}{\partial y^2}+2f\frac{\partial z}{\partial x}+2g\frac{\partial z}{\partial y}+cz=f(x,y) \tag{13.1}$$

The type of differential equation represented by Eq. (13.1) can be determined by the value of $ab-h^2$. The following possibilities may be considered:

(i) When $ab-h^2>0$, the equation is said to be of *elliptic* type.

(ii) When $ab-h^2=0$, the equation is called *parabolic* type.

(iii) When $ab-h^2<0$, the equation is said to be of *hyperbolic* type.

Consider the Laplace's equation, which is given by

$$\frac{\partial^2 z}{\partial x^2}+\frac{\partial^2 z}{\partial y^2}=0 \tag{13.2}$$

Here, $a = 1$, $b = 1$, $h = f = g = c = 0$. Therefore, $ab - h^2 > 0$. Thus, Laplace's equation is of elliptic type.

Consider one-dimensional transient heat-conduction equation given by

$$\frac{\partial^2 T}{\partial x^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \tag{13.3}$$

where t is the time and α is the thermal diffusivity. Here, $a = 1$, $b = 0$, $h = 0$. Therefore, $ab - h^2 = 0$. Hence, this equation is of parabolic type.

Consider the problem of two-dimensional transient heat conduction represented by the following equation:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \tag{13.4}$$

Equation (13.4) is a partial differential equation with three independent variables, namely, x, y, and t. In general, the classification of partial differential equations with more than two independent variables is done on the basis of the eigenvalues of the coefficient matrix of the second-order terms. The general form of a partial differential equation is as follows:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} a_i \frac{\partial u}{\partial x_i} + a_0 u = 0 \tag{13.5}$$

In Eq. (13.5)

$$a_{ij} = a_{ji} \tag{13.6}$$

The matrix of the coefficients of second-order term is given by

$$[X] = \begin{bmatrix} a_{11} & a_{12} \ldots a_{1n} \\ a_{21} & a_{22} \ldots a_{2n} \\ \ldots & \ldots \ldots \\ \ldots & \ldots \ldots \\ a_{n1} & a_{n2} \ldots a_{nn} \end{bmatrix} \tag{13.7}$$

Based on the nature of the matrix $[X]$, the classification of second-order partial differential equations is as follows:

- *Elliptic type* If the eigenvalues of $[X]$ are all of the same sign, the equation is said to be of elliptic type.
- *Parabolic type* If one eigenvalue is zero and the others have the same sign, the equation is of parabolic type.
- *Hyperbolic type* If all eigenvalues except one are of the same sign and no eigenvalue is zero, the equation is of hyperbolic type.
- *Ultrahyperbolic type* If there is more than one positive eigenvalues, more than one negative eigenvalues, and there are no zero eigenvalues, the equation is of ultrahyperbolic type.

Now, for the system represented by Eq. (13.4), $[X]$ is written as

$$[X]=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix} \tag{13.8}$$

This matrix has eigenvalues 1, 1, and 0. This implies that Eq. (13.4) is of parabolic type.

Consider the problem of forced vibration of a rod with damping. The governing equation is given by

$$EA\frac{\partial^2 u}{\partial x^2}+\rho A\frac{\partial^2 u}{\partial t^2}+q(t)=0 \tag{13.9}$$

where E is the Young's modulus of elasticity, A is the area of cross-section, ρ is the density, and q is the forcing load intensity. Here, $a = EA$, $b = \rho A$, and $h = 0$. Therefore, $ab - h^2 > 0$. Thus, this equation is of elliptic type.

13.3 TIME RESPONSE OF A PARABOLIC EQUATION

Consider the following parabolic differential equation describing the one-dimensional transient heat conduction with heat generation in an element of a rod:

$$\frac{\partial^2 T}{\partial x^2}-\frac{1}{\alpha}\frac{\partial T}{\partial t}+\frac{\dot{Q}}{k}=0 \tag{13.10}$$

where α is the thermal diffusivity; the unit of this quantity in SI system is m^2/s. In Eq. (13.10), k is the thermal conductivity and $\dot{Q}$ is the rate of heat generation per unit volume. Note that temperature is a function of time t and coordinate x. While writing this equation in an element we assume that α and k are constant in that element, although they may vary from element to element. Integrating the weighted residual of the function over an element of length h and making it equal to zero, we obtain

$$\int_0^h w\left(\frac{\partial^2 T}{\partial x^2}-\frac{1}{\alpha}\frac{\partial T}{\partial t}+\frac{\dot{Q}}{k}\right)\mathrm{d}x=0 \tag{13.11}$$

where the weight w is a function of x and t. In order to evenly distribute the order of derivative with respect to x on w and T, we integrate the first term of the integrand with respect to x and obtain

$$w\frac{\partial T}{\partial x}\bigg|_0^h-\int_0^h\frac{\partial w}{\partial x}\frac{\partial T}{\partial x}\mathrm{d}x-\int_0^h w\frac{1}{\alpha}\frac{\partial T}{\partial t}\mathrm{d}x+\int_0^h w\frac{\dot{Q}}{k}\mathrm{d}x=0 \tag{13.12}$$

In Eq. (13.12), the highest order of derivative with respect to x is 1; hence, a two-noded C^0 continuity element is enough to approximate the temperature field. Thus, the function representing the temperature and its partial derivatives with respect to x and t are written as

$$T=\lfloor N_1\ \ N_2\rfloor\begin{Bmatrix}T_1(t)\\ T_2(t)\end{Bmatrix};\ \frac{\partial T}{\partial x}=\lfloor N_{1,x}\ \ N_{2,x}\rfloor\begin{Bmatrix}T_1\\ T_2\end{Bmatrix};\ \frac{\partial T}{\partial t}=\lfloor N_1\ \ N_2\rfloor\begin{Bmatrix}\dot{T}_1\\ \dot{T}_2\end{Bmatrix} \tag{13.13}$$

where the shape functions are functions of x and nodal temperatures are functions of time. In Galerkin FEM, the weight function is approximated in the same manner as the primary variable. Hence

$$w = \lfloor w_1 \;\; w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}; \;\; \frac{\partial w}{\partial x} = \lfloor w_1 \;\; w_2 \rfloor \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \tag{13.14}$$

Substituting Eqs (13.13) and (13.14) into Eq. (13.12), we obtain

$$\begin{aligned} &\lfloor w_1 \;\; w_2 \rfloor \begin{Bmatrix} -\left.\dfrac{\partial T}{\partial x}\right|_1 \\ \left.\dfrac{\partial T}{\partial x}\right|_2 \end{Bmatrix} - \lfloor w_1 \;\; w_2 \rfloor \int_0^h \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \;\; N_{2,x} \rfloor \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} \mathrm{d}x \\ &- \lfloor w_1 \;\; w_2 \rfloor \int_0^h \frac{1}{\alpha} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \;\; N_2 \rfloor \begin{Bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{Bmatrix} \mathrm{d}x + \lfloor w_1 \;\; w_2 \rfloor \int_0^h \frac{\dot{Q}}{k} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x = 0 \end{aligned} \tag{13.15}$$

As the nodal weights are arbitrary, the Eq. (13.15) implies that

$$\int_0^h \frac{1}{\alpha} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \;\; N_2 \rfloor \begin{Bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{Bmatrix} \mathrm{d}x + \int_0^h \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \;\; N_{2,x} \rfloor \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} \mathrm{d}x = \int_0^h \frac{\dot{Q}}{k} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} -\left.\dfrac{\partial T}{\partial x}\right|_1 \\ \left.\dfrac{\partial T}{\partial x}\right|_2 \end{Bmatrix} \tag{13.16}$$

This is a system of elemental equations. All elemental equations are assembled in order to yield a global system of equation. In general, the thermal properties may differ from element to element. In one-dimensional heat conduction, heat flux coming out from one element is equal to heat flux going into the adjacent element. Therefore, it is better to write Eq. (13.6) as

$$\int_0^h \frac{k}{\alpha} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \lfloor N_1 \;\; N_2 \rfloor \begin{Bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{Bmatrix} \mathrm{d}x + \int_0^h k \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} \lfloor N_{1,x} \;\; N_{2,x} \rfloor \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix} \mathrm{d}x = \int_0^h \dot{Q} \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \mathrm{d}x + \begin{Bmatrix} -k \left.\dfrac{\partial T}{\partial x}\right|_1 \\ k \left.\dfrac{\partial T}{\partial x}\right|_2 \end{Bmatrix} \tag{13.17}$$

This ensures that internal heat fluxes cancel each other. At a boundary, either the heat flux or the temperature is prescribed, or there is some linear combination of two types of boundary conditions. Regardless of the type of boundary condition, it is possible to eliminate the unknown heat flux at the boundary. Thus, the last term of Eq. (13.17) is not calculated. If the area of cross-section of the rod varies from one element to the other, then the heat fluxes at an interface of the element will not balance each other, instead the heat flow coming out from one element will be equal to the heat flow going into the adjacent element. In order to eliminate internal heat flows, Eq. (13.17) can be written in the following form:

$$\int_0^h \frac{kA}{\alpha}\begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\begin{bmatrix} N_1 & N_2 \end{bmatrix}\begin{Bmatrix} \dot{T}_1 \\ \dot{T}_2 \end{Bmatrix}\mathrm{d}x + \int_0^h kA\begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix}\begin{bmatrix} N_{1,x} & N_{2,x} \end{bmatrix}\begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix}\mathrm{d}x =$$

$$\int_0^h \dot{Q}A\begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\mathrm{d}x + \begin{Bmatrix} -kA\left.\dfrac{\partial T}{\partial x}\right|_1 \\ kA\left.\dfrac{\partial T}{\partial x}\right|_2 \end{Bmatrix} \tag{13.18}$$

Strictly speaking, if the area of cross-section of the rod is not constant, the heat transfer cannot be one-dimensional. However, in many cases, one-dimensional approximation does not introduce much error if there is a gradual change in the area.

After the elemental equations of the form given by Eq. (13.18) are assembled and the boundary conditions are applied, we obtain an equation of the following form:

$$[M]\{\dot{U}\} + [K]\{U\} = \{F(t)\} \tag{13.19}$$

where $\{U\}$ is a vector of nodal temperatures and $\{\dot{U}\}$ is the vector of nodal time derivatives of the temperatures for the transient heat-conduction problem. The right-hand-side vector is due to heat generation and may be a function of time. The meaning of the matrices and vectors will change for the parabolic equations representing other physical problems. Usually, initial conditions will be prescribed in the following form:

$$\{U\}_0 = \{U_0\} \tag{13.20}$$

where subscript '0' on the left-hand side indicates time $t = 0$ and that on the right-hand side indicates prescribed values.

Equation (13.19) is a system of ordinary differential equations with t as independent variable. For carrying out the finite difference approximation of the time derivative, consider the equation

$$(1-\beta)\{\dot{U}\}_s + \beta\{\dot{U}\}_{s+1} = \frac{\{U\}_{s+1} - \{U\}_s}{\Delta t_{s+1}} \quad \text{for } 0 \le \beta \le 1 \tag{13.21}$$

where the subscript s indicates the time step number and $\Delta t_{s+1} = (t_{s+1} - t_s)$. Usually, time interval between two successive increments is kept the same, though this is not necessary. For the time being, we will assume that time interval between two successive intervals is Δt. Thus

$$\{U\}_{s+1} = \{U\}_s + \Delta t\left[(1-\beta)\{\dot{U}\}_s + \beta\{\dot{U}\}_{s+1}\right] \quad \text{for } 0 \le \beta \le 1 \tag{13.22}$$

Let us consider the following cases:

Case 1 $\beta = 0$

This is called the *forward difference* (or *forward Euler*) scheme. In this case

$$\{U\}_{s+1} = \{U\}_s + \Delta t\{\dot{U}\}_s \tag{13.23}$$

It is also called explicit scheme because the value of u at time increment $s + 1$ can be obtained from the known information at time increment s. This scheme is conditionally stable and the order of

accuracy is $O(\Delta t)$. The concept of conditional stability can be understood from the following example. Let us assume that the vector of primary nodal variables is given by

$$\{U\} = \{A\}e^{-ct} \tag{13.24}$$

where $\{A\}$ is a vector of constants and c is a positive constant. Thus

$$\{\dot{U}\} = -c\{A\}e^{-ct} = -c\{U\} \tag{13.25}$$

Therefore

$$\{\dot{U}\}_s = -c\{U\}_s \tag{13.26}$$

Substituting the relation given in Eq. (13.26) in Eq. (13.23), we obtain

$$\{U\}_{s+1} = \{U\}_s - c\Delta t\{U\}_s = (1 - c\Delta t)\{U\}_s \tag{13.27}$$

From Eq. (13.27), it can be understood that if $c\Delta t$ is more than 1, then the magnitudes of primary nodal variables at the $(s + 1)$th time step will be more than those at sth time step. Thus, the magnitudes of primary nodal variables will increase with time, which is contrary to the actual behaviour of exponential decaying function [Eq. (13.24)]. In this situation, the scheme is said to become unstable because the error keeps on increasing with each time step. On the other hand, if $c\Delta t$ is less than 1, then the magnitudes of primary nodal variables at the $(s + 1)$th time step will be less than those at sth time step. Thus, the magnitude of primary nodal variables will decrease with time, this is consistent with the actual behaviour. Thus, the error remains bounded with time and the scheme is called stable. Hence, this scheme will be stable below a certain value of time step Δt, and will be unstable above that value. Such a scheme is called *conditionally stable*.

Stability and accuracy are two different aspects. *Accuracy* of a numerical scheme is a measure of the closeness between the approximate and exact solution, whereas *stability* of the scheme is a measure of the boundedness of the approximate solution with time.

Applying Taylor's expansion, we obtain the following expression:

$$\{U\}_{s+1} = \{U\}_s + \Delta t\{\dot{U}\}_s + \frac{(\Delta t)^2}{2!}\{\ddot{U}\}_s + \frac{(\Delta t)^3}{3!}\{\dddot{U}\}_s + \cdots \tag{13.28}$$

In Euler's method, basically the first two terms of the Taylor series are retained in order to obtain Eq. (13.23). The dominant term in the remaining term is the third term of the Taylor series. Hence, the order of error is $(\Delta t)^2$. This is the order of local truncation error. If the time between 0 and t is divided into n steps, the total truncation error at time t will be of the order of $n\,(\Delta t)^2$ and as n is proportional to $1/(\Delta t)$, the order of error at time will be (Δt). That is why it is said that the order of accuracy in Euler method is $O(\Delta t)$.

Case 2 $\beta = 1$

This is called the *backward difference* (or *backward Euler*) scheme. It is stable and the order of accuracy is $O(\Delta t)$.

Case 3 $\beta = \frac{1}{2}$

This is called *Crank–Nicholson* scheme. It is also stable and the order of accuracy is $O\left((\Delta t)^2\right)$.

Case 4 $\beta = \frac{2}{3}$

This is called Galerkin method. It is also stable and the order of accuracy is $O\left((\Delta t)^2\right)$.

Cases 2, 3, and 4 are known as *implicit methods*. In these methods, we can find $\{U\}$ at time increment $s + 1$, only if we have the information at time increment s as well as $s + 1$.

The finite element equations at two time intervals are given by

$$[M]\{\dot{U}\}_s + [K]_s\{U\}_s = \{F\}_s \tag{13.29}$$

$$[M]\{\dot{U}\}_{s+1} + [K]_{s+1}\{U_{s+1}\} = \{F\}_{s+1} \tag{13.30}$$

We assume that the mass matrix remains constant with respect to time, but stiffness matrix may change. From Eq. (13.21), it follows that

$$\Delta t_{s+1}\beta[M]\{\dot{U}\}_{s+1} + \Delta t_{s+1}(1-\beta)[M]\{\dot{U}\}_s = [M]\left(\{U\}_{s+1} - \{U\}_s\right) \tag{13.31}$$

Eliminating the mass matrix and time derivative on the left-hand side with the help of Eqs (13.29) and (13.30) and dropping the subscript with Δt (for brevity or considering constant time steps), we obtain

$$\Delta t\beta\left(\{F\}_{s+1} - [K]_{s+1}\{U\}_{s+1}\right) + \Delta t(1-\beta)\left(\{F\}_s - [K]_s\{U\}_s\right) = [M]\left(\{U\}_{s+1} - \{U\}_s\right) \tag{13.32}$$

Rearranging the terms in Eq. (13.32), we get

$$\left[[M] + \beta\Delta t[K]_{s+1}\right]\{U\}_{s+1} = \left[[M] - (1-\beta)\Delta t[K]_s\right]\{U\}_s + \Delta t\left[\beta\{F\}_{s+1} + (1-\beta)\{F\}_s\right] \tag{13.33}$$

For $\beta = 0$, the left-hand side becomes $[M]\{U\}_{s+1}$ and the right-hand side requires information only at the sth time step. In this case, if the mass matrix is diagonal, Eq. (13.33) becomes explicit and we can solve for $\{U\}_{s+1}$ directly without inverting. It is for this reason that the lumped mass matrix is sometimes preferred in finite element formulation to consistent mass matrix.

For all numerical schemes, in which $\beta < 1/2$, the β-family of approximation is stable only if the time step satisfies the following (stability) condition [Reddy (2005)]:

$$\Delta t < \Delta t_{cr} \equiv \frac{2}{(1-2\beta)\lambda} \tag{13.34}$$

where λ is the largest eigenvalue of the finite element equations.

13.4 FORCED VIBRATION PROBLEMS

If the frequency of excitations applied to a structure is less than roughly one-third of the structure's lowest natural frequencies of vibration, then the problem may be considered *quasi-static* and the

effect of inertia may be neglected [Cook, et al. (1989)]. If the frequency of excitation is greater than this value, the problem has to be treated as a dynamic problem. There are two types of dynamic problems; they involve either wave propagations or structural dynamics. In wave propagation problems, the frequency of excitation is often much higher compared to that in structural dynamics problems. Time history can be obtained by using model methods and direct integration method. The finite element formulation provides the following equations:

$$[M]\left\{{}^{t+\Delta t}\ddot{U}\right\}+[C]\left\{{}^{t+\Delta t}\dot{U}\right\}+[K]\left\{{}^{t+\Delta t}U\right\}=\left\{{}^{t+\Delta t}R\right\} \tag{13.35}$$

where the left superscripts indicate time.

Equation (13.35) can be directly indicated by *Newmark method* [Bathe (1996)]. The Newmark method is an extension of the linear acceleration method. The following assumptions are used in this method:

$$\left\{{}^{t+\Delta t}\dot{U}\right\}=\left\{{}^{t}\dot{U}\right\}+\left[(1-\delta)\left\{{}^{t}\ddot{U}\right\}+\delta\left\{{}^{t+\Delta t}\ddot{U}\right\}\right]\Delta t \tag{13.36}$$

$$\left\{{}^{t+\Delta t}U\right\}=\left\{{}^{t}U\right\}+\left\{{}^{t}\dot{U}\right\}\Delta t+\left[\left(\frac{1}{2}-\beta\right)\left\{{}^{t}\ddot{U}\right\}+\beta\left\{{}^{t+\Delta t}\ddot{U}\right\}\right]\Delta t^2 \tag{13.37}$$

where β and δ are parameters that can be determined to obtain integration accuracy and stability. When $\delta = 1/2$, $\beta = 1/6$ it corresponds to linear acceleration method. For constant average acceleration, $\delta = 1/2$, $\beta = 1/4$. The average acceleration is $\frac{1}{2}\left(\left\{{}^{t}\ddot{U}\right\}+\left\{{}^{t+\Delta t}\ddot{U}\right\}\right)$. It can be shown that for $\delta \geq 1/2$ and $\beta \geq 1/4$, the method is unconditionally stable and the method is unstable for $\delta < 1/2$.

The step-by-step procedure used to obtain the solution in Newmark integration method is described as follows:

Step 1 Form stiffness matrix $[K]$, mass matrix $[M]$, and damping matrix $[C]$. Start from time $t = 0$.

Step 2 Initialize $\left\{{}^{0}U\right\}, \left\{{}^{0}\dot{U}\right\}$, and $\left\{{}^{0}\ddot{U}\right\}$.

Step 3 Select time step Δt and parameters β and δ and calculate integration constants. For stability, consider the following values:

$$\delta \geq 0.50; \quad \beta \geq 0.25(0.5+\delta)^2 \tag{13.38}$$

The integration constants are given by

$$\begin{aligned} &a_0=\frac{1}{\beta\Delta t^2} \quad a_1=\frac{\delta}{\beta\Delta t} \quad a_2=\frac{1}{\beta\Delta t} \quad a_3=\frac{1}{2\beta}-1 \\ &a_4=\frac{\delta}{\beta}-1 \quad a_5=\frac{\Delta t}{2}\left(\frac{\delta}{\beta}-2\right) \quad a_6=\Delta t(1-\delta) \quad a_7=\delta\Delta t \end{aligned} \tag{13.39}$$

Step 4 Form effective stiffness matrix

$$\left[\hat{K}\right]=[K]+a_0[M]+a_1[C] \tag{13.40}$$

Step 5 Calculate effective loads at time $t + \Delta t$

$$\left\{{}^{t+\Delta t}\hat{R}\right\} = \left\{{}^{t+\Delta t}\mathrm{R}\right\} + [M]\left(a_0\left\{{}^{t}U\right\} + a_2\left\{{}^{t}\dot{U}\right\} + a_3\left\{{}^{t}\ddot{U}\right\}\right) + [C]\left(a_1\left\{{}^{t}U\right\} + a_4\left\{{}^{t}\dot{U}\right\} + a_5\left\{{}^{t}\ddot{U}\right\}\right) \quad (13.41)$$

Step 6 Solve for displacements at time $t + \Delta t$

$$\left[\hat{K}\right]\left\{{}^{t+\Delta t}U\right\} = \left\{{}^{t+\Delta t}\hat{R}\right\} \quad (13.42)$$

Step 7 Calculate accelerations and velocities at time $t + \Delta t$

$$\left\{{}^{t+\Delta t}\ddot{U}\right\} = a_0\left(\left\{{}^{t+\Delta t}U\right\} - \left\{{}^{t}U\right\}\right) - a_2\left\{{}^{t}\dot{U}\right\} - a_3\left\{{}^{t}\ddot{U}\right\} \quad (13.43)$$

$$\left\{{}^{t+\Delta t}\dot{U}\right\} = \left\{{}^{t}\dot{U}\right\} + a_6\left\{{}^{t}\ddot{U}\right\} + a_7\left\{{}^{t+\Delta t}\ddot{U}\right\} \quad (13.44)$$

Step 8 Increment time by Δt, i.e., replace t by $t + \Delta t$ and go to Step 5, if mass, stiffness, and damping matrices are constant. If any of these matrices depends on time, then go to Step 4.

13.5 CONCLUSION

In this chapter, we have discussed the solution of time-dependent problems using direct integration procedure. These procedures use finite difference method along with FEM. We have discussed only some representative methods. For more details, the reader can refer the books listed as References.

SOLVED PROBLEMS

13.1 A 2-m long metal rod is initially at a temperature of 0°C. Its one end is then attached to a very large heat source at 100°C. You may consider that the temperature of the end suddenly rises to 100°C. Take the thermal diffusivity of the metal as 10^{-4} m^2/s. Taking two finite elements and time step as 1000 s, calculate the temperature of the rod at the middle of its length at three time steps. Use Crank–Nicholson scheme.

Solution Consider Eq. (13.33) for parabolic equations

$$\left[[M] + \beta\Delta t[K]_{s+1}\right]\{U\}_{s+1} = \left[[M] - (1-\beta)\Delta t[K]_s\right]\{U\}_s + \Delta t\left[\beta\{F\}_{s+1} + (1-\beta)\{F\}_s\right] \quad (13.1.1)$$

For Crank–Nicholson scheme, $\beta = 0.5$. The differential equation governing the present problem is written as

$$\frac{\partial^2 T}{\partial x^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \quad (13.1.2)$$

We take two elements of equal size with h = 1 m. The shape functions (in terms of local coordinates) in an element are given by

$$N_1 = 1 - \frac{x}{h}, \quad N_2 = \frac{x}{h} \quad (13.1.3)$$

From Eq. (13.17), we get the elemental coefficient matrix associated with the nodal time derivative of temperatures as

$$[m]=\frac{1}{6\alpha}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}=\frac{10^4}{6}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} \tag{13.1.4}$$

Similarly, the coefficient matrix associated with the nodal temperatures is given by

$$[k]=\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix} \tag{13.1.5}$$

The assembly of the two elements yields

$$[M]=\frac{10^4}{6}\begin{bmatrix}2 & 1 & 0\\ 1 & 4 & 1\\ 0 & 1 & 2\end{bmatrix},\quad [K]=\begin{bmatrix}1 & -1 & 0\\ 1 & 2 & 1\\ 0 & 1 & 1\end{bmatrix} \tag{13.1.6}$$

As there is no internal heat generation, the assembled vector $\{F\}$ will have all elements as zero except the first and last term, which are denoted by f_1 and f_3 here. Substituting the relevant values in Eq. (13.1.1), we get

$$\left[\frac{10^4}{6}\begin{bmatrix}2 & 1 & 0\\ 1 & 4 & 1\\ 0 & 1 & 2\end{bmatrix}+500\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & 1\\ 0 & 1 & 1\end{bmatrix}\right]\begin{Bmatrix}T_1\\ T_2\\ T_3\end{Bmatrix}_{s+1}$$

$$=\left[\frac{10^4}{6}\begin{bmatrix}2 & 1 & 0\\ 1 & 4 & 1\\ 0 & 1 & 2\end{bmatrix}-500\begin{bmatrix}1 & -1 & 0\\ -1 & 2 & -1\\ 0 & -1 & 1\end{bmatrix}\right]\begin{Bmatrix}T_1\\ T_2\\ T_3\end{Bmatrix}_s+500\begin{Bmatrix}(f_1)_{s+1}+(f_1)_s\\ 0\\ (f_2)_{s+1}+(f_2)_s\end{Bmatrix}$$

The boundary conditions are T_1 = 100°C and T_3 = 0°C. Thus, first and third equations can be eliminated. The second equation can be written as

$$\frac{10000}{6}T_1+\frac{20000}{3}(T_2)_{s+1}+\frac{10000}{6}T_3-500T_1+1000(T_2)_{s+1}-500T_3$$

$$=\begin{cases}\dfrac{10000}{6}T_1+\dfrac{20000}{3}(T_2)_s+\dfrac{10000}{6}T_3\\ +500T_1-1000(T_2)_s+500T_3\end{cases} \tag{13.1.8}$$

Note that subscripts with T_2 indicate time steps. No subscripts have been used with T_1 and T_3 as these are constants with respect to time. Substituting the values of T_1 and T_3 and simplifying Eq. (13.1.8), we obtain

$$\left(\frac{20000}{3}+1000\right)(T_2)_{s+1}-\left(\frac{20000}{3}-1000\right)(T_2)_s+100000 \tag{13.1.9}$$

or

$$(T_2)_{s+1}=13.0435+0.7391(T_2)_s \tag{13.1.10}$$

Given $(T_2)_0 = 0°C$, from Eq. (13.1.10), we obtain

$$(T_2)_1 = 13.045°C \tag{13.1.11a}$$

$$(T_2)_2 = 22.6866°C \tag{13.1.11b}$$

$$(T_2)_3 = 29.8127°C \tag{13.1.11c}$$

Thus, the temperature keeps increasing. However, it will not cross 50°C. In fact, at steady-state, $(T_2)_{s+1} = (T_2)_s$. Therefore, from Eq. (13.1.10), $(T_2)_s = 49.9943°C$. Here, steady-state temperature comes out to be 49.9943°C instead of 50°C due to round-off errors. It is to be noted that the accuracy of temperature at different time steps can be improved by reducing the time-step size. However, no matter what time-step size is used, the method will be stable. Reader may try to solve the problem with different time-step sizes.

13.2 Prove that for $\delta = 1/2$, $\beta = 1/6$, the Newmark scheme [represented by Eqs (13.36) and (13.37)] corresponds to linear acceleration method.

Solution Denoting the variable time by θ (where $\theta = 0$ at time t), the linear acceleration between time t to $t + \Delta t$ is given by

$$\left\{\frac{\mathrm{d}\dot{U}}{\mathrm{d}t}\right\} = \left\{{}^{t}\ddot{U}\right\} + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\theta \tag{13.2.1}$$

According to Eq. (13.2.1), acceleration is proportional to θ. Its values at time t and $t + \Delta t$ are ${}^{t}\ddot{U}$ and ${}^{t+\Delta t}\ddot{U}$, respectively. Integrating Eq. (13.2.1), we obtain

$$\left\{\dot{U}\right\} = \left\{{}^{t}\ddot{U}\right\}\theta + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\theta^2}{2} + \{C\} \tag{13.2.2}$$

where $\{C\}$ is the vector of constants, which is determined from the condition that at $\theta = 0$

$$\left\{\dot{U}\right\} = \left\{{}^{t}\dot{U}\right\} \tag{13.2.3}$$

Thus

$$\{C\} = \left\{{}^{t}\dot{U}\right\} \tag{13.2.4}$$

Substituting the value of $\{C\}$ in Eq. (13.2.2), we obtain

$$\left\{\dot{U}\right\} = \left\{{}^{t}\ddot{U}\right\}\theta + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\theta^2}{2} + \left\{{}^{t}\dot{U}\right\} \tag{13.2.5}$$

Calculating the velocity at $\theta = \Delta t$, which corresponds to time $t + \Delta t$, we get

$$\left\{{}^{t+\Delta t}\dot{U}\right\} = \left\{{}^{t}\ddot{U}\right\}\Delta t + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\Delta t^2}{2} + \left\{{}^{t}\dot{U}\right\} = \left\{{}^{t}\dot{U}\right\} + \frac{1}{2}\left(\left\{{}^{t+\Delta t}\ddot{U}\right\} + \left\{{}^{t}\ddot{U}\right\}\right) \tag{13.2.6}$$

Equation (13.2.6) can be obtained by substituting $\delta = 1/2$ in Eq. (13.36).

Now, integrating Eq. (13.2.2) with respect to time θ, we get

$$\{U\} = \left\{{}^{t}\ddot{U}\right\}\frac{\theta^2}{2} + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\theta^3}{6} + \left\{{}^{t}\dot{U}\right\}\theta + \{D\} \tag{13.2.7}$$

where $\{D\}$ is the vector of constants, which can be obtained from the condition that at $\theta = 0$

$$\{U\} = \left\{{}^{t}U\right\} \tag{13.2.8}$$

Thus, we can write

$$\{D\} = \left\{{}^{t}U\right\} \tag{13.2.9}$$

Therefore, Eq. (13.2.7) can be written as

$$\{U\} = \left\{{}^{t}\ddot{U}\right\}\frac{\theta^2}{2} + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\theta^3}{6} + \left\{{}^{t}\dot{U}\right\}\theta + \{{}^{t}U\} \tag{13.2.10}$$

Calculating the displacement at $\theta = \Delta t$ from Eq. (13.2.10), we obtain

$$\begin{aligned}\left\{{}^{t+\Delta t}U\right\} &= \left\{{}^{t}\ddot{U}\right\}\frac{\Delta t^2}{2} + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\} - \left\{{}^{t}\ddot{U}\right\}}{\Delta t}\frac{\Delta t^3}{6} + \left\{{}^{t}\dot{U}\right\}\Delta t + \{{}^{t}U\} = \\ &\{{}^{t}U\} + \left\{{}^{t}\dot{U}\right\}\Delta t + \left[\frac{\left\{{}^{\Delta t}\ddot{U}\right\}}{3} + \frac{\left\{{}^{t+\Delta t}\ddot{U}\right\}}{6}\right]\Delta t^2\end{aligned} \tag{13.2.11}$$

Equation (13.2.11) can be obtained by putting $\beta = 1/6$ in Eq. (13.37).

Thus, it is proved that for $\delta = 1/2$, $\beta = 1/6$, Newmark scheme represented by Eqs (13.36) and (13.37) corresponds to linear acceleration method.

EXERCISES

A. Multiple Choice Questions

13.1 Which of the following method is explicit?

(a) $\{U\}_{n+1} = \{U\}_n + \Delta t\{\dot{U}\}_n$ (b) $\{U\}_{n+1} = \{U\}_n + \Delta t\{\dot{U}\}_{n+1}$

(c) $\{U\}_{n+1} = \{U\}_n + \Delta t\left[\{\dot{U}\}_{n+1} + \{\dot{U}\}_n\right]/2$ (d) $\{U\}_n = \{U\}_{n+1} - \Delta t\{\dot{U}\}_{n+1}$

13.2 The order of accuracy of Crank–Nicholson scheme is

(a) $O(\Delta t)$ (b) $O(\Delta t^2)$

(c) $O(\Delta t^3)$ (d) $O(\Delta t^4)$

13.3 The Newmark method uses the following equations:

$$\left\{{}^{t+\Delta t}\dot{U}\right\} = \left\{{}^{t}\dot{U}\right\} + \left[(1-\delta)\left\{{}^{t}\ddot{U}\right\} + \delta\left\{{}^{t+\Delta t}\ddot{U}\right\}\right]\Delta t \tag{13.3.1}$$

$$\left\{{}^{t+\Delta t}U\right\} = \left\{{}^{t}U\right\} + \left\{{}^{t}\dot{U}\right\}\Delta t + \left[\left(\frac{1}{2} - \beta\right)\left\{{}^{t}\ddot{U}\right\} + \beta\left\{{}^{t+\Delta t}\ddot{U}\right\}\right]\Delta t^2 \tag{13.3.2}$$

The method corresponds to average acceleration method for

(a) $\delta = 1/2, \beta = 1/6$ (b) $\delta = 1/2, \beta = 1/2$

(c) $\delta = 1/2, \beta = 1/4$ (d) $\delta = 1/2, \beta = 0$

13.4 Which of the following method for solving first-order differential equation numerically is only conditionally stable?

(a) Backward Euler method (b) Forward Euler method

(c) Crank-Nicholson method (d) Galerkin method

13.5 The Newmark method is unstable for any value of Δt if

(a) $\delta = 1/2, \beta = 1/6$ (b) $\delta = 1/2, \beta = 0$

(c) $\delta = 1/2, \beta = 1/4$ (d) $\delta = \beta = 0$

B. Problems

13.6 A rod is pushed by a force F on a smooth surface (Fig. 13.1). The rigid body displacement of the rod is denoted by a, which is the displacement of node 1. The displacement of the particles relative to node 1 is denoted by u, which is a function of x. The governing equations of motion are written as

$$EA\frac{\partial^2 u}{\partial x^2} - m\frac{\partial^2 (a+u)}{\partial t^2} = 0$$

$$F = \int_0^l m\frac{\partial^2 (a+u)}{\partial t^2}\,\mathrm{d}x$$

where E is Young's modulus of elasticity, A is the area of cross-section, and m is the mass per unit length.

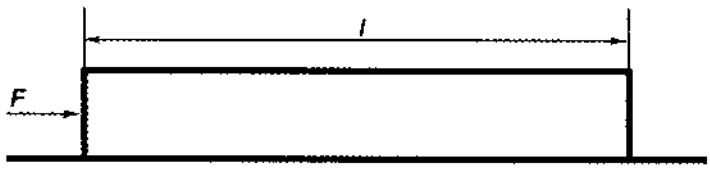


Figure 13.1 *A rod of length l subjected to a force F*

The one-element finite element formulation of this problem leads to

$$[M]\begin{Bmatrix}\ddot{a}\\ \ddot{u}_1\\ \ddot{u}_2\end{Bmatrix} + [K]\begin{Bmatrix}a\\ u_1\\ u_2\end{Bmatrix} = \begin{Bmatrix}F\\ -EA\left.\dfrac{\mathrm{d}u}{\mathrm{d}x}\right|_1\\ 0\end{Bmatrix}$$

Solve this problem on computer by using suitable values of various parameters.

13.7 Show that an explicit problem is only conditionally stable.

13.8 A copper cylinder of diameter 10 cm and length 5 m is heated at one end by a flame of temperature 700°C. The ambient temperature is 20°C. Find out the temperature-time history using FEM. Obtain the required data from standard books on heat transfer.

13.9 Prove that for all numerical schemes, in which $\alpha < 1/2$, the α-family of approximation is stable only if the time step satisfies the following (stability) condition:

$$\Delta t < \Delta t_{cr} \equiv \frac{2}{(1-2\alpha)\lambda}$$

REFERENCES

Bathe, KJ, 1996, *Finite Element Procedures*, Prentice-Hall of India, New Delhi.

Cook, RD, Malkus, DS, and Plesha, ME, 1989, *Concept and Applications of Finite Element Analysis, Third edition*, John Wiley & Sons, Singapore.

Reddy, JN, 2005, *An Introduction to the Finite Element Method, Third edition*, Tata McGraw-Hill, New Delhi.

14
CHAPTER

FEM Formulation for Plate Problem

14.1 Introduction

14.2 Thin-Plate Formulation

14.3 Various Thin-Plate Elements

14.4 Thick-Plate Formulation

14.5 Locking Phenomenon

14.6 An Example

14.7 Conclusion

14.1 INTRODUCTION

A flat plate supports a transverse load applied on it by bending. In the classical theory of plates, certain assumptions are made initially to simplify problems involving plates to two dimensions. Such assumptions concern the linear variation of strains and stresses on lines that are normal to the plane of the plate. The state of deformation of the plate can be described entirely by one quantity; this quantity is defined as the lateral displacement, w, of the 'middle plane' of the plate.

In classical Kirchhoff thin-plate theory, it is assumed that the normal drawn to the mid-plane remains normal to it even after deformation, which implies that cross-sections do not deform. However, in thick-plate theory, deformation of the cross-section is taken into account.

FEM formulation of thin-plate problems involves solving a fourth-order differential equation. The potential energy functional will contain second derivatives of the lateral displacement. Because of this, continuity conditions between elements have to be imposed not only on lateral displacement but also on its first derivatives. This is to ensure that the plate remains continuous and does not 'kink'. If 'kinking' occurs, the second derivative or curvature becomes infinite causing certain infinite terms to appear in the energy expression. Because of the continuity requirements, determination of suitable shape functions for thin plate problems is an extremely complicated task.

The potential energy functional of a thick-plate problem contains only first derivatives. Here, rotations of the cross-section about two axes cannot be determined from the slopes of lateral displacement function. At each node one has to find the lateral displacement w and rotations θ_x and θ_y. Only these three quantities need to be continuous; their first derivatives, on the other hand, can be discontinuous. Thus, in thick-plate problems, same shape functions as used in the plane stress and plane strain problems can be used. However, solving a thin-plate problem using thick-plate formulation poses numerical difficulties. Both the formulations are discussed in this chapter.

14.2 THIN-PLATE FORMULATION

Thin-plate formulation is based on classical Kirchhoff's theory. The following are the main assumptions in this formulation:

- The points on the mid-surface, $z = 0$, move only in the z direction as the plate deforms in bending.
- A line that is straight and normal to the mid-surface before loading is assumed to remain straight and normal to the mid-surface after loading (see line OP in Fig. 14.1). Thus, transverse shear deformation is assumed to be zero. A point not on the mid-surface has displacement components u and v in the x and y directions, respectively. The projection in x–z plane is shown in Fig. 14.1; since w, x and w, y are small, one can easily visualize that

$$u = -zw_{,x} \qquad v = -zw_{,y} \tag{14.1}$$

Hence, the strain–displacement relations are given by

$$\varepsilon_x = u,x = -zw_{,xx}$$

$$\varepsilon_y = v_{,y} = -zw_{,yy}$$

$$\gamma_{xy} = u_{,y} + v_{,x} = -2zw_{,xy} \tag{14.2}$$

It can be easily shown [Cook, et al. (1989)] that for a plate of thickness t with zero initial strain

$$\{M\} = -[D_k]\{\kappa\} \tag{14.3}$$

where

$$\{M\} = [M_x \quad M_y \quad M_{xy}]^{\mathrm{T}} \quad \text{and} \quad \{\kappa\} = [w_{,xx} \quad w_{,yy} \quad 2w_{,xy}]^{\mathrm{T}} \tag{14.4}$$

$$M_x = \int_{-t/2}^{t/2} \sigma_x z \mathrm{d}z \quad M_y = \int_{-t/2}^{t/2} \sigma_y z \mathrm{d}z \quad M_{xy} = \int_{t/2}^{t/2} \sigma_{xy} z \mathrm{d}z \tag{14.5}$$

where σ_x and σ_y are direct stresses and σ_{xy} is the shear stress. Note that M_x and M_y are the bending moments per unit length and M_{xy} is the twisting moment per unit length. For an isotropic plate of thickness t, Young's modulus of elasticity E, and Poisson's ratio ν, we can write

$$[D_k] = \begin{bmatrix} D & \nu D & 0 \\ \nu D & D & 0 \\ 0 & 0 & (1-\nu)D/2 \end{bmatrix} \tag{14.6}$$

where the flexural rigidity of the plate, D, is given by

$$D = \frac{Et^3}{12(1-\nu^2)} \tag{14.7}$$

The strain energy expression is written as

$$U = \int_A \frac{1}{2}\{\kappa\}^{\mathrm{T}}[D_k]\{\kappa\}\,\mathrm{d}A \tag{14.8}$$

Now

$$w = \lfloor N \rfloor \{d\} \tag{14.9}$$

where $\{d\}$ is the column vector containing nodal degrees of freedom in the element and $\lfloor N \rfloor$ is the row vector containing shape functions. Note that nodal degrees of freedom include nodal transverse displacements and the derivatives of transverse displacement. Equation (14.9) has been written in a general way and is applicable for any element. Let us consider a typical element having n degrees of freedom. For this element, Eq. (14.9) is written as

$$w = \lfloor N_1 \quad N_2 \ldots \ldots \quad N_n \rfloor \begin{Bmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_n \end{Bmatrix} \tag{14.10}$$

From Eq. (14.10), the vector $\{\kappa\}$ containing the second derivatives of the transverse displacement can be calculated as

$$\{\kappa\} = \begin{Bmatrix} w_{,xx} \\ w_{,yy} \\ 2w_{,xy} \end{Bmatrix} = \begin{bmatrix} N_{1,xx} & N_{2,xx} & \cdots & \cdots & N_{n,xx} \\ N_{1,yy} & N_{2,yy} & \cdots & \cdots & N_{n,yy} \\ 2N_{1,xy} & 2N_{2,xy} & \cdots & \cdots & 2N_{n,xy} \end{bmatrix} \begin{Bmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_n \end{Bmatrix} \tag{14.11}$$

In the condensed form, the relation given in Eq. (14.11) can be written as

$$\{\kappa\} = [B]\{d\} \tag{14.12}$$

where $[B]$ is the matrix containing the second derivatives of shape functions.

The reader is already familiar with Ritz FEM procedure. In Ritz procedure for linear FEM problems, we often obtain an expression containing quadratic and linear terms, the minimization of which leads to FEM equations. For example, in solid mechanics problems, the potential energy expression is minimized. For minimizing an expression, the first derivatives with respect to nodal degrees of freedom are made zero. It can be observed that the quadratic part of the expression provides the stiffness matrix, while linear part provides the right-hand-side vector. The quadratic part in a potential energy expression comes from the strain energy expression. Thus, differentiation of strain energy expression with respect to nodal degrees of freedom vector provides the stiffness matrix. Substituting Eq. (14.12) in Eq. (14.8), the expression for strain energy of a plate element is obtained as

$$U = \frac{1}{2}\lfloor d \rfloor \int_A [B]^{\mathrm{T}} [D_k] [B] \, \mathrm{d}A \{d\} \tag{14.13}$$

Note that the nodal variables are not functions of x and y, and they can be taken outside the integral sign. As $[D_k]$ is a symmetric matrix, the expression in the integral is also symmetric. Hence, by Eq. (A.2.5) of the Appendix A.2, we can write

$$\frac{\partial U}{\partial \{d\}} = \int_A [B]^{\mathrm{T}} [D_k] [B] \, \mathrm{d}A \{d\} = [k]\{d\} \tag{14.14}$$

Thus, the stiffness matrix is given as

$$[k] = \int_A [B]^{\mathrm{T}} [D_k] [B] \; \mathrm{d}A \tag{14.15}$$

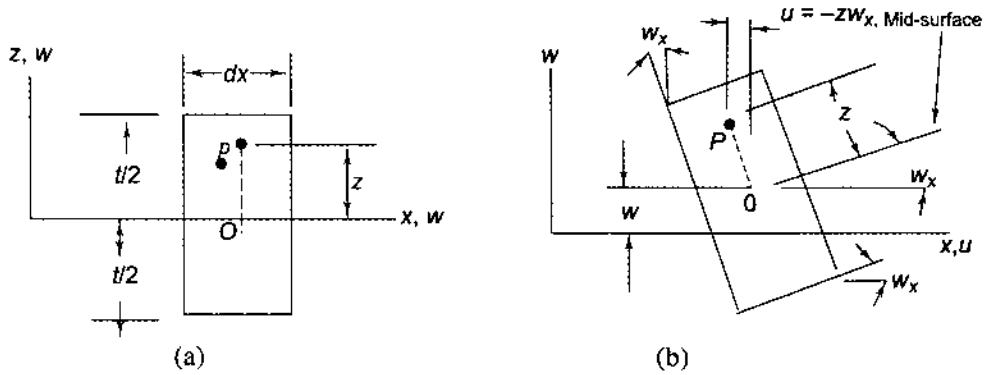


Figure 14.1 *(a) Differential elements of thin plate before loading (b) after loading*

The potential energy due to distributed external load, i.e., due to external pressure is written as

$$V_q = -\int_{A_q} q(x, y)\, w \, \mathrm{d}x\mathrm{d}y \tag{14.16}$$

where A_q is the area on which the external pressure acts. Referring to Eq. (14.9), the transverse deflection w is interpolated inside an element as

$$w = \lfloor N \rfloor \{d\} = \lfloor d \rfloor \{N\} \tag{14.17}$$

Substituting Eq. (14.17) into Eq. (14.16), we obtain

$$V_q = -\lfloor d \rfloor \int_{A_q} q(x, y)\{N\}\mathrm{d}x\mathrm{d}y \tag{14.18}$$

The right-hand-side vector is obtained as

$$\{R\} = -\frac{\partial V_q}{\partial \{d\}} = \frac{\partial\left[\lfloor d \rfloor \int_{A_q} q(x, y)\{N\}\mathrm{d}x\mathrm{d}y\right]}{\partial \{d\}} = \int_{A_q} q(x, y)\{N\}\mathrm{d}x\mathrm{d}y \tag{14.19}$$

using Eqs (A.2.2) and (A.2.3) of Appendix A.2. If there is a concentrated load P at location (x_0, y_0) within A_q, it can be converted into distributed load using a Dirac-delta function defined in two dimensions. Thus, the right-hand-side vector for the concentrated load will be written as

$$\{R\} = \int_A P\delta(x - x_0, y - y_0)\{N\}\mathrm{d}x\mathrm{d}y = P\{N\}_{(x_0, y_0)} \tag{14.20}$$

In Eq. (14.20), we have made use of the following property of Dirac-delta function:

$$\int_A f(x, y)\delta(x - x_0, y - y_0)\mathrm{d}x\mathrm{d}y = f(x_0, y_0) \tag{14.21}$$

provided point (x_0, y_0) is contained in A.

14.3 VARIOUS THIN-PLATE ELEMENTS

As discussed earlier, thin-plate formulation requires both w and its normal slope across an interface to be continuous. This condition increases mathematical and computational difficulties of the formulation. It is, however, relatively simple to obtain shape functions which, while preserving the continuity of w, may violate its slope continuity between elements, even though they satisfy the slope continuity at the node where such continuity is imposed. If these chosen functions can represent 'constant strain' state and, in addition, pass the patch test [Zienkiwicz (1993)], then convergence can still be obtained. Such shape functions are called *non-conforming shape functions* and the associated elements are called *non-conforming elements*. The shape functions satisfying all continuity requirements are called *conforming shape functions* and the associated elements are called *conforming elements*. It has been

found that on many occasions, conforming elements yield lower accuracy compared to non-conforming elements. Subsections 14.3.2–14.3.5 describe the commonly used thin-plate elements. However, before that, we briefly describe the patch test.

14.3.1 Patch Test

In some cases, considerable difficulty is experienced in finding primary variable functions for an element which will automatically be continuous along the whole interface between adjacent elements. The discontinuity of displacement will cause infinite strains at the interfaces. However, if in the limit, as the size of the subdivision decreases continuity is restored, then the formulation will tend to the correct answer. To test whether a non-conforming element will provide satisfactory results, patch test may be conducted.

A set of assembled elements is called a *patch*. Usually, a patch is designed in such a manner so that at least one node of the patch is shared by more than one element. This ensures the testing of assembly routine. The patch is constrained at some nodes to avoid any rigid body motion and the loads are applied at the other nodes, which are expected to provide constant stress or constant strain. In the plate bending problems, three different tests can be conducted.

- In the first test, the loads are applied such that the constant bending moment M_x is obtained.
- In the second test, the loads are applied such that the constant bending moment M_y is obtained.
- In the third test, the loads are applied such that the constant twisting moment M_{xy} is obtained.

The element is said to have passed the patch test if the computer code predicts the expected constant stress, strain, or bending/twisting moments accurately except for round-off errors. Sometimes, a patch may not provide accurate results as such but the accuracy of results can be improved by refinement. This is called a *weak patch test*.

A patch test can also be used to check the stability of the elements. In structural problems, an unstable state provides zero strain energy for a displacement vector that is not a rigid-body motion. Instabilities can be detected by an *eigenvalue test*. In this test, eigenvalues of the stiffness matrix are found. One should obtain zero eigenvalues corresponding to rigid-body modes. Any extra zero eigenvalue indicates instability.

Essentially, patch test is a numerical test for assessing the convergence of a finite element formulation. This test is simpler to conduct compared to other tests. Errors in programming may also get detected with the help of this test. It can be applied to any finite element problem and is not restricted to structural problems. The basic concept is to test the ability of the FEM to provide exactly converged problems to some benchmark problems with known solution.

14.3.2 Rectangular Elements with Corner Nodes

Consider the rectangular element shown in Fig. 14.2; the element has a total of twelve degrees of freedom. Degrees of freedom $w_{,x3}$ and $w_{,y3}$ are slopes of the mid-surface of the plate at node 3. Lateral displacement w of this element has the form [Gallagher (1975)]

$$w = \begin{bmatrix} 1 & x & y & x^2 & xy & y^2 & x^3 & x^2y & xy^2 & y^3 & x^3y & xy^3 \end{bmatrix} \{a\} \tag{14.22}$$

where vector $\{a\}$ contains twelve coefficients, which must be exchanged for the twelve nodal degrees of freedom $\{d\}$ by the usual process. If the degrees of freedom at each nodes are given by w, $\partial w/\partial x$, and $\partial w/\partial y$, then shape functions at the ith nodes are written as

$$\frac{1}{2}\left[(\xi_0+1)(\eta_0+1)(2+\xi_0+\eta_0-\xi^2-\eta^2), b\eta_i(\xi_0+1)(\eta_0+1)^2(1-\eta_0), a\xi_i(\xi_0+1)^2(\eta_0+1)(\xi_0-1)\right] \tag{14.23}$$

with

$$\xi = (x-x_c)/a \qquad \eta = (y-y_c)/b \tag{14.24}$$

$$\xi_0 = \xi\xi_i \qquad \eta_0 = \eta\eta_i$$

where x_c and y_c are the coordinates of the centroid.

It can be shown that this element is non-conforming. In fact, it is impossible to specify simple polynomial expressions for shape functions ensuring full compatibility when only w and its slopes are prescribed at nodes [Irons and Draper (1965)]. However, this element does permit a state of constant strain (curvature) to exist.

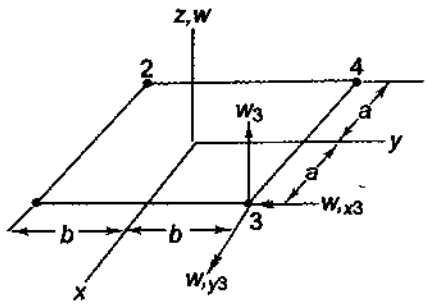


Figure 14.2 *Rectangular element with corner nodes*

14.3.3 Triangular Elements with Corner Nodes

Figure 14.3 shows a triangular node with nine degrees of freedom per node. A nine-term field for w is appropriate for this element. However, a complete cubic contains ten terms. The possible nine-term fields include

$$w = \begin{bmatrix} 1 & x & y & x^2 & y^2 & x^3 & x^2 & x^2y & xy^2 & y^3 \end{bmatrix} \{a\} \tag{14.25}$$

$$w = \begin{bmatrix} 1 & x & y & x^2 & y^2 & xy & y^2 & x^3 & x^2y+xy^2 & y^3 \end{bmatrix} \{a\} \tag{14.26}$$

Equation (14.25) does not contain the term xy. The resulting element is therefore incapable of passing a constant-twist patch test, and is considered unacceptable. Equation (14.26) leads to an

element that lacks geometric isotropy and has poor convergence properties. For certain orientations of the sides of a triangle, the shape functions cannot be obtained since the corresponding transformation matrix becomes singular. An alternative is to add a central node to the formulation and eliminate it by a procedure called *static condensation*. It is briefly explained as follows.

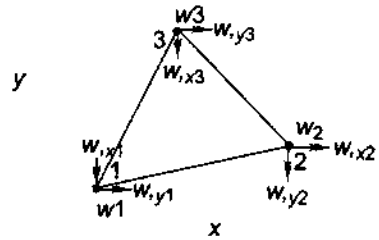


Figure 14.3 *A triangular element with corner nodes*

Suppose the central node is the tenth node and the degree of freedom associated with this node is w_{10}. The vector containing the degrees of freedom associated with other nodes is given by

$$\{d_r\} = \begin{Bmatrix} w_1 \\ \left(\dfrac{\partial w}{\partial x}\right)_1 \\ \left(\dfrac{\partial w}{\partial y}\right)_1 \\ w_2 \\ \left(\dfrac{\partial w}{\partial x}\right)_2 \\ \left(\dfrac{\partial w}{\partial y}\right)_2 \\ w_3 \\ \left(\dfrac{\partial w}{\partial x}\right)_3 \\ \left(\dfrac{\partial w}{\partial y}\right)_3 \end{Bmatrix} \tag{14.27}$$

The final stiffness matrix will be a 10×10 matrix and the load vector will be a vector of size 10. Let us denote the tenth diagonal element of the stiffness matrix by x and the tenth element of the load vector by y. The finite element system of equations of the element can be expressed in the partitioned form as

$$\begin{bmatrix} [A] & \{B\} \\ [C] & x \end{bmatrix} \begin{Bmatrix} \{d_r\} \\ w_{10} \end{Bmatrix} = \begin{Bmatrix} \{R\}_r \\ y \end{Bmatrix} \tag{14.28}$$

where $[A]$ is of size 9×9, $[B]$ is of size 9×1, and $[C]$ is of size 1×9. The right-hand-side vector includes the contribution from external as well as internal loads. However, y is due to external loads only, because corresponding to internal nodes we do not get internal load terms. The mathematical explanation for it is as follows. *In Galerkin method, when integration by part is carried out to reduce the order of derivative, certain boundary terms are obtained. These terms contribute to internal loads at boundary nodes. In Ritz method, while writing the total potential energy terms, we get the terms corresponding to internal loads at the boundary.* The terms due to internal loads get eliminated during assembly except at the boundary of total domain. The internal load terms occurring at boundary of the total domain are tackled by applying the boundary conditions. Equation (14.28) can be written as

$$[A]\{d_r\} + [B]w_{10} = \{R\}_r \tag{14.29a}$$

$$[C]\{d\}_r + x\,w_{10} = y \tag{14.29b}$$

Equation (14.29b) provides

$$w_{10} = \frac{y - [C]\{d\}_r}{x} \tag{14.30}$$

Substituting Eq. (14.30) in Eq. (14.29a), we obtain

$$[A]\{d_r\} + [B]\frac{y - [C]\{d\}_r}{x} = \{R\}_r \tag{14.31}$$

which after simplification becomes

$$\left[[A] - \frac{1}{x}[B][C]\right]\{d_r\} = \{R\}_r - \frac{y}{x}[B] \tag{14.32}$$

Thus, ultimately the degree of freedom associated with central node gets eliminated. This is called *static condensation*.

This procedure would allow a complete cubic to be used but it was found that an element derived on this basis does not pass the patch test [Hrabok and Hrudey (1984)]. The element is too flexible and may sometimes converge to incorrect results.

14.3.4 Quadrilateral and Parallelogram Elements

Henshall, et al. (1972) studied the performance of a non-conforming quadrilateral element and concluded that reasonable accuracy is attainable. Their paper provides detailed information of transformations required for an iso-parametric mapping and the resulting need for numerical integration. Only in the case of a parallelogram, it is possible to achieve constant curvature using natural coordinates. Stiffness matrices of these elements have been worked out by Dawe (1966).

14.3.5 Sixteen-Noded Rectangular Elements

In case of a rectangular element shown in Fig. 14.2, the specification of $\partial^2 w/\partial x \partial y$ as a nodal parameter is always possible as it does not involve 'excessive continuity'. It is easy to show that for

such an element, polynomial shape functions that ensure compatibility can be easily determined. A polynomial expansion involving sixteen constants (equal to the number of nodal parameters) could, for instance, be written retaining terms which do not produce a higher order variation of w or its normal slope than cubic along the sides.

14.4 THICK-PLATE FORMULATION

Thick-plate formulation is based on Mindlin plate theory. The following are the main assumptions of the Mindlin plate theory:

- Stress component along the normal to the mid-plane of the plate is negligible (same as in Kirchhoff plate theory).
- Transverse displacement does not vary along the direction of thickness (same as in Kirchhoff plate theory).
- The normal to the mid-plane remains straight but not necessarily normal to it after deformation.

In view of the Mindlin plate assumptions, for small deformation, we have the following displacement and strain components:

$$
\begin{aligned}
u &= z\theta_y, \qquad v = z\theta_x \\
\varepsilon_x &= z\frac{\partial \theta_y}{\partial x}, \quad \varepsilon_y = z\frac{\partial \theta_x}{\partial y} \\
\gamma_{xy} &= z\left(\frac{\partial \theta_x}{\partial x} + \frac{\partial \theta_y}{\partial y}\right) \\
\gamma_{yz} &= \frac{\partial w}{\partial y} + \theta_x \\
\gamma_{xz} &= \frac{\partial w}{\partial x} + \theta_y
\end{aligned} \tag{14.33}
$$

where θ_x and θ_y are the counter-clockwise rotations about the y and negative x axis, respectively. Since $\sigma_z = 0$, the stress–strain relations for the isotropic linearly elastic material become

$$
\begin{aligned}
\sigma_x &= \frac{E}{1-\nu^2}(\varepsilon_x + \nu\varepsilon_y) \\
\sigma_y &= \frac{E}{1-\nu^2}(\varepsilon_y + \nu\varepsilon_x) \\
\tau_{xy} &= \frac{E}{2(1+\nu)}\gamma_{xy} \\
\tau_{yz} &= \alpha_s \frac{E}{2(1+\nu)}\gamma_{yz}
\end{aligned} \tag{14.34}
$$

$$\tau_{zx} = \alpha_s \frac{E}{2(1+\nu)} \gamma_{zx}$$

where E is Young's modulus, ν is Poisson's ratio, and α_s is the shear correction factor. It is needed because shear stress is not uniform across the thickness.

The strain energy V is given by

$$V = \frac{1}{2} \int_A \int_{-t/2}^{t/2} (\sigma_x \varepsilon_x + \sigma_y \varepsilon_y + \sigma_z \varepsilon_z + \tau_{xy}\gamma_{xy} + \tau_{yz}\gamma_{yz} + \tau_{xz}\gamma_{xz}) \mathrm{d}z \mathrm{d}A \tag{14.35}$$

where t is the thickness of the plate and A is the area of the plate. Substituting the expressions of stresses and strains as given in Eqs (14.33) and (14.34), the expression for V [Eq. (14.35)] becomes

$$V = \frac{1}{2} \int_A \frac{Et}{1-\nu^2} \left[\frac{t^2}{12} \left(\frac{\partial \theta_x}{\partial x} \right)^2 + \frac{\nu t^2}{6} \frac{\partial \theta_y}{\partial y} \frac{\partial \theta_x}{\partial x} + \frac{t^2}{12} \left(\frac{\partial \theta_y}{\partial y} \right)^2 \right] \mathrm{d}A$$

$$+ \frac{1}{2} \int_A \frac{Et}{2(1+\nu)} \left[\frac{t^2}{12} \left(\frac{\partial \theta_x}{\partial y} \right)^2 + \frac{t^2}{12} \left(\frac{\partial \theta y}{\partial y} \right)^2 + \frac{t^2}{6} \frac{\partial \theta_x}{\partial y} \frac{\partial \theta y}{\partial x} \right] \mathrm{d}A$$

$$+ \frac{1}{2} \int_A \frac{Et}{2(1+\nu)} \alpha_s \left[\theta_y^2 + \left(\frac{\partial w}{\partial y} \right)^2 + 2\theta_y \frac{\partial w}{\partial y} + \theta_x^2 + \left(\frac{\partial w}{\partial x} \right)^2 + 2\theta_x \frac{\partial w}{\partial x} \right] \mathrm{d}A \tag{14.36}$$

When the total potential energy term is minimized, the strain energy functional provides the stiffness matrix.

The Mindlin plate element requires C^0 continuity. The types of shape functions used for plane stress and plane strain problems can be used for plates also. For n-noded elements, the primary variables are approximated as

$$w = \sum_{i=1}^{i=n} N_i w_i \tag{14.37}$$

$$\theta_x = \sum_{i=1}^{i=n} N_i \theta_{xi} \tag{14.38}$$

$$\theta_y = \sum_{i=1}^{i=n} N_i \theta_{yi} \tag{14.39}$$

where N_i is Lagrange shape function for i^{th} node.

Following the standard procedure, the elemental stiffness matrix can be obtained as

$$\left[k_{ij}\right] = \int_{A_e} \left[B_{bi}\right]^{\mathrm{T}} \left[D_b\right] \left[B_{bj}\right] dA + \int_{A_e} \left[B_{si}\right]^{\mathrm{T}} \left[D_s\right] \left[B_{sj}\right] \mathrm{d}A \tag{14.40}$$

where

$$[B_{bi}] = \begin{bmatrix} 0 & N_{i,x} & 0 & 0 & 0 \\ 0 & 0 & N_{i,y} & 0 & 0 \\ 0 & N_{i,y} & N_{i,x} & 0 & 0 \end{bmatrix} \tag{14.41}$$

$$[B_{si}] = \begin{bmatrix} N_{i,x} & N_i & 0 & 0 & 0 \\ N_{i,y} & 0 & N_i & 0 & 0 \end{bmatrix} \tag{14.42}$$

$$[D_b] = \frac{Et^3}{12(1-\nu^2)} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{(1-\nu)}{2} \end{bmatrix} \tag{14.43}$$

$$[D_s] = \alpha_s \frac{Et}{2(1+\nu)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{14.44}$$

In the expression given in the right-hand side of Eq. (14.40), the first term indicates stiffness due to bending, while the second term indicates stiffness due to shear.

14.5 LOCKING PHENOMENON

As discussed in the previous section, stiffness matrix is composed of two parts—matrix due to bending and matrix due to shear. In a very thin plate, stiffness matrix due to shear dominates. Hence, the solution gives spurious zero values of bending deformation. This is called *locking of the mesh*. Thus, solving a thin-plate problem by thick-plate formulation may not provide correct results.

One way of reducing the locking phenomenon is to use selective integration scheme. In this scheme, the bending stiffness matrix is computed using full Gauss-quadrature integration, while the shear stiffness matrix is computed using reduced Gauss-quadrature integration. Thus, using a four-noded rectangular element, full (2×2) integration is used for bending stiffness matrix and one Gauss-quadrature point is used for shear stiffness matrix.

14.6 AN EXAMPLE

To illustrate the application of plate elements in stress analysis, an example of a plate clamped at four edges and loaded by uniform pressure is considered (Fig. 14.4). Thin-plate rectangular elements are used in this formulation. The mesh is shown in Fig. 14.5. Figures 14.6 and 14.7 show the contours of σ_x and σ_y, respectively. It is seen that stresses at the edges are greater compared to the stresses at center. This example is a simple one for which even a closed-form solution is available. FEM can be applied to cases involving complicated geometries and non-homogeneous as well as anisotropic materials.

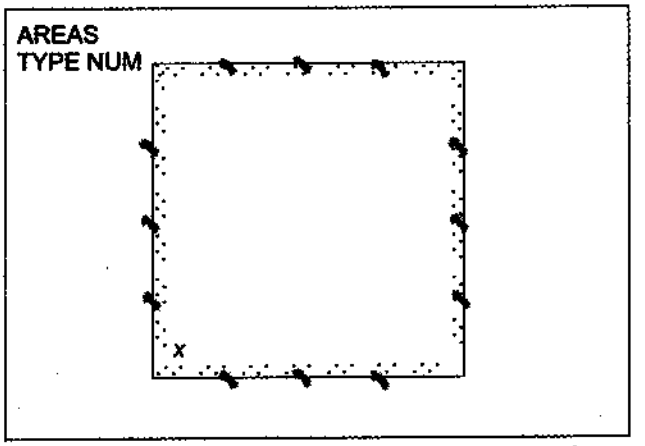


Figure 14.4 *A plate damped at four edges and loaded with uniform pressure*

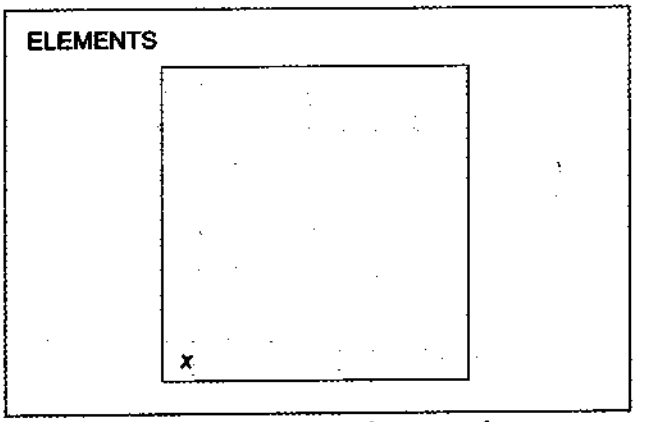


Figure 14.5 *Finite element mesh*

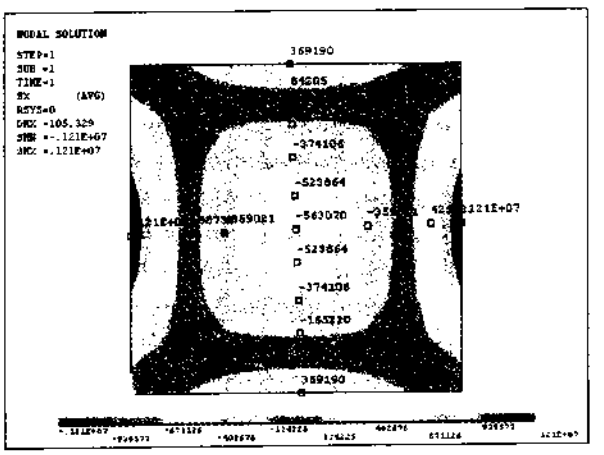


Figure 14.6 *Contours of* σ_x‡

‡ The symbol indicates that the coloured image is provided at the centre of the book.

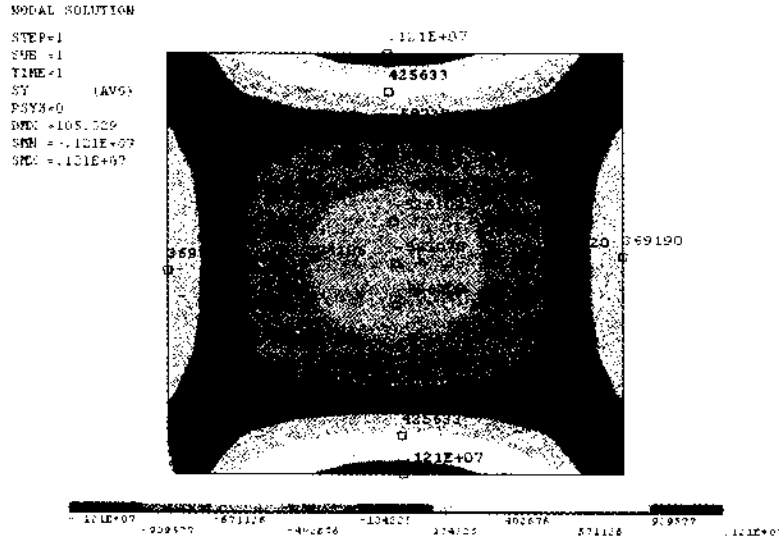


Figure 14.7 *Contours of* $\sigma_y^{\ddagger}$

14.7 CONCLUSION

In this chapter, FEM formulations of thin and thick plates have been discussed. The thin-plate formulation is complicated because of the fourth-order governing differential equation involved. Thick-plate formulation requires only C^0 continuity element; however, it does not provide good results for thin-plate problem due to phenomenon of locking. Locking can be avoided by following a selective integration procedure and it is possible to write a robust code for solving both thick and thin plate problems. It is to be mentioned that for the purpose of stress analysis, even three-dimensional elements can be used. However, solving the plate problem using three-dimensional stress analysis will require a larger memory. Moreover, the system of equations becomes ill-conditioned and poses numerical difficulties.

SOLVED PROBLEMS

14.1 For a four-noded rectangular element, prove that it is impossible to specify simple polynomial expressions for shape functions ensuring full compatibility when only transverse w and its slopes are prescribed at nodes.

Solution Consider the plate element shown in Fig. 14.8. Assume that w, $\partial w/\partial x$, and $\partial w/\partial y$ are prescribed at the nodes. As w and $\partial w/\partial x$ are prescribed along line 1–2, a cubic polynomial can be assumed for the variation of w along line 1–2 as in a beam problem. Thus, along line 1–2, w may be assumed as

$$w = a + bx + cx^2 + dx^3 \tag{14.1.1}$$

For the normal derivative $\partial w/\partial y$ along 1–2, only a linear variation can be assumed, because only the values $\partial w/\partial y$ are prescribed at the two nodes. (We do not know the values of $\partial^2 w/\partial x\partial y$ at the nodes.) Thus, along line 1–2, we can write

$$\frac{\partial w}{\partial y}=\left(1-\frac{x}{h_1}\right)\left(\frac{\partial w}{\partial y}\right)_1+\frac{x}{h_1}\left(\frac{\partial w}{\partial y}\right)_2 \tag{14.1.2}$$

where h_1 is the length of the line 1–2.

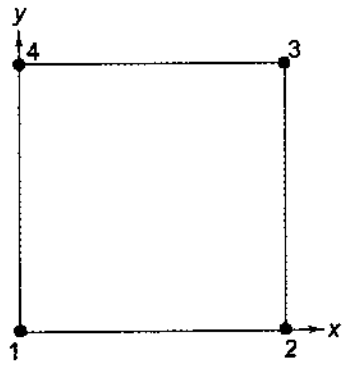


Figure 14.8 *A four-noded plate element*

Along the line 1–4, the normal derivative $\partial w/\partial x$ is a linear function of y. Thus

$$\frac{\partial w}{\partial x}=\left(1-\frac{y}{h_2}\right)\left(\frac{\partial w}{\partial x}\right)_1+\frac{y}{h_2}\left(\frac{\partial w}{\partial x}\right)_4 \tag{14.1.3}$$

where h_2 is the length of the line 1–4.

Differentiating Eq. (14.1.2) with respect to x, we get

$$\frac{\partial^2 w}{\partial x\partial y}=\frac{-\left(\frac{\partial w}{\partial y}\right)_1+\left(\frac{\partial w}{\partial y}\right)_2}{h_1} \tag{14.1.4}$$

Differentiating Eq. (14.1.3) with respect to y, we get

$$\frac{\partial^2 w}{\partial y\partial x}=\frac{-\left(\frac{\partial w}{\partial x}\right)_1+\left(\frac{\partial w}{\partial y}\right)_4}{h_2} \tag{14.1.5}$$

It is clearly seen from Eqs (14.1.4) and (14.1.5) that at the common node 1

$$\frac{\partial^2 w}{\partial x\partial y}\neq\frac{\partial^2 w}{\partial y\partial x} \tag{14.1.6}$$

Thus, we cannot find a suitable polynomial function for this case.

EXERCISES

A. Multiple Choice Questions

14.1 If the thickness of a plate is doubled, its flexural rigidity increases by a factor of
(a) 2 (b) 3 (c) 4 (d) 8

14.2 How many terms does a complete cubic polynomial contain in two dimensions?
(a) 4 (b) 6 (c) 8 (d) 10

14.3 Mindlin plate theory differs from Kirchhoff plate theory in the following respect:
(a) Stress components along the normal to the mid-plane of the plate are assumed to be negligible in Mindlin plate theory.
(b) In Mindlin plate theory, it is assumed that transverse displacements do not vary in the direction of thickness.
(c) In Mindlin plate theory, it is assumed that the normal to the mid-plane remains straight but not necessarily normal to it after deformation.
(d) In Mindlin plate theory, transverse shear deformations are assumed to be zero.

14.4 If the thickness of a thick plate is doubled, then
(a) stiffness due to bending increases by a factor of 8 and stiffness due to shear increases by a factor of 2.
(b) stiffness due to shear increases by a factor of 2 and stiffness due to shear increases by a factor of 8.
(c) both shear stiffness and bending stiffness increases by a factor of 8.
(d) stiffness due to bending increases by a factor of 8 and stiffness due to shear increases by a factor of 4.

14.5 Which is the correct statement among the following?
(a) In Mindlin plate theory, transverse shear deformations are assumed to be zero.
(b) CST elements can be used for solving thin-plate problems.
(c) In thick-plate formulation, the phenomenon of locking is inherently absent.
(d) CST elements can be used for solving thick-plate problems.

B. Problems

14.6 The governing equation for thin-plate bending is given by

$$\frac{\partial^4 w}{\partial x^4} + 2\frac{\partial^4 w}{\partial x^2 \partial y^2} + \frac{\partial^4 w}{\partial y^4} = \frac{q}{D}$$

Obtain the variational form for this. Also obtain FEM equations.

14.7 The flow of an incompressible viscous fluid with negligible inertia is governed by the following differential equation:

$$\frac{\partial^4 \psi}{\partial x^4} + 2\frac{\partial^4 \psi}{\partial x^2 \partial y^2} + \frac{\partial^4 \psi}{\partial y^4} = 0$$

with boundary conditions given as

(i) Essential boundary conditions: $u = u^*$ and $v = v^*$ on Γu

(ii) Natural boundary conditions: $t_x = t_x^*$ and $t_y = t_y^*$

where $\begin{Bmatrix} t_x \\ t_y \end{Bmatrix} = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{bmatrix} \begin{Bmatrix} n_x \\ n_y \end{Bmatrix}$

The variables used are defined as

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\dot{\varepsilon}_{ij}$$

$$u = \psi_{,y} \quad v = -\psi_{,x}$$

where ψ is called the *stream function*. This is similar to the thin-plate equation. Find out the variational form of the equation and obtain FEM formulation.

14.8 For thin-plate FEM analysis, if we choose a four-noded rectangular element with degrees of freedom at the nodes as $w, \partial w/\partial x$ and $\partial w/\partial y$, then this element will be incompatible. Prove this statement mathematically. Also, prove that if we add a degree of freedom, $\frac{\partial^2 w}{\partial x \partial y}$, then this element will become compatible.

14.9 Formulate the problem of vibrations of a thick plate.

14.10 (a) Consider a simply supported thick, square plate loaded with a central concentrated load. Using an FEM package obtain the maximum deflection of the plate.
Solve the same problem by

(b) fixing the plate on one edge and simply supporting on the other three edges.

(c) fixing two edges and simply supporting the other two edges.

(d) fixing three and simply supporting one edge.

(e) fixing all four edges.

Compare the maximum deflection and maximum stresses in all cases.

REFERENCES

Cook, RD, Malkus, DS, and Plesha, ME, 1989, *Concepts and Applications of Finite Element Analysis*, John Wiley & Sons Inc.

Dawe, DJ, 1966, 'Parallelogram element in the solution of rhombic cantilever plate problems', *Journal of Stain Analysis*, Vol. 3.

Gallagher, RH, 1975, *Finite Element Analysis: Fundamentals*, Prentice-Hall, Englewood Cliffs, NJ.

Hrabok, MM. and Hrudey, TM, 1984, 'A review and catalogue of plate bending elements', *Computers & Structures*, Vol. 19, pp. 479–495.

Henshell, RD, Walters, D, and Warburton, GB, 1972, 'A new family of curvilinear plate bending elements for vibration and stability', *Journal of Sound and Vibration*, Vol. 20, pp. 327–343.

Irons, BM, and Draper, JK, 1965, 'Inadequacy of nodal connections in a stiffness solution for plate bending', *J.A.I.A.A.*,Vol. 3.

Zienkiewicz, OC, 1993, *The Finite Element Method*, Tata McGraw-Hill Publishing Company Limited, New Delhi.

15
CHAPTER

Introduction to Non-Linear FEM

15.1 Introduction

15.2 Direct Substitution Method

15.3 Newton–Raphson Method

15.4 Conclusion

15.1 INTRODUCTION

So far we have considered problems in which the coefficient matrix is assumed to be known and is not a function of primary variables. When we consider systems with linear ordinary differential equations or partial differential equations, we obtain constant coefficient matrices in FEM formulation. On the other hand, in the finite element formulation of non-linear differential equations, we obtain coefficient matrices that are dependent on the primary variables. At the same time, there can be many situations where expression on the right-hand side of the equation is a non-linear function of primary variables. Thus, in general, we need to solve systems of equations having the following form:

$$[K(u)]\{u\} = \{R(u)\} \tag{15.1}$$

In solid mechanics problems, non-linearity may arise due to either geometry or material properties. Consider a small deflection of a slender beam under a load intensity q. The Euler–Bernoulli equation for this beam is written as

$$\frac{d^2}{dx^2}\left(EI_{zz}\frac{d^2v}{dx^2}\right) = q \tag{15.2}$$

as mentioned in Chapter 2. If the material is linearly elastic, E is not dependent on the strain or deflection. In non-linearly elastic materials, E may vary with strain; thus, E becomes dependent on the derivative of deflection. Such non-linearity is called *non-linearity due to material*. Now, consider a large deflection of the slender beam. Assume that material is linearly elastic. However, as large-deflection curvature is no longer given by d^2v/dx^2. It is written as

$$\text{Curvature} = \frac{\dfrac{d^2v}{dx^2}}{\left(1+\left(\dfrac{dv}{dx}\right)^2\right)^{3/2}} \tag{15.3}$$

Because of this, the bending moment M is related to deflection in the following manner:

$$M = EI\frac{\dfrac{d^2v}{dx^2}}{\left(1+\left(\dfrac{dv}{dx}\right)^2\right)^{3/2}} \tag{15.4}$$

This is a non-linear relation. Further, in the small-deflection case, moment M is linearly related to q, which is not seen in case of large deflections. Thus, Euler–Bernoulli beam equation for large deflections is non-linear even for a linearly elastic beam material. Such non-linearity is called *non-linearity due to geometry*.

In this chapter, we shall provide an introduction to some of the procedures used for solving non-linear differential equations by FEM. Most of the procedures have been developed on the basis of

well-established methods for solving non-linear equations. Therefore, this chapter provides a brief review of a particular numerical method for solving non-linear equations and then explains how it is applied to solve a system of non-linear equations obtained in the FEM formulation.

15.2 DIRECT SUBSTITUTION METHOD

Suppose there is a non-linear equation $f(x) = 0$ and it is possible to express it in the following form:

$$x = g(x) \tag{15.5}$$

In such a case, the root of the equation $f(x) = 0$ can be obtained iteratively by using the following relation:

$$x_{(i+1)} = g(x_{(i)}) \quad \text{for } i = 0 \text{ to (number of iteration-1)} \tag{15.6}$$

In order to avoid confusion, we are indicating the iteration number as a subscript in brackets. The iterations are carried out till the values of x do not change significantly between two successive iterations or till $f(x)$ is significantly different than 0. The algorithm given by Eq. (15.6) will converge if

$$\left|\frac{dg}{dx}\right| < 1 \tag{15.7}$$

Example 15.1 **The stiffness of a spring is given by the equation**

$$k = 5 + \sqrt{x} \tag{15.8}$$

where k is in N/mm and x denotes deflection in mm. Find out the deflection of the spring for a load of 10 N by using direct substitution method. (The type of spring represented in this problem is called *hardening spring* because its stiffness increases with increasing x.)

Solution The equation governing this problem is written as

$$(5 + \sqrt{x})x = 10 \tag{15.9}$$

Thus

$$x = \frac{10}{5 + \sqrt{x}} \tag{15.10}$$

Assume the initial deflection $x_{(0)}$ to be 0, which gives us $x_{(1)} = 2$. Substituting this value on the right-hand side of Eq. (15.10), we obtain

$$x_{(2)} = \frac{10}{5 + \sqrt{2}} = 1.5590 \tag{15.11}$$

Substituting this value on the right-hand side of Eq. (15.10), we obtain

$$x_{(3)} = \frac{10}{5 + \sqrt{1.5590}} = 1.6004 \tag{15.12}$$

Successive application of this procedure provides

$$x_{(4)} = 1.5962,\ x_{(5)} = 1.5966,\ x_{(6)} = 1.5965, x_{(7)} = 1.5965 \tag{15.13}$$

Thus, the deflection of spring will be 1.5965. We observe that convergence is not monotonic. It is oscillatory.

In this example

$$\frac{dg}{dx} = \frac{d}{dx}\left(\frac{10}{5+\sqrt{x}}\right) = -\frac{5}{\left(5+\sqrt{x}\right)^2\sqrt{x}} \tag{15.14}$$

The magnitude of dg/dx is less than 1 at all other trial solutions, except at $x_{(0)} = 0$. Hence, the solution converges. The oscillatory nature of convergence is due to the negative value of dg/dx.

Example 15.2 The stiffness of a spring is given by the equation

$$k = 5 - \sqrt{x} \tag{15.15}$$

where k is in N/mm and x denotes deflection in mm. Find out the deflection of the spring for a load of 10 N by using direct substitution method. (The type of spring in this problem is called *softening spring* because its stiffness decreases with increasing x.)

Solution The equation governing this problem is written as

$$(5 - \sqrt{x})x = 10 \tag{15.16}$$

Thus

$$x = \frac{10}{5-\sqrt{x}} \tag{15.17}$$

We again start with $x_0 = 0$ and substitute it in Eq. (15.17) to get x_1. The successive application of this procedure provides the following values:

$$\begin{aligned} &x(1) = 2x_{(2)} = 2.7888 \quad x_{(3)} = 3.0030 \quad x_{(4)} = 3.0608 \\ &x_{(5)} = 3.0765 \quad x_{(6)} = 3.0807 \quad x_{(7)} = 3.0818 \quad x_{(8)} = 3.0821 \\ &x_{(9)} = 3.0822 \quad x_{(10)} = 3.0823 \quad x_{(11)} = 3.0823 \end{aligned} \tag{15.18}$$

Thus, the deflection of the spring is 3.0823 mm. We observe that in this case, the convergence is monotonic.

In this example

$$\frac{dg}{dx} = \frac{d}{dx}\left(\frac{10}{5-\sqrt{x}}\right) = \frac{5}{\left(5-\sqrt{x}\right)^2\sqrt{x}} \tag{15.19}$$

Here, dg/dx is always positive, therefore, the convergence is monotonic. However, dg/dx is always less than 1 at all the trial solutions during iterations, except at $x = 0$. At the solution point, its value is 0.2706.

We will now consider a problem in which the solution does not converge.

Example 15.3 The stiffness of a spring is given by the equation

$$k = 1 + 5x^2 \tag{15.20}$$

where k is in N/mm and x denotes deflection in mm. Attempt to find out the deflection of the spring for a load of 10 N by using direct substitution method, as in the last two examples. Thus, show that, in this case, the method does not converge.

Solution The equation governing this problem is written as

$$(1 + 5x^2)x = 10 \tag{15.21}$$

Thus

$$x = \frac{10}{1 + 5x^2} \tag{15.22}$$

We start from the assumption that $x_{(0)} = 0$. Substituting this value on the right-hand side of Eq. (15.22), we get $x_{(1)} = 10$. Substituting $x_{(1)} = 10$, we obtain $x_{(2)} = 0.0200$. Similarly proceeding, we get $x_{(3)} = 9.9800$, $x_{(4)} = 0.0200$, and $x_{(5)} = 9.9800$. Thus, it is observed that the solution starts oscillating from $x = 0.0200$ to $x = 9.9800$.

In this example

$$\frac{dg}{dx} = \frac{d}{dx}\left(\frac{10}{1 + 5x^2}\right) = -\frac{100x}{(1 + 5x^2)^2} \tag{15.23}$$

Although the magnitude of dg/dx remains less than 1 for $x = 0$, 10, and 9.98, its value is more than 1 at $x = 0.0200$. Thus, this example does not pass the convergence test of the direct substitution method. In Example (15.2), dg/dx was more than 1 in the beginning but as the iterations progressed, it became less than 1. The same does not hold true in case of Example 15.3. Here, the value of dg/dx is initially less than 1, but it becomes more than 1 at the second iteration. The approximate solution of this problem is 1.21 mm; at that point, dg/dx is equal to 1.7478.

It is observed that if the magnitude of dg/dx is less than 1 everywhere in the search domain, convergence is assured. In many cases, it may be sufficient if the magnitude of dg/dx is less than 1 near the solution point. In such cases, the initial guess may play an important role. In general, the guess value should be nearer to solution. This ensures faster convergence.

The direct substitution method can also be used to solve systems of non-linear equations. Consider the following system of n equations with n unknowns:

$$\begin{aligned} f_1(x_1, x_2, \ldots, x_n) &= 0 \\ f_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\ldots \\ &\ldots \\ f_n(x_1, x_2, \ldots, x_n) &= 0 \end{aligned} \tag{15.24}$$

In the method of direct substitution, these equations are represented in the following form:

$$\begin{aligned} x_1 &= g_1(x_1, x_2, \ldots, x_n) \\ x_2 &= g_2(x_1, x_2, \ldots, x_n) \\ &\ldots \\ &\ldots \\ x_n &= g_n(x_1, x_2, \ldots, x_n) \end{aligned} \tag{15.25}$$

In vector or matrix notation, Eq. (15.25) can be written as

$$\{x\} = \{g\} \tag{15.26}$$

Using this notation, the direct substitution scheme can be written as

$$\{x\}_{(i+1)} = g\{x\}_{(i)} \qquad \text{for } i = 0 \text{ to (number of iteration-1)} \tag{15.27}$$

It can be shown that if the starting guess solution is sufficiently close to the guess point and if

$$\begin{aligned} \left|\frac{\partial g_1}{\partial x_1}\right| + \left|\frac{\partial g_1}{\partial x_2}\right| + \cdots + \left|\frac{\partial g_1}{\partial x_n}\right| &< 1 \\ \left|\frac{\partial g_2}{\partial x_1}\right| + \left|\frac{\partial g_2}{\partial x_2}\right| + \cdots + \left|\frac{\partial g_2}{\partial x_n}\right| &< 1 \\ &\ldots \\ \left|\frac{\partial g_n}{\partial x_1}\right| + \left|\frac{\partial g_n}{\partial x_2}\right| + \cdots + \left|\frac{\partial g_n}{\partial x_n}\right| &< 1 \end{aligned} \tag{15.28}$$

then the scheme represented by Eq. (15.27) converges.

Now, let us see how this method can be applied to solve non-linear FEM problems. The non-linear system of equations given in Eq. (15.1) generated by finite element procedure, can be written as

$$\{u\} = [K(u)]^{-1}\{R(u)\} \tag{15.29}$$

Equation (15.29) is similar to Eq. (15.26) and can be solved using the scheme given in Eq. (15.27). However, for a large system of equations, obtaining the inverse may not be computationally efficient. Therefore, the direct substitution method is implemented as follows:

- Guess a suitable value for $\{u\}$, say $\{u\}_{(0)}$.
- Using $\{u\}_{(0)}$ calculate $[K(u)]$ and $R(u)$ from Eq. (15.1).
- Obtain the solution $\{u\}_{(1)}$ by any efficient scheme of solving a system of linear equations.
- Using $\{u\}_{(1)}$ again calculate $[K(u)]$ and $R(u)$ from Eq. (15.1) and solve the system of equations for $\{u\}_{(2)}$.
- Repeat this procedure till the solution vectors are significantly different. By taking the difference of two solutions, one can obtain an error vector. For convergence, the Euclidean norm of the error vector (i.e., the square root of the sum of the squares of the all the entries of the vector) should be close to zero. We may also use other convergence criterion, such as the absolute value of any element of the error vector should not exceed a prescribed value.

Sometimes, if the solution starts diverging, an under-relaxation procedure is employed. In this procedure, the obtained solution vector is modified at each iteration as

$$\{u\}_{(i+1)}\Big|_{\text{modified}} = (1-\beta)\{u\}_{(i)} + \beta\{u\}_{(i+1)}, \quad 0 < \beta \le 1 \tag{15.30}$$

where β is the under-relaxation parameter. The suitable value of β is chosen to ensure that at each iteration, some norm of the following error vector is lesser than that at previous iteration

$$\{e\} = [K(u)]\{u\} - \{R(u)\} \tag{15.31}$$

15.3 NEWTON–RAPHSON METHOD

Consider the system of equation given by Eq. (15.1); it can also be written as

$$\{r(u)\} = [K(u)]\{u\} - \{R(u)\} \tag{15.32}$$

where $\{r(u)\}$ denotes the residual vector. In this section, henceforth, $r(u)$, $K(u)$, and $R(u)$ will be denoted by r, K, and R, respectively. Expanding the vector $\{r\}$ by Taylor series about $\{u\}_{(0)}$, a guess solution is given by

$$\{r\} = \{r\}_{(0)} + [J][\{u\} - \{u\}_{(0)}] + \text{higher order terms} \tag{15.33}$$

where $[J]$ is the Jacobian which can be written as

$$[J] = \begin{bmatrix} \dfrac{\partial r_1}{\partial u_1} & \dfrac{\partial r_1}{\partial u_2} \cdots \dfrac{\partial r_1}{\partial u_n} \\ \dfrac{\partial r_2}{\partial u_1} & \dfrac{\partial r_2}{\partial u_2} \cdots \dfrac{\partial r_2}{\partial u_n} \\ & \cdots \\ \dfrac{\partial r_n}{\partial u_1} & \dfrac{\partial r_n}{\partial u_2} \cdots \dfrac{\partial r_n}{\partial u_n} \end{bmatrix} \tag{15.34}$$

Neglecting the higher order terms in Eq. (15.33) and assuming that at $\{u\} = \{u\}_1$, the residual becomes zero and we obtain

$$\{0\} = \{r\}_{(0)} + [J][\{u\}_{(1)} - \{u\}_{(0)}] \tag{15.35}$$

From Eq. (15.35), we get

$$\{u\}_{(1)} = \{u\}_{(0)} - [J]^{-1}\{r\}_{(0)} \tag{15.36}$$

where the inverse of Jacobian is to be calculated at $\{u\}_{(0)}$. Thus, Eq. (15.36) enables us to find a better solution on the basis of a guess solution. This procedure can be repeated in an iterative manner till a converged solution is obtained. In general, the iterative scheme can be represented as

$$\{u\}_{(i+1)} = \{u\}_{(i)} - [J]^{-1}\{r\}_{(i)}, \; i = 0 \text{ to (number of iteration-1)} \tag{15.37}$$

where the inverse of Jacobian is calculated at $\{u\}_{(i)}$. The iterations are stopped when either the norm of $\{r\}$ becomes very small or there is no significant difference between the two solution vectors. It is to be mentioned that like direct substitution method, Newton–Raphson method also does not converge in

all the cases. Sometimes, the under-relaxation procedure may prove to be helpful for Newton–Raphson method as well.

In finite element literature, $[J]$ is also called *tangent stiffness matrix* and is represented as $[K_t]$. In solid mechanics problems, residual vector $\{r\}$ is often called the *unbalanced force vector*. Computation of $[K_t]$ at each iteration may be expensive. Therefore, sometimes the tangent stiffness matrix evaluated at the initial guess point is used in all iterations. This procedure is called modified Newton–Raphson procedure. It requires more iterations for convergence, but computational time in each iteration is reduced.

15.4 CONCLUSION

In this chapter, some methods for solving non-linear FEM problems have been described. The first method—*direct substitution method*—is simple to apply but converges slowly. Newton–Raphson method, on the other hand, converges relatively faster. It can be shown that in direct substitution method, error in an iteration is proportional to that in the previous iteration. Therefore, this method is called *linearly convergent*. Newton-Raphson method is quadratic convergent, which means that error in an iteration is proportional to the square of error in the previous iteration . Each iteration of Newton–Raphson method requires the calculation of inverse of the Jacobian matrix. In modified Newton–Raphson method, the inverse of Jacobain matrix is evaluated at the beginning and used in all iterations. However, it requires more number of iterations to converge. It is important to note that all the methods that are used to solve non-linear FEM equations are highly sensitive to initial guess.

SOLVED PROBLEMS

15.1 The elemental FEM equations for certain problem are given as

$$\begin{bmatrix} u_1 & -u_1 \\ -u_1 & u_2 \end{bmatrix} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix}$$

where the right-hand-side vector is a constant vector. Find out the elemental tangent stiffness matrix.

Solution The residuals for this problem are given as

$$r_1 = u_1^2 - u_1 u_2 - f_1 \quad \text{and} \quad r_2 = -u_1^2 + u_2^2 - f_2$$

The tangent stiffness matrix is given by

$$[J] = \begin{bmatrix} \dfrac{\partial r_1}{\partial u_1} & \dfrac{\partial r_1}{\partial u_2} \\ \dfrac{\partial r_2}{\partial u_1} & \dfrac{\partial r_2}{\partial u_2} \end{bmatrix} = \begin{bmatrix} 2u_1 - u_2 & -u_1 \\ -2u_1 & 2u_2 \end{bmatrix}$$

We observe that although the stiffness matrix is symmetric, the tangent stiffness matrix is not symmetric.

15.2 A linear spring has a stiffness of 9 in appropriate units. It is subjected to a non-linear load of $0.01\ u^3$, where u is the deflection. Obtain the deflection of the spring at equilibrium using Newton-Raphson method. Take initial guess as $u = 25$. Carry out three iterations.

Solution At equilibrium

$$9u = 0.01u^3 \tag{15.2.1}$$

It is seen that $u = 0$ satisfies Eq. (15.2.1). The other equilibrium positions can also be obtained by deriving the closed form solution of the equation. (The solution is $u = \pm 30$.) However, we shall solve it using Newton–Raphson method described in Section 15.3.

From Eq. 15.34, the residual is written as

$$r = 9u - 0.01u^3 \tag{15.2.2}$$

From Eq. (15.36), we can calculate Jacobian J as

$$J = \frac{\mathrm{d}r}{\mathrm{d}u} = 9 - 0.03u^2 \tag{15.2.3}$$

Equation (15.39) provides

$$u_{i+1} = u_1 - J^{-1},\ r_i = u_i - \frac{(9u_i - 0.01u_i^3)}{9 - 0.03u_i^2},\ i = 0, 1, 2$$

Now, we can carry out three iterations as follows with u_0.

Iteration 1

$$u_1 = 25 - \frac{(9 \times 25 - 0.01 \times 25^3)}{9 - 0.03 \times 25^2} = 32.05$$

Iteration 2

$$u_2 = 32.05 - \frac{(9 \times 32.05 - 0.01 \times 32.05^3)}{9 - 0.03 \times 32.05^2} = 30.18$$

Iteration 3

$$u_3 = 30.18 - \frac{(9 \times 30.18 - 0.01 \times 30.18^3)}{9 - 0.03 \times 30.18^2} = 30.00$$

Thus, the answer is $u = 30$. Note that we have retained two decimal places. When you solve this problem using a computer, you will be able to retain many decimal places during iterations.

EXERCISES

A. Multiple Choice Questions

15.1 Among the following four forms, which one will provide converged solution by a direct substitution method with an initial guess of $x = 1$?

(a) $x = \dfrac{5}{x}$ (b) $x = x^2 + x - 5$

(c) $x = \dfrac{x^2 + 1.21}{2.2}$ (d) $x = 5x^2$

15.2 In order to avoid divergence in a non-linear FEM, the primary variable vector $\{u\}_{i+1}$ is modified as

$$\{u\}_{i+1}\big|_{\text{modified}} = \frac{\{u\}_i + \{u\}_{i+1}}{2}$$

This procedure is called

(a) over-relaxation (b) under-relaxation

(c) mean relaxation (d) nodal averaging

15.3 The following equations need to be solved by Newton–Raphson method:

$$x^2 + y - 11 = 0$$
$$x + y^2 - 7 = 0$$

The Jacobian matrix at point (1, 2) is given by

(a) $\begin{bmatrix} -9 & 0 \\ 0 & -2 \end{bmatrix}$ (b) $\begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix}$

(c) $\begin{bmatrix} 2 & 1.5 \\ 1.5 & 4 \end{bmatrix}$ (d) $\begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$

15.4 In solution of FEM problems by Newton–Raphson method, the Jacobian matrix for the system of non-linear equations is called

(a) direct stiffness matrix (b) tangent stiffness matrix

(c) inverse stiffness matrix (d) Hessian matrix

15.5 The advantage of modified Newton–Raphson method over Newton-Raphson method for solving non-linear FEM problems is that

(a) modified Newton–Raphson method requires lesser number of iterations

(b) modified Newton–Raphson method needs less time in each iteration

(c) modified Newton–Raphson method always converges

(d) modified Newton–Raphson method does not require any guess solution

B. Problems

15.6 A linear spring has a stiffness of 9 in appropriate units. It is subjected to a non-linear load of $0.01\ u^3$, where u is the deflection. Obtain the deflection of the spring at equilibrium using Newton–Raphson method. Solve this problem by taking initial guess as $u = 15$, $u = 45$, and $u = 200$. In each case, carry out three iterations.

15.7 The Young's modulus of elasticity of the material of a rod of 1 m length is given as

$$E = 200\ \varepsilon^{0.1}\ \text{GPa}$$

The rod is tapered and its one end is fixed. The area of cross-section of the rod is 2 cm^2 at the fixed end and 1 cm^2 at the other end that is subjected to an axial load of 10 kN. Write a computer program to solve this problem by FEM using direct substitution method and Newton-Raphson method. Generate the results for 10-, 20-, and 30-element meshes. Plot the deflection of rod in each case.

15.8 Table 15.1 gives the variation of thermal conductivity with temperature for a certain type of steel.

Table 15.1 *Variation of Thermal Conductivity with Temperature*

Temperature (°C)	*Thermal Conductivity* (W/m°C)
20	35.1
200	36.8
300	36.8
400	36.2
500	34.6
600	32.0
700	28.5
760	25.8
900	26.6
1000	27.2
1100	27.8
1200	28.4
1300	29.0
1400	29.6

Develop a suitable expression for thermal conductivity as a function of temperature. A square plate of 1 m^2 area is subjected to temperature of 0°C at three edges and 500°C at the fourth edge. Find out the temperature distribution (in the form of contour plots) in the plate using FEM code or package.

15.9 Consider an equation of the following form

$$x = \frac{5}{x}$$

If solved by direct substitution method with initial guess $x = 1$, the convergence will not be obtained. Can this problem be solved with the help of under-relaxation? If yes, show atleast 6 iterations using the procedure.

15.10 Is it possible to increase the rate of convergence by using some sort of over-relaxation procedure? Discuss.

16
CHAPTER

FEM Formulation for Flow Problems

16.1 Introduction
16.2 Strip-Drawing Process: Governing Equations and Boundary Conditions
16.3 Constitutive Relations
16.4 A Note on Different Types of FEM Formulations
16.5 Weak Formulation for Mixed Pressure–Velocity Formulation
16.6 Finite Element Approximation
16.7 Finite Element Equations
16.8 Discretization of Strip
16.9 Application of Boundary Conditions
16.10 Post-Processing
16.11 FEM Formulation of Fluid Flow With Inertia Efect
16.12 Conclusion

16.1 INTRODUCTION

In the previous chapter, we introduced non-linear FEM with special emphasis on solving non-linear problems. In this chapter, we shall study the formulation of two non-linear problems. First, we shall discuss the steady-state drawing of a strip. In the strip-drawing process, the thickness of a strip is reduced by passing it through a die as shown in Fig. 16.1. If the width of the strip is large compared to the contact length of die and strip, the problem can be treated as a plane strain problem, as the spread along the width direction will be negligible. Alternatively, the die itself can restrain the strip along the width direction. The strip-drawing process is not among the most widely used processes. However, we are discussing it first due to the relative simplicity of its finite element formulation. Two exercise problems have been provided at the end of the chapter–one on rolling and another on wire drawing—to help readers grasp the application concepts clearly. In this chapter, we have assumed that the material is rigid-perfectly plastic, i.e., there is no elastic deformation and no strain hardening. In uniaxial tension tests, such materials do not deform till a critical stress called *flow* (or *yield*) *stress* is reached. After that, the material starts deforming plastically. The rate of deformation is related to the applied stress by the constitutive relation of the metal. The behaviour of the material under the action of any form of applied force is modelled using the theory of plasticity. For detailed treatment of theory of plasticity, the reader may refer to Dixit and Dixit (2008). In this chapter, we shall be writing the governing equations without deriving them and obtain the corresponding finite element formulation.

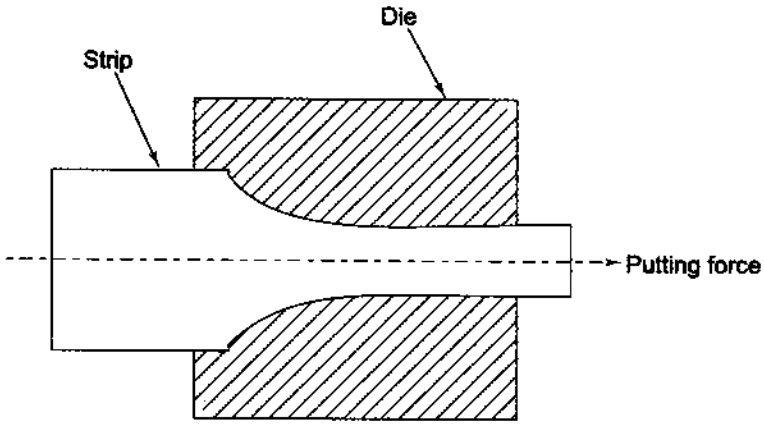


Figure 16.1 *A schematic diagram of the strip-drawing process*

Owing to the assumption regarding the behaviour of the material, the metal-forming problems (namely, strip drawing, rolling, and wire drawing) discussed in this chapter can be treated as problems of flow of a non-Newtonian fluid. In case of non-Newtonian fluids, the shear stress is related to the shear strain rate in a non-linear manner. It is to be observed that there is no ideal perfectly-plastic material and during metal forming, the metals do not completely behave like a fluid. The FEM formulations in which

the metal is modelled like a fluid are called *flow formulations*. Although they have certain limitations, flow formulations often provide quite useful results.

After discussing the flow formulations of metal-forming problems, we discuss the formulation of the famous Navier–Stokes equations of fluid mechanics. In the traditional form of Navier-Stokes equations, the fluid is assumed to be Newtonian. Thus, the shear stress components are linearly related to shear strain rate components. However, because of the presence of inertia terms, the equations still become non-linear. Efficient solution techniques of Navier-Stokes equations form one of the major portions of computational fluid dynamics (CFD) where a number of methods, e.g., finite difference method, finite volume method, finite element method, etc., are studied. In this chapter, only a very brief introduction to FEM formulation will be provided.

To illustrate the concept, the problem of strip drawing has been formulated in great detail. The steady-state strip-drawing process can be considered as the flow of a viscous fluid, where viscosity is dependent on the flow stress and strain rate. After reading this chapter, you will be able to carry out the finite element formulation of Navier–Stokes equations.

16.2 STRIP-DRAWING PROCESS: GOVERNING EQUATIONS AND BOUNDARY CONDITIONS

The plane strain drawing of the strip is considered in this section. Because of the symmetry of the process, it is enough to model the upper half portion of the strip passing through the die. Figure 16.2 shows the domain of the problem. The steady-state process is assumed and the domain shown in Fig. 16.2 is the control volume (a fixed region in space). The formulation in which the domain is a controlled volume is called *Eulerian formulation*. The inlet and exit boundaries are assumed to be sufficiently far away from the die so that the flow velocities at the inlet and exit can be assumed to be along the horizontal direction (normal to the straight boundaries). The die is assumed to be curved and perfectly rigid.

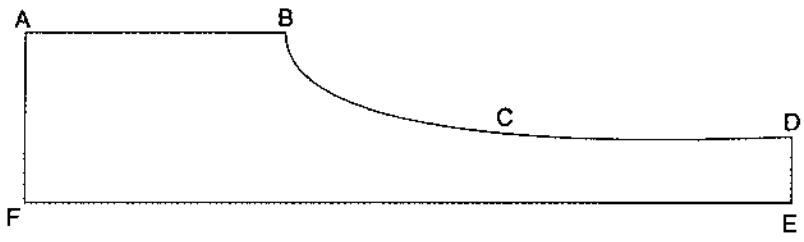


Figure 16.2 *Domain of strip-drawing problem*

For the sake of convenience, we shall use index notation, which was explained in Section 11.4. The index 1 in the subscript is used to indicate x, index 2 indicates y. If an index is repeated in a term, two terms obtained by assigning the values 1 and 2 to this index will have to be summed.

16.2.1 Governing Equations

For the present case, it is assumed that the material is incompressible and is moving with slow speed, and that the temperature effects are negligible. Thus, we need to have one equation for the conservation of mass and two equations for force balance. In Eulerian form, the equations for conservation of mass of an incompressible material are written as

$$\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} = \frac{\partial v_i}{\partial x_i} = v_{i,i} = 0 \tag{16.1}$$

As already explained, x_1 denotes x-axis and x_2 denotes y-axis. Similarly, v_1 denotes the velocity in x_1 direction and v_2 denotes velocity in x_2 direction. Equation (16.1) holds good for steady as well as non-steady flow of an incompressible material. Therefore, this equation may be called *incompressibility constraint equation*.

In index notation, the strain rate tensor is defined as

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) = \frac{1}{2}\left(v_{i,j} + v_{j,i}\right) \tag{16.2}$$

Many authors feel that it is appropriate to call the strain rate tensor as the rate of deformation tensor. This is because only for small deformations, strain rate tensor [Eq. (16.2)] represents the rate of change of the strain tensor, the expression for which is written as

$$\dot{\varepsilon}_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) \tag{16.3}$$

where u_1 and u_2 are the displacements along x and y directions. In this chapter, we shall use strain rate instead of rate of deformation. For large as well as small deformations, $\dot{\varepsilon}_{11}$ represents the rate of change in length of a material arc per unit of current length along the x_1 direction. Similarly, $\dot{\varepsilon}_{22}$ represents the rate of change in length of a material element per unit of current length along the x_2 direction. The components $\dot{\varepsilon}_{12}$ and $\dot{\varepsilon}_{21}$ represent one-half the rate of decrease in the angle between material arcs lying along x_1 and x_2 directions. For a rigid body motion, strain rate tensor will be zero.

In terms of the components of strain-rate tensor, Eq. (16.1) can be written as

$$\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22} = \dot{\varepsilon}_{ii} = 0 \tag{16.4}$$

In the absence of any body force, such as gravity, the force balance along x and y directions provide the following equilibrium equations:

$$\frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{21}}{\partial x_2} = 0 \tag{16.5a}$$

$$\frac{\partial \sigma_{12}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} = 0 \tag{16.5b}$$

where σ_{ij} denotes a stress component. Note that compared to other forces, the force of gravity may be considered to be negligible. In the absence of any body moment, stress tensor is symmetric. Thus

$$\sigma_{12} = \sigma_{21} \tag{16.6}$$

Therefore, Eqs (16.5a) and (16.5b) may be written as

$$\frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{21}}{\partial x_2} = 0 \tag{16.7a}$$

$$\frac{\partial \sigma_{12}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} = 0 \tag{16.7b}$$

Equations (16.4), (16.7a), and (16.7b) are three partial differential equations that need to be solved. However, they contain five unknowns, three stress components, and two strain components. The additional equations will be provided by the constitutive relation of material. Constitutive relation of the material will relate stress to strain rate, which is expressed as the velocity gradient [Eq. (16.2)]. Thus, the differential equations will ultimately contain the velocity components as dependent variables. The velocity components can therefore be treated as primary variables. The strain rate and stresses obtained by taking first derivatives of the velocity vector can be treated as secondary variables. Before proceeding further to obtain FEM formulation, we shall discuss the boundary conditions of the present system.

16.2.2 Boundary Conditions

The boundary conditions specified on any boundary can be classified as *essential* boundary conditions, i.e., the specification of the primary variable vector (such as velocity vector v_i), and *natural* boundary conditions, i.e., the specification of the secondary variable (such as traction vector). The traction on a surface is the force per unit area in the limit of the area tending to zero. In index notation, the components of traction vector can be written as

$$t_i = \sigma_{ji} n_j \tag{16.8}$$

where n_j are the direction cosines of the normal to the surface. Equation (16.8) is called Cauchy's law, in view of the symmetry of the stress tensor, the equation may also be written as

$$t_i = \sigma_{ij} n_j \tag{16.9}$$

Figure 16.2 shows the different boundaries of the strip. The boundary conditions at the different boundaries of the strip are described as follows:

Entry and exit boundaries (AF and DE)

The boundaries *AF* and *DE* are located far away from the die. Therefore, on the surface AF and DE, the boundary condition can be written as

$$v_1 = U_1,\ v_2 = 0 \text{ on AF} \tag{16.10}$$

$$v_1 = U_2,\ v_2 = 0 \text{ on DE} \tag{16.11}$$

where U_1 and U_2 are the inlet and exit velocities, respectively. In strip-drawing process, the strip is pulled through the die and exit speed is controlled. The entry speed is calculated from the continuity requirement as

$$U_1 = \frac{h_2}{h_1} U_2 \tag{16.12}$$

where h_1 and h_2 are the inlet and exit thicknesses, respectively. In place of $v_2 = 0$, it is possible to specify $t_2 = 0$, i.e., the traction on the y-direction is zero, because there is no force in y-direction on these surfaces. However, given a preference, it is better to specify the essential boundary conditions. In FEM, the accuracy of the derivatives is lesser compared to primary variable. Therefore, many a times, the natural boundary condition may not be imposed accurately. On the other hand, the essential boundary conditions are always imposed correctly and one can reduce the size of the system of equations to be solved by eliminating the known degrees of freedom. In any case, one needs to impose the minimum number of boundary conditions needed for eliminating the rigid body modes.

Top free surfaces (AB and CD)

The following boundary conditions may be specified on the top free surfaces:

$$t_1 = 0, v_2 = 0 \text{ on } AB \text{ and } CD \text{ (except some nodes near the die)} \tag{16.13}$$

$$t_1 = 0, t_2 = 0 \quad \text{on some nodes near the die} \tag{16.14}$$

Note that it is possible to specify $t_1 = 0$ and $t_2 = 0$ throughout the top surfaces. However, as argued in the preceding paragraph, it is always better to specify essential boundary conditions. We have assumed that the top free surfaces are straight and parallel to x axis (x_1). This may not be true near the die. As a result of FEM solution, if v_2 comes out to be nearly zero near the die, then it is ascertained that the assumed free boundary is correct. However, if v_2 does not come out to be zero, the boundary may be modified such that there is no velocity component normal to the boundary. This, however, introduces additional computational complexities.

Axis of symmetry (FE)

Because of the symmetry of the system, the following boundary conditions will hold good:

$$t_1 = 0, v_2 = 0 \quad \text{on FE} \tag{16.15}$$

The justification for these boundary conditions is as follows. Assume that t_1 is positive, which means that the upper portion of the strip is being pulled towards exit. From Newton's third law, the upper portion will pull the lower portion towards entry. However, in that case, the symmetry will be destroyed as the top portion is being pulled towards exit and the bottom portion is being pulled towards entry. Similarly, we can argue that if the symmetry is to be preserved, t_1 cannot be negative. Thus, for preservation of symmetry t_1 must be zero. Due to symmetry, the particles on line FE will not have any tendency to go towards a particular direction. Hence, v_2 will be zero.

Die-strip interface (BC)

On the die–strip interface, the component of velocity normal to the die will be zero. Hence

$$v_2 + v_1 \tan\phi = 0 \text{ on } BC \tag{16.16}$$

where ϕ is the inclination of the die surface from horizontal. For a conical die, ϕ is the die semi-angle; for curved die, ϕ is the angle of tangent to a point on the die surface from the surface and it varies from along the length of the die.

The second boundary condition is written as

$$t_s = \mu |t_n| \tag{16.17}$$

The tractions given in Eq. (16.17) are shown in Fig. 16.3, where t_s and t_n are the tangential and normal components of the stress vector t which is defined by $\sigma_{ij} n_j = t_i$ and μ is the coefficient of friction in the case of Coulomb's friction model. Note that as per the sign convention of Fig. 16.3, t_n should come always be negative at the die–strip interface because the die cannot pull the material towards itself.

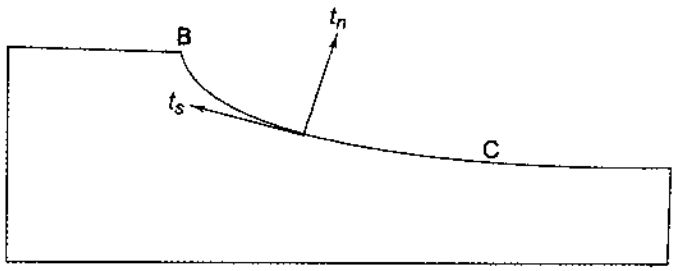


Figure 16.3 *Schematic diagram of die–strip interface BC*

16.3 CONSTITUTIVE RELATIONS

In order to express stress as a function of strain rate, the Cauchy stress is decomposed into two parts

$$\sigma_{ij} = -p\delta_{ij} + S_{ij} \tag{16.18}$$

where p is the negative of the mean normal stress and S_{ij} is the deviatoric part of σ_{ij}. For most fluids, p will be equal to thermodynamic pressure. Therefore, we shall often call the negative of mean normal stress as pressure. The pressure p is obtained from the stress tensor using the following relation:

$$p = -\frac{\sigma_{kk}}{3} \tag{16.19}$$

In Eq. (16.18), δ_{ij} represents Kronecker's delta, which has the following property:

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{16.20}$$

The part $-p\delta_{ij}$ in Eq. (16.18) is also called the *hydrostatic part*.

Similar to stress, the strain-rate tensor can also be decomposed into hydrostatic and deviatoric parts as follows:

$$\dot{\varepsilon}_{ij} = \frac{1}{3}\,\dot{\varepsilon}_{kk}\delta_{ij} + \dot{\varepsilon}'_{ij} \tag{16.21}$$

where $\dot{\varepsilon}'_{ij}$ is the deviatoric part of the strain-rate tensor. For many fluids, the deviatoric part of the stress and deviatoric part of the strain are related as

$$S_{ij} = 2\mu\,\dot{\varepsilon}'_{ij} \tag{16.22}$$

where for Newtonian fluids μ is a constant called the *viscosity* of the fluid. For non-Newtonian fluids, μ is not constant. Substituting Eq. (16.22) in Eq. (16.18), we obtain

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\,\dot{\varepsilon}'_{ij} \tag{16.23}$$

Substituting the expression for deviatoric strain from Eq. (16.21) into Eq. (16.23), we get

$$\sigma_{ij} = \left(-p - \frac{2}{3}\mu\dot{\varepsilon}_{kk}\right)\delta_{ij} + 2\mu\dot{\varepsilon}_{ij} \tag{16.24}$$

Equation (16.14) holds good for most of the compressible and incompressible fluids, where p is the thermodynamic pressure. Given this relation and assuming p to be the thermodynamic pressure, we can show that for fluids following this relation, p is negative of mean direct stress. For this, make index j equal to i in Eq. (16.24) to obtain

$$\sigma_{ii} = \left(-p - \frac{2}{3}\mu\dot{\varepsilon}_{kk}\right)\delta_{ii} + 2\mu\dot{\varepsilon}_{ii} \tag{16.25}$$

Noting that

$$\delta_{ii} = \delta_{11} + \delta_{22} + \delta_{33} = 3 \tag{16.26}$$

Eq. (16.25) can be written as

$$\sigma_{ii} = \left(-p - \frac{2}{3}\mu\dot{\varepsilon}_{kk}\right)\delta_{ii} + 2\mu\dot{\varepsilon}_{ii} = -3p - 2\,\mu\dot{\varepsilon}_{kk} + 2\mu\dot{\varepsilon}_{ii} \tag{16.27}$$

In Eq. (16.27), k is the dummy index, hence it can be replaced by i. Thus, the equation becomes

$$\sigma_{ii} = -3p - 2\,\mu\dot{\varepsilon}_{ii} + 2\mu\dot{\varepsilon}_{ii} = -3p \tag{16.28a}$$

or

$$p = -\frac{1}{3}\,\sigma_{ii} = \frac{-(\sigma_{11} + \sigma_{22} + \sigma_{33})}{3} \tag{16.28b}$$

Thus, the relation given by Eq. (16.24) implies that the mean direct stress is equal to negative thermodynamic pressure.

For some isotropic fluids, the stress and strain rate are related as

$$\sigma_{ij} = (-p + \mu'\dot{\varepsilon}_{kk})\delta_{ij} + 2\mu\dot{\varepsilon}_{ij} \tag{16.29}$$

where μ' is called *second coefficient of viscosity* and μ is called the *first coefficient of viscosity*. However, it has been observed that the relation given by Eq. (16.24) holds good for liquids and gases

except when they are close to critical point. Thus, we shall not be using Eq. (16.29) in this chapter, although it is easy to extend the formulations presented in this chapter to cases which utilize Eq. (16.29).

The constitutive relation given by Eq. (16.24) may be used for metal forming also, despite the fact that μ is not the constant viscosity this time. Here, we shall call μ as the proportionality factor. Considering the metal deformation process as incompressible, the stress is related to strain rate as

$$\sigma_{ij} = -p\delta_{ij} + 2\mu\dot{\varepsilon}_{ij} \tag{16.30}$$

For incompressible materials yielding according to the von Mises criterion, the proportionality factor μ is given by

$$\mu = \frac{\sigma_y}{\sqrt{6\dot{\varepsilon}_{ij}\dot{\varepsilon}_{ij}}} = \frac{\sigma_y}{3\dot{\varepsilon}_{eq}} \tag{16.31}$$

where $\dot{\varepsilon}_{eq}$ is called *equivalent strain rate* and σ_y is called *flow stress* (the stress at which plastic deformation takes place and material starts flowing) of the material. For perfectly plastic materials, flow stress is equal to yield stress.

16.4 A NOTE ON DIFFERENT TYPES OF FEM FORMULATIONS

There are two popular ways of solving the flow problem in a constant volume. One is to carry out FEM formulation with u, v, and p as primary variables. This is known as mixed pressure–volume (p-v) formulation. This is same as the Lagrange-multiplier method with p being the Lagrange multiplier. The second way is to eliminate p using penalty method. In this method, Eq. (16.1) is written as

$$\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} + \frac{p}{\lambda} = 0 \tag{16.32}$$

where λ is a very large number. Ideally for incompressible flow, this number should be infinite. However, in practice, this number should be chosen depending upon the precision of computer. From Eq. (16.32)

$$p = -\lambda\left(\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2}\right) \tag{16.33}$$

Substituting Eq. (16.33) in Eq. (16.31), p gets eliminated from the constitutive relation. When the constitutive relation is substituted in equations of equilibrium, differential equations with velocity components as the dependent variables are obtained.

16.5 WEAK FORMULATION FOR MIXED PRESSURE-VELOCITY FORMULATION

Let v_1, v_2, and p be the functions that satisfy all essential boundary conditions exactly, the weighted residual will thus become

$$\int_A \left[(\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22}) w_p + \left(\frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2} \right) w_1 + \left(\frac{\partial \sigma_{21}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} \right) w_2 \right] \mathrm{d}x_1 \mathrm{d}x_2 = 0 \tag{16.34}$$

where A denotes domain of a typical area element and w_p, w_1, and w_2 are the weight functions.

The expression given in Eq. (16.34) can be written in compact form using the index notation as

$$\int_A \left[\dot{\varepsilon}_{ii} w_p + \sigma_{ij,j} w_i \right] \mathrm{d}x_1 \mathrm{d}x_2 = 0 \tag{16.35}$$

The first part of the integral contains the first-order derivatives, which cannot be weakened further, but the second part contains second-order derivatives that can be integrated by parts. The second part can be written as

$$\int_A \left[\sigma_{ij,j} w_i \right] \mathrm{d}A = \int_A \left[(\sigma_{ij,j} w_i)_{,j} - (\sigma_{ij} w_{i,j}) \right] \mathrm{d}A \tag{16.36}$$

Therefore, Eq. (16.35) becomes

$$\int_A \dot{\varepsilon}_{ii} w_p \mathrm{d}A + \int_A \left[(\sigma_{ij} w_i)_{,j} - (\sigma_{ij} w_{i,j}) \right] \mathrm{d}A = 0 \tag{16.37}$$

From Gauss divergence theorem, we can write

$$\int_A (\sigma_{ij} w_i)_{,j} \mathrm{d}A = \oint_\Gamma \sigma_{ij} w_i n_j \, \mathrm{d}A \tag{16.38}$$

where Γ is the boundary of the element and n_j denotes the direction cosine vector in index notation. According to Cauchy's law, $\sigma_{ij} n_j = t_i$; therefore, Eq. (16.37) can be expressed as

$$\int_A \dot{\varepsilon}_{ii} w_p \mathrm{d}A + \oint_\Gamma t_i w_i \mathrm{d}\Gamma - \int_A (\sigma_{ij} w_{i,j}) \mathrm{d}A = 0 \tag{16.39}$$

The last term in the left-hand side of Eq. (16.39) can be expressed as

$$\int_A \left(\sigma_{ij} w_{i,j} \right) \mathrm{d}A = \int_A \sigma_{ij} \frac{1}{2} \left[\left(w_{i,j} + w_{j,i} \right) + \left(w_{i,j} - w_{j,i} \right) \right] \mathrm{d}A \tag{16.40}$$

Now, consider the expression

$$a = \sigma_{ij} (w_{i,j} - w_{j,i}) \tag{16.41}$$

The reader should note carefully that in Eq. (16.41) a is a scalar, as there is no free index; both i and j are dummy indices. A dummy index can be replaced by any other index without affecting the value of the term. In Eq. (16.41), we interchange i and j and make use of symmetry of σ_{ij} to obtain

$$a = \sigma_{ji} (w_{j,i} - w_{i,j}) = -\sigma_{ij} (w_{i,j} - w_{j,i}) \tag{16.42}$$

Comparing Eq. (16.43) with Eq. (16.41), we get

$$a = -a \quad \text{or} \quad 2a = 0 \tag{16.43}$$

Thus, we can write Eq. (16.41) as

$$\sigma_{ij} (w_{i,j} - w_{j,i}) = 0 \tag{16.44}$$

Therefore, Eq. (16.40) becomes

$$\int_A \left(\sigma_{ij} w_{i,j}\right) \mathrm{d}A = \int_A \sigma_{ij} \frac{1}{2}\left[w_{i,j} + w_{j,i}\right] \mathrm{d}A \tag{16.45}$$

Analogous to Eq. (16.2), we define

$$\dot{\varepsilon}_{ij}(w) = \frac{1}{2}\left(\frac{\partial w_i}{\partial x_j} + \frac{\partial w_j}{\partial x_i}\right) = \frac{1}{2}\left(w_{i,j} + w_{j,i}\right) \tag{16.46}$$

Thus, Eq. (16.45) can be written as

$$\int_A \left(\sigma_{ij} w_{i,j}\right) \mathrm{d}A = \int_A \sigma_{ij} \dot{\varepsilon}_{ij}(w)\, \mathrm{d}A \tag{16.47}$$

Now, using Eq. (16.47), the weighted residual given in Eq. (16.40) becomes

$$\int_A \dot{\varepsilon}_{ii} w_p \, \mathrm{d}A - \int_A \sigma_{ij} \dot{\varepsilon}_{ij}(w)\, \mathrm{d}A - \int_{\Gamma_i} t_i w_i \mathrm{d}\Gamma = 0 \tag{16.48}$$

where Γ_i is the surface at which t_i is prescribed. The first two terms in Eq. (16.49) are first-order derivatives of the velocity component; thus, we have obtained the weak form for further finite element modelling. Using the relationship given by Eq. (16.30) for incompressible material, we can write

$$\begin{aligned} \sigma_{ij} \dot{\varepsilon}_{ij}(w) &= \left(-p\delta_{ij} + 2\mu \dot{\varepsilon}_{ij}\right) \dot{\varepsilon}_{ij}(w) \\ &= -p\dot{\varepsilon}_{ii}(w) + 2\mu \dot{\varepsilon}_{ij} \dot{\varepsilon}_{ij}(w) \end{aligned} \tag{16.49}$$

Note that in the simplification carried out in Eq. (16.50), the following substitution property of Kronecker's delta has been used:

$$a_{ij}\delta_{jk} = a_{ik} \tag{16.50}$$

In unabridged form, Eq. (16.49) is written as

$$\sigma_{ij} \dot{\varepsilon}_{ij}(w) = -p\left[\dot{\varepsilon}_{11}(w) + \dot{\varepsilon}_{22}(w)\right] + 2\mu\left[\dot{\varepsilon}_{11}\dot{\varepsilon}_{11}(w) + 2\dot{\varepsilon}_{12}\dot{\varepsilon}_{12}(w) + \dot{\varepsilon}_{22}\dot{\varepsilon}_{22}(w)\right] \tag{16.51}$$

Substituting Eq. (16.52) in Eq. (16.48), the weak form of weighted residual is obtained as

$$\int_A I_1 \, \mathrm{d}A + \int_A I_2 \, \mathrm{d}A - \int_{\Gamma_1} I_3 \mathrm{d}\Gamma - \int_{\Gamma_2} I_4 \mathrm{d}\Gamma = 0 \tag{16.52}$$

where

$$\begin{aligned} I_1 &= -\varepsilon_{ii} w_p = -\left[\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22}\right] w_p \\ I_2 &= -p\left[\dot{\varepsilon}_{11}(w) + \dot{\varepsilon}_{22}(w)\right] + 2\mu\left[\dot{\varepsilon}_{11}\dot{\varepsilon}_{11}(w) + 2\dot{\varepsilon}_{12}\dot{\varepsilon}_{12}(w) + \dot{\varepsilon}_{22}\dot{\varepsilon}_{22}(w)\right] \\ I_3 &= t_1 w_1 \\ I_4 &= t_2 w_2 \end{aligned} \tag{16.53}$$

Γ_1 and Γ_2 are those parts of the boundary where tractions t_1 and t_2 are specified, respectively. Note that we could have written the last two terms of Eq. (16.52) as the boundary integration over the whole boundary. However, at a boundary where velocity component v_i is specified instead of t_i, the weight function w_i may be taken as zero. Thus, the form given by Eq. (16.52) is justified.

16.6 FINITE ELEMENT APPROXIMATION

In the integrand of the weak form of the governing equation [Eq. (16.52)], velocity is present as a first-order derivative, while pressure is present as a zeroth order derivative. Therefore, C^0 continuity elements are sufficient for modelling and pressure field may be discontinuous. On the boundary also, we obtain terms containing the first derivative of the velocity and pressure. Therefore, for completeness, one may have approximation in the following form:

$$v_1 = a_1 + b_1 x + c_1 y;\ v_2 = a_2 + b_2 x + c_2 y;\ p = q \tag{16.54}$$

where a_1, b_1, c_1, a_2, b_2, c_2, and q are constants. As usual, these constants may be expressed in terms of unknown nodal values of primary variables. However, we avoid taking a discontinuous pressure field and instead choose a four-noded quadrilateral element for pressure. To avoid numerical difficulties, approximation for velocity is chosen one degree higher than pressure variable. This will avoid the problem of dealing with a large number of zeros in the leading diagonal of the assembled matrix. Moreover, as the secondary variables, namely, stress and strain, contain the first derivative of velocity, the velocity solution needs to be more accurate than pressure. Hence, a nine-noded quadrilateral element is chosen for velocity approximation. Thus, there will be nine shape functions for both components of velocity and four shape functions for pressure approximation. For generating coefficient matrices, the elements are transformed to natural coordinates. The approximation for v_1 and v_2 is

$$\begin{Bmatrix} v_1 \\ v_2 \end{Bmatrix} = \begin{bmatrix} N_1 & 0 & N_2 & 0 \ldots N_9 & 0 \\ 0 & N_1 & 0 & N_2 \ldots 0 & N_9 \end{bmatrix} \begin{Bmatrix} (v_1)_1^e \\ (v_2)_1^e \\ \vdots \\ \vdots \\ (v_1)_9^e \\ (v_2)_9^e \end{Bmatrix} = [N_v]\{v^e\} \tag{16.55}$$

In natural coordinates, the expression for the shape functions are given by

$$\begin{aligned}
N_1 &= \frac{1}{4}(\xi^2 - \xi)(\eta^2 - \eta) & N_2 &= \frac{1}{4}(\xi^2 + \xi)(\eta^2 - \eta) \\
N_3 &= \frac{1}{4}(\xi^2 + \xi)(\eta^2 + \eta) & N_4 &= \frac{1}{4}(\xi^2 - \xi)(\eta^2 + \eta) \\
N_5 &= \frac{1}{4}(1 - \xi^2)(\eta^2 - \eta) & N_6 &= \frac{1}{4}(\xi^2 + \xi)(1 - \eta^2) \\
N_7 &= \frac{1}{4}(1 - \xi^2)(\eta^2 + \eta) & N_8 &= \frac{1}{4}(\xi^2 - \xi)(1 - \eta^2) \\
N_9 &= (1 - \xi^2)(1 - \eta^2)
\end{aligned} \tag{16.56}$$

In Galerkin formulation, weight functions for velocity and pressure are approximated using the same shape functions as that of velocity and pressure, respectively. Thus

$$\{w_v\} = \begin{Bmatrix} w_1 \\ w_2 \end{Bmatrix} = [N_v]\{w_v^e\} = \lfloor w_v \rfloor^T \lfloor N_v \rfloor^T \tag{16.57}$$

The approximation for pressure is written as

$$p = \lfloor N_1^p \quad N_2^p \quad N_3^p \quad N_4^p \rfloor \begin{Bmatrix} p_1^e \\ p_2^e \\ p_3^e \\ p_4^e \end{Bmatrix} = \{N_p\}^T \{p^e\} \tag{16.58}$$

where

$$\begin{aligned} N_1^p &= \frac{1}{4}(1-\xi)(1-\eta), \qquad N_2^p = \frac{1}{4}(1+\xi)(1-\eta) \\ N_3^p &= \frac{1}{4}(1+\xi)(1+\eta), \qquad N_4^p = \frac{1}{4}(1-\xi)(1+\eta) \end{aligned} \tag{16.59}$$

The weight function is given by

$$\{w_p\} = \lfloor N_p \rfloor \{w_p^e\} \tag{16.60}$$

In order to model a curved boundary, the geometry is approximated by nine-noded shape functions. Thus, for geometry approximation, the same shape functions can be considered as that of the velocity variable. Therefore

$$x_1 = \lfloor N_1 \quad N_2 \quad \ldots N_9 \rfloor \{x_1^e\} = \lfloor N \rfloor \{x_1^e\} \tag{16.61}$$

where $\{x_1^e\}$ is the vector containing nodal x-coordinates. Similarly

$$x_2 = \lfloor N_1 \quad N_2 \quad \ldots N_9 \rfloor \{x_2^e\} = \lfloor N \rfloor \{x_2^e\} \tag{16.62}$$

where $\{x_2^e\}$ is the vector containing nodal y-coordinates.

As the boundary of the element consists of three nodes, the weight functions at the boundary may be written as

$$w_1 = \lfloor N_1^b \quad N_2^b \quad N_3^b \rfloor \begin{Bmatrix} w_1^{b1} \\ w_1^{b2} \\ w_1^{b3} \end{Bmatrix} = \lfloor N_b \rfloor \{w_1^b\} \tag{16.63}$$

where the superscripts $b1$, $b2$, and $b3$ indicate boundary nodes. This is consistent with the approximation of weight over an element, as on a boundary of a nine-noded quadratic element only three weights will be active. In a similar way

$$w_2 = \lfloor N_1^b \quad N_2^b \quad N_3^b \rfloor \begin{Bmatrix} w_2^{b1} \\ w_2^{b2} \\ w_2^{b3} \end{Bmatrix} = \lfloor N_b \rfloor \{w_2^b\} \tag{16.64}$$

Further, t_1 and t_2 are approximated as

$$t_1 = \lfloor N_1^b \;\; N_2^b \;\; N_3^b \rfloor \begin{Bmatrix} t_1^{b1} \\ t_1^{b2} \\ t_1^{b3} \end{Bmatrix} = \lfloor N_b \rfloor \{ t_1^b \} \tag{16.65}$$

$$t_2 = \lfloor N_1^b \;\; N_2^b \;\; N_3^b \rfloor \begin{Bmatrix} t_2^{b1} \\ t_2^{b2} \\ t_2^{b3} \end{Bmatrix} = \lfloor N_b \rfloor \{ t_2^b \} \tag{16.66}$$

where t_1^{b1} and t_2^{b1} are the values of the traction at the first node in x and y directions, respectively and so on. The shape function at a boundary are given by

$$N_1^b = \frac{1}{2}\left(\zeta^2 - \zeta\right), N_2^b = \frac{1}{2}\left(\zeta^2 + \zeta\right),\; N_3^b = \frac{1}{2}\left(1 - \zeta^2\right) \tag{16.67}$$

where ζ is the natural coordinate on the boundary. Equation (16.67) actually follows from Eq. (16.56), when the fixed coordinate of a boundary is substituted at one boundary. Note that in natural coordinate system, at a boundary of an element one coordinate is always fixed.

16.7 FINITE ELEMENT EQUATIONS

Equation (16.53) has to be expressed in matrix form to obtain the FEM equations. For that purpose, we define

$$\{\dot{\varepsilon}\} = \begin{Bmatrix} \dot{\varepsilon}_{11} \\ \dot{\varepsilon}_{22} \\ \sqrt{2}\,\dot{\varepsilon}_{12} \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial v_1}{\partial x_1} \\ \dfrac{\partial v_2}{\partial x_2} \\ \dfrac{1}{\sqrt{2}}\left(\dfrac{\partial v_1}{\partial x_2} + \dfrac{\partial v_2}{\partial x_1}\right) \end{Bmatrix} \tag{16.68}$$

and

$$\{\dot{\varepsilon}(w)\} = \begin{Bmatrix} \dot{\varepsilon}_{11}(w) \\ \dot{\varepsilon}_{22}(w) \\ \sqrt{2}\,\dot{\varepsilon}_{12}(w) \end{Bmatrix} = \begin{Bmatrix} \dfrac{\partial w_1}{\partial x_1} \\ \dfrac{\partial w_2}{\partial x_2} \\ \dfrac{1}{\sqrt{2}}\left(\dfrac{\partial w_1}{\partial x_2} + \dfrac{\partial w_2}{\partial x_1}\right) \end{Bmatrix} \tag{16.69}$$

The strain rate vectors $\{\dot{\varepsilon}\}$ and $\{\dot{\varepsilon}(w)\}$ contain the first derivatives of velocity components and weights, respectively. They can be approximated as

$$\{\dot{\varepsilon}\} = [B]\{v^e\} \quad \text{and} \quad \{\dot{\varepsilon}(w)\} = [B]\{w_v^e\} \tag{16.70}$$

where

$$[B] = \begin{bmatrix} \frac{\partial N_1}{\partial x_1} & 0 & \frac{\partial N_2}{\partial x_1} & 0 & \cdots & \frac{\partial N_9}{\partial x_1} & 0 \\ 0 & \frac{\partial N_1}{\partial x_1} & 0 & \frac{\partial N_2}{\partial x_1} & \cdots & 0 & \frac{\partial N_9}{\partial x_1} \\ \frac{1}{\sqrt{2}}\frac{\partial N_1}{\partial x_2} & \frac{1}{\sqrt{2}}\frac{\partial N_1}{\partial x_1} & \frac{1}{\sqrt{2}}\frac{\partial N_2}{\partial x_2} & \frac{1}{\sqrt{2}}\frac{\partial N_2}{\partial x_1} & \cdots & \frac{1}{\sqrt{2}}\frac{\partial N_9}{\partial x_2} & \frac{1}{\sqrt{2}}\frac{\partial N_9}{\partial x_1} \end{bmatrix} \tag{16.71}$$

Further, we can write

$$\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22} = \{1 \;\; 1 \;\; 0\}\{\dot{\varepsilon}\} = \{m\}^{\mathrm{T}}[B]\{v^e\} \tag{16.72}$$

where

$$\{m\} = \begin{Bmatrix} 1 \\ 1 \\ 0 \end{Bmatrix} \tag{16.73}$$

Similarly

$$\dot{\varepsilon}_{11}(w) + \dot{\varepsilon}_{22}(w) = \{m\}^{\mathrm{T}}[B]\{w_v^e\} = \{w_v^e\}^{\mathrm{T}}[B]^{\mathrm{T}}\{m\} \tag{16.74}$$

Substituting Eqs (16.70), (16.72), and (16.74) in Eq. (16.52), we obtain

$$\begin{aligned} &\int_A -\{w_p^e\}^{\mathrm{T}}\{N^p\}\{m\}^{\mathrm{T}}[B]\{v^e\}\mathrm{d}x_1\mathrm{d}x_2 \\ &\quad + \int_A -\{w_v^e\}^{\mathrm{T}}[B]^{\mathrm{T}}\{m\}\{N^p\}^{T}\{p^e\}\mathrm{d}x_1\mathrm{d}x_2 \\ &\quad + \int_A 2\mu\{w_v^e\}^{\mathrm{T}}[B]^{\mathrm{T}}[B]\{v^e\}\mathrm{d}x_1\mathrm{d}x_2 = \int_{\Gamma_1} t_1 w_1 \mathrm{d}\Gamma + \int_{\Gamma_2} t_2 w_2 \mathrm{d}\Gamma \end{aligned} \tag{16.75}$$

It is convenient to use the following notations:

$$\begin{aligned} \left[K_{pv}^e\right] &= \int_A -\{N^p\}\{m\}^{\mathrm{T}}[B]\mathrm{d}x_1\mathrm{d}x_2 \\ \left[K_{vp}^e\right] &= \int_A -[B]^{\mathrm{T}}\{m\}\{N^p\}^{\mathrm{T}}\mathrm{d}x_1\mathrm{d}x_2 = \left\{K_{pv}^e\right\}^{\mathrm{T}} \\ \left[K_{vv}^e\right] &= \int_A 2\mu[B]^{\mathrm{T}}[B]\mathrm{d}x_1\mathrm{d}x_2 \end{aligned} \tag{16.76}$$

At this stage, note that if instead of the weak form given by Eq. (16.34), we would have taken the following weak form:

$$\int_A \left[-(\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22}) w_p + \left(\frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2} \right) w_1 + \left(\frac{\partial \sigma_{21}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} \right) w_2 \right] \mathrm{d}x_1 \mathrm{d}x_2 = 0 \tag{16.77}$$

we would have got

$$[K^e_{vp}] = -[K^e_{pv}]^{\mathrm{T}} \tag{16.78}$$

Thus, the system of equations would not have been symmetric.

Substituting Eqs (16.63) to (16.66) in Eq. (16.75) and assembling such expressions for all elements, the finite element equations for the system become

$$\sum_{e=1}^{ne} \left\{ w^e \right\}^{\mathrm{T}} \left[K^e \right] \left\{ \delta^e \right\} = \sum_{b=1}^{nb_1} \left\{ w_1^b \right\}^{\mathrm{T}} \left\{ f_1^b \right\} + \sum_{b=1}^{nb_2} \left\{ w_2^b \right\}^{\mathrm{T}} \left\{ f_2^b \right\} \tag{16.79}$$

where

$$\begin{aligned} \left\{ f_1^b \right\} &= \int_{\Gamma_1^b} \left\{ N_b \right\} \left\{ N_b \right\}^{\mathrm{T}} \left\{ t_1^b \right\} \mathrm{d}\Gamma \\ \left\{ f_2^b \right\} &= \int_{\Gamma_2^b} \left\{ N_b \right\} \left\{ N_b \right\}^{\mathrm{T}} \left\{ t_2^b \right\} \mathrm{d}\Gamma \end{aligned} \tag{16.80}$$

ne is the number of elements, nb_1 and nb_2 are the number of boundary elements on Γ_1 and Γ_2 and

$$\left\{ w^e \right\} = \begin{Bmatrix} \left\{ w_p^e \right\} \\ \left\{ w_v^e \right\} \end{Bmatrix}, \left\{ \delta^e \right\} = \begin{Bmatrix} \left\{ p^e \right\} \\ \left\{ v^e \right\} \end{Bmatrix} \text{ and } \left[K^e \right] = \begin{bmatrix} [0] & \left[K^e_{pv} \right] \\ \left[K^e_{vp} \right] & \left[K^e_{vv} \right] \end{bmatrix} \tag{16.81}$$

For the purpose of numerical evaluation, the variables of the area integrals in Eq. (16.76) are transformed to the natural coordinates (ξ, η) using the following transformation:

$$\int_{A^e} (\ldots) \mathrm{d}x_1 \mathrm{d}x_2 = \int_{-1}^{+1} \int_{-1}^{+1} (\ldots) \mid J \mid \mathrm{d}\xi \mathrm{d}\eta \tag{16.82}$$

where $|J|$ is the determinant of the Jacobian and is given by

$$|J| = \begin{vmatrix} \dfrac{\partial x_1}{\partial \xi} & \dfrac{\partial x_2}{\partial \xi} \\ \dfrac{\partial x_1}{\partial \eta} & \dfrac{\partial x_2}{\partial \eta} \end{vmatrix} \tag{16.83}$$

Similarly, the boundary integrals are transformed by the relation

$$\int_{\Gamma} (\ldots) \mathrm{d}\Gamma = \int_{-1}^{+1} (\ldots) \mid J_b \mid \mathrm{d}\zeta \tag{16.84}$$

where $|J_b|$ is the Jacobian for boundary element and is given by

$$\mid J_b \mid = \sqrt{\left(\frac{\mathrm{d}x_1}{\mathrm{d}\zeta} \right)^2 + \left(\frac{\mathrm{d}x_2}{\mathrm{d}\zeta} \right)^2} \tag{16.85}$$

Along the boundary, the geometry is approximated using one-dimensional quadratic shape functions. All elemental matrices are evaluated using 3×3 Gauss quadrature. Similarly, the elemental vectors are evaluated using three-point Gauss quadrature.

Equation (16.79) can be used to obtain finite element equations of the system. For that purpose, we take the finite element weighted residual form for a typical element as given below.

$$\left\{w^e\right\}^{\mathrm{T}}\left[K^e\right]\left\{\delta^e\right\} = \left\{w_1^b\right\}^{\mathrm{T}}\left\{f_1^b\right\} + \left\{w_2^b\right\}^{\mathrm{T}}\left\{f_2^b\right\} \tag{16.86}$$

In Eq. (16.86), terms $\{w_1^b\}^{\mathrm{T}}\{f_1^b\}$ and $\{w_2^b\}^{\mathrm{T}}\{f_2^b\}$ need to be expressed in the form containing $\{w^e\}$. It is easy to do that. For example, for the sake of simplicity, consider that $\{w_1^b\}$ is $\{w_2 \;\; w_3\}^{\mathrm{T}}$, is $\{f_2 \;\; f_3\}^{\mathrm{T}}$ and $\{w^e\}$ is $\{w_1 \;\; w_2 \;\; w_3 \;\; w_4\}^{\mathrm{T}}$. Then

$$\left\{w_1^b\right\}^{\mathrm{T}}\{f_1^b\} = w_2 f_2 + w_3 f_3 = \lfloor w_1 \;\; w_2 \;\; w_3 \;\; w_4 \rfloor \begin{Bmatrix} 0 \\ f_2 \\ f_3 \\ 0 \end{Bmatrix} = \{w^e\}^{\mathrm{T}} \begin{Bmatrix} 0 \\ f_2 \\ f_3 \\ 0 \end{Bmatrix}$$

Once all the terms contain the arbitrary term $\{w^e\}$, it can be removed to obtain the elemental equations. These elemental equations will contain the term corresponding to tractions on the internal surfaces of the element. However, during assembly, such terms will nullify each other and ultimately the right-hand-side vector will contain a large number of zeros, with non-zero terms corresponding to boundary nodes. These non-zero terms get tackled by boundary conditions.

Alternatively, the assembled finite element weighted residual form can be written as

$$\{W\}^T[K]\{\Delta\} = \{W\}^T\{F\} \tag{16.87}$$

where $\{W\}$, $[K]$, $\{\Delta\}$ represent the global vector of nodal values of weight function, global (assembled) coefficient matrix, and global vector of nodal values of pressure and velocity, respectively. $\{F\}$ denotes the global right hand side vector. Since weight functions are arbitrary, the final FEM expression will be

$$[K]\{\Delta\} = \{F\} \tag{16.88}$$

Some readers may have difficulty in visualizing how Eq. (16.87) is obtained from Eq. (16.79). It is briefly explained in the following paragraph.

Consider that nine nodes per element are chosen for velocity components and four nodes for pressure. Thus, in an element, there are 22 degrees of freedom. The coefficient matrix $[K^e]$ has a size of 22×22, $\{\delta^e\}$ has a size of 22×1, and $\{w^e\}^{\mathrm{T}}$ has a size of 1×22. Thus, the matrix multiplication $\{w^e\}^{\mathrm{T}}[K^e]\{\delta^e\}$ yields a matrix of size 1×1, i.e., one value. Assume that total global degrees of freedom are 576. The reader is familiar from the first chapter that an element matrix can be expressed in the global form. In the global form, the entries of elemental coefficient matrix are placed at appropriate locations of the matrix with the remaining locations having zeros. If $\{\Delta\}$ represents 576 global degrees of freedom and $\{W\}$ represents 576 nodal weights (including 22 elemental weights), it can be easily seen that

$$\{w^e\}^{\mathrm{T}}[K^e]\{\delta^e\} = \{W\}[K^e]_g\{\Delta\} \tag{16.89}$$

where $[K^e]_g$ is the elemental matrix written in global form and is of size 576×576. Therefore

$$\sum_{e=1}^{ne} \left\{w^e\right\}^T \left[K^e\right]\left\{\delta^e\right\} = \{W\}\sum_{e=1}^{ne}\left[K^e\right]_g \{\Delta\} \tag{16.90}$$

However

$$\sum_{e=1}^{ne}\left[K^e\right]_g = [K] \tag{16.91}$$

where $[K]$ is global stiffness matrix. Thus

$$\sum_{e=1}^{ne}\left\{w^e\right\}^T \left[K^e\right]\left\{\delta^e\right\} = \{W\}[K]\{\Delta\} \tag{16.92}$$

In the similar manner, it can be shown that

$$\sum_{b=1}^{nb_1}\left\{w_1^b\right\}^T \left\{f_1^b\right\} + \sum_{b=1}^{nb_2}\left\{w_2^b\right\}^T \left\{f_2^b\right\} = \{W\}^T\{F\} \tag{16.93}$$

16.8 DISCRETIZATION OF STRIP

In finite element modelling of drawing process where a curved die has been chosen, mesh generation can be done as shown in the Fig. 16.4. The mesh selected for the current problem consists of 56 elements. More elements have been selected in the deformation zone. This is the pre-processing stage of the problem. For velocity approximation, a nine-noded element is chosen, while for pressure approximation, a four-noded element is selected.

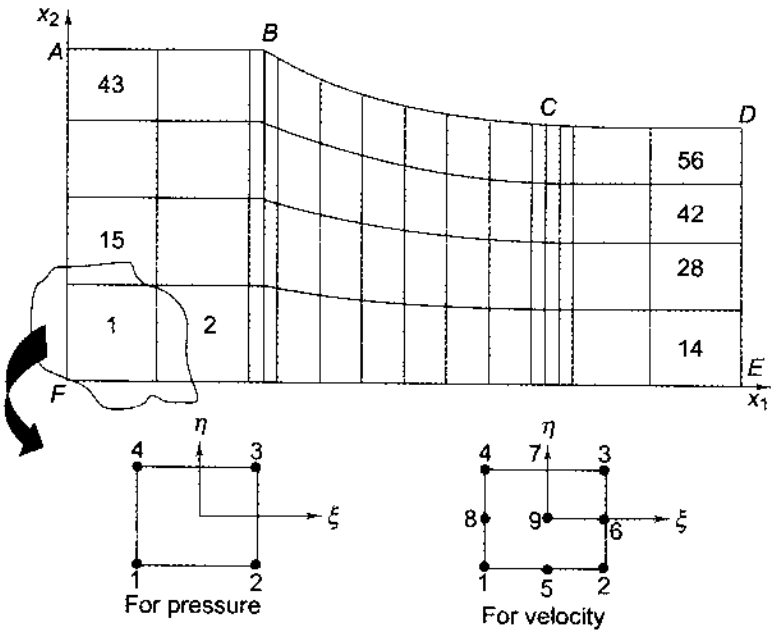


Figure 16.4 *A typical 56 element mesh and nodal structure for pressure and velocity*

16.9 APPLICATION OF BOUNDARY CONDITIONS

The friction conditions at a typical node on the tool–work interface (Fig. 16.5) is given by Eq. (16.17).

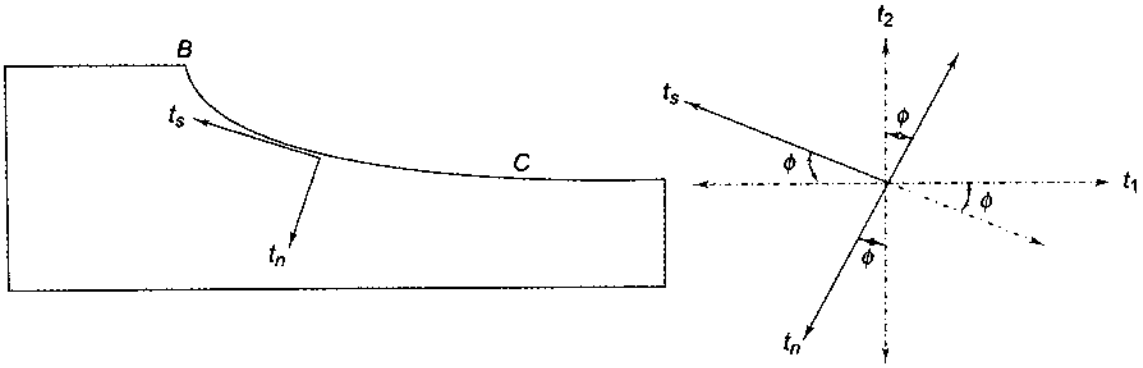


Figure 16.5 *Shear and normal components of traction at tool–work interface of a strip*

Using the Eq. (16.17), we can express ts and tn in the form of t_1 and t_2 as follows:

$$t_s = -t_1 \cos\phi + t_2 \sin\phi$$
$$t_n = -t_1 \sin\phi - t_2 \cos\phi \tag{16.94}$$

Thus, putting the decomposed form t_s and t_n into Eq. (16.17), we obtain the following relations:

$$-t_1 (\cos\phi + t_2 \sin\phi) = \mu(-t_1 \sin\phi - t_2 \cos\phi) \tag{16.95}$$

or

$$(-t_1 \cos\phi - \mu \sin\phi) + t_2 (\sin\phi + \mu \cos\phi) = 0 \tag{16.96}$$

or

$$t_1(1 + \mu \tan\phi) - t_2(\tan\phi + \mu) = 0 \tag{16.97}$$

The expression given in Eq. (16.97) for an element of FEM can be written as

$$\begin{Bmatrix} \left(t_b^1\right)_1 \\ \left(t_b^2\right)_1 \\ \left(t_b^3\right)_1 \end{Bmatrix} (1 + \mu \tan\phi) - \begin{Bmatrix} \left(t_b^1\right)_2 \\ \left(t_b^2\right)_2 \\ \left(t_b^3\right)_2 \end{Bmatrix} (\tan\phi + \mu) = 0 \tag{16.98}$$

After multiplying Eq. (16.98) by

$$\int_{-1}^{+1} \{N_b\}\{N_b\}^{\mathrm{T}} \mid J_b \mid \mathrm{d}\zeta \tag{16.99}$$

it becomes

$$\{f_1^b\}(1 + \mu \tan\phi) - \{f_2^b\}(\tan\phi + \mu) = \{0\} \tag{16.100}$$

The expression given in Eq. (16.100) holds good at all nodes of the element. At the middle node, say k (global node number), there is no contribution from the neighbouring elements and therefore, in terms of the global right-hand side of the vector, Eq. (16.100) can be expressed as

$$\{F\}_{(d_p+2k-1)}(1+\mu\tan\phi)-\{F\}_{(d_p+2k)}(\tan\phi+\mu)=\{0\} \tag{16.101}$$

where d_p is the total number of pressure variables. For end nodes of the element also, the equation will be similar. In that case, the contribution from the two elements will be added.

The procedure for applying the condition given in Eq. (16.101) at the node k is as follows:

- Replace $(d_p + 2k - 1)$th row of the global coefficient matrix $[K]$ by the linear combination $(1 + \mu \tan \phi) \times (d_p + 2k - 1)^{\text{th}}$ row of $[K] - (\tan\phi + \mu) \times \{(d_p + 2k)$th row of $[K]$.
- Make the $(d_p + 2k - 1)$th row of global right-hand-side vector $\{F\}$ zero.
  The velocity boundary condition at the node is applied by replacing $(d_p + 2k)$th row of $[K]$ by Eq. (16.16). For that the following steps are carried out:
- Make $(d_p + 2k)$th row of $[K]$ zero.
- Set $(d_p + 2k)d_p + 2k - 1)$th element of $[K]$ to $\tan\phi$ and $(d_p + 2k, d_p + 2k)$th element of $[K]$ to 1.
- Make $(d_p + 2k)$th row of $\{F\}$ to zero.

The essential boundary conditions at the other boundaries are applied as follows:

- In the stiffness matrix, all the elements of the rows and columns corresponding to the specified degree of freedom except the diagonal term are made equal to zero.
- The diagonal term is replaced by unity.
- The right-hand-side vector element corresponding to the specified degree of freedom is replaced by the specified value and other elements are modified by subtracting from them the products of corresponding coefficient element and specified value.

After imposing the boundary conditions, the resulting non-linear algebraic equations are solved using a method suitable to handle ill-conditioned system of equations. The Householder method is one such method. The solution is obtained in the form of primary variables, i.e., nodal pressure and velocity.

16.10 POST-PROCESSING

The evaluation of the secondary variables such as drawing force, die pressure, separation force and strain, strain-rate contours are termed as post-processing of the finite element method. There are certain precautions to be followed in post-processing step. Drawing force can be calculated by integrating the stresses across the thickness or dividing the total power by the velocity. The latter method provides a more accurate solution because errors get subdued due to integration. Stresses should preferably be calculated at Gauss points corresponding to two-Gauss-point formula. Interfacial tangential stress should be calculated by multiplying the coefficient of friction to interfacial normal stress. This way the error will be less than if they have been obtained directly from the stresses.

16.11 FEM FORMULATION OF FLUID FLOW WITH INERTIA EFECT

Navier–Stokes equations for two-dimensional flow of viscous and incompressible fluids are given by

$$v_1 \frac{\partial v_1}{\partial x_1} + v_2 \frac{\partial v_1}{\partial x_2} = -\frac{1}{\rho}\frac{\partial p}{\partial x_1} + \nu\left(\frac{\partial^2 v_1}{\partial x_1^2} + \frac{\partial^2 v_1}{\partial x_2^2}\right) \tag{16.102}$$

$$v_1 \frac{\partial v_2}{\partial x_1} + v_2 \frac{\partial v_2}{\partial x_2} = -\frac{1}{\rho}\frac{\partial p}{\partial x_2} + \nu\left(\frac{\partial^2 v_2}{\partial x_1^2} + \frac{\partial^2 v_2}{\partial x_2^2}\right) \tag{16.103}$$

Here, ρ is the density and ν is the kinematic viscosity, which is assumed to be constant, throughout. The continuity equations are given by

$$\frac{\partial v_1}{\partial x_1} + \frac{\partial v_2}{\partial x_2} = 0 \tag{16.104}$$

In index notations, Eqs (16.102) and (16.103) are written as

$$v_j v_{i,j} = -\frac{1}{\rho} p_{,i} + \nu v_{i,jj} \tag{16.105}$$

Similarly, Eq. (16.104) is written as

$$v_{i,i} = 1 \tag{16.106}$$

We shall modify Eq. (16.105) in order to make the form of the equation similar to Eq. (16.5). In view of Eq. (16.106), we can write

$$v_{j,ji} = v_{j,ij} = 0 \tag{16.107}$$

Equation (16.105) can be modified as

$$\rho v_j v_{i,j} = \quad p_{,i} + \mu v_{i,jj} + \mu v_{j,ij} \tag{16.108}$$

Making use of the substitution property of Kronecker's delta, Eq. (16.108) can be written as

$$\rho v_j v_{i,j} = \left(-p\delta_{ij} + \mu v_{i,j} + \mu v_{j,i}\right)_{,j} \tag{16.109}$$

Using Eq. (16.2) in Eq. (16.109), we obtain

$$\rho v_j v_{i,j} = \left(-p\delta_{ij} + 2\mu\dot{\varepsilon}_{ij}\right)_{,j} \tag{16.110}$$

Using Eq. (16.31), we can write Eq. (16.110) as

$$\rho v_j v_{i,j} = \sigma_{ij,j} \tag{16.111}$$

Making the integration of weighed residual of Eq (16.106) and (16.111) over the domain of area A equal to zero, we get

$$I = \int_A \left(w_p v_{i,i} + w_i \sigma_{ij,j} - w_i \rho v_j v_{i,j}\right) \mathrm{d}A \tag{16.112}$$

where w_p, w_1, and w_2 are the weight functions. In Eq. (16.112), we shall reduce the order of derivative of the middle term as was done in strip-drawing problem. If we try to reduce the order of derivative in the first and third terms, then although there will be no derivative on velocity, there will be an introduced derivative term on the weight function. This may be tolerable; however, there will be some terms on the boundary without a physical meaning. Hence, we reduce only the middle term. With this, boundary terms containing the traction will be obtained.

The weak form of weighted residual is written as

$$\int_A I_1 \, dA + \int_A I_2 \, dA + \int_A I_3 dA - \int_{\Gamma_1} I_4 d\Gamma - \int_{\Gamma_2} I_5 d\Gamma = 0 \tag{16.113}$$

where

$$\begin{aligned}
I_1 &= -\varepsilon_{ii} w_p = -\left[\dot{\varepsilon}_{11} + \dot{\varepsilon}_{22}\right] w_p \\
I_2 &= -p\left[\dot{\varepsilon}_{11}(w) + \dot{\varepsilon}_{22}(w)\right] + 2\mu\left[\dot{\varepsilon}_{11}\dot{\varepsilon}_{11}(w) + 2\dot{\varepsilon}_{12}\dot{\varepsilon}_{12}(w) + \dot{\varepsilon}_{22}\dot{\varepsilon}_{22}(w)\right] \\
I_3 &= \rho\left[\left(v_1 \frac{\partial v_1}{\partial x_1} + v_2 \frac{\partial v_1}{\partial x_2}\right) w_1 + \left(v_1 \frac{\partial v_2}{\partial x_1} + v_2 \frac{\partial v_2}{\partial x_2}\right) w_2\right] \\
I_4 &= t_1 w_1 \\
I_5 &= t_2 w_2
\end{aligned} \tag{16.114}$$

As a result of finite element approximation, the equation obtained will be similar to that obtained for the strip-drawing problem. Only the submatrix $[K_{vv}]$ will be different. It is given by

$$[K_{vv}]^e = \int_A \left(\rho v_1 [N_v]^T [B_x] + \rho v_2 [N_v]^T [B_y] + 2\mu [B]^T [B]\right) dA \tag{16.115}$$

where

$$[B_x] = \begin{bmatrix} \frac{\partial N_1}{\partial x} & 0 & \frac{\partial N_2}{\partial x} & 0 & \ldots\ldots & \frac{\partial N_9}{\partial x} & 0 \\ 0 & \frac{\partial N_1}{\partial x} & 0 & \frac{\partial N_2}{\partial x} & 0 \ldots & 0 & \frac{\partial N_9}{\partial x} \end{bmatrix} \tag{16.116}$$

and

$$[B_y] = \begin{bmatrix} \frac{\partial N_1}{\partial y} & 0 & \frac{\partial N_2}{\partial y} & 0 & \ldots\ldots & \frac{\partial N_9}{\partial y} & 0 \\ 0 & \frac{\partial N_1}{\partial y} & 0 & \frac{\partial N_2}{\partial y} & 0 \ldots & 0 & \frac{\partial N_9}{\partial y} \end{bmatrix} \tag{16.117}$$

16.12 CONCLUSION

The flow formulation has been extensively used for the analysis of metal-forming process considering the material as rigid plastic. The steady-state drawing process has been solved using mixed pressure–velocity

FE formulation. In the mixed pressure–velocity finite element formulation, no pressure boundary conditions are employed and therefore the pressure field is computed in this method needs further refinement. FEM is preferred to other methods in the analysis of drawing process, as it can easily incorporate non-homogeneity of deformation, process-dependent material properties and different friction models. Finally, formulation of fluid flow problem with inertia effect is carried out.

SOLVED PROBLEMS

16.1 Per unit volume power dissipated due to plastic deformation is $\sigma_{ij}\dot{\varepsilon}_{ij}$. Considering the material as incompressible, prove that

$$\sigma_{ij}\dot{\varepsilon}_{ij} = S_{ij}\dot{\varepsilon}_{ij}$$

where S_{ij} is the deviatoric component of σ_{ij}.

Solution Using Eq. (16.30), we can write

$$\sigma_{ij}\dot{\varepsilon}_{ij} = (-p\delta_{ij} + S_{ij})\dot{\varepsilon}_{ij} = -p\dot{\varepsilon}_{ij} + S_{ij}\dot{\varepsilon}_{ij}$$

Now remember that δ_{ij} is non-zero (and equal to 1) only if i and j are equal. Therefore

$$\dot{\varepsilon}_{ij}\delta_{ij} = \dot{\varepsilon}_{11}\delta_{11} + \dot{\varepsilon}_{22}\delta_{22} + \dot{\varepsilon}_{33}\delta_{33} = \dot{\varepsilon}_{11} + \dot{\varepsilon}_{22} + \dot{\varepsilon}_{33} = \dot{\varepsilon}_{ii}$$

For incompressible material, $\dot{\varepsilon}_{ii} = 0$

Hence

$$\sigma_{ij}\dot{\varepsilon}_{ij} = S_{ij}\dot{\varepsilon}_{ij}$$

16.2 The velocity gradient $v_{i,j}$ at a point is represented in the matrix form as

$$\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} 10^{-4}\,\mathrm{s}^{-1}$$

Find out the strain-rate tensor and its deviatoric part.

Solution If $v_{i,j}$ is represented by the matrix given in the problem, $v_{i,j}$ will be represented by its transpose. Converting Eq. (16.2) in the matrix form and substituting the values of velocity gradient and its transpose we get strain-rate as

$$[\dot{\varepsilon}] = \frac{1}{2}\left(\begin{bmatrix} 2 & 3 \\ 4 & 6 \end{bmatrix} + \begin{bmatrix} 2 & 4 \\ 3 & 6 \end{bmatrix}\right)10^{-4} = \begin{bmatrix} 2 & 3.5 \\ 3.5 & 6 \end{bmatrix}10^{-4} \tag{16.2.1}$$

Writing Eq. (16.2.1) in matrix form, we get

$$[\dot{\varepsilon}] = \frac{\dot{\varepsilon}_{kk}}{3}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + [\dot{\varepsilon}'] \tag{16.2.2}$$

Note that it is essential to use a three-dimensional form for computing the deviatoric strain-rate. In three-dimensional form, Eq. (16.2.1) is written as

$$[\dot{\varepsilon}] = \begin{bmatrix} 2 & 3.5 & 0 \\ 3.5 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} 10^{-4} \tag{16.2.3}$$

Now

$$\dot{\varepsilon}_{kk} = \dot{\varepsilon}_{11} + \dot{\varepsilon}_{22} + \dot{\varepsilon}_{33} = (2 + 6 + 0) \times 10^{-4} = 8 \times 10^{-4} \tag{16.2.4}$$

Substituting Eqs (16.2.3) and (16.2.4) in Eq. (16.2.2), we get

$$\begin{bmatrix} 2 & 3.5 & 0 \\ 3.5 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} 10^{-4} = \frac{8}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} 10^{-4} + [\dot{\varepsilon}']$$

From this, we get the value of deviatoric strain-rate as

$$[\dot{\varepsilon}'] = \begin{bmatrix} 2 & 3.5 & 0 \\ 3.5 & 6 & 0 \\ 0 & 0 & 0 \end{bmatrix} 10^{-4} - \frac{8}{3} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} 10^{-4} = \begin{bmatrix} -2/3 & 3.5 & 0 \\ 3.5 & 10/3 & 0 \\ 0 & 0 & -8/3 \end{bmatrix}$$

EXERCISES

A. Multiple Choice Questions

16.1 Which of the following is the most appropriate statement about a metal-forming process?

(a) During the process, metal flows like a Newtonian fluid.

(b) While solving the metal-forming problem by flow formulation, it is usual to assume that the material is incompressible.

(c) Metal-forming processes are always of plane strain type.

(d) FEM flow formulation of metal-forming problems provides linear equations.

16.2 If v_i is the component of the velocity field of a fluid in index notation, then the strain rate is given by

(a) $v_{i,j}$ (b) $v_{i,i}$

(c) $\frac{1}{2}(v_{i,j} + v_{j,i})$ (d) $\frac{1}{2}(v_{i,j} - v_{j,i})$

16.3 Which of the following expressions is equal to zero?

(a) $\frac{1}{2}\sigma_{ij}(v_{i,j} + v_{j,i})$ (b) $\frac{1}{2}\sigma_{ij}(v_{i,j} - v_{j,i})$

(c) $\sigma_{ij}\dot{\varepsilon}_{ij}$ (d) $\sigma_{ij}\sigma_{ij}$

16.4 At the line of symmetry of a cold rolling process,

(a) normal (to the line of symmetry) components of the traction and velocity are zero.
(b) the normal component of velocity and the tangential component of traction are zero.
(c) the tangential component of velocity and the normal component of traction are zero.
(d) tangential components of the traction and velocity are zero.

16.5 The velocity (in m/s) of a particle as a function of x is given as

$$v = 5 - x$$

where the position x is measured in meters. At $x = 1$, the acceleration is equal to

(a) 0 (b) -1
(c) 4 (d) -4

B. Problems

16.6 In the rolling process, the governing equations are given as

$$\bar{\dot{\varepsilon}}_{11} + \bar{\dot{\varepsilon}}_{22} = 0$$

$$\frac{\partial \bar{\sigma}_{11}}{\partial \bar{x}_1} + \frac{\partial \bar{\sigma}_{12}}{\partial \bar{x}_2} = 0$$

$$\frac{\partial \bar{\sigma}_{21}}{\partial \bar{x}_1} + \frac{\partial \bar{\sigma}_{22}}{\partial \bar{x}_2} = 0$$

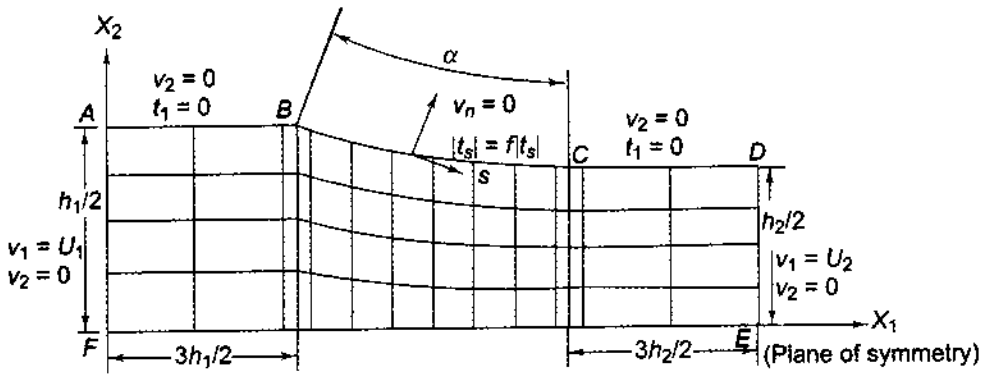


Figure 16.6 *Essential boundary conditions for the domain discussed in Problem 16.6*

The essential boundary conditions are shown in Fig. 16.6. The two natural boundary conditions at the roll-strip interface are given as

$$v_2 + v_1 \tan\phi = 0 \quad \text{and} \quad |t_s| = \mu |t_n|$$

where ϕ is the angular position of the point on the surface, μ is the Coloumb's coefficient of friction, and t_s and t_n are the tangential and normal components of the stress vector t. Due to the existence of neutral point, friction conditions will change. From point B to neutral point the

tangential traction on the strip is directed from left to right and afterwards its direction reverses. Explain how you will apply these boundary conditions accordingly in the FEM modelling of rolling process.

16.7 The governing equations for the steady, constant volume flow in axisymmetric problem is given as

$$\dot{\varepsilon}_{rr} + \dot{\varepsilon}_{\theta\theta} + \dot{\varepsilon}_{zz} = 0$$

$$p\left(v_r \frac{\partial v_r}{\partial r} + v_z \frac{\partial v_r}{\partial z}\right) - \left(\frac{1}{r}\frac{\partial(r\sigma_{rr})}{\partial r} + \frac{\partial \sigma_{rz}}{\partial z} - \frac{\sigma_{\theta\theta}}{r}\right) = 0$$

$$p\left(v_r \frac{\partial v_z}{\partial r} + v_z \frac{\partial v_z}{\partial z}\right) - \left(\frac{1}{r}\frac{\partial(r\sigma_{rz})}{\partial r} + \frac{\partial \sigma_{zz}}{\partial z}\right) = 0$$

The boundary conditions for the domain of the present problem are shown in Fig. 16.7.

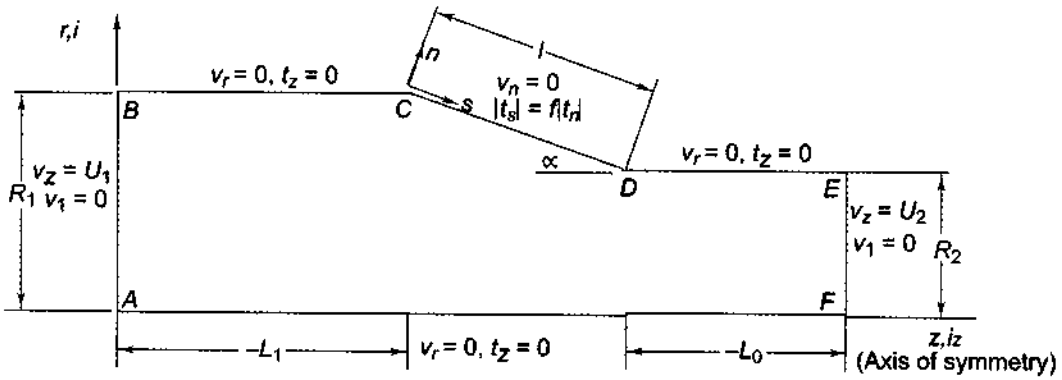


Figure 16.7 *Boundary conditions for the system discussed in Problem 16.7 [Dixit and Dixit (1995)].*

Carry out the FEM modelling of this problem using these governing equations and boundary conditions

16.8 The governing equation of a spring is given as

$$F = (k_0 + kx)x$$

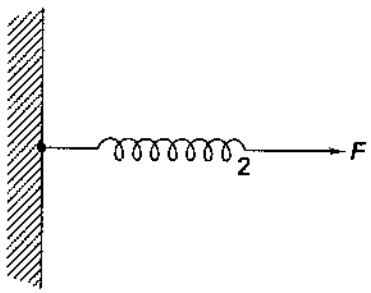


Figure 16.8 *Force acting on a spring*

where $k_0 = 1000$ N/m, $k = 100$ N/m^2, $F = 2000$ N

The equations can be written as

$$k = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} -F \\ 0 \end{Bmatrix}$$

Solve the problem by finite element method.

16.9 The governing equation of a large deflection bending of elastic beam is

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left\{EA\left(\frac{\partial u}{\partial x} + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2\right)\right\} - f = 0$$

$$-\frac{\mathrm{d}^2}{\mathrm{d}x^2}\left\{EI\frac{\partial^2 w}{\partial x^2}\right\} - \frac{\mathrm{d}}{\mathrm{d}x}\left\{EA\left(\frac{\partial u}{\partial x} + \frac{1}{2}\left(\frac{\partial w}{\partial x}\right)^2\right)\right\} - q = 0$$

where u is the longitudinal displacement, w is the transverse deflection, E is modulus of elasticity, A is the cross-sectional area, and f and q are the axial distributed and transverse loading, respectively. Carry out the FEM formulation to solve this problem.

REFERENCES

Dixit, US, and Dixit, PM, 1995, 'An analysis of the steady-state wire drawing of the strain hardening materials', *J. Mater. Process.* Technol., 47, pp.201–229.

Dixit, US, and Dixit, PM, 2008, *Modeling of Metal Forming and Machining Processes By Finite Element and Soft Computing Methods*, Springer–Verlag, London.

17

CHAPTER

Error Analysis in Finite Element Method

17.1 Introduction

17.2 Error Measures

17.3 Types of Error Estimates

17.4 Conclusion

17.1 INTRODUCTION

Finite element method (FEM) provides an approximate solution of differential equations. There is a need to calculate the error, which is defined as the difference between the exact solution and the approximate solution (i.e., FEM solution). Errors in finite element solutions are divided mainly into three categories:

(i) *Domain approximation error*, which is due to the approximation of the domain.

(ii) *Quadrature and finite arithmetic errors*, which are due to the numerical evaluation of integrals and the numerical computation on a computer.

(iii) *Solution approximation error*, which is due to the approximation of the solution.

In the formulation of FEM, usually the displacement or primary variable field is approximated by polynomials. This approximation is the main source of error in the solution. As this error is inherent in the method, the amount of error in the solution must be determined in order to judge the quality of the results obtained. This information regarding error can be used to improve the results (by improving the primary variable approximation), and to monitor the convergence of the solution. The errors decrease as the size of the subdivision (h) gets smaller or as the order of the polynomial (p) in the trial function used increases.

In this chapter, various error measures are described. Two common estimates of error are *a priori* and *a posteriori* error estimates; a discussion regarding these is presented in the following subsections. The recovery method of estimating error is also discussed in detail.

17.2 ERROR MEASURES

The subject of error estimation for finite element solutions and a consequent adaptive analysis, in which the approximation is successively refined to reach predetermined standards of accuracy, is central to the effective use of finite element codes for practical engineering analysis. The main problems in error estimation are the cost of computations and implementation of such computations into an existing code structure. Various error measures are explained with the help of the following example.

Consider the second-order differential equilibrium equation

$$\boldsymbol{Lu} - \boldsymbol{q} = \boldsymbol{S}^{\mathrm{T}}[D]\boldsymbol{Su} - \boldsymbol{q} = \boldsymbol{0} \tag{17.1}$$

in a domain Ω, with prescribed displacement

$$\boldsymbol{u} = \bar{\boldsymbol{u}} \text{ on the boundary } \Gamma_u \tag{17.2}$$

and prescribed derivatives

$$\boldsymbol{Su} = \bar{\boldsymbol{\varepsilon}} \text{ on the boundary } \Gamma_t \tag{17.3a}$$

with

$$\Gamma = \Gamma_u \cup \Gamma_t \tag{17.3b}$$

where $\boldsymbol{L}$ is linear second-order differential operator matrix, $\boldsymbol{S}$ is the first-order differential operator matrix, $\boldsymbol{q}$ is a constant vector, $[D]$ is an arbitrary matrix, and $\boldsymbol{u}$ is primary variable vector. The boldfaced letters indicate matrices or vectors. The following equation is an example of a one-dimensional system:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q=0 \tag{17.4}$$

with boundary conditions

$$\text{at } x=0,\ u=0;\ \ \text{at } x=l,\ \frac{\mathrm{d}u}{\mathrm{d}x}=0 \tag{17.5}$$

It can be seen that in this case

$$L=\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}}{\mathrm{d}x}\right),\ S=\frac{\mathrm{d}}{\mathrm{d}x}\ \text{ and }\ D=EA \tag{17.6}$$

In a finite element approximation, we obtain the approximating equations by a standard Galerkin process (or equivalently by minimizing the potential energy) to obtain

$$[K]\{u\}-\{f\}=0 \tag{17.7}$$

where $[K]=\int_{\Omega}(\boldsymbol{SN})^{\mathrm{T}}[D](\boldsymbol{SN})\mathrm{d}\Omega$ is the elemental stiffness matrix and $\{f\}=\int_{\Omega}\boldsymbol{N}^{\mathrm{T}}\boldsymbol{q}\,\mathrm{d}\Omega+\int_{\Gamma_t}\boldsymbol{N}^{\mathrm{T}}\bar{\boldsymbol{t}}\,\mathrm{d}\Gamma$ is the load vector. Here, $\boldsymbol{N}$ is the shape function (interpolation function) matrix. The approximate solution $\hat{\boldsymbol{u}}$ for an element is related to nodal solution $\boldsymbol{u}^{ne}$ in the following manner:

$$\hat{\boldsymbol{u}}=\boldsymbol{N}\,\boldsymbol{u}^{ne} \tag{17.8}$$

The derivatives are calculated as

$$\hat{\boldsymbol{\varepsilon}}=(\boldsymbol{SN})\boldsymbol{u}^{ne} \tag{17.9}$$

The approximate solutions $\hat{\boldsymbol{u}}$ and $\hat{\boldsymbol{\varepsilon}}$ differs from the exact values $\boldsymbol{u}$ and $\boldsymbol{\varepsilon}$ and the difference is defined as

. Thus, primary variable and derivative errors are

$$\boldsymbol{e}=\boldsymbol{u}-\hat{\boldsymbol{u}} \tag{17.10a}$$

$$\boldsymbol{e}_{\boldsymbol{\varepsilon}}=\boldsymbol{\varepsilon}-\hat{\boldsymbol{\varepsilon}} \tag{17.10b}$$

The specification of local error in Eqs (17.10a) and (17.10b) is generally inconvenient and occasionally misleading. For this reason, various 'norms' representing some integral quantity are often introduced to measure the error. The most common measures are the *energy norm* and $\mathbf{L}_2$ *norm*. The energy norm for general problems is defined as

$$\|\boldsymbol{e}\|=\left(\int_{\Omega}\boldsymbol{e}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{e}\,d\Omega\right)^{\frac{1}{2}} \tag{17.11}$$

where $\boldsymbol{e}$ is given by Eq. (17.10a).

A more direct measure is the $\mathbf{L_2}$ norm, which can be associated with the errors in any quantity. Thus for a displacement $\boldsymbol{u}$, the $\mathbf{L_2}$ norm of error $\boldsymbol{e}$ is written as

$$\|\boldsymbol{e}\|_{L_2} = \left(\int_{\Omega} \boldsymbol{e}^{\mathrm{T}} \boldsymbol{e}\, \mathrm{d}\Omega\right)^{\frac{1}{2}} \tag{17.12}$$

and for derivatives

$$\|\boldsymbol{e}_{\varepsilon}\|_{L_2} = \left(\int_{\Omega} (\boldsymbol{e}_{\varepsilon})^{\mathrm{T}} (\boldsymbol{e}_{\varepsilon})\, \mathrm{d}\Omega\right)^{\frac{1}{2}} \tag{17.13}$$

The *root mean square* (RMS) error for the whole domain Ω is given by

$$e_u^{rms} = \left(\frac{\|\boldsymbol{e}\|_{L_2}^2}{\boldsymbol{\Omega}}\right)^{\frac{1}{2}} \tag{17.14}$$

Similarly, for derivatives, the RMS error can be written as

$$e_{\varepsilon}^{rms} = \left(\frac{\|\boldsymbol{e}_{\varepsilon}\|_{L_2}^2}{\boldsymbol{\Omega}}\right)^{\frac{1}{2}} \tag{17.15}$$

The error for the whole domain is obtained by summing the element contributions. Thus

$$\|\boldsymbol{e}\|^2 = \sum_{i=1}^{m} \|\boldsymbol{e}\|_i^2 \tag{17.16}$$

where i represents the element number and m denotes the total number of elements. For an optimal mesh, it is desirable that the contribution to the square of the norm is equal for all elements. The relative percentage error η can be given by

$$\eta = \frac{\|\boldsymbol{e}\|}{(\text{strain energy of the element})^{1/2}} \times 100\% \tag{17.17}$$

The strain energy of the element is the integration of dot product of the strain ε and stress $[D]\,\varepsilon$ over the element.

17.3 TYPES OF ERROR ESTIMATES

The quantities that are used for error estimation in finite element analysis can be divided into two main categories: *a priori* error estimates and *a posteriori* estimates. Finding out the error before the solution is called *a priori* and after the solution is called *a posteriori*. Various error estimates are described in detail in this section.

17.3.1 A *priori* Error Estimates

In *a priori* error estimation, the effect of the proposed improved solution is predicted without actually determining the solution. A *priori* error estimates provide only a qualitative description on the rate of

convergence of the finite element solution. Hence, it is difficult to directly make use of this type of error estimates in a mesh-refinement process of adaptive mesh generation, which usually requires a quantitative description of the error distribution as input information. However, such error estimates provide an excellent tool for predicting the convergence rate during the adaptive refinement process.

Convergence during h-refinement of the mesh

For improving the accuracy of finite element solution, the mesh can be refined. One way of refinement is by reducing the size of the elements. This is called h-refinement. In h-refinement, the degree of polynomial interpolation function (p) is kept constant. This means that in h-refinement, number of elements of the same type (but reduced size) is increased in a given domain. If the mesh is refined uniformly and the size of the elements (h) approaches zero, the error estimate is given in the form

$$\|e\| \leq Ch^p \tag{17.18}$$

where C is a constant. Equation (17.18) assumes the absence of singularities. If singularities are present, due to abrupt change of cross-section in solid mechanics problems, etc., then the convergence rate will get slower. For one-dimensional cases, the error estimate can be derived from Taylor series expansion. For two-dimensional problems, h is approximately proportional to the inverse of the square root of the total degrees of freedom of the mesh. Hence

$$\|e\| \leq CN^{-\frac{p}{2}} \tag{17.19}$$

where N is the total number of degrees of freedom of the mesh. The mesh is called *optimal* if it is designed in such a way that the error is equally distributed in each element and the influence of the singularities is eliminated.

Convergence during p-refinement of the mesh

For p-version refinement, the degree of interpolation polynomial p is increased keeping the mesh size fixed. The error estimate is given by

$$\|e_u\| \leq CN^{-\beta} \tag{17.20}$$

where β is a positive constant dependent on smoothness of the exact solution and quality of the mesh. Hence, the rate of convergence will depend graetly on the design of the mesh. If a properly designed mesh is used, an exponential rate of convergence can be obtained. If the mesh is not well designed, the performance will be affected especially when singularities are present. The convergence rate of the p-refinement is generally better than that of the uniform h-refinement but care should be taken for the element size near singularities as oscillation of derivatives of solution may occur.

Convergence during hp-refinement of the mesh

In hp-refinement, the mesh size, h, is refined simultaneously with the increase of value of p. A properly designed hp-refinement scheme provides faster convergence than either p-refinement or h-refinement

scheme. However, the scheme is not useful if it is not well-designed. In *hp*-refinement scheme, the *h*-refinement can be used to isolate singularities, while *p*-refinement delivers exponential rate of convergence.

17.3.2 A *Posteriori* Error Estimates

In *a posteriori* error estimation, the error norms are based on already determined solution and hence the effect of increasing polynomial order or decreasing element size is known and not estimated as is done in *a priori* error estimation. In addition, because the error is found on a local basis and then summed globally, *a posteriori* error analyses allow for element level refinement, which is advantageous. There are many methods for the estimation of *a posteriori* error analysis, which are computationally inexpensive and simple to calculate. Owing to the reasons mentioned above, *a posteriori* error estimator has become more popular and is seen to be effective and helps in fast convergence during the refinement process.

Error estimates by recovery

Most of the *a posteriori* error estimators take a recovered solution as the exact solution for the estimation of error. The recovered solution is obtained from FEM solution and is expected to be more accurate than the FEM solution. There are a number of methods to obtain a recovered solution as we shall be studying in the following subsections. The recovered solutions can also be taken as the final result of an FEM analysis. Error estimators formulated by replacing the exact solution with the recovered solution are sometimes called *recovery-based error estimators*.

The accuracy or the quality of the error estimator is measured by the effectivity index θ, which is defined as the ratio between recovery-based error estimator and the actual error estimator. Ideally, the value of effectivity index should be 1. In practice, it may be less or more than 1. If the error norm based on the recovered solution is 90% of the actual error norm, then the effectivity index will be 0.9. Generally, as the mesh refinement progresses, effectivity index keeps on moving closer to 1.

Zienkiewicz-Zhu error estimate

Zienkiewicz and Zhu (1987) proposed an error estimate, which became popular as Zienkiewicz–Zhu (ZZ) error estimator. Later the same authors proposed another error estimator which is popular as *superconvergent patch recovery* (SPR) technique [Zienkiewicz and Zhu (1992)]. In this section, we shall briefly describe the ZZ method published in 1987. The main advantages of using ZZ error estimator over the other types of error estimators is the simplicity of its implementation and its cost effectiveness. In practical finite element computations, some smoothing procedure will always be employed at the post-processing stage of the computing process to recover the derivatives of the finite element solutions in order to achieve more acceptable approximations. Using such recovered derivatives, the ZZ error estimator can be calculated at a fraction of the total cost of the computation. However, the quality and the reliability of the error estimator is obviously dependent on the accuracy of the recovered solutions, and therefore, on the smoothing procedures.

To obtain acceptable results for stress, we generally resort to a nodal averaging or projection process in which it is assumed that the recovered derivative $\{\sigma^*\}$ is interpolated by the same function as the displacement

$$\{\sigma^*\} = [N]\{\bar{\sigma}^*\} \tag{17.21}$$

where $\{\sigma^*\}$ is improved nodal derivative and

$$\int_\Omega [N]^T \left(\{\sigma^*\} - \{\hat{\sigma}\}\right) d\Omega = 0 \tag{17.22}$$

where $\{\sigma^*\}$ is elemental FEM derivative. Substituting Eq. (17.21) in Eq. (17.22), we get

$$\int_\Omega [N]^T \left([N]\{\bar{\sigma}\}^* - \{\hat{\sigma}\}\right) d\Omega = 0 \tag{17.23}$$

Thus

$$\int_\Omega [N]^T [N] d\Omega \{\bar{\sigma}^*\} = \int_\Omega [N]^T \{\hat{\sigma}\} d\Omega \tag{17.24}$$

The system of equations given in Eq. (17.24) can be written as

$$[A]\{\bar{\sigma}^*\} = \{b\} \tag{17.25}$$

Equation (17.25) can be solved to obtain recovered nodal derivatives, from where the recovered derivative can be computed at any point using Eq. (17.21).

It is intuitively obvious that $\{\sigma^*\}$ is in fact a better approximation than $\{\hat{\sigma}\}$ and we shall use it to estimate the error $\{e_\sigma\}$ as follows

$$e_\sigma \approx \{\sigma^*\} - \{\hat{\sigma}\} \tag{17.26}$$

Equation (17.26) can be used for computing various error norms.

Residual method

For the differential equations given in Eq. (17.1), the residual is defined as

$$\boldsymbol{r} = \boldsymbol{L}\hat{\boldsymbol{u}} - \boldsymbol{q} = \boldsymbol{L}\hat{\boldsymbol{u}} - \boldsymbol{L}\boldsymbol{u} = \boldsymbol{L}(\hat{\boldsymbol{u}} - \boldsymbol{u}) \tag{17.27}$$

Using Eq. (17.10a), we can write

$$\boldsymbol{r} = -\boldsymbol{L}\boldsymbol{e} \tag{17.28}$$

The energy norm can then be defined as

$$\|\boldsymbol{e}\| = \left(-\int_\Omega \boldsymbol{e}^T \boldsymbol{r}\, d\Omega\right)^{\frac{1}{2}} \tag{17.29}$$

In FEM, residuals generally become infinite at the element interfaces. To avoid this difficulty, the domain is divided into interior element domain and element interface zones. Inside the element, no difficulty arises in determining residuals. At the element interface, the discontinuity in the first derivative of primary variables (known as *jump discontinuity*) forms the basis of error. Suitable weightage can be given to errors inside the element and at the interface for estimating the overall error. More details of various variants of residual method can be found in Zienkiewicz, et al. (2005).

Error estimator for hierarchic p-refinement

In the hierarchic p-refinement scheme, the degree of interpolation polynomial is increased by adding extra shape functions in the element, while preserving the previous shape functions. For example, for two-noded, one-dimensional C^0 continuity element, the interpolation function is given as

$$u = N_1 u_1 + N_2 u_2 \tag{17.30}$$

where shape functions in natural coordinates are written as

$$N_1 = \frac{1-\xi}{2}; \quad N_2 = \frac{1+\xi}{2} \tag{17.31}$$

The coordinates of the two nodes are -1 and $+1$ in natural coordinates. It is clear that the interpolation function given by Eq. (17.30) is linear. We can make a quadratic interpolation function by taking a three-noded Lagrangian element, in which all the shape functions will be polynomials of degree 2. However, such p-refinements will not be called hierarchic p-refinement. In hierarchic p-refinement, the first two shape functions are the same linear function, whereas the third shape function is a quadratic shape function. Thus, a suitable hierarchic interpolation polynomial may be written as

$$u = N_1 u_1 + N_2 u_2 + N_3 a \tag{17.32}$$

where

$$N_1 = \frac{1-\xi}{2}; \quad N_2 = \frac{1+\xi}{2}, \quad N_3 = (1-\xi^2) \tag{17.33}$$

Note that the constant a in Eq. (17.32) does not have the same physical meaning as u_1 and u_2. Equation (17.32) provides $u = u_1$ at node 1 and $u = u_2$ at node 2 as desired.

Zienkiewicz, et al. (1987) have discussed a procedure for error estimation procedure for hierarchic p-refinement scheme. In this method, we temporarily introduce some degrees of freedom associated with the shape functions of order $p + 1$ into the previous refinement system, where p is the current highest order of the shape function. Let us consider that the displacement field of an element is given by

$$\{u\} = [N]\{d\} \tag{17.34}$$

Consider that at the refinement step m, the element displacement fields of an element of order p are given by

$$\{u_p\} = [N_n]\{d_n\} \tag{17.35}$$

where $[N_n]$ contains all existing nodal and hierarchical shape functions and $\{d_n\}$ is the vector of known element displacements. The corresponding finite element equation is written as

$$[K_{nn}]\{d_n\} = \{f_n\} \tag{17.36}$$

If all new hierarchical shape functions of order $p + 1$, $[N_h]$, are temporarily added into the element, Eqs (17.35) and (17.36) are upgraded to

$$\{u_{p+1}\} = [N_n]\{d_n\} + [N_h]\{d_h\} \tag{17.37}$$

and

$$\begin{bmatrix} [K_{nn}] & [K_{nh}] \\ [K_{hn}] & [K_{hh}] \end{bmatrix} \begin{Bmatrix} \{d_n\} \\ \{d_h\} \end{Bmatrix} = \begin{Bmatrix} \{f_n\} \\ \{f_h\} \end{Bmatrix} \tag{17.38}$$

respectively, where $\{d_h\}$ are the temporarily added element displacements, which correspond to the shape function $[N_h]$. The stiffness matrix has been divided into sub-stiffness matrices corresponding to degrees of freedom. Note that $\{f_h\}$ is the vector of element loads on the added degree of freedom. Assuming that the vector $\{d_n\}$ remains unchanged due to refinement, the approximate values of $\{d_h\}$ can be obtained from the second equation of Eq. (17.38) as

$$\{d_h\} = [K_{hh}]^{-1}\left(\{f_h\} - [K_{hn}]\{d_n\}\right) \tag{17.39}$$

To obtain the approximate value of each component of $\{d_h\}$, Eq. (17.39) may be simply written as

$$(d_h)_j = \frac{\left(f_h - [K_{nh}]\{d_n\}\right)_j}{(K_{hh})_j} \tag{17.40}$$

where j is the index number of the entry of $\{d_h\}$ and $(K_{hh})_j$ is the diagonal term of $[K_{hh}]$ at the particular j. The difference between the two fields (before and after adding the higher order shape functions) is defined as the estimated displacement error. The error in displacement fields can be expressed as

$$\{e_u\} = \{u_{p+1}\} - \{u_p\} \tag{17.41}$$

Superconvergent patch recovery technique

Superconvergent patch recovery (SPR) technique proposed by Zienkiewicz and Zhu (1992) is a recovery technique in which the recovered solution is obtained based on the values at certain points called superconvergent points. Superconvergent points are points at which the values predicted by numerical methods are expected to be more accurate than the rest of the points. The recovered solution that is calculated on the basis of superconvergent points is considered to be exact for the purpose of estimation of the error. Implementation of the SPR method is very simple and cost effective. This can be widely used for linear, quadratic, and cubic elements for both one- and two-dimensional problems. It has been understood that the derivatives of the primary variables are more accurate at certain points called Barlow points [Barlow (1976)]. In one-dimensional C^0 continuity elements, Barlow points coincide with certain Gauss points. For two-noded rod elements, Barlow point corresponds to the point of one-Gauss-point formula. For three-noded rod elements, Barlow points correspond to points of two-Gauss-point formula. For four-noded rod elements, Barlow points correspond to points of three-Gauss-point formula. In general, for C^0 continuity elements, the best accuracy for the derivatives of the primary variable is expected at the Gauss points corresponding to the order of polynomial used in solution. For beam elements, the derivatives are generally calculated at the points corresponding to two-Gauss-point formula. For estimating the derivatives at the element boundaries or nodes, the derivative values at the Gauss points may be extrapolated. However, the gradients at nodes and element boundaries which are obtained by extrapolation of the gradients at Gauss points from the

finite element analysis do not possess the inter-element continuity and are of low accuracy. To obtain a better gradient field, the SPR technique has been proposed by Zienkiewicz and Zhu (1992).

In the SPR technique, we take a patch of elements surrounding a node, called 'patch' assembly point. Some typical one-dimensional and two-dimensional patches are shown in Fig. 17.1. The nodes of the elements (shown by filled circles in the figure) and superconvergent points (shown by crosses) are shown. Patch assembly node is shown by the bigger circle. In the figure, one-dimensional patches consist of two elements and two-dimensional patches consist of four elements. We denote the gradient by $\{q\}$. In this recovery process, it is assumed that the accurate nodal values belong to a polynomial expansion $\{q_p^*\}$ of the same complete order p as that present in the shape function $\{N\}$ and which is valid over an element patch surrounding the particular assembly node considered. Such a 'patch' represents a union of elements containing this vertex node. This polynomial expansion will be used for each component of $\{q_p^*\}$, denoted as q_p^*, and one can get

$$q_p^* = \lfloor P \rfloor \{a\} \tag{17.42}$$

where $\lfloor P \rfloor$ contains the appropriate polynomial terms and $\{a\}$ is a set of unknown parameters. For one-dimensional elements, if a gradient component is approximated by a pth degree polynomial, then $\lfloor P \rfloor$ is given by

$$\lfloor P \rfloor = \lfloor 1 \;\; x \;\; x^2 \;\; x^3 \;\; \ldots \;\; x^p \rfloor \tag{17.43}$$

and

$$\{a\} = \lfloor a_1 \;\; a_2 \;\; \ldots \;\; a_{p+1} \rfloor^{\mathrm{T}} \tag{17.44}$$

Similarly, for two-dimensional linear approximation (as done in three-noded triangular element), $\lfloor P \rfloor$ is given by

$$\lfloor P \rfloor = [1 \;\; x \;\; y] \tag{17.45}$$

For two-dimensional quadratic approximation (as done in six-noded triangular element), $\lfloor P \rfloor$ is given by

$$\lfloor P \rfloor = \left[1 \;\; x \;\; y \;\; x^2 \;\; xy \;\; y^2\right] \tag{17.46}$$

For a bilinear quadrilateral type approximation, we can write

$$\lfloor P \rfloor = \lfloor 1 \;\; x \;\; y \;\; xy \rfloor \tag{17.47}$$

The similar forms for other higher order expansions can be used.

The determination of the unknown parameters $\{a\}$ of the approximation given in Eq. (17.42) is best made by ensuring a least square fit of this to the set of superconvergent or high-accuracy sampling points existing in the patch considered if such points are available. To obtain this, one can minimize

$$\begin{aligned} F(\{a\}) &= \sum_{i=1}^{n} \left(q_h(x_i, y_i) - q_p^*(x_i, y_i)\right)^2 \\ &= \sum_{i=1}^{n} \left(q_h(x_i y_i) - \lfloor P(x_i, y_i) \rfloor \{a\}\right)^2 \end{aligned} \tag{17.48}$$

where (x_i, y_i) are the global co-ordinates of sampling points, $n = mk$ is the total number of sampling points, and k is the number of the sampling points on each element m_j ($m_j = 1, 2, 3 \ldots m$) of the element patch. In Eq. (17.48), q_h denotes the value obtained by FEM at a superconvergent point. The minimization condition of $F(\{a\})$ implies that $\{a\}$ satisfies

$$\sum_{i=1}^{n} \lfloor P(x_i, y_i) \rfloor^{\mathrm{T}} \lfloor P(x_i, y_i) \rfloor \{a\} = \sum_{i=1}^{n} \lfloor P(x_i, y_i) \rfloor^{\mathrm{T}} \{q_h(x_i, y_i)\} \tag{17.49}$$

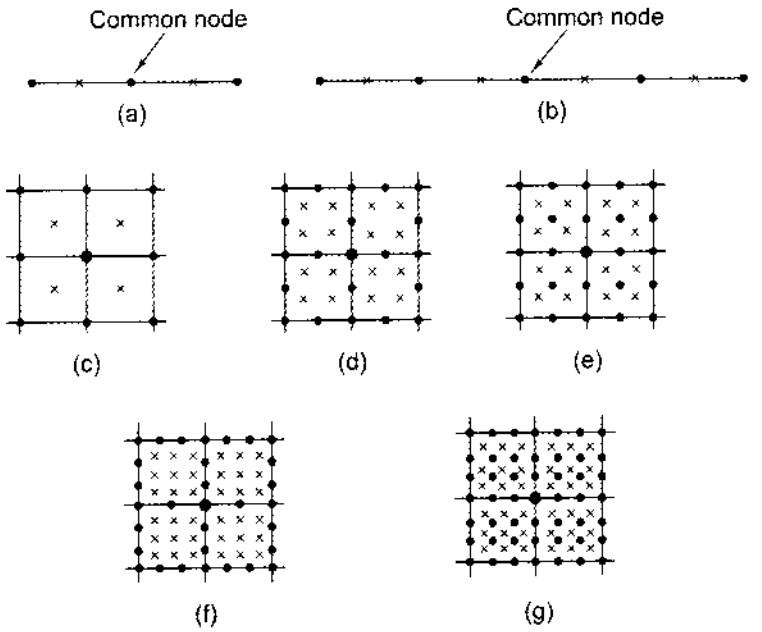


Figure 17.1 *Superconvergent points (represented by crosses) for patches of (a) two-noded one-dimensional elements (b) three-noded, one-dimensional elements (c) four-noded quadrilateral elements (d) eight-noded quadrilateral elements (e) twelve-noded quadrilateral elements (f) sixteen-noded quadrilateral elements*

This can be solved in matrix form as

$$\{a\} = [A]^{-1} \{b\} \tag{17.50}$$

where

$$[A] = \sum_{i=1}^{n} \lfloor P(x_i, y_i) \rfloor^{\mathrm{T}} \lfloor P(x_i, y_i) \rfloor \quad \text{and} \quad \{b\} = \sum_{i=1}^{n} \lfloor P(x_i, y_i) \rfloor^{\mathrm{T}} q_h(x_i, y_i) \tag{17.51}$$

Once the parameters $\{a\}$ are determined, the recovered nodal values are calculated by insertion of appropriate co-ordinates into the expression for q_p^* [Eq. (17.42)].

It is to be mentioned that as far as possible, Eq. (17.42) should be used for interpolation and not for extrapolation. It will be observed that element patches will overlap for internal mid-side nodes and nodes in the element interior. This means that such recovered nodal values are frequently evaluated from two or more patches. In these cases, average values of the patches are used for the nodes. For the boundary nodes, there are insufficient numbers of nodes for a patch. For one such element situation (corner node) the size of the patch is insufficient for determination of the parameters $\{a\}$ and the corner node values are determined from an interior patch by extrapolation. This is shown in Fig. 17.2.

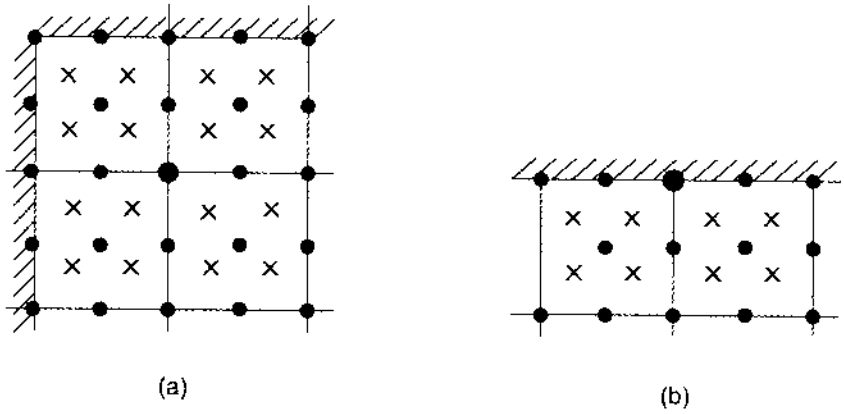


Figure 17.2 *Patches for recovery at boundary nodes: (a) patch consisting of four elements (b) patch consisting of two elements*

Higher order approximation of primary variables

As discussed in the previous section, in SPR technique, a smooth polynomial is fitted by obtaining the sampling data for derivatives from so-called superconvergent points. Sometimes, it may be difficult to assess which points in a particular element are superconvergent points. On the other hand, it has been observed that the values of primary variables are more accurate at the nodes. If a higher degree polynomial is fitted on the basis of nodal values, it is expected to give more accurate results. This procedure used by some researchers may be called SPR for displacement [Zienkiewicz, et al. (2005)]. Here we introduce a similar method called the higher order approximation of primary variables (HOAPV) method [Karimulla, et al. (2003)]. It can be shown that the finite element solution obtained by the linear interpolation functions can be improved by using HOAPV. In this, the solution for the primary variable itself is obtained by fitting a polynomial one degree higher than the finite element polynomial and a blended function of such fitted functions are used. The higher degree polynomial can be used to estimate the error by residual method. For example, consider the one-dimensional, second-order differential equation. If one has nodal solution data, then fitting a second-degree polynomial will ensure to provide finite residue at each point of the domain, which can be used as an estimate of the error.

A typical node inside the mesh is surrounded by two nodes on each side as shown in Fig. 17.3a. A quadratic polynomial can be fitted on that node passing through three points. The particular node is taken as one point and the adjacent nodes on both sides as other two points. At node j, the quadratic polynomial is of the form

$$T^j = a^j + b^j x + c^j x^2 \tag{17.52}$$

Similarly, the quadratic polynomial on $(j+1)^{\text{th}}$ node is

$$T^{j+1} = a^{j+1} + b^{j+1} x + c^{j+1} x^2 \tag{17.53}$$

Now, the temperature function in between the nodes j and $(j+1)$ can be written as

$$T = N_1 T^j + N_2 T^{j+1} \tag{17.54}$$

where N_1 and N_2 are standard linear shape functions.

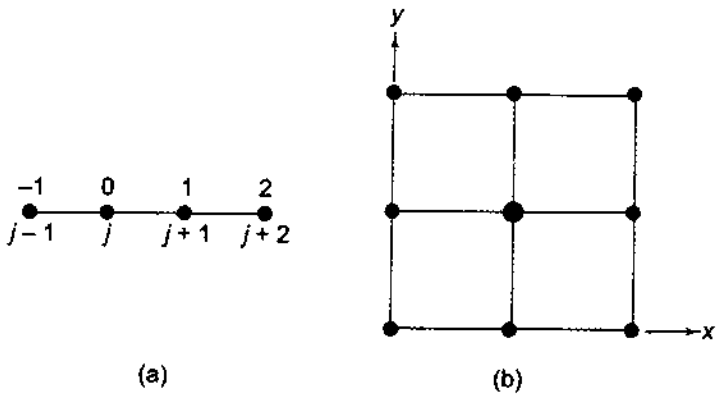


Figure 17.3 *Patches for obtaining the higher degree functions for (a) two-noded, one-dimensional element (b) four-noded quadrilateral elements*

The above scheme is based on curve-fitting using parabolic bending scheme. It can be easily shown that the above approximation yields highly accurate results compared to standard post-processing procedure. To derive the expression for error, take the coordinates of the nodes $j-1$, j, $j+1$ and $j+2$ as -1, 0, 1, and 2, respectively. Assume that the exact polynomial Texact is given by the following equation

$$T_{\text{exact}} = a + bx + cx^2 + dx^3 \tag{17.55}$$

Since the fitted polynomial given by Eq. (17.52) is exact at the nodes $j-1$, j, $j+1$, it is equal to the exact value of primary variables at these three points. Hence, the following set of equations is obtained:

$$\begin{aligned} a^j - b^j + c^j &= a - b + c - d \\ a^j &= a \\ a^j + b^j + c^j &= a + b + c + d \end{aligned} \tag{17.56}$$

Solution of Eq. (17.56) yields

$$a^j = a,\ b^j = b + d,\ c^j = c \tag{17.57}$$

Therefore, the approximation function T^j is given by

$$T^j = a + (b+d)x + cx^2 = a + bx + cx^2 + dx \tag{17.58}$$

In the same way, T^{j+1} given by Eq.(17.53) is exact at the nodes $j(x = 0)$, $j + 1$ $(x = 1)$, and $j + 2$ $(x = 2)$, and hence is equal to the exact value of primary variables at these three points. Therefore, the coefficients in Eq. (17.53) are given by

$$a^{j+1} = a,\ b^{j+1} = b - 2d \text{ and } c^{j+1} = c + 3d \tag{17.59}$$

Thus, the approximating function T^{j+1} is given by

$$T^{j+1} = a + bx + cx^2 + (3x^2 - 2x)d \tag{17.60}$$

Now, in the element $j - (j + 1)$, N_1 is $(1 - x)$ and N_2 is x. Hence, from Eq. 17.54, we obtain

$$T = (1-x)T_j + xT_{j+1} = a + bx + cx^2 + (x - 3x^2 + 3x^3)d \tag{17.61}$$

Subtracting the equation for exact solution from the Eq. (17.61), error in the primary variable becomes

$$e_T = (2x^3 - 3x^2 + x)d \tag{17.62}$$

The roots of the polynomial inside the brackets in Eq. (17.62) are 0, 1, and 0.5, indicating that blended function gives exact value not only at the nodes, but also at the middle point. One can also find the $\mathbf{L}_2$ norm

$$\|e_T\|_{L_2} = \left(\int_0^1 \left[\left(2x^3 - 3x^2 + x\right)d \right]^2 \mathrm{d}x \right)^{\frac{1}{2}} \tag{17.63}$$

The first derivative will be exact at points $\frac{1}{2} \pm \frac{1}{2\sqrt{3}}$ and the second derivative at the middle of the element. There can be various ways to find out error and the mesh refinement strategy. The simplest way to find out the error is to calculate residue r by putting back the expression for T in the differential equation.

A procedure has been developed for two-dimensional problems using Laplace or Poisson equation [Karimulla, et al. (2003)]. A two-dimensional finite element mesh using four-noded elements is shown in Fig. 17.3(b). A typical node i inside the mesh is usually a member of four elements. There are a total of nine nodes in these four elements; hence, a quadratic at the ith node may be constructed using nine points. The quadratic is of the form given below:

$$f_i = \alpha + \alpha_2 x + \alpha_3 y + \alpha_4 xy + \alpha_5 x^2 + \alpha_6 y^2 + \alpha_7 xy^2 + \alpha_8 x^2 y + \alpha x^2 y^2 \tag{17.64}$$

Approximation of the primary variable inside an element is carried out in the following manner:

$$f^e = N_1 f_1 + N_2 f_2 + N_3 f_3 + N_4 f_4 \tag{17.65}$$

In which N_1, N_2, etc. are the linear shape functions and f_i is quadratic function given by Eq. (17.64) defined for the ith node. Note that the ith node is at the center surrounded by other eight

nodes, when a polynomial f_i is constructed using nine nodes. In case of heat transfer problems, the primary variables (temperature) T can be written as

$$T^e = N_1T_1 + N_2T_2 + N_3T_3 + N_4T_4 \tag{17.66}$$

These primary variables themselves are the improved ones and hence gradients found using these values are also improved ones.

There may be situations where all the four nodes of the element may not be at the center of the patch of the elements. This situation is tackled in the following manner. If only one node is available over which the polynomial can be defined, then at all the other nodes, the same function is taken. If the functions can be defined in two adjacent nodes, then the same functions are used on the corresponding opposite nodes. If the polynomial is defined on the diagonal nodes, then in the other two diagonal nodes, average value of the function is taken. If the polynomial can be defined on three nodes, then in the fourth node, average of two adjacent nodes is taken for the coefficients of the polynomial.

17.4 CONCLUSION

In this chapter, a brief introduction of error analysis has been provided. The error analysis is very important for adaptive refinement. It also provides an idea about the quality of the solution. Error analysis is still a hot area of research. In particular, the suitable recovery techniques for triangular elements are still being explored.

SOLVED PROBLEMS

17.1 The following table provides the values of stress component σ_{xx} at five superconvergent points of a two-dimensional element:

Superconvergent point	σ_{xx} (MPa)
(0, 0)	200
(5, 5)	230
(5, –5)	270
(–5, 5)	300
(–5, –5)	160

From this data, obtain a best-fit expression of the following form:

$$\sigma_{xx} = a + bx + cy + dxy \tag{17.1.1}$$

Solution Assume, for present, that the data at all the points satisfies the expression given in Eq. (17.1.1). We get the following five equations:

$$a = 250 \tag{17.1.2a}$$

$$a + 5b + 5c + 25d = 230 \tag{17.1.2b}$$

$$a + 5b - 5c - 25d = 270 \tag{17.1.2c}$$

$$a - 5b + 5c - 25d = 300 \tag{17.1.2d}$$

$$a - 5b - 5c + 25d = 160 \tag{17.1.2e}$$

In the matrix form, the system of equations given in Eq. (17.1.2) can be written as

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 5 & 5 & 25 \\ 1 & 5 & -5 & -25 \\ 1 & -5 & 5 & -25 \\ 1 & -5 & -5 & 25 \end{bmatrix} \begin{Bmatrix} a \\ b \\ c \\ d \end{Bmatrix} = \begin{Bmatrix} 200 \\ 230 \\ 270 \\ 300 \\ 160 \end{Bmatrix} \tag{17.1.3}$$

or

$$[P]\{a\} = \{\sigma\} \tag{17.1.4}$$

As there are five equations and only four unknowns, there may be no solution satisfying all the equations. However, we can find a solution in which error is minimized in the least square sense. A simple way to do that is as follows.

Multiply both sides of Eq. (17.1.4) by the transpose of the coefficient matrix to get

$$[P]^{\mathrm{T}}\,[P]\{a\} = [P]^{\mathrm{T}}\,\{\sigma\} \tag{17.1.5}$$

Then

$$\{a\} = ([P]^{\mathrm{T}}[P])^{-1}[P]^{\mathrm{T}}\,\{\sigma\} \tag{17.1.6}$$

or

$$\begin{Bmatrix} a \\ b \\ c \\ d \end{Bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 5 & 5 & -5 & -5 \\ 0 & 5 & -5 & 5 & -5 \\ 0 & 25 & -25 & -25 & 25 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 5 & 5 & 25 \\ 1 & 5 & -5 & -25 \\ 1 & -5 & 5 & -25 \\ 1 & -5 & -5 & 25 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 5 & 5 & -5 & -5 \\ 0 & 5 & -5 & 5 & -5 \\ 0 & 25 & -25 & -25 & 25 \end{bmatrix} \begin{Bmatrix} 200 \\ 230 \\ 270 \\ 300 \\ 160 \end{Bmatrix}$$

$$= \begin{Bmatrix} 232 \\ 2 \\ 5 \\ -1.8 \end{Bmatrix}$$

Thus, the best-fit expression is given by

$$\sigma_{xx} = 232 + 2x + 5y - 1.8xy$$

17.2 Prove that in a two-noded bar element used for solving the problem of a rod subjected to axial loads, the strains are expected to be the most accurate at the mid-point of the element.

Solution It is well-known that nodal values in the FEM solution of rod problems can be determined exactly, if domain approximation, quadrature and finite arithmetic errors are absent. In general, the

nodal values in an FEM solution are expected to be most accurate. A two-noded bar element approximates the displacement field by a linear function. Assume that the exact solution is of quadratic polynomial form such that

$$u_{\text{exact}} = a + bx + cx^2 \tag{17.2.1}$$

Let the two nodes of the element be at $x = 0$ and $x = h$. FEM determined values at these nodes can be obtained by substituting the values of x in Eq. (17.2.1). Thus

$$u_1 = a, \quad u_2 = a + bh + ch^2 \tag{17.2.2}$$

The strains calculated by FEM will be

$$\varepsilon_{\text{FEM}} = \frac{u_2 - u_1}{h} = b + ch \tag{17.2.3}$$

The exact strain is given by

$$\varepsilon_{\text{exact}} = \frac{\mathrm{d}u_{\text{exact}}}{\mathrm{d}x} = b + 2cx \tag{17.2.4}$$

The FEM calculated strain will be equal to exact strain if

$$b + 2cx = b + ch \tag{17.2.5}$$

which gives $x = h/2$. Thus, the strain is exact at the mid-point of the element, if the exact solution is of quadratic polynomial form. In general, the strain may be expected to be the most accurate at the mid-point.

EXERCISES

A. Multiple Choice Questions

17.1 If the error vector is $\{0.3\ 0.4\}^{\mathrm{T}}$, the $\mathbf{L}_2$ norm of the error is

(a) 0.3 (b) 0.4
(c) 0.7 (d) 0.5

17.2 In a one-dimensional FEM mesh consisting of three-noded C^0 continuity elements, if the size of the element is reduced by a factor of 2, the error norm is expected to reduce by a factor of

(a) 2 (b) 3
(c) 4 (d) 8

17.3 If a four-noded element is used for the estimation of temperature to solve a two-dimensional heat conduction problem, the gradient is expected to be the most accurate

(a) at the centroid of the element
(b) everywhere
(c) at the Gauss points corresponding to two-Gauss-point formula
(d) at the Gauss points corresponding to three-Gauss-point formula

17.4 If a four-noded element is used for solving the two-dimensional temperature problems, the temperatures are expected to be the most accurate at

(a) nodes
(b) Gauss point corresponding to one-Gauss-point formula
(c) Gauss points corresponding to two-Gauss-point formula
(d) Gauss points corresponding to three-Gauss-point formula

17.5 If an eight-noded element is used for solving the two-dimensional temperature problems, the temperatures are expected to be the most accurate at
(a) nodes
(b) Gauss point corresponding to one-Gauss-point formula
(c) Gauss points corresponding to two-Gauss-point formula
(d) Gauss points corresponding to three-Gauss-point formula

B. Problems

17.6 Prove that in solving the problem of rod subjected to axial load using three-noded Lagrangian element, longitudinal stress is expected to be accurate corresponding to the points of two-point-Gauss-quadrature formula.

17.7 A square plate is subjected to the temperature of 500 K at three sides and 0 K at the fourth side. Solve this problem using FEM by discretizing the domain into 16 equal elements. Find out the recovered solution using SPR technique. Using the recovered solution, find out the error of each element and overall error of the solution. You are already familiar with the exact solution. Find out the error using the exact solution. Compare it with the estimated error.

17.8 Solve the problem of a plate with hole subjected to tensile load in one direction and find out the error with coarse and fine mesh.

17.9 The governing equation for certain heat conduction problem is written as

$$k\frac{\mathrm{d}^2T}{\mathrm{d}x^2}+q=0$$

where T is a temperature as a function of x, k is thermal conductivity, and q is the rate of heat generation per unit volume. For formulation purpose, an element of length h is transformed to natural coordinates ξ having the coordinate of the first node as -1 and that of second as $+1$. The interpolation function for temperature is given as

$$T=N_1T_1+N_2T_2+(1-\xi^2)a$$

where T_1 and T_2 are temperatures at the two nodes and a is some constant. Find out the elemental stiffness matrix and elemental load vector for this problem. Also, carry out the error analysis.

REFERENCES

Barlow, J, 1976, 'Optimal stress locations in finite element models', *International Journal of Numerical Methods for Engineering*, Volume 10, pp. 243–251.

Karimulla, SK, Dwivedy, SK, and Dixit, US, 'An efficient post-processing strategy for finite element analysis of heat transfer problems', *CD Proceedings Recent Advances in Heat and Mass Transfer*, January 6–8, 2002, IIT Guwahati, India.

Zienkiewicz, OC, and Taylor, RL, 1989, *The Finite Element Method, Volume 1, Fourth Edition*, McGraw-Hill, New York.

Zienkiewicz, OC, and Taylor, RL, and Zhu, JZ, 2005, *The Finite Element Method: Its Basis & Fundamentals, Sixth edition*, Butterworth-Heinemann, Oxford.

Zienkiewicz, OC, and Zhu, JZ, 1987, 'A simple error estimator and adaptive procedure for practical engineering analysis', *International Journal Numerical Methods for Engineering*, 24, 337–357.

Zienkiewicz, OC, and Zhu, JZ, 1992, 'The super convergent patch recovery and a posteriori error estimation in the finite element method', Part I & Part II, *International Journal Numerical Methods for Engineering*, 33, 1331–1364.

Zienkiewicz, OC, Gargo, JP, De, SR, and Kelly, DW, 1983, 'The hierarchical concept in finite element analysis', *Computers & Structures*, Volume 16, pp. 53–65.

18
CHAPTER

Solving FEM Problems on A Computer

18.1 Introduction

18.2 Developing an FEM Code

18.3 Finite Element Packages

18.4 Conclusion

18.1 INTRODUCTION

Generally problems of finite element method (FEM) are solved by considering a large number of elements and nodes; therefore, it becomes essential to use a computer for solving such problems. The finite element formulation of a problem is carried out manually and the expressions for elemental coefficient matrix and right-hand-side vector are evaluated using a numerical integration routine. The assembly of the elemental matrices and vectors is carried out by a subroutine for assembly and the resulting equations are solved by an equation solver. Finally, post-processing is carried out.

Nowadays, there are a number of finite element packages that are able to solve standard problems. These packages have the capability of graphical display. Industrial users prefer to use these packages in order to save the time required for writing a code. On the other hand, researchers find it convenient to write their own codes for the sake of flexibility in modifying the code for incorporating newer features. Many a times, researchers make use of preprocessor and post-processor of a commercial package and write their own code for the processor part of FEM. In this chapter, we shall first describe the procedure for writing an FEM code and then provide a brief note on the use of commercial FEM packages.

18.2 DEVELOPING AN FEM CODE

A finite element code should be made in a modular fashion, which means each subtask should be carried out in a subroutine or function. The variables should be named in such a manner that their meaning is explicit. A large number of comments should be written in the code to clarify each step of the code. If a code is developed following these tips, then the debugging of the code becomes an easy task.

Generally, a finite element code can be divided into the following three parts:

(i) preprocessor
(ii) main program
(iii) post-processor

18.2.1 Preprocessor

Preprocessor is a program that processes an input data to produce an output which is then used as the input to the processor of the FEM code. The preprocessor needs information about geometry, materials, loading, and boundary conditions. Here, we use the word *loading* to indicate the interaction of the problem domain with outside world. Thus, for solid mechanics problems, loading may refer to forces and moments, while for heat-conduction problems, it may refer to heat flow. Internal heat generation is analogous to body forces in solid mechanics. Generally, a preprocessor performs the following tasks:

(i) generates a finite element mesh
(ii) takes material properties and loading as input

These tasks are described in detail as follows:

Mesh Generation

In this task of the preprocessor, a finite element mesh is generated. The input about the type of elements and number of elements may be obtained from the user. There are different techniques for mesh generation [Zienkiwicz, et al. (2005)]. The fully automatic mesh generation creates a mesh without user intervention. In the semi-automatic method, the model domain is subdivided manually into simple sub-regions called multi-blocks. The mesh generation in each multi-block is carried out automatically. Mesh can be generated manually as well; but this becomes a tedious process if the number of elements is large.

There are two types of mesh: *structured and unstructured*. A *structured mesh* is a mesh that has a regular arrangement of elements and can be defined by specifying the parameter of the arrangement. It resembles a lattice structure of metals. Each node inside a structured mesh is generally connected to the same number of elements. On the other hand, an *unstructured mesh* is a mesh that has an irregular arrangement of elements. Each element and its connections to adjacent cell are defined separately. At different nodes inside the mesh, different number of elements may meet, although it need not be the case always. Figure 18.1(a) shows a structured mesh with rectangular elements. Here, the node number formed by the intersection of ith vertical line (measured from left to right) and jth horizontal line (measured from top to bottom) is given by

$$k = 5j + i - 5 \tag{18.1}$$

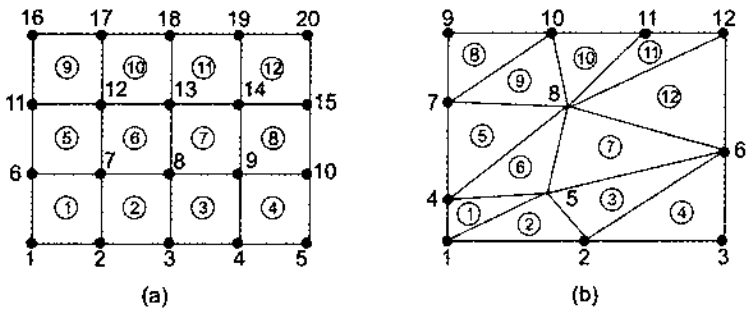


Figure 18.1 *Examples of finite element meshes: (a) structured mesh (b) unstructured mesh*

By varying i from 1 to 5 and j from 1 to 4, we can get the global number of all the nodes. Also, for each element, if k is the global node number of the first node, the global node numbers of the second, third, and fourth nodes will be written as $k+1$, $k+6$, and $k+5$, respectively. The element numbers can also be assigned systematically based on which vertical and horizontal lines intersect to form the first node of the element. Thus, the topology and node numbers of mesh can be expressed in the form of formulae containing i and j, the identification numbers of vertical and horizontal lines. Therefore, the mesh shown in Fig. 18.1(a) is a structured mesh. If this mesh is converted into a mesh of triangular elements by dividing each rectangular element into triangles with one diagonal as the interface between the triangles, the resulting mesh will also be called a structured mesh. Figure 18.1(b) shows an unstructured mesh consisting of triangular elements. Note that in Fig. 18.1(a), all non-boundary nodes are shared by four elements. On the other hand, in Fig. 18.1(b), node 5 is shared by five elements

and node 8 is shared by seven elements. There is no regular arrangement of elements. Note that it is not necessary that elements in the unstructured mesh are triangular only. The mesh consisting of elements that are all quadrilateral can also be unstructured.

For getting introduced to mesh-generation methods, the reader may refer to the following review papers: Ho-Le (1988) and Shepard, et al. (1988). The meshes that are commonly used in two-dimensional space consist of triangular or quadrilateral elements, while in three-dimensional space, tetrahedron and brick elements are the most commonly used. Many mesh generators first produce set points called nodes and then join them to form triangles. This is most efficiently carried out by a procedure called *Delaunay triangulation*. This triangulation procedure maximizes the sum of the smallest angles of the triangles. Triangles having very small angles are avoided as they may create numerical difficulties during FEM solution. The concept of Delaunay triangulation can be extended to three dimensions also. Heighway (1983) has devised a technique to convert a mesh of triangular elements into a mesh of quadrilateral elements by combining every pair of adjacent triangles into a quadrilateral. Another popular technique for mesh generation is the *advancing front method*. In this method, the mesh is constructed by progressively adding the elements starting at the boundaries. The process results in a front, which is the boundary between the meshed and unmeshed region. This method is suitable for the creation of mesh consisting of triangular or tetrahedron elements.

The output of a mesh-generation routine is the coordinates of the nodes and connectivity matrix. With this information, a mesh can be constructed graphically. The nodal coordinates of the elements will be required by the processor routine.

Taking Material Properties and Loading as Input

Inputs regarding material properties and loading are collected from the user and stored in a particular location. These inputs are then used by the processor and post-processor routines. Since material properties may differ from element to element, a routine is often written to accept the material properties element wise. Similarly, loading may also differ from element to element; the load may be a distributed load or may be concentrated at a point. For the case of a concentrated load, it is desirable to have a node at the point of application of the load.

18.2.2 Main Program

The main program is designed to include evaluation of coefficient matrices and right-hand-side vectors, assembly, application of boundary conditions, and solution of linear system of equation. To be consistent with the names preprocessor and post-processor, we could call the main program as processor. However, the word 'processor' is used more popularly for certain other applications, such as the central processing unit of a computer. The main program contains the following routines:

Routine for Evaluation of Element Coefficient Matrix

This routine generates the coefficient matrix. The nodal coordinates of the element and element properties are provided as input to this routine. With this information, the coefficient matrix may be

generated. Rarely do we obtain closed form for various entries of the coefficient matrix. Instead numerical integration is carried out. Often Gauss-quadrature is used for this purpose. The weights and points of Gauss-quadrature formulae can be stored in a subroutine.

Routine for Evaluation of Right-Hand-Side Vector

The right-hand-side vector is evaluated in the same way as coefficient matrix. In solid mechanics problems, the evaluation of right-hand-side vector requires information about the applied forces, whereas in heat-transfer problems, information regarding heat generation and heat flux is required.

Routine for Assembly

This routine carries out the assembly of coefficient matrices and right-hand-side vectors. For two-dimensional problems, the assembly procedure is explained in detail in Chapter 8. A pseudo-code is also provided there.

Routine for Applying Boundary Conditions

These routines apply the essential and natural boundary conditions. Sometimes, there may be special type of boundary conditions. For applying such special boundary conditions, a separate routine may be written.

Routine for Solving the Equations

After the successful application of boundary conditions, a linear system of equations is obtained. This system can be solved by a number of methods. One of the simplest methods is Gauss elimination. It consists of two parts: in the first part, the coefficient matrix is reduced to an upper triangular form by carrying out the elementary row operation. As a result of row operations, the right-hand-side vector also gets modified. The second part consists of *back substitution*. Suppose, the unknown variables are $x_1, x_2, \ldots, x_n$; in back substitution, first the value of x_n is obtained from the last equation. The obtained value is then substituted in the second last equation and the value of x_{n-1} is obtained. The process is continued till we reach the first equation and substitute the values of $x_2, x_3 \ldots, x_n$ in the first equation in order to obtain the value of x_1.

Usually, a Gauss-elimination routine has a provision of pivoting to ensure that zeroes are not present at the diagonal positions of the coefficient matrix. Even if there are no zeroes at the diagonal positions initially, they may be created during row operations. A strategy is adopted for arranging the equations so as to put the coefficient of largest magnitude on the diagonal at each step. This is called *pivoting*. Complete pivoting may require both row and column interchange. When a column interchange is done, the unknown variables $(x_1, x_2, \ldots, x_n)$ need to be renamed. To avoid this effort, often only partial pivoting is carried out. Partial pivoting places a coefficient of larger magnitude on the diagonal by row interchanges only. The diagonal element that results after pivoting is called the *pivot element*. All the entries in the row have to be divided by the pivot element. A C++ function for carrying out Gauss elimination is provided in Appendix A.3.

18.2.1 Post-processor

In an FEM code, a post-processor usually contains the following routines:

Routine for Interpolation of Primary Variables from the Nodal Data

The output of the main program is in the form of discrete values at the nodes. The post-processor may convert them into continuous form by multiplying the row vector of shape functions with the column vector of nodal values. Once the continuous form is available, the values of the variable at different points can be obtained for displaying it graphically.

Routine for Obtaining the Gradient of Primary Variable

Once the continuous form of the variable is available, it can be differentiated to obtain the gradients of the variable. However, the accuracy of the derivatives is lesser than the accuracy of primary variable in FEM solution. Therefore, special care is taken to obtain the derivative solution. First, the derivatives are computed at points where they are expected to be more accurate than the rest of the points in the element. Then, the interpolation is carried out by taking the values of the derivatives at the expected high-accuracy points. For many elements, the expected most-accuracy points are the Gauss points. For four-noded Lagrangian elements, the most-accuracy point is the Gauss point corresponding to one-Gauss-point formula, which is the centroid of the element. For nine-noded Lagrangian elements, the expected most-accuracy points are the Gauss points corresponding to two-Gauss-point formula. Sometimes, values of the derivatives obtained from the most-accuracy points are extrapolated to the nodes. If a node is shared by a number of elements, the values are extrapolated from all those elements and an average is taken as the proper value. This procedure is called *nodal averaging*.

Routine for the Estimation of Error

We have already described a number of a *posteriori* methods for the estimation of error in Chapter 7. The codes based on these methods can be prepared. The user can get an idea about how much mesh refinement is required on the basis of error estimation.

In Appendix A.4, a pseudo-code for solving steady-state, two-dimensional, heat-conduction problem is presented. The code does not contain post-processing routines. It has been deliberately made simpler such that it contains only the minimum required features to enable beginners to understand it fully and prepare them for writing their own codes.

18.3 FINITE ELEMENT PACKAGES

The following are some of the commercial finite element packages available (their Web sites have also been provided):

- ANSYS (http://www.ansys.com)
- COMSOL (http://www.comsol.com)

- SolidWorks Simulation Premium (www.solidworks.com)
- NISA (http://www.nisasoftware.com/)
- MSC NASTRAN (http://www.mscsoftware.com/)
- ALGOR (http://www.algor.com/)
- ABAQUS (http://www.simulia.com)
- MARC (http://www.mscsoftware.com/)
- Pro/ENGINEER MECHANICA (http://www.ptc.com/)

These packages are supported with powerful graphics. One can solve a problem interactively with the help of the menu or write a program in the package-specific language. Different packages use different terminology for the commands. Therefore, it is important to get familiar with the terminology used by a package before using it for analysis. It is also important to know the features and flexible options offered by the package.

Typical finite element analysis using a package involves the following three distinct steps:

(i) Preprocessing
(ii) Applying loads and boundary conditions
(iii) Viewing the results and post-processing

Preprocessing

Usually, preprocessing (also called model-building) takes the maximum time among the three steps. In this step, the job name and analysis type are first specified. Then, a preprocessor is used to define the element types, real constants of the element, material properties, and the model geometry. Often, the preprocessor is a separate software or a module which is supplied along with the main software. Model geometry can be created in the preprocessor of the package or can be imported from other solid modelling packages.

A typical sequence of operations in this step is as follows:

- *Define units* The units such as SI, SI (mm), US unit (ft), and US unit (inch) need to be defined in each package. They should be consistent for all input parameters.
- *Define element type* The element type needs to be defined as beam, plane, or solid. The packages have an element library from which one can choose the elements. For example, ANSYS has got more than 150 elements in its library. Beam4 has six structural degrees of freedom and can be modelled in three-dimensional space. Plane77 is an eight-noded quadrilateral element suitable for two-dimensional space; it has one thermal degree of freedom per node.
- *Provide real constants for the elements* These include geometrical properties of the material such as area of cross-section and area moment.
- *Specify the material* Generally, a material library is available in the package from where one can select the material. The material properties can also be specified manually. The material can be linear or non-linear and isotopic, orthotropic, or anisotropic. Further, the properties can be constant or dependent on temperature. The material properties can be given as a function of strain, temperature, and strain rate. One can add some new material in the material library.

- *Model the geometry* This can be done in two ways. One way is by creating or importing a solid model and creating a mesh. Another way is by directly generating the mesh by specifying the nodal coordinates and connectivity.

Applying loads and obtaining the solutions

In the terminology of some packages, application of loads also means applying the boundary conditions. The term *load* also has a general meaning; for example, in heat-transfer problems, it may refer to heat flux. In this step, the user will first specify analysis type, such as thermal or structural, and analysis option, such as steady-state and transient. After that, loads and boundary conditions will be specified; the loads may be concentrated or distributed. Essential boundary condition will be applied depending on the analysis type. Then, the job is submitted for analysis. A message is displayed when the analysis is complete.

Viewing the results and post-processing

From the main menu bar one can choose the field output that needs to be displayed. Then, one can select the command for viewing contours. One can magnify (zoom) certain regions to get a better view. Graphs can also be plotted. In many packages, it is also possible to see animations. Data in the form of tables (lists) can be obtained.

18.4 CONCLUSION

Finite element method is essentially a computational technique; thus, computer skills are extremely important in effectively using FEM. Readers can write their own codes in their favourite computer language. In order to refrain from being partial to any computer language, we have explained the basic concepts of programming using pseudo-code. Commercial FEM packages can also be used to solve standard problems. Many free software are also available for this purpose. Some of the Web sites providing free software are mentioned as follows:

- http://www.cemtool.com
- http://www.freefem.org,
- http://www-rocq.inria.fr/OpenFEM/
- http://www.numeritek.com/

SOLVED PROBLEMS

18.1 Illustrate the procedure for partial pivoting, conversion of system of equations to upper triangular form, and back substitution in Gauss-elimination method by solving the following equations:

$$x + y + z = 6 \tag{18.1.1}$$

$$2x + 3y + 4z = 20 \tag{18.1.2}$$

$$-5x + y + z = 0 \tag{18.1.3}$$

Solution We shall illustrate partial pivoting, conversion of system of equations to upper triangular form, and back substitution in Gauss-elimination method one by one.

Partial pivoting We interchange Eqs 18.1.1 and 18.1.3 so that the coefficient matrix of system of equations has the largest magnitude element among the first column. Thus, after partial pivoting, the system of equation in matrix form is given as

$$\begin{bmatrix} -5 & 1 & 1 \\ 2 & 3 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 \\ 20 \\ 6 \end{Bmatrix} \tag{18.1.4}$$

Conversion to upper triangular form First, we shall make all sub-diagonal elements in the first column zero. Here, –5 is the value of the pivot element. We carry out the following row operations:

(i) Multiply the first row by –2/5 and subtract from the second row.

(ii) Multiply the first row by –1/5 and subtract from the third row.

With these operations, we get the following system of equations:

$$\begin{bmatrix} -5 & 1 & 1 \\ 0 & \frac{17}{5} & \frac{22}{5} \\ 0 & \frac{6}{5} & \frac{6}{5} \end{bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 \\ 20 \\ 6 \end{Bmatrix} \tag{18.1.5}$$

Now, we observe the elements of the second column and note that the diagonal element is of the largest magnitude (among the elements of the column). Therefore, we do not need pivoting. The value of the pivot element is 17/5. We multiply the elements of the second row by (6/5)/(17/5) and subtract from the third row in order to make sub-diagonal element zero. We obtain

$$\begin{bmatrix} -5 & 1 & 1 \\ 0 & \frac{17}{5} & \frac{22}{5} \\ 0 & 0 & -\frac{6}{17} \end{bmatrix} \begin{Bmatrix} x \\ y \\ z \end{Bmatrix} = \begin{Bmatrix} 0 \\ 20 \\ -\frac{18}{17} \end{Bmatrix} \tag{18.1.6}$$

Now, the upper-triangular form has been obtained. In this, all the elements below the diagonal are zero.

Back substitution In back-substitution process, we solve the equations in a reverse sequence, from last equation to first equation. The third equation is given by

$$-\frac{6}{17} z = -\frac{18}{17} \tag{18.1.7}$$

providing $z = 3$. The second equation is given by

$$\frac{17}{5} y + \frac{22}{5} z = 20, \quad \text{or} \quad \frac{17}{5} y = 20 - \frac{22}{5} z \tag{18.1.8}$$

Substituting the value of z in Eq. (18.1.8), we get $y = 2$. The first equation is given by

$$-5x + y + z = 0 \quad \text{or} \quad -5x = -y - z \tag{18.1.9}$$

Substituting the values of y and z in it, we get $x = 1$.

Thus, the solution of the system of equation given above is $x = 1$, $y = 2$, $z = 3$.

18.2 A square plate of area 1 m^2 is discretized into four elements as shown in Fig. 18.2 for studying two-dimensional heat conduction. After the solution, nodal temperatures were obtained as follows:

Node	1	2	3	4	5	6	7	8	9
Temperature (°C)	500	500	500	500	375	500	250	0	250

Find out the gradient of temperature at node 5 by nodal averaging method. Show all the steps as will be carried out in a post-processing module.

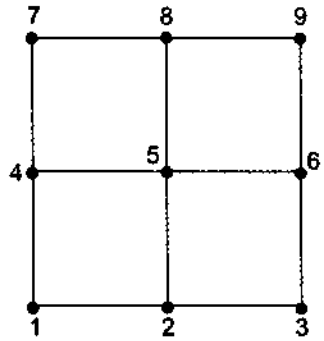


Figure 18.2 *A mesh consisting of four elements*

Solution First, the derivatives of the temperature with respect to natural coordinates are estimated. The temperature field in an element is given by

$$T = \frac{1}{4}(1-\xi)(1-\eta)T_1 + \frac{1}{4}(1+\xi)(1-\eta)T_2 + \frac{1}{4}(1+\xi)(1+\eta)T_3 + \frac{1}{4}(1-\xi)(1+\eta)T_4 \tag{18.2.1}$$

From Eq. (18.2.1), we obtain

$$\frac{\partial T}{\partial \xi} = \frac{(T_2 + T_3 - T_1 - T_4)}{4} + \frac{(T_3 + T_1 - T_2 - T_4)}{4}\xi \tag{18.2.2}$$

$$\frac{\partial T}{\partial \eta} = \frac{(T_3 + T_4 - T_1 - T_2)}{4} + \frac{(T_3 + T_1 - T_2 - T_4)}{4}\xi \tag{18.2.3}$$

To find out the derivative with respect to physical coordinates, we need to find the Jacobian. In the physical coordinates, elements are square (with one side equal to 0.5 m). The area of the element in physical coordinates is equal to 0.25, where as that in natural coordinates is 4. Therefore, the determinant of Jacobian will be 0.25/4 = 1/16. As the physical element is square both the diagonal entries of the Jacobian matrix will be equal and off-diagonal elements will be zero. Thus, the Jacobian is given by

$$[J]=\begin{bmatrix}\frac{1}{4} & 0\\ 0 & \frac{1}{4}\end{bmatrix} \tag{18.2.4}$$

Note that in a general computer program, Jacobian will not be calculated like this. Instead, the procedure outline discussed in Section 10.5 will be adopted. The inverse of Jacobian matrix is given by

$$[J]^{-1}=\begin{bmatrix}4 & 0\\ 0 & 4\end{bmatrix} \tag{18.2.5}$$

Then

$$\begin{Bmatrix}\frac{\partial T}{\partial x}\\ \frac{\partial T}{\partial y}\end{Bmatrix}=[J]^{-1}\begin{Bmatrix}\frac{\partial T}{\partial \xi}\\ \frac{\partial T}{\partial \eta}\end{Bmatrix}=\begin{bmatrix}4 & 0\\ 0 & 4\end{bmatrix}\begin{Bmatrix}\frac{\partial T}{\partial \xi}\\ \frac{\partial T}{\partial \eta}\end{Bmatrix}=\begin{Bmatrix}4\frac{\partial T}{\partial \xi}\\ 4\frac{\partial T}{\partial \eta}\end{Bmatrix} \tag{18.2.6}$$

Node 5 is shared by all four elements. Therefore, we calculate the gradients for all elements.

For element 1–2–5–4 Local node number is 3 having $\xi = 1$ and $\eta = 1$. Also, $T_1 = 500$, $T_2 = 500$, and $T_3 = 375$, $T_4 = 500$. Substituting these values in Eqs. (18.2.2) and (18.2.3), we get

$$\frac{\partial T}{\partial \xi}=-\frac{125}{2},\ \frac{\partial T}{\partial \eta}=-\frac{125}{2} \tag{18.2.7}$$

Hence, from Eq. (18.2.6), we obtain

$$\begin{Bmatrix}\frac{\partial T}{\partial x}\\ \frac{\partial T}{\partial y}\end{Bmatrix}=\begin{Bmatrix}-250\\ -250\end{Bmatrix} \tag{18.2.8}$$

For element 2–3–6–5 Local node number is 4 having $\xi = -1$ and $\eta = 1$. Also, $T_1 = 500$, $T_2 = 500$, $T_3 = 500$, and $T_4 = 375$. Substituting these values in Eqs (18.2.2) and (18.2.3), we get

$$\frac{\partial T}{\partial \xi}=\frac{125}{2},\ \frac{\partial T}{\partial \eta}=-\frac{125}{2} \tag{18.2.9}$$

Hence, from Eq. (18.2.6), we obtain

$$\begin{Bmatrix}\frac{\partial T}{\partial x}\\ \frac{\partial T}{\partial y}\end{Bmatrix}=\begin{Bmatrix}250\\ -250\end{Bmatrix} \tag{18.2.10}$$

For element 5–6–9–8 Local node number is 1 having $\xi = -1$ and $\eta = -1$. Also, $T_1 = 375$, $T_2 = 500$, $T_3 = 250$, and $T_4 = 0$. Substituting these values in Eqs (18.2.2) and (18.2.3), we get

$$\frac{\partial T}{\partial \xi} = \frac{250}{4}, \quad \frac{\partial T}{\partial \eta} = -\frac{750}{4} \tag{18.2.11}$$

Hence, from Eq. (18.2.6), we obtain

$$\begin{Bmatrix} \dfrac{\partial T}{\partial x} \\ \dfrac{\partial T}{\partial y} \end{Bmatrix} = \begin{Bmatrix} 250 \\ -750 \end{Bmatrix} \tag{18.2.12}$$

For element 4–5–8–7 Local node number is 2 having $\xi = 1$ and $\eta = -1$. Also, $T_1 = 500$, $T_2 = 375$, $T_3 = 0$, and $T_4 = 250$. Substituting these values in Eqs (18.2.2) and (18.2.3), we get

$$\frac{\partial T}{\partial \xi} = -\frac{250}{4}, \quad \frac{\partial T}{\partial \eta} = -\frac{750}{4} \tag{18.2.13}$$

Hence, from Eq. (18.2.6), we get

$$\begin{Bmatrix} \dfrac{\partial T}{\partial x} \\ \dfrac{\partial T}{\partial y} \end{Bmatrix} = \begin{Bmatrix} -250 \\ -750 \end{Bmatrix} \tag{18.2.14}$$

The average value of the gradient at node 5 is written as

$$\begin{Bmatrix} \dfrac{\partial T}{\partial x} \\ \dfrac{\partial T}{\partial y} \end{Bmatrix} = \frac{1}{4}\left[\begin{Bmatrix} -250 \\ -250 \end{Bmatrix} + \begin{Bmatrix} 250 \\ -250 \end{Bmatrix} + \begin{Bmatrix} 250 \\ -750 \end{Bmatrix} + \begin{Bmatrix} -250 \\ -750 \end{Bmatrix}\right] = \begin{Bmatrix} 0 \\ -500 \end{Bmatrix} \text{°C/m} \tag{18.2.15}$$

EXERCISES

A. Multiple Choice Questions

18.1 A mesh consisting of triangular elements is
(a) always structured
(b) always unstructured
(c) may be structured or unstructured
(d) cannot be classified into structured or unstructured

18.2 Delaunay triangulation
(a) minimizes the sum of the largest angles of the triangles
(b) minimizes the sum of the smallest angles of the triangles
(c) maximizes the sum of the largest angles of the triangles
(d) always generates isosceles triangles

18.3 Partial pivoting places a coefficient of larger magnitude on the diagonal by

(a) row interchanges only (b) column interchanges only

(c) row and column interchanges (d) scaling the columns

18.4 Complete pivoting places a coefficient of larger magnitude on the diagonal by

(a) row interchanges only (b) column interchanges only

(c) row and column interchanges (d) scaling the columns

18.5 Advancing front method is a

(a) mesh-generation method (b) post-processing method

(c) method to solve linear equations (d) method to solve non-linear equations

B. Problems

18.6 By carrying out the complete (total) pivoting, solve the following system of equations:

$$4x - y + 2z = 15$$

$$5x - 7y + 9z = 8$$

$$x - 2y - 3z = -5$$

18.7 List five popular FEM packages. Visit their websites and write a brief note on the capabilities of individual packages.

18.8 Write a program that generates a uniform structured mesh consisting of rectangular elements in a rectangular plate. The input of the program will be

(a) length and width of the plate

(b) number of divisions in length and width direction

The output of the program should be

(i) connectivity matrix

(ii) nodal coordinates

(iii) graphical display.

18.9 A square plate of area 1 m^2 is subjected to a temperature of 100°C on three sides and 0°C on the fourth side. Top and bottom surfaces of the plate are insulated, such that there is a two-dimensional heat flow. Find out the temperature and heat-flux distribution in the plate using some FEM packages to which you have access. (Do not use pirated packages.) Compare those packages in terms of

(a) ease of solving

(b) overall speed

(c) accuracy

(d) display of solution.

18.10 Make a flow chart of FEM program for solving the plane stress problem; it should have a provision for adaptively refining the mesh based on the error analysis.

REFERENCES

Heighway, EA, 1983, 'A mesh generator for automatically subdividing irregular polygons into quadrilaterals', *IEEE transaction on Mag.*, Vol. Mag. 19, pp. 2535–2538.

Ho-Le, K, 1988, 'Finite element mesh generation method: A review and classification', *Computer Aided Design*, Vol. 20, pp. 27–38.

Shepard, MS, Grice, KR, Lo, JA, and Schroeder, WJ, 1988, 'Trends in automatic 3-D mesh generation', *Computers & Structures*, Vol. 30, pp. 421–429.

Zienkiewicz, OC, Taylor, RL, and Zhu, JZ, 2005, *The Finite Element Method: Its Basis & Fundamentals, Sixth edition*, Butterworth-Heinemann, Oxford.

19
CHAPTER

Miscellaneous Topics

19.1 Introduction
19.2 Finite Element Solutions versus Exact Solutions
19.3 Accuracy of Derivatives of Primary Variables
19.4 Essential and Natural Boundary Conditions
19.5 Mesh Refinement
19.6 Effect of Geometry of an Element on FEM
19.7 Using FEM in Fracture Mechanics
19.8 Infinite Elements
19.9 Ill-Conditioned Systems
19.10 Patch Test
19.11 Three-Dimensional Finite Elements
19.12 Conclusion

19.1 INTRODUCTION

In the last eighteen chapters, we have covered various topics of finite element method. This material is sufficient for an introductory course on FEM covered in a semester. After understanding the basic concepts of FEM from this book, one can refer other textbooks and papers for further knowledge. One can also apply the knowledge gained in the previous chapter to many new areas. Finite element method is still an important area of research and newer developments are taking place continuously. In this chapter, we shall touch upon some miscellaneous topics that need to be discussed and conclude the first course on FEM in engineering. In this chapter, there are some topics which have not been described at all in the previous chapters, and some other topics which have been revisited.

19.2 FINITE ELEMENT SOLUTIONS VERSUS EXACT SOLUTIONS

Finite element method is a numerical technique that is generally used for solving differential and integral equations. Similar to other numerical techniques, it usually provides approximate solutions. In FEM, the error between the approximate and exact solutions can be reduced by mesh refinement. One may be interested to know if the method provides an upper bound or lower bound of a problem. We shall throw some light on this aspect. Let us focus on the problems of elastic structural mechanics. In general, the finite element procedure tends to make the structure stiffer than the actual one. This is because of the fact that we are imposing constraints on the structure by way of forcing the deflection field to follow a certain pattern. Hence, deflections obtained are expected to be lesser than the actual deflections. Similarly, natural frequencies are expected to be more than the actual natural frequencies. Thus, as we refine a mesh in structural problems, we approach the exact deflections from lower deflections. Similarly, with mesh refinement, we approach the exact natural frequencies from higher natural frequencies.

This can be understood mathematically In the following manner: Consider the finite element formulation in which $[K]$ is the stiffness matrix, $\{u\}$ is the nodal deflection vector, and $\{P\}$ is the load vector. The system of equation to be solved is given by

$$[K]\{u\} = \{P\} \tag{19.1}$$

Total potential energy is given by

$$\Pi = \frac{1}{2}\{u\}^{\mathrm{T}}[K]\{u\} - \{u\}^{\mathrm{T}}\{P\} \tag{19.2}$$

In the expression given in Eq. (19.2), the first term on the right-hand side of the equation denotes strain energy, while the second term denotes the potential due to applied loads. Substituting Eq. (19.1) in Eq. (19.2), the total potential energy expression becomes

$$\Pi = \frac{1}{2}\{u\}^{\mathrm{T}}\{P\} - \{u\}^{\mathrm{T}}\{P\} = -\frac{1}{2}\{u\}^{\mathrm{T}}\{P\} \tag{19.3}$$

The potential energy given by this expression will be more than or equal to the actual potential energy because the potential energy gets minimized in the actual system. Hence

$$-\frac{1}{2}\{u\}^{\mathrm{T}}\{P\} \geq -\frac{1}{2}\{u_{\text{exact}}\}^{\mathrm{T}}\{P\} \tag{19.4}$$

or

$$\{u\}^{\mathrm{T}}\{P\} \leq \{u_{\text{exact}}\}^{\mathrm{T}}\{P\} \tag{19.5}$$

Equation (19.5) also implies that if there are n nodes and a load P_1 is applied at the first node, then

$$u_1 P_1 \leq (u_1)_{\text{exact}} P_1 \tag{19.6}$$

If the load is in positive direction, then the deflection will be in positive direction and

$$u_1 \leq (u_1)_{\text{exact}} \tag{19.7}$$

On the other hand, if the load is in negative direction, then the deflection will be in negative direction and

$$u_1 \geq (u_1)_{\text{exact}} \tag{19.8}$$

In any case, the magnitude of the deflection at node 1 will always be less than the exact deflection at that node. However, if external loads are acting at more than one point, it cannot be said that deflection at each node will be lower than the exact deflection. We can only say that inequality given by Eq. (19.6) will hold good.

It is interesting to note that in many cases, nodal values predicted by FEM are exact, although interpolation by these values may be approximate. There is a belief that in FEM, the primary solution at the nodes is more accurate than that at the rest of the places. It is difficult to provide a general proof for this statement. However, on the basis of the paper of Tong (1969), we shall show that it is possible to get the exact nodal displacement by FEM for a rod subjected to axial loads even if the complete interpolated solution is approximate.

Let us consider the rod problem discussed in Chapter 5. The differential equation governing the system given by Eq. (5.49) is reproduced here as follows:

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q=0 \tag{19.9}$$

Making the integral of weighted residual inside an element of length h zero, we obtain

$$\int_0^h w\left(\frac{\mathrm{d}}{\mathrm{d}x}\left(EA\frac{\mathrm{d}u}{\mathrm{d}x}\right)+q\right)\mathrm{d}x=0 \tag{19.10}$$

In Galerkin FEM, we integrate the first term of Eq. (19.10) by parts so as to distribute the order of derivative evenly on w and u. Thus, the weak form of Eq. (19.10) contains the first-order derivative on w as well as u. However, here we integrate by parts twice and make the derivative on w as second order. In doing so, we assume EA to be constant inside the element. Thus, we obtain

$$wEA\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_0^h-\frac{\mathrm{d}w}{\mathrm{d}x}EAu\bigg|_0^h+\int_0^h\frac{\mathrm{d}^2w}{\mathrm{d}x^2}EAu\mathrm{d}x+\int_0^h qw\mathrm{d}x=0 \tag{19.11}$$

If w is linear as is the case in Galerkin FEM with two-noded elements, then the second derivative of w vanishes and Eq. (19.11) can be written as

$$wEA\frac{\mathrm{d}u}{\mathrm{d}x}\bigg|_0^h - \frac{\mathrm{d}w}{\mathrm{d}x}EAu\bigg|_0^h + \int_0^h qw\mathrm{d}x = 0 \tag{19.12}$$

Assume that node 1 is at $x = 0$ and node 2 is at $x = h$ and denote the values at these nodes by suffixing the node numbers. Then, Eq. (19.12) can be written as

$$w_2EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_2 - w_1EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_1 - \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)_2 EAu_2 + \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)_1 EAu_2 + \int_0^h qw\mathrm{d}x = 0 \tag{19.13}$$

As w is linear, the value of dw/dx is the same everywhere in the element and hence the suffix can be dropped from it. Thus, Eq. (19.13) can be written as

$$w_2EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_2 - w_1EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_1 - \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)EAu_2 + \left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)EAu_2 + \int_0^h qw\mathrm{d}x = 0 \tag{19.14}$$

Equation (19.14) can be written as

$$\lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} - EA\left(\frac{\mathrm{d}w}{\mathrm{d}x}\right)\lfloor -1 \quad 1 \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} + \int_0^h qw\mathrm{d}x = 0 \tag{19.15}$$

Expressing the linear function w in terms of shape functions, we have

$$w = \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix} \tag{19.16}$$

where

$$N_1 = 1 - \frac{x}{h}; \quad N_2 = \frac{x}{h} \tag{19.17}$$

Therefore

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_{1,x} \\ N_{2,x} \end{Bmatrix} = \lfloor w_1 \quad w_2 \rfloor \frac{1}{h}\begin{Bmatrix} -1 \\ 1 \end{Bmatrix} \tag{19.18}$$

Substituting Eqs (19.16) and (19.18) in Eq. (19.15), we get

$$\lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} -EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ EA\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} - EA\lfloor w_1 \quad w_2 \rfloor \frac{1}{h}\begin{Bmatrix} -1 \\ 1 \end{Bmatrix}\lfloor -1 \quad 1 \rfloor \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} + \int_0^h q\lfloor w_1 \quad w_2 \rfloor \begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\mathrm{d}x = 0 \tag{19.19}$$

In view of the nodal weights being arbitrary, Eq. (19.19) implies

$$\begin{Bmatrix} -EA\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ EA\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} - \frac{EA}{h}\begin{Bmatrix} -1 \\ 1 \end{Bmatrix}\begin{bmatrix} -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} + \int_0^h q\begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\mathrm{d}x = 0 \tag{19.20}$$

After simplification, Eq. (19.20) can be written as

$$\frac{EA}{h}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \int_0^h q\begin{Bmatrix} N_1 \\ N_2 \end{Bmatrix}\mathrm{d}x + \begin{Bmatrix} -EA\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_1 \\ EA\left(\dfrac{\mathrm{d}u}{\mathrm{d}x}\right)_2 \end{Bmatrix} \tag{19.21}$$

Note that Eq. (19.21) is the system of FEM elemental equations. You can compare it with Eq. (5.59) and verify that both are same, although Eq. (19.21) has been derived in a different manner. In this derivation, no assumption was made regarding the form of u inside the element and Eq. (19.21) is not dependent on it. In fact, from Eq. (19.9) to Eq. (19.21), all the equations are valid for exact u. Hence, the nodal values u_1 and u_2 will be exact provided the integral on right-hand side of Eq. (19.21) is evaluated exactly. [Note that after assembly and application of load, the internal load vector will vanish. Nevertheless if after obtaining the exact nodal values, the first derivatives of displacement are calculated at the nodes using Eq. (19.21), we shall get exact values. However, we often take the first derivative in a two-noded element as $(u_2 - u_1)/h$, which may not be exact at the nodes.]

In all physical problems in which the governing differential equation is of a form similar to Eq. (19.19), the nodal values will be exact for a finite element analysis with two-noded elements. In fact, it can be shown that in finite element analysis of Euler-Bernoulli beam, in which the governing equation is given by (Eq. 6.1) as

$$EI\frac{\mathrm{d}^4 v}{\mathrm{d}x^4} = q \tag{19.22}$$

the values of slopes and deflections will be exact by finite element analysis with two-noded beam elements. The weight function has to be cubic.

19.3 ACCURACY OF DERIVATIVES OF PRIMARY VARIABLES

Deflections are expected to be most accurate at the nodes, while stresses are expected to be most accurate at certain points called *Barlow points* [Barlow (1976)], which often coincide with the Gauss quadrature points. In two-noded, one-dimensional elements, stresses are expected to be accurate at the center. Similarly, in a three-noded, one-dimensional element with end-point coordinates of -1 and 1, stresses are expected to be accurate at coordinates $\pm 1/\sqrt{3}$. In beam elements, stresses are expected to be accurate at natural coordinates $\pm 1/\sqrt{3}$ that correspond to points of two-Gauss-point formula. A similar condition is applicable in case of elements of higher dimensions. In four-noded rectangular

elements, stresses are expected to be most accurate at the center. In eight- and nine-noded rectangular elements, the four Barlow points are given by ($\pm 1/\sqrt{3}, \pm 1/\sqrt{3}$). However, a rigorous proof for higher dimensional elements is not available.

Actually, the proof for most accuracy points is carried out in the following manner. We assume that nodal values of primary derivatives are accurate and the exact solution can be represented by a polynomial function one degree higher than the FEM approximation. This is the case when mesh is sufficiently refined. Thus, derivative values at Barlow points are expected to be highly accurate in a sufficiently refined mesh only. Derivative values at Barlow points may be interpolated (or extrapolated to some extent) for obtaining the estimate of derivatives at other points.

19.4 ESSENTIAL AND NATURAL BOUNDARY CONDITIONS

In structural mechanics, the essential boundary conditions correspond to prescribed displacement and rotations. If in the integral of the variational form, the highest derivative of a state variable (with respect to a space coordinate) is m, the order of derivative in the essential boundary conditions is at most $(m - 1)$. The second class of boundary conditions, namely the natural boundary conditions are also called force boundary conditions (as opposed to forced boundary conditions, which some authors have used for essential boundary conditions) because in structural mechanics, natural boundary conditions correspond to prescribed boundary forces and moments. The highest derivatives in these boundary conditions are of order m to $(2m - 1)$.

Essential boundary conditions are also known as *Dirichlet* or *geometric boundary conditions* and natural boundary conditions are also known as *Neumann boundary conditions*. Robin boundary conditions are the weighted combination of Dirichlet boundary conditions and Neumann boundary conditions. For example, the following boundary condition, corresponding to rod supported by a spring, is a Robin boundary condition:

$$EA\frac{\mathrm{d}u}{\mathrm{d}x} + ku = 0 \tag{19.23}$$

19.5 MESH REFINEMENT

The solution accuracy is very much dependent on the type of mesh. Modifying the mesh with a view to obtain a more accurate solution is called *mesh refinement*. There are four methods of carrying out the mesh refinement. They are discussed below.

(i) ***h*-refinement** In this refinement method, the size of the element is reduced. Consequently, the number of elements in a domain increases. This refinement method is most commonly used. However, in this method, the mesh has to be generated again.

(ii) ***p*-refinement** In this refinement method, element sizes remain the same but the degree of polynomial approximation in an element increases. The advantages of this method include faster convergence and avoidance of new mesh generation at each refinement stage.

(iii) ***hp*-refinement** This is the combination of both the above methods. It is a very effective method for most of the problems. However, developing a suitable strategy for this type of refinement is still a challenging task.

(iv) ***r*-method** In this method, nodes are relocated and a new mesh is constructed. The size of the elements and degrees of approximation remain the same. This is not a very commonly used method.

Refinement should be carried out in systems where the primary variables are expected to be polynomial functions of high degrees. For example, if a plate contains a hole, then the mesh should be refined near the hole. If a cantilever beam problem is solved using constant strain triangular (CST) elements, then there should be enough elements in the vertical directions also because the strain is linear in vertical direction and CST gives only constant strain. Therefore, many elements are needed in the vertical direction in order to properly represent the strain.

19.6 EFFECT OF GEOMETRY OF AN ELEMENT ON FEM

If the shape of the element is not proper, computational difficulties are introduced. For example, while using CST elements, three angles of the triangle should be close to 60°. Similarly, in a rectangular element, the ratio of higher side to lower side should not be more than 10. A square element is better. As the aspect ratio (ratio of higher side to lower side) increases, contribution due to parasitic shear increases and locking phenomenon gets activated. Parasitic shear is the presence of spurious shear when a rectangular element is subjected to pure bending. It is briefly explained below.

Consider a square element in natural coordinate system, which is subjected to pure bending. The deformed and undeformed elements are shown in Fig. 19.1(a), deformed elements are denoted by thick line, while undeformed elements are shown by thin line. As a result of pure bending, there is no displacement of the element in η direction. The displacements of nodes 1, 2, 3, and 4 in ξ direction are assumed as c, $-c$, c, and $-c$, respectively, where c indicates the extent of the bending. After deformations, the nodes occupy the positions indicated by 1′, 2′, 3′ and 4′ respectively. Inside the element, the interpolated displacement function u is given as

$$u = N_1 u_1 + N_2 u_2 + N_3 u_3 + N_4 u_4 \tag{19.24}$$

Substituting the nodal values and expression for shape functions (Eq. 11.39) in Eq. (19.24), we get

$$u = \frac{(1-\xi)(1-\eta)}{4}c - \frac{(1+\xi)(1-\eta)}{4}c + \frac{(1+\xi)(1+\eta)}{4}c - \frac{(1-\xi)(1-\eta)}{4}u_4 = \xi\eta c \tag{19.25}$$

Obviously, v is zero throughout the element. Now, let us calculate the shear stress in the element. The shear stress is given by

$$\gamma = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \tag{19.26}$$

We note that for this element

$$\frac{\partial u}{\partial \eta} = \xi c \quad \text{and} \quad \frac{\partial v}{\partial \xi} = 0 \tag{19.27}$$

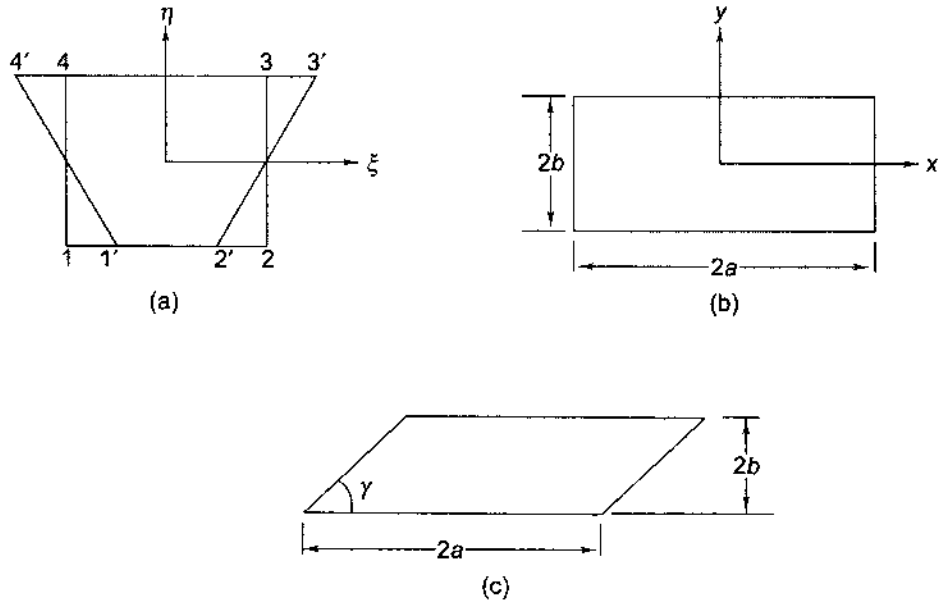


Figure 19.1 *Concept of parasitic shear (a) deformation of an element in bending in natural coordinates (b) undeformed element in physical coordinates (c) deformed element in physical coordinates*

For calculating the shear strains, derivatives with respect to physical coordinates are to be taken. Assume that in physical coordinates, the element is rectangular as shown in Fig. 19.1 (b). The physical and natural coordinates are related as

$$x = \sum_{i=1}^{4} N_i x_i, \quad y = \sum_{i=1}^{4} N_i y_i \tag{19.28}$$

In our case

$$x = a\xi \quad \text{and} \quad y = b\eta \tag{19.29}$$

Now, we can use Jacobian for transforming the derivatives from natural to physical coordinates. However, let us do it more fundamentally. By chain rule, we have

$$\frac{\partial u}{\partial \eta} = \frac{\partial u}{\partial x}\frac{\partial x}{\partial \eta} + \frac{\partial u}{\partial y}\frac{\partial y}{\partial \eta} \tag{19.30}$$

$$\frac{\partial v}{\partial \varepsilon} = \frac{\partial v}{\partial x}\frac{\partial x}{\partial \xi} + \frac{\partial v}{\partial y}\frac{\partial y}{\partial \xi} \tag{19.31}$$

Substituting Eq. (19.29) in Eqs (19.30) and (19.31) we get

$$\frac{\partial u}{\partial \eta} = b\frac{\partial u}{\partial y} \quad \text{or} \quad \frac{\partial u}{\partial y} = \frac{1}{b}\frac{\partial u}{\partial \eta} \tag{19.32}$$

and

$$\frac{\partial v}{\partial \varepsilon} = a\frac{\partial v}{\partial x} \quad \text{or} \quad \frac{\partial v}{\partial x} = \frac{1}{a}\frac{\partial v}{\partial \varepsilon} \tag{19.33}$$

Substituting Eqs (19.32) and (19.33) in Eq. (19.26) and noting that $v = 0$, we get

$$\gamma = \frac{1}{b}\frac{\partial u}{\partial \eta} \tag{19.34}$$

Substituting Eq. (19.27) in Eq. (19.34), we get

$$\gamma = \varepsilon\frac{c}{b} \tag{19.35}$$

Equation (19.35) shows that even in the case of pure bending, a rectangular element generates a shear strain except at $\xi = 0$. This spurious shear strain is called *parasitic shear strain* and the corresponding stress as *parasitic shear stress*. This occurs because in pure bending, top and bottom lines of the rectangular are constrained to remain straight as shown in Fig. 19.1(c) and do not become curved (as they should be).

Equation (19.35) shows that if b is very small, the parasitic shear strain will be very high. For the same area of the element, if b very small, the proportion of the energy due to shear strain to total elastic energy of the element will increase. Therefore, b should not be very small. In this case, neutral axis of the bending was horizontal. For a bending, in which the neutral axis is vertical, we can carry out similar analysis and conclude that a should be very small. In practice, loading can be of any type. Therefore, both a and b should be almost equal. In any case, neither a nor b should be very small. If any of these dimensions is very small, spurious portion of the energy due to parasitic shear will dominate. This phenomenon is called *locking*. One way to avoid this is to use only one Gauss point for estimating the stiffness matrix due to shear. At the point corresponding to one-Gauss-point formula, parasitic shear strain will be 0.

In a quadrilateral element, angle between two sides should not be much greater than 90°. In an eight-noded, rectangular element, the middle node should lie between 1/4th to 3/4th of length of the side from the end nods. It should preferably be at the midpoint. However, in solving fracture mechanics problems, in some elements, one node is deliberately kept at 1/4th position of the length of the side. This is explained in the next section.

19.7 USING FEM IN FRACTURE MECHANICS

As mentioned earlier, FEM is a numerical method used for solving differential and integral equations. It has been applied to a number of physical problems, where the governing differential equations are

available. Linear elastic fracture mechanics (LEFM) is one of the areas where governing differential equations are well-established. However, except for simple geometries, the closed-form expressions for stress distributions cannot be obtained. FEM, therefore, finds a role in solving LEFM problems.

Most of the applications of FEM have been in determining stress intensity factor. The stress intensity factor is proportional to far-field stress and square root of crack length and is denoted by K. In SI system, its unit is N $m^{-3/2}$. When the stress intensity factor reaches a critical value called *fracture toughness*, fracture impends. Fracture toughness is a material property; thus, it is important for designers to determine stress intensity factor for producing high–reliability products.

The stress intensity factor is not the only way of predicting the onset of fracture. One can also find out the energy release rate G, which is defined as energy release per unit increase in area, during crack growth. The energy requirement R for a crack to grow per unit area extension is called *crack resistance*. For a crack to grow and become critical, firstly, the energy release rate should be greater than crack resistance, and secondly, the rate of increase of G with crack length should be more than the rate of increase of R with crack length. The parameter G is applied to brittle or less ductile material. J-integral has been developed to deal with ductile as well as brittle materials. J-integral represents a way to calculate energy per unit surface area in a material. In fact, G is a special case of J-integral.

Plane stress and plane strain problems of fracture mechanics may be solved by using plane stress and plane strain FEM formulations, respectively. The value of stresses and displacements near the crack tip can be used for finding out the stress intensity factors. In general, the displacement components (in r–θ coordinate system) near the crack tips are given by

$$u_i = K_I\sqrt{\frac{r}{2\pi}}f_i^I(\theta) + K_{II}\sqrt{\frac{r}{2\pi}}f_i^{II}(\theta) + K_{III}\sqrt{\frac{r}{2\pi}}f_i^{III}(\theta) \tag{19.36}$$

and stress components are given by

$$\sigma_{ij} = \frac{K_I}{\sqrt{2\pi r}}g_{ij}^I(\theta) + \frac{K_{II}}{\sqrt{2\pi r}}g_{ij}^{II}(\theta) + \frac{K_{III}}{\sqrt{2\pi r}}g_{ij}^{III}(\theta) \tag{19.37}$$

where f and g are functions of θ, K denotes the stress intensity factor, and I, II, and III denote the three modes of the fracture. For example, for mode 1 problem of plane strain, the various components can be written as

$$\sigma_{11} = \frac{K_I}{(2\pi r)^{1/2}}\cos\frac{\theta}{2}\left[1 - \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right] \tag{19.38a}$$

$$\sigma_{22} = \frac{K_I}{(2\pi r)^{1/2}}\cos\frac{\theta}{2}\left[1 + \sin\frac{\theta}{2}\sin\frac{3\theta}{2}\right] \tag{19.38b}$$

$$\sigma_{12} = \frac{K_I}{(2\pi r)^{1/2}}\sin\frac{\theta}{2}\cos\frac{\theta}{2}\cos\frac{3\theta}{2} \tag{19.38c}$$

$$u_1 = \frac{K_I}{\mu}\left(\frac{r}{2\pi}\right)^{1/2}\cos\frac{\theta}{2}\left[1 - 2\nu + \sin^2\frac{\theta}{2}\right] \tag{19.38d}$$

$$u_2 = \frac{K_I}{\mu}\left(\frac{r}{2\pi}\right)^{1/2} \sin\frac{\theta}{2}\left[1 - 2\nu + \cos^2\frac{\theta}{2}\right] \tag{19.39}$$

where μ is the shear modulus and ν is the Poisson's ratio.

We can evaluate K_I from stresses as

$$K_I = \sigma_{ij}\frac{\sqrt{2\pi r}}{g_{ij}^I(\theta)} \tag{19.40}$$

or from displacement field as

$$K_I = u_i\frac{\sqrt{2\pi}}{\sqrt{r}f_i^I(\theta)} \tag{19.41}$$

Since the stress gradient near the crack tip is very high and stress becomes singular at the crack tip, a fine mesh is required. An alternate method is the *J-integral method*. The J-integral is given by

$$J = \int_\Gamma \left(W\mathrm{d}y - t_i\frac{\partial u_i}{\partial x}\mathrm{d}s\right) \tag{19.42}$$

where

$$W = \int_0^\epsilon \sigma_{ij}\mathrm{d}\varepsilon_{ij} \tag{19.43}$$

is the strain energy density and t_i is the traction. Here, crack propagation direction is along x.

In general, a very fine mesh is required near the crack tip due to large stress gradient. The stress field displays a singularity of order $r^{-1/2}$ [see Eq. (19.37)]. An element having side nodes can be made to display a $r^{-1/2}$ stress/strain singularity. Consider a quarter point bar element (Fig. 19.2), in which the third node is at a distance of $L/4$ from one end. This element is called *quarter point element*. By using

$$x = N_1x_1 + N_2x_2 + N_3x_3 \tag{19.44}$$

where N_1, N_2, and N_3 are the Lagrangian shape functions for a three-noded element, this element yields

$$x = \frac{L}{4}(1+\xi)^2 \tag{19.45}$$

or

$$\xi = 2(x/L)^{1/2} - 1 \tag{19.46}$$

The $[B]$ matrix (the row vector containing the derivatives of shape function) is written as

$$[B] = \left[\left(\frac{2}{L} - \frac{3}{2(Lx)^{1/2}}\right)\left(\frac{2}{L} - \frac{1}{2(Lx)^{1/2}}\right)\left(\frac{4}{L} + \frac{2}{(Lx)^{1/2}}\right)\right] \tag{19.47}$$

Note that pre-multiplication of $[B]$ matrix with nodal displacement vector will give the strain as

$$\varepsilon = \frac{\mathrm{d}u}{\mathrm{d}x} = \left(\frac{2}{L} - \frac{3}{2(Lx)^{1/2}}\right)u_1 + \left(\frac{2}{L} - \frac{1}{2(Lx)^{1/2}}\right)u_2 + \left(\frac{2}{L} - \frac{1}{2(Lx)^{1/2}}\right)u_3 \tag{19.48}$$

As is clear from Eq. (19.48), in this element, the stresses vary as $x^{-1/2}$. Thus, there is a singularity at $x = 0$, where the crack tip may be located.

The six-noded triangular element can display singularity if side nodes are moved to quarter points near the mid-side (Fig. 19.3). The rectangular quarter point elements (Fig. 19.4) display the $r^{-1/2}$ singularity along two sides and the diagonal.

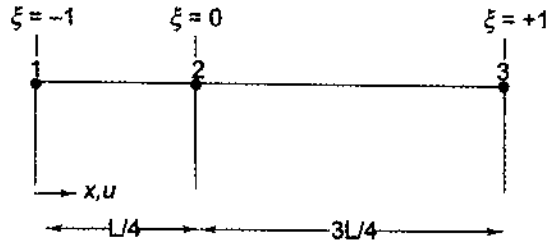


Figure 19.2 *Three-node quarter-point bar element*

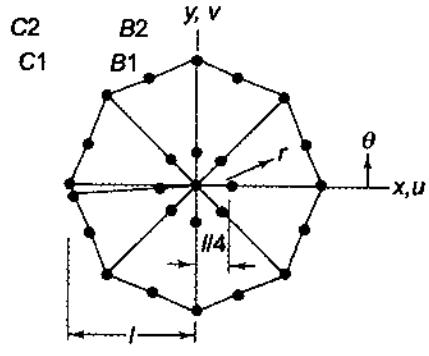


Figure 19.3 *Mesh of quarter plate elements around the crack tip*

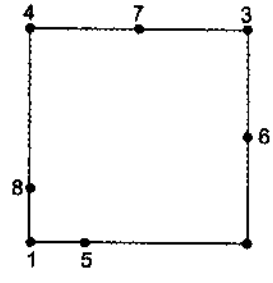


Figure 19.4 *A rectangular quarter point element*

19.8 INFINITE ELEMENTS

Many stress analysis problems deal with unbounded media, for example, the problem of a load supported on the ground. In this problem, a finite element model must be terminated somewhere short of infinity. The analysis then becomes expensive and many elements have to be taken. In order to solve such problems economically, a special type of element called *infinite element* can be used. The shape functions of this element will cause the stresses to reach to zero at infinity.

Two types of shape functions can be used: decay shape functions and growth shape functions. *Decay shape functions* approach zero as coordinate approaches infinity, whereas *growth shape functions* grow to infinity at one particular node. Decay shape functions are applied to stresses while the normal shape functions are used for geometry. On the other hand, growth shape functions are used to approximate geometry, while the stresses are approximated by the normal shape functions.

As an example of an infinite element [Zienkiewicz, et al. (1983)], consider the following mapping from natural coordinate ξ to physical coordinate x:

$$x = \frac{-\xi}{1-\xi} x_0 + \frac{1}{1-\xi} x_2 \tag{19.49}$$

At $\xi = 1$, Eq. (19.49) provides $x = \infty$. At $\xi = -1$

$$x = \frac{x_0 + x_2}{2} \tag{19.50}$$

Thus, the finite element in natural coordinates with end coordinates as -1 and 1 gets transformed into an infinite element with end coordinates as $(x_0 + x_2)/2$ and ∞. As the finite element formulation and programming is done in natural coordinates, use of interpolation function given by Eq. (19.49) can solve the problems of infinite extent.

19.9 ILL-CONDITIONED SYSTEMS

Consider the set of equations

$$\begin{bmatrix} 1.00 & -1.00 \\ -1.00 & 1.02 \end{bmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 4.00 \\ -2.00 \end{Bmatrix} \tag{19.51}$$

for which

$$\begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 104 \\ 100 \end{Bmatrix} \tag{19.52}$$

and a very similar set of equations

$$\begin{bmatrix} 1.00 & -1.00 \\ -1.00 & 1.01 \end{bmatrix} \begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 4.00 \\ -2.00 \end{Bmatrix} \tag{19.53}$$

for which

$$\begin{Bmatrix} x \\ y \end{Bmatrix} = \begin{Bmatrix} 204 \\ 200 \end{Bmatrix} \tag{19.54}$$

About 1% change in one coefficient has changed the results by a factor of almost two. Both these sets of equations [Eqs (19.51) and (19.53)] are ill-conditioned, which implies that their solutions are sensitive to small changes in the coefficient matrix and the right-hand-side vector.

In finite element method, an ill-conditioned system is produced because of the physics of the problem. A simple example of an ill-conditioned system is provided by Cook, et al. (1989). Consider a structure with two degrees of freedom which has linear springs of stiffness k_1 and k_2 (Fig. 19.5). FEM formulation provides

$$\begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1 + k_2 \end{bmatrix} = \begin{Bmatrix} P \\ 0 \end{Bmatrix} \tag{19.55}$$

The rows of stiffness matrix are almost linearly dependent if $k_1 >> k_2$ and hence the stiffness matrix becomes ill-conditioned. This is not the case if $k_2 >> k_1$.

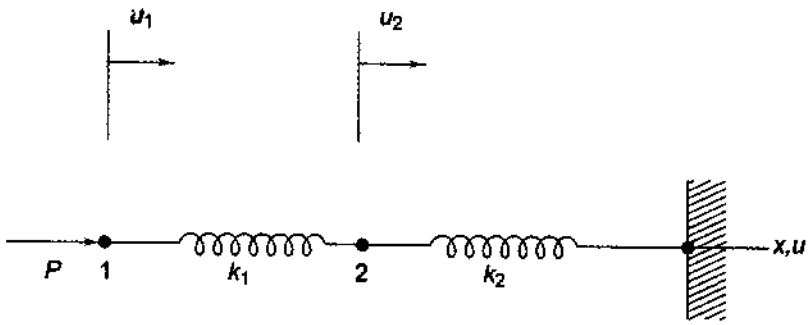


Figure 19.5 *An example that can yield ill-conditioned system if $k_1 >> k_2$*

A quantitative measure of ill-conditioning is the *condition number*. An ideal well-conditioned system has a condition number of 1. The higher the condition number, the more ill-conditioned the system is. However, there are different definitions of the condition number and its value depends on the definition used for calculating it. For a real, symmetric coefficient matrix $[A]$, one definition of condition number $k([A])$is

$$k([A]) = \left| \frac{\text{Maximum eigenvalue}}{\text{Miniumum eigenvalue}} \right| \tag{19.56}$$

For Eq. (19.55), for the case when $k_1 >> k_2$, the approximate coefficient matrix $[A]$ is given as

$$[A] = \begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_1 \end{bmatrix} \tag{19.57}$$

which has the eigenvalues as 0 and $2k_1$. Thus, from Eq. (19.56), the condition number is infinite. In actual case, k_2 is a small non-zero quantity. Hence, the condition number will not be infinite but a very large value. For the case when $k_2 >> k_1$, the coefficient matrix is given by

$$[A]=\begin{bmatrix} k_1 & -k_1 \\ -k_1 & k_2 \end{bmatrix} \tag{19.58}$$

It can be seen that the coefficient matrix given in Eq. (19.38) will have lesser value of condition number than the coefficient matrix given by Eq. (19.57).

In structural analysis, the major cause of ill-conditioning is the large difference in stiffnesses, with the stiffer region being supported by the more flexible region. Physically, the stiffer region has one or more displacement states that are almost rigid-body motions within a more flexible supporting structure. The limiting case is a structure without any supports: it has only rigid-body motion in static analysis, and its stiffness matrix is singular.

The following measures should be adopted for avoiding ill-conditioning of systems:

(i) It is recommended that double-precision arithmetic be used for all phases of finite element analysis.

(ii) Attempt should be to develop a model without large discrepancies in stiffness. A comparatively stiff region may be modelled as perfectly rigid by use of a constraint transformation. On the other hand, perhaps a stiff region can be made more flexible.

(iii) One possibility in modelling is to use relative motions, rather than absolute motions, for troublesome degrees of freedom. For example, in Fig. 19.4, if we introduce the degree of freedom. $u_r = u_1 - u_2$, we would obtain the following equation instead of Eq. (19.55):

$$\begin{bmatrix} k_1 & 0 \\ 0 & k_2 \end{bmatrix}\begin{Bmatrix} u_r \\ u_2 \end{Bmatrix}=\begin{Bmatrix} P \\ P \end{Bmatrix} \tag{19.59}$$

This system is well-conditioned for all values of k_1 and k_2. For $k_1 > k_2$, the condition number of the system of equations in Eq. (19.59) is given by k_1/k_2.

19.10 PATCH TEST

In some cases, considerable difficulty is experienced in finding displacement functions for an element which will automatically be continuous along the whole interface between adjacent elements. The discontinuity of displacement will cause infinite strains at the interfaces. However, if, in the limit, as the size of the subdivision decreases continuity is restored, then formulation will tend to the correct answer. This condition is always reached if

(i) a constant strain condition automatically ensures displacement continuity.

(ii) the constant strain criterion is satisfied, i.e., if nodal displacements are compatible with a constant strain condition such constant strain will in fact be obtainable.

To test whether a non-conforming element will provide satisfactory result, a patch test is conducted. In this test, nodal displacements corresponding to any state of constant strain are imposed on an arbitrary patch of elements. If nodal equilibrium is simultaneously achieved without the imposition of external, nodal forces and if a state of constant stress is obtained, then clearly no external work has been lost through inter-element discontinuity. Elements that pass such a patch test will converge.

The patch test is equally applicable in case of non-structural problems. In general, the patch of elements is subjected to loading for which the exact solution is known. The elements pass the patch test if the finite element solution is the same as the exact solution. Patch test also helps in detecting programming errors. For further details, refer Zienkiewicz, et al., (2005).

19.11 THREE-DIMENSIONAL FINITE ELEMENTS

The three-dimensional finite element formulation is carried out in a manner similar to the two-dimensional formulation. Many a times, the governing equations in two and three dimensions appear to be almost same when expressed using matrix notation. For example, the mass matrix in two-dimensional plane stress problem is given by

$$[m^e] = \int_{\Omega^e} \rho t [N]^{\mathrm{T}} [N] \, \mathrm{d}x\mathrm{d}y \tag{19.60}$$

and the mass matrix for three-dimensional solid is given by

$$[m^e] = \int_{V^e} \rho [N]^{\mathrm{T}} [N] \, \mathrm{d}x\mathrm{d}y\,\mathrm{d}z \tag{19.61}$$

Of course, the matrix $[N]$ is different in Eqs (19.60) and (19.61). In view of it we shall not describe the FEM formulation for a three-dimensional problem here, but will give an example of it in Solved Problem 19.2. In the following paragraphs, we shall describe some three-dimensional elements and will provide the expressions for their shape functions.

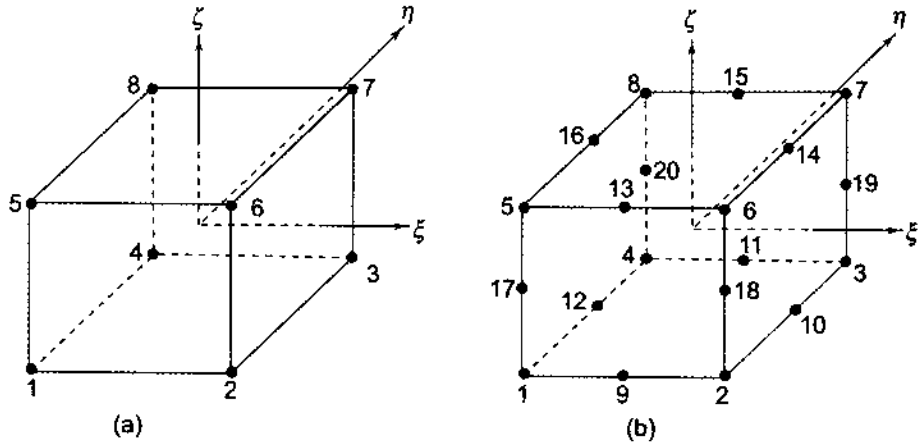


Figure 19.6 *Brick elements: (a) 8-noded (b) 20-noded*

A three-dimensional brick element shown in Fig. 19.6a consists of a cube of side 2 units in natural coordinates. The origin of the natural coordinates system $\xi - \eta - \zeta$ is at the centroid of the cube. With tri-linear approximation for primary variables, the shape functions are given by

$$
\begin{aligned}
N_1 &= \frac{1}{8}(1-\xi)(1-\eta)(1-\zeta) & N_2 &= \frac{1}{8}(1+\xi)(1-\eta)(1-\zeta) \\
N_3 &= \frac{1}{8}(1+\xi)(1+\eta)(1-\zeta) & N_4 &= \frac{1}{8}(1-\xi)(1+\eta)(1-\zeta) \\
N_5 &= \frac{1}{8}(1-\xi)(1-\eta)(1+\zeta) & N_6 &= \frac{1}{8}(1+\xi)(1-\eta)(1+\zeta) \\
N_7 &= \frac{1}{8}(1+\xi)(1+\eta)(1+\zeta) & N_8 &= \frac{1}{8}(1-\xi)(1+\eta)(1+\zeta)
\end{aligned}
\tag{19.62}
$$

A three-dimensional, 20-noded brick element is shown in Fig. 19.6b. The shape functions of this element are as follows:

(a) *Mid-side nodes*

$$
\begin{aligned}
N_9 &= \frac{(1-\xi^2)(1-\eta)(1-\zeta)}{4} & N_{10} &= \frac{(1-\eta^2)(1+\xi)(1-\zeta)}{4} \\
N_{11} &= \frac{(1-\xi^2)(1+\eta)(1-\zeta)}{4} & N_{12} &= \frac{(1-\eta^2)(1-\xi)(1-\zeta)}{4} \\
N_{13} &= \frac{(1-\xi^2)(1-\eta)(1+\zeta)}{4} & N_{14} &= \frac{(1-\eta^2)(1+\xi)(1+\zeta)}{4} \\
N_{15} &= \frac{(1-\xi^2)(1+\eta)(1+\zeta)}{4} & N_{16} &= \frac{(1-\eta^2)(1-\xi)(1+\zeta)}{4} \\
N_{17} &= \frac{(1-\zeta^2)(1-\xi)(1-\eta)}{4} & N_{18} &= \frac{(1-\zeta^2)(1+\xi)(1-\eta)}{4} \\
N_{19} &= \frac{(1-\zeta^2)(1+\xi)(1+\eta)}{4} & N_{20} &= \frac{(1-\zeta^2)(1-\xi)(1+\eta)}{4}
\end{aligned}
\tag{19.63}
$$

(b) *Corner nodes*

$$
\begin{aligned}
N_1 &= \frac{1}{8}(1-\xi)(1-\eta)(1-\zeta) - \frac{1}{2}(N_9 + N_{12} + N_{17}) \\
N_2 &= \frac{1}{8}(1+\xi)(1-\eta)(1-\zeta) - \frac{1}{2}(N_9 + N_{10} + N_{18}) \\
N_3 &= \frac{1}{8}(1+\xi)(1+\eta)(1-\zeta) - \frac{1}{2}(N_{10} + N_{11} + N_{19}) \\
N_4 &= \frac{1}{8}(1-\xi)(1+\eta)(1-\zeta) - \frac{1}{2}(N_{11} + N_{12} + N_{20}) \\
N_5 &= \frac{1}{8}(1-\xi)(1-\eta)(1+\zeta) - \frac{1}{2}(N_{13} + N_{16} + N_{17}) \\
N_6 &= \frac{1}{8}(1+\xi)(1-\eta)(1+\zeta) - \frac{1}{2}(N_{13} + N_{14} + N_{18}) \\
N_7 &= \frac{1}{8}(1+\xi)(1+\eta)(1+\zeta) - \frac{1}{2}(N_{14} + N_{15} + N_{19}) \\
N_8 &= \frac{1}{8}(1-\xi)(1+\eta)(1+\zeta) - \frac{1}{2}(N_{15} + N_{16} + N_{20})
\end{aligned}
\tag{19.64}
$$

It can be easily verified that all these shape functions are zero at all nodes except at one node, where they adopt a value 1. Also, the sum of the shape functions is 1. One can also use a 27-noded brick element that is a natural extension to a nine-noded quadrilateral element.

Tetrahedral elements are shown in Fig. 19.7. It is common to use four natural coordinates ξ_1, ξ_2, ξ_3, and ξ_4 for a tetrahedron, out of which only three are independent as they satisfy the following relation:

$$\xi_1 + \xi_2 + \xi_3 + \xi_4 = 1 \tag{19.65}$$

The coordinates of four vertices of the tetrahedron in natural coordinates are expressed as (1, 0, 0, 0), (0, 1, 0, 0), (0, 0, 1, 0), and (0, 0, 0, 1), respectively. We can express the coordinates of a point P, in terms of natural coordinates ξ_1, ξ_2, ξ_3, and ξ_4 as

$$\xi_1 = \frac{\text{Volume of tetrahedron P234}}{\text{Volume of tetrahedron 1234}} \quad \xi_2 = \frac{\text{Volume of tetrahedron P134}}{\text{Volume of tetrahedron 1234}}$$
$$\xi_3 = \frac{\text{Volume of tetrahedron P124}}{\text{Volume of tetrahedron 1234}} \quad \xi_4 = \frac{\text{Volume of tetrahedron F123}}{\text{Volume of tetrahedron 1234}} \tag{19.66}$$

The volume V of a tetrahedron is given by

$$V = \frac{1}{6}\begin{vmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \\ x_4 & y_4 & z_4 & 1 \end{vmatrix} \tag{19.67}$$

where the coordinates of vertex i are (x_i, y_i, z_i). In terms of natural coordinates, the shape functions of a four-noded tetrahedral element (Fig. 19.7a) are

$$N_1 = \xi_1, \quad N_2 = \xi_2, \quad N_3 = \xi_3, \quad N_4 = \xi_4 \tag{19.68}$$

The shape functions of a ten-noded tetrahedral element [Fig. (19.7b)] are

(a) *Mid-side nodes*

$$N_5 = 4\xi_1\xi_2 \quad N_6 = 4\xi_2\xi_3 \quad N_7 = 4\xi_1\xi_3$$
$$N_8 = 4\xi_1\xi_4 \quad N_9 = 4\xi_2\xi_4 \quad N_{10} = 4\xi_3\xi_4 \tag{19.69}$$

(b) *Corner nodes*

$$N_1 = \xi_1(2\xi_1 - 1) \quad N_2 = \xi_2(2\xi_2 - 1)$$
$$N_3 = \xi_3(2\xi_3 - 1) \quad N_4 = \xi_4(2\xi_4 - 1) \tag{19.70}$$

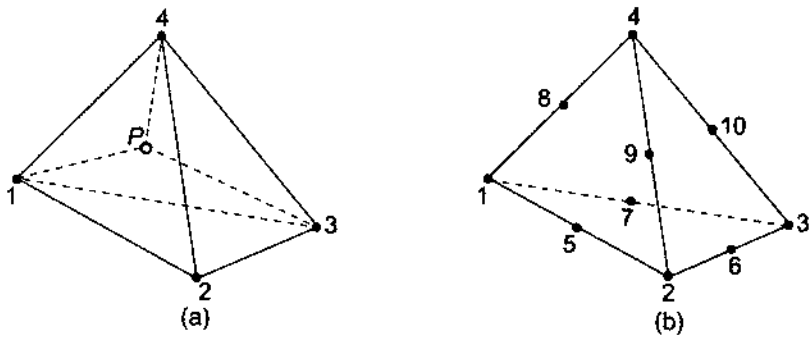


Figure 19.7 *Tetrahedral elements with (a) four nodes (b) ten nodes*

19.12 CONCLUSION

In this chapter we have discussed a number of topics in FEM. The reader can go through the references provided at the end of this chapter for further details. With this we conclude the book. After carefully reading the book and solving the exercise problems, the reader will be able to refer other advanced books on FEM as well as research papers.

SOLVED PROBLEMS

19.1 Consider a three-noded bar element with three nodes at location $x = 0$, 0.1, and 1. This element is used for solving the problem of bar subjected to axial loads. An isoparametric formulation is used. At what point of the element will the strain become infinite? What is the problem with this type of element?

Solution In physical coordinates, an element is shown in Fig. 19.8.

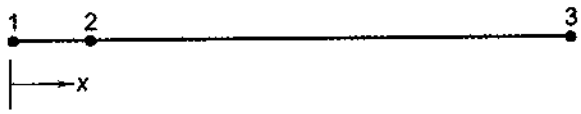


Figure 19.8 *An element in physical coordinates*

In natural coordinates, the shape functions of a three-noded element are given as

$$N_1 = \frac{\xi(\xi-1)}{2}, N_2 = 1-\xi^2,\ N_3 = \frac{\xi(\xi+1)}{2}$$

In iso-parametric formulation, the geometry and physical variables are approximated in the same manner. Thus

$$x = N_1x_1 + N_2x_2 + N_3x_3 = (1-\xi^2)0.1 + \frac{\xi(\xi+1)}{2} = 0.1 + 0.5\xi + 0.4\xi^2 \tag{19.1.1}$$

$$u = N_1u_1 + N_2u_2 + N_3u_3 = \frac{\xi(\xi-1)}{2}u_1 + (1-\xi^2)u_2 + \frac{\xi(\xi+1)}{2}u_3 \tag{19.1.2}$$

From Eq. (19.1.1), we can write

$$\frac{dx}{d\xi} = 0.5 + 0.8\xi \tag{19.1.3}$$

Using Eqs (19.1.2) and (19.1.3), the strain is calculated as

$$\varepsilon = \frac{du}{dx} = \frac{du}{d\varepsilon}\frac{d\varepsilon}{dx} = \frac{(\xi-0.5)u_1 - 2\xi u_2 + (\xi+0.5)u_3}{0.5+0.8\xi} \tag{19.1.4}$$

At $\xi = -(5/8)$, the denominator of Eq (19.1.4) will become 0, thus producing infinite strain. From Eq. (19.1.1), the value of x at this point is -0.0562. Thus, this point does not exist in physical coordinate.

This element is not suitable for FEM formulation. In this element, certain points in natural coordinates do not have a representation in the physical coordinates. For example, $\xi = -0.5$ gets mapped to non-existent point $x = -0.05$. Also, as Eq. (19.1.1) is a quadric equation, corresponding to a physical coordinate, there is a possibility of having two points in natural coordinates. For example, point $x = 0$, gets mapped to point -1 and $-1/4$ in natural coordinates.

19.2 Obtain the finite element formulation of a three-dimensional heat-conduction problem.

Solution The governing differential equation for a heat-conduction problem is given as

$$\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right)+\dot{Q}=0 \tag{19.2.1}$$

where $\dot{Q}$ is the rate of heat generation per unit volume. Let the approximate temperature function in an element be T. For brevity, we shall not use superscript e on the temperature. The residual in the element is given by

$$R=\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right)+\dot{Q} \tag{19.2.2}$$

Making the weighted integral of the residual over the element volume V zero:

$$\int_V w\left\{\frac{\partial}{\partial x}\left(k_x \frac{\partial T}{\partial x}\right)+\frac{\partial}{\partial y}\left(k_y \frac{\partial T}{\partial y}\right)+\frac{\partial}{\partial z}\left(k_z \frac{\partial T}{\partial z}\right)+\dot{Q}\right\}\mathrm{d}V=0 \tag{19.2.3}$$

By applying the divergence theorem, the weak form of Eq. (19.2.3) is obtained as

$$\int_V\left(k_x \frac{\partial w}{\partial x}\frac{\partial T}{\partial x}+k_y \frac{\partial w}{\partial y}\frac{\partial T}{\partial y}+k_z \frac{\partial w}{\partial z}\frac{\partial T}{\partial z}\right)\mathrm{d}V=\int_V w\dot{Q}\mathrm{d}V+\oint_S\left(k_x \frac{\partial T}{\partial x}n_x+k_y \frac{\partial T}{\partial y}n_y+k_z \frac{\partial T}{\partial z}n_z\right)\mathrm{d}S \tag{19.2.4}$$

where S is the surface surrounding the volume.

From the weak form, it is seen that a linear or higher degree function ensuring C^0 continuity at the interface is good enough. One can choose an eight-noded brick element or a four-noded tetrahedral element. Higher degree approximations are also permissible. Let

$$T=\lfloor N\rfloor\{T^{ne}\} \tag{19.2.5}$$

where $\{T^{ne}\}$ is the nodal temperature vector. In Galerkin FEM, weight is also approximated in the same manner as temperature. Thus

$$w=\lfloor w^{ne}\rfloor\{N\} \tag{19.2.6}$$

Substituting Eqs (19.2.5) and (19.2.6) in Eq. (19.2.4), we get

$$\begin{aligned}&\int_V\lfloor w^{ne}\rfloor\left(k_x\{N_{,x}\}\lfloor N_{,x}\rfloor+k_y\{N_{,y}\}\lfloor N_{,y}\rfloor+k_z\{N_{,z}\}\lfloor N_{,z}\rfloor\right)\mathrm{d}V\{T^{ne}\}=\\&\lfloor w^{ne}\rfloor\left(\int_V\dot{Q}\{N\}\mathrm{d}V+\oint_S\{N\}\left(k_x \frac{\partial T}{\partial x}n_x+k_y \frac{\partial T}{\partial y}n_y+k_z \frac{\partial T}{\partial z}n_z\right)\mathrm{d}S\right)\end{aligned} \tag{19.2.7}$$

As nodal weights are arbitrary, Eq. (19.27) implies that

$$\int_V \left(k_x\{N_{,x}\}\lfloor N_{,x}\rfloor + k_y\{N_{,y}\}\lfloor N_{,y}\rfloor + k_z\{N_{,z}\}\lfloor N_{,z}\rfloor\right)\mathrm{d}V\{T^{ne}\} = \int_V \dot{Q}\{N\}\mathrm{d}V + \oint_S \{N\}\left(k_x\frac{\partial T}{\partial x}n_x + k_y\frac{\partial T}{\partial y}n_y + k_z\frac{\partial T}{\partial z}n_z\right)\mathrm{d}S \tag{19.2.8}$$

The expression given in Eq. (19.2.8) represents FEM elemental equations. The first term on the left-hand side of the expression is called the *elemental coefficient matrix* (analogous to stiffness matrix):

$$[k]^e = \int_V \left(k_x\{N_{,x}\}\lfloor N_{,x}\rfloor + k_y\{N_{,y}\}\lfloor N_{,y}\rfloor + k_z\{N_{,z}\}\lfloor N_{,z}\rfloor\right)\mathrm{d}V$$

The first term on the right-hand side provides the vector due to heat generation (analogous to the load vector):

$$\{f\}^e = \int_V \dot{Q}\{N\}\mathrm{d}V$$

The last term is analogous to internal load vector, whose elements nullify each other at inside the element and are taken care of by boundary conditions at external boundaries.

Usually, the shape functions are expressed in natural coordinates. The following relation is used to obtain the derivative in physical coordinates:

$$\begin{Bmatrix} \dfrac{\partial N_i}{\partial \xi} \\ \dfrac{\partial N_i}{\partial \eta} \\ \dfrac{\partial N_i}{\partial \varsigma} \end{Bmatrix} = \begin{bmatrix} \dfrac{\partial x}{\partial \xi} & \dfrac{\partial y}{\partial \xi} & \dfrac{\partial z}{\partial \xi} \\ \dfrac{\partial x}{\partial \eta} & \dfrac{\partial y}{\partial \eta} & \dfrac{\partial z}{\partial \eta} \\ \dfrac{\partial x}{\partial \varsigma} & \dfrac{\partial y}{\partial \varsigma} & \dfrac{\partial z}{\partial \varsigma} \end{bmatrix} \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \\ \dfrac{\partial N_i}{\partial z} \end{Bmatrix} = [J] \begin{Bmatrix} \dfrac{\partial N_i}{\partial x} \\ \dfrac{\partial N_i}{\partial y} \\ \dfrac{\partial N_i}{\partial z} \end{Bmatrix} \tag{19.2.9}$$

Integrals are transformed using determinant of Jacobian as in two-dimensional formulation. For example, the right-hand-side vector due to heat generation in brick element becomes

$$\{f\}^e = \int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} \dot{Q}\{N\}|J|\mathrm{d}\xi\mathrm{d}\eta\mathrm{d}\varsigma \tag{19.2.10}$$

where the shape functions are expressed in natural coordinates.

The rest of the procedure, namely, assembly, application of boundary conditions, and solution is same as in the case of two-dimensional formulation.

EXERCISES

A. Multiple Choice Questions

19.1 Consider the following differential equation:

$$\frac{d^2u}{dx^2} + x^4 = 0 \quad 0 < x < 1$$

with boundary conditions

$$\text{At } x = 0, \quad u = 0$$

$$\text{At } x = 1, \quad \frac{du}{dx} = 0$$

If this problem is solved using FEM by taking only one element, but evaluating the integrals involved in the formulation exactly, then

(a) FEM solution will be exact everywhere in the domain
(b) FEM solution will be exact at $x = 1$
(c) FEM solution will be inexact everywhere
(d) FEM solution will be less than the exact solution at $x = 1$

19.2 Consider the following differential equation:

$$\frac{d^6u}{dx^6} + x^2 = 0, \quad 0 < x < 1$$

The essential boundary conditions for this differential equation will be in the form of prescribed values of

(a) u, du/dx, d^2u/dx^2
(b) u, du/dx, d^2u/dx^2, d^3u/dx^3, d^4u/dx^4, d^5u/dx^5
(c) u, du/dx, d^4u/dx^4
(d) d^2u/dx^2, d^3u/dx^3, d^4u/dx^4, d^5u/dx^5

19.3 Which of the following mesh refinement method does not require a new mesh to be constructed?

(a) h-refinement
(b) p-refinement
(c) hp-refinement
(d) r-refinement

19.4 In solving the pure bending problem with plane stress FEM, a rectangular element generates a spurious shear that is called

(a) locking shear
(b) false shear
(c) parasitic shear
(d) external shear

19.5 Two end nodes of a one-dimensional bar element have the coordinates as $x = 0$ and $x = 1$. If a third node is placed at location $x = 3/4$, the element will provide infinite stress at

(a) $x = 0$
(b) $x = 1/4$
(c) $x = 3/4$
(d) $x = 1$

B. Problems

19.6 Mathematically show that a quarter-point element can be used for solving the problems where the derivative of the primary variable becomes infinite at a singular point.

19.7 In solving an axial rod problem using FEM with three-noded elements, the middle node need not be exactly at the middle. Find out the zone in which the middle node can be placed.

19.8 Take a three-noded, one-dimensional element 1–2–3, in which the third node is at infinity. Show that the geometry can be approximated as

$$x = N_1 x_1 + N_2 x_2$$

where

$$N_1 = -\frac{2\xi}{1-\xi}; \quad N_2 = \frac{1+\xi}{1-\xi}$$

Approximate the primary variable u by a three-noded Lagrangian shape function and obtain the expression for d*u*/d*x*.

19.9 Show that in a CST element, if one angle is very small, the resulting stiffness matrix will be ill-conditioned.

REFERENCES

Barlow, G, 1976, 'Optimal stress locations in finite element models', *Int. J. Numer. Meth. Engg*, Vol. 10, pp. 243–251.

Cook, RD, Malkus, DS, and Plesha, ME, 1989, *Concepts and Application of Finite Element Analysis, Third edition*, John Wiley & Sons, Singapore.

Tong, P, 1969, 'Exact solution of certain problems by the finite element method', *J. AIAA*, Vol. 7, pp. 179–180.

Zienkiewicz, OC, Emson, C, and Bettess, P, 1983, 'A novel boundary infinite element', *Int. J. Numer. Meth. Engg*, Vol. 19, pp. 393–404.

Zienkiewicz, OC, Taylor, RL, and Zhu, JZ, 2005, *The Finite Element Method: Its Basis & Fundamentals, Sixth edition*, Butterworth–Heinemann, Oxford.

A

APPENDIX

Appendix A

A.1 Divergence Theorem

A.2 The Rule of Differentiation with Respect to a Vector

A.3 Solving the System of Equations $[A]\{x\} = \{b\}$

A.4 Pseudocode for Solving a Two-Dimensional Heat-Conduction Problem

A.1 DIVERGENCE THEOREM

If V be the volume of a three-dimensional region bounded by a closed regular surface S, then for a vector field u defined in V and S:

$$\int_V (\operatorname{div} \boldsymbol{u}) \mathrm{d}V = \int_V \nabla \bullet \boldsymbol{u} \, \mathrm{d}V = \int_S (\boldsymbol{u} \bullet \hat{\boldsymbol{n}}) \mathrm{d}S \tag{A.1.1}$$

where $\hat{\boldsymbol{n}}$ is the unit outward normal to S.

Divergence theorem is a special case of Stokes's theorem. In Eq. (A.1.1), if V is used to represent a two-dimensional region, then the divergence theorem remains valid with S representing the boundary of the region.

It can be easily shown [Chandrasekharaiah and Debnath (1994)] that a consequence of divergence theorem is

$$\int_V \nabla \phi \mathrm{d}V = \int_S \phi \hat{\boldsymbol{n}} \, \mathrm{d}S \tag{A.1.2}$$

where ϕ is a scalar field defined in V and on S. For the three dimensional case, in the component form, Eq. (A.1.2) implies

$$\int_V \frac{\partial \phi}{\partial x} n_x \, \mathrm{d}V = \int_S \phi n_x \mathrm{d}S; \quad \int_V \frac{\partial \phi}{\partial y} n_y \, \mathrm{d}V = \int_S \phi n_y \mathrm{d}S \text{ and } \int_V \frac{\partial \phi}{\partial z} n_z \, \mathrm{d}V = \int_S \phi n_z \mathrm{d}S \tag{A.1.3}$$

where n_x, n_y and n_z are the direction cosines of the normal.

For the two dimensional case, in the component form:

$$\int_S \frac{\partial \phi}{\partial x} n_x \, \mathrm{d}S = \int_\Gamma \phi n_x \mathrm{d}\Gamma \quad \text{and} \quad \int_V \frac{\partial \phi}{\partial y} n_y \, \mathrm{d}S = \int_S \phi n_y \mathrm{d}\Gamma \tag{A.1.4}$$

where Γ is the boundary of the area S.

A.2 THE RULE OF DIFFERENTIATION WITH RESPECT TO A VECTOR

Suppose there is a function F of n independent variables $x_1, x_2, \ldots, x_n$. The differentiation of this function with respect to each of the variables can be represented by a column vector as follows:

$$\frac{\partial F}{\partial \{x\}} = \begin{Bmatrix} \dfrac{\partial F}{\partial x_1} \\ \dfrac{\partial F}{\partial x_2} \\ \vdots \\ \vdots \\ \vdots \\ \dfrac{\partial F}{\partial x_n} \end{Bmatrix} \tag{A.2.1}$$

Now, if F is a linear function given by

$$F = l_1x_1 + l_2x_2 + \cdots + l_nx_n = \lfloor x_1\ x_2 \ldots x_n \rfloor \begin{Bmatrix} l_1 \\ l_2 \\ \vdots \\ \vdots \\ l_n \end{Bmatrix} = \lfloor x \rfloor \{l\} \tag{A.2.2}$$

then it can be easily verified that

$$\frac{\partial F}{\partial \{x\}} = \{l\} \tag{A.2.3}$$

[If there was a constant term in Eq. (A.2.2), it would not have made any difference to derivatives.]

If S is given by

$$F = \lfloor x \rfloor [A] \{x\} \tag{A.2.4}$$

where $[A]$ is a symmetric coefficient matrix, then it can be verified that

$$\frac{\partial F}{\partial \{x\}} = 2[A]\{x\} \tag{A.2.5}$$

What happens if $[A]$ is non-symmetric? In that case, we rewrite Eq. (A.2.4) by decomposing $[A]$ into a symmetric and a skew-symmetric part:

$$F = \lfloor x \rfloor \left\{ \frac{1}{2}\left([A] + [A]^{\mathrm{T}}\right) + \frac{1}{2}\left([A] - [A]^{\mathrm{T}}\right) \right\} \{x\} \tag{A.2.6}$$

As $\lfloor x \rfloor [A] \{x\}$ is a scalar expression, its transpose will be same. Thus

$$\lfloor x \rfloor [A] \{x\} = \lfloor x \rfloor [A]^{\mathrm{T}} \{x\} \tag{A.2.7}$$

Using this relation given in Eq. (A.2.7), F can be written as

$$F = \lfloor x \rfloor \left\{ \frac{1}{2}\left([A] + [A]^{\mathrm{T}}\right) \right\} \{x\} \tag{A.2.8}$$

With this, the coefficient matrix of the expression becomes symmetric; using Eq. (A.2.5), we get

$$\frac{\partial F}{\partial \{x\}} = \left([A] + [A]^{\mathrm{T}}\right) \{x\} \tag{A.2.9}$$

A.3 SOLVING THE SYSTEM OF EQUATIONS $[A]\{x\} = \{b\}$

The following function of C++ solves the system of equations $[A]\{x\}=\{b\}$. Input to the function is the coefficient matrix $[A]$, designated as 'a' in the program, right-hand-side vector $\{b\}$, and number of the equations. The solution is stored in $\{b\}$, that is, the output of the program in the form of $\{b\}$ represents $\{x\}$.

```
//*********************************************
void gauss(double **a,double *b,int eqn)
{int k,row,r,s;
 int i,j;
 double sum ,c,*temp,btemp, large=-1e-15;
 temp = new double[eqn];
 for (k=0;k<eqn-1;k++)

 for (i=k+1;i<eqn;i++)
        {
        //pivoting begins
        if (a[k][k]==0)
        {for(r=k+1;r<eqn;r++)
{if((a[r][k]>large)&&(a[r][k]!=0)) {large=a[r][k];row=r;}}

        btemp=b[k];b[k]=b[row];b[row] =btemp;
        for(s=0;s<eqn;s++) {temp[s] = a[k][s];
                                      a[k][s]= a[row][s];
                                      a[row][s]=temp[s];
                                      }
        }
        //end of pivoting
        large=-1e-15;

        c= a[i][k]/a[k][k];
        for(j=k+1;j<eqn;j++)    a[i][j]=a[i][j]- c*a[k][j] ;
        b[i]=b[i] - c * b[k] ;
        }
 }

b[eqn-1]=b[eqn-1]/a[eqn-1][eqn-1];
for (i=eqn-2;i>-1;i--)
{k = i+1;
   sum =0 ;
   for(j=k;j<eqn;j++)
         sum+=a[i][j]*b[j];
 b[i] = (b[i]- sum)/a[i][i];
 }
 }
//*****************************************************************
```

A.4 PSEUDOCODE FOR SOLVING A TWO-DIMENSIONAL HEAT CONDUCTION PROBLEM

The following pseudocode helps in solving a steady-state, two-dimensional heat-conduction problem:

```
Main()
{
input(x,y,cv,nnode,ntelm)  /*calling function providing nodal
coordinates and connectivity matrix. Here,  'nnode' is the total
number of nodes, 'ntelm' is the total number of elements, 'x', 'y'
denotes the coordinates  and 'cv' denotes the connectivity matrix.
*/
glomatrix (gk,ntelm,nnode,x,y,icv) /*calling global stiffness
matrix 'gk' */
bcond(nnode,gk,rhs) /*applying boundary conditions, 'rhs' being
the right hand side vector*/
solution(gk,tem,rhs,nnode)  /* solving the system of linear
equations, 'tem' being the temperature */
for (i=1; i<=nnode; i=i+1) {
     printf (x[i],y[i],tem[i]))
}

/* Function input starts here. It prepares a mesh of four-noded
rectangular elements in a rectangular plate.*/
input(x,y,icv,nnode,ntelm) {
f1=fopen("conn", "w");      /*open the file "conn" for writing the
connectivity matrix */
f2=fopen("coor", "w"); /*open the file "coor" for writing the
nodal coordinates */
f3=fopen("input", "r");  /*open the file "input" for reading the
length and width of the plate and the number of divisions in length
and width direction */
fscanf(f3,l,w);   /*reading length and width from "input" file*/
fscanf(f3,nx,ny); /*reading the number of division along length
and width direction from "input" file*/
h1=x1/nx;
h2=y1/ny;
for(i=1; i<=nx+1; i=i+1)
{
for(j=1; j<=ny+1; j=j+1)
{
```

```
    k=(i-1)*(ny+1) + j;
    x[k]=h1*(i-1);
    y[k]=h2*(j-1);
    }
    }
nnode= (nx+1)*(ny+1)
for (i=1; i<=nnode; i=i+1)
{
fprintf(f2, x[i], y[i]); /*write the nodal coordinates in file
"coor"*/
}
fclose(f2); /*closing the file*/

/* Now we generate CONECTIVITY MATRIX "cv" *. Note that we generate
a structured mesh./

for (i=1; i<=nx; i=i+1)
{
for (j=1; j<=ny; j=j+1)
{
n=(i-1)*ny+j;
cv[n, 1]=j+(i-1)*(ny+1);
cv[n, 2]=cv[n, 1]+(ny+1);
cv[n, 3]=cv[n, 2]+1;
cv(n, 4)=cv[n, 1]+1;
}
}
ntelm=nx*ny;
for (i=1; i<=ntelm; i=i+1)
{
fprintf(f1, cv[i, 1], cv[i, 2], cv[i, 3], cv[i, 4]); /*write
global node numbers in connectivity matrix file "conn"*/
}
fclose(f1); /*closing the file*/
} /*Function "input" ends here. */

/*Function for generating global stiffness matrix "gk" here. Input
to this function is arguments "ntelm", "nnode", "x", "y" and "cv".
*/
 glomatrix (gk,ntelm,nnode,x,y,icv){
```

```
/* First  we initialize the global matrix. */
for (i=1; i<=nnode; i=i+1)
{
for (j=1; j<=nnode; j=j+1)
{
gk[i, j]=0.0;
}
}

/* Now, we start calling the elemental matrices one by one and put
their entries in the appropriate position in global stiffness
matrix. */
for (i=1; i<=ntelm; i=i+1)
{
 elematrix (i,ak,x,y,icv,nnode,ntelm); /* Here "ak" denotes 4-by-
 4 stiffness matrix. By calling this matrix, you have to get
 elemental stiffness matrix. */
for (i1=1; i1<=4; i1=i1+1)
{
for (j1=1; j1<=4; j1=j1+1)
{
p=cv(i,i1); /*global node corresponding to i1-th node of element
"i" */
q=cv(i,j1);  /*global node corresponding to j1-th node of element
"i" */
gk(ip,iq)=gk(ip,iq)+ak(i1,j1);
}
}
}
} /*Function "glomatrix" ends here. */

/*Function "elematrix" starts here. The purpose of the routine is
to provide the stiffness matrix of an element. */
elematrix (ielm,ak,x,y,icv,nnode,ntelm)
{
 m=2;  /*We decide to use two Gauss point formula in both x and y
 direction. */
/* First, we initialize the local stiffness matrix. */
for (i=1; i<=4; i=i+1)
{
```

```
for (j=1; j<=4; j=j+1)
{
ak(i,j)=0.0;
}
}
gauss2(); /*We extract the data required for Gauss quadrature from
gauss2 subroutine. The i-th Gauss-point of m-point formula is
denoted by "xg[m, i]" and corresponding weight "w[m, i]. These can
be made global variable. */
for (i=1; i<=m; i=i+1)
{
xi=xg(m,i);
for(i1=1; i1<=4; i1=i1+1)
{
for(j1=1; j1<=4; j1=j1+1)
{
temp(i1,j1)=0.0;
}
}
for (j=1; j<=m; j=j+1)
{
eta=xg(m, j);
   shap(xi, eta); /*Information about shape function and its
   derivative is obtained by calling this routine.*/
   bmat(xi,eta,ielm,x,y,icv,b,bt,detj,nnode,ntelm) /* This
   routine gives B matrix and its transpose. It also gives the
   determinant of Jacobian.*/
   mult(bt,4,2,b,4,prod); /* BTB is calculated in this routine. */
for (i1=1; i1<=4; i1=i1+1)
{
for (j1=1; j1<=4; j1=j1+1)
{
prod(i1,j1)=detj*prod(i1,j1)*w(m,j);
}
}
for (i1=1; i1<=4; i1=i1+1)
{
for (j1=1; j1<=4; j1=j1+1)
{
```

```
temp(i1,j1)=temp(i1,j1)+prod(i1,j1);
}
}
}
for (i1=1; i1<=4; i1=i1+1)
{
for (j1=1; j1<=4; j1=j1+1)
{
ak(i1,j1)=ak(i1,j1)+temp(i1,j1)*w(m,i);
}
}
}
} /* Function "elematrix" ends here. */

/* This function contains the data for numerical integration. The
i-th Gauss-point of m-point formula is denoted by x[m,i] and
corresponding weight by w[m,i]. */
 gauss2()
{
x[1, 1]=0.0;
w[1, 1]=2.0;
x[2, 1)=-.577350269189626
x[2,2]=-x(2,1)
w[2,1]=1.0
w[2,2]=1.0
x[3,1]=-.77459666924148
x[3,2]=0.0
x[3,3]=.774596669241483
w[3,2]=8.0/9.0
w[3,1]=5.0/9.0
w[3,3]=w(3,1)
x[4,2]=-.339981043584856
x[4,1]=-.861136311594053
x[4,3]=-x(4,2)
x[4,4]=-x(4,1)
w[4,2]=.652145154862546
w[4,3]=w(4,2)
w[4,1]=.347854845137454
w[4,4]=w(4,1)
x[5,1]=0.00000000000000
```

```
x[5,2]=.538469310105683
x[5,3]=-x(5,2)
x[5,4]=.906179845938684
x[5,5]=-x(5,4)
w[5,1]=.568888888888889
w[5,2]=.478628670499365
w[5,3]=w[5,2]
w[5,4]=.236926885056189
w[5,5]=w[5,4]
} /*Function Gauss2 ends here. */

/* Function for calculating shape functions and its derivatives
starts here. */
shap(xi,eta)
{
/* calculation for 4 noded shape functions and derivatives */
/* enp = shape function for four noded */
/* dngta = differentiation of shape function with respect to xi */
/* dneta = differentiation of shape fns with respect to eta */
gtai[1]=-1.0;
gtai[2]=1.0;
gtai[3]=1.0;
gtai[4]=-1.0;
etai[1]=-1.0;
etai[2]=-1.0;
etai[4]=1.0;
etai[3]=1.0;
for (i=1; i<=4; i=i+1)
{
enp[1, i]=(1.+etai[i]*eta)*(1.+gtai[i]*xi)/4.0;
dngta[1, i]=(1.+etai[i]*eta)*gtai[i]/4.0;
dneta[1,i]=(1.+gtai[i]*xi)*etai[i]/4.0;
}
} /*Function "shap" ends here.

/* Function for calculation B-matrix (denoted by 'b') and its
transpose (denoted by 'bt') starts here. It essentially contains
the derivatives of shape functions. */
bmat(xi,eta,ielm,x,y,icv,b,bt,detj,nnode,ntelm)
{
```

```
for (i=1; i<=4; i=i+1)
{
for (j=1; j<=2; j=j+1)
{
bt[i, j]=0.0;
b[j, i]=0.0;
}
}

shap(xi, eta);
for(i=1; i<=4; i=i+1)
{
bi[1, i]=dngta[1, i]; /*dngta is the derivative of shape function
with respect to 'xi' and comes from function shape */
bi[2, i]=dneta[1, i];  /*dneta is the derivative of shape function
with respect to 'eta' and comes from function shape */
}
 jacbi(gta,eta,ielm,x,y,icv,detj,nnode,ntelm); /* This function
 is called for getting the Jacobian matrix and its determinant. */
 mult(aijac,2,2,bi,4,b); /* In this function by multiplication of
 Jacobian matrix with bi-matrix, B-matrix is obtained */
for (i=1; i<=4; i=i+1)
{
for (i=1; i<=4; i=i+1)
{
bt[i, j]=b[j, i];
}
}
}/*Function "bmat" end.

/*Function for calculation of Jacobian and its inverse starts here
*/
/* gacobi=Jacobian */
/* aijac =inverse of Jacobian */
/* detj=determinant of Jacobian */
    jacbi(xi,eta,ielm,x,y,icv,detj,nnode,ntelm)
{
for(i=1; i<=2; i=i+1)
{
for(j=1; j<=2; j=j+1)
```

```
{
    gacobi[i, j]=0.0;
}
}
 shap(xi,eta);
for(i=1; i<=4; i=i+1)
{
gacobi[1, 1]=gacobi[1, 1]+dngta[1, i]*x[cv[ielm, i]];
gacobi[1, 2]=gacobi[1, 2]+dngta[1, i]*y[cv[ielm, i]];
gacobi[2, 1]=gacobi[2, 1]+dneta[1, i]*x[cv[ielm, i]];
gacobi[2, 2]=gacobi[2, 2]+dneta[1, i]*y[cv[i,ielm]];
}
detj=gacobi[1, 1]*gacobi[2,2]-gacobi[2,1]*gacobi[1,2];
aijac[1, 1]=gacobi[2, 2]/detj;
aijac[1, 2]=-gacobi[1, 2]/detj;
aijac[2, 1]=-gacobi[2, 1]/detj;
aijac[2, 2]=gacobi[1, 1]/detj;
}/* Function "Jacbi" ends here.*/

/*Function "mult" starts here. It multiplies two matrices and
return the product. */
mult(an,m,l,ant,n,prod)
{
/* Intialization
for (i=1; i<=m; i=i+1)
{
for (j=1; j<=m; j=j+1)
{
prod[i, j]=0.0
for (k=1; k<=l; k=k+1)
{
prod(i,j)=prod(i,j)+an(i,k)*ant(k,j);
}
}
}
}

/* Function for applying essential boundary conditions, starts
here. */
bcond(nnode,gk,rhs)
```

```
{
/* no_ess=number of essential boundary conditions */
fscanf(f3, no_ess);
for (i=1; i<= no_ess; i=i+1)
{
fscanf(f3, node, temp); /*At global number "node", temperature
"temp" has been prescribed. */
}
for ( j=1; j<=nnode; j=j+1)
{
gk(node, j)=0.0;
}
gk(node, node)=1.0;
rhs(node)=temp;
fclose(f3); /*closing the file*/
}

/* Solution by Gauss elimination method is carried out in the
following function.*/
/* SOLUTION BY GAUSS-ELIMINATION METHOD */
 solution(aa,x,gforce,ia)
{
   n=ia;
   m=n+1;
   l=n-1;
for(i=1; i<=ia; i=i+1)
{
for(j=1; j<=ia; j=j+1)
{
gsf[i, j]=aa[i, j];
}
gsf[i, ia+1]=gforce[i]
}
for(k=1; k<=l; k=k+1)
{
   jj=k;
   big=abs(gsf[k,k]);
   kp1=k+1;
/*SEARCH FOR LARGEST POSSIBLE PIVOT ELEMENT IN THE COLUMN*/
for(i=kp1; i<=n; i=i+1)
```

```
{
ab=abs(gsf[i, k]);
If(big<ab)
big=ab;
else
continue;
/* DECISION OF NECESSITY OF ROW INTERCHANGE  */
/* The symbol "!=" means not equal to */
       if(jj !=k)

/*   ROW INTERCHANGE
for(j=k; j<=m; j=j+1)
{
temp=gsf(jj,j);
gsf(jj,j)=gsf(k,j);
gsf(k,j)=temp;
}
else
continue;

/*   CALCULATION OF ELEMENTS OF NEW MATRIX */
for(i=kp1; i<=n; i=i+1)
{
quot=gsf(i,k)/gsf(k,k);
for(j=kp1; j<=m; j=j+1)
{
y=gsf(i,j)-quot*gsf(k,j);
gsf(i,j)=y;
}
}
for(i=1; i<=n; i=i+1)
{
gsf(i,k)=0.0;
}
/* FIRST STEP IN BACK SUBSTITUTION */
x(n)=gsf(n,m)/gsf(n,n);

/*   REMAINDER OF BACK-SUBSTITUTION PROCESS */
for(nm=1; nm<=l; nm=nm+1)
{
```

```
sum=0.0;
i=n-nm;
ip1=i+1;
for(j=ip1; j<=n; j=j+1)
{
sum=sum+gsf[i,j]*x(j);
}

x[i]=(gsf[i,m]-]sum)/gsf[i,i];
}
}/*Function ends here. */
```

REFERENCE

Chandrasekharaiah, DS, and Lokenath Debanath, 1994, *Continuum Mechanics*, Academic Press, Boston.

ANSWERS TO MULTIPLE CHOICE QUESTIONS

Chapter 1

1.1 (c)

Explanation: One can observe that matrix (iii) is the assembly of the following elemental matrices:

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix}$$

In the rod subjected to axial rod problem, the assembled global stiffness matrix should be symmetric with non-zero diagonal elements and the elements of each row (and column) added together should give 0. Obviously, matrix (i) and (ii) do not represent global stiffness matrix.

1.2 (d)

Explanation: In the matrix

$$\begin{bmatrix} 0 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 0 \\ 10 \end{Bmatrix}$$

the first global equation has been replaced by the equation corresponding to the boundary condition.

1.3 (d)

Explanation: In the matrix

$$\begin{bmatrix} 1 & 0 & 1 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 10 \end{Bmatrix}$$

the first global equation has been replaced by the equation corresponding to the boundary condition.

1.4 (a)

Explanation: The cubic interpolation provides $v = x^3$. Thus, at $x = 0.5$, $v = 1/8$.

1.5 (b)

Explanation: The vertical rigid body displacement, say $v_1 = v_2 = 1$ and $\theta_1 = \theta_2 = 0$, should provide zero forces and moments. Therefore, for all rows, the sum of first and third element will be 0.

Chapter 2

2.1 (d)

Explanation: The functional can be written as

$$\Pi = \int_0^{10} \left\{ \frac{1}{2}\left(\frac{\mathrm{d}u(x)}{\mathrm{d}x}\right)^2 - 5u\delta(x-5) \right\} \mathrm{d}x$$

where $\delta(x-5)$ is Dirac-delta function whose value is 0 everywhere except at $x = 5$, where it becomes infinite. One property of the Dirac-delta function, $\delta(x-\bar{x})$, is that

$$\int_a^b f(x)\delta(x-\bar{x})\,\mathrm{d}x = f(\bar{x}) \quad \text{for } a < x < b$$

Using this property, we have been able to change the form of the function. Now, from Euler–Lagrange equation given in Eq. (2.21), we obtain the differential equation.

2.2 (b)

Explanation: We can obtain the differential equation from Euler–Lagrange equation given in Eq. (2.21).

2.3 (a)

Explanation: The total potential energy is in this case is the sum of strain energy of the beam, energy stored in the spring and work-potential of the load.

2.4 (a)

Explanation: The strain energy of the spring is

$$\int_0^x (4+1000x)x\,\mathrm{d}x = 2x^2 + \frac{10^3 x^3}{3}$$

2.5 (b)

Explanation: Option (a) is wrong because the actual work done is $kxdx$; option (b) is correct as the virtual displacement does not change the forces; option (c) is wrong because the virtual work may be negative; and option (d) is wrong as the applied forces may do some work. In equilibrium, the virtual work done by all the forces together is 0.

Chapter 3

3.1 (c)

Explanation: Ritz method works on the variational (weak) form, which takes into account the natural boundary conditions. Thus, in this method only the essential boundary conditions need to be satisfied. The other three methods work on the strong form of the differential equation, requiring the satisfaction of essential as well as natural boundary conditions.

3.2 (d)

Explanation: Option (a) provides incomplete approximation as the term $x^3(x-1)$ is missing. Options (b) and (c) do not satisfy the essential boundary conditions.

3.3 (a)

Explanation: Both the boundary conditions are satisfied in this method.

3.4 (a)

Explanation: Although it is not a complete approximation, it will provide the exact solution as the final solution is of this form. In general, one should take only the complete approximation.

3.5 (b)

$$\int_D \delta(x-\bar{x})R\mathrm{d}x = R(\bar{x})$$

Thus, making the weighted residual 0 amounts to making the residual zero at point $R(\bar{x})$, which is done in collocation method.

Chapter 4

4.1 (a)

Explanation: At the vertex, the slope is not continuous; however, the function is continuous. Hence, it is having C^0 continuity.

4.2 (d)

Explanation: The curvature contains the second order derivatives. Hence, if curvature is continuous, the function is C^2 continuous.

4.3 (c)

Explanation: The variational form of the differential equation will contain the third order derivative at the boundaries. Hence, a polynomial of third degree should be taken as the interpolation function.

4.4 (a)

Explanation: Use Eq. (4.55) to solve this problem.

4.5 (c)

Explanation: We follow the Galerkin procedure. Integration by parts is not needed. We shall get unsymmetrical stiffness matrix.

Chapter 5

5.1 (a)

Explanation: Use Eq. (5.15) to obtain the right hand side vector representing heat generation.

5.2 (b)

Explanation: The right hand side vector is $\int_0^h q\lfloor N_1\, N_2\, N_3 \rfloor^{\mathrm{T}}\,\mathrm{d}x$. The N_1, N_2 and N_3 are Lagrangian shape function and may be calculated using Eq. (4.17).

5.3 (a)

Explanation: Due to symmetry, both shear stress and shear strain will be equal to zero.

5.4 (b)

Explanation: Make an analogy with heat transfer problem.

5.5 (c)

Hint: For this element $u = N_3 u_3 = \frac{2}{h^2}\left(x^2 - \frac{h}{2}x\right)$. Differentiate it with respect to x to make it equal to zero.

Chapter 6

6.1 (a)

Explanation: The load vector is

$$\int_0^h q\lfloor N_1\ N_2\ N_3\ N_4 \rfloor^{\mathrm{T}}\,\mathrm{d}x = \int_0^h P\delta\left(x - \frac{h}{4}\right)\lfloor N_1\ N_2\ N_3\ N_4 \rfloor^{\mathrm{T}}\,\mathrm{d}x = \lfloor N_1\ N_2\ N_3\ N_4 \rfloor^{\mathrm{T}}_{x=h/4}$$

6.2 (a)

Hint: Refer to Solved Problem 6.1.

6.3 (d)

Hint: Refer to the expressions for shape functions.

6.4 (c)

Hint: Use $y = N_1 y_1 + N_2\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)_1 + N_3 y_2 + N_2\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)_2$ and differentiate thrice.

6.5 (b)

Chapter 7

7.1 (d)

Explanation: Use $u_1' = u_1 \cos\theta + v_1 \sin\theta$; $v_1' = -u_1 \sin\theta + v_1 \cos\theta$. Put $u_1 = 3$, $v_1 = 4$ and $\theta = 30°$. Thus, $u_1' = \frac{3\sqrt{3}}{2} + 2$; $v_1' = -\frac{3}{2} + 2\sqrt{3}$.

7.2 (b)

Explanation: It is an orthogonal matrix and therefore its inverse is equal to its transpose.

7.3 (d)

Explanation: If we impose non-zero rigid body displacements on the truss, no forces will be produced and the product of stiffness matrix and displacement will be 0. Therefore, although the matrix is positive definite, it is not strictly positive definite.

7.4 (a)

Hint: Follow the assembly procedure.

7.5 (c)

Explanation: The angle between the lines will not change due to rotation of the coordinate system.

Chapter 8

8.1 (b)

Explanation: Heat flux will be zero. There may be a variation along the x-axis.

8.2 (b)

Hint: Use Eq. (8.42) to solve this problem.

8.3 (a)

Hint: Use Eq. (8.8)

8.4 (d)

Hint: Use Eqs (8.39) and (8.40)

8.5 (a)

Hint: Use Eqs (8.39) and (8.40) and differentiate.

Chapter 9

9.1 (a)

9.2 (b)

9.3 (d)

Explanation: Using $2n - 1 = 10$, we get $n = 5.5$. Hence, six Gauss points would be appropriate.

9.4. (a)

Hint: Refer to Table 9.2.

9.5. (a)

Explanation: There is a requirement of two Gauss points along x and one Gauss point along y. Hence, a total of two Gauss points are required.

Chapter 10

10.1 (c)

Explanation: In super-parametric formulation, the geometry is approximated by higher order interpolation function.

10.2 (b)

Explanation: In iso-parametric formulation, the average nodal temperature will be at average nodal coordinates.

10.3 (c)

Hint: Refer Section 10.4 and use the equations provided for the shape functions.

10.4 (a)

Explanation: For a rectangular element, the Jacobian is defined as the ratio between the areas of the domains in physical and natural coordinate systems.

10.5 (d)

Explanation: For a cubic element, the Jacobian is defined as the ratio between the volumes of the domains in physical and natural coordinate system.

Chapter 11

11.1 (b)

Hint: Use Equation (11.36).

11.2 (d)

Hint: Use Equation (11.36).

11.3 (d)

Hint: Use Equation (11.30).

11.4 (a)

Explanation: Constant strain triangle element provides constant strain throughout the element.

11.5 (c)

Hint: Use Equation (11.37).

Chapter 12

12.1 (a)

Explanation: The consistent mass matrix will provide positive kinetic energy for non-zero velocity vector.

12.2 (c)

Explanation: Differentiate Eq. (12.93) and make the differentiation equal to zero.

12.3 (d)

Explanation: The units of $\alpha[K]$ and $\beta[M]$ are equal as these are added. Hence, the unit of α/β are is equal to that of mass divided by stiffness. Thus, it is equal to unit of square of time period.

12.4 (b)

Explanation: Use the formula

$$[m^e] = \int_0^h M\{N\}\lfloor N \rfloor \delta\left(x - \frac{h}{2}\right) \mathrm{d}x$$

where $\delta\left(x - \frac{h}{2}\right)$ is the Dirac-delta function, which is 0 everywhere except at $x = h/2$.

12.5 (b)

Explanation: The stiffer the beam, the greater is the natural frequency. Thus, option (b) denotes the correct order.

Chapter 13

13.1 (a)

Explanation: Only in option (a) can the primary variables at $(n + 1)$th time step be calculated from the information at nth time step.

13.2 (b)

Hint: It has been mentioned in the text. However, you may derive it from Taylor series expansion.

13.3 (c)

Hint: It has been mentioned in the text and can be easily derived.

13.4 (b)