

Using Finite Elements in Mechanical Design

Dr. J. Toby Mottram and
Dr. Christopher T. Shaw

McGRAW-HILL BOOK COMPANY

London · New York · St Louis · San Francisco · Auckland · Bogotá · Caracas
Lisbon · Madrid · Mexico · Milan · Montreal · New Delhi · Panama · Paris
San Juan · São Paulo · Singapore · Sydney · Tokyo · Toronto

Published by
McGRAW-HILL Book Company Europe
Shoppenhangers Road, Maidenhead, Berkshire, SL6 2QL, England
Telephone 01628 23432
Facsimile 01628 770224

British Library Cataloguing in Publication Data

Mottram, J. Toby

Using finite elements in mechanical design

1. Finite element method 2. Mechanical engineering – Data processing
3. Design, Industrial 4. Structural design – Data processing 5. Computer-aided
design 6. Mechanics, Analytic

I. Title II. Shaw, Christopher T.

620'.0042'0285

ISBN 0-07-709093-4

Library of Congress Cataloging-in-publication Data

Mottram, J. Toby (James Toby),

Using finite elements in mechanical design / J. Toby Mottram and
Christopher T. Shaw.

p. cm.

Includes index.

ISBN 0-07-709093-4

1. Finite element method–Data processing. 2. Engineering design–
Data processing. 3. Machine design–Data processing. 4. Computer
-aided design.

I. Shaw, Christopher T. (Christopher Thomas). II. Title.

TA347.F5M68 1996

620.1'05'0285–dc20

95-46042 CIP

McGraw-Hill

A Division of The McGraw-Hill Companies



Copyright © 1996 McGraw-Hill International (UK) Ltd. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of McGraw-Hill International (UK) Limited.

1234 CUP 99876

Typeset by Keyword Typesetting Services Ltd, Wallington, Surrey
Printed and bound in Great Britain at the University Press, Cambridge

Printed on permanent paper in compliance with ISO Standard 9706

This book is dedicated:

by JTM to his wife, Liz, and sons, Thomas and Jamie

by CTS to the memory of his father, Albert

CONTENTS

Preface	xi
Notation	xiv
1 Relationship Between Design and Finite Elements	1
1.1 The Design Process	2
1.2 Evaluating Designs	4
1.3 Using Computers for Design Evaluation	4
1.4 Typical Design Situations	5
1.5 Historical Review of Obtaining Structural Design Solutions by Computer	6
1.6 Overview of the Modern Finite Element Method	13
1.7 Industrial Implementation and Research in Finite Elements	14
1.8 References and Further Reading	17
2 Structural Mechanics	18
2.1 The Use of Load-Bearing Structures	18
2.2 Engineering Stress and Strain	20
2.3 Some Simple Situations	37
2.4 Describing Problems in Structural Mechanics	42
2.5 References and Further Reading	46
3 Finite Element Solutions of the Equations	47
3.1 Developing the Finite Element Methodology	47
3.2 Satisfying Equilibrium and Compatibility	53
3.3 Use of Elements with Different Spatial Dimensions	54
3.4 Overview of Methods for Calculating Element Equations	56
3.5 The Role of Displacement Distributions	58
3.6 The Direct Approach to Element Formulation	59
3.7 Isoparametric Elements	64
3.8 Other Types of Elements	78
3.9 Producing a Solution from the Element Equations	82
3.10 References and Further Reading	90
4 Implementing a Computer-based Analysis Procedure	91
4.1 A Process for Analysing a Structure	91
	vii

4.2	Generic Software Packages for Stress Analysis	94
4.3	Available Computer Hardware Systems	97
4.4	Matching Hardware Types to Analyses	100
4.5	Acquiring the Technologies	101
5	Building the Geometry Description and the Mesh	111
5.1	Think First—Compute Later	111
5.2	Overview of Computer Models of Geometry	112
5.3	Wireframe Models	113
5.4	Surface Models	114
5.5	Solid Models	115
5.6	Choice of Geometry Description	117
5.7	Overview of Mesh Generation	118
5.8	Advice on Choosing Element Types	119
5.9	Determining a Strategy for Distributing the Elements	123
5.10	Types of Mesh Structure	125
5.11	Building Meshes	128
5.12	Adapting a Mesh to Improve a Solution	134
5.13	Checking a Mesh	137
6	Obtaining and Analysing the Results	140
6.1	Specifying Material Properties	141
6.2	Finding the Boundaries of the Mesh	142
6.3	Imposing the Boundary Conditions on the Mesh	146
6.4	Controlling the Solution	149
6.5	Running the Solver	150
6.6	Troubleshooting	150
6.7	The Results Generated by the Solver	151
6.8	Using Computer Graphics to Evaluate Solutions	152
6.9	Checking a Solution for Accuracy	159
6.10	Making Refinements	159
6.11	Failure Criteria	160
7	Some Example Problems	161
7.1	The Brazilian or Split Tensile Test	161
7.2	A Pressure Vessel	174
8	More Advanced Problems	192
8.1	Composite Material Open-sectioned Beam	192
8.2	Design Example with Submodels	209
9	Industrial Case Studies	231
9.1	Analysing a Bicycle Frame	231
9.2	Design of a Connecting Rod	239

9.3	A Composite Suspension Arm	246
9.4	Displacement Analysis of a Vehicle Body Shell	249
10	Solving More Complex Problems	252
10.1	Transient Problems	253
10.2	Vibration Problems	254
10.3	Optimization	255
10.4	Nonlinear Problems	257
10.5	Calculating Thermal Effects within Structures	260
10.6	References and Further Reading	261
	References and Further Reading	262
	Index	266

PREFACE

MOTIVATION

This book is aimed at those who are new to using finite element software for mechanical or structural design. There are many textbooks that explain the theory behind the use of finite elements in this context, but few that delve into the subtlety of the modelling process itself. While an understanding of the theory is vital to the successful use of finite elements, emphasis will be given here to the practicalities of modelling. As well as describing the modelling process, examples—both simple test cases and real engineering problems—will be solved.

The text will consider in detail only the linear elastic small deformation behaviour of structures. This restriction is necessary to limit the length of the book, but it is justified as much of the analysis carried out by industry falls squarely within this category. Some insight is given into more advanced analysis techniques at the end of the book.

OBJECTIVES

After reading the material, a reader will be able to:

- understand the ways in which mechanical structures and devices transmit the loads applied to them
- appreciate the historical development of the finite element method
- appreciate some of the theoretical background to the fundamental equations that govern stress, strain and displacements
- follow the analysis process for a design
- model the geometry of a design on a computer
- define the mechanical properties of materials used in design (an open-ended problem similar to the modelling of turbulence in fluids)
- begin to use finite element software to predict stresses and displacements within designs for given applied loads
- have an understanding of the computer technology required to carry out an analysis.

ORGANIZATION

Chapter 1 looks at the design process, how designs are evaluated, the use of computers in this process and a historical review of the development of the finite element method. Chapter 2 then develops the background to continuum mechanics as it relates to the transmission of forces through a solid material, developing the governing equations for equilibrium, strain and compatibility as well as looking at material properties. Chapter 3 discusses the use of finite elements in producing numerical analogues to the equations that can be solved on a computer. After this more technical material, the main core of the text, follows as subsequent chapters look at the use and acquisition of the necessary hardware and software and at the process of using the finite element method to obtain design information. This is done by looking at this process in a series of stages: thinking about the design problem; creating a computer model of the geometry of the design; building a finite element mesh; applying boundary conditions; obtaining a solution; and analysing the results. Finally, Chapter 10 looks at advanced topics.

Throughout the text SI units have been used, with the exception that mm have been used for length in places. This usage should be clear when the text is read.

USING THE BOOK

This book is intended to be both an introductory guide and a working reference. Students at both undergraduate and postgraduate level who are studying courses on finite element methods may want to read the chapters in sequence, but those who are modelling may just at some stage want information on a particular aspect of the process of evaluating designs. For these readers we have tried to keep the individual chapters as self-contained as possible so that they may where necessary be read in isolation. We have also attempted to keep the mathematics to a minimum and to keep the introduction to the finite element method as general as possible so that a wide range of people from student to practising engineer will find the book of use.

We hope that readers will find this book a suitable bridge between the complementary areas of continuum mechanics, numerical analysis and computing that together enable computers to be used to evaluate designs. If some of the pain that is currently found when building this bridge is reduced then our effort will have been worth while.

Note that to keep production costs to a minimum, only monochrome figures have been used.

FURTHER STUDY

For some readers the presentation here will suffice, but others may want to learn more about processes both in mathematical and physical terms. For those who want to continue their learning into structural behaviour, several ways are open. The first is simply to observe structures and think about why they deform as they do, or why they are designed as they are. The second is to read other textbooks, which can be of great help. Appropriate references are given throughout this book to suitable texts. Finally, courses are often run by universities in the field of structural mechanics, and it may be that one of these may provide the necessary insight.

ACKNOWLEDGEMENTS

Both authors are grateful to the professional help given by staff at McGraw-Hill during the development of the material and to the reviewers who provided such helpful comments. Encouragement has come from many places and this book has been enriched by the examples provided by both contacts in industry and students. In particular, we are grateful to Michael Bradley for developing the model of the bicycle frame as a third-year undergraduate project, to Geriant Williams and Ralph Smith at Jaguar for providing the con-rod example, to William Turner and Andy Ellis at the Motor Industry Research Association for the vehicle body shell example and to Martyn Pinfold of the Warwick Manufacturing Group, University of Warwick and Geoff Calvert of the Rover Group, for the composite suspension arm.

Thanks are also owed to Prentice Hall for permission to use a modified form of material that has already appeared in Shaw (1992). Indeed, they actively encouraged this use!

Finally, thanks are owed to SDRC, and in particular our contact there, Andy Burton, for the provision of the I-DEAS software at the educational rate and for the encouragement given when this book was just a figment of our imagination.

Dr. J. Toby Mottram and Dr. Christopher T. Shaw
University of Warwick

NOTATION

ITALIC SYMBOLS

a	side length of quadrilateral element, mm
a_i	generalized coefficients in Rayleigh–Ritz method
A	area, mm ²
b	side length of quadrilateral element, mm
C_{ij}	elastic stiffnesses, N mm ⁻²
d	diameter of disk, lever arm, mm
D	diameter of bar, mm; objective function
E	Young's modulus, kN mm ⁻²
F	resultant force, N
G	shear modulus, kN mm ⁻²
h	length of real element, mm
I_a	second moment of area, mm ⁴
I_p	polar second moment of area, mm ⁴
J	determinant of [J] (the Jacobian)
J_Γ	boundary Jacobian
k	elastic stiffness, N mm ⁻¹
K_s	shear correction factor
K_{ij}	stiffness term, i is node number, j is nodal displacement
l, m, n	directional cosines
L	length, beam element, beam span, mm
m	mass, kg
M	moment, N mm; total number of elements
n	number of degrees of freedom; number of Gauss points in numerical integration schemes
N_i	shape function terms in matrix [N]
O	order, e.g. $O(h^2)$ is a term of order h^2
P	line load, N; number of concentrated load components
q	pressure load per unit length, N mm ⁻¹ per unit length
Q	heat source, °C mm ⁻²
r	radius, mm
R	radius of curvature, mm
S	surface area, mm ²
S_{ij}	elastic compliances, N ⁻¹ mm ²
t	thickness of plate, mm; time, s
T	temperature, °C; torque, N mm

u, v, w	displacement components in x -, y -, z -directions, respectively, mm
V	volume, mm^3
w_{nl}, w_{nk}	weighted factors in numerical integration scheme
x, y, z	Cartesian coordinate system
X, Y, Z	body force components per unit volume in x -, y -, z -directions respectively, N mm^{-3}
$\bar{X}, \bar{Y}, \bar{Z}$	surface traction components in x -, y -, z -directions respectively, N mm^{-2}

GREEK SYMBOLS

α_i	generalized coefficient(s)
γ	shear strain (e.g. γ_{xy} , γ_{yz} , γ_{xz})
δ	small bit of, finite element degree of freedom
Δ	change in length, mm, displacements due to external concentrated loads
Δt	increment in time, s
ε	direct strain (e.g. ε_x , ε_y , ε_z)
η	isoparametric coordinate, damping factor(s)
θ	angle of twist per unit length, rotational degrees of freedom, rad; weighting factor
ν	Poisson's ratio
ξ	isoparametric coordinate, damping factor
ξ_{nk}, η_{nl}	sample points in numerical integration scheme
Π_p	total potential energy functional
ρ	mass density, kg m^{-3}
σ	direct stress (e.g. σ_x , σ_y , σ_z), N mm^{-2}
τ	shear stress (e.g. τ_{xy} , τ_{yz} , τ_{xz}), N mm^{-2}
Φ	Airy's stress function
ω	natural frequency, Hz; relaxation factor

SUBSCRIPTS AND SUPERSCRIPTS

a	axial loading
b	bending
d	damped
e	element
e	element number
i, j, k, l	labels for nodes
m, n, o, p	labels for nodes
max	maximum
min	minimum
n	number of degrees of freedom, number of simultaneous equations

xvi NOTATION

nat	natural
new	new value
next	next value
old	old value
O	initial value
p	concentrated load component
s	shear
U	ultimate, upper
vonMises	von Mises' stress
w	web
x, y, z	Cartesian coordinates
Y	yield
1,2,3	principal values

MATHEMATICAL SYMBOLS

$[\]$	rectangular or square matrix
$\{ \ }$	vector or column matrix
$[\]^T$	matrix transpose
$[\]^{-1}$	matrix inverse
∂	partial differential
d	total differential
$\{a\}$	coefficients in Rayleigh–Ritz method
$\{b\}$	general vector
$[A]$	Vandermonde matrix, general square matrix, system matrix
$[B]$	spatial derivative(s) of the field variable(s) are $[B]\{\delta^e\}$
$[C]$	damping matrix
$[D]$	material stiffness(es) matrix
$[f]$	displacement function(s) matrix
$\{F^e\}$	element force vector = $\{F_q^e\} + \{R^e\}$
$\{F\}$	global force vector
$\{F_q^e\}$	consistent element force vector due to distributed loading
$\{F(t)\}$	forcing function vector
$[I]$	unit matrix
$[J]$	the Jacobian matrix
$[k]$	element stiffness matrix
$[K]$	global or structure stiffness matrix
$[L]$	lower triangular matrix
$[M^e]$	element mass matrix
$[M]$	mass matrix
$[N]$	element shape function matrix
$\{r\}$	residual errors vector
$\{R\}$	vector of applied load(s) corresponding to displacement components

$\{\mathbf{u}\}$	displacement components
$[\mathbf{U}]$	upper triangular matrix
$\{\mathbf{x}\}$	general vector
$\{\mathbf{X}\}$	body force(s) per unit volume vector
$\{\bar{\mathbf{X}}\}$	surface traction(s) vector
$\{\alpha\}$	generalized coefficient vector
$\{\delta\}$	global nodal degree of freedom vector
$\{\delta^e\}$	element nodal degree of freedom vector
$\{\tilde{\delta}\}$	nodal amplitude(s) vector
$\left\{\frac{\partial \delta}{\partial t}\right\}$	velocity vector of degrees of freedom
$\left\{\frac{\partial^2 \delta}{\partial t^2}\right\}$	acceleration of degrees of freedom
$\{\Delta_p\}$	displacement(s) associated with P external concentrated load(s) vector
$\{\epsilon\}$	strain vector
$\{\sigma\}$	stress vector
$[\partial]$	partial differential operator matrix

RELATIONSHIP BETWEEN DESIGN AND FINITE ELEMENTS

In mechanical design some form of strength analysis is usually required as part of the design process. Traditionally this has been done by simple engineering calculations, but as product performance becomes more important and as designs become more complex so these simple methods have become inadequate. With the increase in computational capacity and the availability of software that can predict loading for complex geometries and material behaviour so there has been an explosion in the use of the finite element method both in academia and in industry. To understand this growth and its implication for design, it is necessary to look at what the design process is all about, with particular reference to the field of mechanical engineering. By looking at the design process, it is possible to look at how design solutions are evaluated and at where computer technology fits into this process. Then the historical development of the finite element method can be discussed, to illustrate the benefits of using this technology.

From looking at the design process it will become clear that to appreciate fully both the background theory of the finite element method and the ways in which it is used, analysts must have some understanding of the mechanics that structures use to transmit loads. These mechanics can then be translated to a mathematical form suitable for solution by computer. Chapter 2 develops the understanding and Chapter 3 explains the translation before the remaining chapters show how the finite element analysis is carried out in practice.

1.1 THE DESIGN PROCESS

All around there are examples of artefacts that have been designed with their mechanical properties in mind. For example, computers have a structure that must be capable of transmitting the weight of the cathode ray tube and its associated electromagnets through to the desk. While doing this the structure must also not deform too much. Equally, the disk drive attached to the computer must be able to withstand the loads imposed by the spinning disk and the moving heads that read the data. Regardless of the product being considered, the process undertaken by its designers from first thoughts to final design will be similar.

There has been much research into the stages that are followed during the design process, but many people describe the process in similar terms. In particular, the key stages in the process (see Fig. 1.1) can be identified as follows:

- *Recognition of a need*, which may well come from market research or a request to tender in terms of a design brief for a new or modified product or from other sources such as the simple intuition of a designer.
- *Definition of the problem or specification*, where, from the starting point of a design brief or whatever, a full technical specification of the desired product must be developed through consultation between designers and those who want the product, for example those who sell into the market-place or the end-

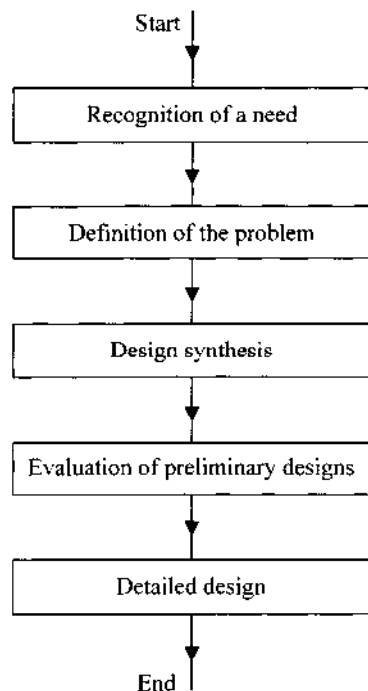


Figure 1.1 The design process.

users. Such a specification should contain all the relevant technical information describing the product, covering a variety of design considerations such as function, materials, appearance, environmental effect, product life, reliability, safety, interchangeability, standardization of parts, maintenance and service requirements and costs, together with any constraints that the design must meet. Further, the criteria that the design is to meet must be detailed. It is worth noting here that as well as essential items any desirable items must be specified too.

- *Design synthesis*, during which a variety of solutions are generated as rough ideas. Here the geometry of the product is most important and can be determined not only by the function of the product but also by the manufacturing process or the material to be used.
- *Evaluation of preliminary designs*, when each design is tested in some sense against the specification. For example, this analysis of a design might include simple physical testing or computer modelling, and should enable unsuitable designs to be filtered out. Also optimization might be achieved through a review of the designs, such that at the end of this stage a design that has the capacity to meet the specification in full and is best in some sense should be available.
- *Detailed design*, when the winning design is specified in full, i.e. in terms of size, materials, tolerances and shape, ready for manufacture.

A simpler version of this might consist of just four phases; clarification, concept design, embodiment design and detail design, but, regardless of the words used to describe the process, three key activities can be seen to be carried out by designers. These are:

- *generation* of solutions to the design problem
- *evaluation* of solutions to see if they meet the design specification
- *communication* of solutions, traditionally in the form of drawings, to many people such as customers and manufacturers.

In Fig. 1.1 the process progresses neatly from stage to stage. In reality, however, it is an inherent feature of the process that some form of iteration will take place. For example, problems might be found during the later stages because there is a lack of definition within the specification. If this lack of definition did not exist then there would be little or no scope for engineers and designers to be ingenious in generating solutions. However, it might mean that a review of the design specification is necessary and so the process loops back towards the start again. Also, none of the design solutions may meet the specification and so a rethink may be necessary.

When several designs are found that meet the specification the process changes into the search for the so-called optimal design. This is the one that is seen as best in some sense. Again iteration might be required to refine the best design still further.

It is worth noting that this might be considered a traditional view of the design process, as recent thinking considers the embodiment of the design process within the whole product life cycle. This is known as *simultaneous* or *concurrent*

4 USING FINITE ELEMENTS IN MECHANICAL DESIGN

engineering, where the design process takes into account not only the market into which the product must be launched, but also the manufacture of the subassemblies of the product and the final assembly. By doing this costly mistakes should be avoided and the product lead time reduced.

1.2 EVALUATING DESIGNS

In the evaluation process, all designs have to be checked for conformance with the specification. Clearly the specification has to be sufficiently detailed for the evaluation process to determine which designs meet the specification and which do not.

The list of areas covered in a typical specification indicates that various evaluation techniques may need to be used. In particular, the function category may include:

- strength under load
- displacement of the structure under load
- thermal behaviour
- fluid flow properties
- electrical/electronic behaviour.

All of these can be described as properties of the physical behaviour of the design. This book, however, is about the use of finite element methods in mechanical design—the design of the physical structure to carry an applied load or loads. Consequently the text is concerned with the use of continuum mechanics in determining such things as

- the strength of a structure
- the displacement of parts of the structure when loaded
- the effect of heat on internal stresses and displacement
- the optimal (minimum) thicknesses of material for a given displacement
- the fatigue life of a structure (or product life)
- the dynamic response of a structure
- crash worthiness.

1.3 USING COMPUTERS FOR DESIGN EVALUATION

For computer technology to be adopted at any stage of the design process there must be some benefit to the process as a whole. As the acquisition of the technology is fairly expensive these benefits have to be compared to the investment. For example, the following skills are required by a designer or engineer:

- *Analysis* Where physical or numerical experiments are carried out and the results are used for decision-making.

- *Reasoning* Where logic and intellect are used to produce a reasoned argument for following certain decisions. This is often developed through experience in engineering design.
- *Storage and handling of data* Where information is processed, both as input to a system and as output.
- *Error handling* Where errors to a given tolerance must be detected and corrected.

By adding computer technology to the armoury of the designer, the best human qualities of the designer can be linked by the best qualities of the computer. For example, humans excel at reasoning, developed from experience, error handling and the practical aspects of analysis. Equally, computers are far better at the storage and handling of data and the numerical aspects of analysis where repetitive calculations have to be made. The combination of these two, i.e. designers who use modern computer technology, should be able to produce design solutions more quickly and more accurately than would be possible if designers were to work without computers.

In summary, the benefits of using computer-aided design tools are as follows:

- *increased design efficiency and effectiveness* because repetitive tasks are carried out by the computer
- *simplification of the design process* by using an integrated data storage system, allowing many people access to relevant information when they want it
- *economy of material and labour* through a reduction in the amount of prototype building and testing required
- *better documentation* through computer-generated drawings, bills of materials, parts lists, work schedules and so on

1.4 TYPICAL DESIGN SITUATIONS

Section 1.2 listed some typical properties of a structure that might be of interest to a designer. In both aircraft and road vehicle design these properties need to be known before the first prototype flies or is driven on the road. Looking at each in turn:

- *The strength of a structure* needs to be known so that failure loads can be predicted. It is no use having an aircraft that cannot support its own weight when fully laden with fuel and passengers. Consequently calculations are carried out to predict the internal stresses in the wing structure when the aircraft is generating sufficient lift for certain manoeuvres plus an additional safety factor of load. At this condition the internal stresses should not lead to any failure of the material.
- *The displacement of parts of the structure when loaded* needs to be known. For example, the deflection of the wing tips of an aircraft when the wing is full of fuel will determine, in part, the design of the undercarriage.

- *The effect of heat on internal stresses and displacement* is particularly important in the area around the engine or engines. Hence, the displacements and loads due to the heating of the structure both of the engine itself and the surrounding vehicle body need to be determined.
- *The optimal (minimum) thicknesses of material for a given displacement* can be crucial to the handling characteristics of a road vehicle. For example, when designing an open-top version of a saloon car the removal of the roof makes the body shell much more flexible. Engineers need to know how much extra material will be needed in other parts of the structure to return the body shell to an acceptable level of stiffness. By calculating the displacements of some extremity of the shell under a given load for various configurations, the stiffness of the shell can be predicted as a function of the material added.
- *The fatigue life of a structure (or product life)* can be predicted by calculating the mean stress in the structure and the operating stress range seen at a known frequency. This might occur, for example, when an aircraft lands and standard design tables give the life of the structure before cracking initiates.
- *The dynamic response of a structure* is calculated to enable any resonances of the structure to be determined. This occurs if the structure is excited at frequencies close to the natural frequencies of the structure.
- *Crash worthiness* is particularly important for road vehicles and, in fact, a standard crash test must be carried out on at least one vehicle of each type before approval is given for the vehicles to be sold to the public. Predictions of the behaviour of the body shell are routinely made to assist in the design of the vehicle substructure that must absorb the energy of impact.

From these examples it can be seen that the ways in which mechanical designs carry and transmit applied loads to the supports are very important and need to be considered at an early stage in the design process. As the geometries are so complex, manual methods of calculation have proved difficult to apply and so a combination of manual and computer methods has become common.

1.5 HISTORICAL REVIEW OF OBTAINING STRUCTURAL DESIGN SOLUTIONS BY COMPUTER

1.5.1 The Computer Methods Available

In Chapter 2 the equations of equilibrium (2.5), compatibility (2.12) and stress-strain (2.15) will be developed for problems in linear static elasticity with small deformations. It is the solution of these partial differential equations with appropriate boundary conditions, i.e. loadings and restraints, that provides information that is useful to engineers when carrying out mechanical design. When the problem has simple geometry and simple boundary conditions these partial differential equations can be solved to provide classical solutions. In these cases the solution functions, for example Airy's stress functions in Sec. 2.3.2, are described

in the form of series expansions and at every point of the structure they satisfy equilibrium, compatibility and the boundary conditions.

Most engineering designs, however, are too complex for this classical approach. For example, a structure may have spatially dependent material properties if different materials are used; the geometry may be irregular in some sense or the boundary conditions may be complex. In all these examples no solution functions exist and so solutions can be achieved only by resorting to an approximate numerical method.

There are three numerical methods that are commonly used to solve partial differential equations throughout a three-dimensional domain. These methods are:

- the finite element method
- the finite difference method
- the boundary integral or boundary element method.

Of the three methods, the finite element method has become the most widely used when solving structural problems in both industry and academia. There has been a sustained research effort into the finite element method since 1960 and numerous commercial finite element programs are now available. Since then the range of problems being routinely solved has grown, starting with the original application of finding static elastic solutions, to embracing even those problems which possess large deformations and dynamic plasticity.

It is worth a note here to explain some of the philosophy behind the finite element method. Essentially, any problem can be split up into any number of smaller problems. With the finite element method this is done by considering that a complex geometrical shape is made up of a number of simpler shapes. For example, a circle might be approximated by a series of triangles in an attempt to calculate the area of the circle. This is known as *spatial discretization*, with each simple shape being known as an *element* and the whole collection of elements being known as a *mesh*. Within each element the relevant property of the element is predicted, say its area in the case of the triangles approximating a circle, or the relationship between forces and displacements for a structural element. This is done without any reference to other elements in the mesh. Here, people talk of forming the *element equations*, often by assuming known values of properties at fixed points on the elements known as *nodes*. Then the properties of all the elements and the interactions between them are taken into account by *assembling* the element equations and finding a solution to them. In the case of the area of a circle, the element equations calculate the area of the triangles and the solution process adds these together to predict the area of the triangle. Clearly, as more triangles are considered, the area predicted approximates ever more closely the area of the circle, which is known as the *convergence* of the solution.

While the interested reader will find many texts, conferences proceedings, journal papers and review articles concerned with applications of the finite difference method and boundary element method to structural problems, this book focuses on the finite element method. As will be seen from this historical review,

developments in mathematics, engineering and computing as shown in Table 1.1 have led to the finite element method's predominant position among the numerical solution methods. Inevitably, the review introduces mathematical terms that may not be familiar to the reader and a better insight into the historical developments may come after reading Chapters 2 and 3. The review is positioned here, to set the scene for the development of the underpinning principles of elasticity and the finite element method.

1.5.2 Some Developments Before 1945

The governing equations of compatibility, equilibrium and stress-strain relationships through material properties were developed well before 1830 by Lamé and Clapeyron. At this time digital computers did not exist and numerical methods for solving partial differential equations were only in their infancy, and so there were very few ways of obtaining useful engineering solutions. As will be discussed in Sec. 2.3, the only way forward for engineers was either to consider simple

Table 1.1 Developments in mathematics, engineering and computing (1830–1980)

Year	Mathematics	Engineering	Computing
1830	Governing equations developed	Use of classical methods	
1840			
1850			
1860			
1870	Calculus of variations		
1880		Use of subregions in structure	
1890		Direct stiffness method for frames	
1900			
1910	Rayleigh–Ritz method		
1920	Galerkin method		
1930			
1940		Use of triangular regions	Digital computer
1950		Use of bars Stiffness method	
1960		Energy methods	
1970		Commercial software appears	
1980			

structures such as bars and beams or to use the stress function approach for a two-dimensional and, on occasions, a three-dimensional continuum.

To progress further, developments in the engineering approach to finding solutions, in the mathematics of solving partial differential equations and in computing technology needed to be made. The first part of the jigsaw was produced by Rayleigh in 1870 when he developed a method for solving partial differential equations based on the minimization of potential energy through the calculus of variations. The only formal difference between this method and the Rayleigh–Ritz method, which was a mathematical stage in the development of the finite element method, is the number of unknowns. In the Rayleigh method the assumed distribution for the unknowns (e.g. displacements of a structure) have a single term only, whereas the Rayleigh–Ritz method can have as many terms as necessary.

Advances soon followed in the handling of the structural problem. In the 1890s engineers moved away from looking at the structure as a single unit, the approach of the method of classical elasticity, where the stress functions are found that satisfy the boundary conditions and also describe exactly the stresses, strains and displacements throughout the whole of the single structure. The new approach was to model the continuum, or structure, as an assembly of simple, structural subregions whose behaviour could be readily described by algebraic equations. Such a spatial discretisation process is natural to engineers and has been used extensively in structural problems ever since.

The most frequently used of these methods has been the *direct stiffness method* associated with structural engineering where buildings and bridges are often designed from slender members. In this area typical problems are analysed as plane frames (skeletal structures), i.e. using one-dimensional beam, torsion and bar members, as will be introduced in Sec. 2.3, and later as finite elements, as will be described in Sec. 3.6.1. Hence it can be seen that, in principle, the method allows complex structures to be analysed providing that the numerical manipulation to solve the algebraic simultaneous equations can be dealt with by hand. Here the development of iterative procedures, such as the moment distribution by Cross (1930) and the relaxation method by Southwell (1940), has benefited the design engineer by enabling solutions to be produced for modest numbers of unknowns without the digital computer. When the computer became available it was logical that engineers would implement the direct stiffness method and this has led to a number of refinements. Traditionally, the connections between members have been assumed to be fully fixed, but recently torsional spring elements have been included to model practical connections in steel frames, known as 'semi-rigid' connections. Refinement of the method allows the analysis to include secondary effects due to changes in the position of the loading as the structure is deformed.

In 1909 Ritz developed the Rayleigh–Ritz method of solving partial differential equations and because of its importance this will be presented in Sec. 3.1.2. This is a generalization of the earlier variational method of Rayleigh. In this method, engineering problems involving the structural continuum can be solved

by applying the principle of minimum potential energy to the single body problem, using a simple continuous approximating distribution for the unknowns, which were usually the displacements. Galerkin (1915) showed how such approximate solutions could be derived directly from partial differential equations, leading to the powerful *weighted residuals method* known as the Galerkin method. This is suited to field problems for which the partial differential equations are known but for which a variational statement cannot be found. Examples of such problems are incompressible viscous fluid flow and electromagnetic field theory.

At around the same time, Richardson (1910) used the finite difference method to approximate partial differential equations directly by using the values of the unknowns at points of an imaginary grid to form the derivatives. He used the technique to perform a plane strain analysis of the Aswan Dam. While finite difference procedures have been developed to solve complex problems, they have not become as popular as finite element procedures. There are a number of reasons for this, but in the main the finite difference method is not as good at modelling complex geometries as the finite element method owing to its need for a rectangular grid of points.

It was not until Courant (1943) proposed the idea of using locally defined simple linear distributions defined by specific values at the nodes, in his case for stresses, over a triangular subregion that the forerunner to modern finite elements was developed. His contribution preceded the modern computer and therefore lay dormant until its rediscovery long after the establishment of finite elements.

Clearly, engineers wanted to model a continuum using an assembly of simple structural elements, thereby reducing the problem to a set of simultaneous equations. While the earliest models preceded the modern computer, they were limited to modelling plane frame problems with regular grids of bars or beams. This approach was used by Hrenikoff (1941) and McHenry (1943) who both proposed to approximate the material of an elastic continuum by a lattice of simple bars.

1.5.3 Aircraft Design as a Driving Force (post-1945)

One of the main driving forces in the late 1940s for the development of structural analysis techniques was the need to predict the structural behaviour of aircraft wings. With known materials and simple wing forms the existing analysis methods were adequate, but the increase in aircraft speeds towards and beyond supersonic led to swept and delta wings being proposed. Analysis of these structures was not possible with the existing methods and so new methods had to be developed.

When the stiffness of each subregion is based on an assumed simple stress distribution determined using the forces at the nodes, which are the unknowns to be found, the method is known as the *force matrix method*. However, if the displacements of each node are taken as the unknowns, with an associated simple distribution of displacement assumed throughout the subregion, the method is called the *stiffness method*, and this method was developed by Levy (1953). It was further developed by Turner, Clough, Martin and Topp (1956) who made

the major advance of discretizing the continuum fully using an assembly of continuum subregions (or elements) of arbitrary triangular (Fig. 1.2) or quadrilateral form. The domain of the problem, the volume of the physical material itself, was divided into a number of subregions which were considered to be interconnected only at vertices or corners, which are nodes used to define the continuum (Fig. 1.3). Both force and stiffness matrix methods assume a simple piecewise continuous distribution for the unknowns throughout the element. Usually, polynomial functions are used, but occasionally sine and cosine functions are also used. These functions within an element are known within the finite element method by a variety of names such as trial functions or interpolation functions, but here they will be referred to as *shape functions*. Their existence and the restrictions on their form are fundamental to the success of the finite element method and so these will be discussed further in Sec. 3.5.

While there are a number of methods to determine the characteristics of finite elements—the *element stiffness matrix*, $[k]$, and the *element force vector*, $\{F^e\}$, in solid mechanics—by 1960, energy methods which assume a simple displacement

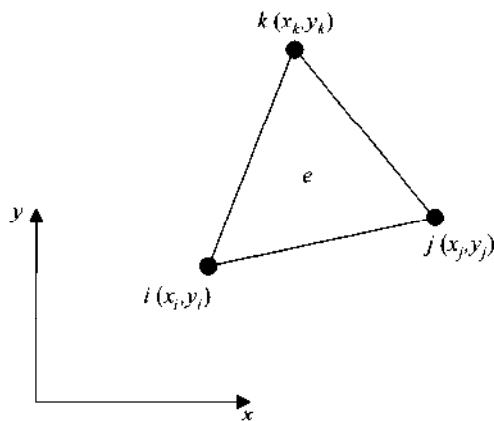


Figure 1.2 A two-dimensional three-noded triangular element.

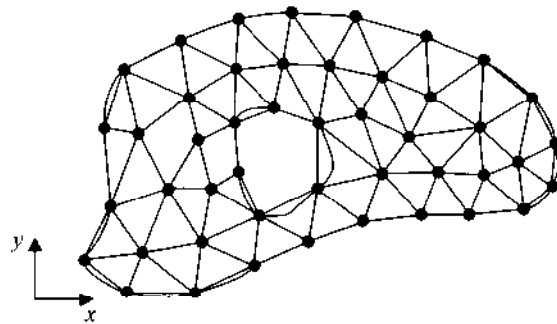


Figure 1.3 An irregularly shaped plate discretized by three-noded triangular elements.

distribution throughout an element had become the standard technique. An important contribution was made by Argyris and Kelsey (1960) who used energy theorems in modelling single triangular and quadrilateral aircraft panels, but the first comprehensive treatment of rectangular finite elements generated by an energy method had been given by Szmelter (1958).

These energy methods came out of a branch of mathematics known as the *calculus of variations*, developed by Rayleigh (1870), and so it can be seen that the wheel has nearly turned full circle, combining mathematics with engineering approaches to produce a workable method for engineering problems in structural design. An explanation and interpretation of this branch of mathematics for the finite element method is outside the scope of this text.

If there is a conservative structural system, i.e. one where the work done by internal forces and the work done by external loads are independent of the path taken, it can be shown that the equilibrium configuration is found by analysis of the potential energy of the system. From the calculus of variations, the variational formulation gives the *principle of minimum potential energy* for small displacement theory of elasticity. In Chapter 3, the finite element representation of this principle will be developed for $\{\mathbf{k}\}$ and $\{\mathbf{F}^*\}$ and their whole-structure equivalents $[\mathbf{K}]$ and $\{\mathbf{F}\}$. This principle can be regarded as equivalent to the virtual work principle.

So it can be seen that the energy content of a conservative system can be expressed in terms of its configuration without reference to how the system reached its present configuration. Potential or total potential energy includes the strain energy of deformation and the potential possessed by the loading, by virtue of its having the capacity to do work if it moves through a distance. The *principle of minimum potential energy* (*principle of stationary potential energy*) states that

Among all admissible configurations of a conservative system, those that satisfy the equations of equilibrium make the potential energy stationary with respect to small admissible variations of displacement.

It is this property of the potential energy that allows us to determine finite element characteristics. Note that the principle is not limited to linear elastic behaviour, but its extension to nonlinear behaviour will be covered only briefly in Chapter 10. Expressions for the potential energy of a structure are integral expressions, called *functionals*. Such functionals implicitly contain the differential equations that state the problem. Whereas the differential equations of classical elasticity (Chapter 2) state a problem in a *strong form* the integrals state it in a *weak form*. The strong form states conditions that must be met at every point in the structure, whereas the weak form states conditions that must be met only in an average sense. This statement indicates that a loss in accuracy must be expected in a finite element analysis, and that it will not necessarily give identical results to a classical solution, if the latter exists. In Sec. 3.1 the steps taken when using the energy method to obtain a general matrix expression for the $[\mathbf{k}]$ and the $\{\mathbf{F}^*\}$, as defined later in Eq. (1.1), will be discussed. Further details on functionals and their interpretation are to be found in many of the texts referred to in this chapter and Chapters 2 and 3.

1.6 OVERVIEW OF THE MODERN FINITE ELEMENT METHOD

The work of Turner *et al.* (1956) can be considered as the birth of the 'finite element method', and the name was first used by Clough (1960) who saw a model as consisting of a *finite number of elements* (or subregions). Now, however, the modelling approach of the finite element method must be discussed so that the basic mathematical method which determines the finite element characteristics can be produced. Here, finite elements based on simple stress shape functions (i.e. the force method) will not be dealt with, as the most commonly used elements are based on displacement shape functions (i.e. the stiffness method).

The basic idea behind the finite element method is to construct a structure from a *finite* number of elements, and, as will be demonstrated, these elements may be one-, two- or three-dimensional. In Sec. 3.6.2, the development of the stiffness matrix $[k]$ for the three-noded two-dimensional triangular element, as shown in Fig. 1.2, will be shown. When a two-dimensional structure (plane stress and plane strain are defined in Sec. 2.3.2) is constructed of hundreds or thousands of these non-overlapping straight-sided triangles connected at their vertices, it can be seen that all planar geometries can be accommodated. Figure 1.3 shows how an irregularly shaped plate can be discretized into a number of finite elements. At the vertices of the element in Figure 1.2 there are three nodes (i, j, k). These nodes define the elements that discretize the continuum; forces are transmitted by these nodes from one element to the next. The forces acting at the nodes are uniquely defined by the displacements of these nodes, the distributed loading acting on the element, and its initial strain (e.g. temperature, shrinkage, 'lack of fit'). To find the stresses, strains and displacements within the interior of the element, simple displacement functions (the assumed simple displacement distributions) with nodal displacements as the primary unknowns are used to provide a means of interpolation.

For those readers familiar with the direct stiffness method, recall that the behaviour of an element is defined by the matrix equation

$$\{F^e\} = [k]\{\delta^e\} \quad (1.1)$$

in which $\{F^e\}$ is the *nodal force vector* and $\{\delta^e\}$ is the *nodal displacement vector* containing the element degrees of freedom. The symmetric, positive definite, matrix $[k]$ is the *element stiffness matrix*, whose physical meaning is contained in the following statement:

the j th column of $[k]$ is the vector of loads that must be applied to nodal displacements in order to maintain the deformation state associated with unit values of displacement j while all other nodal displacements are zero.

For the three-noded triangular element in Fig. 1.2, the nodal displacements can be considered to be made up of two independent displacements. Similarly, two independent forces can be assumed to act at each node. These independent quantities are usually defined in terms of a Cartesian coordinate system, i.e. both forces and displacements have components in the x - and y -directions, such that at node i the

nodal displacements are u_i and v_i and the nodal forces are F_{xi} and F_{yi} . By the same reasoning, the element vectors $\{\mathbf{F}^e\}$ and $\{\delta^e\}$ have six terms and thus $[\mathbf{k}]$ is a 6 by 6 matrix. Each nodal displacement in $\{\delta^e\}$ is one of the degrees of freedom for the element and so our triangular element has a total of 6. It is the formulation of the terms in the finite element characteristics $[\mathbf{k}]$ and $\{\mathbf{F}^e\}$ that is fundamental to the finite element method, and Sec. 3.4 discusses in detail the methods available.

There are two principal reasons why the method has become the preferred numerical method to solve linear static structural problems using many discrete points (i.e. nodes) in modelling, and these are:

1. Once the stiffness and nodal force vector matrices in terms of geometry, nodal coordinates and material properties for an element type are known, the calculation of their terms for each occurrence of that element in the model is straightforward.
2. It is a simple process to take the individual element stiffness matrix and nodal force vector and then construct and solve the algebraic equations for the whole structure.

In matrix notation the equation representing the whole structural problem is

$$\{\mathbf{F}\} = [\mathbf{K}]\{\delta\} \quad (1.2)$$

in which both the *vector of nodal force*, $\{\mathbf{F}\}$, and the *vector of nodal displacements*, $\{\delta\}$, have n terms. $[\mathbf{K}]$ is the *stiffness matrix* (sometimes prefixed by *global* or *structure*) having n by n terms. It is not uncommon for models of commercial problems to have the number of degrees of freedom n in the order of tens of thousands, these being the combination of all the individual element degrees of freedom (each of which provides components in their $\{\delta^e\}$).

The procedures used in commercial software to generate the element characteristics and algebraic equations, to apply boundary conditions to them, and to solve for nodal displacements and element stress will be given later in Secs 3.6–3.9.

1.7 INDUSTRIAL IMPLEMENTATION AND RESEARCH IN FINITE ELEMENTS

Having given an engineering insight into the historical development of the finite element method, it is now time to discuss some of the other advances since 1870 that have contributed to making the method so successful in industry and academia. From the above discussion it can be seen that any structural problem can, in principle, be solved using the finite element method if a mathematical method is available to derive the element stiffness and element force matrices for a structural problem under consideration. From our knowledge of classical elasticity, as will be outlined in Chapter 2, it seems sensible to apply this method to the calculation of finite element equations. A classical solution requires its functions (e.g. Airy's stress functions) to satisfy, at the same time, not only the differential equations but also both displacement (or *essential*) and the stress (*non-essential* or *natural*)

boundary conditions. Consequently, solutions are found only for a complete structure and are limited to structures with simple geometries and boundary conditions. Such a restriction means that classical elasticity cannot be used to find the $[k]$ for a small subregion in the interior of a structure.

An important method for deriving element stiffness matrices $[k]$ is the direct method that is based on physical reasoning. It is limited to simple linear elastic elements such as a bar and a pure bending beam with constant properties along their lengths. Note that such elements are identical to those used in the direct stiffness matrix method. Attention to this simple method is often found in texts because it enhances the reader's understanding, but real structures cannot often be modelled using such simple elements so alternative methods are sought to develop $[k]$ when shape, material properties and behaviour are more complex.

Both variational and weighted residual methods have been developed to derive $[k]$ using piecewise continuous distributions for the nodal unknowns. For structural mechanics problems it is observed that the finite element characteristics are identical whichever of the two methods is used.

An interesting revelation to come out of the historical development before 1960 is that many methods, in different branches of mathematics, were independently developed, and that these were later brought together in the finite element method as the digital computer made its impact in permitting a direct solution of quite large systems of algebraic equations. As the 1960s progressed major advances took place in development and research, and these were accompanied by a growth in the number of publications from 10 in 1961 to 531 in 1969. The first textbook on the subject was written in 1967 by Zienkiewicz and Cheung. An exponential growth in research, development and publications continued in the 1970s. Specific journals for papers on numerical methods in engineering were created and their number and size have continued to grow ever since. Commercial finite element programs appeared in the market-place in the 1970s and the pace at which the finite element method has been allowed to develop mirrors the increasing capacity of suitable computing hardware up to the current situation described in Chapter 4.

The status today is that the annual contribution of papers continues to grow and it is estimated that some 5000 papers are published each year. Also, while mathematicians continue to put the method on firm theoretical ground, engineers continue to find new applications and extensions in many fields. The range of problems solved by structural finite element analyses has become extensive such that commercial programs can, for example, analyse successfully the behaviour of highly nonlinear plastic structures under impact. NAFEMS (National Agency for Finite Element Methods and Standards) in the United Kingdom has estimated that some of the major companies writing finite element software now have an annual turnover of some £50m each. From this it is estimated that the total annual investment is around £1 billion.

Future research will, no doubt, concentrate on tools to aid the design engineer such as error estimation and adaptive analysis. The number of algebraic equations to model a problem will also get larger and this will concentrate development on

stable iterative procedures, the use of parallel processors and a return to modelling with simple elements.

Recently there has been a development in the monitoring and controlling of programs. Users of finite element packages may be so impressed by the versatility of the pre- and post-processing features available that modelling and analysis limitations are ignored. As many of the commercial programs do not have accuracy estimation built-in, so users must learn by experience to produce accurate solutions. It can be reasonably assumed that all packages contain bugs either in terms of implementing wrong modelling assumptions or by introducing incorrect coding. To help users with problems when solving in real commercial situations, organizations such as NAFEMS and similar bodies in the United States have been formed. The publication of benchmark examples by these authorities for a range of problems has enabled users to test the accuracy of their own modelling technique when using software.

1.8 REFERENCES AND FURTHER READING

Interested readers might wish to look at some key texts. For example, the design process is discussed by Deutschman, Mitchels and Wilson (1975), Shigley and Mitchell (1983) and Pahl and Beitz (1988), and Lamit (1994) gives a good account of the design variables that need to be considered. Tizzard (1994) is an up-to-date text that discusses concurrent and simultaneous engineering as well as the skills needed by engineers. Similarly, Onwubiko (1989) looks at the benefits of computer-aided design.

Works on basic structural problems include Clapeyron and Lamé (1833), Airy (1863), Maxwell (1872), Rayleigh (1870), Ritz (1909), Galerkin (1915) and Timoshenko (1953). Slender members are discussed by Coates, Coutie and Kong (1988).

Key works on numerical methods include Richardson (1910), Cross (1930), Southwell (1940) and Shaw (1992).

For a review of the development of finite element methods, see Hrenikoff (1941), Courant (1943), McHenry (1943), Levy (1953), Turner *et al.* (1956), Szmelter (1958), Argyris and Kelsey (1960), Clough (1960), Zienkiewicz and Cheung (1967), Martin and Carey (1973), Washizu (1982), Norrie and de Vries (1976), Reddy (1984), Cook, Malkus and Plesha (1989), Zienkiewicz and Taylor (1989), NAFEMS (1992) and Cook (1995).

This chapter discusses the ways in which structures behave when they are subjected to loading. At first the development will be intuitive, but then a mathematical development of the description of a structure's behaviour will be given. Here the class of behaviour considered is that of small deformation of structures of a single material that is isotropic and generally linear elastic. Once a general mathematical description is available, the description may be simplified in certain situations and some of these simplifications will be outlined. In particular, it will be shown that manipulation of the general description leads to the simple formulae that engineers are taught during their first courses in structural mechanics. It is the solution of the equations of this same general description that computational methods such as the finite element method have been developed to solve. The chapter concludes with an outline of how problems in structural design are described.

Development of the finite element equations will be covered in Chapter 3 and the presentation of advanced topics such as time-dependent and nonlinear behaviour will be left until Chapter 10.

2.1 THE USE OF LOAD-BEARING STRUCTURES

There are examples of the use of load-bearing structures all around, but do people ever think about what the structures are actually doing? As the answer to this is probably 'no', the behaviour of load-bearing structures must be considered in some detail. Take, for example, a floor supporting a weight such as a person sitting on a chair, or shelves that are used to carry heavy books. In each of

these cases the structure acts to hold items, and itself, in place against the effects of gravity loading.

All objects have a mass, and an object's weight is the effect of gravity on the mass. This weight is a body force (i.e. internal load) and it acts on and is resisted by the structures mentioned above. If the structure can provide an equal and opposite force to that due to gravity, then the total force on the object will be zero and, if it is at rest, the object will remain at rest and stay in position.

Commonly, there are two parts to the load. A structure's self-weight is known as the *dead load*. For all but massive structures (e.g. bridges and aerospace vehicles) the dead load is small compared to that part of the load which is applied externally to the structure. This latter load is known as the *live load*. In the case of the books on the bookshelf the load's cause is the effect of gravity. In other circumstances, however, the live load can be due to other effects such as wind, snow, impact, acceleration and friction. For example, in machines many forces are generated by the operation of the machine. Consider a bicycle, where power from the rider is used to propel the bike-rider combination. Gravity acts on the rider and the bike, and so the wheels have to resist a gravitational dead load acting on the ground. However, the rider also exerts a resultant force on the pedals to provide the motive power, and so the pedal axle must resist this force and its associated moment which in turn has to be resisted by the crank and so on. Eventually the force exerted by the rider must be transmitted through the pedal, crank, gears and chain to the rear wheel. Hence, it can be seen that structures must not only resist forces but they must transmit them as well.

Let us consider an elastic band. If the band is pulled it resists the force that is exerted, and the greater the pull the greater the resistance. Structures may be thought of as being like a whole series of elastic bands or even springs, where forces between the atoms or molecules of the structure resist the deflection of the structure due to the external forces acting on it. When a bar is pulled, it is under tension, and if a bar is pushed it is placed under compression. A beam can also be bent by applying a moment to the beam that acts perpendicularly to its longitudinal axis, or a circular bar can be placed under pure torsion by twisting it, i.e. exerting a moment that acts in the plane perpendicular to the longitudinal axis of the bar. However, no matter which way it is loaded, be it under tension, compression, bending or torsion, the force exerted on the structure is resisted by the internal forces generated by moving the atoms or molecules of the material of the structure.

With an elastic band the deflection is large, but often structures are not thought of as deflecting under load. However, let us consider a shelf full of books, where it is quite often noticeable that the shelf deflects several millimetres because of weight of the books. This deflection is the result of the atoms or molecules in the shelf being moved apart or moved together. When this happens the interatomic forces that are generated provide the necessary force to oppose the weight of the books, and clearly the deflection depends on the properties of the structure's material.

From these simple examples the key ideas underlying the analysis of structures can be listed. If a structure is considered that has external loading applied to

it, and the structure neither moves as a rigid body after loading nor fails, i.e. it resists load by deflecting only a small amount, then it is known that:

1. The structure must be in *equilibrium*, with the loading being matched by appropriate reactions where the structure is restrained, such that the total force and moment acting on the structure are zero. The structure is therefore either at rest or travelling with a constant velocity.
2. The structure will deflect a small amount as the atoms or molecules of the material move and so generate forces inside the structure. Hence the structure is *stressed* internally, with each piece of material in local equilibrium and subject to *strains* which are a measure of the deformation of the structure and will be defined later.
3. The topology of the structure remains the same even though it has deflected. That is, the material of the structure moves such that the form remains the same, and it is said that *compatibility* of the structure is maintained.
4. The size of the deflections is determined by the *material elastic constant properties* of the structure.

Now, having studied the physical nature of structural mechanics the mathematical rules that govern the behaviour can be determined.

2.2 ENGINEERING STRESS AND STRAIN

Having established that all structures deform when subjected to a set of external and restraining forces that are in equilibrium, and that these forces are transmitted through the body, quantities are needed to describe the internal response, or state, of the structure. These quantities can be defined by considering the concepts of engineering stress and strain. The analysis of stress is essentially a branch of statics that is concerned with the detailed description of the way in which the stress (or intensity of force) at a point in the structure varies, whereas the analysis of strain is essentially a branch of geometry which deals with the deformation of an assemblage of particles. The variation of stress and strain throughout a structure, unless it is fracturing, is found to be continuous and described by single-valued functions. These functions cannot be found readily except for cases where the geometries and loading are simple, as discussed in Sec. 2.3.

2.2.1 Stress at a Point

The understanding of stress at a point is fundamental to stress analysis. In order to specify the forces acting within a structure an arbitrary plane can be chosen which passes through the material. On this plane the point O is located at the centroid of a small flat surface of area δA , as shown in Fig. 2.1(a). The *positive side* of the plane is defined by the outward normal to the plane ON , as shown, and the plane in the opposite direction is the *negative side*.

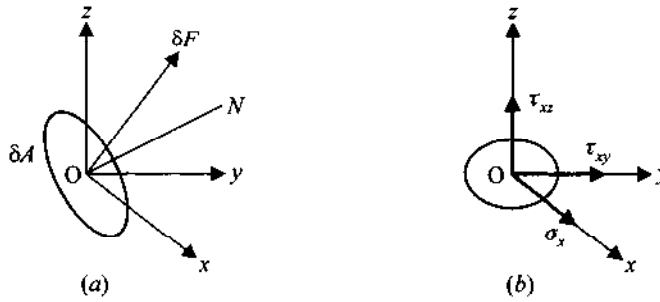


Figure 2.1 The forces and stresses on a small area: (a) forces, (b) stresses.

At all particles on the surface of δA the material on one side of the surface exerts a force upon the material on the other side, so that equilibrium is maintained if a cut is made along our arbitrary plane and these forces are applied at the cut. The resultant of the forces exerted by the material on the positive side of δA is a force δF . A force δF also acts on the material on the negative side of δA in the opposite direction to maintain equilibrium. Note that the resultant force is assumed to have its line of action through point O and to have produced no couple. Now, the total stress at the point O across the plane whose normal is in the direction ON is defined by

$$\text{Total stress}_{ON} = \lim_{\delta A \rightarrow 0} \frac{\delta F}{\delta A} \quad (2.1)$$

a vector whose magnitude has dimensions of force per unit area. In the physical situation, for every point O of the body and every plane through each point, a vector of total stress exists. Hence it can be seen that, to be strictly accurate, stress is not a vector quantity as not only the magnitude and direction of the vector but also the plane on which the stress acts must be defined. Stress is therefore a tensor, its complete description depending on the vectors of force (δF) and the outward normal to the surface of action (a vector in the direction ON).

2.2.2 Types of Stresses

It is convenient when developing the mathematical theory of stress analysis to set up a rectilinear coordinate system, such as the Cartesian, polar or spherical coordinate systems. Here the Cartesian system will be used as this is the most common in classical elasticity and the finite element method. Now place point O at the origin of three mutually perpendicular right-handed axes Ox , Oy and Oz . Let the normal ON be in the direction Ox as in Fig. 2.1(b). The surface δA lies in the $y-z$ plane. The total stress can be resolved into three components σ_x , τ_{xy} and τ_{xz} in directions Ox , Oy and Oz respectively. The component σ_x , which acts in the direction normal to the plane, is called the *direct* or *normal stress*. The other two components τ_{xy} and τ_{xz} , which act in the plane, are called *shear stresses*.

Two subscripts are needed to define a shear stress; the first subscript denotes the direction of the normal to the small area δA and the second subscript the direction in which the shear stress acts. Only one subscript is needed to define the direct stress since the direction is the same as that of the normal to the surface. This sign convention is related to the deformational influence of the stress. If σ is positive (Fig. 2.1b) it is called a tensile stress (i.e. the stress tends to stretch the material on the positive side of the surface away from that on the negative side), and if it is negative it is called a compressive stress.

The above argument can be extended to establish the three-dimensional stresses at a point. Figure 2.2 represents an infinitesimal rectangular parallelepiped (cuboid) with its sides parallel to a Cartesian coordinate system. The surface δA is now each of the six sides of the parallelepiped. The length of each side is taken to be so small that the stress components are uniformly distributed on each face and that they do not vary from the negative to the positive side. A three-dimensional state of stress at a point is shown in Fig. 2.2, where the directions of the stresses shown are defined to be positive.

The centroid of the parallelepiped is at C . A total of nine stress components define the state of stress at point C . By taking moments and considering equilibrium, it can be shown that the shear stresses have to be complementary. Hence

$$\begin{aligned}\tau_{xy} &= \tau_{yx} \\ \tau_{yz} &= \tau_{zy} \\ \tau_{zx} &= \tau_{xz}\end{aligned}\tag{2.2}$$

The six quantities σ_x , σ_y , σ_z , τ_{xy} , τ_{yz} and τ_{zx} are therefore both sufficient and necessary to define the stresses acting on the coordinate planes through point C .

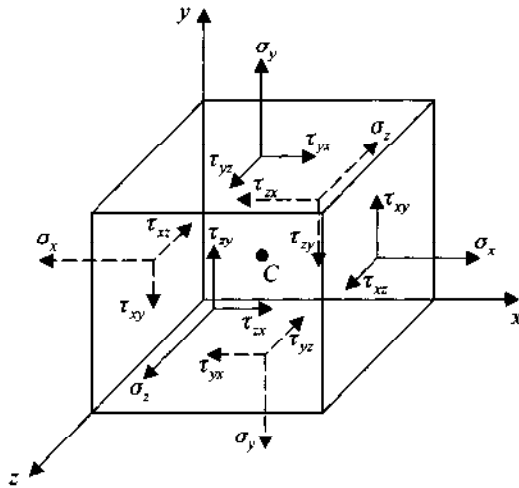


Figure 2.2 The stresses on an infinitesimal volume.

The parallelepiped in Fig. 2.2 can be rotated to any orientation within the structure about its centroid C . At each orientation the stresses, both direct and shear, acting on its faces are in static equilibrium. This gives mathematical relationships for these stresses in terms of the angles of rotations and stress components referenced to the Cartesian coordinate system. A visual representation for the variation of the stress components on the faces as the parallelepiped is rotated is obtained using Mohr's circle for three-dimensional stress. At one specific orientation it is found that there are no shear stress components on any of the faces. Then the three direct stress components that exist are known as the principal stresses (Fig. 2.3). These are given the notation of σ_1 , σ_2 , and σ_3 , where the subscripts define the axes of the principal coordinate system. Their directions are along the principal axes, and the planes on which they act are the principal planes. The maximum shearing stress acts on one of the planes at 45° to the planes on which the principal stresses act, and its value is given by

$$\tau_{\max} = \frac{\sigma_{\max} - \sigma_{\min}}{2} \quad (2.3)$$

where the direct stresses are the maximum and minimum principal stresses (i.e. the maximum is the most tensile, and the minimum is the most compressive). For example, in Fig. 2.3 $\tau_{\max} = \tau_2$ when $\sigma_1 > \sigma_2 > \sigma_3$. It is important to understand that principal stresses are readily derived by transformation from Cartesian values and that they are the parameters in conventional failure criteria (Sec. 2.2.6) that are often the useful output of a finite element analysis.

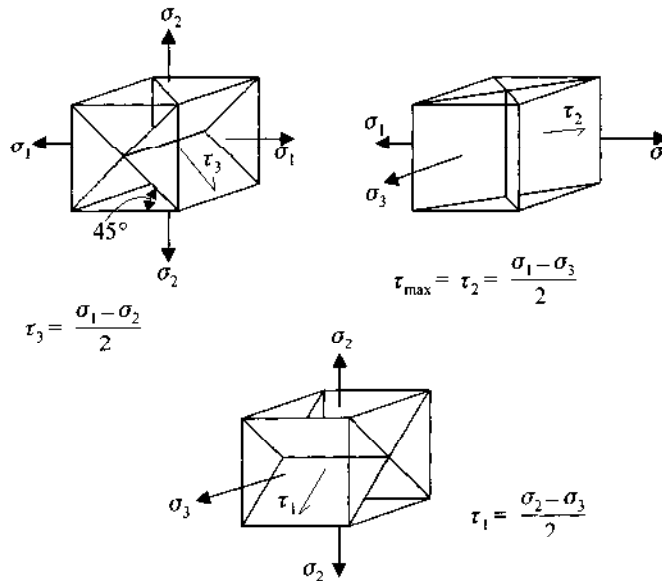


Figure 2.3 The principal stresses.

2.2.3 Differential Equations of Equilibrium

If the lengths of the sides of the infinitesimal parallelepiped in Fig. 2.2 are allowed to increase, it is now not a point that is considered but a small subregion enclosing a certain volume of material. In the general case, the direct and shear stresses on opposite faces are not equal. This variation in stresses is indicated in Fig. 2.4. Note that the shear stresses are defined using the notation on the left-hand side of Eq. (2.2). Thus if the direct stress on the negative-facing plane of constant x is σ_x then the direct stress acting on the positive-facing plane at $x + \delta x$ is, from the first two terms of a Taylor series expansion, $\sigma_x + (\partial\sigma_x/\partial x)\delta x$. It is assumed that the stresses acting on each face of the subregion do not vary over the surface such that their lines of action go through the centroid, as shown in the Fig. 2.4.

To establish a relationship for equilibrium the forces acting on each face of the subregion must be considered. It is the six stress components that produce these surface forces. For example, the resultant force in the x -direction on the negative-facing plane of constant x is $\sigma_x \delta y \delta z$. In the process of increasing the size of the rectangular parallelepiped from a point to a subregion material is introduced into the analysis of volume $\delta x \delta y \delta z$. This provides forces distributed over the volume of the subregion, such as gravitational forces, magnetic forces, or in the case of motion, inertia forces, known as *body forces*. The body forces per unit volume are resolved into the three Cartesian directions and the components are denoted by X , Y , and Z .

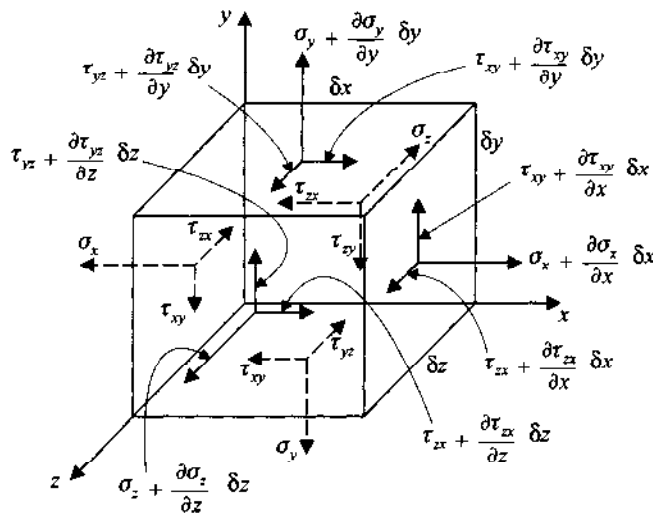


Figure 2.4 The stresses on the faces of a volume element.

In Fig. 2.4 the subregion is in equilibrium, acted upon by the surface forces and the components of the body forces, which are not shown. Considering equilibrium in the x -direction

$$\begin{aligned} \left(\sigma_x + \frac{\partial \sigma_x}{\partial x} \delta x \right) \delta y \delta z - \sigma_x \delta y \delta z + \left(\tau_{xy} + \frac{\partial \tau_{xy}}{\partial y} \delta y \right) \delta x \delta z \\ - \tau_{xy} \delta x \delta z + \left(\tau_{zx} + \frac{\partial \tau_{zx}}{\partial z} \delta z \right) \delta x \delta y - \tau_{zx} \delta x \delta y + X \delta x \delta y \delta z = 0 \end{aligned} \quad (2.4)$$

If similar expressions are written for the y - and z -directions, and then simplified, then

$$\begin{aligned} X + \frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} + \frac{\partial \tau_{zx}}{\partial z} &= 0 \\ Y + \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} + \frac{\partial \tau_{yz}}{\partial z} &= 0 \\ Z + \frac{\partial \tau_{zx}}{\partial x} + \frac{\partial \tau_{yz}}{\partial y} + \frac{\partial \sigma_z}{\partial z} &= 0 \end{aligned} \quad (2.5)$$

If no body forces exist, then X , Y , and Z are all zero. This is often the situation in mechanical design as the weight of a structure is much less than the forces it must transmit, but note that the body forces have to be included in the analysis when their magnitudes are of similar order to those of the external forces.

2.2.4 Strain at a Point

The stresses described in Sec. 2.2.2 cause small linear and angular displacements in the deformable structure. These small displacements are generally defined in terms of engineering strains, which are also known as Lagrangian strains. Note that there are a number of definitions for strain and each has its appropriate mathematical foundations. Engineering direct strains define the change in length produced by direct stresses while engineering shear strains define the change in angle produced by shear stresses. These strains are denoted, with appropriate subscripts, by ϵ and γ respectively, and they have the same sign convention as the associated stresses (i.e. σ_x and ϵ_x , τ_{xy} and γ_{xy} , etc.).

The displacement of any portion of a loaded structure can first be resolved into components u , v and w parallel to the x -, y - and z -axes, and the functions that describe the variation of these displacements over the volume of the structure are continuous and single-valued. Hence, the displacement in the x -direction, u , is defined by a function of x , y and z . It will be assumed that these components of displacement are very small quantities, and so, as was seen when the stress at a point was analysed in Sec. 2.2.2 an approximation for the strain can be made. Provided that the subregion of material being considered is sufficiently small, the strain within the material may be treated as a constant or homogeneous, giving the definition in the limit to the concept of strain at a point.

Consider the very small subregion with sides of length δx , δy and δz shown in Fig. 2.5(a) which is part of the undeformed structure. Choose three mutually perpendicular lines PA , PB and PC at a point P . If the structure undergoes a small deformation with displacement components u , v , and w at point P , it will move to P' , as shown in Fig. 2.5(b). Similarly, A goes to A' , B to B' and so on. The displacement in the x -direction of the point A on the x -axis is approximately $u + (\partial u / \partial x) \delta x$ (i.e. the first two terms of a Taylor series expansion). The

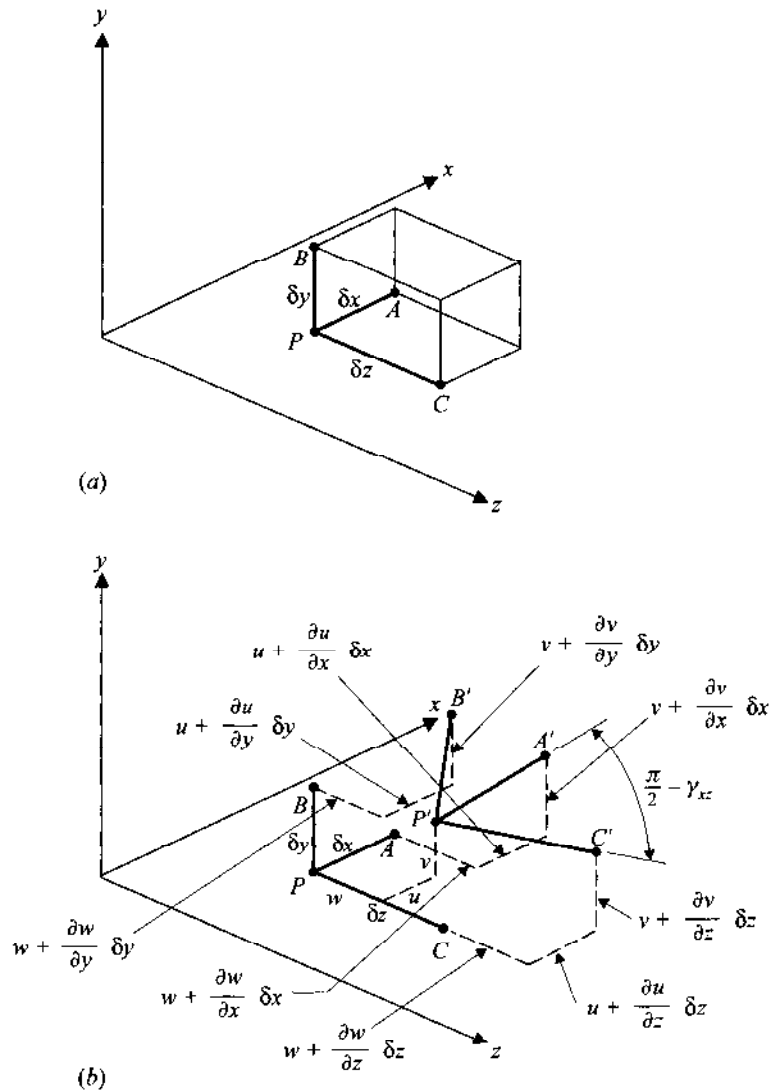


Figure 2.5 Deformation of a cuboid: (a) the undeformed shape, (b) the deformed shape.

remaining displacement components of the point A and those of the points B and C can also be found and are shown in Fig. 2.5(b).

Note that the approximation, in which second- and higher-order terms of the Taylor expansion are neglected, is consistent with the assumption of constant strains (i.e. constant displacement gradient) over the subregion. It is also consistent with the straight line PA in the undeformed structure becoming the displaced straight line $P'A'$ following deformation. The second-order terms would be included in the development of a geometric nonlinear analysis.

If a line of length L_0 undergoes a change in length ΔL then direct strain is defined as

$$\varepsilon = \frac{\Delta}{L_0} = \frac{L - L_0}{L_0} \quad (2.6)$$

where L is the final length of the line. This is known as the *engineering direct strain* and is acceptable for strains of magnitude less than 0.1. Note that the *true strain* is required when the extension of the current length is considerable. For true strain the change in length is referred to the instantaneous length, rather than to the original length.

The change in length of the line PA is $P'A' - PA$ so that the direct strain at P in the x -direction is obtained from

$$\varepsilon = \frac{P'A' - PA}{PA} = \frac{P'A' - \delta x}{\delta x} \quad (2.7)$$

From considerations of geometry and applying the binomial expansion

$$P'A' = \delta x \left(1 + \frac{\partial u}{\partial x} \right) \quad (2.8)$$

in which second- and higher-order terms of $\partial u / \partial x$ are neglected. Substituting for length $P'A'$ in (2.7)

$$\left. \begin{aligned} \varepsilon_x &= \frac{\partial u}{\partial x} \\ \text{and similarly} \quad \varepsilon_y &= \frac{\partial v}{\partial y} \\ \varepsilon_z &= \frac{\partial w}{\partial z} \end{aligned} \right\} \quad (2.9a)$$

Engineering shear strain at a point is defined as the change in the angle between two mutually perpendicular lines intersecting at the point. Therefore, if the shear strain in the x - z plane is γ_{xz} then the angle between the displaced lines $P'A'$ and $P'C'$ in Fig. 2.5(b) is $\pi/2 - \gamma_{xz}$ radians. From the geometry and displacements in Fig. 2.5(b) it can be shown that

$$\begin{aligned}
\gamma_{xz} &= \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x} \\
\gamma_{yz} &= \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \\
\gamma_{xy} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}
\end{aligned} \tag{2.9b}$$

The shear strains are complementary and so $\gamma_{xy} = \gamma_{yx}$, $\gamma_{yz} = \gamma_{zy}$, $\gamma_{zx} = \gamma_{xz}$. As strains are tensor quantities, like stresses, it follows that there are three principal strains (i.e. ε_1 , ε_2 and ε_3) whose magnitudes and directions can be obtained by transformation. This gives Mohr's circle for strain. It is worth noting that when the material has the same mechanical properties in all directions, i.e. it is isotropic, then the principal stresses and principal strains are aligned.

It is convenient when formulating the finite element stiffness matrix (Chapter 3) to use matrix notation and so the strain-displacement relationships (2.9a,b) can be rewritten as $\{\varepsilon\} = [\partial]\{u\}$ or in full

$$\begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 \\ 0 & 0 & \frac{\partial}{\partial z} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial z} & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} & 0 & \frac{\partial}{\partial x} \end{bmatrix} \begin{Bmatrix} u \\ v \\ w \end{Bmatrix} \tag{2.10}$$

2.2.5 Compatibility Equations

Equations (2.9), or (2.10), express the six components of strain at a point in a structure in terms of the three components of displacement u , v and w at that point. For the strains to be compatible over the volume then the material cannot overlap or form gaps (i.e. no voids are created). It follows then that the components of displacement can be given by

$$\begin{aligned}
u &= f_1(x, y, z) \\
v &= f_2(x, y, z) \\
w &= f_3(x, y, z)
\end{aligned} \tag{2.11}$$

where the functions are single-valued and continuous. If gaps are produced due to deformation, the displacements are discontinuous and separate functions either side of a gap are necessary. The existence of just three singled-valued functions for displacement is a statement of the continuity or compatibility of displacement.

Since the six strains described by (2.9) are functions of only three displacements, they cannot be independent. By differentiating these equations twice and substituting, the following six compatibility equations are derived:

$$\begin{aligned}
 \frac{\partial^2 \gamma_{xy}}{\partial x \partial y} &= \frac{\partial^2 \epsilon_x}{\partial y^2} + \frac{\partial^2 \epsilon_y}{\partial x^2} \\
 \frac{\partial^2 \gamma_{yz}}{\partial y \partial z} &= \frac{\partial^2 \epsilon_y}{\partial z^2} + \frac{\partial^2 \epsilon_z}{\partial y^2} \\
 \frac{\partial^2 \gamma_{zx}}{\partial x \partial z} &= \frac{\partial^2 \epsilon_z}{\partial x^2} + \frac{\partial^2 \epsilon_x}{\partial z^2} \\
 2 \frac{\partial^2 \epsilon_x}{\partial y \partial z} &= \frac{\partial}{\partial x} \left(-\frac{\partial \gamma_{yz}}{\partial x} + \frac{\partial \gamma_{zx}}{\partial y} + \frac{\partial \gamma_{xy}}{\partial z} \right) \\
 2 \frac{\partial^2 \epsilon_y}{\partial x \partial z} &= \frac{\partial}{\partial y} \left(\frac{\partial \gamma_{yz}}{\partial x} - \frac{\partial \gamma_{zx}}{\partial y} + \frac{\partial \gamma_{xy}}{\partial z} \right) \\
 2 \frac{\partial^2 \epsilon_z}{\partial x \partial y} &= \frac{\partial}{\partial z} \left(\frac{\partial \gamma_{yz}}{\partial x} + \frac{\partial \gamma_{zx}}{\partial y} - \frac{\partial \gamma_{xy}}{\partial z} \right)
 \end{aligned} \tag{2.12}$$

To gain further insight into the principle of compatibility, imagine a structure subdivided into a large number of small rectangular parallelepiped subregions before deformation. On loading the structure these subregions will be deformed into a system of general parallelepipeds. While the subregions are now deformed, they must be connected to each other in exactly the same way as before, i.e. the topology of the structure remains the same. If the components of strain satisfy the compatibility equations, then this situation will be forced to occur, and therefore the six equations of (2.12) must be satisfied when solving any problems in small-displacement elasticity.

2.2.6 Structural Materials and Hooke's Law

One consideration of the functional design of mechanical structures is that the materials used have adequate strength and stiffness. These materials are referred to as structural materials (e.g. steels, aluminium and titanium alloys, engineering composites, timbers and plastics). To carry out a successful stress analysis for the purpose of design analysts must have a knowledge of the material properties, in particular the elastic constants (e.g. Young's modulus and Poisson's ratio) and strengths. Other properties such as thermal conductivity, wear resistance and corrosion resistance may be relevant to the product's function, but are not of concern in this treatment of the fundamental principles.

Throughout most of the text, an important assumption is made that a structure consists of material which is continuous, homogeneous and isotropic. A body is homogeneous if it has identical properties at all points and it is considered to be isotropic when its properties do not vary with direction or orientation. Except for Chapter 10 the development of stress analysis will be limited to materials that are

linear elastic. While a number of traditional structural materials such as steels, aluminium alloys, particulate composites (such as concrete) and some plastics appear to meet these conditions when viewed on the macroscopic scale, it is apparent when they are viewed on the microscopic scale that they are anything but homogeneous and isotropic. For example, metals are generally made up of more than one phase, in the form of small crystal grains with different properties (strengths and moduli), such that they are actually heterogeneous. The reason why the stress analysis equations given here are still valid for the behaviour of metal structures is that the very large number of heterogeneous grains are uniformly distributed over the volume; thus making the polycrystalline material statistically homogeneous and isotropic. As a rule of thumb, a structure satisfies the requirement if its dimensions are at least 20 times the largest dimension of any grain or inclusion. Since most structures are considerably larger than the microstructural detail it is acceptable to use the assumptions stated.

A property which varies with orientation is said to be an anisotropic property. Metals become anisotropic when they are deformed severely in a particular direction, as happens in rolling and forging. Other examples of structural materials with anisotropic properties are fibre-reinforced composites, reinforced concrete, oriented plastics and timbers.

Having looked at the assumptions, let us return to the mathematical development. So far nine equations (2.5) and (2.9), or (2.5) and (2.12), have been produced for a deformable structure, involving fifteen unknowns (six stresses, six strains and three displacements). For a solution further equations must be found. These equations are provided by six material relationships which describe the stress-strain behaviour of the material. Note that the derivation of (2.5) and (2.9) does not make any assumption about this behaviour. It follows, then, that these fundamental equations are applicable to any type of deformable structure no matter how complex the material behaviour is.

The procedure for establishing the stress-strain behaviour utilizes certain simple experiments (e.g. in a standard testing machine) in which both stress and strain are measured and the relationship between them determined. This approach is used to produce idealized mathematical formulae relating stress and strain from which the response of any structures under more complicated deformation can be calculated. Note that the properties of an engineering material vary enormously with temperature, pressure, rate of strain and fatigue, so that any formula is only a reasonable approximation to the response of the material under limited conditions. For now, however, consider only the case of a material under normal laboratory conditions, which is an example of the ideal situation, being simple and one-dimensional.

Suppose that a bar of constant cross-sectional area is stretched in tension. The rate of stretching is such that fracture occurs within several minutes. Measurements from the test are therefore not affected either by a high strain rate or by the length of time the load is applied. Such a tension test is used to measure the tensile elastic constants of metals, plastics, fibre reinforced composites and so on, and is conducted in accordance with international and national

standards. It is often known as a *coupon test*. The measured quantities are engineering direct stress σ_a (where subscript 'a' denotes pure axial loading), which is the load applied divided by the original cross-sectional area and engineering direct strain ϵ_a which is $(L - L_0)/L_0$ (Eq. 2.6), where L_0 is the original length and L is the length when the stress is σ_a . If σ_a is plotted against ϵ_a , stress-strain curves such as those shown in Fig. 2.6 are obtained. Note that the area of the bar changes during the test, so that strictly σ_a as defined is not the true

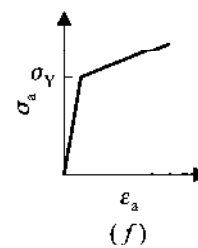
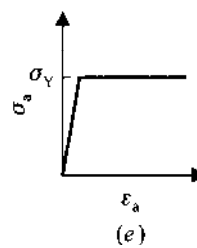
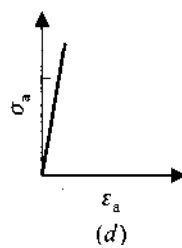
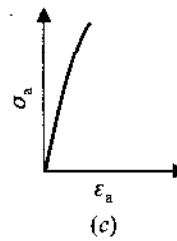
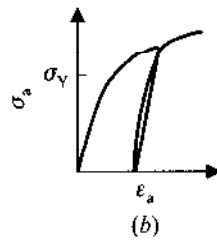
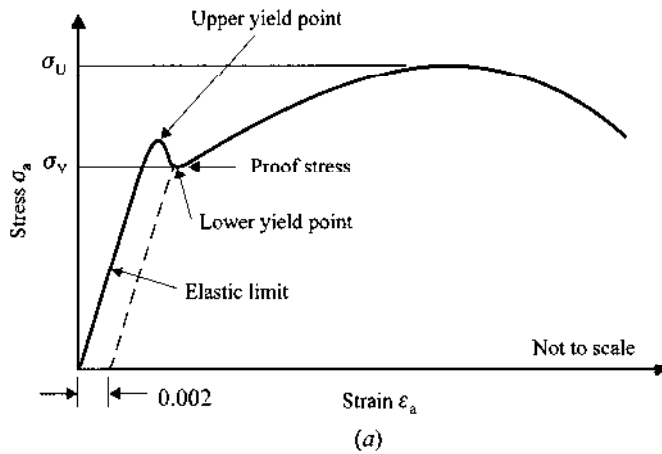


Figure 2.6 Some stress-strain curves for structural materials: (a) ductile steel, (b) aluminium alloy, (c) unidirectional fibre-reinforced composite material, (d) linear elastic model, (e) perfect linear elastic-plastic model, (f) linear elastic-strain hardening model.

stress. However, for small strains of magnitude less than 0.02 (i.e. ± 2 per cent) the difference between the stresses is small enough to be ignored, and the theory of infinitesimal strain described in Sec. 2.2.4 remains acceptable.

It is found that for an applied stress below a certain limit, if the stress is reduced to zero the bar returns to its original dimensions, i.e. there is no permanent deformation. This is the property of the material known as *elasticity*, and the range of stresses in which there is no permanent deformation is the *elastic range*. The most widely used structural materials, such as metals, composites and timbers, have an initial elastic stress-strain response which for mathematical modelling purposes can be assumed to be linear. This provides Fig. 2.6(d), the idealization of the perfectly linear elastic material. It is of paramount importance in mechanical design because many structures are designed to operate with low values of stress within them so that a linear stress analysis and thus the principle of superposition will be valid.

The linear relationship between stress and strain for a bar in tension (or compression) gives the definition of the modulus of elasticity, E , which is also known as Young's modulus, from

$$\sigma_a = E \varepsilon_a \quad (2.13)$$

Hence E is defined by the slope of the stress-strain curve in the linear elastic region, and its value depends upon the particular structural material being used. For materials that do not possess a linear elastic stress-strain range the definition for E is changed such that (2.13) may still be used. The quoted E for plastics is the *secant* modulus of elasticity. Table 2.1 presents values for a variety of material properties including Young's modulus for a number of structural materials. Unlike strength, it is insensitive to the precise microstructure. For example, steels have a value for E of 205–215 kN mm⁻² (GPa) while the yield strength can range from 250 to 1900 N mm⁻² (MPa). Equation (2.13) is commonly known as Hooke's law, but it is a very limited version of the generalized Hooke's law as it relates only the direct stress and strain developed in a bar subjected to pure axial loading.

As the bar undergoes longitudinal stretching there is an accompanying lateral contraction. Equally, lateral expansion is observed if the bar is compressed. In the linear elastic range, it is found experimentally that the lateral direct strains, say in the y - and z -directions, are related to the longitudinal direct strain, in the x -direction, by a constant, ν , such that

$$\varepsilon_y = \varepsilon_z = -\nu \varepsilon_x = -\frac{\nu \sigma_x}{E} \quad (2.14)$$

where ν is called Poisson's ratio (see Table 2.1).

For the three-dimensional state of stress, each of the six stress components is expressed as a linear function of six components of strain. This is the generalization of Hooke's law and can be written as

Table 2.1 Typical material properties of structural materials

Material	Young's modulus E kN mm^{-2} (GPa)	Poisson's ratio ν	Yield stress (or yield strength) σ_Y N mm^{-2} (MPa)	Ultimate strength ¹ σ_U N mm^{-2} (MPa)
Carbon steels	205	0.30	250–350	400–700
1.5% Mn steels	205	0.3	320–500	540–850
Stainless steels	215	0.3	280–1900	520–1970
Cast iron (malleable)	170	0.25		270–570
Aluminium cast alloys	71	0.30	55–250	145–330
Concrete	15–40	0.10		15–70 ²
Timbers ³	3.5–17			20–130
Plastics	0.06–11			5–130
Unidirectional fibre reinforced polymers ⁴	35–350	0.25–0.35		400–2000

Notes

1. Tensile strength unless otherwise stated

2. Compression strength

3. Parallel to grain

4. Parallel to fibres

$$\sigma_x = C_{11}\varepsilon_x + C_{12}\varepsilon_y + C_{13}\varepsilon_z + C_{14}\gamma_{xy} + C_{15}\gamma_{yz} + C_{16}\gamma_{zx}$$

$$\sigma_y = C_{21}\varepsilon_x + \dots$$

$$\sigma_z = C_{31}\varepsilon_x + \dots$$

$$\tau_{xy} = C_{41}\varepsilon_x + \dots$$

$$\tau_{yz} = C_{51}\varepsilon_x + \dots$$

$$\tau_{zx} = C_{61}\varepsilon_x + \dots$$

(2.15)

in which the terms C_{ij} are the components of the *elastic stiffness* matrix (also known as the *material property* matrix). In matrix notation, (2.15) can be written as $\{\sigma\} = [\mathbf{D}]\{\varepsilon\}$.

The matrix $[\mathbf{D}]$ is symmetric and its 36 terms C_{ij} are functions of the elastic constants for the material. However, some of these terms are often zero. For example, an isotropic, homogeneous material has 12 nonzero terms, which are functions of E and ν only. The zero values occur because:

1. Pure direct stress cannot give shear with respect to the same coordinate, i.e. σ_x is not a function of τ_{xy} or τ_{zx} .
2. Pure shear cannot give tension or compression with respect to the same coordinates, i.e. τ_{xy} is not a function of σ_x or σ_y .

3. Pure shear produces stresses only in the plane of the applied stress, i.e. τ_{xy} is not a function of τ_{yz} or τ_{zx} .

For the isotropic, homogeneous case

$$\varepsilon_x = f(\sigma_x, \sigma_y, \sigma_z) \quad \gamma_{xy} = f(\tau_{xy}) \quad (2.16)$$

and so on.

For ε_x , one can write

$$\varepsilon_x = S_{11}\sigma_x + S_{12}\sigma_y + S_{13}\sigma_z \quad (2.17)$$

where S_{ij} s are known as *compliances*. Equations (2.13) and (2.14) and experimental data give $S_{11} = 1/E$ and $S_{12} = S_{13} = -\nu/E$. It follows that the three-dimensional stress-strain relationships (2.15) can be written in the matrix form:

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \sigma_z \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \\ \gamma_{xy} \\ \gamma_{yz} \\ \gamma_{zx} \end{Bmatrix} \quad (2.18)$$

The material property matrix, $[\mathbf{D}]$ in (2.18) is for an isotropic material. Equations (2.18) are often presented in the more convenient strain-stress form of

$$\begin{aligned} \varepsilon_x &= \frac{1}{E} [\sigma_x - \nu(\sigma_y + \sigma_z)] & \gamma_{xy} &= \tau_{xy} \frac{2(1+\nu)}{E} \\ \varepsilon_y &= \frac{1}{E} [\sigma_y - \nu(\sigma_x + \sigma_z)] & \gamma_{yz} &= \tau_{yz} \frac{2(1+\nu)}{E} \\ \varepsilon_z &= \frac{1}{E} [\sigma_z - \nu(\sigma_x + \sigma_y)] & \gamma_{zx} &= \tau_{zx} \frac{2(1+\nu)}{E} \end{aligned}$$

Note that $E/2(1+\nu)$ is G , the shear modulus of elasticity. Remember, however, that the above relationships (2.18) are valid only when the material is isotropic and the stress is in the linear elastic range.

Formulae in terms of material elastic constants for the C_{ij} s for anisotropic linear elastic materials (e.g. fibre-reinforced composites) and for nonlinear stress-strain behaviour (e.g. plasticity of steels) are also available. Should the finite element analysis require such material stress-strain relationships the user should consult specialist texts and software vendors' manuals.

The values for the elastic constants determine how much deformation occurs. In other words, a structure under a given set of boundary conditions deforms twice as much if the Young's modulus is halved. In certain mechanical designs it is necessary to limit the deformation of the structure, as a deformation greater than some known value is seen as a failure. For example, an automobile body shell

must be sufficiently rigid so that the stresses transmitted from it to the windscreen glass are not large enough for the glass to crack; or for a vehicle without a roof, say an open-top sports car, the body must not deflect too much so that the handling of the vehicle is affected.

Although the deformation of a structure changes with choice of material, the stresses at any point remains unaltered as long as deflections are small. This is because stress is a measure of the intensity of force and this is independent of the material. However, all materials have limiting stresses in tension, compression and shear which cause the structure to fail mechanically. These failures can be due to the onset of yielding, gross cross-sectional plasticity (or ductile collapse), fatigue mechanisms and brittle fracture, but they can be accounted for in a design if the strengths of the material and their interactions are known. Table 2.1 presents typical strength data for some structural materials under normal laboratory conditions.

To illustrate how these strengths of materials are determined let us return to the stress-strain curves in Fig. 2.6. The stress-strain behaviour of structural materials differs considerably. As a result the strength of a solid requires careful definition. First, consider a ductile metal (i.e. carbon steels) with a distinct deviation in the stress-strain curve (Fig. 2.6a). The point where this occurs is the *upper yield point*. The stress then drops to a value called the *lower yield point* before increasing again as the strain continues. The corresponding yield strain is usually small, of the order of 0.001 (0.1 per cent). Other metals (e.g. aluminium alloys) do not have a clearly defined yield point and their stress-strain curves are as in Fig. 2.6(b). As the *yield stress*, σ_Y (sometimes referred to as *yield strength*), could therefore have a number of definitions it is usually taken to be *0.2 per cent proof stress*. The 0.2 per cent proof stress (Fig. 2.6a) is defined as that stress which results in a 0.2 per cent offset; the stress given by drawing a line parallel to the linear part of the curve through the 0.002 strain value. Also shown in Fig. 2.6(a) is the ultimate tensile strength, σ_U , the ultimate or maximum strength attained during a test, and also listed in Table 2.1.

Deformation at stresses above yield is plastic deformation, and for ductile metals this may be idealized either as perfectly plastic (Fig. 2.6e) or strain hardening (Fig. 2.6f) depending on the behaviour of the material. If the material is stressed beyond the yield point and the stress is removed there will be permanent deformation. Fig. 2.6(b) shows that when the load is reapplied it is found that the yield stress increases. For mechanical design, a knowledge of the strength is of utmost importance. Although σ_U seems to be the most reported strength for ductile materials, the yield or proof stress value is of greater importance as this tells designers when the material is no longer elastic and so they can determine the dimensions of the structure necessary to transmit the applied loads without causing failure. The ultimate strength provides a factor of safety against overload and mechanical failure.

In the mechanical design of structures the design must be fit for its purpose. This means that failure due to the expected loading is to be avoided over the design life of the structure. If the material of the structure is a ductile metal there should not, usually, be any gross cross-sectional yielding. When all, or part, of a structure is transmitting forces predominantly by compressive stresses there is the

likelihood that failure of the material is preceded by an instability mode of structural failure, and there are procedures to design against such behaviour. If instability does not occur before material failure then a material failure criterion may be used. There is at present no theoretical way of determining the relationship between the stresses and yielding for a three-dimensional state of stress with yielding in the uniaxial tension test. However, using experimental evidence and mathematical insight two yield criteria have now become generally accepted for predicting the onset of yielding in ductile metals (Dieter, 1984). They are the von Mises' or distortional energy and the maximum shear stress or Tresca's criteria. Both criteria depend on the principal stresses (see Sec. 2.2.2) and yield strength. The former is often a *default* output parameter of commercial finite element software. For this reason it is important to remember that the Cartesian stresses, shown in Fig. 2.2, determined in a finite element analysis can be used to determine the principal stresses and their planes.

The above discussion on material strength applies to metals. For plastics, strength is identified as the stress σ_Y at which the stress-strain curve becomes noticeably nonlinear. Typically, the strain is 0.001 (0.1 per cent). Other engineering materials can be classed as brittle in which little, if any, plastic deformation takes place. If a bar of brittle material such as a unidirectional fibre-reinforced polymer composite is tested, the stress-strain curve shown in Fig. 2.6(c) is obtained, where the curve is practically linear up to the point of ultimate failure when the material yields. In these cases a typical strain at failure can exceed 0.015 and for laminated plates used in the aerospace industry load-carrying capacity is limited by a maximum strain of 0.002.

For metals the compression stress-strain curve up to the yield point is very similar to the tension curve. This is not the case with other structural materials. For ceramics, glasses and concretes the strength depends strongly on the type of loading. In tension the fracture strength is one-tenth of the crushing strength in compression. Table 2.1 quotes some typical values. A full understanding of the strength of a material generally requires the determination of its stress-strain curves in tension, compression and shear. These curves provide several measures of strength according to the relevant mode of failure. Moreover, the material properties of structural materials have a different degree of variability which further complicates the definition of strength. It is the authors' opinion that specialist advice, usually from the material manufacturer or supplier, be obtained on material strength and that values, like Table 2.1, published in texts on selection of materials be used principally for guidance only. It is worth emphasizing that it does not matter how accurate the finite element representation of the structure and its boundary conditions is, the accuracy of the results will only be as good as the material property data available.

2.2.7 Boundary Conditions

The differential equations of equilibrium (2.5) must be satisfied at all interior points in a deformable structure under a three-dimensional force system. The

stress components vary over the volume of the structure, and when the boundary (i.e. the surface of the structure) is considered, then these stress components must be in equilibrium with the external forces there. These external forces are always distributed over some area of the boundary such that there is a stress distribution at the surface and these stresses are known as surface tractions. Hence, the external forces may be regarded as a continuation of the internal stress distribution.

If the rectilinear components of the surface forces per unit area are denoted by \bar{X} , \bar{Y} , \bar{Z} , it can readily be shown (Timoshenko and Goodier, 1988) that the boundary conditions for a three-dimensional body are

$$\begin{aligned}\bar{X} &= \sigma_x l + \tau_{xy} m + \tau_{zx} n \\ \bar{Y} &= \tau_{xy} l + \sigma_y m + \tau_{yz} n \\ \bar{Z} &= \tau_{zx} l + \tau_{yz} m + \sigma_z n\end{aligned}\tag{2.19}$$

where l , m and n are the direction cosines of the angles that a normal to the surface of the body makes with the x -, y - and z -axes, respectively.

For example, if the boundary of the structure has a plane surface, with an x -directed normal, then $\bar{X} = \sigma_x$, $\bar{Y} = \tau_{xy}$ and $\bar{Z} = \tau_{zx}$. Under this condition the applied surface traction components \bar{X} , \bar{Y} and \bar{Z} are balanced by internal stresses σ_x , τ_{xy} and τ_{zx} respectively.

It is usual in the classical methods of stress analysis to specify the loading boundary conditions in terms of surface forces (tractions or pressures), but it is interesting to note that the boundary conditions of a structure may also be given in terms of the displacement components. This is standard practice in a finite element analysis when the element formulation is based on assumed displacement fields (see Chapter 3). When displacement boundary conditions are given, the equilibrium equations (2.5) express the situation in terms of strains, through use of the stress-strain relationships (2.15) and in terms of displacements (2.9).

2.3 SOME SIMPLE SITUATIONS

When introducing the equations of equilibrium (2.5), compatibility (2.12), stress-strain for an isotropic material (2.15) and boundary conditions (2.19), it was stated that the exact solution to a problem must satisfy each of these sets of equations. One of the classical methods for solving the differential equations is to use ordinary differential calculus. Exact solutions for three-dimensional problems are very difficult to obtain because of the large number of unknowns (six stresses, six strains and three displacements) and, often, the complexity of the geometry of the structure. One way of making it easier to obtain a solution is to reduce the number of dimensions of the problem, and hence the unknowns. By doing this solutions for a large number of practical, if simple, structural problems can be obtained. Such solutions have provided the well-known design formulae for structural members such as bars, beams and thin plates.

While the differential calculus method is suitable for problems that are statically determinate, it is often not suitable if the structure is statically indeterminate. In this case a classical solution may be obtained by using an energy method.

2.3.1 One-dimensional Situations and Saint-Venant's Principle

The simplest structure to illustrate the classical solution using differential calculus is a weightless bar, where the body forces are zero, subjected to pure axial loading. This situation models, for small deformation, the gauge length in a tensile (or compressive, providing there is no buckling instability) coupon test discussed in Sec. 2.2.6.

Let the bar be positioned such that its longitudinal centroidal axis lies along the x -axis of a Cartesian system whose origin is at the left-hand end. It has constant circular cross-sectional area A , diameter D , length L , and external load F acting in the x -direction along the centroidal axis. Note that the cross-sectional shape need not be circular but must be constant along the length of the bar.

Equilibrium equations (2.5) are satisfied by taking

$$\sigma_x = \text{constant} = \frac{F}{A} \quad (2.20)$$

and

$$\sigma_y = \sigma_z = \tau_{xy} = \tau_{yz} = \tau_{zx} = 0$$

It is evident that the boundary conditions for the curved surface of the bar at

$$z^2 + y^2 = D^2/4 \quad x = 0 \text{ to } L \quad (2.21)$$

where there are no external forces, are satisfied, as the only load acts at the ends of the bar. The boundary conditions on the two ends whose surfaces lie in the yz -plane is simply $\sigma_x = \bar{X} = F/A$, from (2.19). Hence, there is a uniform distribution of stress over the cross-section of the bar if the tensile stresses are uniformly distributed over the ends. If this condition is invalidated by the load being applied in some different way, while maintaining the same overall equilibrium (Fig. 2.7), the solution to the problem is no longer exact. As this is the situation in practically every structural problem it would at first appear that the usefulness of the theory is strictly limited. To overcome this difficulty Saint-Venant's principle is invoked. It states:

While statically equivalent systems of forces acting on a body produce substantially different local effects the stresses at sections distant from the surface of the loading are essentially the same.

Figure 2.7 shows two bars with the same resultant axial load F . Each bar has the same average stress (F/A) at section BB , a distance usually taken to be greater than the minimum dimension of the surface to which the load is applied. However, at section AA closer to the ends, the stress distribution is locally altered owing to the precise nature of the loading. Therefore, the theory may be applied

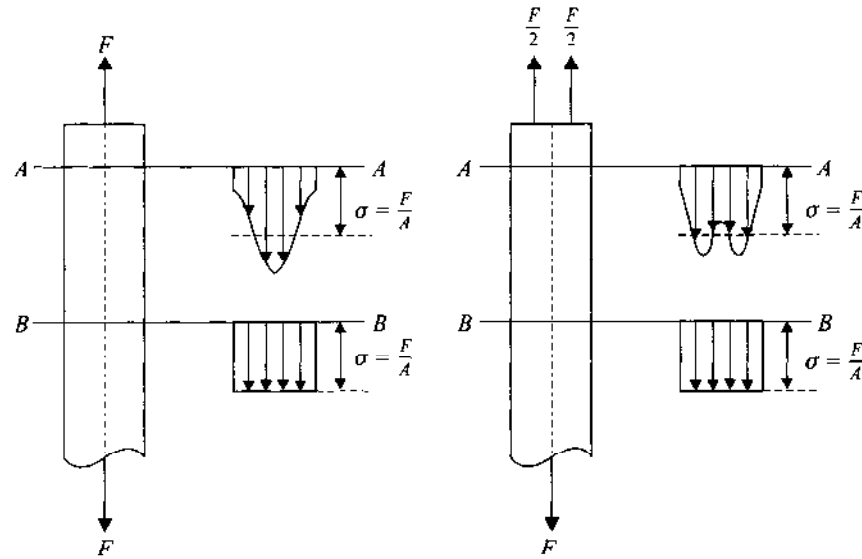


Figure 2.7 Saint-Venant's principle.

to those parts of a structure away from loading surfaces and restraints. In real structures the determination of local stresses which will include higher values than average is best accomplished using a two- or three-dimensional finite element analysis as will be seen in Chapters 7, 8 and 9.

Compatibility equations (2.12) are identically satisfied as the stress-strain equations give

$$\varepsilon_x = \text{constant} = \frac{\sigma_x}{E} = \frac{F}{AE} \quad (2.22)$$

with the other five strains zero. From (2.9) it is found that, for the one-dimensional situation, $\varepsilon_x = du/dx$. On substitution for the axial strain and integrating along the length of the bar, the familiar expression for the change of length is obtained

$$\Delta L = \frac{FL}{AE} \quad (2.23)$$

The exact solution for the axially loaded bar uses a modelling approach which lumps the properties of the structure along the centroidal axis, thus effectively reducing the three-dimensional problem to a one-dimensional problem. The same modelling approach is used to provide solutions for the pure torsion of cylinders and tubes, i.e.

$$\frac{T}{J_p} = \frac{\tau}{r} = \frac{G\theta}{L} \quad (2.24)$$

where T is the torque, I_p is the polar moment of inertia, τ is the shear stress, r is the radial distance from the axis of twist, G is the shear modulus, θ is the angle of twist per unit length and L is the length of shaft, and for pure bending of beams, i.e.

$$\frac{M}{I_a} = \frac{\sigma}{\pm y} = \frac{E}{R} \quad (2.25)$$

where M is the bending moment, I_a is the second moment of area, σ is the direct stress in the x -direction along the beam, y is the distance from the neutral axis, E is Young's modulus and R is the radius of curvature. The beam solution is not an exact one as it fails to model the secondary deformation that occurs in the plane perpendicular to the beam's length. Equation (2.25) is used to derive the standard bending deflection of beams. If the loading is other than pure bending, for example uniformly distributed loading, then there is additional deformation of the beam owing to the presence of shear.

Each of the three one-dimensional members introduced here is used in the direct matrix stiffness method and the finite element method to analyse structural frames. The latter method did indeed develop from the former in the late 1950s to analyse non-frame-type continuum structures, as discussed in Sec. 1.5.3.

2.3.2 Two-dimensional Situations

In many situations it is possible to consider structures which behave two-dimensionally over much of their volume. For example, consider a structure with a constant cross-section which lies in the xy -plane. With boundary conditions and body forces independent of z the problem is two-dimensional. Now, there are eight unknown quantities (i.e. three stresses, three strains and two displacements), each of which is a function of the x and y spatial coordinates.

With no body forces the equilibrium equations (2.5) become

$$\frac{\partial \sigma_x}{\partial x} + \frac{\partial \tau_{xy}}{\partial y} = 0 \quad \frac{\partial \tau_{xy}}{\partial x} + \frac{\partial \sigma_y}{\partial y} = 0 \quad (2.26)$$

and the compatibility equations (2.12) become

$$\frac{\partial^2 \gamma_{xy}}{\partial x \partial y} = \frac{\partial^2 \epsilon_x}{\partial y^2} + \frac{\partial^2 \epsilon_y}{\partial x^2} \quad (2.27)$$

Also, the strain-displacement relationships from (2.9) are

$$\epsilon_x = \frac{\partial u}{\partial x} \quad \epsilon_y = \frac{\partial v}{\partial y} \quad \gamma_{xy} = \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \quad (2.28)$$

Two linear elastic stress-strain relationships can be developed from the three-dimensional version given in (2.18) depending on the assumed size of the z -dimension.

In plane stress problems of plates or flat sheet material, where the thickness is small compared with the other dimensions, the surfaces of the sheet are frequently free of both normal stresses, σ_z , and shear stresses, τ_{yz} and τ_{zx} . It is therefore assumed

that these stress components remain zero throughout the thickness of the sheet, or at least that they are negligibly small compared with the other applied stresses. The state of stress in the sheet is then two-dimensional and the other nonzero stress components σ_x , σ_y and τ_{xy} may be averaged over the thickness. From (2.18)

$$\varepsilon_z = \frac{-\nu(1+\nu)}{(1-\nu^2)}(\varepsilon_x + \varepsilon_y)$$

and

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix} \quad (2.29)$$

In contrast to plane stress, in which the dimension in z is small compared to that in the x - and y -directions, the plane strain condition is characterized by a large dimension in z , for example with long tubes, dams and mines. In this situation it is assumed that the z -displacement component w is zero at every cross-section. The strain components ε_z , γ_{yz} , and γ_{zx} therefore vanish. Moreover, with ε_z set to zero, the stress σ_z can be expressed in terms of σ_x and σ_y . From (2.18)

$$\sigma_z = \nu(\sigma_x + \sigma_y)$$

and

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{Bmatrix} \quad (2.30)$$

and for plane strain problems a slice of unit thickness is usually considered. The general matrix form (2.29) and (2.30) is $\{\sigma\} = [D]\{\varepsilon\}$, in which matrix $[D]$ is 3 by 3.

Faced with eight unknowns and eight equations, it is of great advantage to reduce this number. A classical approach is to seek special functions that automatically satisfy some of the equations. The concept of a stress function is now introduced which has the property of representing all the stresses in such a way as to satisfy the equations of equilibrium (2.26). The Airy stress function $\Phi(x, y)$ (Fenner, 1986; Love, 1944; Timoshenko and Goodier, 1988), is defined as being related to the stresses by

$$\sigma_x = \frac{\partial^2 \Phi}{\partial y^2} \quad \sigma_y = \frac{\partial^2 \Phi}{\partial x^2} \quad \tau_{xy} = -\frac{\partial^2 \Phi}{\partial x \partial y} \quad (2.31)$$

Starting with the compatibility equation (2.27) and substituting for strains in terms of stresses using either (2.29) or (2.30), followed by substitution for stresses in terms of the Airy stress function Φ , the governing biharmonic equation is obtained

$$\frac{\partial^4 \Phi}{\partial x^4} + 2 \frac{\partial^4 \Phi}{\partial x^2 \partial y^2} + \frac{\partial^4 \Phi}{\partial y^4} = 0 \quad \text{or} \quad \nabla^4 \Phi = 0 \quad (2.32)$$

Equilibrium, compatibility and stress-strain relationships are satisfied by (2.31) and (2.32). Providing that a stress function Φ , often in the form of a series function, can be found to satisfy the biharmonic equation and the boundary conditions for the problem, the stresses are given by (2.31). Strains and displacements follow from the relevant equations and boundary conditions.

This classical method is restricted to problems with simple geometry and simple loading. However, the restrictions do not detract from the power of the technique to provide designers with relevant solutions to relatively complex problems. Such solutions are used to benchmark the performance of a two-dimensional finite element analysis. This will be illustrated in Chapter 7 using the so-called Brazilian test (or split tensile test) for brittle materials (John, 1992) which consists of a disc with diametrical opposite compression loading.

It is practical to solve simple three-dimensional problems by the classical method but because of the increased number of unknowns the mathematics becomes very complicated.

2.4 DESCRIBING PROBLEMS IN STRUCTURAL MECHANICS

In the preceding sections the theory of linear elastic structures has been considered. Before developing the equations of the finite element method in Chapter 3 it is worth considering the specification of a structural problem that must be made regardless of the analysis tool to be used. Readers should be aware that finite elements are not always the best analytical tool for a given problem. For example, an approximate but usable solution might well be found by the application of some standard formula. Equally, finite difference methods are ideal for shells of revolution and boundary methods are ideal for problems with boundaries at infinity. It is through careful specification of the problem that the analyst should have sufficient information to be able to decide upon the best analysis tool for the particular problem under consideration.

When thinking about the structural problem, the analyst must determine a specification of the problem. This should be a clear exposition of the reasoning that is driving the analysis, outlining what information is required at the end of the analysis and as much as is known about the physical problem itself. This specification then becomes the main source of information for the analyst.

2.4.1 Producing a Specification

A specification for a structural problem must be sufficiently detailed so that the analyst can obtain from it all the information necessary to define the problem. This information comes from a good understanding of the engineering problem which the analyst must obtain by discussion with the people who require the results of the simulation. In particular the analyst must know the following:

- the requirements of the analysis
- the geometry of the problem

- the possible range of materials to be used
- the physical loads and restraints on the structure.

Requirements of the analysis Analysing a structure can be an expensive business. If someone wants to commission a computational analysis then considerable expense will probably be involved as access to computer hardware must be provided along with the necessary software and highly trained analysts to produce the solution. Even if computers are not used the analysts must still be paid and, consequently, there must be good reasons for carrying out the analysis. The analysts must therefore explore these reasons first, by talking to the people who need the results of the simulation, such as design engineers.

The reasons for an analysis are many and varied. In Sec. 1.4 examples of where computers might be used to calculate the behaviours of structures were considered, i.e. strength, displacement, effects of heat, optimum material thicknesses, fatigue, dynamic response and crash worthiness. Further, each individual application has its own peculiarities. For example, there might be known limits for stress or displacement determined from experience or to meet legislative requirements, and analysts must know the appropriate form of the results that the simulation should produce; often they must do this from only a vague description of an engineering problem.

Once analysts know the reasoning behind the structural problem it is easier to plan ahead so that any computational model produces the necessary information. One further benefit of this discussion between analysts and their clients is that they get to know each other and their respective problems. Such an understanding can help the analysis process to be brought to a successful conclusion, especially if things do not quite go as planned.

At the end of this initial part of the specification phase analysts should have a list of the data that the computational model must produce. This could include the following:

- the peak stresses in the structure
- the displacement of certain parts of the structure
- the fatigue life of the structure
- the time history of the deformation or damage growth
- the energy absorbed during deformation.

Clearly, assessment of the suitability of results should be made with the clients.

At this stage it might be that the analysts judge that computational methods are not suitable for the physical problem being considered, perhaps because of the unreliability of numerical methods in this case, or because the cost is prohibitive when compared to physical experiment, for example. It is better to make this clear to the clients at this stage so as to prevent a loss of face or embarrassment later. Some pointers as to the suitability of numerical methods will be given in Chapters 7–9.

Specifying the geometry of the structure When looking at any structural problem it is important to be able to describe the geometry of the structure either in detail or as an approximation.

Various sources of geometrical data can be available and these can be used by the analyst to describe the bounding surfaces. For example, this data might come from:

- analytical descriptions of shapes in two dimensions given by such things as points, lines, arcs and splines
- engineering drawings
- databases created by computer-aided design (CAD) systems
- physical three-dimensional models
- measurements taken from existing components.

From such sources most of the bounding surfaces of the geometry may be determined precisely. This information might be used when building the mesh of points inside the domain, but during the specification stage it is sufficient to know roughly where these surfaces are in relation to each other and how they fit together. A simple sketch might help to show this.

Determining the material properties Having developed an understanding of the geometry of the problem the actual material properties for the structure need to be found. These might include the elastic constants, Young's modulus and Poisson's ratio, the material strengths and the effects of anisotropy as outlined in Sec. 2.2.6.

Defining the applied loading and restraints Once the geometry and material properties of the problem are understood the analyst must think about the way in which the structure will be loaded for the particular analysis. In mathematical terms this involves defining appropriate boundary conditions for the problem and, depending on the analysis, initial conditions. The first step here is to determine the external loads themselves in terms of their values as concentrated forces and pressures and the positions of their application on the structure. Then the positions where the structure is supported must be determined. Possibilities include a full restraint support where all displacements are set to zero, or a sliding restraint support where the motion of the boundary is constrained to be in a plane or along a line.

2.4.2 Physical Behaviour of the Structure

From all of the above information it should be possible to sketch out how the structure will behave physically. In this text only linear elastic small displacement problems are considered in the main, and so consideration must be given in design to the structural response having buckling instabilities and nonlinearities leading to large deflection and plasticity. If these are present in regions of interest then appropriate action should be taken as will be discussed in Chapter 10.

However as a guide:

If nonlinearities are localized and occur away from the regions of interest then the analyst is at liberty to ignore the behaviour and treat the problem as small displacement linear elastic.

Note also that a static analysis will provide results independent of the real behaviour of the structure and its material. In other words, if the loads are high enough the analysis will output displacements and stress values that would trigger, if activated, a nonlinear or instability response in the structure. For example, a long slender solid steel rectangular-shaped column, fixed at one end, free at the other end and subjected to a pure axial compressive load, i.e. there is no moment, is a classical bifurcation buckling problem of Euler. However, a static analysis of this problem results in ever-increasing displacement and stress as the load increases and there is no buckling when the load passes the critical Euler buckling value. Equally, stresses will continue to increase even when they exceed, many fold, the yield (and ultimate) strength of the steel.

Note that when modelling, real imperfections that always exist in a structure and its loading and supports are often ignored completely. Some account of this simplification needs to be made if buckling instabilities are a likely mode of failure.

2.5 REFERENCES AND FURTHER READING

In his classic book, Gordon (1976) analyses structural problems in some detail and in a readable style. Standard and more mathematical texts on the derivation of the governing equations and classical solutions are Love (1944), Timoshenko and Goodier (1988), Gere and Timoshenko (1991) and Fenner (1986). However, a useful summary of the formulae, facts, and principles pertaining to the classical methods is given in the compendium commonly known as *Roark's Formulas for Stress and Strain* Young (1989).

Material properties are discussed by John (1992) and Dieter (1986). However, anisotropy is covered by Calladine and Christopher (1985), Dieter (1986), Halpin (1992) and software vendors' manuals. Finally, the selection of materials is discussed by Ashby (1992) and Charles and Crane (1989).

FINITE ELEMENT SOLUTIONS OF THE EQUATIONS

Chapter 1 discussed structural design problems together with the historical development of the finite element method, while Chapter 2 discussed the governing equations for linear elastic structures. Consequently, the mathematical techniques for producing numerical solutions on a computer can now be reviewed.

Our starting point is energy methods, as more than 88 per cent of the continuum elements are derived by applying the principle of minimum potential energy over an element. The way the displacement of a structure is approximated is crucial to the finite element formulation, and so some time will be spent discussing these distributions. Direct methods of element equation formulation will also be discussed.

Whatever method is used to form the element equations, integration of functions must be performed through an element. This can be done analytically for elements of simple shape, but for many elements numerical integration must be used and so this is also discussed before looking at special element types. Finally, the ways in which the numerical solution of the equations is produced are discussed.

3.1 DEVELOPING THE FINITE ELEMENT METHODOLOGY

3.1.1 An Expression for Potential Energy

When an elastic structure is loaded by some external forces or moments its potential energy is increased. This potential energy is made up of the internal strain energy due to deformation and the potential of the loads that act within the

structure (e.g. body forces) or on its surface. The strain energy is generated as atoms are moved from their equilibrium positions as a result of deformation of the structure.

Some mathematical way of describing this process must now be found. In Chapter 2 the stress-strain relationship was defined, see (2.15) and (2.18), of which

$$\{\sigma\} = [D](\{\varepsilon\} - \{\varepsilon_0\}) + \{\sigma_0\} \quad (3.1)$$

is a modified form to account for any initial strains ε_0 and stresses σ_0 .

As strain energy is defined mathematically as one half of the product of stress and strain per unit volume, (3.1) can be used to develop an expression for this energy. By considering a linear elastic structure that carries conservative loads (i.e. the work done by the loads is path independent), and has volume V and surface area S , the general expression for the potential energy, Π_p , can be written as the functional (Washizu, 1982):

$$\begin{aligned} \Pi_p = \int_V \left(\frac{1}{2} \{\varepsilon\}^T [D] \{\varepsilon\} - \{\varepsilon\}^T [D] \{\varepsilon_0\} + \{\varepsilon\}^T \{\sigma_0\} \right) dV \\ - \int_V \{\mathbf{u}\}^T \{\mathbf{X}\} dV - \int_S \{\mathbf{u}\}^T \{\bar{\mathbf{X}}\} dS - \sum_{p=1}^P \{\Delta_p\}^T \{\mathbf{R}_p\} \end{aligned} \quad (3.2)$$

where, for a Cartesian coordinate system:

$\{\mathbf{u}\}$ is $[u \ v \ w]^T$, the displacement components throughout the structure (Sec. 2.2.4).

$\{\varepsilon\}^T$ is $[\varepsilon_x \ \varepsilon_y \ \varepsilon_z \ \gamma_{xy} \ \gamma_{yz} \ \gamma_{zx}]$, the strain components, given in terms of displacements by (2.10).

$[D]$ is the material property matrix, for example, $[D]$ in (2.18) is for an isotropic material.

$\{\varepsilon_0\}^T$ is $[\varepsilon_x^0 \ \varepsilon_y^0 \ \varepsilon_z^0 \ \gamma_{xy}^0 \ \gamma_{yz}^0 \ \gamma_{zx}^0]$, the initial strain components due to, for example, free expansion, shrinkage or moisture swelling.

$\{\sigma_0\}^T$ is $[\sigma_x^0 \ \sigma_y^0 \ \sigma_z^0 \ \tau_{xy}^0 \ \tau_{yz}^0 \ \tau_{zx}^0]$, the initial stress components due to, for example, pre-stressing or residual stresses.

$\{\mathbf{X}\}^T$ is $[X \ Y \ Z]$, the body forces per unit volume (Eqs 2.5).

$\{\bar{\mathbf{X}}\}^T$ is $[\bar{X} \ \bar{Y} \ \bar{Z}]$, the surface tractions (Sec. 2.2.7 and Eqs 2.19).

$\{\Delta_p\}$ is the vector of P translational and rotational displacements corresponding to their P external concentrated forces and moments in vector $\{\mathbf{R}_p\}$.

Now it can be seen that in (3.2) there are six terms. The first comes from (3.1) and is the main term as it accounts for the potential energy due to the current levels of stress and strain. The second and third terms also come from (3.1) and these account for the energy changes due to any initial stresses and strains. The fourth term represents the energy due to the presence of body forces, the fifth term that due to surface tractions and the final term that due to any external concentrated forces and moments.

Note that the integrals in (3.2) that contain $\{\mathbf{X}\}$ and $\{\bar{\mathbf{X}}\}$ are the work done (hence potential lost) by body forces and surface tractions as the structure deforms. Integration is performed only over the portions of volume V and surface area S where body forces and surface tractions act. They (i.e. the integrals in (3.2) that contain vectors $\{\mathbf{X}\}$ and $\{\bar{\mathbf{X}}\}$, and the potential energy calculated after solving these integrals), can be regarded as potential changes of $-Xu - Yv - Zw$ per unit volume and $-\bar{X}u - \bar{Y}v - \bar{Z}w$ per unit surface area, respectively. The final term represents the potential lost by concentrated forces and/or moments moving as the structure deformed and it can be written as $-\Delta_1 R_1 - \Delta_2 R_2 - \Delta_3 R_3 \dots - \Delta_P R_P$.

3.1.2 The Rayleigh–Ritz Method

Equation (3.2) relates the potential energy of a deformed structure to the strains within the structure and to the initial stresses and strains, the applied external forces, the body forces, the surface tractions and the material properties. Clearly, everything except the strains and stresses will be known from the definition of the structural problem.

When any structure is deformed, the continuum that makes up the structure moves from the equilibrium position to some other position. To describe this deformation in full the displacement at all positions within the continuum must be known, and it is said that the problem has an infinite number of degrees of freedom, i.e. an infinite number of points as discussed in Sec. 2.2. To make the problem amenable to solution, the number of degrees of freedom must be reduced to a manageable number. This can be done using the Rayleigh–Ritz method.

This method involves the approximation of the displacement components u , v and w for the whole structure by distributions which have the form

$$u = \sum_{i=1}^l a_i f_i(x, y, z) \quad v = \sum_{i=l+1}^m a_i f_i(x, y, z) \quad w = \sum_{i=m+1}^n a_i f_i(x, y, z) \quad (3.3)$$

in which the coefficients a_i , whose values are yet to be determined, are known as *generalized coordinates*. The functions f_i are known as *admissible functions* and are usually polynomials. For a function to be admissible it must satisfy compatibility conditions, Eqs (2.12) and displacement boundary conditions. It is not required that any of the functions f_i satisfy exactly stress (loading) boundary conditions, but if they do so then accuracy is improved. Here, the analyst must determine both the number of terms and the form of the functions in order to achieve the desired solution accuracy. Hence the unknowns are simply the generalized coordinates a_i .

To determine the n unknown generalized coordinates the following procedure is applied. The assumed displacement distributions are substituted into the strain–displacement expressions (2.9) to find strains $\{\epsilon\}$, then the functional (3.2) is used to evaluate Π_p . Thus the potential energy of the structure becomes dependent on the coordinates a_i alone. Now the principle of minimum potential energy is

applied, as discussed in Sec. 1.5.3. To find the minimum energy the differential of Π_p with respect to the all generalized coordinates must be set to zero to find the stationary values. Hence the deformed equilibrium configuration of the structure in terms of n algebraic equations is derived from the minimization as

$$\left\{ \frac{\partial \Pi_p}{\partial a_i} \right\} = \{0\} \quad \text{for} \quad i = 1, 2, \dots, n \quad (3.4)$$

Equations (3.4) are found to be stiffness equations and can be written in the matrix form $[\mathbf{K}]\{\mathbf{a}\} = \{\mathbf{F}\}$. This form is similar to Eq. (1.2), except that the unknown degrees of freedom are generalized coordinates and not nodal displacements. Numerical values for the coordinates a_i are obtained from solving (3.4). When substituted into the expressions in (3.3) they provide the approximate displacement distributions in the structure. Differentiation of the displacements u , v and w yields strains which, when substituted into the stress-strain relationships, will give stresses.

3.1.3 Using Local Displacement Distributions

Recall from the historical development in Sec. 1.5 that, by the 1940s, engineers and mathematicians had realized that it was not possible to find admissible functions, f_i , capable of describing the exact deformation when a structure had general geometry, boundary conditions and material properties. Such a realization was the catalyst for developing different numerical methods that treated a structure as a sum of smaller subregions (or elements).

As integration can be thought of as a summation process, it follows that by discretizing the structure into a finite number of elements and summing each of their potential energy contributions the method will approximate the total potential energy, as given by the single body functional (3.2). When implementing this method, the finite element method, the potential energy functional (3.2) must be rewritten as a summation over all the elements that make up the structure. The last term in (3.2) is not amenable to this and so it is treated as follows.

The displacement vector $\{\Delta_p\}$ in the last term is replaced by the vector of all displacements at all nodes (the degrees of freedom) in the model $\{\delta\}$. Note that nodes should be placed where the external concentrated forces and/or moments have been applied so that the displacements in $\{\Delta_p\}$ can be readily expressed in terms of nodal displacements. The vector of applied forces and moments $\{\mathbf{R}_p\}$, is replaced by the global vector of nodal applied forces, $\{\mathbf{R}\}$, which has terms associated with each of the displacements in $\{\delta\}$. Note that the forces and/or moments in $\{\mathbf{R}_p\}$ may have to be transformed before they are inserted into $\{\mathbf{R}\}$, such that the energy terms $-\sum_{p=1}^P \{\Delta_p\}^T \{\mathbf{R}_p\}$ and $-\sum_{\delta=1}^n \{\delta\}^T \{\mathbf{R}\}$ are equivalent.

Many terms in both $\{\delta\}$ and $\{\mathbf{R}\}$ will be zero, as nonzero terms are generated only where there are external applied loads.

Having transformed the last term in (3.2), the element form of the remaining terms can be considered. For the single body problem in Sec. 3.1.2 it was seen that the deformation could be approximated, although not necessarily accurately, by assuming the Rayleigh–Ritz displacement distributions of expressions (3.3). Now a similar approach is needed for a single element to provide admissible simple displacement distributions that can be used to generate the element characteristics, the element stiffness matrix $[\mathbf{k}]$ and the element force vector $\{\mathbf{F}^e\}$. For now it is sufficient to assume that the appropriate displacement distributions for an element can be interpolated from the element nodal displacements $\{\delta^e\}$. This can be defined in matrix form as

$$\{\mathbf{u}\} = [\mathbf{N}]\{\delta^e\} \quad (3.5)$$

where $[\mathbf{N}]$ is the shape function matrix, standard forms of which for various elements will be presented later.

Now, the derivation of the element representation of (3.2) can be completed as follows. In Chapter 2 the relationship between strains and displacements is given by (2.9) and in matrix form by (2.10). This can be rewritten as

$$\{\boldsymbol{\varepsilon}\} = [\boldsymbol{\partial}]\{\mathbf{u}\} \quad (3.6)$$

in which $[\boldsymbol{\partial}]$ is the differential operator matrix which, for a three-dimensional problem, is given by (2.10). Definition of the element strains in terms of the element nodal displacements is, therefore, given by

$$\{\boldsymbol{\varepsilon}\} = [\mathbf{B}]\{\delta^e\} \quad \text{where} \quad [\mathbf{B}] = [\boldsymbol{\partial}][\mathbf{N}] \quad (3.7)$$

when (3.5) is substituted into (3.6).

Substitution of the element expressions (3.5) and (3.7) for $\{\mathbf{u}\}$ and $\{\boldsymbol{\varepsilon}\}$ into the functional (3.2), and summing the contributions from M elements, gives the potential energy of the structure as

$$\Pi_p = \frac{1}{2} \sum_{e=1}^M \{\delta^e\}^T [\mathbf{k}] \{\delta^e\} - \sum_{e=1}^M \{\delta^e\}^T \{\mathbf{F}_q^e\} - \{\delta\}^T \{\mathbf{R}\} \quad (3.8)$$

Here, the first term comes from the first term in (3.2), the second term is a composite of the second, third, fourth and fifth terms in (3.2) and the final term is in the modified form discussed previously in this section. Hence it can be seen that from (3.8) the element stiffness matrix is defined by

$$[\mathbf{k}] = \int_{V^e} [\mathbf{B}]^T [\mathbf{D}] [\mathbf{B}] dV \quad (3.9)$$

and that the *consistent* (or *kinematically equivalent*) element force vector is defined as

$$\{\mathbf{F}_q^e\} = \int_{V^e} [\mathbf{B}]^T [\mathbf{D}] \{\boldsymbol{\varepsilon}_0\} dV + \int_{V^e} [\mathbf{B}]^T \{\boldsymbol{\sigma}_0\} dV + \int_{V^e} [\mathbf{N}]^T \{\mathbf{X}\} dV + \int_{S^e} [\mathbf{N}]^T \{\bar{\mathbf{X}}\} dS \quad (3.10)$$

where V^e and S^e are the volume and surface area of an element respectively.

Equation (3.10) shows how certain distributed loads can be transformed into equivalent nodal forces associated with the nodal displacements used to describe the deformation of the structure. The word 'consistent' is used to indicate that the transformation is consistent with the shape functions N_i of (3.5) used to generate $\{\mathbf{k}\}$, and that under certain conditions (see Sec. 3.7.7) the work done by the actual loading is equal to that in the finite element analysis. To complete an element force vector in (1.1), the external concentrated forces and/or moments from $\{\mathbf{R}\}$ may be added to the consistent force vector to yield

$$\{\mathbf{F}^e\} = \{\mathbf{R}^e\} + \{\mathbf{F}_q^e\} \quad (3.11)$$

Usually this step is done when the characteristics of each element are assembled into the governing simultaneous equations. Equation (3.11) has been given here to show that an element contribution to $\{\mathbf{F}\}$ has two different components and is not due to (3.10) alone.

Of the two element characteristics the expression for $\{\mathbf{k}\}$ is the more important. Although the expression (3.10) for $\{\mathbf{F}_q^e\}$ looks formidable in typical static problems, many of the terms $\{\epsilon^0\}$, $\{\sigma^0\}$, $\{\mathbf{X}\}$, $\{\bar{\mathbf{X}}\}$, contributing to the consistent force vector are zero. Consequently, body forces, initial strains and initial stresses are generally ignored in the rest of the text to simplify the presentation. To find out how all these energy terms alter the formulation of $\{\mathbf{F}_q^e\}$ interested readers are referred to the general texts in Sec. 3.10.

In making this simplification expression (3.10) reduces to a single term for the surface tractions:

$$\{\mathbf{F}_q^e\} = \int_{S^e} [\mathbf{N}]^T \{\bar{\mathbf{X}}\} dS \quad (3.12)$$

To complete the element derivation the algebraic equations that are solved to yield the displacements must be obtained. Every displacement in the element vector $\{\delta^e\}$ also exists in the global (or structure) vector $\{\delta\}$ of displacements. As a consequence of the assembly process which must include each of the M element contributions, the last two terms in (3.8) can be combined to give

$$\Pi_p = \frac{1}{2} \{\delta\}^T [\mathbf{K}] \{\delta\} - \{\delta\}^T \{\mathbf{F}\} \quad (3.13)$$

where

$$[\mathbf{K}] = \sum_{e=1}^M [\mathbf{k}] \quad \text{and} \quad \{\mathbf{F}\} = \sum_{e=1}^M \left(\{\mathbf{F}_q^e\} + \{\mathbf{R}^e\} \right) \quad (3.14)$$

Now, minimization of Π_p to find its stationary state with respect to a small variation in each of the nodal displacements δ in $\{\delta\}$ gives

$$\left\{ \frac{\partial \Pi_p}{\partial \delta} \right\} = \{\mathbf{0}\} \quad (3.15)$$

the solution to which is

$$[\mathbf{K}]\{\delta\} = \{\mathbf{F}\}$$

i.e. (1.2) again. Solving the n simultaneous equations in (1.2) yields the n degrees of freedom in $\{\delta\}$. After these primary unknown displacements have been determined, strain and stresses in an element may be calculated from (3.7) and

$$\{\sigma\} = [\mathbf{D}][\mathbf{B}]\{\delta^e\} \quad (3.16)$$

respectively.

It is important to appreciate that the process used in the finite element method, as given by (1.2), is universal and is therefore independent of the nature of any static element type used to model a structural problem.

3.2 SATISFYING EQUILIBRIUM AND COMPATIBILITY

Chapter 2 discussed what must be achieved for an exact solution of a static problem. It was shown that in an elasticity solution every small subregion enclosing a volume of material is in static equilibrium (Sec. 2.2.3) and compatibility is satisfied throughout (Sec. 2.2.5). In Sec. 3.1 a finite element methodology was developed which discretized the problem by using displacement distributions to interpolate nodal unknowns within an element. Therefore it is expected that an approximate finite element solution will not meet the necessary requirements in every sense. The degree to which equilibrium and compatibility may be satisfied at nodes, across interelement boundaries and within each element is as follows:

1. *Equilibrium of nodal forces and moments is satisfied* Structural equations $[\mathbf{K}]\{\delta\} = \{\mathbf{F}\}$ are equilibrium equations at nodes. Therefore, the solution for nodal displacements is such that nodal forces have zero resultant at every node.
2. *Compatibility prevails at nodes* Elements are compatible at nodes to the extent of the nodal degrees of freedom they share. This general statement allows for hinge or roller connections and is not restricted to permanent connection as is the physical situation for nodes representing a point in the continuum.
3. *Equilibrium is usually not satisfied across interelement boundaries* When examining element stress results, the stresses in adjacent elements do not necessarily have the same values along the common boundary or at common nodes. For a properly designed mesh these discrepancies are small and can become insignificant when the mesh is refined in the limit.
4. *Compatibility may or may not be satisfied across interelement boundaries* Incompatibility between elements, when it occurs, should tend to zero as more and more elements are used to model a structure. Indeed, this *must* be true if such an element is to solve real problems.

5. *Equilibrium is usually not satisfied within elements.* In general, to satisfy the differential equations of equilibrium (2.5), at every point in an element demands a relationship between the nodal displacements $\{\delta^e\}$ that usually does not exist from the solution of (1.2). The exact solution does exist for certain element types as the size of the element becomes smaller and smaller.
6. *Compatibility is satisfied within elements.* This is achieved if an element displacement distribution, as defined by (3.5), is continuous and single-valued.

The first four of these conditions result from the process of discretizing the continuum—the structure—into elements, whereas the last two conditions relate to the elements themselves. It is important that any inherent discretization error is not increased significantly, but conditions 3 and 4 infer that the accuracy of solution is a function of the mesh construction. However, conditions 5 and 6 lead to a number of requirements that should be satisfied when choosing the polynomial expressions that define an element's displacement distributions. This will be discussed further in Sec. 3.5.2.

3.3 USE OF ELEMENTS WITH DIFFERENT SPATIAL DIMENSIONS

When modelling any structural problem, the geometry must be split into a variety of elements. To do this, elements are used which essentially have one of the five basic forms shown in Table 3.1. These range from a single lumped mass to a three-dimensional volume and are used to model the situations listed in the table. Programmed within the largest commercial finite element codes, which can model various elasticity problems (both linear and nonlinear) and a variety of other field problems, such as thermal conduction, might be more than 100 different elements. A typical range of such elements are shown in Table 3.2. The

Table 3.1 The basic element geometries







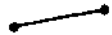

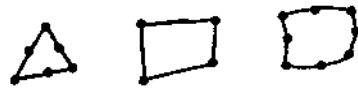
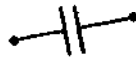

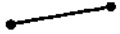
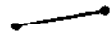
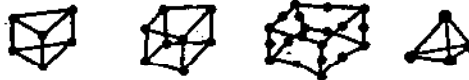


Dimensionality	Type	Geometry
Point	Mass	
Line	Spring, beam, bar, spar, gap, torsion	
Area	2D continuum, axisymmetric continuum, plate or flat shell	
Curved area	Generalized shell	
Volume	3D continuum	

Table 3.2 Typical range of elements available in commercial software packages

Element type	Degrees of freedom	Representation
Mass	—	
2D bar	u, v	
2D beam	v, θ_z	
2D continuum plane stress plane strain axisymmetric	u, v	
2D interface	u, v	
Axisymmetric shell	u, v, θ_z	
3D bar	u, v, w	
3D beam	$v, v, w, \theta_x, \theta_y, \theta_z$	
3D solid	u, v, w	
3D shell	$u, v, w, \theta_x, \theta_y, \theta_z$	
3D interface	u, v, w	

number and nature of the nodal degrees of freedom, both displacement and rotation, for each element are given in the table. Note that the rotation (or slope) degrees of freedom for plate bending and shell elements are obtained by differentiation of a displacement component.

In the physical world all structural forms must be three-dimensional, so why are one- and two-dimensional elements found in finite element codes? To answer this some aspects of modelling must be considered.

The simplest two-dimensional element has a triangular shape and three corner nodes. Its three-dimensional equivalent is a tetrahedron with four corner nodes. Both element types are shown in Table 3.2. If it is assumed that a situation

can be modelled as a plane strain problem either element type can be used, but the computational work involved is vastly different. To appreciate the relative magnitude of the computation when solving the simultaneous equations (1.2) for a problem meshed with three-dimensional elements compared to two-dimensional elements, let us consider a simple example. Here, it is assumed that the accuracy of both element types are comparable.

If an adequate analysis of a problem with two-dimensional elements requires a mesh with 20 nodes in each direction, i.e. 400 nodes in total, the total number of simultaneous equations is 800 as there are 2 degrees of freedom per node. Further, the half-bandwidth (see Sec. 3.9) of the matrix $[K]$ in (1.2) is 20 nodes or 40 variables. For an equivalent three-dimensional representation a cube must be used with 20 nodes in each direction, i.e. 8000 nodes in total. Now the total number of simultaneous equations is 24 000 as each node has 3 degrees of freedom. Furthermore, the bandwidth is now 20 nodes squared (400 nodes) or 1200 variables. Given that with the usual solution techniques discussed in Sec. 3.9 the computational effort is roughly proportional to the number of equations and to the square of the bandwidth, the computational effort required to solve the three-dimensional problem is some 27 000 times greater. There are further disadvantages to using a three-dimensional mesh as there are inherent round-off errors which grow as the number of degrees of freedom increases. Hence, it is not surprising that analysts prefer to model the problem with the element type that, for the specified accuracy needed to design, requires the least amount of preparation time and computational resources. Sometimes the only suitable element type is three-dimensional and then the analyst must decide how many degrees of freedom are needed for an acceptable model and compare this to the number that the code and hardware can handle.

It follows from this that, in commercial packages, there are many elements that are one- or two-dimensional. Such elements represent certain three-dimensional forms where simplifications can be made to reduce the dimension of the form. This modelling approach was shown in Sec. 2.3.1 to yield one-dimensional solutions. For a bar member it is found that (2.23), giving the change in length, satisfies the theory of elasticity exactly, whereas in the case of the beam member, although the solution (2.25) is not exact, the secondary deformation neglected in the theory is of little consequence when analysing a structure of slender beams. Similarly, in Sec. 2.3.2 an in-plane loaded structure is modelled as a two-dimensional body by making the assumption that either the out-of-plane direct strain or stress is zero.

3.4 OVERVIEW OF METHODS FOR CALCULATING ELEMENT EQUATIONS

From Eq. (1.1) for any element it can be seen that for each element the stiffness matrix $[k]$ and nodal force vector $\{F^e\}$ will have a different number of terms.

For static analysis the matrix $[k]$ is formulated using (3.9), with $\{F^e\}$ being obtained from (3.10) with the addition of the element concentrated forces from vector $\{R\}$ as shown in (3.11). It can be seen from (3.9) and (3.10) that an element formulation needs definitions of the three basic matrices $[N]$, $[B]$ and $[D]$. As has been seen in Table 3.2, there are numerous classes of element and for each class there are a number of different ways to derive their element characteristics.

There is not scope here to cover the formulation of all types element, and so there will be a concentration on the fundamental features of how element characteristics are derived, placing emphasis on those features that affect accuracy. Sections 3.4–3.8 give sufficient background to enable the reader to refer to, and understand, specialized texts devoted to element formulation.

To derive the matrix $[k]$ and the vector $\{F_q^e\}$ a variety of methods may be used. One method is to apply physical reasoning, and this gave the direct stiffness method before computational methods were available. Unfortunately, it is limited to a few one-dimensional element types and so will not be discussed further.

Another method, known as the *direct approach*, considers the integrals in (3.9) and (3.10) directly. To formulate these integrals this method uses polynomial displacement distributions (known as displacement functions) which are dependent on a local element coordinate system or on a global Cartesian coordinate system. This approach was used by the early finite element method pioneers in the 1950s and 1960s when formulating the first two-dimensional elements, and will be discussed in detail in Sec. 3.6.

In two- and three-dimensional problems elements can have complicated geometries with curved boundaries. In these cases it is very difficult, if not impossible, to define appropriate displacement functions over such shapes to satisfy the requirements necessary for accuracy in the finite element method. To overcome this difficulty a method using isoparametric elements was developed by Taig in 1958 (Robinson, 1985). This is the method used in most commercial software as it can routinely deal with elements having curved sides and/or surfaces (see Sec. 3.7).

Attempts have also been made to formulate the element equations using assumed stress distributions and the *principle of complementary energy* (Washizu, 1982). This method overestimates the strain energy and therefore give an *upper bound* solution with underestimated stresses that are safe for design purposes. However, pure equilibrium elements, as they are known, are difficult to create for all but the simplest element types and it is considered that further development in this field is unlikely within commercial software.

A compromise is the formulation of *stress-hybrid elements* that are based neither on the pure displacement method nor on the pure equilibrium method. Here, it has been found feasible to mix the two methods both inside an element and across interelement boundaries. A number of finite element packages now offer a few hybrid elements and they have been found to perform well. However, as it is more likely that packages have displacement elements, this text concentrates on their formulation and accuracy.

3.5 THE ROLE OF DISPLACEMENT DISTRIBUTIONS

3.5.1 Shape Functions

Formulation of the matrix $[k]$ and the vector $\{F_q^e\}$ is crucially dependent on the local shape functions in (3.5) that are used in interpolating the nodal displacements throughout the element. In fact, there is not a free choice of terms N_i in a shape function matrix $[N]$, as for some choices the method fails. In practice the choice of these functions for a specific element type is further restricted, if they are to satisfy all, or most, of the requirements given in Sec. 3.5.2. It is also clear that, whatever method is used to create them, the shape functions must have unit value at node i and zero value at all other nodes on the element. This ensures that the displacements at each node have the correct value.

An element's displacement distribution is nonzero (except at isolated points) in only the small part of the continuum which forms the element, being zero everywhere else in the structure. It is this property that produces so many zeros in the matrix elements of the global stiffness matrix $[K]$. This matrix is said to be sparse providing assembly has been optimized and this allows relatively efficient solutions of the very large set of simultaneous equations to be found.

Using shape functions leads to an approximation for the displacements over most of the element and it is hoped that this does not differ too much from reality. It can be shown that the strain energy in a structure, found using (3.8), is less than the true strain energy. Moreover, as the number of degrees of freedom is increased, the calculated strain energy approaches the exact value. A solution by the displacement finite element method is therefore referred to as a *lower bound* solution, or a solution in which the structure is found to be *over-stiff*. This over-stiff solution is valid globally but does not apply to every point in the structure. Stresses are overestimated when the numerical method has an over-stiff representation and this is not really what the analyst likes when optimizing a mechanical design.

3.5.2 Requirements to be Met by Displacement Distributions

Certain requirements must be satisfied by any assumed displacement distribution for the finite element method to work, and for many element types this is easily achieved. These requirements are given now since these underpin all element formulations based on the integrals in (3.9) and (3.10).

1. *Number of terms in the polynomial* The number of terms in a polynomial expression selected to represent the unknown displacements must at least equal the number of degrees of freedom associated with the element.
2. *Differentiability* The assumed displacement distribution and its derivatives should be continuous within an element. Since simple polynomial expressions are usually used, no difficulty occurs here and this requirement is inherently satisfied.

3. *Rigid-body modes* The assumed displacement distributions should allow rigid body displacements without invoking strain in an element. To achieve this requirement all polynomial expressions must include a constant term.
4. *Constant strain* The assumed displacement distributions should allow for all states of uniform strain (and stress) in an element. To achieve this requirement the second term in the polynomial expression must be linear. Note that if a sequence of approximate solutions is obtained using more elements for each solution, then these solutions should become closer and closer to any exact solution. For this so-called convergence to occur the strains in an element must approach a constant value. This requirement gives reliable modelling when an element becomes very small. To test that an element meets this requirement, the patch test has been developed; details of its application are given in texts listed in Sec. 3.10.
5. *Compatibility* The chosen displacement distributions should provide internal element compatibility and ideally maintain continuity of displacement (and rotation when bending is modelled) between elements. If all compatibility conditions are satisfied the element is said to be *conforming*. For plate bending and shell elements this requirement gives some difficulty, particularly as it has to be satisfied while also meeting requirement 1.

Convergence of a solution is guaranteed if the *completeness* requirements 2 to 4 are satisfied. The completeness of an element is assured if the polynomial expressions are of high enough degree and if no terms are omitted. Compatibility, requirement 5, must be met in the limit of element refinement and is met with fewer elements by most commonly used elements.

3.6 THE DIRECT APPROACH TO ELEMENT FORMULATION

3.6.1 A Bar Element

Figure 3.1 shows an element for a bar. For this one-dimensional element it will be assumed that it has constant properties along its length L . It is aligned with a local x -axis (and generally oriented with respect to the structure's global coordinate system) and its displacement can be defined in terms of the local axial displacement component u (i.e. parallel to the local x -axis). At each end of the element a node is placed giving it just two degrees of freedom, u_i and u_j , in the vector $\{\delta^e\}$, where subscripts i and j are used to indicate the element's left- and right-hand end nodes respectively. The element need not be given any real physical meaning here, other than representing a subregion of a physical bar.

It is now necessary to choose a mathematical form to represent the displacement over the element $u(x)$. A polynomial is the simplest form for the shape functions and, considering the requirements listed in Sec. 3.5.2, it is apparent that there is only one acceptable polynomial, namely

$$u(x) = u = \alpha_1 + \alpha_2 x = [\mathbf{f}]\{\alpha\} \quad (3.17)$$

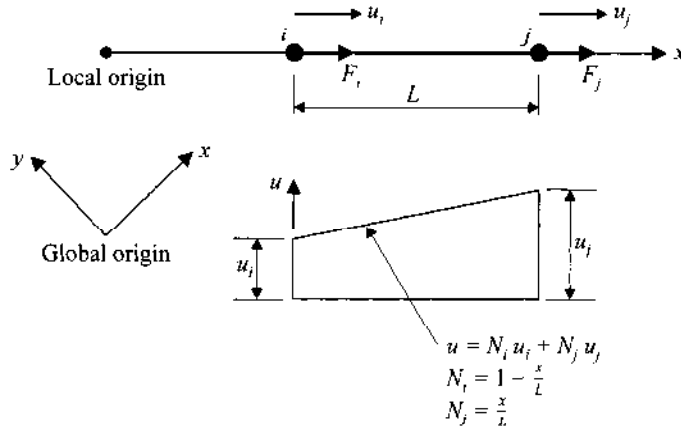


Figure 3.1 A two-noded linear bar element.

in which $\{f\}$ is simply $[1 \ x]$ and the coefficients α_i are known as *generalized coefficients* which are dependent on the nodal displacements and coordinates. Expressions such as (3.17) are often referred to as *displacement functions*. Inserting the nodal boundary conditions of $u(x) = u_i$ at $x = 0$ and $u(x) = u_j$ at $x = L$ into (3.17) gives

$$\begin{Bmatrix} u_i \\ u_j \end{Bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & L \end{bmatrix} \begin{Bmatrix} \alpha_1 \\ \alpha_2 \end{Bmatrix} \quad \text{or} \quad \{\delta^e\} = [A]\{\alpha\} \quad (3.18)$$

in which $[A]$ is the Vandermode matrix. Combining (3.17) and (3.18) the shape function matrix $[N]$ is defined from

$$\{u\} = [f][A]^{-1}\{\delta^e\} = [N]\{\delta^e\} \quad (3.19)$$

Equation (3.17) is therefore $u = u_i + ((u_j - u_i)/L)x$ which, as shown in Fig. 3.1, is a linear variation in the displacement from u_i to u_j . $[N]$ is the first of the three basic matrices on which the matrices $[k]$ and $\{F_q^e\}$ depend and for the bar element is $[1 - x/L \ x/L]$. Note that $N_i = 1$ and $N_j = 0$ when $x = 0$ and that $N_i = 0$ and $N_j = 1$ when $x = L$, as required.

The second basic matrix is $[B]$ and it is derived from (3.7) (a form of (2.9))

$$\{\epsilon\} = [B]\{\delta^e\} = [f'] [A]^{-1} \{\delta^e\} \quad (3.20)$$

in which f' indicates differentiation of terms (here with respect to x). For the bar element this is

$$\epsilon_x = \frac{du}{dx} = \begin{bmatrix} -\frac{1}{L} & \frac{1}{L} \end{bmatrix} \begin{Bmatrix} u_i \\ u_j \end{Bmatrix}$$

in which the matrix $[B]$ is $[-1/L \ 1/L]$.

Finally, the third basic matrix $[\mathbf{D}]$ is derived from the one-dimensional stress-strain relationship (2.13). For the case where there is no initial strain or stress

$$\{\boldsymbol{\sigma}\} = [\mathbf{D}]\{\boldsymbol{\varepsilon}\} \quad \text{or} \quad \sigma_x = E\varepsilon_x \quad (3.21)$$

in which $[\mathbf{D}]$ has only one term E .

To complete the formulation of the matrix $[\mathbf{k}]$ using the integral in (3.9) it is noted that the terms in matrices $[\mathbf{B}]$ and $[\mathbf{D}]$ are constant such that the integration is simply

$$A \int_{x=0}^L dx = AL \quad (3.22)$$

where A is the constant cross-sectional area. Now straightforward mathematical manipulation gives the 2 by 2 element matrix $[\mathbf{k}]$.

A similar approach enables the terms in the element consistent force vector $\{\mathbf{F}_q^e\}$ to be derived using (3.10) when one or more distributed loadings is present. Adding to the consistent nodal forces the concentrated nodal forces gives the element vector $\{\mathbf{F}^e\}$ and thus equation (1.1) ($\{\mathbf{F}^e\} = [\mathbf{k}]\{\boldsymbol{\delta}^e\}$) has the form

$$\begin{Bmatrix} F_i \\ F_j \end{Bmatrix} = \frac{EA}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{Bmatrix} u_i \\ u_j \end{Bmatrix} \quad (3.23)$$

for a bar element with constant properties.

Building a mesh using bar elements, to model a bar with axial loading or a pin-jointed frame with loads concentrated at the joints, is carried out using the methods discussed in Chapter 5. However, as most of the elements are not positioned at the origin of the global axis system in the x -direction, transformations must be used to transform the local element into a global element before the system of equations is assembled. Note that an inclined bar element may have up to six degrees of freedom, as each node can have nonzero u -, v - and w -displacement components. Although a node may move in any global coordinate direction, a bar can carry only axial load, as the pin joint cannot transmit a moment, and thereby equation (2.23) is always valid.

Assembly of the algebraic equations from each element in (3.14) generates the global equations (1.2) (i.e. $\{\mathbf{F}\} = [\mathbf{K}]\{\boldsymbol{\delta}\}$). Methods of solution for the primary unknown nodal displacements ($\{\boldsymbol{\delta}\}$) are given in Sec. 3.9. Nodal displacements can then be extracted from the total set for the problem and, following appropriate transformation to the local axis system, the constant element stress (σ_x) determined by combining (3.20) and (3.21). It is worth remembering throughout the rest of the discussion on element formulation that the methods for assembly and solution of the algebraic equations are independent of the element formulation method, this being one of the most important advantages of the finite element method.

3.6.2 A Three-noded Triangular Element

A second example of the direct approach method illustrates how Turner *et al.* (1956) derived the characteristics for the three-noded triangular element shown in

Figs 1.2 and 1.3. They were interested in analysing aerospace structures consisting of thin flat plates with loads purely in the one plane (the x - y plane) and stress components uniform through the thickness. The element has straight sides and is assumed to have constant thickness. It has two displacement components u and v depending on the spatial global coordinates x and y . Such an element is often referred to in the literature as a plane stress or plane strain element. In order to satisfy requirements 1 to 5 of Sec. 3.5.2, the displacement functions are taken to be the complete linear polynomials

$$\begin{aligned} u &= \alpha_1 + \alpha_2 x + \alpha_3 y \\ v &= \alpha_4 + \alpha_5 x + \alpha_6 y \end{aligned} \quad (3.24)$$

The same three basic matrices, as derived for the bar element Sec. 3.6.1, need to be obtained before the element characteristics can be obtained. A three-noded triangular element has six generalized coefficients in (3.24) and, therefore, six nodal degrees of freedom. To create matrix $[A]$, now 6 by 6, the same procedure as discussed for the bar is used, by inserting into (3.24) the nodal displacements (u_i, v_i, u_j, \dots) and their corresponding global coordinates (x_i, y_i, x_j, \dots). The 3 by 3 $[D]$ matrix, for an isotropic material, is obtained from either the plane stress or plane strain equations (2.29) or (2.30). Matrix $[B]$ is obtained by using (3.24) with the strain-displacement relationships for the two-dimensional strain components $\varepsilon_x, \varepsilon_y$ and γ_{xy} . Matrix $[B]$ has, just as with the bar element, constant terms for an element of constant thickness, t . It follows from this that the integral over the volume will be simply tA (i.e. area \times thickness) and so (3.9) has the simple form $[k] = At[B]^T[D][B]$ in which $[B] = [f'(x, y)][A]^{-1}$. For this element it is not too difficult to derive explicitly the 6 by 6 matrix $[k]$. The element's consistent force vector $\{F_q^e\}$ is derived by following a similar procedure using (3.10).

Solving the simultaneous equations in (1.2) gives, for each element, six nodal displacements. These displacements, when inserted in (3.16) determine the element stresses ($\sigma_x, \sigma_y, \tau_{xy}$), which are constant within the element and usually assigned to the centroid of the element. Alternatively, the stresses can be determined at the nodes by an averaging process of the values in adjoining elements.

One reason why this element has not been accepted as the universal element for irregular geometries is the development of the isoparametric element discussed in Sec. 3.7.1. Chapter 7 shows results for a plane stress example problem to illustrate the modelling ability of linear triangular elements.

3.6.3 Higher-order and Other Elements

The assumed simple displacement distributions for the bar (3.17) and the plane stress or plane strain (3.24) elements give a linear displacement variation and constant strain, and hence stress, in an element. Such an element is known as a *linear, simple or low-order element*. A *refined or higher-order element* is obtained by placing more nodes along the edges of an element and increasing the degree of the polynomial to account for the additional degrees of freedom. In the direct

approach method the element edges are straight and the nodes have to be equally spaced. Such a procedure modifies a three-noded linear triangular element to a six-noded quadratic, nine-noded cubic or even higher-order elements.

The displacement functions for the six-noded triangular quadratic element use complete polynomials with those terms up to and including the quadratic terms. These are

$$\begin{aligned} u &= \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2 \\ v &= \alpha_7 + \alpha_8 x + \alpha_9 y + \alpha_{10} x^2 + \alpha_{11} xy + \alpha_{12} y^2 \end{aligned} \quad (3.25)$$

Here the displacement components vary quadratically and the strain (and stress) vary linearly within the element. It is therefore to be expected that this element gives better results where there are high strain (stress) gradients than does the three-noded element of equivalent size. Again, the three basic matrices $[N]$, $[B]$ and $[D]$ are derived from which the 12 by 12 matrix $[k]$ and the 6 by 1 vector $\{F_q^e\}$ can be formulated.

Displacement functions (3.24) and (3.25) are *complete polynomials* (see requirements 1 to 4 in Sec. 3.5.2), which can be derived using *Pascal's triangle*. The form of the polynomials is such that there is geometric isotropy, i.e. the polynomials are balanced with respect to the x - and y -directions. If each side of the element has the same number of nodes the performance of these elements is independent of their orientation in the mesh and this is the reason why the element formulation is based on the global rather than a local coordinate system. All five requirements listed in Sec. 3.5.2 are satisfied by such displacement functions and convergence is therefore ensured. Hence these elements are conforming, or compatible. Triangular elements therefore have desirable properties and are a popular choice in analyses where the shape of the structure is irregular.

A family of two-dimensional straight-sided rectangular elements can also be formulated by the direct approach. Unlike the family of triangular elements their displacement functions are *incomplete polynomials* and so they do not give geometric isotropy. Because of their rectangular shape these elements are not as versatile as the equivalent triangular elements, as will be shown through examples in Chapters 7–9. They are, however, the preferred choice when modelling a structure with a topologically rectangular shape.

Finally, the matrix $[k]$ and the vector $\{F_q^e\}$ for three-dimensional eight-noded and twenty-noded solid brick elements and the four-noded tetrahedron (Table 3.2) can be derived by the direct approach and these elements have comparable properties to the linear and quadratic plane stress elements. Note that these matrices now have 24 and 60 terms associated with the number of degrees of freedom in these two elements.

From this discussion of the direct approach method it is apparent that the mathematical operations to formulate characteristics for two- and three-dimensional elements become increasingly more complicated as the order of the element (i.e. of the polynomial displacement functions) increases. Moreover, the

inverse of matrix $[A]$ may not exist. Even if it does, then the integration in (3.9) and (3.10) can often be very troublesome, requiring special mathematical procedures.

3.7 ISOPARAMETRIC ELEMENTS

Element types discussed so far have functions said to be of class C^0 , meaning that the displacement components are continuous along interelement boundaries. Slopes, curvatures and higher-order derivatives of the displacements are not. The strain–displacement relationships and the stress–strain relationships, as given in Chapter 2, for C^0 elements show that strains and stresses depend on the first derivative (i.e. the rate of change of displacement components) and thus strains (and stresses) are discontinuous across interelement boundaries.

A second popular class of functions, C^1 , for elements modelling bending deformation as shown in Table 3.2, will be discussed in Sec. 3.8.1. Here, the functions are those that have continuous displacement and slope along their interelement boundaries. These functions are used for beams, flat plate structures and general shell structures; this differs from the plane stress or plane strain simplification in that in-plane direct strains and direct stresses due to bending are assumed to vary linearly through the thickness of the structure.

The most popular and widely available of the C^0 elements, *isoparametric elements*, will be discussed.

3.7.1 Isoparametric Formulation of the Element Stiffness Matrix

There are several features of the isoparametric method that make it different from, and often more desirable than, the direct approach method. In the isoparametric method, a *master*, or parent, set of shape functions is first developed using the direct approach and then the master set is mapped or transformed onto each of the real elements in a mesh. Master elements are available for all of the elements in Tables 3.1 and 3.2. Their shape functions in matrix $[N]$ of (3.5) are defined implicitly rather than explicitly such that there is now a greater reliance on numerical procedures to evaluate any integrals. Closed-form integration is possible in some special cases; but the expressions tend to be lengthy and so calculation is tedious and error-prone. However, the use of numerical integration schemes to integrate (3.9) and (3.10) for an element's characteristics can lead to different problems, and some of these will be presented.

For each element type the master element has a constant size and shape. For example, in two dimensions, triangular elements have the master element shown on the left in Fig. 3.2(a) and quadrilateral elements have the master element shown on the left in Fig. 3.2(b). Note that the master elements are defined in the natural coordinate system ξ and η . For the triangle it can be seen that it is placed in the range 0 to 1, whereas the quadrilateral master ranges from -1 to $+1$. Figure 3.2 also shows two curved-sided real elements, a six-node C^0 -quadratic

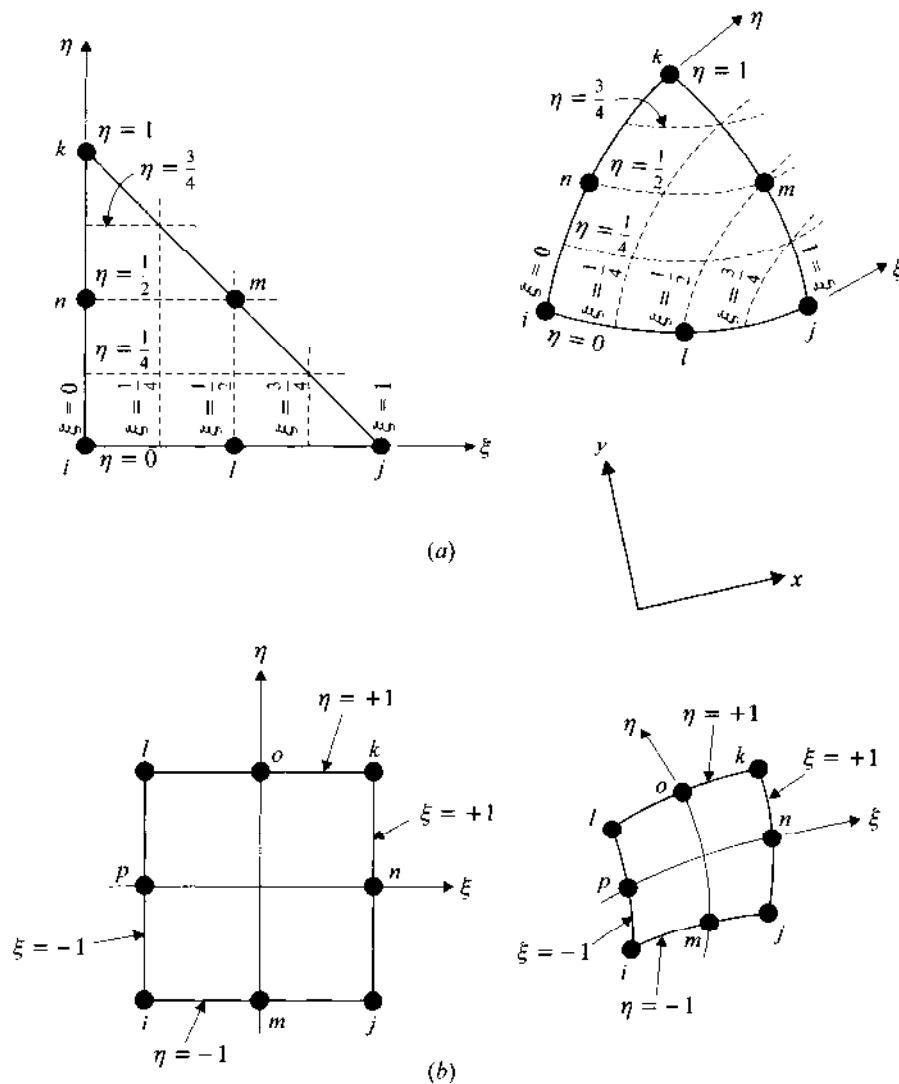


Figure 3.2 Isoparametric elements—master and real forms: (a) for six-noded C^0 quadratic triangular element, (b) for eight-noded C^0 quadratic quadrilateral element.

triangle and an eight-node C^0 -quadratic quadrilateral. When transformed using appropriate shape functions these curved-sided elements in the global coordinate system become the appropriate master element in the natural coordinate system.

Note that 'parametric' refers to the use of the mapping and that 'iso', which means equal, refers to the fact that the mapping functions between global and natural coordinates are chosen to be the same as the shape functions used in (3.5). There are also C^0 sub- and super-parametric, as well as 'nonstandard' hierarchical

elements which are used less frequently. Although C^1 elements for bending are not isoparametric, their element characteristics can be formulated using a master element and a natural coordinate system. Such elements have shape functions, for example, from the *Hermitian* family. They will not be discussed further here.

To explain how the isoparametric method is used to derive the stiffness matrix $[k]$, the example of a bar element with constant properties will be repeated. This linear C^0 element has two nodes at each end. Its master element and a typical real element are shown in Fig. 3.3. The master element is placed in a natural coordinate system ξ such that node i is at $\xi = -1$ and node j is at $\xi = +1$. Hence the length of the master element is 2. Each real element will have some length L (defined as $x_j - x_i$) and any point x along its length is mapped from the master element by the expression

$$x = x(\xi) = \frac{1}{2}(1 - \xi)x_i + \frac{1}{2}(1 + \xi)x_j \quad (3.26)$$

and so it can be seen that $N_i = \frac{1}{2}(1 - \xi)$ and $N_j = \frac{1}{2}(1 + \xi)$ which are the shape functions for the element.

The element displacement distribution can be defined by interpolating nodal displacements with the same shape functions, so

$$u = u(\xi) = \frac{1}{2}(1 - \xi)u_i + \frac{1}{2}(1 + \xi)u_j \quad (3.27)$$

Now (3.27) can be written in the general form of (3.5) as

$$\{u\} = \begin{bmatrix} \frac{1}{2}(1 - \xi) & \frac{1}{2}(1 + \xi) \end{bmatrix} \begin{Bmatrix} u_i \\ u_j \end{Bmatrix} \quad \text{i.e. } \{u\} = [N]\{\delta^e\}$$

The displacement u has the same linear form as that in (3.17), the starting point in the direct approach method, indicating that (3.27) satisfies the five requirements necessary for the element to provide convergence and compatibility given in Sec. 3.5.2. The element is seen to be isoparametric because the same shape function matrix $[N]$ is used for interpolation of both *element geometry* and *element*

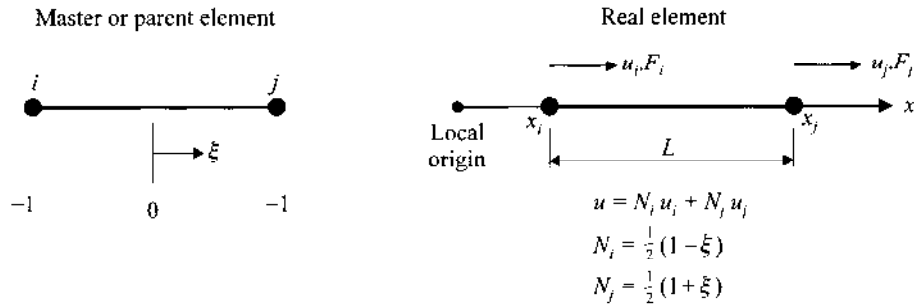


Figure 3.3 A two-noded C^0 linear isoparametric bar element.

displacements, i.e. when evaluating either the x -coordinate or the displacement u at any point along the master bar element the ξ -coordinate of that point is substituted into either (3.26) or (3.27).

Formulation of the stiffness matrix $[\mathbf{k}]$ requires that the matrix $[\mathbf{B}] = [\partial][\mathbf{N}]$ (see (2.10)) be known. Note that with the isoparametric method the basic shape function matrix $[\mathbf{N}]$ has been derived directly and, therefore, the need to derive and invert the Vandermonde matrix $[\mathbf{A}]$ is removed, which is one of the major advantages of the method. The constant axial strain in the bar is, therefore, given by

$$\epsilon_x = \frac{du}{dx} = \left(\frac{d}{dx} [\mathbf{N}] \right) \begin{Bmatrix} u_i \\ u_j \end{Bmatrix} \quad \text{where} \quad \frac{d}{dx} = \frac{d\xi}{dx} \frac{d}{d\xi} \quad (3.28)$$

in which $[\partial]$ is the single term d/dx . To evaluate $d/dx[\mathbf{N}]$ the chain rule must be invoked because $[\mathbf{N}]$ is expressed in terms of ξ rather than in terms of x used to define strain and displacement. The term $d\xi/dx$ is not immediately available and first its *inverse* must be calculated from (3.26). For a one-dimensional element, the Jacobian J is defined as $dx/d\xi$ and so

$$J = \frac{d}{d\xi} [\mathbf{N}] \begin{Bmatrix} x_i \\ x_j \end{Bmatrix} = -\frac{1}{2}x_i + \frac{1}{2}x_j = \frac{L}{2} \quad (3.29)$$

as $x_j - x_i = L$. For the bar element shown in Fig. 3.3, J can be regarded as a scale factor that relates the physical length of the real bar element dx to corresponding length of the master bar element $d\xi$; that is, $dx = J d\xi$.

Equation (3.9) for the bar element can now be written in two alternative forms as

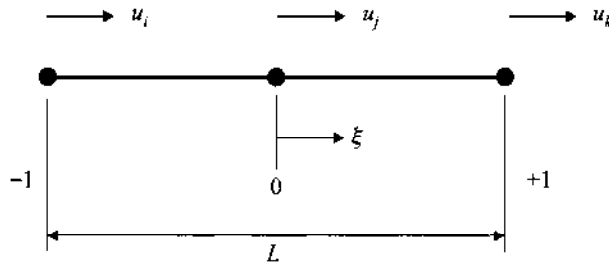
$$[\mathbf{k}] = AE \int_{x_i}^{x_j} [\mathbf{B}]^T [\mathbf{B}] dx = AE \int_{-1}^{+1} [\mathbf{B}]^T [\mathbf{B}] J d\xi \quad (3.30)$$

For the simple bar element under consideration matrix $[\mathbf{B}]$ in (3.30) has the same constant terms in both integral forms as it is independent of both x and ξ . As a result numerical integration is not necessary for this element, a special case, and integration of (3.30) can be achieved explicitly to yield the matrix $[\mathbf{k}]$ in (3.23).

3.7.2 Lagrangian Elements

By placing nodes along the length of the straight bar element, a family of elements can be built up which are known as *Lagrangian elements* after the form of their shape functions. Figure 3.4 shows the next element in the series, the C^0 -isoparametric quadratic bar element, and gives its shape functions. The element displacement, u , is now interpolated from three nodal values and (3.5) can be written as

$$u = \frac{1}{2}\xi(\xi - 1)u_i + (1 - \xi^2)u_j + \frac{1}{2}\xi(\xi + 1)u_k \quad (3.31)$$



Shape functions $N_i = \frac{1}{2} \xi(\xi - 1)$, $N_j = (1 - \xi^2)$, $N_k = \frac{1}{2} \xi(\xi + 1)$

Figure 3.4 A three-noded C^0 quadratic isoparametric bar element.

Note that node j is the interior node and that the shape functions are $N_i = \frac{1}{2} \xi(\xi - 1)$ and so on.

It is not necessary, as is the case with the direct approach method, to place the interior node at mid-length. If the third node is not at the centre, the mapping is nonuniform, i.e. the mapping is distorted, thus in the ξ -coordinate for the real element becomes compressed at one end and stretched at the other. To check that acceptable mapping occurs everywhere in the element it is required that the Jacobian $J > 0$. Note that if J , that is equal to $(L/2 + \xi(x_i - 2x_j + x_k))$, is negative the mapping is folding back on itself, which is clearly unacceptable. Hence, the interior node must always be located within $\frac{1}{4}L$ of the centre for the quadratic element shown in Fig. 3.4, but that in practice the best location is at the centre. If the third node is at a quarter point it produces a mathematical stress singularity and it is this feature of quadratic elements that allows the formulation of a cheap fracture mechanics element. For a cubic element, the standard location for the two interior nodes are at the $\frac{1}{3}$ points and there is a range either side of this in which J remains positive for the one-to-one mapping. It is common practice in commercial software to place interior nodes at locations known to provide the most well-behaved element and so the control of their positions is often outside the influence of the analyst.

3.7.3 Numerical Integration in One Dimension

With the procedure in Sec. 3.7.1, the element stiffness matrix for a quadratic element can be derived from (3.30) using (3.31) for the displacement u . As matrix $[B]$ and J , if the interior node is not at the centre, are now functions of ξ , it is necessary to evaluate the terms of the 3 by 3 matrix $[k]$ by numerical integration.

The process of numerical integration transforms an analytical integral into the sum of a finite number of terms. Note the similarity with the philosophy of the general finite element method where a continuum has been discretized into the

sum of a finite number of elements (or subregions). Hence, for a one-dimensional integral such as that in (3.30), with n sampling points

$$\int_{x_j}^{x_{j+1}} f(x) dx = \int_{-1}^{+1} I(\xi) d\xi \equiv \sum_{l=1}^n w_{nl} I(\xi_{nl}) \quad (3.32)$$

Here, l is the number of the sampling points being considered, w_{nl} are weighted factors and ξ_{nl} are points at which the integrand is evaluated, called *quadrature points*, *integration points*, *sampling points* or *Gauss points*. In the finite element formulation $I(\xi)$ represents the terms, depending on element type, from the matrix product $[\mathbf{B}]^T [\mathbf{D}] [\mathbf{B}] J$.

To evaluate (3.32) there are many well-established quadrature formulae available. Examples of these include the Newton–Cotes rules, such as the well-known Simpson's rule, or the trapezoidal rule in which the sampling points are equally spaced. Experience has shown, however, that Gauss quadrature rules are more precise than these. For n sampling points in (3.32), the Gauss rules integrate exactly a polynomial of degree $2n - 1$. When the limits of the integration are -1 to $+1$ the rule is the *Gauss–Legendre rule*, although, as will be seen shortly, this is not restricted to one dimension.

In Gauss–Legendre quadrature, the positions of the sampling points ξ_{nl} are the zeros of the n th-degree Legendre polynomial and are often known as Gauss points. Both weights w_{nl} and the Gauss points ξ_{nl} have numerical values that are given in Table 3.1 for values of n up to 3. Note that the Gauss points are located symmetrically with respect to the centre of the integration interval ($\xi = 0$) and that symmetrically equivalent points have the same weights, e.g. when n is 2 $w_{22} = w_{21} = 1$. Table 3.3 gives also the order of accuracy as a power of the element length $h \equiv L$.

To approximate integral (3.32) in the simplest way only one sampling point can be used, i.e. $n = 1$. From Table 3.3 this gives $\xi_{11} = 0$ and $w_{11} = 2$ to give (3.32) as

$$\int_{-1}^{+1} I(\xi) d\xi \equiv w_{11} I(\xi_{11}) = 2I(0) \quad (3.33)$$

Table 3.3 Gauss points and associated weights for Gauss–Legendre quadrature

Number of Gauss points, n	Accuracy of quadrature	Gauss points, ξ_{nl}	Weights, w_{nl}
1	$O(h^2)$	$\xi_{11} = 0$	$w_{11} = 2$
2	$O(h^4)$	$\xi_{21} = -\xi_{22} = -0.57735$	$w_{11} = w_{22} = 1$
3	$O(h^6)$	$\xi_{31} = -\xi_{33} = -0.77460$ $\xi_{32} = 0$	$w_{31} = w_{33} = 0.55555$ $w_{32} = 0.88888$

in which $I(0)$ is the value of the integrand at the midpoint and 2 is the length of the interval. Thus the area under a curve $I(\xi)$ is approximated by a rectangular area and this is exact if curve $I(\xi)$ describes a straight line of finite slope. If the integrand is not of this form then a better approximation for the integral is to take $n = 2$ to give the integral as

$$\int_{-1}^{+1} I(\xi) d\xi \equiv w_{21}I(\xi_{12}) + w_{22}I(\xi_{22}) = I(-0.57735) + I(0.57735) \quad (3.34)$$

This two-point rule integrates a cubic polynomial exactly. Note that when deriving the terms in matrix $[k]$ using (3.32) the matrix terms $I(\xi)$ have a constant denominator if the terms in $[B]$ are constant (e.g. for the bar elements in Figs 3.3 and 3.4 (if the third node is at the centre)). Then and only then is numerical integration exact. With matrix $[B]$ depending on $1/J$ this can be true only when J is constant. If, however, the element is distorted from its master shape the terms in $I(\xi)$ are the ratio of two polynomials in ξ and numerical integration is generally inexact, regardless of the rule used. In other words, the ratio of two polynomials is not a polynomial in general. Distortion of an element from its parent shape always introduces discretization errors into the finite element analysis. This aspect of modelling has been taken into account when developing the examples for Chapters 7 to 9.

Clearly, the number of Gauss points can be increased until acceptable accuracy is obtained. However, there will be an optimum rule and therefore the question of the best order of quadrature must be considered. There is no simple answer, although one-dimensional elements are much less complicated to evaluate than two- and three-dimensional elements. According to Burnett (1987) many of the quadrature rules commonly used in commercial codes for a given element type have evolved more from numerical experience than from theoretical arguments, the latter being developed to explain the success or failure of the former.

3.7.4 Serendipity Elements and the Comparison with Lagrangian Elements

It might seem that two-dimensional elements could have a whole host of shapes such as general polygons or conic sections (circles and ellipses). However, of all possible geometries, only three-sided and four-sided polygons (with straight or curved sides) seem to be well suited to application of the isoparametric method. Families of C^0 -isoparametric triangular and quadrilateral elements have been created following similar arguments to those used in the direct approach method. Both these element shapes have a family of elements based on Lagrangian shape functions. There is also a second family of quadrilateral elements which do not have interior nodes and whose shape functions are known as *serendipity shape functions*. Here, 'serendipity' means that the shape functions have been determined by inspection. The very high order Lagrangian

and serendipity elements, say quintic or above, at present are mostly of academic interest, as commercial packages usually allow elements only up to the order of cubic elements.

Shape functions for these element types are given in many texts and so only the C^0 -isoparametric serendipity shape functions for the quadratic eight-noded element in Fig. 3.2(b) are reproduced here. They are

$$\begin{aligned}
 N_i &= \frac{1}{4}(1 - \xi)(1 - \eta)(-\xi - \eta - 1) \\
 N_j &= \frac{1}{4}(1 + \xi)(1 - \eta)(\xi - \eta - 1) \\
 N_k &= \frac{1}{4}(1 + \xi)(1 + \eta)(\xi + \eta - 1) \\
 N_l &= \frac{1}{4}(1 - \xi)(1 + \eta)(-\xi + \eta - 1) \\
 N_m &= \frac{1}{2}(1 - \xi^2)(1 - \eta) \\
 N_n &= \frac{1}{2}(1 + \xi)(1 - \eta^2) \\
 N_o &= \frac{1}{2}(1 - \xi^2)(1 + \eta) \\
 N_p &= \frac{1}{2}(1 - \xi)(1 - \eta^2)
 \end{aligned} \tag{3.35}$$

The majority of software packages provide two- and three-dimensional elements which do not have interior nodes. There is some disagreement between authors on the merits of Lagrangian and serendipity elements and on the usefulness of interior nodes. Burnett (1987) considers Lagrangian quadrilateral elements to have two disadvantages; that the degree of polynomial completeness is low for the order of the element and that the presence of interior nodes can create problems when generating a mesh. Zienkiewicz and Taylor (1989) also state that the Lagrangian family is limited not only because of the large number of internal nodes but also because of their poorer curve-fitting properties. For these reasons they advocate the use of the serendipity family instead. In contrast, Cook *et al.* (1989) state that the Lagrangian shape functions, because they 'do not leave out the middle-part' of Pascal's triangle, allow better accuracy than is found using the equivalent serendipity shape functions. Furthermore, the nine-noded Lagrangian quadrilateral element is much less sensitive than the eight-noded serendipity quadrilateral element to distortion of its geometry and to placing 'midside' nodes away from the centre. When consulting the user manuals for commercial software, the analyst often finds that the method used to formulate the equations for an element is not given and this makes assessment of the code's performance more difficult. Such conflicting guidance in the standard texts illustrates the importance of the activities of NAFEMS and, in particular, their benchmarks.

3.7.5 Numerical Integration in Two Dimensions

Clearly, it must be possible to integrate functions over two and three dimensions. This is a simple extension of the work over one dimension given in Sec. 3.7.3 and so only the two-dimensional version is given here. Over two dimensions, numerical integration is given by

$$\iint f(x, y) dx dy = \int_{-1}^{+1} \int_{-1}^{+1} I(\xi, \eta) d\xi d\eta \equiv \sum_{k=1}^n \sum_{l=1}^n w_{nk} w_{nl} I(\xi_{nk}, \eta_{nl}) \quad (3.36)$$

where n Gauss points are used in each of the two coordinate directions, k and l are the indicators for the sampling point being considered and $I(\xi, \eta)$ is the integrand from the stiffness matrix $[\mathbf{k}]$ which is again obtained from the matrix multiplication $[\mathbf{B}]^T [\mathbf{D}] [\mathbf{B}] J$. Previously it was noted that the Jacobian J for a one-dimensional element is just a single term, $d/d\xi = [\mathbf{J}]d/dx$. In two dimensions J becomes the determinant of the 2 by 2 Jacobian matrix

$$[J(\xi, \eta)] = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} \quad (3.37)$$

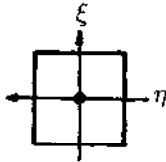
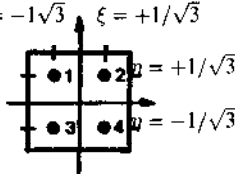
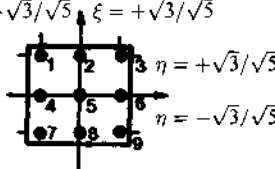
Equation (3.36) can be used with the Gauss-Legendre rule for a quadrilateral element as the integration limits are from -1 to $+1$. Now, the sampling points form a two-dimensional array and the weight applied at each Gauss point can be seen in (3.36) to be a product of the one-dimensional weights given in Table 3.3. Table 3.4 gives values for the Gauss points ξ_{nk} and η_{nl} and the weights $w_{nk} \times w_{nl}$ for values of n up to 3.

Triangular elements, because of their shape, cannot have integration limits of -1 to $+1$ as shown in Fig. 3.2(a) and so the Gauss-Legendre rule is inappropriate. Consequently, another Gaussian integration scheme has to be available to derive the matrix $[\mathbf{k}]$ and the vector $\{\mathbf{F}_q^e\}$ for the Lagrangian family of triangular elements.

3.7.6 Accuracy of Isoparametric Elements

Now a number of the features of a quadrilateral element and numerical integration which relate to the accuracy of calculation of the stiffness matrix $[\mathbf{k}]$ will be discussed. Note that these features are similar when considering a triangular element. As usual, the criterion for an acceptable element is that the mapping is one-to-one, which is achieved as long as J is positive and not zero. To achieve this, two restrictions must be applied to the shape of an element. For an eight-noded quadratic element as shown in Fig. 3.5, the interior corners must have an included angle of less than 180° and the midside nodes must be placed within $\pm L/4$ of the centre position on a side. Note here the similarity to the restrictions on a quadratic bar element discussed previously. It has also been found, through

Table 3.4 Gauss points and associated weights for two-dimensional product-type Gauss-Legendre quadrature

n	Number of Gauss points $n \times n$	Accuracy of quadrature	Gauss points ξ_{nk}, η_{nl}	Weights $w_{nk} \times w_{nl}$
1	1 (1 × 1)	$O(h^2)$		4 (= 2 × 2) at centre
2	4 (2 × 2)	$O(h^4)$		1 (= 1 × 1) at points 1, 2, 3, 4
3	9 (3 × 3)	$O(h^6)$		25/81 (= 5/9 × 5/9) at points 1, 3, 7, 9 40/81 (= 5/9 × 8/9) at points 2, 4, 6, 8 64/81 (= 8/9 × 8/9) at point 5

numerical experiments, that accuracy is seriously impaired when the curvature of a side is so great that the circular arc defined by the three nodes of a side subtends an angle of 180° or greater. This is also shown in Fig. 3.5.

In finite element models the most frequently used element shape has straight sides with the nodes equally spaced along the sides of the element. For example, quadratic elements, in this case, have their midside nodes placed centrally. Also, when the shape of a quadrilateral element is a parallelogram it can be shown that the J is simply one-quarter of the area of the element, that there is no mapping distortion and that numerical integration using (3.36) can be exact. The next most frequently used shape has three straight sides and one curved side. In this case, where there is some distortion of the element, J is nonzero and acceptable, provided that the midside nodes on the curved side remain at their preferred locations or are within the prescribed limits as shown in Fig. 3.5.

Only infrequently does an element have more than one curved side and this is usually at a corner of a structure. The ability of an element's shape to model exactly the shape of the surface of a structure depends on the description of the surface geometry itself. If the surface is flat the lowest-order element to fit exactly is a linear element. However, if the surface has a parabolic shape then a quadratic

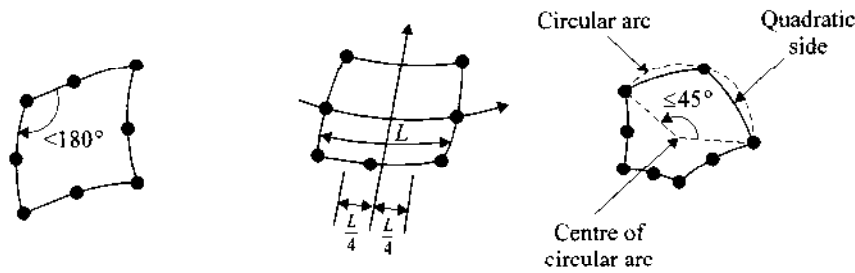


Figure 3.5 Some limitations on C^0 quadratic isoparametric elements for mapping problems.

element is sufficient. In this case, the use of linear elements may provide a poor approximation to the surface geometry unless many such elements are used. In fact, whenever the geometry of a surface is described by a polynomial of higher order than an element's displacement distributions (i.e. the shape functions for isoparametric elements) modelling inaccuracy is incurred. In modern CAD packages the order of a polynomial can be around 20 and so there is often a modelling error when a structure is discretized into elements.

Distortion can also exist if the shape of an element is an extremely narrow parallelogram. The stiffness matrix $[k]$ for such an element is ill-conditioned because two of its interior angles are close to 180° . In addition, the area of the element and hence its J tend to zero. This leads the element to have a much greater stiffness than a less distorted element and, in turn, is a further source of ill-conditioning leading to round-off errors when solving (1.2) by the techniques discussed in Sec. 3.9.

It follows from this discussion that, when creating a mesh, elements that are not too dissimilar from their master shape (Fig. 3.2) are preferred. The following guidelines are often applied:

- The aspect ratio of an element should not be greater than 5 to 1 (or 10 to 1).
- Interior corner angles should be within 20° or 30° of a right angle. If this condition can not be achieved then a triangular element would be better.

In an attempt to maintain solution accuracy, some commercial packages enforce a strict limit on the maximum and minimum allowable values for the interior corner angles. Beyond this limit the solution will not proceed. Further, a second more generous limit is often set beyond which a warning is given to alert the analyst to the fact that element performance in terms of accuracy is being lost. Acceptable element shapes can also be enforced during the automatic creation of an unstructured mesh (see Chapter 5).

Extension of the procedures outlined here to formulate the stiffness matrix $[k]$ for three-dimensional solid brick elements (eight-noded linear and twenty-noded quadratic) does not pose any further fundamental difficulties. Brick elements usually have no interior node and their shape functions belong to the serendipity family. Other shapes of elements that are commonly available include the tetrahedron and wedge.

As alluded to when introducing numerical integration, it is often unclear which quadrature rule should be used when forming the stiffness matrix $[k]$. Now, *full integration* is defined as taking place when a quadrature rule is used that is sufficient to provide the exact integral of all terms if the element is undistorted. Note that the same full integration does not exactly integrate all the terms if the sides are curved, for the reasons already given above. Full integration is the only sure way to avoid pitfalls such as mesh instability, where the elements deform in such a way that the strains at the Gauss points are zero and hence these do not contribute to the strain energy in (3.9). Further details on mesh instabilities (which have nothing to do with structural buckling) and the various methods to control them are given in standard texts.

A lower-order quadrature rule, called *reduced integration*, may be desirable for two reasons:

1. Since the expense of generating matrix $[k]$ by numerical integration is proportional to the number of Gauss points used (i.e. to the value of n), using lower-order quadrature leads to lower computational cost.
2. A lower-order rule tends to soften the stiffness of an element, thus compensating for the over-stiff behaviour inherent with an assumed displacement distribution. Softening comes about because certain of the higher-order polynomial terms vanish at the Gauss points so that they make no contribution to strain energy. This has the additional benefit of reducing the values of stresses that are overestimated by displacement elements.

Experience has shown that for an isoparametric element the best quadrature rule is usually the lowest that computes volume correctly and does not produce mesh instabilities. For example, a Gauss rule with order $n = 2$ (reduced integration) is favoured for eight-noded quadratic quadrilateral elements and for eight-noded linear brick elements. There are, therefore, four and eight Gauss points, respectively. Note that all numerical integration rules are valid only in the limit of the element becoming small and hence a loss in accuracy must be incurred when elements are large.

3.7.7 Isoparametric Formulation of the Element Consistent Force Vector

Throughout the discussion of the formulation of element characteristics there has been a concentration on the element stiffness matrix $[k]$, virtually ignoring the element consistent force vector $\{F_q^e\}$. Most of the features discussed with regard to formulating $[k]$ are true when deriving $\{F_q^e\}$. As explained in Sec. 3.1.3, the term 'consistent' follows from the fact that the distributed loading is transformed into a set of nodal forces that are consistent with the shape functions in matrix $[N]$ of (3.5). It is found that when the sides of an element are straight and any edge nodes are at their preferred locations then the force vector is exact in terms of the external work done by the distributed loading.

Let us now consider an eight-noded isoparametric quadrilateral element, whose shape functions are given in (3.35), having one external edge with a surface

pressure applied to it. To simplify the explanation of how to derive $\{\mathbf{F}_q^e\}$, the element is assumed to be rectangular with sides of length a and b , to have constant thickness, to have the midnodes at the midpoints and to have the loaded edge parallel to the ξ and global x -coordinates (i.e. on the side with nodes i , m and j). The applied pressure is constant and has the value q_y with units of Nmm^{-1} per unit thickness. Body forces, initial stresses and initial strains are assumed to be zero. Figure 3.6(a) shows the situation described.

Recalling expression (3.12) for $\{\mathbf{F}_q^e\}$ and noting that the only nonzero distributed load term in the surface traction vector $\{\bar{\mathbf{X}}\}$ is $\bar{Y}(=q_y)$, it is now

$$\{\mathbf{F}_q^e\} = \int_{-1}^{+1} [\mathbf{N}(\xi, -1)]^T \begin{Bmatrix} 0 \\ q_y \end{Bmatrix} J_T(\xi, -1) d\xi = \sum_{l=1}^l w_{nl} \left([\mathbf{N}(\xi_n, -1)]^T \begin{Bmatrix} 0 \\ q_y \end{Bmatrix} J_T(\xi_n, -1) \right) \quad (3.38)$$

in which $J_T(\xi, \eta)$ is known as the *boundary Jacobian* since it is derived from the ratio of differential arc lengths (Burnett, 1987). For the simple case in Fig. 3.6(a), $J_T(\xi, -1)$ is the constant $a/2$. Equation (3.38) is readily solved either directly or by numerical integration (using the rules in Table 3.3) to yield the following nodal force terms:

$$F_{yi} = \frac{1}{6} a q_y \quad F_{ym} = \frac{4}{6} a q_y \quad F_{yj} = \frac{1}{6} a q_y \quad (3.39)$$

The vector $\{\mathbf{F}_q^e\}$ consists of 16 terms for this element as there are 2 degrees of freedom per node, but all the other 13 terms here are zero. It can be shown that the consistent element forces are those which, if applied in the opposite sense as

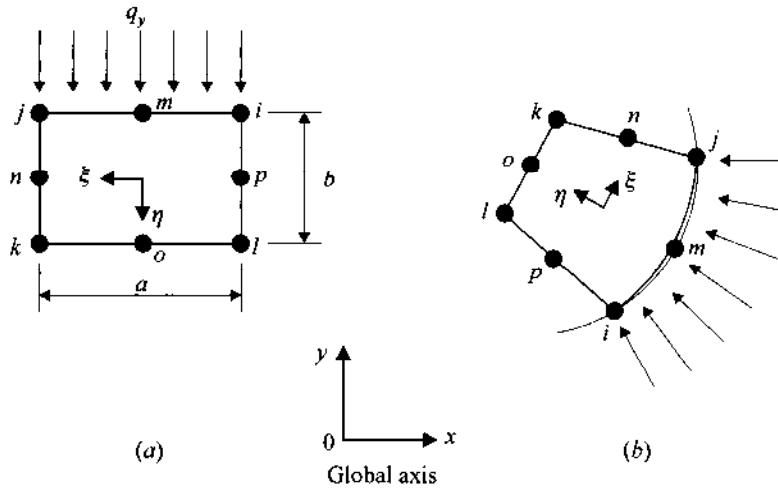


Figure 3.6 Evaluation of element consistent force terms for quadrilateral elements with an edge pressure: (a) on a straight edge, (b) on a curved edge.

constraints, would keep all the nodal displacements zero in the presence of the true loading. Note that by making the nodal forces consistent, the midside node attracts four times the force value of the corner nodes and not twice the value if the total load aq_j is apportioned between the nodes on the basis of a lumped surface area. If the latter method is used it is known as *lumping* or *inconsistent loading* (Cook *et al.*, 1989). The same result is obtained if the loaded side is one of the three straight sides of a six-noded triangular element.

If the third node is moved away from the centre or if the element has a curved side, as shown in Fig. 3.6(b), the boundary Jacobian $J_T(\xi, \eta)$ and \bar{X} and \bar{Y} are no longer constant (Burnett, 1987). These factors and any lack of element fit to the surface geometry of a structure (Fig. 3.6b) provide a different set of nodal forces (with F_{xi} , F_{xm} , F_{xj} also nonzero if the side is curved). Such nodal forces, although consistent with the shape functions of the element, do not provide the exact external work. It can be seen that, as when formulating $[k]$, any distortion to the master element causes a loss in the accuracy of $\{F_q^e\}$ and hence in the finite element solution. A similar procedure is followed to determine the nodal forces due to other distributed loads. Those nodal forces due to body forces, initial stresses and initial strains are derived from volume integrals so that the procedures discussed when deriving $[k]$ are applicable.

The above argument for the development of $\{F_q^e\}$ can be extended to generalized shell and flat plate bending elements with a loading in the form of a face (or surface) distributed pressure.

3.7.8 Calculating Stresses with Isoparametric Elements

Once $[k]$ and $\{F_q^e\}$ for an element have been derived, and then calculated for each element in a mesh, it is not difficult to create the global system of equations, as defined by $\{F\} = [K]\{\delta\}$, Eq. (3.14), that describes the full structural problem. Before solving these simultaneous equations for the nodal displacements $\{\delta\}$ it is necessary to eliminate, by a *condensation technique* described by Cook *et al.* (1989), those equations associated with the interior nodal degrees of freedom of higher-order Lagrangian elements. The equations generated using isoparametric elements which do not have interior nodes require no such treatment and their element characteristics are dealt with in the same way as those formulated by the direct approach method.

Having solved (1.2), the nodal displacements $\{\delta^e\}$ of each element are extracted from $\{\delta\}$ and inserted into (3.16) to evaluate the stresses within an element. If initial values are neglected, then the stresses are given by

$$\{\sigma\} = [D][B]\{\delta^e\}$$

in which the stresses (e.g. σ_x , σ_y and so on) are a function of the global coordinates and the $[B]$ matrix is in terms of natural coordinates. This combination of coordinate systems poses the question of where in the element the stresses should be calculated. For isoparametric elements, stresses (especially shear stresses) are most accurate at the Gauss points of a quadrature rule one order lower than that

for full integration of $[k]$ (Burnett, 1987; Cook *et al.*, 1989). Because stresses are dependent on strains and these are derived from differentials of the element displacements, it may be expected that stresses will be less accurate than the primary unknowns, the nodal displacements, and this is true for elements formulated using the direct approach method. However, under certain conditions the stresses for isoparametric elements are not less accurate than displacements. These stresses are *superaccurate* or *superconvergent* at the Gauss points because they have there the same degree of accuracy as displacements. Nodal stresses are evaluated by interpolating and/or extrapolating the stresses at Gauss points.

Whatever method is used to formulate an element's characteristics, the calculation of stresses is always relative to a rectilinear coordinate system and often the system is Cartesian. As discussed in Sec. 2.2.6 the acceptance of a structural design can be aided by the use of an appropriate failure criterion such as von Mises' criterion. Such criteria, be they stress or strain based, are functions of the principal values (Sec. 2.2.2 and 2.2.4) and these are obtained by transformation from the Cartesian values. It will be seen in Chapters 6–9 how element stresses are manipulated to produce contour plots, this being the most popular method used to display results.

3.8 OTHER TYPES OF ELEMENTS

3.8.1 Elements for Bending

These one- and two-dimensional elements are used to model beams and flat plate structures in bending by assuming that in-plane direct strains vary linearly through the thickness of the structure and that there is no axial strain component. However, bending and in-plane elements can be combined to produce flat shell elements. These are not considered to be the best shell elements.

Now it is found that both displacement and slope components have to be continuous along interelement boundaries, while curvatures and higher-order derivatives are not continuous. To illustrate the features of this class of element the three basic matrices in (3.9) and (3.10) will be derived for a pure bending beam element (i.e. neglecting shear deformation due to through-thickness shearing). This one-dimensional element (see Table 3.2) can have the same form as the bar element in Fig. 3.1 which has two end nodes. In this development it is assumed that the element has constant flexural rigidity EI_u (I_u is the second moment of area about the axis of bending) along its length L . Its centroidal axis is aligned with the local x -axis and the y -axis is vertically downwards. Each end node has two degrees of freedom, one transverse deflection and one slope (or rotation), giving four degrees of freedom (i.e., $v_i, \theta_i = dv/dx_{(x=i)}, v_j$ and $\theta_j = dv/dx_{(x=j)}$), in vector $\{\delta^e\}$.

Enforcement of requirements 1 to 5 in Sec. 3.5.2 defines the displacement function to be

$$v = \alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 \quad (3.40)$$

from which matrix $[A]$ (and $[A]^{-1}$) is derived by inserting into expression (3.40) the nodal coordinates and nodal displacements.

Derivation of matrices $[B]$ and $[D]$ is different to that for class C^0 elements as it is convenient to have the stress-strain relationship (3.1) without initial values, in the form of the well-known *moment-curvature* expression

$$M(x) = -EI_a \frac{d^2 v}{dx^2} \quad \text{i.e. } \{\sigma\} = [D]\{\epsilon\} \quad (3.41)$$

which shows that when bending is being modelled the terms in the strain matrix $\{\epsilon\}$ are *curvatures* and the terms in the stress matrix $\{\sigma\}$ are *bending moments*. Note that the strain is given by $\epsilon_x = \pm y \, d^2 v/dx^2$. Matrix $[D]$ for the beam element has the single term EI_a and matrix $[B]$ is derived from the second derivative of displacement function (3.41) by following the procedure given for the bar element. The exact element stiffness matrix can then be readily derived from (3.9) as the integration can be done explicitly. The form of $[k]$ for the beam element is identical to the *slope-deflection* equations used in the direct stiffness method.

After Eq. (1.2) for a whole beam problem have been solved, the four element nodal displacements are used to evaluate the bending moment anywhere along the element. The direct stress that determines the resistance of the beam is then calculated from the well-known bending relationship (2.25), i.e., $M/I_a = \sigma_x/\pm y$, where the sign convention is that positive moment M causes the beam to sag.

Extension to two-dimensional flat plate bending elements (Fig. 3.1) follows a similar procedure to that discussed for two-dimensional plane stress and plane strain elements. However, it is not practical to have a displacement function for the vertical displacement, $w(x, y)$, that is compatible and this leads to *nonconforming elements*. Some plate elements have been tuned to remedy problems due to shear and locking. In practice these elements are found to work well providing sufficient elements are used to minimize the error due to the lack of interelement compatibility. However, as with the case of shell elements, no outright champion has yet been formulated.

3.8.2 Special Elements

The scope of this text allows detailed discussion of only the basic features of the linear elastic static finite element method. There are many other features which need attention as they represent further evidence of the diversity and versatility of the method. The purpose of this section is to convey a number of these features and to direct the reader to references giving further details. In addition to the references given here, information on these features, when they are available, can be found in the user manuals of commercial analysis software.

With reference to Tables 3.1 and 3.2 it is found that a number of element types have not yet been dealt with. Of these types there are the one-dimensional elements to model a rigid link, a linear spring, a beam with offset, a tapered beam and a Mindlin shear deformable beam (i.e. for deep beams).

Axisymmetric solid and axisymmetric shell element types (both thick and thin) are often used in problem-solving and a major application for these elements is in the analysis of pressure vessels and tanks. Axisymmetric elements are for problems concerning the stress distribution in structures of revolution under axisymmetric loading. By symmetry, an element for the solid of revolution is two-dimensional and the formulation of its element characteristics is analogous to that for plane stress and plane strain elements.

Shell elements can be in two or three dimensions. They are used to model structures with curved surfaces in space that have loading acting normal to the surface. The properties of a shell are lumped to act at the midsurface of the shell to develop the mathematical theory from which element characteristics are derived in the usual way. Often a shell is thin in comparison with its span and then classical thin shell theory is applicable. However, when it is necessary to include in the analysis the effects of transverse shear deformation and perhaps of through-thickness direct stress, the theory is that of a thick shell. Note that the Mindlin plate theory allows transverse displacement and rotations to be independent such that the element formulation only requires C^0 continuity. This has a major advantage of simplifying the shape functions in (3.5) needed to formulate the element characteristics. General shell elements display bending and membrane (in-plane) deformations, with the latter dominating deformation if the structure is to be effective. It is readily seen that flat plate elements (in-plane and bending) are in fact degenerate forms of a generalized thin shell element. Some packages therefore do not provide separate flat plate elements and expect the analyst to use the available shell elements to model such cases.

Two-dimensional classical shell theory produces equations that are difficult to solve. The equations in terms of displacements are often complicated unless many approximations are made (Calladine, 1983). There is no consensus as to which approximations are acceptable, so it is found that there a number of shell theories (e.g. Donnell (1933), Flügge (1973) and Vlasov (1964)). All these theories are limited to small displacements. To formulate element characteristics it is helpful if displacement distributions are chosen, Eq. (3.5), that will satisfy all five requirements presented in Sec. 3.5.2—for shell elements this is difficult, if not impossible. Elements for shells are therefore the most difficult elements to derive, and because of the various shell theories and approaches to the problem there are a number of elements available. Problems with three- and four-sided elements based on thin shell analysis have led to a move away from classical shell formulation and towards treating shell elements as a special version of the three-dimensional solid element, when the through-thickness dimension becomes relatively small. These quadratic isoparametric elements that simulate behaviour of the Mindlin shell element are not without their problems. For example, they do not pass the patch test and so one of the functions of NAFEMS is to publicize the performance of shell elements either singly or in groups. They are, however, favoured by the analysis community.

Elements described so far have sides with the same number of nodes. Packages do offer so-called *transition elements* where, for example, one side of a quadratic quadrilateral element does not have a third node and therefore, along

that edge, behaves as a linear element. This is achieved by forcing the midside node to lie on a straight line between corner nodes using constraint equations so that no gap can open up and continuity is maintained. Transition elements are used in mesh construction when the element order changes. A situation when this may be desirable is where, because strain (and stress) gradients are high, regions in a structure need quadratic elements for an accurate solution, but where in the adjacent regions linear elements are acceptable as the gradients there are much lower. By integrating into a mesh both linear and quadratic elements the numbers of degrees of freedom, and so the round-off errors and the computational cost of solving the problem, can be minimized for a given accuracy.

Special element types exist to model cracks in a fracture mechanics analysis and to model an infinite media such as the ground supporting a structure. Other specialized elements are available which do not represent a continuum but which have their place when modelling practical problems. Such elements include the lumped mass, gap or interface elements, damper elements, spring elements of zero length and constraint elements.

Finite element packages offer a range of material properties for an element type. Here, for convenience, isotropic materials (Sec. 2.2.6) have been chosen to show how the material's elastic constants (i.e., E and ν) in matrix $[D]$ are used in the formulation of an element stiffness matrix terms and in the calculation of the element stresses. Some packages allow all 36 C_{ij} terms (see Eq. (2.15)) in $[D]$ to be independently defined such that any anisotropic material can be described. A majority of packages are less flexible but, because of their importance in structural applications, have the facility to model laminated materials. Laminates often have a sheet structure and consist of multi-layers of a fibre-reinforced material (Halpin, 1992), and it is therefore plate and shell elements (thin and thick) that have this material option. For a general laminate, the properties attached to a general shell element are formulated using lamination theory, details of which are to be found in texts on fibre-reinforced composite materials.

When developing the finite element method equations in Sec. 3.2.2 it was stated that initial strains and stresses can be generated because of the thermal expansion property of a material. It is often the situation that a load-bearing structure has residual stresses at room temperature after cooling down from the processing temperature, e.g. hot rolled steel, or has to operate at a temperature other than room temperature, and for the finite element analysis to be accurate it is imperative to account for any thermal effects. Finite element packages therefore offer a range of thermal elements to solve heat transfer problems. To determine the overall loading for a structural analysis the results from such an analysis are used to establish the thermally generated loadings that are added to the other loading types (see (3.10) and (3.12)).

In Chapter 10, both buckling and dynamic analysis will be discussed when they are performed using the linear small-displacement elements developed in this section. This will be followed by the introduction of the finite element method in analysing problems that are time-dependent and nonlinear in terms of either geometry or their material properties.

3.9 PRODUCING A SOLUTION FROM THE ELEMENT EQUATIONS

In the previous sections of this chapter the ways in which numerical analogues of the governing partial differential equations can be produced have been shown. For all types of elements, equations are produced for each element which relate deformation of the whole element to the discrete displacements and forces at the nodes of the element. These equations know nothing of the behaviour of the full structure and so, now, there must be a discussion as to how these equations are combined, a process known as assembly, and then processed such that the boundary conditions of the problem are applied before a static solution for the full structure is found.

3.9.1 Assembly into a Global Matrix

On each element of the structure, an equation or equations can be developed to describe the behaviour over the element. For example:

$$[\mathbf{k}]\{\delta^e\} = \{\mathbf{F}^e\}$$

which is (1.1) again. Here, $[\mathbf{k}]$ is the stiffness matrix for the element, $\{\delta^e\}$ is the vector of nodal displacements and $\{\mathbf{F}^e\}$ is the vector of forces at the nodes. Such equations are always singular and so they have to be combined together to form a global set of matrix equations and then the boundary conditions for the whole structure can be applied to remove any singularities. This works only when the structure has no mechanisms.

Figure 3.7 shows a simple mesh of two-dimensional plane stress triangular elements with two degrees of freedom per node. The mesh consists of three triangular elements and five nodes in total. Element equations can be written for each element in the form

$$\begin{bmatrix} a_{11}^e & a_{12}^e & a_{13}^e & a_{14}^e & a_{15}^e & a_{16}^e \\ a_{21}^e & a_{22}^e & a_{23}^e & a_{24}^e & a_{25}^e & a_{26}^e \\ a_{31}^e & a_{32}^e & a_{33}^e & a_{34}^e & a_{35}^e & a_{36}^e \\ a_{41}^e & a_{42}^e & a_{43}^e & a_{44}^e & a_{45}^e & a_{46}^e \\ a_{51}^e & a_{52}^e & a_{53}^e & a_{54}^e & a_{55}^e & a_{56}^e \\ a_{61}^e & a_{62}^e & a_{63}^e & a_{64}^e & a_{65}^e & a_{66}^e \end{bmatrix} \begin{bmatrix} u_i^e \\ v_i^e \\ u_j^e \\ v_j^e \\ u_k^e \\ v_k^e \end{bmatrix} = \begin{bmatrix} F_{ix}^e \\ F_{iy}^e \\ F_{jx}^e \\ F_{jy}^e \\ F_{kx}^e \\ F_{ky}^e \end{bmatrix} \quad (3.42)$$

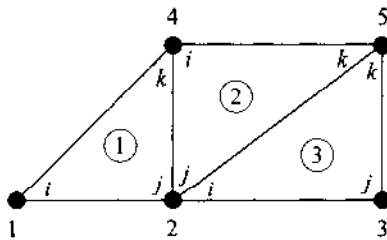


Figure 3.7 A simple mesh of triangular elements.

where the superscript e refers to the number of the element and the other notation is the same as that used before and in Fig. 3.7.

These element equations can be modified, for each element, to relate to the 10 degrees of freedom global problem by using the global node numbers shown in Fig. 3.7. For example, local node k on element 1 is global node 4, local node k on element 2 is global node 5 and so on. These expanded element equations are:

$$\begin{aligned}
 & \begin{bmatrix} a_{11}^1 & a_{12}^1 & a_{13}^1 & a_{14}^1 & 0 & 0 & a_{17}^1 & a_{18}^1 & 0 & 0 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 & a_{24}^1 & 0 & 0 & a_{27}^1 & a_{28}^1 & 0 & 0 \\ a_{31}^1 & a_{32}^1 & a_{33}^1 & a_{34}^1 & 0 & 0 & a_{37}^1 & a_{38}^1 & 0 & 0 \\ a_{41}^1 & a_{42}^1 & a_{43}^1 & a_{44}^1 & 0 & 0 & a_{47}^1 & a_{48}^1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{71}^1 & a_{72}^1 & a_{73}^1 & a_{74}^1 & 0 & 0 & a_{77}^1 & a_{78}^1 & 0 & 0 \\ a_{81}^1 & a_{82}^1 & a_{83}^1 & a_{84}^1 & 0 & 0 & a_{87}^1 & a_{88}^1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \\ u_5 \\ v_5 \end{bmatrix} = \begin{bmatrix} F_{1x}^1 \\ F_{1y}^1 \\ F_{2x}^1 \\ F_{2y}^1 \\ 0 \\ 0 \\ F_{4x}^1 \\ F_{4y}^1 \\ 0 \\ 0 \end{bmatrix} \\
 & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33}^2 & a_{34}^2 & 0 & 0 & a_{37}^2 & a_{38}^2 & a_{39}^2 & a_{310}^2 \\ 0 & 0 & a_{43}^2 & a_{44}^2 & 0 & 0 & a_{47}^2 & a_{48}^2 & a_{49}^2 & a_{410}^2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{73}^2 & a_{74}^2 & 0 & 0 & a_{77}^2 & a_{78}^2 & a_{79}^2 & a_{710}^2 \\ 0 & 0 & a_{83}^2 & a_{84}^2 & 0 & 0 & a_{87}^2 & a_{88}^2 & a_{89}^2 & a_{810}^2 \\ 0 & 0 & a_{93}^2 & a_{94}^2 & 0 & 0 & a_{97}^2 & a_{98}^2 & a_{99}^2 & a_{910}^2 \\ 0 & 0 & a_{103}^2 & a_{104}^2 & 0 & 0 & a_{107}^2 & a_{108}^2 & a_{109}^2 & a_{1010}^2 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \\ u_5 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ F_{2x}^2 \\ F_{2y}^2 \\ 0 \\ 0 \\ F_{4x}^2 \\ F_{4y}^2 \\ F_{5x}^2 \\ F_{5y}^2 \end{bmatrix} \quad (3.43) \\
 & \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{33}^3 & a_{34}^3 & a_{35}^3 & a_{36}^3 & 0 & 0 & a_{39}^3 & a_{310}^3 \\ 0 & 0 & a_{43}^3 & a_{44}^3 & a_{45}^3 & a_{46}^3 & 0 & 0 & a_{49}^3 & a_{410}^3 \\ 0 & 0 & a_{53}^3 & a_{54}^3 & a_{55}^3 & a_{56}^3 & 0 & 0 & a_{59}^3 & a_{510}^3 \\ 0 & 0 & a_{63}^3 & a_{64}^3 & a_{65}^3 & a_{66}^3 & 0 & 0 & a_{69}^3 & a_{610}^3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{93}^3 & a_{94}^3 & a_{95}^3 & a_{96}^3 & 0 & 0 & a_{99}^3 & a_{910}^3 \\ 0 & 0 & a_{103}^3 & a_{104}^3 & a_{105}^3 & a_{106}^3 & 0 & 0 & a_{109}^3 & a_{1010}^3 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \\ u_5 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ F_{2x}^3 \\ F_{2y}^3 \\ F_{3x}^3 \\ F_{3y}^3 \\ 0 \\ 0 \\ F_{5x}^3 \\ F_{5y}^3 \end{bmatrix}
 \end{aligned}$$

Note that now the degrees of freedom are u and v at each of the five nodes and that each element has zeros in both the left-hand side matrix and the right-hand side vector where the position corresponds to a global node that is not attached to

the element in question. These equations are then summed to give the final equation, which is

$$\begin{bmatrix}
 a_{11}^1 & a_{12}^1 & a_{13}^1 & a_{14}^1 & 0 & 0 & a_{17}^1 & a_{18}^1 & 0 & 0 \\
 a_{21}^1 & a_{22}^1 & a_{23}^1 & a_{24}^1 & 0 & 0 & a_{27}^1 & a_{28}^1 & 0 & 0 \\
 a_{31}^1 & a_{32}^1 & a_{33}^1 + a_{33}^2 & a_{34}^1 + a_{34}^2 & a_{35}^1 & a_{36}^1 & a_{37}^1 + a_{37}^2 & a_{38}^1 + a_{38}^2 & a_{39}^1 + a_{39}^2 & a_{310}^1 + a_{310}^2 \\
 a_{41}^1 & a_{42}^1 & a_{43}^1 + a_{43}^2 & a_{44}^1 + a_{44}^2 & a_{45}^1 & a_{46}^1 & a_{47}^1 + a_{47}^2 & a_{48}^1 + a_{48}^2 & a_{49}^1 + a_{49}^2 & a_{410}^1 + a_{410}^2 \\
 0 & 0 & a_{53}^3 & a_{54}^3 & a_{55}^3 & a_{56}^3 & 0 & 0 & a_{59}^3 & a_{510}^3 \\
 0 & 0 & a_{63}^3 & a_{64}^3 & a_{65}^3 & a_{66}^3 & 0 & 0 & a_{69}^3 & a_{610}^3 \\
 a_{71}^1 & a_{72}^1 & a_{73}^1 + a_{73}^2 & a_{74}^1 + a_{74}^2 & 0 & 0 & a_{77}^1 + a_{77}^2 & a_{78}^1 + a_{78}^2 & a_{79}^1 & a_{710}^1 \\
 a_{81}^1 & a_{82}^1 & a_{83}^1 + a_{83}^2 & a_{84}^1 + a_{84}^2 & 0 & 0 & a_{87}^1 + a_{87}^2 & a_{88}^1 + a_{88}^2 & a_{89}^1 & a_{810}^1 \\
 0 & 0 & a_{93}^2 + a_{93}^3 & a_{94}^2 + a_{94}^3 & a_{95}^2 & a_{96}^2 & a_{97}^2 & a_{98}^2 & a_{99}^2 + a_{99}^3 & a_{910}^2 + a_{910}^3 \\
 0 & 0 & a_{103}^2 + a_{103}^3 & a_{104}^2 + a_{104}^3 & a_{105}^2 & a_{106}^2 & a_{107}^2 & a_{108}^2 & a_{109}^2 + a_{109}^3 & a_{1010}^2 + a_{1010}^3
 \end{bmatrix}
 \begin{bmatrix}
 u_1 \\
 v_1 \\
 u_2 \\
 v_2 \\
 u_3 \\
 v_3 \\
 u_4 \\
 v_4 \\
 u_5 \\
 v_5
 \end{bmatrix}
 =
 \begin{bmatrix}
 F_{1x} \\
 F_{1y} \\
 F_{2x} \\
 F_{2y} \\
 F_{3x} \\
 F_{3y} \\
 F_{4x} \\
 F_{4y} \\
 F_{5x} \\
 F_{5y}
 \end{bmatrix}
 \quad (3.44)$$

To carry out this process all that is required is for the element assembly software to know, for each element, the global node number of each of the nodes attached to it. This is the connectivity list. Simple algorithms can then be used to loop over each element in the mesh, placing the elements of the local element equations into the global left-hand side matrix and the global right-hand side vector. This gives the governing equations in the form of (1.2). Note that (3.44) is sparse even for this simple example. The zeros in the global stiffness matrix come about naturally during the assembly process, confirming the fact that certain nodes have no direct effect on other nodes. For example, global node 3 has no effect on global nodes 1 and 4.

3.9.2 Applying Boundary Conditions

At this stage in the process, having completed the formation of the global matrices, the boundary conditions of the problem can be applied. Typically this involves the specification of applied nodal forces and displacement restraints. Figure 3.8 shows the same mesh as Fig. 3.7 with the addition of boundary conditions. Here, it can be seen that at nodes 1 and 2 the v -displacement is zero, that at node 3 both the u - and

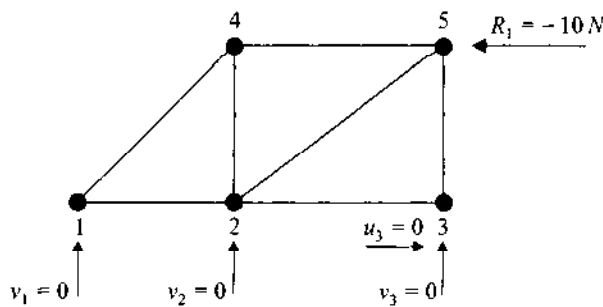


Figure 3.8 Boundary conditions for the simple mesh.

v -displacements are zero and that at node 5 there is an applied concentrated force of 10 N in the negative x -direction. To impose the displacement constraints v_1 , v_2 , v_3 and u_3 must be set to zero. This is forced on the matrix system by setting all values of the row and column associated with these degrees of freedom in the left-hand matrix to zero except for the diagonal which is set to one. Then the right-hand vector has to be modified. To do this (3.11) is used, where the element right-hand vector is made up of the contribution due to the applied external loads and any internal forces as defined by (3.10). In this case all the internal forces are zero and the only applied load is at the fifth node in the negative x -direction. Hence the right-hand vector becomes all zero except for the applied load, and then the appropriate values for the restraints, which are zero in this case, multiplied by the column values that have been zeroed from the left-hand side matrix are subtracted.

This gives

$$\begin{bmatrix} a_{11} & 0 & a_{13} & 0 & 0 & 0 & a_{17} & a_{18} & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ a_{31} & 0 & a_{33} & 0 & 0 & 0 & a_{37} & a_{38} & a_{39} & a_{310} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ a_{71} & 0 & a_{73} & 0 & 0 & 0 & a_{77} & a_{78} & a_{79} & a_{710} \\ a_{81} & 0 & a_{83} & 0 & 0 & 0 & a_{87} & a_{88} & a_{89} & a_{810} \\ 0 & 0 & a_{93} & 0 & 0 & 0 & a_{97} & a_{98} & a_{99} & a_{910} \\ 0 & 0 & a_{103} & 0 & 0 & 0 & a_{107} & a_{108} & a_{109} & a_{1010} \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ u_3 \\ v_3 \\ u_4 \\ v_4 \\ u_5 \\ v_5 \end{bmatrix} = \begin{bmatrix} 0 - a_{17} \cdot 0 - a_{18} \cdot 0 \\ 0 \\ 0 - a_{32} \cdot 0 - a_{34} \cdot 0 - a_{35} \cdot 0 - a_{36} \cdot 0 \\ 0 \\ 0 \\ 0 \\ 0 - a_{72} \cdot 0 - a_{74} \cdot 0 \\ 0 - a_{82} \cdot 0 - a_{84} \cdot 0 \\ R - a_{94} \cdot 0 - a_{95} \cdot 0 - a_{96} \cdot 0 \\ 0 - a_{104} \cdot 0 - a_{105} \cdot 0 - a_{106} \cdot 0 \end{bmatrix} \quad (3.45)$$

As an aside, note that this process has mimicked the mathematical imposition of boundary conditions in that where the displacement is restrained the effect of any applied loads is overwritten.

3.9.3 Solving the Simultaneous Equations

The global equations must now be solved. For linear static problems without any time variation, this means solving just one set of linear simultaneous equations. Note that this is not the situation if gap elements are present as an iterative procedure is necessary. There are many ways of solving the equations, and each solver will have its own way, or ways, of finding a solution from the equations. For real structural problems the number of degrees of freedom is typically of the order of thousands or tens of thousands and so robust direct methods are used to find solutions. However, as problems become more complex so the computational effort needed to find a solution starts to rise as a proportion of the time taken to set up the element equations. When the solution of the equations consumes a large amount of computational effort, there are great benefits to be gained from using iterative methods to solve the simultaneous equations.

Solving any set of simultaneous equations is the process of finding a vector $\{\mathbf{x}\}$ that satisfies the general matrix equation

$$[\mathbf{A}]\{\mathbf{x}\} = \{\mathbf{b}\} \quad (3.46)$$

For finite element problems, (3.46) is directly equivalent to (1.2). To find a solution, the inverse of the matrix $[A]$ must be determined and then both sides of (3.46) are premultiplied by this inverse to give

$$\{x\} = [A]^{-1}\{b\} \quad (3.47)$$

When the number of equations is relatively small then the inverse of the matrix $[A]$ can be found using *direct methods*. Typically, a version of the method called LU decomposition is used. However, to reduce the requirements on computer memory the frontal method as described by Lewis and Ward (1991) and Taylor and Hughes (1981) has been developed. This combines the process of assembly and solution, handling the matrices element after element. It will not be discussed further here.

In LU decomposition the matrix $[A]$ is decomposed into two other matrices:

$$[A] = [L][U] \quad (3.48)$$

where $[L]$ is a lower triangular matrix and $[U]$ is an upper triangular matrix. Once the matrix $[A]$ has been decomposed into $[L]$ and $[U]$ the solution is easy to find. This is because the simple triangular structure makes it easy to find the inverses of these matrices. Depending on the symmetry properties of the matrix $[A]$, the method has a number of variants, such as the Crout method, for nonsymmetric matrices and the Cholesky method for symmetric matrices.

As a rule of thumb, the time taken for solution using LU decomposition is proportional to the cube of the number of unknowns for a fully populated matrix, or to the product of the number of unknowns and the half-bandwidth squared for a sparse matrix. In some texts the terms 'half-bandwidth' and 'bandwidth' are synonymous. As was seen for the simple problem shown in Fig. 3.7, some of the terms in the matrix are zero because of the structure of the mesh. For example, node 1 has no direct connection to node 3 or node 5 and the matrix structure reflects this. For meshes of realistic problems many values in the matrix are zero, and the half-bandwidth is the maximum distance between the diagonal and the first nonzero element of a column in the upper triangle of the left-hand-side matrix. As the matrices are usually symmetrical, and even nonsymmetric ones have a symmetrical structure, this is the same as the maximum distance between the diagonal and the first nonzero element of a row in the lower triangle. Figure 3.9 shows this for a typical matrix.

Owing to the nature of the solution effort, if a matrix is very large then direct methods take a long time to produce a solution. In cases of complex structural analysis with nonlinearity and time variance, such large matrices are often produced. To reduce the time taken *iterative methods* are usually used. With these methods some guess at the solution vector $\{x\}$ is made and then updated using the vector $\{x\}$ and the coefficients of the matrix $[A]$ and vector $\{b\}$. In fact, many iterative schemes can be used, so to illustrate the use of these methods let us consider the solution of (3.46). For a system of three equations:

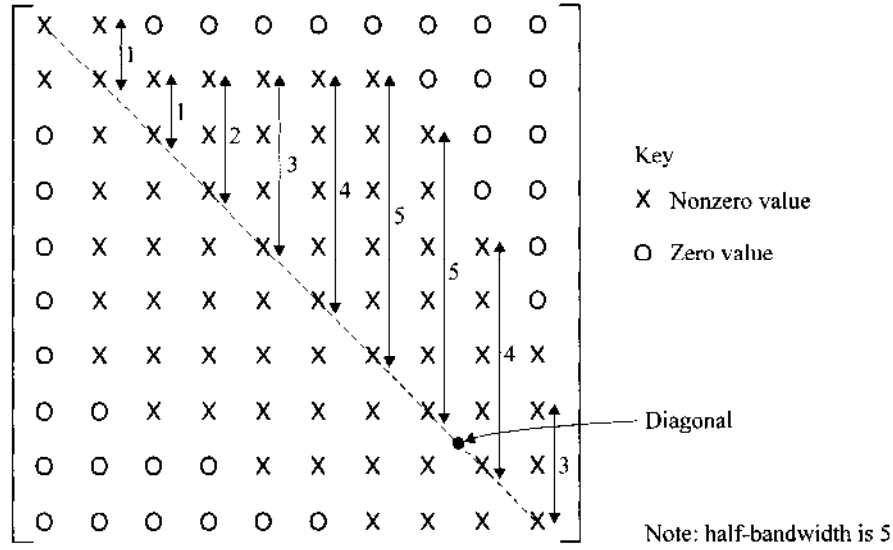


Figure 3.9 The half-bandwidth of a sparse matrix.

$$\begin{aligned}
 a_{11}x_1 + a_{12}x_2 + a_{13}x_3 &= b_1 \\
 a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \\
 a_{31}x_1 + a_{32}x_2 + a_{33}x_3 &= b_3
 \end{aligned}
 \tag{3.49}$$

The simplest iterative schemes are those due to Jacobi and Gauss-Seidel. In both of these methods the equations are transformed to give

$$\begin{aligned}
 x_1 &= \frac{1}{a_{11}}(b_1 - a_{12}x_2 - a_{13}x_3) \\
 x_2 &= \frac{1}{a_{22}}(b_2 - a_{21}x_1 - a_{23}x_3) \\
 x_3 &= \frac{1}{a_{33}}(b_3 - a_{31}x_1 - a_{32}x_2)
 \end{aligned}
 \tag{3.50}$$

Now it can be seen that the terms on the diagonal of $[A]$, i.e. a_{ii} , must not be zero for these methods to work. Fortunately, zero values are not generated in static analysis using the finite element method. In the Jacobi method, the right-hand side of (3.50) is taken to be the known values at the n th iteration and the left-hand side to be the new values at the $n + 1$ th iteration. This gives

$$\begin{aligned}
 x_1^{n+1} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^n - a_{13}x_3^n) \\
 x_2^{n+1} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^n - a_{23}x_3^n) \\
 x_3^{n+1} &= \frac{1}{a_{33}}(b_3 - a_{31}x_1^n - a_{32}x_2^n)
 \end{aligned}
 \tag{3.51}$$

whereas the Gauss-Seidel method continually updates the right-hand side where it can to give

$$\begin{aligned}x_1^{n+1} &= \frac{1}{a_{11}}(b_1 - a_{12}x_2^n - a_{13}x_3^n) \\x_2^{n+1} &= \frac{1}{a_{22}}(b_2 - a_{21}x_1^{n+1} - a_{23}x_3^n) \\x_3^{n+1} &= \frac{1}{a_{33}}(b_3 - a_{31}x_1^{n+1} - a_{32}x_2^{n+1})\end{aligned}\quad (3.52)$$

The next level of complexity involves point relaxation methods. These methods operate on the error in the solution vector $\{\mathbf{x}\}$ which is known as the *residual error* and is defined as

$$\{\mathbf{r}\} = \{\mathbf{b}\} - [\mathbf{A}]\{\mathbf{x}\} \quad (3.53)$$

This should become smaller from iteration to iteration. To modify the previous method (3.52) is taken and x_i^n is both added to and subtracted from the right-hand side, to give

$$\begin{aligned}x_1^{n+1} &= x_1^n + \left[\frac{1}{a_{11}}(b_1 - a_{11}x_1^n - a_{12}x_2^n - a_{13}x_3^n) \right] \\x_2^{n+1} &= x_2^n + \left[\frac{1}{a_{22}}(b_2 - a_{21}x_1^{n+1} - a_{22}x_2^n - a_{23}x_3^n) \right] \\x_3^{n+1} &= x_3^n + \left[\frac{1}{a_{33}}(b_3 - a_{31}x_1^{n+1} - a_{32}x_2^{n+1} - a_{33}x_3^n) \right]\end{aligned}\quad (3.54)$$

Now the expressions in square brackets are the terms of the residual $\{\mathbf{r}\}$, and as these should tend to zero as the process proceeds, attempts can be made to accelerate the process by multiplying the right-hand side by a *relaxation factor*, ω , to give

$$\begin{aligned}x_1^{n+1} &= x_1^n + \left[\frac{\omega}{a_{11}}(b_1 - a_{11}x_1^n - a_{12}x_2^n - a_{13}x_3^n) \right] \\x_2^{n+1} &= x_2^n + \left[\frac{\omega}{a_{22}}(b_2 - a_{21}x_1^{n+1} - a_{22}x_2^n - a_{23}x_3^n) \right] \\x_3^{n+1} &= x_3^n + \left[\frac{\omega}{a_{33}}(b_3 - a_{31}x_1^{n+1} - a_{32}x_2^{n+1} - a_{33}x_3^n) \right]\end{aligned}\quad (3.55)$$

For most systems of equations the value of ω can be set to a value somewhere between 1 and 2. In this case the method is the *successive overrelaxation method*. If ω is unity then the method becomes the original Gauss-Seidel method.

Line relaxation methods where subsystems of the equations are used to provide an update to a number of values at one time can also be used. Also, more advanced methods are continuously being developed, as further research is carried out. Such research is needed to reduce the computational effort required to solve large systems of equations on supercomputers. Advanced methods include Stone's

(Smith, 1985) strongly implicit procedure, preconditioning methods and conjugate gradient methods which can be seen as matrix manipulation procedures, and multigrid methods which calculate the solution on a series of coarse and fine grids in space, swapping between the grids in such a way that any errors are smoothed out.

3.10 REFERENCES AND FURTHER READING

Some of the key texts that cover the mathematics of the finite element method are Burnett (1987), Cook *et al.* (1989), Stasa (1986) and Zienkiewicz and Taylor (1989). Of these texts, that by Burnett provides the most rigorous and detailed discussion of the mathematical aspects of the finite element method for solid mechanics. Other general works are Cook (1995), NAFEMS (1986) and Hughes and Hinton (1986). The work of Taig is mentioned by Robinson (1985), although no work was published on Taig's method until Irons (1966). For the Rayleigh–Ritz method see Ritz (1909) and Reddy (1984).

Numerical solution methods are described by the authors above as well as by Smith (1985) and Hoffman (1992).