

Nam-Ho Kim

Introduction to Nonlinear Finite Element Analysis



Introduction to Nonlinear Finite Element Analysis

Nam-Ho Kim

Introduction to Nonlinear Finite Element Analysis



Springer

Nam-Ho Kim
Department of Mechanical
and Aerospace Engineering
University of Florida
Gainesville, FL, USA

Additional material to this book can be downloaded from <http://extras.springer.com>

ISBN 978-1-4419-1745-4 ISBN 978-1-4419-1746-1 (eBook)
DOI 10.1007/978-1-4419-1746-1
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2014951702

© Springer Science+Business Media New York 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my family

Preface

The finite element method (FEM) is one of the numerical methods for solving differential equations that describe many engineering problems. The FEM, originated in the area of structural mechanics, has been extended to other areas of solid mechanics and later to other fields such as heat transfer, fluid dynamics, and electromagnetism. In fact, FEM has been recognized as a powerful tool for solving partial differential equations and integrodifferential equations, and in the near future, it may become the numerical method of choice in many engineering and applied science areas. One of the reasons for FEM's popularity is that the method results in computer programs versatile in nature that can solve many practical problems with least amount of training.

The availability of undergraduate- and advanced graduate- level FEM courses in engineering schools has increased in response to the growing popularity of the FEM in industry. In the case of linear structural systems, the methods of modeling and solution procedure are well established. Nonlinear systems, however, take different modeling and solution procedures based on the characteristics of the problems. Accordingly, the modeling and solution procedures are much more complicated than that of linear systems, although there are advanced topics in linear systems such as complex shell formulations.

Researchers who have studied and applied the linear FEM cannot apply the linearized method to more complicated nonlinear problems such as elastoplastic or contact problems. However, many textbooks in the nonlinear FEMs strongly emphasize complicated theoretical parts or advanced topics. These advanced textbooks are mainly helpful to students seeking to develop additional nonlinear FEMs. However, the advanced textbooks are oftentimes too difficult for students and researchers who are learning the nonlinear FEM for the first time.

One of the biggest challenges to the instructor is finding a textbook appropriate to the level of the students. The objective of this textbook is to simply introduce the nonlinear finite element analysis procedure and to clearly explain the solution procedure to the reader. In contrast to the traditional textbooks which treat a vast amount of nonlinear theories comprehensively, this textbook only addresses the

representative problems, detailed theories, solution procedures, and the computer implementation of the nonlinear FEM. Especially by using the MATLAB programming language to introduce the nonlinear solution procedure, those readers who are not familiar with FORTRAN or C++ programming languages can easily understand and add his/her own modules to the nonlinear analysis program.

The textbook is organized into five chapters. The objective of Chap. 1 is to introduce basic concepts that will be used for developing nonlinear finite element formulations in the following chapters. Depending on the level of the students or prerequisites for the course, this chapter or a part of it can be skipped. Basic concepts in this chapter include vector and tensor calculus in Sect. 1.2, definition of stress and strain in Sect. 1.3, mechanics of continuous bodies in Sect. 1.4, and linear finite element formulation in Sect. 1.5. A MATLAB code for three-dimensional finite element analysis with solid elements will reinforce mathematical understanding.

Chapter 2 introduces nonlinear systems of solid mechanics. In Sect. 2.1, fundamental characteristics of nonlinear problems are explained in contrast to linear problems, followed by four types of nonlinearities in solid mechanics: material, geometry, boundary, and force nonlinearities. Section 2.2 presents different methods of solving a nonlinear system of equations. Discussions on convergence aspects, computational costs, load increment, and force-controlled vs. displacement-controlled methods are provided. In Sect. 2.3, step-by-step procedures in solving nonlinear finite element analysis are presented. Section 2.4 introduces NLFEA, a MATLAB code for solving nonlinear finite element equations. NLFEA can handle different material models, such as elastic, hyperelastic, and elastoplastic materials, as well as large deformation. Section 2.5 summarizes how commercial finite element analysis programs control nonlinear solution procedures. This section covers Abaqus, ANSYS, and NEi Nastran programs.

Chapter 3 presents theoretical and numerical formulations of nonlinear elastic materials. Since nonlinear elastic material normally experiences a large deformation, Sect. 3.2 discusses stress and strain measures under large deformation. Section 3.3 shows two different formulations in representing large deformation problems: total Lagrangian and updated Lagrangian. In particular, it is shown that these two formulations are mathematically identical but different in computer implementation and interpreting material behaviors. Critical load analysis is introduced in Sect. 3.4, followed by hyperelastic materials in Sect. 3.5. Different ways of representing incompressibility of elastic materials are discussed. The continuum form of the nonlinear variational equation is discretized in Sect. 3.6, followed by a MATLAB code for a hyperelastic material model in Sect. 3.7. Section 3.8 summarizes the usage of commercial finite element analysis programs to solve nonlinear elastic problems, particularly for hyperelastic materials. In hyperelastic materials, it is important to identify material parameters. Section 3.9 presents curve-fitting methods to identify hyperelastic material parameters using test data.

Different from elastic materials, some materials, such as steels or aluminum alloys, show permanent deformation when a force larger than a certain limit (elastic limit) is applied and removed. This behavior of materials is called plasticity.

When the total strain is small (infinitesimal deformation), it is possible to assume that the total strain can be additively decomposed into elastic and plastic strains. Sections 4.2 and 4.3 are based on infinitesimal elastoplasticity. In a large structure, even if the strain is small, the structure may undergo a large rigid-body motion due to accumulated deformation. In such a case, it is possible to modify infinitesimal elastoplasticity to accommodate stress calculation with the effect of rigid-body motion. Since the rate of Cauchy stress is not independent of rigid-body motion, different types of rates, called objective stress rates, are used in the constitutive relation, which is discussed in Sect. 4.4. When deformation is large, the assumption of additive decomposition of elastic and plastic strains is no longer valid. A hyperelasticity-based elastoplasticity is discussed in Sect. 4.5, in which the deformation gradient is multiplicatively decomposed into elastic and plastic parts and the stress-strain relation is given in the principal directions. This model can represent both geometric and material nonlinearities during large elastoplastic deformation. Section 4.6 is supplementary to Sect. 4.5, as it derives several expressions used in Sect. 4.5. Section 4.7 summarizes the usage of commercial finite element analysis programs to solve elastoplastic problems.

When two or more bodies collide, contact occurs between two surfaces of the bodies so that they cannot overlap in space. Metal formation, vehicle crash, projectile penetration, various seal designs, and bushing and gear systems are only a few examples of contact phenomena. In Sect. 5.2, simple one-point contact examples are presented in order to show the characteristics of contact phenomena and possible solution strategies. In Sect. 5.3, a general formulation of contact is presented based on the variational formulation. Section 5.4 focuses on finite element discretization and numerical integration of the contact variational form. Three-dimensional contact formulation is presented in Sect. 5.5. From the finite element point of view, all formulations involve use of some form of a constraint equation. Because of the highly nonlinear and discontinuous nature of contact problems, great care and trial and error are necessary to obtain solutions to practical problems. Section 5.6 presents modeling issues related to contact analysis, such as selecting slave and master bodies, removing rigid-body motions, etc.

This textbook details how the nonlinear equations are solved using practical computer programs and may be considered an essential course for those who intend to develop more complicated nonlinear finite elements. Usage of commercial FEA programs is summarized at the end of each chapter. It includes various examples in the text using Abaqus, ANSYS, NEi Nastran, and MATLAB program. Depending on availability and experience of the instructor, any program can be used as part of homework assignments and projects. The textbook website will maintain up-to-date examples with the most recent version of the commercial programs. Each chapter contains a comprehensive set of homework problems, some of which require commercial FEA programs.

Prospective readers or users of the text are graduate students in mechanical, civil, aerospace, biomedical, and industrial engineering and engineering mechanics as well as researchers and design engineers from the aforementioned fields.

The author is thankful to the students who took Advanced Finite Element Analysis course at the University of Florida and used the course package that had the same material as in this book. The author is grateful for their valuable suggestions especially regarding the example and exercise problems. Finally, special thanks to my daughter, Hyesu Grace Kim, for her outstanding work editing the manuscript.

Gainesville, FL, USA
July 2014

Nam-Ho Kim

Contents

1 Preliminary Concepts	1
1.1 Introduction	1
1.2 Vector and Tensor Calculus	3
1.2.1 Vector and Tensor	3
1.2.2 Vector and Tensor Calculus	11
1.2.3 Integral Theorems	12
1.3 Stress and Strain	14
1.3.1 Stress	15
1.3.2 Strain	26
1.3.3 Stress–Strain Relationship	31
1.4 Mechanics of Continuous Bodies	36
1.4.1 Boundary-Valued Problem	37
1.4.2 Principle of Minimum Potential Energy	38
1.4.3 Principle of Virtual Work	46
1.5 Finite Element Method	50
1.5.1 Finite Element Approximation	50
1.5.2 Finite Element Equations for a One-Dimensional Problem	54
1.5.3 Finite Element Equations for 3D Solid Element	61
1.5.4 A MATLAB Code for Finite Element Analysis	67
1.6 Exercises	73
References	79
2 Nonlinear Finite Element Analysis Procedure	81
2.1 Introduction to Nonlinear Systems in Solid Mechanics	81
2.1.1 Geometric Nonlinearity	85
2.1.2 Material Nonlinearity	87
2.1.3 Kinematic Nonlinearity	89
2.1.4 Force Nonlinearity	90

2.2	Solution Procedures for Nonlinear Algebraic Equations	91
2.2.1	Newton–Raphson Method	93
2.2.2	Modified Newton–Raphson Method	101
2.2.3	Incremental Secant Method	103
2.2.4	Incremental Force Method	109
2.3	Steps in the Solution of Nonlinear Finite Element Analysis	114
2.3.1	State Determination	114
2.3.2	Residual Calculation	115
2.3.3	Convergence Check	116
2.3.4	Linearization	116
2.3.5	Solution	117
2.4	MATLAB Code for a Nonlinear Finite Element Analysis Procedure	119
2.5	Nonlinear Solution Controls Using Commercial Finite Element Programs	132
2.5.1	Abaqus	133
2.5.2	ANSYS	134
2.5.3	NEiNastran	136
2.6	Summary	137
2.7	Exercises	138
	References	140
3	Finite Element Analysis for Nonlinear Elastic Systems	141
3.1	Introduction	141
3.2	Stress and Strain Measures in Large Deformation	142
3.2.1	Deformation Gradient	142
3.2.2	Lagrangian and Eulerian Strains	145
3.2.3	Polar Decomposition	150
3.2.4	Deformation of Surface and Volume	154
3.2.5	Cauchy and Piola–Kirchhoff Stresses	158
3.3	Nonlinear Elastic Analysis	161
3.3.1	Nonlinear Static Analysis: Total Lagrangian Formulation	162
3.3.2	Nonlinear Static Analysis: Updated Lagrangian Formulation	174
3.4	Critical Load Analysis	179
3.4.1	One-Point Approach	181
3.4.2	Two-Point Approach	181
3.4.3	Stability Equation with Actual Critical Load Factor	182
3.5	Hyperelastic Materials	183
3.5.1	Strain Energy Density	184
3.5.2	Nearly Incompressible Hyperelasticity	190
3.5.3	Variational Equation and Linearization	197
3.6	Finite Element Formulation for Nonlinear Elasticity	200
3.7	MATLAB Code for Hyperelastic Material Model	205

3.8	Nonlinear Elastic Analysis Using Commercial Finite Element Programs	211
3.8.1	Usage of Commercial Programs	211
3.8.2	Modeling Examples of Nonlinear Elastic Materials	214
3.9	Fitting Hyperelastic Material Parameters from Test Data	221
3.9.1	Elastomer Test Procedures	222
3.9.2	Data Preparation	224
3.9.3	Curve Fitting	227
3.9.4	Stability of Constitutive Model	229
3.10	Summary	231
3.11	Exercises	232
	References	239
4	Finite Element Analysis for Elastoplastic Problems	241
4.1	Introduction	241
4.2	One-Dimensional Elastoplasticity	242
4.2.1	Elastoplastic Material Behavior	242
4.2.2	Finite Element Formulation for Elastoplasticity	247
4.2.3	Determination of Stress State	250
4.3	Multidimensional Elastoplasticity	265
4.3.1	Yield Functions and Yield Criteria	266
4.3.2	Von Mises Yield Criterion	272
4.3.3	Hardening Models	275
4.3.4	Classical Elastoplasticity Model	280
4.3.5	Numerical Integration	290
4.3.6	Computational Implementation of Elastoplasticity	299
4.4	Finite Rotation with Objective Integration	308
4.4.1	Objective Tensor and Objective Rate	309
4.4.2	Finite Rotation and Objective Rate	314
4.4.3	Incremental Equation for Finite Rotation Elastoplasticity	317
4.4.4	Computational Implementation of Finite Rotation	321
4.5	Finite Deformation Elastoplasticity with Hyperelasticity	325
4.5.1	Multiplicative Decomposition	325
4.5.2	Finite Deformation Elastoplasticity	326
4.5.3	Time Integration	330
4.5.4	Return-Mapping Algorithm	333
4.5.5	Consistent Algorithmic Tangent Operator	336
4.5.6	Variational Principles for Finite Deformation	337
4.5.7	Computer Implementation of Finite Deformation Elastoplasticity	338
4.6	Mathematical Formulas from Finite Elasticity	343
4.6.1	Linearization of Principal Logarithmic Stretches	343
4.6.2	Linearization of the Eigenvector of the Elastic Trial Left Cauchy-Green Tensor	344

4.7	MATLAB Code for Elastoplastic Material Model	345
4.8	Elastoplasticity Analysis of Using Commercial Finite Element Programs	350
4.8.1	Usage of Commercial Programs	350
4.8.2	Modeling Examples of Elastoplastic Materials	355
4.9	Summary	359
4.10	Exercises	360
	References	365
5	Finite Element Analysis for Contact Problems	367
5.1	Introduction	367
5.2	Examples of Simple One-Point Contact	369
5.2.1	Contact of a Cantilever Beam with a Rigid Block	369
5.2.2	Contact of a Cantilever Beam with Friction	374
5.3	General Formulation for Contact Problems	378
5.3.1	Contact Condition with Rigid Surface	379
5.3.2	Variational Inequality in Contact Problems	382
5.3.3	Penalty Regularization	385
5.3.4	Frictionless Contact Formulation	389
5.3.5	Frictional Contact Formulation	393
5.4	Finite Element Formulation of Contact Problems	398
5.4.1	Contact Between a Flexible Body and a Rigid Body	398
5.4.2	Contact Between Two Flexible Bodies	404
5.4.3	MATLAB Code for Contact Analysis	406
5.5	Three-Dimensional Contact Analysis	408
5.6	Contact Analysis Procedure and Modeling Issues	412
5.6.1	Contact Analysis Procedure	413
5.6.2	Contact Modeling Issues	417
5.7	Exercises	423
	References	426
Index		427

Chapter 1

Preliminary Concepts

1.1 Introduction

The finite element method (FEM) is one of the numerical methods for solving differential equations that describe many engineering problems. The FEM originated from the structural mechanics discipline and has since been extended to other areas of solid mechanics as well as heat transfer, fluid dynamics, and electromagnetism. In fact, FEM has been recognized as a powerful tool for solving partial differential equations and integrodifferential equations, and in the near future, it may become the numerical method of choice in many engineering and applied science areas. One of the many reasons for the popularity of the FEM is that a minimal amount of training is required to solve many practical problems with the aid of versatile computer programs.

The availability of undergraduate- and advanced graduate-level FEM courses in engineering schools has increased in response to the growing popularity of the FEM in industry. In the case of linear structural systems, the methods of modeling and solution procedure are well established. Nonlinear systems, however, take different modeling and solution procedures based on the characteristics of the problems. Accordingly, the modeling and solution procedures are much more complicated than that of linear systems, although there are advanced topics in linear systems such as complex shell formulations.

Researchers who have studied and applied the linear FEM cannot apply the linearized method to more complicated nonlinear problems such as elastoplastic or contact problems. However, many textbooks in the nonlinear FEMs strongly emphasize complicated theoretical parts or advanced topics. These advanced textbooks are mainly helpful to researchers seeking to develop additional nonlinear FEMs. However, the advanced textbooks are oftentimes too difficult for students and researchers who are learning the nonlinear FEM for the first time.

The objective of this textbook is to simply introduce the nonlinear finite element analysis procedure and to clearly explain the solution procedure to the reader.

In contrast to the traditional textbooks which treat a vast amount of nonlinear theories comprehensively, this textbook only addresses the representative problems, detailed theories, solution procedures, and the computer implementation of the nonlinear FEM. Especially by using the MATLAB programming language to introduce the nonlinear solution procedure, those readers who are not familiar with FORTRAN or C++ programming languages can easily understand and add his/her own modules to the nonlinear analysis program. This textbook details how the nonlinear equations are solved using practical computer programs and may be considered an essential course for those who intend to develop more complicated nonlinear finite elements.

The objective of this chapter is to introduce basic concepts that will be used for developing nonlinear finite element formulations in the following chapters. Basic concepts in this chapter include vector and tensor calculus in Sect. 1.2, definition of stress and strain in Sect. 1.3, mechanics of continuous bodies in Sect. 1.4, and linear finite element formulation in Sect. 1.5. Technical contents in this chapter are by no means rigorous or complete. The readers are referred to advanced textbooks for detailed explanations and rigorous derivations.

A relatively simple theory is introduced in Sect. 1.4 that can formulate the structural equilibrium using the energy principle. Since all conservative systems have potential energy, the energy principle may be applied to find the structural equilibrium. Structural equilibrium, by the principle of minimum total potential energy, is considered to be a stationary configuration in which the potential energy of the structural system is minimized. Since the potential energy of many structural problems is the positive definite quadratic function of a state variable, such as displacement, the stationary condition yields a unique global minimum solution. The stationary condition is further developed to a variational method for a conservative system. An important result is then shown, namely, that if the solution for a differential equation exists, then that solution is the minimizing solution of the total potential energy. In addition, the structural problem may have a natural solution that minimizes the total potential energy even if the structural differential problem does not have a solution. The energy principles presented in Sect. 1.4 will be restricted to small strains and displacements so that strain–displacement relationships can be expressed in terms of linear equations; such displacements and corresponding strains obviously have additive properties. A nonlinear elastic stress–strain relationship will be discussed in Chap. 3 of this text.

The energy-based formulation of the potential problem is generalized to the principle of virtual work, which can handle arbitrary constitutive relations. The principle of virtual work is the equilibrium of the work done by both internal and external forces with the small, arbitrary, virtual displacements that satisfy kinematic constraints. For a conservative system, the same results are obtained as with the principle of minimum total potential energy. The unified approach to various structural problems is made possible by introducing energy-bilinear and load-linear forms. As long as energy-bilinear and load-linear forms share the same properties, then all structural problems in this text can be treated in the same manner, even structural problems with different differential operators. The existence and

uniqueness of a solution can be shown through rigorous mathematical proofs. The concept of Sobolev space and the bounded property of an energy-bilinear form are required in the proof. However, in this text, such rigorous mathematical proofs are avoided and corresponding references are cited.

1.2 Vector and Tensor Calculus

Since vector and tensor calculus are extensively used in computational mechanics, it is worth reviewing some fundamental concepts and recalling some important results that will be used in this book. A brief summary of concepts and results pertinent to the development of the subject is provided within Sect. 1.2 of the text for the convenience of students. For a thorough understanding of the mathematical concepts, readers are advised to refer to standard textbooks, e.g., Kreyszig [1] and Strang [2].

1.2.1 Vector and Tensor

Cartesian vector: In general, a vector is defined as a collection of scalars. A Cartesian vector is a Euclidean vector defined using Cartesian coordinates. Each vector, in this text, is considered to be a column vector unless otherwise specified. A Cartesian vector in two- or three-dimensional space is denoted by a bold typeface:

$$\mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} \quad \text{or} \quad \mathbf{u} = \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix}, \quad (1.1)$$

where u_1 , u_2 , and u_3 are components of the vector \mathbf{u} in the x -, y -, and z -coordinates, respectively, as shown in Fig. 1.1. To save space, the above column vector \mathbf{u} can be written as $\mathbf{u} = \{u_1, u_2, u_3\}^T$, in which $\{\cdot\}^T$ denotes the *transpose* of a vector.

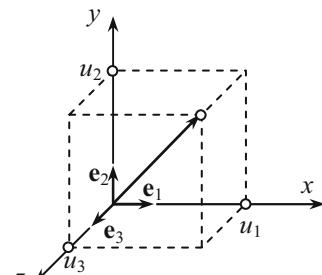


Fig. 1.1 Three-dimensional Cartesian vector

The above three-dimensional Cartesian vector can also be denoted using a unit vector in each coordinate direction. Let $\mathbf{e}_1 = \{1, 0, 0\}^T$, $\mathbf{e}_2 = \{0, 1, 0\}^T$, and $\mathbf{e}_3 = \{0, 0, 1\}^T$ be the unit vectors in the x -, y -, and z -direction, respectively. Then,

$$\mathbf{u} = u_1 \mathbf{e}_1 + u_2 \mathbf{e}_2 + u_3 \mathbf{e}_3.$$

In the above equation, \mathbf{e}_1 , \mathbf{e}_2 , and \mathbf{e}_3 are called basis vectors. Any vector in the three-dimensional space V can be represented by a linear combination of the basis vectors, e.g., $\mathbf{w} = w_1 \mathbf{e}_1 + w_2 \mathbf{e}_2 + w_3 \mathbf{e}_3$, for all $\mathbf{w} \in V$. For notational convenience, the following summation notation will be used throughout the text:

$$\mathbf{u} = u_j \mathbf{e}_j,$$

where $j = 1, 2$, and 3 for three dimensions or $j = 1$ and 2 for two dimensions. In this notation, the summation is specified over the range of the repeated index, j . Note that an index can only be repeated once in a term; therefore, the term $u_i v_i \mathbf{e}_i$ is an improper instance of index notation. The repeated index is called a dummy index because it disappears after summation; therefore, $u_j \mathbf{e}_j = u_i \mathbf{e}_i$.

Using the summation notation, the inner product of two Cartesian vectors can be calculated by

$$\begin{aligned} \mathbf{u} \cdot \mathbf{v} &= (u_i \mathbf{e}_i) \cdot (v_j \mathbf{e}_j) \\ &= u_i v_j (\mathbf{e}_i \cdot \mathbf{e}_j) \\ &= u_i v_j \delta_{ij} \\ &= u_i v_i. \end{aligned} \tag{1.2}$$

In the above derivation, δ_{ij} is the Kronecker delta symbol, which is defined as

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j. \end{cases} \tag{1.3}$$

Using this property, it is straightforward to verify that $v_j \delta_{ij} = v_i$; i.e., the Kronecker delta symbol replaces the repeated index with the non-repeated one. Also note that $\delta_{jj} = 2$ and $\delta_{jj} = 3$ for two and three dimensions, respectively, because summation is specified by the repeated index j .

Cartesian components of a vector can be obtained by using the inner product with the basis vectors, e.g.,

$$\mathbf{e}_j \cdot \mathbf{v} = \mathbf{e}_j \cdot (v_i \mathbf{e}_i) = v_i \delta_{ij} = v_j.$$

Since this is equivalent to projecting the vector onto the axis of a coordinate, it is also called a projection.

The magnitude of a vector can be calculated by taking the square root of the inner product of the vector itself as

$$\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}. \quad (1.4)$$

In general, the magnitude of a vector is called a norm.

Cartesian tensor: The component form of a vector in Eq. (1.1) has a single index, i.e., u_j . In general, it is possible to have multiple indices; for example, the components of a matrix, a_{ij} , have two indices. The notion of a Cartesian tensor is a generalization of a vector; i.e., a vector is called a rank-1 tensor. Then, it is possible to define a rank-2 tensor, a rank-3 tensor, etc. In addition, a scalar can be considered as a rank-0 tensor. The rank of a tensor can be determined by the number of indices; for example, the components of a rank-4 tensor have four indices, as C_{ijkl} . A basic rank-2 tensor is the identity tensor, which is defined by $\mathbf{1} = [\delta_{ij}]$. In matrix notation, the rank-2 identity tensor corresponds to a 3×3 identity matrix. In particular, a rank-2 Cartesian tensor is often called a matrix. For example, a stress is a rank-2 tensor, whose components are defined as

$$[\sigma_{ij}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}. \quad (1.5)$$

A Cartesian tensor can be represented by a component array in terms of a basis (\mathbf{e}_i). For example, a rank-2 Cartesian tensor can be written as

$$\mathbf{T} = T_{ij} \mathbf{e}_i \otimes \mathbf{e}_j, \quad (1.6)$$

where the symbol, \otimes , is called the dyadic product, which increases the rank by 1. A higher rank tensor can be defined by using multiple dyadic products. Since \mathbf{e}_i is a rank-1 tensor, $\mathbf{e}_i \otimes \mathbf{e}_j$ and $\mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_k \otimes \mathbf{e}_l$ yield a rank-2 and rank-4 tensor, respectively. The transpose of \mathbf{T} can be defined as $\mathbf{T}^T = T_{ji} \mathbf{e}_i \otimes \mathbf{e}_j$. Note that the summation rule should be applied for the repeated indices. In this definition, the stress tensor can be defined as $\boldsymbol{\sigma} = \sigma_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$, and the matrix in Eq. (1.5) is the Cartesian components of the stress tensor. The following identities are a direct consequence of the definition of the dyadic product:

$$\begin{aligned} \mathbf{u} \otimes \mathbf{v} &\neq \mathbf{v} \otimes \mathbf{u}, \\ (\alpha \mathbf{u}) \otimes \mathbf{v} &= \alpha(\mathbf{u} \otimes \mathbf{v}), \\ \mathbf{u} \otimes (\mathbf{v} + \mathbf{w}) &= \mathbf{u} \otimes \mathbf{v} + \mathbf{u} \otimes \mathbf{w}, \\ (\mathbf{u} \otimes \mathbf{v}) \cdot \mathbf{w} &= (\mathbf{v} \cdot \mathbf{w}) \mathbf{u}, \\ \mathbf{u} \cdot (\mathbf{v} \otimes \mathbf{w}) &= (\mathbf{u} \cdot \mathbf{v}) \mathbf{w}. \end{aligned} \quad (1.7)$$

Note that the inner product is applied to the closest two vectors. The inner product between rank- m and rank- n tensors yields a rank- $(m+n-2)$ tensor; therefore, the

Table 1.1 Comparison of different notations

Direct tensor notation	Tensor component notation	Matrix notation
$\alpha = \mathbf{a} \cdot \mathbf{b}$	$\alpha = a_i b_i$	$\alpha = \mathbf{a}^T \mathbf{b}$
$\mathbf{A} = \mathbf{a} \otimes \mathbf{b}$	$A_{ij} = a_i b_j$	$\mathbf{A} = \mathbf{ab}^T$
$\mathbf{b} = \mathbf{A} \cdot \mathbf{a}$	$b_i = A_{ij} a_j$	$\mathbf{b} = \mathbf{Aa}$
$\mathbf{b} = \mathbf{a} \cdot \mathbf{A}$	$b_j = a_i A_{ij}$	$\mathbf{b}^T = \mathbf{a}^T \mathbf{A}$

inner product reduces the rank by 2. Table 1.1 compares three different notations used in this text. For convenience, the symbol “.” can often be omitted for the inner product, i.e., $\mathbf{A} \cdot \mathbf{B} = \mathbf{AB}$.

Example 1.1 (Inner product of two tensors) Consider the inner product of two rank-2 tensors: $\mathbf{C} = \mathbf{A} \cdot \mathbf{B}$. Using the dyadic representation method as in Eq. (1.6), calculate the Cartesian components of \mathbf{C} in terms of that of \mathbf{A} and \mathbf{B} .

Solution In the dyadic representation, the two tensors can be written as $\mathbf{A} = A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j$ and $\mathbf{B} = B_{kl} \mathbf{e}_k \otimes \mathbf{e}_l$. Therefore, the inner product between them can be expressed as

$$\begin{aligned}\mathbf{C} &= \mathbf{A} \cdot \mathbf{B} \\ &= (A_{ij} \mathbf{e}_i \otimes \mathbf{e}_j) \cdot (B_{kl} \mathbf{e}_k \otimes \mathbf{e}_l) \\ &= A_{ij} B_{kl} \delta_{jk} \mathbf{e}_i \otimes \mathbf{e}_l \\ &= A_{ik} B_{kl} \mathbf{e}_i \otimes \mathbf{e}_l.\end{aligned}$$

Therefore, the components of \mathbf{C} become $C_{il} = A_{ik} B_{kl}$. Note that the same components of \mathbf{C} can be obtained by matrix multiplication between the components of \mathbf{A} and \mathbf{B} . ■

Symmetric and skew tensors: Rank-2 symmetric and skew tensors can be defined as

- Symmetric tensor:

$$\mathbf{S} = \mathbf{S}^T. \quad (1.8)$$

- Skew tensor:

$$\mathbf{W} = -\mathbf{W}^T. \quad (1.9)$$

It is noted that every rank-2 tensor can be uniquely decomposed into a symmetric and a skew tensor, as

$$\mathbf{T} = \mathbf{S} + \mathbf{W}, \quad (1.10)$$

where

$$\mathbf{S} = \frac{1}{2}(\mathbf{T} + \mathbf{T}^T), \quad (1.11)$$

$$\mathbf{W} = \frac{1}{2}(\mathbf{T} - \mathbf{T}^T). \quad (1.12)$$

Note that the skew tensor \mathbf{W} has zero diagonal components and $W_{ij} = -W_{ji}$. The symmetric part of a tensor is often written as $\mathbf{S} = \text{sym}(\mathbf{T})$, while the skew part is written as $\mathbf{W} = \text{skew}(\mathbf{T})$.

Example 1.2 (Symmetric and skew part of displacement gradient) A displacement gradient, $\nabla \mathbf{u}$, is a rank-2 tensor. Calculate the symmetric and skew part of the displacement gradient.

Solution The components of the displacement gradient can be defined as

$$\nabla \mathbf{u} = \left[\begin{matrix} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \end{matrix} \right] = \left[\begin{matrix} \frac{\partial u_1}{\partial x_1} & \frac{\partial u_1}{\partial x_2} & \frac{\partial u_1}{\partial x_3} \\ \frac{\partial u_2}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_2}{\partial x_3} \\ \frac{\partial u_3}{\partial x_1} & \frac{\partial u_3}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \\ \frac{\partial u_1}{\partial x_1} & \frac{\partial u_2}{\partial x_2} & \frac{\partial u_3}{\partial x_3} \end{matrix} \right].$$

Then, the symmetric and skew parts can be obtained as

$$\text{sym}(\nabla \mathbf{u}) = \left[\begin{matrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \\ \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{\partial u_2}{\partial x_2} & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) \\ \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) & \frac{\partial u_3}{\partial x_3} \end{matrix} \right],$$

$$\text{skew}(\nabla \mathbf{u}) = \left[\begin{matrix} 0 & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} - \frac{\partial u_2}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} - \frac{\partial u_3}{\partial x_1} \right) \\ \frac{1}{2} \left(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2} \right) & 0 & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} - \frac{\partial u_3}{\partial x_2} \right) \\ \frac{1}{2} \left(\frac{\partial u_3}{\partial x_1} - \frac{\partial u_1}{\partial x_3} \right) & \frac{1}{2} \left(\frac{\partial u_3}{\partial x_2} - \frac{\partial u_2}{\partial x_3} \right) & 0 \end{matrix} \right].$$

Note that $\text{sym}(\nabla \mathbf{u})$ is called the strain tensor, while $\text{skew}(\nabla \mathbf{u})$ is called the spin tensor. ■

Contraction and trace: The contraction operator is defined between two tensors and can be considered as a double inner product. For two rank-2 tensors, the contraction is defined as

$$\mathbf{a} : \mathbf{b} = a_{ij}b_{ij} = a_{11}b_{11} + a_{12}b_{12} + \cdots + a_{32}b_{32} + a_{33}b_{33}. \quad (1.13)$$

Note that the result becomes a scalar. In general, the contraction operator reduces four ranks from the sum of ranks of two tensors. Similar to the magnitude of a vector, the magnitude (or, norm) of a rank-2 tensor can be defined using the contraction operator as

$$\|\mathbf{a}\| = \sqrt{\mathbf{a} : \mathbf{a}}. \quad (1.14)$$

In solid mechanics, the constitutive equation of an elastic material is often given as a linear relationship between stress and strain. Since stress and strain are rank-2 tensors, the elastic modulus must be defined in terms of rank-4 tensors as

$$\boldsymbol{\sigma} = \mathbf{D} : \boldsymbol{\epsilon}, \quad \sigma_{ij} = D_{ijkl}\epsilon_{kl}, \quad (1.15)$$

where D_{ijkl} is a rank-4 tensor that represents the elastic modulus.

The trace of a tensor is part of the contraction operator in which a pair of indices is under the inner product. In the case of a rank-2 tensor, the trace can be defined as

$$tr(\mathbf{A}) = A_{ii} = A_{11} + A_{22} + A_{33}, \quad (1.16)$$

where $tr(\cdot)$ stands for the trace operator. In the tensor notation, the trace can be written as $tr(\mathbf{A}) = \mathbf{A} : \mathbf{1} = \mathbf{1} : \mathbf{A}$.

Example 1.3 (Contraction of a symmetric tensor) Let \mathbf{A} be a rank-2 symmetric tensor. Show that $\mathbf{A} : \mathbf{W} = 0$ and $\mathbf{A} : \mathbf{T} = \mathbf{A} : \mathbf{S}$, where \mathbf{T} is a rank-2 nonsymmetric tensor, whose symmetric and skew parts are, respectively, \mathbf{S} and \mathbf{W} .

Solution The contraction between a symmetric and a skew tensor becomes

$$\mathbf{A} : \mathbf{W} = A_{ij}W_{ij} = -A_{ij}W_{ji} = -A_{ji}W_{ji} = -\mathbf{A} : \mathbf{W}.$$

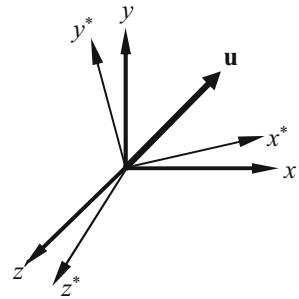
In the second equality, the definition of a skew tensor is used, while the definition of a symmetric tensor is used in the third equality. From the above relation, it is obvious that $\mathbf{A} : \mathbf{W} = 0$.

For a nonsymmetric tensor, \mathbf{T} , it can be decomposed into a symmetric and a skew part:

$$\mathbf{A} : \mathbf{T} = \mathbf{A} : (\mathbf{S} + \mathbf{W}) = \mathbf{A} : \mathbf{S}.$$

Note that $\mathbf{A} : \mathbf{W} = 0$ is used. ■

Fig. 1.2 Representation of a vector in two coordinate systems



Orthogonal tensor: An important rank-2 tensor is an orthogonal tensor, which represents the rotation of a vector or coordinate system. Consider a vector, \mathbf{u} , in Fig. 1.2 with two different coordinate systems. The vector can be represented by the bases of each of the two coordinate systems as

$$\mathbf{u} = u_i \mathbf{e}_i = u_j^* \mathbf{e}_j^*.$$

Then, using the two bases, an orthogonal tensor, $\beta = [\beta_{ij}]$, can be defined as

$$\beta_{ij} = \mathbf{e}_i^* \otimes \mathbf{e}_j. \quad (1.17)$$

This orthogonal tensor represents the rotational relation between the two coordinate systems. It is straightforward to show that (no sum on j)

$$\beta_{ij} \mathbf{e}_j = (\mathbf{e}_i^* \otimes \mathbf{e}_j) \cdot \mathbf{e}_j = \mathbf{e}_i^*.$$

In a similar way, $\mathbf{e}_j = \beta_{ij} \mathbf{e}_i^*$. Then, using an easy calculation, it is possible to show that

$$\begin{aligned} \mathbf{u}^* &= \beta \mathbf{u}, \\ \mathbf{u} &= \beta^T \mathbf{u}^*. \end{aligned} \quad (1.18)$$

In the above equation, the inner product symbol “.” is omitted, which will be commonly excluded from this book. Using the above relation, it is easy to show that $\beta^T \beta = \beta \beta^T = \mathbf{1}$, which is the property of an orthogonal tensor. The coordinate transformation of a rank-2 tensor can be written as

$$\mathbf{T}^* = \beta \mathbf{T} \beta^T, \quad T_{ij}^* = \beta_{ik} T_{kl} \beta_{jl}. \quad (1.19)$$

Note that the above equation is not a rotation of a tensor but a rotation of the coordinate system.

Permutation: The permutation symbol has three indices, but it is not a tensor. It is used to

$$e_{ijk} = \begin{cases} 1 & \text{if } ijk \text{ are an even permutation : } 123, 231, 312 \\ -1 & \text{if } ijk \text{ are an odd permutation : } 132, 213, 321 \\ 0 & \text{otherwise} \end{cases}.$$

Note that the permutation is zero when any of two indices have the same value. The permutation symbol will be used for several important derivations. The following identity can be useful in deriving the determinant of a tensor:

$$e_{ijk}e_{lmk} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}. \quad (1.20)$$

Another usage of the permutation symbol is for a vector product of two vectors,

$$\mathbf{u} \times \mathbf{v} = \mathbf{e}_i e_{ijk} u_j v_k. \quad (1.21)$$

Note that the output of a vector product is another vector that is orthogonal to the two vectors.

Dual vector of a skew tensor: A rank-2 skew tensor has only three independent components. Therefore, it is possible that a skew tensor can be defined using a vector with the permutation symbol as $W_{ij} = -e_{ijk}w_k$, where the components of \mathbf{W} and \mathbf{w} are given as

$$\mathbf{W} = \begin{Bmatrix} 0 & W_{12} & W_{13} \\ -W_{12} & 0 & W_{23} \\ -W_{13} & -W_{23} & 0 \end{Bmatrix} \Rightarrow \mathbf{w} = \begin{Bmatrix} -W_{23} \\ W_{13} \\ -W_{12} \end{Bmatrix}.$$

In addition, for any skew tensor, \mathbf{W} , and vector, \mathbf{u} , the following property can be shown:

$$\mathbf{u} \cdot \mathbf{Wu} = \mathbf{u} \cdot \mathbf{W}^T \mathbf{u} = -\mathbf{u} \cdot \mathbf{Wu} = 0.$$

In the first equality, since $(\mathbf{u} \cdot \mathbf{Wu})$ is a scalar, it is equivalent to its transpose $(\mathbf{u} \cdot \mathbf{W}^T \mathbf{u})$. The above relation reveals that \mathbf{Wu} and \mathbf{u} are orthogonal. Then, we can obtain the following relation:

$$W_{ij}u_j = -e_{ijk}w_ku_j = e_{ikj}w_ku_j.$$

Note that the last term is simply the definition of the vector product in Eq. (1.21). Therefore, in tensor notation,

$$\mathbf{Wu} = \mathbf{w} \times \mathbf{u}. \quad (1.22)$$

In the above equation, \mathbf{w} is called a dual vector of the skew tensor \mathbf{W} . For a given skew tensor, the dual vector can be obtained using $w_i = -\frac{1}{2}e_{ijk}W_{jk}$. In practice, the usage of the vector product is inconvenient because of the permutation symbol. However, the above relation makes it possible to convert the vector product into the inner product between a skew tensor and a vector.

1.2.2 Vector and Tensor Calculus

Gradient: Many governing equations of structural mechanics include the derivative of a field variable with respect to spatial coordinates. Here, a “field” means a function in the space, such as a temperature or displacement of a structure. The field variable can be a scalar, vector, or tensor. Therefore, it is a good idea to clearly define the gradient operator using the tensor notation. The gradient operator is defined as a vector (or, rank-1 tensor), as

$$\nabla = \frac{\partial}{\partial \mathbf{x}} = \mathbf{e}_i \frac{\partial}{\partial x_i}. \quad (1.23)$$

For example, the gradient of a scalar field, $\phi(\mathbf{x})$, can be written as

$$\nabla \phi = \text{grad} \phi = \mathbf{e}_i \frac{\partial \phi}{\partial x_i}, \quad (1.24)$$

which is a vector. The gradient of a vector field, $\mathbf{u}(\mathbf{x})$, can be defined as

$$\nabla \mathbf{u} = \left(\mathbf{e}_i \frac{\partial}{\partial x_i} \right) \otimes (u_j \mathbf{e}_j) = \frac{\partial u_j}{\partial x_i} \mathbf{e}_i \otimes \mathbf{e}_j. \quad (1.25)$$

Note that the gradient of a vector is a rank-2 tensor. For the purpose of notational convenience, the subscribed comma will often be used for the gradient, i.e., $v_{i,j} = \partial v_i / \partial x_j$. The following divergence is defined so that the gradient of a vector produces a scalar quantity:

$$\nabla \cdot \mathbf{u} = \left(\mathbf{e}_i \frac{\partial}{\partial x_i} \right) \cdot (u_j \mathbf{e}_j) = \frac{\partial u_i}{\partial x_i} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}. \quad (1.26)$$

The Laplace operator can be defined using the inner product of two gradient operators as

$$\nabla^2 = \nabla \cdot \nabla = \left(\mathbf{e}_i \frac{\partial}{\partial x_i} \right) \cdot \left(\mathbf{e}_j \frac{\partial}{\partial x_j} \right) = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_i} = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}. \quad (1.27)$$

Example 1.4 (Divergence of a stress tensor) Let σ be the stress tensor given in Eq. (1.5). The force equilibrium of an infinitesimal component can be written as $\nabla \cdot \sigma = 0$. Write the force equilibrium equation in component form.

Solution By replacing the vector \mathbf{u} in Eq. (1.26) with the stress tensor, the divergence of the stress tensor becomes

$$(\nabla \cdot \sigma)_j = \frac{\partial \sigma_{ij}}{\partial x_i}.$$

By expanding the above equation for all components, and by putting the divergence equal to zero, we can obtain the following differential equation for equilibrium:

$$\begin{cases} \frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{21}}{\partial x_2} + \frac{\partial \sigma_{31}}{\partial x_3} = 0 \\ \frac{\partial \sigma_{12}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2} + \frac{\partial \sigma_{32}}{\partial x_3} = 0. \\ \frac{\partial \sigma_{13}}{\partial x_1} + \frac{\partial \sigma_{23}}{\partial x_2} + \frac{\partial \sigma_{33}}{\partial x_3} = 0 \end{cases}$$

It will be shown later that the stress tensor is symmetric. Therefore, $\sigma_{12} = \sigma_{21}$, $\sigma_{23} = \sigma_{32}$ and $\sigma_{13} = \sigma_{31}$. ■

1.2.3 Integral Theorems

Many equations in solid mechanics are expressed in terms of differential equations. For example, the force equilibrium of an infinitesimal component can be expressed in terms of partial differential equations. Since these differential equations are satisfied at every point in the domain of a structure, they are integrated over the entire domain; the result is an integral equation. In this section, several theorems that are useful in deriving the integral equations of solid mechanics are introduced. The proof of each theorem is out of the scope of this text. Interested readers are referred to the text by Hildebrand [3].

Divergence theorem: The divergence theorem is a special case of Green's theorem for a tensor field. The divergence theorem relates a domain integral to a boundary integral around the domain. Let Ω be a domain bounded by Γ . If a tensor, \mathbf{A} , has continuous partial derivatives in the domain, the integral of the divergence of \mathbf{A} over the domain can be converted into the integral over the boundary, as

$$\iint_{\Omega} \nabla \cdot \mathbf{A} d\Omega = \int_{\Gamma} \mathbf{n} \cdot \mathbf{A} d\Gamma, \quad (1.28)$$

where \mathbf{n} is the outward unit normal vector of the boundary, Γ . A variant of the divergence theorem is the gradient theorem in which the inner product is replaced with the dyadic product, as

$$\iint_{\Omega} \nabla \mathbf{A} d\Omega = \int_{\Gamma} \mathbf{n} \otimes \mathbf{A} d\Gamma.$$

Reynolds transport theorem: The Reynolds transport theorem is related to the time derivative of an integral equation over a domain in which the integrand, as well as the domain, varies as a function of time. Consider integrating $\mathbf{f} = \mathbf{f}(\mathbf{x}, t)$ over the time-dependent domain, $\Omega(t)$, that is bounded by $\Gamma(t)$. Then, the time derivative of the integral of $\mathbf{f}(\mathbf{x}, t)$ over the domain, $\Omega(t)$, can be expressed as

$$\frac{d}{dt} \iint_{\Omega} \mathbf{f} d\Omega = \iint_{\Omega} \frac{\partial \mathbf{f}}{\partial t} d\Omega + \int_{\Gamma} (\mathbf{n} \cdot \mathbf{v}) \mathbf{f} d\Gamma, \quad (1.29)$$

where $\mathbf{n}(\mathbf{x}, t)$ is the outward unit normal vector to the boundary and $\mathbf{v}(\mathbf{x}, t)$ is the velocity of the boundary. The first term on the right-hand side (RHS) is called the partial derivative, and the second term is called the convective term. Note that the integral on the left-hand side (LHS) is solely a function of time, so that the total derivative is used.

Integration-by-parts: Integration-by-parts is a theorem that relates the integral of a product of functions to the integral of their derivative and antiderivative. In the one-dimensional case, if $u(x)$ and $v(x)$ are two continuously differentiable functions in the domain (a, b) , then the integration-by-parts can be stated as

$$\int_a^b u(x)v'(x) dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x) dx.$$

The above relation can be extended to the two- or three-dimensional case. Let Ω be the domain of integral with the boundary, Γ . Then, the integration-by-parts can be written as

$$\iint_{\Omega} \frac{\partial u}{\partial x_i} v d\Omega = \int_{\Gamma} u v n_i d\Gamma - \iint_{\Omega} u \frac{\partial v}{\partial x_i} d\Omega,$$

where n_i is the components of the unit normal vector directed outward to the boundary, Γ . Replacing the scalar function, v , in the above formula with a vector, \mathbf{v}_i , and summing over i give the following vector formula:

$$\iint_{\Omega} \nabla u \cdot \mathbf{v} d\Omega = \int_{\Gamma} u(\mathbf{v} \cdot \mathbf{n}) d\Gamma - \iint_{\Omega} u \nabla \cdot \mathbf{v} d\Omega. \quad (1.30)$$

By replacing u with the constant 1 in the above formula, the divergence theorem in Eq. (1.28) can be obtained. For the purpose of continuum mechanics, the following Green's identity can be obtained by replacing \mathbf{v} with ∇v in the above formula:

$$\iint_{\Omega} \nabla u \cdot \nabla v d\Omega = \int_{\Gamma} u \nabla v \cdot \mathbf{n} d\Gamma - \iint_{\Omega} u \nabla^2 v d\Omega. \quad (1.31)$$

One of the important reasons for using integration-by-parts is to relax the requirement of differentiability. In the above formula, for example, the RHS requires that $v(\mathbf{x})$ must be a twice differentiable function, while the LHS is well defined with the first-order partial derivative of $v(\mathbf{x})$. The additional requirement of differentiability has been shifted to $u(\mathbf{x})$.

Example 1.5 (Divergence theorem) Integrate $\int_S \mathbf{F} \cdot \mathbf{n} dS$, where \mathbf{F} is a vector field given as $\mathbf{F} = 2x\mathbf{e}_1 + y^2\mathbf{e}_2 + z^2\mathbf{e}_3$ and S is the area of the surface of unit sphere ($x^2 + y^2 + z^2 = 1$), whose unit normal vector is \mathbf{n} .

Solution Using the divergence theorem,

$$\begin{aligned} \int_S \mathbf{F} \cdot \mathbf{n} dS &= \iint_{\Omega} \nabla \cdot \mathbf{F} d\Omega \\ &= 2 \iint_{\Omega} (1 + y + z) d\Omega \\ &= 2 \iint_{\Omega} d\Omega + 2 \iint_{\Omega} y d\Omega + 2 \iint_{\Omega} z d\Omega \\ &= 2 \iint_{\Omega} d\Omega \\ &= \frac{8\pi}{3}. \end{aligned}$$

In the above equation, the integral of odd functions is zero because of symmetry in the domain. ■

1.3 Stress and Strain

In the elementary mechanics of materials or physics courses, stress is defined as force per unit area. While such a notion is useful and sufficient to analyze one-dimensional structures under a uniaxial state of stress, a complete understanding of the state of stress in a three-dimensional body requires a thorough understanding of the concept of stress at a point. Similarly, strain is defined as the change in length per original length of a one-dimensional body. However, the concept of strain at a point in a three-dimensional body is quite interesting and is required for a complete understanding of the deformation a solid undergoes. While stresses and strains are concepts developed by engineers for better understanding of the

physics of deformation of a solid, the relation between stresses and strains is phenomenological in the sense that it is something observed and described as a simplified theory. Robert Hooke [4] was the first to establish the linear relation between stresses and strains in an elastic body. Although he explained his theory for one-dimensional objects, his theory later became the generalized Hooke's law that relates the stresses and strains in three-dimensional elastic bodies.

1.3.1 Stress

Surface traction: Consider a solid subjected to external forces and in static equilibrium, as shown in Fig. 1.3. We are interested in the state of stress at a point, P , in the interior of the solid. We cut the body of the solid into two halves by passing an imaginary plane through P . The unit vector normal to the plane is denoted by \mathbf{n} [see Fig. 1.3b]. The left side of the body is in equilibrium because of the external forces, \mathbf{f}_1 , \mathbf{f}_2 , and \mathbf{f}_3 , and also the internal forces acting on the cut surface. Surface traction is defined as the internal force per unit area or the force intensity acting on the cut plane. In order to measure the intensity or traction, specifically at P , we consider the force, $\Delta\mathbf{F}$, acting over a small area, ΔA , that contains point, P . Then the surface traction, $\mathbf{t}^{(\mathbf{n})}$, acting at the point, P , is defined as

$$\mathbf{t}^{(\mathbf{n})} = \lim_{\Delta A \rightarrow 0} \frac{\Delta\mathbf{F}}{\Delta A}. \quad (1.32)$$

In Eq. (1.32), the right superscript (\mathbf{n}) , is used to denote the fact that this surface traction is defined on a plane whose normal is \mathbf{n} . It should be noted that at the same point P , the traction vector, \mathbf{t} , would be different on a different plane passing through P . It is clear from Eq. (1.32) that the units of the traction vector are force

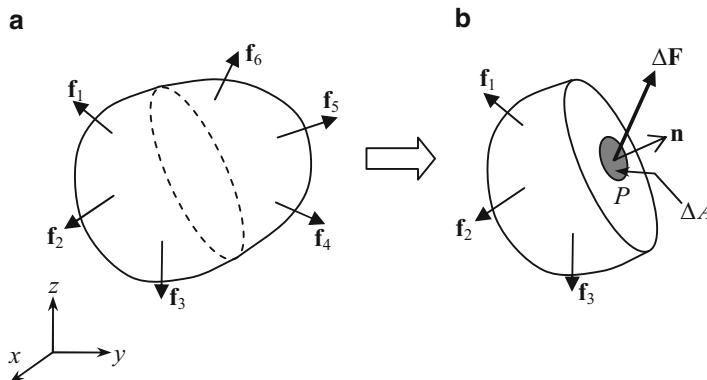
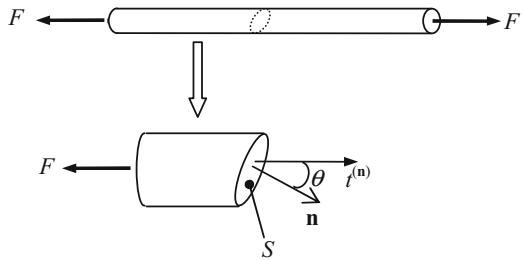


Fig. 1.3 Surface traction acting on a plane at a point

Fig. 1.4 Equilibrium of a uniaxial bar under axial force



per unit area, the same as that of pressure. Since $\mathbf{t}^{(n)}$ is a vector, one can resolve it into components and write it as

$$\mathbf{t}^{(n)} = t_1 \mathbf{e}_1 + t_2 \mathbf{e}_2 + t_3 \mathbf{e}_3. \quad (1.33)$$

Example 1.6 (Stress in an inclined surface) Consider a uniaxial bar with the cross-sectional area $A = 2 \times 10^{-4} \text{ m}^2$, as shown in Fig. 1.4. If an axial force, $F = 100 \text{ N}$, is applied to the bar, determine the surface traction on the plane whose normal is at an angle, θ , from the axial direction.

Solution To simplify the analysis, let us assume that the traction on the plane is uniform; i.e., the stresses are equally distributed over the cross section of the bar. In fact, this is the fundamental assumption in the analysis of bars. The force on the inclined plane, S , can be obtained by integrating the constant surface traction, $t^{(n)}$, over the plane, S . In this simple example, direction of the surface traction, $t^{(n)}$, must be opposite to that of the force, F . Since the member is in static equilibrium, the integral of the surface traction must be equal to the magnitude of the force, F :

$$F = \iint_S t^{(n)} dS = t \iint_S dS = t \frac{A}{\cos \theta},$$

$$\therefore t = \frac{F}{A} \cos \theta = 500 \cos \theta \frac{\text{N}}{\text{m}^2} = 500 \cos \theta \text{ Pa.}$$

Note that the unit of traction is Pascal (Pa or N/m²). It is clear that the surface traction depends on the direction of the normal to the plane. ■

Stress tensor and Cartesian components: Since the surface traction at a point varies depending on the direction of the normal to the plane, one can obtain an infinite number of traction vectors, $\mathbf{t}^{(n)}$, and the corresponding normal and shear stresses for a given state of stress at a point. Fortunately, the state of stress at a point can be completely characterized by defining traction vectors on three mutually perpendicular planes passing through the point. From the knowledge of $\mathbf{t}^{(n)}$ acting on three orthogonal planes, one can determine $\mathbf{t}^{(n)}$ on any arbitrary plane passing through the same point. For convenience, these planes are taken as the three planes that are normal to the x -, y -, and z -axes.

Let us denote the traction vector on the yz -plane, which is normal to the x -axis, as $\mathbf{t}^{(x)}$. The surface traction can be represented using its components that are parallel to the coordinate directions as

$$\mathbf{t}^{(x)} = t_1^{(x)} \mathbf{e}_1 + t_2^{(x)} \mathbf{e}_2 + t_3^{(x)} \mathbf{e}_3. \quad (1.34)$$

It may be noted that $t_1^{(x)}$ in Eq. (1.34) is the normal stress and $t_2^{(x)}$ and $t_3^{(x)}$ are the shear stresses in the y - and z -directions, respectively. In contemporary solid mechanics, the stress components in Eq. (1.34) are denoted by σ_{11} , σ_{12} and σ_{13} , where σ_{11} is the normal stress and σ_{12} and σ_{13} are components of shear stress. In this notation, the first subscript denotes the plane on which the stress component acts—in this case the plane normal to the x -axis or simply the x -plane—and the second subscript denotes the direction of the stress component. We can repeat this exercise by passing two more planes, normal to y - and z -axes, respectively, through the point, P . Thus, the surface tractions acting on the plane normal to y -plane will be σ_{21} , σ_{22} and σ_{23} . The stresses acting on the z -plane can be written as σ_{31} , σ_{32} and σ_{33} . In solid mechanics, the symbols, σ and τ , are often used for normal and shear stresses, respectively. However, in this text, the same symbol, σ , will be used for both normal and shear stresses. The stresses can be distinguished using their indices.

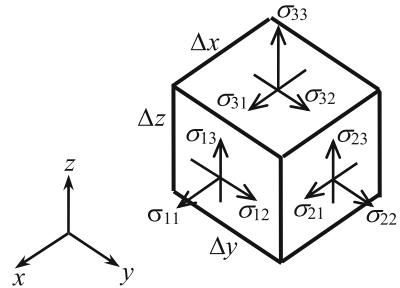
The stress components acting on the three planes can be depicted using a cube, as shown in Fig. 1.5. It must be noted that this cube is not a physical cube and, hence, has no dimensions. The six faces of the cube represent the three pairs of planes which are normal to the coordinate axes. The top face, for example, is the $+z$ -plane and then the bottom face is the $-z$ -plane or whose normal is in the $-z$ -direction. Note that the three visible faces of the cube in Fig. 1.5 represent the three positive planes, i.e., planes whose normal are the positive x -, y -, and z -axes. On these faces, all tractions are shown in the positive direction. For example, the stress component, σ_{23} , is the traction on the y -plane acting in the positive z -direction. By using these Cartesian stress components, the rank-2 stress tensor can be defined as

$$\boldsymbol{\sigma} = \sigma_{ij} \mathbf{e}_i \otimes \mathbf{e}_j, \quad (1.35)$$

where σ_{ij} represents the Cartesian components of a stress tensor, which is defined in the matrix form in Eq. (1.5). The stress tensor in Eq. (1.35) completely characterizes the state of stress at a given point.

The sign convention of stress is different from that of regular force vectors. Stress components, in addition to disclosing the direction of the force, contain information of the surface on which they are defined. A stress component is positive when both the surface normal and the stress component are either in the positive or in the negative coordinate direction. For example, if the surface normal is in the positive direction and the stress component is in the negative direction, then the stress component has a negative sign. Positive and negative normal stresses are called tensile and compressive stresses, respectively. The shear stress is positive when it is

Fig. 1.5 Stress components in Cartesian coordinate system



acting in the positive coordinate direction upon a positive face of the stress cube. The positive directions of all the stress components are shown in Fig. 1.5.

Symmetry of stress tensor: The nine components of the stress tensor can be reduced to six components using the symmetry property of the stress tensor. Consider the infinitesimal cube in Fig. 1.5, which is in equilibrium. In contrast to the previous section, let us assume that the cube has a very small finite dimension. The direction of the shear stress, σ_{12} , on the positive x -plane is in the positive y -direction, while on the positive y -plane, the direction of the shear stress, σ_{21} , is in the positive x -direction. As the body is in static equilibrium, the sum of the moments about the z -axis must be equal to zero; this implies that the shear stresses σ_{12} and σ_{21} must be equal to each other. The same is true for the moment equilibrium on x - and y -axes:

$$\sigma_{12} = \sigma_{21}, \quad \sigma_{23} = \sigma_{32}, \quad \sigma_{13} = \sigma_{31}.$$

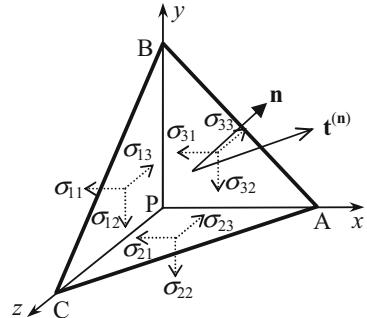
Therefore, the components of the stress tensor in Eq. (1.5) are revised using symmetry as

$$[\sigma_{ij}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix}. \quad (1.36)$$

Thus, we need only six components to fully represent the stress at a point. In some instances, stress at a point is written as a 6×1 pseudo vector as shown below:

$$\{\sigma\} = \left\{ \begin{array}{c} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{12} \\ \sigma_{23} \\ \sigma_{13} \end{array} \right\}. \quad (1.37)$$

Fig. 1.6 Surface traction and stress components acting on faces of an infinitesimal tetrahedron, at a given point P



Cauchy's Lemma: Knowledge of the six stress components is necessary in order to determine the components of the surface traction, $\mathbf{t}^{(n)}$, acting on an arbitrary plane with a normal vector, \mathbf{n} . Let \mathbf{n} be the unit normal vector of the plane on which we want to determine the surface traction. For convenience, we choose P as the origin of the coordinate system, as shown in Fig. 1.6, and consider a plane parallel to the intended plane which passes at an infinitesimally small distance, h , away from P . Note that the normal to the face, ABC, is also \mathbf{n} . We will calculate the tractions on the plane formed by ABC and then take the limit, as h approaches zero. We will consider the equilibrium of the tetrahedron, PABC. If A is the area of the triangle, ABC, then the areas of triangles PAB, PBC, and PAC are given by An_z , An_x , and An_y , respectively. Let $\mathbf{t}^{(n)} = t_1^{(n)}\mathbf{e}_1 + t_2^{(n)}\mathbf{e}_2 + t_3^{(n)}\mathbf{e}_3$ be the surface traction acting on the face, ABC.

From the definition of surface traction in Eq. (1.32), the force on the surface can be calculated by multiplying the stresses with the surface area. Since the tetrahedron should be in equilibrium, the sum of the forces acting on its surfaces should be equal to zero. The force balance in the x -direction yields

$$\sum F_1 = t_1^{(n)}A - \sigma_{11}An_1 - \sigma_{21}An_2 - \sigma_{31}An_3 = 0.$$

In the above equation, we have assumed that the stresses acting on a surface are uniform; this will not be true if the size of the tetrahedron is not small. However, the tetrahedron is infinitesimally small, which is the case as h approaches zero. By dividing the above equation by A , we obtain the following relation:

$$t_1^{(n)} = \sigma_{11}n_1 + \sigma_{21}n_2 + \sigma_{31}n_3.$$

Similarly, the force balance in the y - and z -directions yields

$$\begin{aligned} t_2^{(n)} &= \sigma_{12}n_1 + \sigma_{22}n_2 + \sigma_{32}n_3, \\ t_3^{(n)} &= \sigma_{13}n_1 + \sigma_{23}n_2 + \sigma_{33}n_3. \end{aligned}$$

From the above equations, it is clear that the surface traction acting on the surface whose normal is \mathbf{n} can be determined if the six stress components are available. By using tensor notation, we can write the above equations as

$$\mathbf{t}^{(\mathbf{n})} = \mathbf{n} \cdot \boldsymbol{\sigma}. \quad (1.38)$$

Due to the symmetry of the stress tensor, the above relation is equivalent to $\mathbf{t}^{(\mathbf{n})} = \boldsymbol{\sigma} \cdot \mathbf{n}$. The surface traction, $\mathbf{t}^{(\mathbf{n})}$, remains unchanged for all surfaces which pass through the point, P , and have the same normal vector, \mathbf{n} , at P ; i.e., surfaces which have a common tangent at P will have the same surface traction. This means that the stress vector is only a function of the normal vector, \mathbf{n} , and is not influenced by the curvature of the internal surfaces. From this observation, Cauchy's Lemma [5], also called the Cauchy reciprocal theorem, states that the surface tractions acting on opposite sides of the same surface are equal in magnitude and opposite in direction, i.e.,

$$\mathbf{t}^{(\mathbf{n})} = -\mathbf{t}^{(-\mathbf{n})}, \quad (1.39)$$

which can easily be shown using Eq. (1.38).

Normal stress and shear stress: The surface traction, $\mathbf{t}^{(\mathbf{n})}$, defined by Eq. (1.38) does not generally act in the direction of \mathbf{n} ; i.e., $\mathbf{t}^{(\mathbf{n})}$ and \mathbf{n} are not necessarily parallel to each other. Thus, we can decompose the surface traction into two components, one parallel to \mathbf{n} and the other perpendicular to \mathbf{n} , which will lie on the plane. The component normal to the plane or parallel to \mathbf{n} is called the normal stress and is denoted by σ_n . The other component parallel to the plane or perpendicular to \mathbf{n} is called the shear stress and is denoted by τ_n .

The normal stress can be obtained from the inner product of $\mathbf{t}^{(\mathbf{n})}$ and \mathbf{n} (see Fig. 1.7) as

$$\sigma_n = \mathbf{t}^{(\mathbf{n})} \cdot \mathbf{n} = \mathbf{n} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} \quad (1.40)$$

and shear stress can be calculated from the relation

$$\tau_n = \sqrt{\|\mathbf{t}^{(\mathbf{n})}\|^2 - \sigma_n^2}. \quad (1.41)$$

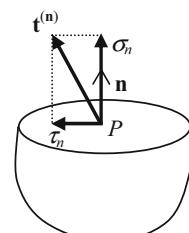


Fig. 1.7 Normal and shear stresses at a point P

Example 1.7 (Normal and shear stresses on a plate) The state of stress at a particular point in the xyz coordinate system is given by the following stress components:

$$\boldsymbol{\sigma} = \begin{bmatrix} 3 & 7 & -7 \\ 7 & 4 & 0 \\ -7 & 0 & 2 \end{bmatrix}.$$

Determine the normal and shear stresses on a surface passing through the point and parallel to the plane given by the equation $4x - 4y + 2z = 2$.

Solution To determine the surface traction, $\mathbf{t}^{(n)}$, it is necessary to determine the unit vector normal to the plane. From solid geometry, the normal to the plane is found to be in the direction of $\mathbf{d} = \{4, -4, 2\}^T$ with a magnitude of $\|\mathbf{d}\| = 6$. Thus, the unit normal vector becomes

$$\mathbf{n} = \left\{ \frac{2}{3}, -\frac{2}{3}, \frac{1}{3} \right\}^T.$$

The surface traction can be obtained as

$$\mathbf{t}^{(n)} = \boldsymbol{\sigma} \cdot \mathbf{n} = \frac{1}{3} \begin{bmatrix} 3 & 7 & -7 \\ 7 & 4 & 0 \\ -7 & 0 & 2 \end{bmatrix} \cdot \begin{Bmatrix} 2 \\ -2 \\ 1 \end{Bmatrix} = \begin{Bmatrix} -5 \\ 2 \\ -4 \end{Bmatrix}.$$

By using Eqs. (1.40) and (1.41), the normal and shear stresses can be obtained as

$$\begin{aligned} \sigma_n &= \mathbf{t}^{(n)} \cdot \mathbf{n} = -5 \times \frac{2}{3} - 2 \times \frac{2}{3} - 4 \times \frac{1}{3} = -6, \\ \|\mathbf{t}^{(n)}\|^2 &= 5^2 + 2^2 + 4^2 = 45, \\ \tau_n &= \sqrt{\|\mathbf{t}^{(n)}\|^2 - 6^2} = 3. \end{aligned}$$

■

Mean stress and stress deviator: The stress in Eq. (1.36) can be decomposed into hydrostatic pressure and deviatoric stress. The former is related to the change in volume, while the latter is related to the change in shape. The hydrostatic pressure, often called the mean stress, can be defined using the trace of the stress tensor as

$$p = \sigma_m = \frac{1}{3} \text{tr}(\boldsymbol{\sigma}) = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}). \quad (1.42)$$

Note that the hydrostatic pressure is invariant on coordinate transformation in Eq. (1.19), that is, for $\boldsymbol{\sigma}$ in xyz coordinates and $\boldsymbol{\sigma}'$ for $x'y'z'$ coordinates, $\text{tr}(\boldsymbol{\sigma}) = \text{tr}(\boldsymbol{\sigma}')$. Therefore, the mean stress has the property of frame indifference.

On the other hand, the stress deviator is defined by subtracting the mean stress from the original stress tensor as

$$\mathbf{s} = \boldsymbol{\sigma} - \sigma_m \mathbf{1} = \begin{bmatrix} \sigma_{11} - \sigma_m & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} - \sigma_m & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} - \sigma_m \end{bmatrix}. \quad (1.43)$$

Note that $\text{tr}(\mathbf{s}) = 0$. Therefore, the stress deviator is called trace-free. The mean stress and stress deviator are important in representing the plastic behavior of a material beyond the yield point.

For a formal definition, the stress deviator can be defined by contracting the original stress with the unit deviatoric tensor of rank-4:

$$\mathbf{s} = \mathbf{I}_{\text{dev}} : \boldsymbol{\sigma},$$

where \mathbf{I}_{dev} is defined as

$$\mathbf{I}_{\text{dev}} = \mathbf{I} - \frac{1}{3} \mathbf{1} \otimes \mathbf{1}, \quad (1.44)$$

where \mathbf{I} is a unit symmetric tensor of rank-4, which is defined as $\mathbf{I}_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$. Note that since \mathbf{I}_{dev} is trace-free, it is easy to show that $\mathbf{I}_{\text{dev}} : \mathbf{1} = \mathbf{0}$. In addition, the unit deviatoric tensor preserves a deviatoric tensor, that is, $\mathbf{I}_{\text{dev}} : \mathbf{s} = \mathbf{s}$ for a deviatoric rank-2 tensor \mathbf{s} .

Principal stresses: The normal and shear stresses acting on a plane, which passes through a given point in a solid, change as the orientation of the plane is changed. Then a natural question is: Is there a plane on which the normal stress becomes the maximum? Similarly, we would also like to find the plane on which the shear stress attains a maximum. These questions have significance in predicting the failure of the material at a point. In the following, we will provide some answers to the above questions, without furnishing the proofs. The interested reader is referred to books on continuum mechanics, e.g., Malvern [6] or Boresi [7] for a more detailed treatment of the subject.

It can be shown that, at every point in a solid, there are at least three mutually perpendicular planes on which the normal stress attains an extremum (maximum or minimum) value. On all of these planes, the shear stresses vanish. Thus, the traction vector, $\mathbf{t}^{(n)}$, will be parallel to the normal vector, \mathbf{n} , on these planes, i.e., $\mathbf{t}^{(n)} = \sigma_n \mathbf{n}$. Of these three planes, one plane corresponds to the global maximum value of the normal stress and another corresponds to the global minimum. The third plane will carry the intermediate normal stress. These special normal stresses are called the *principal stresses* at that point, the planes on which they act are called the principal stress planes and the corresponding normal vectors are called the principal stress directions. The principal stresses are denoted by σ_1 , σ_2 , and σ_3 , such that $\sigma_1 \geq \sigma_2 \geq \sigma_3$.

Based on the above observations, the principal stresses can be calculated, as follows. When the normal direction to a plane is the principal direction, the surface normal and the surface traction are in the same direction, i.e., $(\mathbf{t}^{(n)} \parallel \mathbf{n})$. Thus, the surface traction on a plane can be represented by the product of the normal stress, σ_n , and the normal vector, \mathbf{n} , as

$$\mathbf{t}^{(n)} = \sigma_n \mathbf{n}. \quad (1.45)$$

By combining Eq. (1.45) with Eq. (1.38) for the surface traction, we obtain

$$\boldsymbol{\sigma} \cdot \mathbf{n} = \sigma_n \mathbf{n}. \quad (1.46)$$

Equation (1.46) represents the eigenvalue problem, where σ_n is the eigenvalue and \mathbf{n} is the corresponding eigenvector. Equation (1.46) can be rearranged as

$$(\boldsymbol{\sigma} - \sigma_n \mathbf{1}) \cdot \mathbf{n} = \mathbf{0}. \quad (1.47)$$

In the component form, the above equation can be written as

$$\begin{bmatrix} \sigma_{11} - \sigma_n & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} - \sigma_n & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} - \sigma_n \end{bmatrix} \begin{Bmatrix} n_1 \\ n_2 \\ n_3 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \\ 0 \end{Bmatrix}. \quad (1.48)$$

Note that a solution, $\mathbf{n} = \mathbf{0}$, is not only a trivial solution to the above equation, but also not physically possible as $\|\mathbf{n}\|$ must be equal to unity. The above set of linear simultaneous equations will have a nontrivial physically meaningful solution if and only if the determinant of the coefficient matrix is zero, i.e.,

$$\begin{vmatrix} \sigma_{11} - \sigma_n & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} - \sigma_n & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} - \sigma_n \end{vmatrix} = 0. \quad (1.49)$$

By expanding this determinant, we obtain the following cubic equation in terms of σ_n :

$$\sigma_n^3 - I_1 \sigma_n^2 + I_2 \sigma_n - I_3 = 0, \quad (1.50)$$

where

$$\begin{aligned} I_1 &= \sigma_{11} + \sigma_{22} + \sigma_{33}, \\ I_2 &= \begin{vmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{vmatrix} + \begin{vmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{23} & \sigma_{33} \end{vmatrix} + \begin{vmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{13} & \sigma_{33} \end{vmatrix} \\ &= \sigma_{11}\sigma_{22} + \sigma_{22}\sigma_{33} + \sigma_{33}\sigma_{11} - \sigma_{12}^2 - \sigma_{23}^2 - \sigma_{13}^2, \\ I_3 &= |\boldsymbol{\sigma}| = \sigma_{11}\sigma_{22}\sigma_{33} + 2\sigma_{12}\sigma_{23}\sigma_{13} - \sigma_{11}\sigma_{23}^2 - \sigma_{22}\sigma_{13}^2 - \sigma_{33}\sigma_{12}^2. \end{aligned} \quad (1.51)$$

In the above equation, I_1 , I_2 , and I_3 are the three invariants of the stress, which can be shown to be independent of the coordinate system. The three roots of the cubic equation (1.50) correspond to the three principal stresses. We will denote them by σ_1 , σ_2 , and σ_3 in the order of $\sigma_1 \geq \sigma_2 \geq \sigma_3$.

Once the principal stresses have been computed, we can substitute them, one at a time, into Eq. (1.48) to obtain \mathbf{n} . We will get a principal direction that will be denoted as \mathbf{n}^1 , \mathbf{n}^2 , and \mathbf{n}^3 , which each corresponds to a principal value. Note that \mathbf{n} is a unit vector, and hence its components must satisfy the following relation:

$$\|\mathbf{n}^i\|^2 = (n_1^i)^2 + (n_2^i)^2 + (n_3^i)^2 = 1, \quad i = 1, 2, 3. \quad (1.52)$$

It can be shown that the planes on which the principal stresses act are mutually perpendicular. Let us consider any two principal directions \mathbf{n}^i and \mathbf{n}^j , with $i \neq j$. If σ_i and σ_j are the corresponding principal stresses, then they satisfy the following equations:

$$\begin{aligned} \boldsymbol{\sigma} \cdot \mathbf{n}^i &= \sigma_i \mathbf{n}^i, \\ \boldsymbol{\sigma} \cdot \mathbf{n}^j &= \sigma_j \mathbf{n}^j. \end{aligned} \quad (1.53)$$

By multiplying the first equation by \mathbf{n}^j and the second equation by \mathbf{n}^i , we obtain

$$\begin{aligned} \mathbf{n}^j \cdot \boldsymbol{\sigma} \cdot \mathbf{n}^i &= \sigma_i \mathbf{n}^j \cdot \mathbf{n}^i, \\ \mathbf{n}^i \cdot \boldsymbol{\sigma} \cdot \mathbf{n}^j &= \sigma_j \mathbf{n}^i \cdot \mathbf{n}^j. \end{aligned} \quad (1.54)$$

Considering the symmetry of $\boldsymbol{\sigma}$ and the rule for inner product, one can show that $\mathbf{n}^j \cdot \boldsymbol{\sigma} \cdot \mathbf{n}^i = \mathbf{n}^i \cdot \boldsymbol{\sigma} \cdot \mathbf{n}^j$. Then subtracting the first equation from the second in Eq. (1.54), we obtain

$$(\sigma_i - \sigma_j) \mathbf{n}^i \cdot \mathbf{n}^j = 0. \quad (1.55)$$

This implies that if the principal stresses are distinct, i.e., $\sigma_i \neq \sigma_j$, then

$$\mathbf{n}^i \cdot \mathbf{n}^j = 0, \quad (1.56)$$

which means that \mathbf{n}^i and \mathbf{n}^j are orthogonal. The three planes, on which the principal stresses act, are mutually perpendicular.

There are three different possibilities for principal stresses and directions:

- (a) σ_1, σ_2 , and σ_3 are distinct \Rightarrow principal stress directions are three unique mutually orthogonal unit vectors.
- (b) $\sigma_1 = \sigma_2 \neq \sigma_3 \Rightarrow \mathbf{n}^3$ is a unique principal stress direction, and any two orthogonal directions on the plane that is perpendicular to \mathbf{n}^3 are the other principal directions.
- (c) $\sigma_1 = \sigma_2 = \sigma_3 \Rightarrow$ any three orthogonal directions are principal stress directions. This state of stress is called hydrostatic or isotropic state of stress.

Example 1.8 (Principal stresses and principal directions) For the Cartesian stress components given below, determine the principal stresses and principal directions.

$$\boldsymbol{\sigma} = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{bmatrix}.$$

Solution Setting the determinant of the coefficient matrix to zero yields

$$\begin{vmatrix} 3 - \sigma_n & 1 & 1 \\ 1 & -\sigma_n & 2 \\ 1 & 2 & -\sigma_n \end{vmatrix} = 0.$$

By expanding the determinant, we obtain the following characteristic equation:

$$(3 - \sigma_n)(\sigma_n^2 - 4) - (-\sigma_n - 2) + (2 + \sigma_n) = -(\sigma_n + 2)(\sigma_n - 1)(\sigma_n - 4) = 0.$$

Three roots of the above equation are the principal stresses. They are

$$\sigma_1 = 4, \quad \sigma_2 = 1, \quad \sigma_3 = -2.$$

For the case when $\sigma_n = \sigma_3 = -2$, we may obtain the following simultaneous equations, by using the form of Eq. (1.48):

$$\begin{aligned} 5n_x + n_y + n_z &= 0, \\ n_x + 2n_y + 2n_z &= 0, \\ n_x + 2n_y + 2n_z &= 0. \end{aligned}$$

We note that the three equations are not independent; in fact, the second and third equations are identical. From the first two equations, we can obtain the following ratios between components:

$$n_x : n_y : n_z = 0 : -1 : 1.$$

By using Eq. (1.52), a unique solution of the following form can be obtained:

$$\mathbf{n}^{(3)} = \frac{1}{\sqrt{2}} \begin{Bmatrix} 0 \\ -1 \\ 1 \end{Bmatrix}.$$

The same process can be repeated for σ_1 and σ_2 to obtain the following two principal directions:

$$\mathbf{n}^{(1)} = \frac{1}{\sqrt{6}} \begin{Bmatrix} -2 \\ -1 \\ -1 \end{Bmatrix}, \quad \mathbf{n}^{(2)} = \frac{1}{\sqrt{3}} \begin{Bmatrix} 1 \\ -1 \\ -1 \end{Bmatrix}.$$

Note that all principal directions are mutually perpendicular. ■

1.3.2 Strain

When a solid is subjected to forces, it deforms. A measure of the deformation is provided by strains. Imagine an infinitesimal line segment in an arbitrary direction which passes through a point in a solid. As the solid deforms, the length of the line segment changes. The strain, specifically the normal strain, in the original direction of the line segment is defined as the change in length divided by the original length. However, the strain at the same point will be different in different directions. In the following, the concept of strain in a three-dimensional body is developed.

Figure 1.8 shows a body before and after deformation. Let the points P , Q , and R , in the undeformed body move to P' , Q' , and R' , respectively, after deformation. For the convenience of notation, the three coordinate directions are denoted by

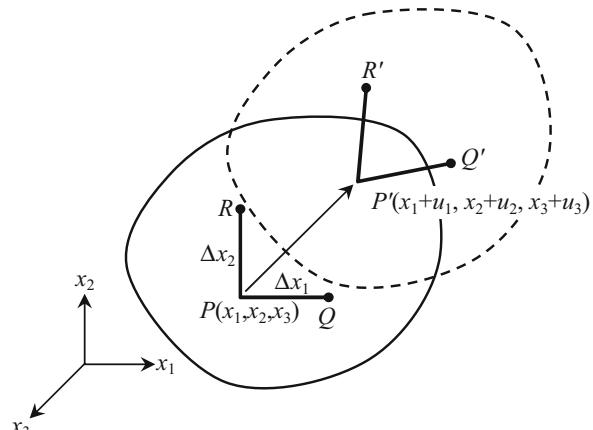


Fig. 1.8 Deformation of line segments

$x_1x_2x_3$ coordinates instead of using the xyz coordinates. The displacement of P can be represented by three displacement components, u_1 , u_2 , and u_3 in the x_1 -, x_2 -, and x_3 -directions. Thus, the coordinates of P' are $(x_1 + u_1, x_2 + u_2, x_3 + u_3)$. The functions $u_1(x_1, x_2, x_3)$, $u_2(x_1, x_2, x_3)$, and $u_3(x_1, x_2, x_3)$ are components of a vector field that is referred to as the deformation field or the displacement field. The displacements of the point, Q , will be slightly different from that of P . They can be written as

$$\begin{aligned} u_1^Q &= u_1 + \frac{\partial u_1}{\partial x_1} \Delta x_1, \\ u_2^Q &= u_2 + \frac{\partial u_2}{\partial x_1} \Delta x_1, \\ u_3^Q &= u_3 + \frac{\partial u_3}{\partial x_1} \Delta x_1. \end{aligned} \quad (1.57)$$

Similarly, displacements of the point, R , are

$$\begin{aligned} u_1^R &= u_1 + \frac{\partial u_1}{\partial x_2} \Delta x_2, \\ u_2^R &= u_2 + \frac{\partial u_2}{\partial x_2} \Delta x_2, \\ u_3^R &= u_3 + \frac{\partial u_3}{\partial x_2} \Delta x_2. \end{aligned} \quad (1.58)$$

The coordinates of P , Q , and R before and after deformation are as follows:

$$\begin{aligned} P &: (x_1, x_2, x_3), \\ Q &: (x_1 + \Delta x_1, x_2, x_3), \\ R &: (x_1, x_1 + \Delta x_2, x_3), \\ P' &: (x_1 + u_1^P, x_2 + u_2^P, x_3 + u_3^P) = (x_1 + u_1, x_2 + u_2, x_3 + u_3), \\ Q' &: \left(x_1 + \Delta x_1 + u_1^Q, x_2 + u_2^Q, x_3 + u_3^Q \right) \\ &= \left(x_1 + \Delta x_1 + u_1 + \frac{\partial u_1}{\partial x_1} \Delta x_1, x_2 + u_2 + \frac{\partial u_2}{\partial x_1} \Delta x_1, x_3 + u_3 + \frac{\partial u_3}{\partial x_1} \Delta x_1 \right), \\ R' &: \left(x_1 + u_1^R, x_2 + \Delta x_2 + u_2^R, x_3 + u_3^R \right) \\ &= \left(x_1 + u_1 + \frac{\partial u_1}{\partial x_2} \Delta x_2, x_2 + \Delta x_2 + u_2 + \frac{\partial u_2}{\partial x_2} \Delta x_2, x_3 + u_3 + \frac{\partial u_3}{\partial x_2} \Delta x_2 \right). \end{aligned}$$

The length of the line segment $P'Q'$ can be calculated as

$$P'Q' = \sqrt{\left(x_1^{P'} - x_1^{Q'} \right)^2 + \left(x_2^{P'} - x_2^{Q'} \right)^2 + \left(x_3^{P'} - x_3^{Q'} \right)^2}. \quad (1.59)$$

By substituting for the coordinates of P' and Q' , we obtain

$$\begin{aligned} P'Q' &= \Delta x_1 \sqrt{\left(1 + \frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_1}\right)^2 + \left(\frac{\partial u_3}{\partial x_1}\right)^2} \\ &= \Delta x_1 \left(1 + 2\frac{\partial u_1}{\partial x_1} + \left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_1}\right)^2 + \left(\frac{\partial u_3}{\partial x_1}\right)^2\right)^{1/2} \\ &\approx \Delta x_1 \left(1 + \frac{\partial u_1}{\partial x_1} + \frac{1}{2}\left(\frac{\partial u_1}{\partial x_1}\right)^2 + \frac{1}{2}\left(\frac{\partial u_2}{\partial x_1}\right)^2 + \frac{1}{2}\left(\frac{\partial u_3}{\partial x_1}\right)^2\right). \end{aligned} \quad (1.60)$$

It may be noted that we have used a two-term binomial expansion in deriving an approximate expression for the change in length. In this chapter, we will consider only small deformations such that all deformation gradients are very small when compared to unity, e.g., $\partial u_1/\partial x_1 \ll 1$, $\partial u_2/\partial x_1 \ll 1$. Then we can neglect the higher-order terms in Eq. (1.60) to obtain

$$P'Q' \approx \Delta x_1 \left(1 + \frac{\partial u_1}{\partial x_1}\right). \quad (1.61)$$

Now we invoke the definition of normal strain as the ratio of the change in length to the original length in order to derive the expression for strain as

$$\varepsilon_{11} = \frac{P'Q' - PQ}{PQ} = \frac{\partial u_1}{\partial x_1}. \quad (1.62)$$

Thus, the normal strain, ε_{11} , at a point can be defined as the change in length per unit length of an infinitesimally long line segment, originally parallel to the x_1 -axis. Similarly, we can derive normal strains in the x_2 - and x_3 -directions as

$$\varepsilon_{22} = \frac{\partial u_2}{\partial x_2}, \quad \varepsilon_{33} = \frac{\partial u_3}{\partial x_3}. \quad (1.63)$$

The shear strain, say γ_{12} , is defined as the change in angle between a pair of infinitesimal line segments that were originally parallel to the x_1 - and x_2 -axes. From Fig. 1.8, the angle between PQ and $P'Q'$ can be derived as

$$\theta_1 = \frac{x_2^{Q'} - x_2^Q}{\Delta x_1} = \frac{\partial u_2}{\partial x_1}. \quad (1.64)$$

Similarly, the angle between PR and $P'R'$ is

$$\theta_2 = \frac{x_1^{R'} - x_1^R}{\Delta x_2} = \frac{\partial u_1}{\partial x_2}. \quad (1.65)$$

Using the aforementioned definition of shear strain,

$$\gamma_{12} = \theta_1 + \theta_2 = \frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}. \quad (1.66)$$

Similarly, we can derive shear strains in the x_2x_3 - and x_3x_2 -planes as

$$\begin{aligned} \gamma_{23} &= \frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2}, \\ \gamma_{13} &= \frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3}. \end{aligned} \quad (1.67)$$

The shear strains, γ_{12} , γ_{23} , and γ_{13} , are called engineering shear strains. From the definition in Eq. (1.66), it is clear that $\gamma_{12} = \gamma_{21}$. We define tensorial shear strains as

$$\begin{aligned} \varepsilon_{12} &= \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right), \\ \varepsilon_{23} &= \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right), \\ \varepsilon_{13} &= \frac{1}{2} \left(\frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \right). \end{aligned} \quad (1.68)$$

It may be noted that the tensorial shear strains are one-half of the corresponding engineering shear strains. It can be shown that the normal strains and the tensorial shear strains transform from one coordinate system to another by following the tensor transformation rule in Eq. (1.19).

In the general three-dimensional case, the strain tensor can be defined using the dyadic product, as

$$\boldsymbol{\varepsilon} = \varepsilon_{ij} \mathbf{e}_i \otimes \mathbf{e}_j, \quad (1.69)$$

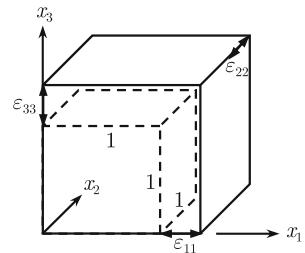
where the components of the strain tensor are defined as

$$[\boldsymbol{\varepsilon}] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{12} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{13} & \varepsilon_{23} & \varepsilon_{33} \end{bmatrix}. \quad (1.70)$$

As is clear from the definition in Eq. (1.68), the strain tensor is symmetric. Similar to the stress tensor, the symmetric strain tensor can be represented as a pseudo vector

$$\{\boldsymbol{\varepsilon}\} = \begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ 2\varepsilon_{12} \\ 2\varepsilon_{23} \\ 2\varepsilon_{13} \end{Bmatrix} = \begin{Bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \gamma_{12} \\ \gamma_{23} \\ \gamma_{13} \end{Bmatrix}. \quad (1.71)$$

Fig. 1.9 Volume change of a unit cube



The six components of strain completely define the deformation at a point. Since strain is a tensor, it has properties similar to the stress tensor. For example, the normal strain in any arbitrary direction at that point and also the shear strain in any arbitrary plane passing through the point can be calculated using the same process as in the stress tensor. Similarly, the transformation of strain, principal strains, and corresponding principal strain directions can be determined using the procedures we described for stresses.

Decomposition of strain: The strain tensor can be decomposed into a volumetric and a distortional part. The former changes the volume of an infinitesimal element, while the latter changes the shape of the element. For volumetric strain, consider a unit cube in Fig. 1.9, which undergoes three normal strains (ϵ_{11} , ϵ_{22} , and ϵ_{33}). Since there is no shape change, all shear strains are zero, for now. Then, the deformed volume

$$\epsilon_V = \frac{V - V_0}{V_0} = (1 + \epsilon_{11})(1 + \epsilon_{22})(1 + \epsilon_{33}) - 1 \approx \epsilon_{11} + \epsilon_{22} + \epsilon_{33}.$$

Since the magnitudes of strain components are small, the higher-order terms may be ignored. Therefore, the volumetric strain can be defined as

$$\epsilon_V = \epsilon_{11} + \epsilon_{22} + \epsilon_{33} = \epsilon_{kk}. \quad (1.72)$$

Or, in tensor notation, $\epsilon_V = \mathbf{1} : \boldsymbol{\epsilon}$, with $\mathbf{1}$ being the rank-2 identity tensor. Note that the volumetric strain is a scalar that is three times the value of the average normal strain.

The deviatoric part of strain can be defined by subtracting the average normal strain from the diagonal components of the original strain. The deviatoric strain tensor can be defined as

$$\begin{aligned} \boldsymbol{\epsilon} &= \boldsymbol{\epsilon} - \frac{1}{3}\epsilon_V\mathbf{1}, \\ \epsilon_{ij} &= \epsilon_{ij} - \frac{1}{3}\epsilon_V\delta_{ij}. \end{aligned} \quad (1.73)$$

For a formal definition, the deviatoric strain can be defined by contracting the original strain with the unit deviatoric tensor of rank-4.

$$\mathbf{e} = \mathbf{I}_{\text{dev}} : \boldsymbol{\varepsilon},$$

where \mathbf{I}_{dev} is the unit deviatoric tensor of rank-4, defined in Eq. (1.44).

1.3.3 Stress–Strain Relationship

Finding a relationship between the loads acting on a structure and its deflection has been of great interest to scientists since the seventeenth century [8]. Robert Hooke, Jacob Bernoulli, and Leonard Euler are some of the pioneers who developed various theories to explain the bending of beams and stretching of bars. Forces applied to a solid create stresses within the body in order to satisfy equilibrium. These stresses also cause deformation or strains. Accumulation of strains over the volume of a body manifests as deflections or a gross deformation of the body. Hence, it is clear that a fundamental knowledge of the relationship between stresses and strains is necessary in order to understand the global behavior. Navier tried to explain deformations considering the forces between neighboring particles in a body, as they tend to separate and come closer. Later this approach was abandoned in favor of Cauchy's stresses and strains. Robert Hooke was the first one to propose the linear uniaxial stress–strain relation, which states that the stress is proportional to strain. Later, the general relation between the six components of strains and stresses called the generalized Hooke's law was developed. The generalized Hooke's law states that each component of stress is a linear combination of strains. It should be mentioned that stress–strain relations are called phenomenological models or theories as they are based on commonly observed behavior of materials which may be verified through experimentation. Only recently, with the advancement of computers and computational techniques, have we started to model the behavior of materials based on the first principles or based on the fundamentals of atomistic behavior. This new field of study is called computational materials, and it involves techniques such as molecular dynamic simulations and multiscale modeling. Stress–strain relations are also called constitutive relations as they describe the constitution of the material.

A cylindrical test specimen is loaded along its axis as shown in Fig. 1.10. This type of loading ensures that the specimen is subjected to a uniaxial state of stress. If the stress–strain relation of the uniaxial tension test in Fig. 1.10 is plotted, then a typical ductile material may show a behavior as in Fig. 1.11. The explanation of the terms in the figure is summarized in Table 1.2.

After the material yields, the shape of the structure permanently changes. Hence, many engineering structures are designed such that the maximum stress is smaller than the yield stress of the material. Under this range of the stress, the stress–strain



Fig. 1.10 Uniaxial tension test

Fig. 1.11 Stress–strain diagram for a typical ductile material in tension

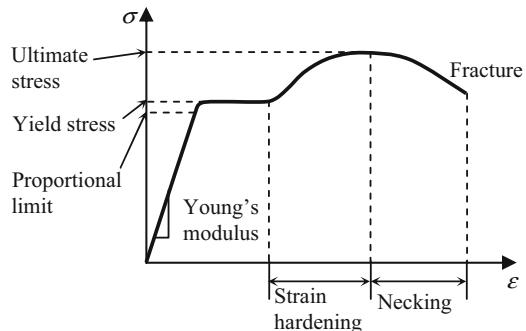


Table 1.2 Explanations of uniaxial tension test

Terms	Explanation
Proportional limit	The greatest stress for which the stress is still proportional to the strain
Elastic limit	The greatest stress without resulting in any permanent strain on release of stress
Young's modulus	Slope of the linear portion of the stress–strain curve
Yield stress	The stress required to produce 0.2 % plastic strain
Strain hardening	A region where more stress is required to deform the material
Ultimate stress	The maximum stress the material can resist
Necking	Cross section of the specimen reduces during deformation

relation can be linearly approximated. The main interest of this text is to study the behavior of materials beyond the linear relation. However, in this and the following sections, we will focus on the linear relationship between stress and strain.

Stress–strain relationship for isotropic material: The one-dimensional stress–strain relation can be extended to the three-dimensional state of stress. The linear elastic material means the relationship between the stress and strain is linear. Since both stress and strain tensors are rank-2, the relationship between them requires a rank-4 tensor. For a general linear elastic material, the stress–strain relationship can be written as

$$\boldsymbol{\sigma} = \mathbf{D} : \boldsymbol{\epsilon}, \quad \sigma_{ij} = D_{ijkl} \epsilon_{kl}. \quad (1.74)$$

The rank-4 tensor, \mathbf{D} , is called the elasticity tensor. A general rank-4 tensor in three dimensions has 81 components. Since the stress and strain tensors are symmetric, it can be shown that \mathbf{D} must be symmetric¹; hence, the number of independent coefficients or elastic constants for an anisotropic material is only 21.

¹ More specifically, \mathbf{D} has major and minor symmetry.

Many composite materials that are naturally occurring, such as wood or bone, and man-made materials, such as fiber-reinforced composites, can be modeled as an orthotropic material with nine independent elastic constants. Some composites are transversely isotropic and require only five independent elastic constants. For isotropic materials, the 21 constants in the elasticity tensor can be expressed in terms of two independent constants called engineering elastic constants. Therefore, the elasticity tensor for an isotropic material can be written as

$$\mathbf{D} = \lambda \mathbf{1} \otimes \mathbf{1} + 2\mu \mathbf{I}, \quad (1.75)$$

where λ and μ are Lame's constants. In fact, μ is also called the shear modulus. The Lame's constants are related to the nominal engineering elastic constants: Young's modulus, E , and Poisson's ratio, ν , as

$$\begin{aligned} \nu &= \frac{\lambda}{2(\lambda + \mu)}, & E &= \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}, \\ \lambda &= \frac{E\nu}{(1 + \nu)(1 - 2\nu)}, & \mu &= \frac{E}{2(1 + \nu)}. \end{aligned} \quad (1.76)$$

Using index notation, the components of the rank-4 elasticity tensor can be written as $D_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk})$. As the stress and strain tensors are decomposed into volumetric and deviatoric parts, the elasticity tensor can also be decomposed as

$$\mathbf{D} = \left(\lambda + \frac{2}{3}\mu \right) \mathbf{1} \otimes \mathbf{1} + 2\mu \mathbf{I}_{\text{dev}}, \quad (1.77)$$

where \mathbf{I}_{dev} is the unit deviatoric tensor of rank-4 (see Eq. (1.44)). The advantage of decomposing volumetric and deviatoric parts is that it is possible to make the stress-strain relationship from the decomposed parts:

$$p = \frac{1}{3} \text{tr}(\boldsymbol{\sigma}) = \left(\lambda + \frac{2}{3}\mu \right) \text{tr}(\boldsymbol{\epsilon}), \quad (1.78)$$

$$\mathbf{s} = 2\mu \mathbf{I}_{\text{dev}} : \boldsymbol{\epsilon} = 2\mu \boldsymbol{\epsilon}. \quad (1.79)$$

Stress-strain relationship using vector notation: Although the tensor notation is clear, it can be sometimes cumbersome, especially when implementing it in a computer code. Therefore, it would be desirable to express the stress-strain relationship using the pseudo vectors of stress and strain in Eqs. (1.37) and (1.71). When the stress-strain relation is linear, the relationship can be written in the matrix form as

$$\{\boldsymbol{\sigma}\} = [\mathbf{D}] \cdot \{\boldsymbol{\epsilon}\}, \quad (1.80)$$

where $[\mathbf{D}]$ is the elasticity matrix, defined as

$$[\mathbf{D}] = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2}-\nu & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2}-\nu & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2}-\nu \end{bmatrix}. \quad (1.81)$$

Plane stress and plane strain: The general three-dimensional stress-strain relations in Eq. (1.74) can be simplified for certain special situations that often occur in practice. The two-dimensional stress-strain relations can be categorized into two cases: plane stress and plane strain.

Many practical structures consist of thin plate-like components in order to be efficient. Assume that a thin plate is parallel to the xy -plane. If we assume that the top and bottom surfaces of the plate are not subjected to any significant forces in the z -direction, i.e., the plate is subjected to forces in its plane only, then the transverse stresses (stresses with a z subscript) vanish on the top and bottom surfaces, i.e., $\sigma_{13} = \sigma_{23} = \sigma_{33} = 0$ on the top and bottom surfaces. If the thickness is small compared to the lateral dimensions of the plate, then we can assume that the aforementioned transverse stresses are approximately zero through the entire thickness. The plate is then said to be in a state of plane stress where all stresses are parallel to the xy -plane and normal to the z -axis. In order to derive the stress-strain relations for the state of plane stress, we set $\sigma_{13} = \sigma_{23} = \sigma_{33} = 0$ to obtain

$$\{\boldsymbol{\sigma}\} = \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1}{2}(1-\nu) \end{bmatrix} \begin{Bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \gamma_{12} \end{Bmatrix}. \quad (1.82)$$

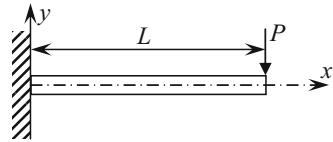
Similar to plane stress, one can define a state of plane strain in which the strains with a z subscript are all equal to zero. This situation corresponds to a structure whose deformation in the z -direction is constrained (i.e., $u_3 = 0$), so that the following relation holds: $\epsilon_{13} = \epsilon_{23} = \epsilon_{33} = 0$. Plane strain can also be used if the structure is infinitely long in the z -direction. In order to derive the stress-strain relations for the state of plane strain, we set $\epsilon_{13} = \epsilon_{23} = \epsilon_{33} = 0$ to obtain

$$\{\boldsymbol{\sigma}\} = \begin{Bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & \frac{1}{2}-\nu \end{bmatrix} \begin{Bmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \gamma_{12} \end{Bmatrix}. \quad (1.83)$$

Note that the normal stress, σ_{33} , is not zero in the plane-strain problem, but can be calculated from ϵ_{11} and ϵ_{22} :

$$\sigma_{33} = \frac{E\nu}{(1+\nu)(1-2\nu)}(\epsilon_{11} + \epsilon_{22}). \quad (1.84)$$

Fig. 1.12 Cantilever beam bending problem



Although the plane-stress and plane-strain problems are quite different in the engineering perspective, they are not much different from the viewpoint of a computer program. Both problems have only three components of stress and strain. The only difference is the elasticity matrix in Eqs. (1.82) and (1.83). Therefore, most computer programs do not distinguish between plane-stress and plane-strain problems. They use the same code, but with different elasticity matrix.

Example 1.9 (Stress distribution of a cantilevered beam) The displacement field for the thin beam, shown in Fig. 1.12, only considers bending

$$u(x, y) = \frac{P}{EI} \left(Lx - \frac{x^2}{2} \right) y - \frac{\nu P}{6EI} y^3,$$

$$v(x, y) = \frac{-\nu P}{2EI} (L - x) y^2 - \frac{P}{EI} \left(\frac{Lx^2}{2} - \frac{x^3}{6} \right),$$

where P is the applied force at the tip, I is the area moment of inertia about the bending axis, and L is the length of the beam. For an isotropic material with Young's modulus, E , and Poisson's ratio, ν , determine the entire stress field.

Solution Since the thickness of the beam is small, we can assume the plane-stress condition along the z -direction. From the definition of strain,

$$\epsilon_{xx} = \frac{\partial u}{\partial x} = \frac{P}{EI} (L - x) y,$$

$$\epsilon_{yy} = \frac{\partial v}{\partial y} = \frac{-\nu P}{EI} (L - x) y,$$

$$\gamma_{xy} = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} = \left[\frac{\nu P y^2}{2EI} - \frac{P}{EI} \left(Lx - \frac{x^2}{2} \right) \right] + \left[\frac{P}{EI} \left(Lx - \frac{x^2}{2} \right) y - \frac{\nu P y^2}{2EI} \right] = 0.$$

Substituting into Eq. (1.82) yields the stress field

$$\sigma_{xx} = \frac{E}{1 - \nu^2} \left[\frac{P}{EI} (L - x) y - \frac{\nu^2 P}{EI} (L - x) y \right] = \frac{P}{I} (L - x) y,$$

$$\sigma_{yy} = \frac{E}{1 - \nu^2} \left[-\frac{\nu P}{EI} (L - x) y + \frac{\nu P}{EI} (L - x) y \right] = 0,$$

$$\tau_{xy} = 0.$$

Since the normal stress, σ_{xx} , changes linearly in the y -direction, the stress field represents the bending of a beam. ■

1.4 Mechanics of Continuous Bodies

Mathematical models of many structural problems are formulated as differential equations that are satisfied at every point in the domain. These differential equations are usually obtained from the three fundamental laws of mechanics: conservation of mass, conservation of linear momentum, and conservation of angular momentum. The conservation of mass can be easily satisfied for a Lagrangian description of the problem, and the conservation of an angular momentum results in the symmetry of the stress tensor. Thus, the conservation of linear momentum, which is a differential equation used to satisfy the force equilibrium, is the major consideration in the structural problem.

Force equilibrium is imposed on an arbitrary infinitesimal element of a structure in order to obtain a boundary-valued problem. The smoothness of the solution in the boundary-valued problem depends on the order of the differential equation. For example, truss and continuum problems require continuous second-order derivatives of the solution, while beam and plate bending problems require continuous fourth-order derivatives. However, this section will show that these orders of differentiability are not necessary in order to represent many types of mechanical behaviors. In contrast, the variational approach reduces the solution's smoothness requirements and provides a general interpretation of the solution. The variational formulation that has been mathematically obtained can be rigorously related to a virtual work or energy principle in mechanics.

A complete mathematical theory related to the existence and uniqueness of the solution was developed by Aubin [9] and Fichera [10] using the Sobolev space and the properties of a bounded elliptic linear operator. However, the mathematical comprehension of this functional analysis requires a good deal of effort, with some physical insights. By contrast, a relatively simple theory is available that can formulate the structural problem using the energy principle. If the structural system is conservative, then it has a potential energy. Structural equilibrium is considered by the principle of minimum potential energy to be a stationary configuration of the potential energy. Since the potential energy of many structural problems is the positive definite quadratic function of a state variable (i.e., displacement), the stationary condition yields a unique global minimum solution.

In Sect. 1.4.2, the principle of minimum potential energy and a variational method are developed for a conservative structural system. An important result is then shown; namely, that if the solution for a structural differential equation exists, then that solution is the minimizing solution of the potential energy. In addition, even if the structural differential problem does not have a solution, the solution that minimizes the potential energy may exist and would provide a natural solution to the structural problem. The energy principles presented here will be restricted to small strains and displacements so that strain–displacement relationships can be expressed in linear equations; such displacements and corresponding strains, obviously, have additive properties. A nonlinear elastic stress–strain relationship will be discussed in Chap. 3.

The energy-based formulation of the structural problem in Sect. 1.4.2 is generalized to the principle of virtual work in Sect. 1.4.3, which can handle arbitrary constitutive relations. The principle of virtual work is the equilibrium of the work done by both internal and external forces with the small arbitrary virtual displacements that satisfy kinematic constraints. For a conservative system, the results obtained from the principle of virtual work are the same as the results obtained using the principle of minimum potential energy in Sect. 1.4.2.

1.4.1 Boundary-Valued Problem

Balance of linear momentum: A body in Fig. 1.13 is in static equilibrium under the applied body force, \mathbf{f}^b , and the surface traction, $\mathbf{t}^{(n)}$. The domain, inside of the body, is denoted by Ω , whose boundary is Γ . The balance of the linear moment can be stated as

$$\iint_{\Omega} \mathbf{f}^b d\Omega + \int_{\Gamma} \mathbf{t}^{(n)} d\Gamma = 0. \quad (1.85)$$

Using the property in Eq. (1.38) and the divergence theorem in Eq. (1.28), the second term in the forgoing equation can be converted into the integral over the domain as

$$\iint_{\Omega} \mathbf{f}^b d\Omega = - \int_{\Gamma} \mathbf{n} \cdot \boldsymbol{\sigma} d\Gamma = - \iint_{\Omega} \nabla \cdot \boldsymbol{\sigma} d\Omega.$$

Since both integrals are written over the same domain, they can be combined as

$$\iint_{\Omega} (\nabla \cdot \boldsymbol{\sigma} + \mathbf{f}^b) d\Omega = 0.$$

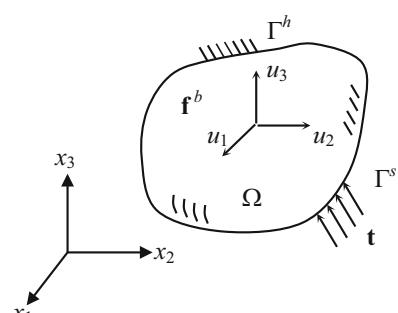


Fig. 1.13 Deformable body under equilibrium

Therefore, the balance of the linear momentum can be written at every point in the domain as

$$\nabla \cdot \boldsymbol{\sigma} + \mathbf{f}^b = 0, \quad \mathbf{x} \in \Omega. \quad (1.86)$$

The equilibrium state (solution) of the body must satisfy the local momentum balance equation in Eq. (1.86) as well as boundary conditions. Note that the balance of the angular momentum becomes identical to the symmetry of the stress tensor, which is similar to the process in Eq. (1.36).

Boundary-valued problem: Consider the linear elastic structure in Fig. 1.13 under the applied surface-traction \mathbf{t} on the boundary, Γ^s , and under the body force, \mathbf{f}^b , in the domain. The whole boundary, Γ , is decomposed into $\Gamma = \Gamma^h \cup \Gamma^s$ and $\Gamma^h \cap \Gamma^s = \emptyset$. The motion of the structure is fixed (or prescribed) on the essential boundary, Γ^h . The purpose of the boundary-valued problem is to find a displacement that satisfies

$$\begin{aligned} \nabla \cdot \boldsymbol{\sigma}(\mathbf{u}) + \mathbf{f}^b &= 0, & \mathbf{x} \in \Omega, \\ \mathbf{u} &= 0, & \mathbf{x} \in \Gamma^h, \\ \boldsymbol{\sigma} \cdot \mathbf{n} &= \mathbf{t}, & \mathbf{x} \in \Gamma^s, \end{aligned} \quad (1.87)$$

where \mathbf{n} is an outward unit normal vector to the surface, Γ^s . A constitutive relation is required, such as the one in Eq. (1.74), in order to make the boundary-valued problem complete. The boundary condition on Γ^h is called the displacement boundary condition, whereas the boundary condition on Γ^s is called the traction boundary condition. Equation (1.87) is often called the strong form because the differential equation must be satisfied at every point, $\mathbf{x} \in \Omega$, and the solution, \mathbf{u} , must be smooth enough such that its second-order derivatives are continuous, i.e., $\mathbf{u} \in [C^2(\Omega)]^3$. Although Eq. (1.87) only includes a divergence operator, which is a first-order derivative, the problem has second-order derivatives because strain contains derivatives of displacements.

In order to solve for the strong form in Eq. (1.87), the first step is to construct trial solutions that automatically satisfy a part of the boundary-valued problem, and then, the solution that satisfies the remaining conditions is found. For example, the trial solutions that satisfy the differential equation and traction boundary condition are called the statically admissible stress field. The trial solutions that satisfy the displacement boundary condition are called the kinematically admissible displacement field. Since the admissible stress field is difficult to construct, the admissible displacement field is often used to solve for the strong form.

1.4.2 Principle of Minimum Potential Energy

Principle of minimum potential energy: Due to the applied load, the elastic structure experiences deformation (or displacement), as described by $\mathbf{u}(\mathbf{x}) = \{u_1, u_2, u_3\}^T$ for $\mathbf{x} \in \Omega$. The structure resists any deformation by generating internal

forces. In general, each internal force is proportional to the amount of deformation. For a given applied load, if the internal force is smaller than the applied force, then the structure continues to deform in order to equilibrate the two forces. Many structural problems consist of computing the displacement due to force equilibrium conditions between the applied load and internal forces.

If the concept of structural force equilibrium is extended to the energy formulation, then a good deal of physical insight can be obtained. Let the displacements be used as state variables of the problem being considered. The internal force, generated during deformation, can be thought of as the energy that is stored in the structure. As the structure deforms, not only does the internal force increase, but the energy of the structure also increases. This stored energy is called the strain energy of the structure, which is defined as

$$U(\mathbf{u}) \equiv \frac{1}{2} \iint_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{u}) d\Omega, \quad (1.88)$$

where the components of the strain tensor are defined as

$$\epsilon_{ij}(\mathbf{u}) = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) = \frac{1}{2} (u_{i,j} + u_{j,i}), \quad (1.89)$$

and the constitutive relation for an elastic material in Eq. (1.74) is assumed. In Eq. (1.89), the subscribed comma represents the derivative with respect to the spatial coordinate, i.e., $u_{i,j} = \partial u_i / \partial x_j$. The strain energy, $U(\mathbf{u})$, is the energy required to produce the displacement, \mathbf{u} . For elastic problems, since $U(\mathbf{u})$ does not depend on the path chosen for deformation, $U(\mathbf{u})$ is a function of the displacement, \mathbf{u} , only.

If forces are applied to the structure and the structure deforms in the direction of the applied forces, then work is done by the applied forces. The work done by the applied load can be defined as

$$W(\mathbf{u}) = \iint_{\Omega} \mathbf{u} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma^s} \mathbf{u} \cdot \mathbf{t} d\Gamma. \quad (1.90)$$

The first integral in Eq. (1.90) represents the work done by the body force, \mathbf{f}^b , while the second integral is the work done by the surface traction, \mathbf{t} . The integrals are evaluated over the whole domain, Ω , and over the traction boundary, Γ^s . If any concentrated force, \mathbf{f} , is applied externally, then the surface traction term in Eq. (1.90) may include the Dirac delta measure. Note that $U(\mathbf{u})$ is a quadratic function of \mathbf{u} , while $W(\mathbf{u})$ is a linear function of \mathbf{u} .

Since the strain energy, $U(\mathbf{u})$, is independent of the deformation path, it is the potential energy that is stored in the structure. If the applied force in Eq. (1.90) is conservative, then Eq. (1.90) defines the negative value of the potential energy generated by the applied loads. The applied load is considered to be conservative if it is independent of deformation, such that the work done by a system of applied forces in traversing any closed path in the displacement space has to be zero.

The potential energy of the structure is the difference between the strain energy and the work done by the applied loads, written as

$$\begin{aligned}\Pi(\mathbf{u}) &= U(\mathbf{u}) - W(\mathbf{u}) \\ &= \frac{1}{2} \iint_{\Omega} \boldsymbol{\sigma}(\mathbf{u}) : \boldsymbol{\epsilon}(\mathbf{u}) d\Omega - \iint_{\Omega} \mathbf{u} \cdot \mathbf{f}^b d\Omega - \int_{\Gamma^s} \mathbf{u} \cdot \mathbf{t} d\Gamma.\end{aligned}\quad (1.91)$$

The principle of minimum potential energy is as follows: for all displacements that satisfy the boundary conditions, known as kinematically admissible displacements, those which satisfy the boundary-valued problem in Eq. (1.87), if they exist, make the total potential energy in Eq. (1.91) stationary on

$$D_A = \left\{ \mathbf{u} \in [C^2(\Omega)]^3 \mid \mathbf{u} = 0 \text{ on } \mathbf{x} \in \Gamma^h, \boldsymbol{\sigma} \cdot \mathbf{n} = \mathbf{t} \text{ on } \mathbf{x} \in \Gamma^s \right\}. \quad (1.92)$$

The principle of minimum total potential energy provides a generalized solution to the differential equation. For the generalized solution that minimizes the total potential energy in Eq. (1.91), the solution space, D_A , of Eq. (1.92) has to be extended so that the potential energy in Eq. (1.91) can be well defined. This space is called the space of finite energy or the space of kinematically admissible displacements, defined as

$$\mathbb{Z} = \left\{ \mathbf{u} \in [H^1(\Omega)]^3 \mid \mathbf{u} = 0 \text{ on } \mathbf{x} \in \Gamma^h \right\}, \quad (1.93)$$

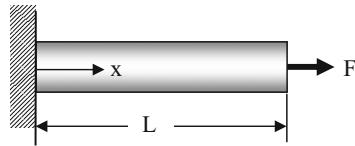
where $H^1(\Omega)$ is the Sobolev space of order 1.² The generalized solution belongs to the space, \mathbb{Z} . It is important to point out that the traction boundary condition is not required to define the space of kinematically admissible displacements because it is included in the work done by the applied load in Eq. (1.91). Thus, it is easier to construct the space of kinematically admissible displacements than it is to construct D_A . Generally, if we let the order of differential equation equal $2m$, then the boundary conditions that contain $(m-1)$ th-order derivatives are called the essential boundary conditions and derivatives of a higher order than $(m-1)$ are called the natural boundary conditions.

Example 1.10 (Equilibrium of a bar) The bar in Fig. 1.14 has length, L ; Young's modulus, E ; and cross-sectional area, A . Assume that displacement is in the form of $u(x) = c_1x + c_2$; calculate the displacement (1) by solving the governing differential equation and (2) by using the stationary condition of the potential energy.

Solution It is important to identify boundary conditions first. Since the bar is fixed at the left end, it has zero displacement at $x=0$, which corresponds to the displacement boundary condition. On the other hand, a force is applied at the

² $H^m(\Omega)$ is the Sobolev space of the order m , whose functions are continuously differentiable up to $m-1$, and m th partial derivatives are integrable.

Fig. 1.14 Equilibrium of a bar



right end of the bar. In the view of Eq. (1.87), the force can be considered as an integral of a constant surface traction, t . In addition, using the relation of $F = tA = A\sigma_{11} = EA\varepsilon_{11} = EAu'(L)$, the differential equation and boundary conditions can be written as

$$\begin{aligned} EAu'' &= 0 \quad x \in [0, L], \\ u &= 0 \quad x = 0, \\ EAu'(L) &= F \quad x = L. \end{aligned}$$

In the above equations, ' and '' are used for the first-order and second-order derivative with respect to x . At $x=L$, the unit normal is $\mathbf{n} = \{1, 0, 0\}^T$ and the stress tensor only has the σ_{11} component.

- (1) The governing differential equation is integrated twice to obtain

$$EAu(x) = c_1x + c_2.$$

After applying the two boundary conditions, we have $c_1 = F$ and $c_2 = 0$. Therefore, the displacement becomes

$$u(x) = \frac{Fx}{EA}.$$

- (2) To construct kinematically admissible displacements, the candidates must satisfy the displacement boundary condition, $u(0) = 0$. Therefore, the candidate displacements become $u(x) = c_1x$. The objective is to find a value of c_1 that makes the potential energy of the bar stationary. By considering only σ_{11} and ε_{11} for the bar, the strain energy and the work done by the applied load, respectively, become

$$\begin{aligned} U &= \frac{1}{2} \int_0^L EA(u')^2 dx = \frac{1}{2} EALc_1^2, \\ W &= Fu(L) = FLc_1. \end{aligned}$$

The potential energy can be obtained from $\Pi = U - W$. The stationary condition of the potential energy can be obtained by differentiating the potential energy with respect to c_1 and making the result equal to zero:

$$\frac{d\Pi}{dc_1} = \frac{d}{dc_1}(U - W) = EALc_1 - FL = 0,$$

which yields the unknown coefficient, c_1 , that minimizes the potential energy, as

$$c_1 = \frac{F}{EA} \Rightarrow u(x) = \frac{Fx}{EA}.$$

Note that the solution of (2) is identical to that of (1). ■

Virtual displacement: The virtual displacement, or the variation of \mathbf{u} , is a small arbitrary perturbation of the real displacement \mathbf{u} in \mathbb{Z} . Let the virtual displacement be $\boldsymbol{\eta}(\mathbf{x})$ with a small scalar, τ , so that the perturbed state is $\mathbf{u} + \tau\boldsymbol{\eta}$. Since the perturbed state has to be in \mathbb{Z} , the perturbation $\boldsymbol{\eta}(\mathbf{x})$ has to vanish at its essential boundary; i.e., $\boldsymbol{\eta}(\mathbf{x})$ satisfies all homogeneous essential boundary conditions (see Fig. 1.15). Therefore, $\boldsymbol{\eta}(\mathbf{x})$ also belongs to the space, \mathbb{Z} . The function, $\boldsymbol{\eta}(\mathbf{x})$, can be thought of as a variation of the displacement, \mathbf{u} . In order to show this, consider $\mathbf{u} + \tau\boldsymbol{\eta}$ that is defined on \mathbb{Z} . For a sufficiently small τ , the variation of the displacement can be defined as

$$\delta\mathbf{u} = \lim_{\tau \rightarrow 0} \frac{1}{\tau} [(\mathbf{u} + \tau\boldsymbol{\eta}) - (\mathbf{u})] = \frac{d}{d\tau} (\mathbf{u} + \tau\boldsymbol{\eta}) \Big|_{\tau=0} = \boldsymbol{\eta} \equiv \bar{\mathbf{u}}. \quad (1.94)$$

Since the variation, $\boldsymbol{\eta}$, is related to the displacement, \mathbf{u} , the notation, $\bar{\mathbf{u}}$, is used instead of $\delta\mathbf{u}$ to denote the variation of the displacement, \mathbf{u} , in this text. This notational system is preferred in order to avoid an excessive usage of δ , which typically denotes the Dirac delta measure as well as the Kronecker delta symbol.

An important property of the variation is that it is independent of the differentiation with respect to space coordinates. For example, consider the variation of strain:

$$\delta \left(\frac{d\mathbf{u}}{dx} \right) = \frac{d(\delta\mathbf{u})}{dx}.$$

Variational equation: The principle of minimum potential energy, discussed in the previous section, requires obtaining a stationary condition for the total potential energy. This principle is closely related to the variational formulation.

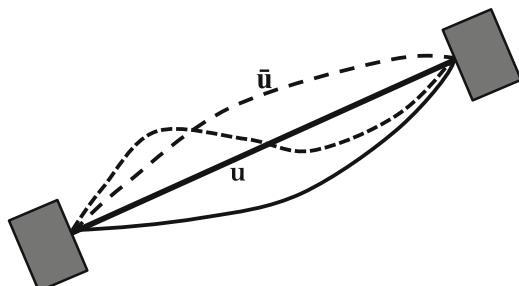


Fig. 1.15 Virtual displacement as a perturbation of real displacement

Let displacement, $\mathbf{u} \in \mathbb{Z}$, be the solution to the structural problem that uniquely minimizes $\Pi(\mathbf{u})$. Consider a functional Π that is defined in \mathbb{Z} . For a sufficiently small τ , if the limit

$$\delta\Pi(\mathbf{u}; \bar{\mathbf{u}}) = \lim_{\tau \rightarrow 0} \frac{1}{\tau} [\Pi(\mathbf{u} + \tau\bar{\mathbf{u}}) - \Pi(\mathbf{u})] = \left. \frac{d}{d\tau} \Pi(\mathbf{u} + \tau\bar{\mathbf{u}}) \right|_{\tau=0} \quad (1.95)$$

exists, then it is called the first variation of Π at \mathbf{u} in the direction of $\bar{\mathbf{u}}$. If this limit exists for every $\bar{\mathbf{u}} \in \mathbb{Z}$, Π is said to be differentiable (i.e., Fréchet differentiable) at \mathbf{u} .

If a functional has a first variation, then quantitative criteria can be defined for its minimization. The focus here is on the necessary conditions for the extrema of the functional. Presume that \mathbf{u} is such that

$$\Pi(\mathbf{u}) \leq \Pi(\mathbf{w}), \quad (1.96)$$

for all \mathbf{w} in \mathbb{Z} ; then \mathbf{u} is said to minimize Π over \mathbb{Z} . If Eq. (1.96) holds for all \mathbf{w} in \mathbb{Z} that satisfy $\|\mathbf{w} - \mathbf{u}\| \leq d$, for some $d > 0$, Π is said to have a relative minimum value at \mathbf{u} .

From Eq. (1.96), for any $\bar{\mathbf{u}} \in \mathbb{Z}$ and for any sufficiently small τ , if Π has a relative minimum at \mathbf{u} , then

$$\Pi(\mathbf{u}) = \min_{\tau} \Pi(\mathbf{u} + \tau\bar{\mathbf{u}}) = \Pi(\mathbf{u} + \tau\bar{\mathbf{u}})|_{\tau=0}, \quad (1.97)$$

that is, for fixed \mathbf{u} and $\bar{\mathbf{u}}$, the real value function, $\Pi(\mathbf{u} + \tau\bar{\mathbf{u}})$, of the real parameter, τ , is a minimum at $\tau = 0$. If the functional has a first variation, then $\Pi(\mathbf{u} + \tau\bar{\mathbf{u}})$ is a differentiable function of τ , and a necessary condition for a minimum of Π at \mathbf{u} is

$$\delta\Pi(\mathbf{u}; \bar{\mathbf{u}}) = \left. \frac{d}{d\tau} \Pi(\mathbf{u} + \tau\bar{\mathbf{u}}) \right|_{\tau=0} = 0, \quad (1.98)$$

for all $\bar{\mathbf{u}} \in \mathbb{Z}$. The notation, $\delta\Pi(\mathbf{u}; \bar{\mathbf{u}})$, represents a variation of Π at \mathbf{u} in the direction of $\bar{\mathbf{u}}$. Thus, the principle of minimum potential energy is equivalent to the condition of Eq. (1.98) for all kinematically admissible $\bar{\mathbf{u}}$.

The potential energy in Eq. (1.91) is composed of the strain energy and the work done by the applied load. Thus, a variational formulation of the structural problem can be written, using the first variation of $\Pi(\mathbf{u})$, as

$$\delta\Pi(\mathbf{u}; \bar{\mathbf{u}}) = \delta U(\mathbf{u}; \bar{\mathbf{u}}) - \delta W(\mathbf{u}; \bar{\mathbf{u}}) = 0, \quad (1.99)$$

for all $\bar{\mathbf{u}}$ in \mathbb{Z} . Equation (1.99) is called the variational equation of the structural problem under consideration. The first term in Eq. (1.99) is obtained from the definition of $U(\mathbf{u})$ in Eq. (1.88) and the stress-strain relation in Eq. (1.74) as

$$\delta U(\mathbf{u}; \bar{\mathbf{u}}) = \iint_{\Omega} \boldsymbol{\varepsilon}(\bar{\mathbf{u}}) : \mathbf{D} : \boldsymbol{\varepsilon}(\mathbf{u}) d\Omega \equiv a(\mathbf{u}, \bar{\mathbf{u}}), \quad (1.100)$$

where $a(\mathbf{u}, \bar{\mathbf{u}})$ is called the energy-bilinear form since it is bilinear with respect to its two arguments \mathbf{u} and $\bar{\mathbf{u}}$. Note that $e_{ij}(\bar{\mathbf{u}})$ is made the same as $e_{ij}(\mathbf{u})$ in Eq. (1.89) by substituting $\bar{\mathbf{u}}$ into \mathbf{u} . Thus, the energy-bilinear form is symmetric with respect to its arguments. The variation of the work done by the applied load can be written as

$$\delta W(\mathbf{u}; \bar{\mathbf{u}}) = \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma^s} \bar{\mathbf{u}} \cdot \mathbf{t} d\Gamma \equiv \ell(\bar{\mathbf{u}}), \quad (1.101)$$

where $\ell(\bar{\mathbf{u}})$ is called the load-linear form. Only conservative loads are considered such that $\ell(\bar{\mathbf{u}})$ is independent of displacement. Thus, the variational formulation of the structural problem in Eq. (1.99) can be written as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (1.102)$$

where $\forall \bar{\mathbf{u}} \in \mathbb{Z}$ represents “for all $\bar{\mathbf{u}}$ in \mathbb{Z} .” If the load-linear form on the right-hand side of Eq. (1.102) is continuous in the space, \mathbb{Z} , and if the energy-bilinear form on the left-hand side of Eq. (1.102) is positive definite on \mathbb{Z} , then Eq. (1.102) has a unique solution, $\mathbf{u} \in \mathbb{Z}$.

If a solution exists to the differential equation, then it is also the solution to the variational equation, Eq. (1.102). However, a solution to Eq. (1.87) may not exist if the distributed function \mathbf{t} is a Dirac delta measure, which means that \mathbf{t} is the applied point load. Nevertheless, the variational equation, Eq. (1.102), still has a solution in this case, which is called a generalized solution.

Example 1.11 (Equilibrium of a spring) Consider a spring component, which is fixed at one end and under an applied force, f , at the other end. Calculate the displacement, u , at the load application point.

Solution The potential energy of the spring can be written as

$$\Pi(u) = \frac{1}{2}ku^2 - fu.$$

If the displacement is perturbed by $u + \tau\bar{u}$, the perturbed potential energy can be written as

$$\Pi(u + \tau\bar{u}) = \frac{1}{2}k(u + \tau\bar{u})^2 - f(u + \tau\bar{u}).$$

Then the variation of the potential energy can be obtained by differentiating the potential energy with respect to τ , as

$$\frac{d}{d\tau}[\Pi(u + \tau\bar{u})] = k(u + \tau\bar{u})\bar{u} - f\bar{u}.$$

Now, if the variation of potential energy is evaluated at the original state, i.e., at $\tau = 0$,

$$\frac{d}{d\tau}[\Pi(u + \tau\bar{u})] \Big|_{\tau=0} = ku\bar{u} - f\bar{u} = 0.$$

Since the displacement variation, \bar{u} , is arbitrary, the above variational equation yields

$$ku = f$$

which is the same as the equilibrium of the spring component. ■

Example 1.12 (Equilibrium of a bar) For the bar in Example 1.10, calculate the displacement, $u(x)$, using the principle of minimum potential energy. Assume that the virtual displacement is in the same form as the displacement.

Solution As with Example 1.10, the displacement can be assumed to be $u(x) = cx$. In this form of displacement, the coefficient, c , is the only unknown; that is, the solution is sought only from linear functions. Obtaining a solution, $u(x)$, is equivalent to determining the coefficient, c . Since the virtual displacement shares the same property with the displacement, it is natural to assume that the virtual displacement has the following form: $\bar{u}(x) = \bar{c}x$, where \bar{c} is the coefficient of virtual displacement. Therefore, the arbitrary virtual displacements, \bar{u} , imply that the coefficient, \bar{c} , is arbitrary.

The variation of strain energy can be obtained

$$\begin{aligned}\delta U &= \frac{d}{d\tau} \left[\frac{1}{2} \int_0^L EA \left[(u + \tau\bar{u})' \right]^2 dx \right]_{\tau=0} \\ &= \frac{1}{2} \int_0^L 2EA(u + \tau\bar{u})' \bar{u}' dx \Big|_{\tau=0} \\ &= \int_0^L EAu' \bar{u}' dx.\end{aligned}$$

After substituting $u(x) = cx$ and $\bar{u}(x) = \bar{c}x$ into the above expression, the variation of strain energy becomes

$$\delta U = EALcc\bar{c}.$$

In the case of the work done by applied load, the variation becomes

$$\begin{aligned}\delta W &= \frac{d}{d\tau} [F(u(L) + \tau\bar{u}(L))]_{\tau=0} \\ &= F\bar{u}(L) \\ &= FL\bar{c}.\end{aligned}$$

Therefore, the necessary condition in Eq. (1.98) can be written as

$$\delta\Pi = \delta U - \delta W = \bar{c}(EALc - FL) = 0$$

for all $\bar{c} \in R$, with R being the space of real numbers. Therefore, the unknown coefficient, c , can be solved from the above variational equation. Finally, the displacement can be expressed by

$$u(x) = cx = \frac{Fx}{EA}$$

which is identical to that of Example 1.10. ■

1.4.3 Principle of Virtual Work

The variational formulation provided by Eq. (1.102), obtained from the principle of minimum potential energy, is limited in solving elastic problems. In the principle of virtual work, the constitutive relations, including the elastoplasticity, can be quite general since we are not assuming that potential energy exists. Let the differential problem in Eq. (1.87) be satisfied and let the integration-by-parts be justified. Consider a virtual displacement, $\bar{\mathbf{u}}$, that satisfies the displacement boundary condition, i.e., $\bar{\mathbf{u}} = 0$ on Γ^h . Note that the displacement variation, $\bar{\mathbf{u}}$, in Eq. (1.94) is related to the displacement, \mathbf{u} . Even if the same notation, $\bar{\mathbf{u}}$, is used here, the virtual displacement is considered a small arbitrary continuous field that satisfies the problem's kinematic constraints, while the applied load is kept constant. Since the differential equation (1.87) is satisfied in the domain, Ω , by multiplying $\bar{\mathbf{u}}$ on both sides of the differential equation and integrating it, we have

$$\iint_{\Omega} \bar{\mathbf{u}} \cdot (\nabla \cdot \boldsymbol{\sigma} + \mathbf{f}^b) d\Omega = 0, \quad (1.103)$$

for any $\bar{\mathbf{u}}$ in \mathbb{Z} . In Eq. (1.103), equilibrium of the structural problem is sought in the sense of integration. The point-wise requirement of differential equations has no meaning in the variational approach. Since the differential equation, Eq. (1.87), is obtained from the force equilibrium relation, the term, $\nabla \cdot \boldsymbol{\sigma} + \mathbf{f}^b$, represents the unbalanced force, and Eq. (1.103) represents the virtual work done by the system during virtual displacement. Thus, structural equilibrium is considered a vanishing condition of the virtual work. After integrating by parts, the principle of virtual work is obtained by using the symmetric property of the stress tensor, $\boldsymbol{\sigma}$; the boundary conditions of Eq. (1.87); and the constitutive relation in Eq. (1.74) as

$$\begin{aligned} & \iint_{\Omega} \bar{\mathbf{u}} \cdot (\nabla \cdot \boldsymbol{\sigma} + \mathbf{f}^b) d\Omega \\ &= - \iint_{\Omega} \nabla \bar{\mathbf{u}} : \boldsymbol{\sigma} d\Omega + \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma^h \cup \Gamma^s} \bar{\mathbf{u}} \cdot \boldsymbol{\sigma} \cdot \mathbf{n} d\Gamma \\ &= - \iint_{\Omega} \nabla \bar{\mathbf{u}} : \boldsymbol{\sigma} d\Omega + \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma^s} \bar{\mathbf{u}} \cdot \mathbf{t} d\Gamma \\ &= - \iint_{\Omega} \boldsymbol{\epsilon}(\bar{\mathbf{u}}) : \mathbf{D} : \boldsymbol{\epsilon}(\mathbf{u}) d\Omega + \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma^s} \bar{\mathbf{u}} \cdot \mathbf{t} d\Gamma \\ &= 0. \end{aligned}$$

where $\sigma \cdot \mathbf{n} = \mathbf{t}$ on Γ^s . By using definitions of the energy-bilinear form and the load-linear form, the principle of virtual work can be stated as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (1.104)$$

Equation (1.104) is the same as the variational formulation in Eq. (1.102). In the principle of virtual work, the left side of Eq. (1.104) is interpreted as the virtual work done by internal force, while the right side is seen as virtual work done by the external applied force. Thus, Eq. (1.104) states that the structure is in equilibrium when internal and external virtual works are equal during all virtual displacements.

In the derivation of the principle of virtual work in Eq. (1.103), it is assumed that the differential equation is satisfied at every point within the structure, which is an unnecessary requirement. Further, consider a virtual work,

$$\delta W = \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma} \bar{\mathbf{u}} \cdot \mathbf{t} d\Gamma. \quad (1.105)$$

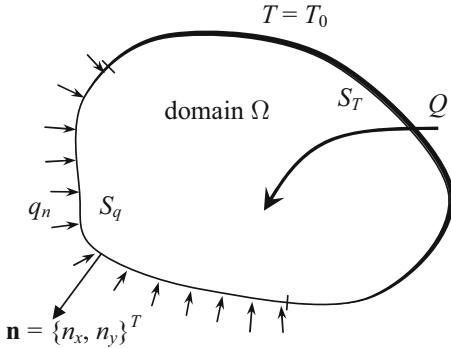
Since $\bar{\mathbf{u}} = \mathbf{0}$ on Γ^h , the whole boundary $\Gamma = \Gamma^h \cup \Gamma^s$ is used instead of Γ^s . Using the relation of $\mathbf{t} = \sigma \cdot \mathbf{n}$ and Gauss' theorem, the virtual work in Eq. (1.105) can be extended to

$$\begin{aligned} \iint_{\Omega} \bar{\mathbf{u}} \cdot \mathbf{f}^b d\Omega + \int_{\Gamma} \bar{\mathbf{u}} \cdot \mathbf{t} d\Gamma &= \iint_{\Omega} \bar{\mathbf{u}} \cdot (\mathbf{f}^b + \nabla \cdot \sigma) d\Omega \\ &\quad + \iint_{\Omega} \sigma : \boldsymbol{\epsilon}(\bar{\mathbf{u}}) d\Omega, \end{aligned} \quad (1.106)$$

where, again, the symmetric property of the stress tensor is used. The first integral on the RHS of the above equation is the same as in Eq. (1.103), which vanishes. Thus, the same principle of virtual work as in Eq. (1.104) is obtained. A subtle difference in this approach is that it is unnecessary to assume a point-wise satisfaction of the differential equation. As long as the first integral on the RHS vanishes, the principle of virtual work is well defined.

The difference between the variational formulation and the principle of virtual work cannot be clearly seen from the conservative system or from the linear elastic structural problem. However, in developing the variational formulation, we assumed that potential energy existed in the structure. Thus, the variational formulation is limited to elastic problems. For most of the problems discussed in Chap. 4, the potential energy of the structural problem does not exist. For those problems, the principle of virtual work has to be used. However, proving the existence and uniqueness of a solution in Eq. (1.104) is a difficult procedure that goes beyond the scope of this text. For a proof of the existence and uniqueness of a solution, the reader is referred to the articles of Aubin [9] and Fichera [11].

Fig. 1.16 Two-dimensional heat transfer problem



Example 1.13 (Heat transfer problem) Consider a two-dimensional heat transfer problem in Fig. 1.16 under the heat source, Q , over the domain, the heat flux, q_n , on the boundary, S_q , and the prescribed temperature, T_0 , on the boundary, S_T . The governing differential equation of the steady-state heat transfer problem is

$$\frac{\partial}{\partial x} \left(k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial T}{\partial y} \right) + Q = 0 \quad (1.107)$$

with the following boundary conditions:

$$\begin{cases} T = T_0 \text{(prescribed)} & \text{on } S_T \\ q_n = -f_x n_x - f_y n_y & \text{on } S_q \end{cases} \quad (1.108)$$

where Fourier law of heat transfer is given as

$$\begin{aligned} f_x &= -k_x \frac{\partial T}{\partial x}, \\ f_y &= -k_y \frac{\partial T}{\partial y}. \end{aligned}$$

Using the principle of virtual work, calculate the variational equation of the steady-state heat transfer problem.

Solution We associate the above differential equation with test functions (or “virtual temperatures”) in the space \mathbb{Z} that satisfies the essential boundary condition on S_T , where

$$\mathbb{Z} = \left\{ \bar{T} \in H^{(1)}(\Omega) \mid \bar{T}(\mathbf{x}) = 0, \forall \mathbf{x} \in S_T \right\}. \quad (1.109)$$

By multiplying Eq. (1.107) with an arbitrary element of \mathbb{Z} and then integrating over the domain, Ω , we obtain

$$\int_{\Omega} \left[\frac{\partial}{\partial x} \left(k_x \frac{\partial T}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial T}{\partial y} \right) + Q \right] \bar{T} \, d\Omega = 0, \quad \forall \bar{T} \in \mathbb{Z},$$

where the LHS of the equation can be integrated by parts; in addition, we can use the property that $\bar{T} = 0$ on S_T . Thus, in the variational formulation of the steady-state heat flow problem, the generalized solution, T , must satisfy the following equation:

$$\int_{\Omega} \left(k_x \frac{\partial T}{\partial x} \frac{\partial \bar{T}}{\partial x} + k_y \frac{\partial T}{\partial y} \frac{\partial \bar{T}}{\partial y} \right) \, d\Omega = \int_{\Omega} \bar{T} Q \, d\Omega + \int_{S_q} \bar{T} q_n \, dS_q, \quad \forall \bar{T} \in \mathbb{Z}. \quad (1.110)$$

■

Example 1.14 (Beam problem) The governing differential equation of the beam component is

$$EI \frac{d^4 v}{dx^4} = f(x), \quad x \in [0, L], \quad (1.111)$$

where $f(x)$ is the distributed load. In the case of a cantilevered beam, the boundary condition can be given by

$$v(0) = \frac{dv}{dx}(0) = \frac{d^2 v}{dx^2}(L) = \frac{d^3 v}{dx^3}(L) = 0. \quad (1.112)$$

Using the principle of virtual work, derive the variational equation.

Solution By multiplying the governing differential equation by a virtual displacement, $\bar{v} \in \mathbb{Z}$, and integrating over the domain, we obtain

$$\int_0^L EI \frac{d^4 v}{dx^4} \bar{v} \, dx = \int_0^L f \bar{v} \, dx, \quad \forall \bar{v} \in \mathbb{Z},$$

where \mathbb{Z} is the space of kinematically admissible displacements that satisfy the essential boundary conditions:

$$\mathbb{Z} = \left\{ \bar{v} \in H^{(2)}[0, L] \mid \bar{v}(0) = \frac{d\bar{v}}{dx}(0) = 0 \right\}. \quad (1.113)$$

Since the order of the differential equation is 4, integration-by-parts is performed twice to make the order of differentiation between the displacement, v , and the virtual displacement, \bar{v} , the same. Now we have

$$\int_0^L EI \frac{d^2 v}{dx^2} \frac{d^2 \bar{v}}{dx^2} \, dx + \left[EI \frac{d^3 v}{dx^3} \bar{v} \right]_0^L - \left[EI \frac{d^2 v}{dx^2} \frac{d\bar{v}}{dx} \right]_0^L = \int_0^L f \bar{v} \, dx, \quad \forall \bar{v} \in \mathbb{Z}.$$

Using the boundary conditions in Eq. (1.112) and the virtual displacement in Eq. (1.113), the two boundary terms in the above equation vanish. Thus, the final expression of the variational equation becomes

$$\int_0^L EI \frac{d^2\bar{v}}{dx^2} \frac{d^2v}{dx^2} dx = \int_0^L f\bar{v} dx, \quad \forall \bar{v} \in \mathbb{Z}. \quad (1.114)$$

■

1.5 Finite Element Method

In general, it is difficult to find an analytic solution that satisfies the variational equation in the previous section. Instead, the FEM divides the entire domain into a set of simple sub-domains or *finite elements*. The finite elements are connected with adjacent elements by sharing their nodes. Then within each finite element, the solution is approximated in a simple polynomial form.

FEMs for structural analysis require knowledge of the behavior of each element in the structure. In this section, a structural analysis based on the finite element approach is introduced using three-dimensional solid elements. Finite element formulations for other structural elements, such as bars, beams, and plates, can be found in many textbooks, such as Bathe [12] or Hughes [13]. Apart from the more intricate algebra that is required for more complex elements, the basic approach for deriving element equations is identical to the process illustrated in this section. Once each element is described, the governing equations of the entire structure may then be derived.

1.5.1 Finite Element Approximation

Differential equations and variational equations, introduced in the previous section, are difficult to solve, except for a handful of simple cases. When the geometry is complicated, it is not trivial to solve for $u(x)$ analytically. Since the solution that satisfies the differential equation and boundary conditions can have a complicated expression, an infinite series solution may need to be employed. In the FEM, instead of solving the variational equation analytically, an approximate solution is sought. The approximate solution $u(x)$ is expressed as a sum of a number of functions that are called *trial functions*:

$$u(x) = \sum_{i=1}^n c_i \phi_i(x), \quad (1.115)$$

where n is the number of terms used, $\phi_i(x)$ are known *trial functions*, and c_i are coefficients to be determined by minimizing error between the true and the approximate solution. Since the approximate solution is a linear combination of the trial functions, the accuracy of approximation depends on them.

The trial functions and coefficients are chosen such that $u(x)$ must satisfy the essential boundary conditions of the problem; that is, $u(x)$ must belong to the space of kinematically admissible displacements, \mathbb{Z} . Therefore, if the solution to the variational equation is approximated by a series of functions in the entire domain of the problem, it is difficult to obtain the trial functions that satisfy the essential boundary conditions. An important idea of the FEM is to divide the entire domain into a set of simple sub-domains or *finite elements* and then to apply the approximation in Eq. (1.115) on the element level. Then, it is unnecessary to build the trial functions that satisfy the essential boundary conditions. Instead, only those elements that include the essential boundary conditions need to have a special treatment. The finite elements are connected with adjacent elements by sharing their nodes. Then within each finite element, the solution is approximated using a simple polynomial form. For example, let us assume that the domain is one-dimensional and the exact solution is given as a dashed curve in Fig. 1.17. When the entire domain is divided into sub-domains (finite elements), it is possible to approximate the solution using piecewise continuous linear polynomials as shown in Fig. 1.17. Within each element, the approximate solution is linear. Two adjacent elements have the same solution value at the shared node. As can be seen in the figure, when more numbers of elements are used, the approximate piecewise linear solution will converge to the exact solution. In addition, the approximation can be more accurate if higher-order polynomials are used in each element.

Various types of finite elements can be used, depending on the domain that needs to be discretized and the order of polynomials that are used to approximate the solution. Table 1.3 illustrates several types of finite elements that are often used in one-, two-, and three-dimensional problems.

After dividing the domain into finite elements, the integrations in the variational equations are performed over each element. For example, let us assume that the one-dimensional domain $(0, 1)$ is divided into ten equal-sized finite elements. Then, the integral can be written as a summation of integrals over each element:

$$\int_0^1 \square dx = \int_0^{0.1} \square dx + \int_{0.1}^{0.2} \square dx + \cdots + \int_{0.9}^1 \square dx$$

where \square is the integrand.

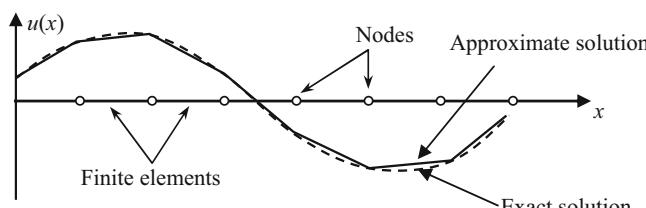
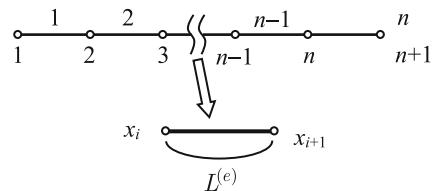


Fig. 1.17 Piecewise linear approximation of the solution for one-dimensional problem

Table 1.3 Different types of finite elements

Element	Name
	1D linear element
	2D triangular element
	2D rectangular element
	3D tetrahedron element
	3D hexahedron element

Fig. 1.18 Domain discretization of one-dimensional problem

After the domain is divided into a set of simple-shaped elements, the solution within an element is approximated in the form of simple polynomials. Let us consider a one-dimensional case in which the domain is discretized by n number of elements, as shown in Fig. 1.18. For this specific example, each element is composed of two end nodes. The trial solution is constructed in the element using the solution values at these nodes.

For example, element e connects two nodes at $x = x_i$ and $x = x_{i+1}$. If we want to interpolate the solution using two nodal values, then the linear polynomial is the appropriate choice because it has two unknowns. Thus, the solution is approximated by

$$u(x) = a_0 + a_1x, \quad x_i \leq x \leq x_{i+1}. \quad (1.116)$$

Note that the trial solution in the above equation is only defined within element e . Although we can determine two coefficients, a_0 and a_1 , they do not have a physical meaning. Instead, the unknown coefficients, a_0 and a_1 , in Eq. (1.116) will be expressed in terms of the nodal solutions, $u(x_i)$ and $u(x_{i+1})$. Although we do not know these nodal solutions yet, they will be determined later. By substituting these two nodal values, we have

$$\begin{cases} u(x_i) = u_i = a_0 + a_1x_i \\ u(x_{i+1}) = u_{i+1} = a_0 + a_1x_{i+1} \end{cases}$$

where u_i and u_{i+1} are the solution values at the two end nodes. Then, by solving the above equation, the two unknown coefficients, a_0 and a_1 , can be represented by the nodal solution, u_i and u_{i+1} . After substituting the two coefficients into Eq. (1.116), the approximate solution can be expressed in terms of the nodal solutions as

$$u(x) = \underbrace{\frac{x_{i+1} - x}{L^{(e)}} u_i}_{N_1(x)} + \underbrace{\frac{x - x_i}{L^{(e)}} u_{i+1}}_{N_2(x)},$$

where $L^{(e)} = x_{i+1} - x_i$ is the length of element e . Now, the approximate solution for $u(x)$ in Eq. (1.116) can be rewritten as

$$u(x) = N_1(x)u_i + N_2(x)u_{i+1}, \quad x_i \leq x \leq x_{i+1}, \quad (1.117)$$

where the functions $N_1(x)$ and $N_2(x)$ are called *interpolation functions* for obvious reasons. The expression in Eq. (1.117) shows that the solution, $u(x)$, is interpolated using its nodal values, u_i and u_{i+1} . At $x = x_i$, $N_1(x) = 1$, and at $x = x_{i+1}$, $N_1(x) = 0$, while at $x = x_{i+1}$, $N_2(x) = 1$ and at $x = x_i$, $N_2(x) = 0$. Interpolation functions $N_1(x)$ and $N_2(x)$ are also called *shape functions*, a term used in solid mechanics, as the functions describe the deformed shape of a solid or structure.

Note that the approximate solution in Eq. (1.117) is similar to that of Eq. (1.115). In this case, the interpolation function corresponds to the trial function. The difference is that the approximation in Eq. (1.117) is written in terms of solution values at nodes, whereas the coefficients c_i in the approximation in Eq. (1.115) do not have any physical meanings. In addition, the interpolation in Eq. (1.117) is limited to the current element, while the approximation in Eq. (1.115) is over the entire domain.

In order to explain the accuracy of the approximation, the interpolated solution and its gradients for two continuous elements are illustrated in Fig. 1.19. Note that in this particular interpolation, the solution is approximated by a piecewise linear function and its gradient is constant within an element. Accordingly, the gradients are not continuous at the element interface. In structural problems, the solution, $u(x)$, often represents displacement of the structure and its gradient is stress or strain. Thus, the approximation yields a continuous displacement, but discontinuous stress and strain between elements. Many commercial finite element programs provide the stress values at nodes and display a smooth change of stresses across elements. However, users must be careful because these nodal stress values are the average of values for different elements connected to a node.

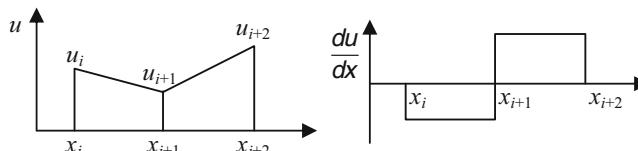


Fig. 1.19 Interpolated solution and its gradient

1.5.2 Finite Element Equations for a One-Dimensional Problem

Once the finite element approximation is available using interpolation functions, it can be used to solve for variational equations. For the sake of explanation, a one-dimensional problem is discussed in this section. Many engineering problems, such as the deformation of a beam and heat conduction in a solid, can be described using a differential equation. The differential equation along with boundary conditions is called the *boundary-valued problem*. A simple, one-dimensional example of a boundary-valued problem is

$$\begin{aligned} \frac{d^2u}{dx^2} + p(x) &= 0, \quad 0 \leq x \leq 1, \\ u(0) = 0 & \\ \left. \frac{du}{dx} \right|_{(1)} &= 0 \end{aligned} \quad \text{Boundary conditions.} \quad (1.118)$$

The above differential equation describes the displacements in a uniaxial bar subjected to a distributed force $p(x)$ along its axis. The first boundary condition prescribes the value of the solution at a given point and is called the *essential boundary condition*. The term displacement boundary condition or kinematic boundary condition is also used in the context of solid mechanics. On the other hand, the second boundary condition prescribes the value of derivative, du/dx , at $x=1$, and is called the *natural boundary condition*. In solid mechanics, the term *force boundary condition* or *stress boundary condition* is also used.

As with the previous section, the principle of virtual work can be used by multiplying the differential equation with a virtual displacement, $\bar{u}(x)$, and integrating over the domain as

$$\int_0^1 \left(\frac{d^2u}{dx^2} + p \right) \bar{u} dx = 0.$$

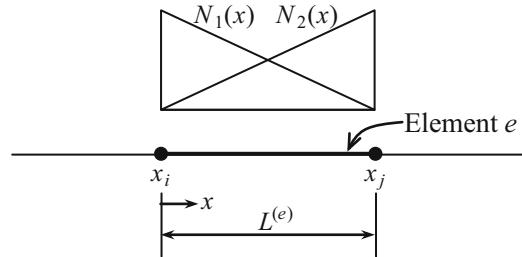
The virtual displacement belongs to the space of kinematically admissible displacements, defined as

$$\mathbb{Z} = \left\{ \bar{u} \in H^{(1)}[0, 1] \mid \bar{u}(0) = 0 \right\}. \quad (1.119)$$

Since the function $p(x)$ is known, we will take the term containing it to the RHS and then use integration-by-parts to the term on the LHS to reduce the order of differentiation of $u(x)$:

$$\left. \frac{du}{dx} \bar{u} \right|_0^1 - \int_0^1 \frac{du}{dx} \frac{d\bar{u}}{dx} dx = - \int_0^1 p \bar{u} dx. \quad (1.120)$$

Fig. 1.20 One-dimensional finite element with interpolation functions



The boundary terms on the LHS of the above equation can be simplified by using the property that $\bar{u}(0) = 0$ at $x = 0$ and the boundary condition of $du(1)/dx = 0$ at $x = 1$. Note that the above variational equation is also satisfied in the element level, in which the boundary terms are the values at the element boundary.

Now, finite element approximation is introduced to solve for the above variational equation. We apply the approximation to one element at a time. Let us consider a general element, say element e , in Fig. 1.20 which has two nodes, say i and $i+1$.

The approximate solution within element e is given by

$$u^{(e)}(x) = u_i N_1(x) + u_{i+1} N_2(x) = \mathbf{N}^{(e)} \cdot \mathbf{d}^{(e)}, \quad (1.121)$$

where $\mathbf{d}^{(e)} = [u_i, u_{i+1}]^T$ is the vector of nodal solutions, and $\mathbf{N}^{(e)} = [N_1, N_2]$ is the vector of interpolation functions for element e from Eq. (1.117). One can also verify that the above interpolation functions yield

$$\begin{aligned} u^{(e)}(x_i) &= u_i, \\ u^{(e)}(x_{i+1}) &= u_{i+1}, \end{aligned}$$

where u_i and u_{i+1} are nodal solutions at nodes i and $i+1$, respectively. The above equation is an important property of interpolation.

Since the solution, $u^{(e)}(x)$, and the virtual displacement, $\bar{u}^{(e)}(x)$, belongs to the same space, \mathbb{Z} , it is natural to use the same interpolation functions to approximate the virtual displacement:

$$\bar{u}^{(e)}(x) = \bar{u}_i N_1(x) + \bar{u}_{i+1} N_2(x) = \mathbf{N}^{(e)} \cdot \bar{\mathbf{d}}^{(e)}, \quad (1.122)$$

where $\bar{\mathbf{d}}^{(e)} = [\bar{u}_i, \bar{u}_{i+1}]^T$ is the vector of virtual displacements for element e .

From Eq. (1.121) the derivative of $u^{(e)}(x)$ is obtained as

$$\frac{du^{(e)}}{dx} = u_i \frac{dN_1}{dx} + u_{i+1} \frac{dN_2}{dx}.$$

The above equation can be written in a matrix form as

$$\frac{du^{(e)}}{dx} = \begin{bmatrix} \frac{dN_1}{dx} & \frac{dN_2}{dx} \end{bmatrix} \begin{Bmatrix} u_i \\ u_{i+1} \end{Bmatrix} = \begin{bmatrix} -\frac{1}{L^{(e)}} & \frac{1}{L^{(e)}} \end{bmatrix} \begin{Bmatrix} u_i \\ u_{i+1} \end{Bmatrix} = \mathbf{B}^{(e)} \cdot \mathbf{d}^{(e)}, \quad (1.123)$$

where $L^{(e)}$ denotes the length of the element. The same equation can be used to approximate the derivatives of virtual displacement by replacing the nodal displacements with the nodal virtual displacements.

Now, we apply the finite element approximation into the variational equation on the element level. By substituting Eqs. (1.122) and (1.123) into Eq. (1.120), we obtain

$$\bar{\mathbf{d}}^{(e)\top} \left[\int_{x_i}^{x_j} \mathbf{B}^{(e)\top} \mathbf{B}^{(e)} dx \right] \mathbf{d}^{(e)} = \bar{\mathbf{d}}^{(e)\top} \int_{x_i}^{x_j} \mathbf{N}^{(e)\top} p(x) dx + \bar{\mathbf{d}}^{(e)\top} \begin{Bmatrix} -\frac{du}{dx}(x_i) \\ +\frac{du}{dx}(x_{i+1}) \end{Bmatrix}.$$

The above variational equation must be satisfied for all virtual displacements, $\bar{u}^{(e)}(x)$. Since element e does not belong to any boundary, the virtual displacements can be any function with the smoothness requirement in Eq. (1.119). In the view of the interpolation scheme in Eq. (1.122), arbitrary virtual displacements can be represented by selecting arbitrary nodal values of \bar{u}_i and \bar{u}_{i+1} . Therefore, if the above equation is rearranged by $\bar{u}_i A + \bar{u}_{i+1} B = 0$, then A and B must vanish individually, from which we can get the following equation for element e :

$$\left[\int_{x_i}^{x_j} \mathbf{B}^{(e)\top} \mathbf{B}^{(e)} dx \right] \mathbf{d}^{(e)} = \int_{x_i}^{x_j} \mathbf{N}^{(e)\top} p(x) dx + \begin{Bmatrix} -\frac{du}{dx}(x_i) \\ +\frac{du}{dx}(x_{i+1}) \end{Bmatrix}.$$

The above equation can be written in matrix form as

$$[\mathbf{k}^{(e)}] \{ \mathbf{d}^{(e)} \} = \{ \mathbf{f}^{(e)} \} + \begin{Bmatrix} -\frac{du}{dx}(x_i) \\ +\frac{du}{dx}(x_{i+1}) \end{Bmatrix}, \quad (1.124)$$

where the coefficient matrix $[\mathbf{k}^{(e)}]$ and the vector $\{ \mathbf{f}^{(e)} \}$ are defined as

$$[\mathbf{k}^{(e)}]_{2 \times 2} = \int_{x_i}^{x_{i+1}} \mathbf{B}^{(e)\top} \mathbf{B}^{(e)} dx = \frac{1}{L^{(e)}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (1.125)$$

and

$$\left\{ \mathbf{f}^{(e)} \right\} = \int_{x_i}^{x_{i+1}} p(x) \begin{Bmatrix} N_1(x) \\ N_2(x) \end{Bmatrix} dx. \quad (1.126)$$

In solid mechanics, the coefficient matrix in Eq. (1.125) is called the element stiffness matrix and the vector in Eq. (1.126) the element force vector. One can derive an equation similar to Eq. (1.124) for each element $e = 1, 2, \dots, N_E$, where N_E is the number of elements.

The RHS of these equations contain terms that are derivatives at the nodes $du(x_i)/dx$ and $du(x_{i+1})/dx$, which are not generally known. However, the second equation for element e can be added to the first equation of element $e+1$ to eliminate this derivative term. To illustrate this point, consider the equations for elements 1 and 2. Two element matrix equations are

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}^{(1)} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix}^{(1)} + \begin{Bmatrix} -\frac{du}{dx}(x_1) \\ +\frac{du}{dx}(x_2) \end{Bmatrix} \quad (1.127)$$

and

$$\begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}^{(2)} \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} f_2 \\ f_3 \end{Bmatrix}^{(2)} + \begin{Bmatrix} -\frac{du}{dx}(x_2) \\ +\frac{du}{dx}(x_3) \end{Bmatrix}. \quad (1.128)$$

We want to combine these two matrix equations into one, which is called the *assembly process*. The assembled matrix equation will have three unknowns: u_1 , u_2 , and u_3 . Therefore, the assembled coefficient matrix will be 3×3 . Equation (1.127) will be added to the first two rows, while Eq. (1.128) will be added to the last two rows. When the second equation in Eq. (1.127) and the first equation in Eq. (1.128) are added together, the boundary term, $du(x_2)/dx$, is canceled. Thus, the assembled matrix equation becomes

$$\begin{bmatrix} k_{11}^{(1)} & k_{12}^{(1)} & 0 \\ k_{21}^{(1)} & k_{22}^{(1)} + k_{11}^{(2)} & k_{12}^{(2)} \\ 0 & k_{21}^{(2)} & k_{22}^{(2)} \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \end{Bmatrix} = \begin{Bmatrix} f_1^{(1)} \\ f_2^{(1)} + f_2^{(2)} \\ f_3^{(2)} \end{Bmatrix} + \begin{Bmatrix} -\frac{du}{dx}(x_1) \\ 0 \\ +\frac{du}{dx}(x_3) \end{Bmatrix}. \quad (1.129)$$

This process can be continued for successive elements, and the $2 \times N_E$ equations for the N_E elements will reduce to $N_E + 1$ equations. In fact $N_E + 1 = N_D$, which is equal to the number of nodes. The N_D equations will take the form

$$\begin{aligned} & \left[\begin{array}{ccccc} k_{11}^{(1)} & k_{12}^{(1)} & 0 & \cdots & 0 \\ k_{21}^{(1)} & k_{22}^{(1)} + k_{11}^{(2)} & k_{12}^{(2)} & \cdots & 0 \\ 0 & k_{21}^{(2)} & k_{22}^{(2)} + k_{11}^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & k_{21}^{(N_E)} & k_{22}^{(N_E)} \end{array} \right] \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{Bmatrix}_{(N_D \times 1)} \\ &= \left\{ \begin{array}{l} f_1^{(1)} \\ f_2^{(1)} + f_2^{(2)} \\ f_3^{(2)} + f_3^{(3)} \\ \vdots \\ f_N^{(N_E)} \end{array} \right\}_{(N_D \times 1)} + \left\{ \begin{array}{l} -\frac{du}{dx}(x_1) \\ 0 \\ 0 \\ \vdots \\ +\frac{du}{dx}(x_N) \end{array} \right\}_{(N_D \times 1)}. \end{aligned} \quad (1.130)$$

In compact form, the above equation is written as

$$[\mathbf{K}]\{\mathbf{d}\} = \{\mathbf{F}\}. \quad (1.131)$$

In general, the global matrix $[\mathbf{K}]$, will be singular and hence the equations cannot be solved directly. However, the matrix will be nonsingular after implementing the boundary conditions. It may be noted that there are N_D unknowns in the N_D equations. At the boundaries ($x=0$ and $x=1$), either u (the essential boundary condition) or du/dx (the natural boundary condition) will be specified. We will illustrate the method in the following example.

Example 1.15 (Three-element solution of a differential equation) Using three elements of equal length, solve the differential equation given below for $p(x)=x$.

$$\begin{aligned} & \frac{d^2u}{dx^2} + p(x) = 0, \quad 0 \leq x \leq 1, \\ & \left. \begin{array}{l} u(0) = 0 \\ u(1) = 0 \end{array} \right\} \quad \text{Boundary conditions.} \end{aligned}$$

Solution Since the elements are of equal length, each element has the length of $L^{(e)} = \frac{1}{3}$. Substituting in Eq. (1.125) the element stiffness matrices for the three elements can be derived as

$$[\mathbf{k}^{(e)}]_{2 \times 2} = \frac{1}{L^{(e)}} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix}, \quad (e = 1, 2, 3).$$

Note that the element stiffness matrices for the three elements are identical. Now the variable in $p(x) = x$ is substituted in Eq. (1.126) to calculate the element force vectors for the three elements and can be derived as

$$\begin{aligned}\{\mathbf{f}^{(e)}\} &= \int_{x_i}^{x_{i+1}} p(x) \begin{Bmatrix} N_1(x) \\ N_2(x) \end{Bmatrix} dx \\ &= \frac{1}{L^{(e)}} \int_{x_i}^{x_{i+1}} \begin{Bmatrix} x(x_{i+1} - x) \\ x(x - x_i) \end{Bmatrix} dx \\ &= L^{(e)} \begin{Bmatrix} \frac{x_i}{3} + \frac{x_{i+1}}{6} \\ \frac{x_i}{6} + \frac{x_{i+1}}{3} \end{Bmatrix}, \quad (e = 1, 2, 3).\end{aligned}$$

Substituting for the element lengths and nodal coordinates

$$\begin{Bmatrix} f_1^{(1)} \\ f_2^{(1)} \end{Bmatrix} = \frac{1}{54} \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}, \quad \begin{Bmatrix} f_2^{(2)} \\ f_3^{(2)} \end{Bmatrix} = \frac{1}{54} \begin{Bmatrix} 4 \\ 5 \end{Bmatrix}, \quad \text{and} \quad \begin{Bmatrix} f_3^{(3)} \\ f_4^{(3)} \end{Bmatrix} = \frac{1}{54} \begin{Bmatrix} 7 \\ 8 \end{Bmatrix}.$$

Now, the global matrix, $[\mathbf{K}]$, and vector, $\{\mathbf{F}\}$, can be assembled using Eq. (1.130) as

$$\begin{bmatrix} 3 & -3 & 0 & 0 \\ -3 & 3+3 & -3 & 0 \\ 0 & -3 & 3+3 & -3 \\ 0 & 0 & -3 & 3 \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{Bmatrix} = \begin{Bmatrix} 1 \\ 54 \\ 2 \\ 54 \end{Bmatrix} - \frac{du}{dx}(0) + \begin{Bmatrix} 4 \\ 54 \\ 5 \\ 54 \end{Bmatrix} + \frac{du}{dx}(1) + \begin{Bmatrix} 7 \\ 54 \\ 8 \\ 54 \end{Bmatrix} + \frac{du}{dx}(1)$$

In the view of the variational equation, the virtual displacements at the essential boundary should be zero, which correspond to the first and last rows in this example, that is, $\bar{u}_1 = \bar{u}_4 = 0$. Therefore, the first and last rows of the above assembled matrix equation are unnecessary. In practice, we discard the first and last rows, as we do not know the RHS of these equations (striking-the-rows). Furthermore, we note that $u_1 = u_4 = 0$. Thus, these two variables are removed, and the first and last columns of matrix $[\mathbf{K}]$ are deleted (striking-the-columns). Then, the four global equations reduce to two equations

$$\begin{bmatrix} 6 & -3 \\ -3 & 6 \end{bmatrix} \begin{Bmatrix} u_2 \\ u_3 \end{Bmatrix} = \frac{1}{9} \begin{Bmatrix} 1 \\ 2 \end{Bmatrix}.$$

Solving the above matrix equation, we obtain $u_2 = 4/18$ and $u_3 = 5/18$. Then using the interpolation functions in Eq. (1.121), the approximate solution at each element can be expressed as

$$u(x) = \begin{cases} \frac{4}{27}x, & 0 \leq x \leq \frac{1}{3} \\ \frac{4}{81} + \frac{1}{27}\left(x - \frac{1}{3}\right), & \frac{1}{3} \leq x \leq \frac{2}{3} \\ \frac{5}{81} - \frac{5}{27}\left(x - \frac{2}{3}\right), & \frac{2}{3} \leq x \leq 1 \end{cases}. \quad (1.132)$$

The exact solution can be obtained by integrating the governing differential equation twice and applying the two essential boundary conditions to solve for the constants:

$$u(x) = \frac{1}{6}x(1-x^2). \quad (1.133)$$

The exact and approximate solutions are plotted in Fig. 1.21. The value of the approximate solution at nodes 2 and 3 coincides with that of the exact solution, but it is actually a coincidence. Otherwise, one can note that the three-element solution is a poor approximation of the exact solution, and more elements are needed to obtain a more accurate solution. This is because the finite element solution is a linear function between nodes, whereas the exact solution is a cubic polynomial in x . ■

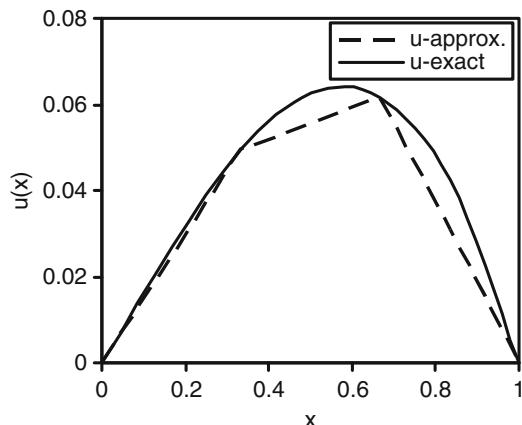


Fig. 1.21 Comparison of exact and approximate solution for Example 3.6

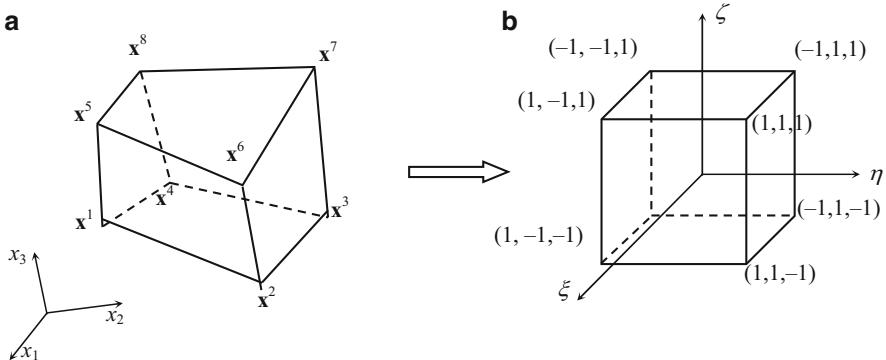


Fig. 1.22 Eight-node three-dimensional isoparametric solid element. (a) Finite element. (b) Reference element

1.5.3 Finite Element Equations for 3D Solid Element

Isoparametric mapping: There are many ways to interpolate a solid component using finite elements. Here, only the eight-node isoparametric hexahedral element is taken as an example. For a more detailed discussion of solid elements, refer to the additional literature on this topic by Zienkiewicz [14], Hughes [15], and Bathe [16]. Figure 1.22 depicts a three-dimensional, eight-node, isoparametric solid element. The element consists of eight nodes and three DOFs at each node. The sequence of node numbers should be given in the same order with the one given in Fig. 1.22a. Since different elements have different shapes, it would not be a trivial task to develop the interpolation functions for individual elements. Instead, the concept of mapping to the reference element will be used. The physical element in Fig. 1.22a will be mapped into the reference element shown in Fig. 1.22b. The physical element is defined in $x_1-x_2-x_3$ coordinates, while the reference element is defined in $\xi-\eta-\zeta$ coordinates. The reference element is a cube with the length of each edge being 2 and has the center at the origin. Although the physical element can have the first node at any corner, the reference element always has the first node at $(-1, -1, -1)$.

The interpolation functions are defined in the reference element so that different elements have the same interpolation function. The only difference is the mapping relation between the two elements. Let $\mathbf{u}_I = [u_{I1}, u_{I2}, u_{I3}]^T$ be the displacement vector at node $I = 1, \dots, 8$ and let $\xi_I = [\xi_I, \eta_I, \zeta_I]^T$ be the corresponding reference coordinate. For the isoparametric element, the coordinate and the displacement of the element can be expressed by

$$\mathbf{x}(\xi) = \sum_{I=1}^8 N_I(\xi) \mathbf{x}_I \quad (1.134)$$

and

$$\mathbf{u}(\xi) = \sum_{I=1}^8 N_I(\xi) \mathbf{u}_I, \quad (1.135)$$

where \mathbf{x}_I is the nodal coordinate and $N_I(\xi)$ is the isoparametric shape function, defined as

$$N_I(\xi) = \frac{1}{8}(1 + \xi\xi_I)(1 + \eta\eta_I)(1 + \zeta\zeta_I), \quad (1.136)$$

where (ξ_I, η_I, ζ_I) are the values of the reference coordinate corresponding to node I , whose values are ± 1 , as shown in Fig. 1.22. Since the above shape functions are Lagrange interpolation functions, they satisfy the same properties as that of the one-dimensional element, that is, N_1 is equal to 1 at node 1 and 0 at other nodes.

The solid element is defined by the coordinates of eight corner nodes: $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_8$. These eight corner nodes are mapped into the eight corner nodes of the reference element. In addition, every point in the physical element is also mapped into a point in the reference element. The mapping relation is one to one such that every point in the reference element also has a mapped point in the physical element. This mapping relation is called an *isoparametric mapping* because the same shape functions are used for interpolating geometry as well as displacements. The above mapping relation is explicit in terms of x_1, x_2 , and x_3 , which means that when ξ, η , and ζ are given, x_1, x_2 , and x_3 can be calculated explicitly from Eq. (1.134). The reverse relation is not straightforward.

Jacobian of mapping: The idea of using the reference element is convenient because it is unnecessary to build different shape functions for different elements. The same shape functions can be used for all elements. However, it has its own drawbacks. The strain energy requires the derivative of displacement, i.e., strains. As can be seen in Eq. (1.89), the strains are defined as derivatives of displacements in the physical coordinates. Since displacements are interpolated using shape functions, it is necessary to differentiate the shape functions with respect to physical coordinates. Since the shape functions are defined in the reference coordinates, differentiation with respect to the physical coordinates is not straightforward. In this case, we use a Jacobian relation and the chain rule of differentiation for that purpose.

The transformation from physical to reference elements can be defined using a mapping relation. The Jacobian matrix of the mapping can be obtained by taking the derivative of Eq. (1.134) as

$$\mathbf{J}_{3 \times 3} = \frac{\partial \mathbf{x}}{\partial \xi} = \sum_{I=1}^8 \mathbf{x}_I \frac{\partial N_I(\xi)}{\partial \xi}. \quad (1.137)$$

Note that $\partial N_I / \partial \xi$ is a (1×3) row vector. The Jacobian matrix is used to relate the derivatives of shape functions between physical and reference coordinates. From the fact that $\xi = \xi(x_1, x_2, x_3)$, $\eta = \eta(x_1, x_2, x_3)$, and $\zeta = \zeta(x_1, x_2, x_3)$, we can write the derivatives of N_I as follows:

$$\begin{aligned}\frac{\partial N_I}{\partial \xi} &= \frac{\partial N_I}{\partial x_1} \frac{\partial x_1}{\partial \xi} + \frac{\partial N_I}{\partial x_2} \frac{\partial x_2}{\partial \xi} + \frac{\partial N_I}{\partial x_3} \frac{\partial x_3}{\partial \xi}, \\ \frac{\partial N_I}{\partial \eta} &= \frac{\partial N_I}{\partial x_1} \frac{\partial x_1}{\partial \eta} + \frac{\partial N_I}{\partial x_2} \frac{\partial x_2}{\partial \eta} + \frac{\partial N_I}{\partial x_3} \frac{\partial x_3}{\partial \eta}, \\ \frac{\partial N_I}{\partial \zeta} &= \frac{\partial N_I}{\partial x_1} \frac{\partial x_1}{\partial \zeta} + \frac{\partial N_I}{\partial x_2} \frac{\partial x_2}{\partial \zeta} + \frac{\partial N_I}{\partial x_3} \frac{\partial x_3}{\partial \zeta}.\end{aligned}$$

Using the matrix form, the above equation can be written as

$$\left\{ \frac{\partial N_I}{\partial \xi} \quad \frac{\partial N_I}{\partial \eta} \quad \frac{\partial N_I}{\partial \zeta} \right\} = \left\{ \frac{\partial N_I}{\partial x_1} \quad \frac{\partial N_I}{\partial x_2} \quad \frac{\partial N_I}{\partial x_3} \right\} \begin{bmatrix} \frac{\partial x_1}{\partial \xi} & \frac{\partial x_1}{\partial \eta} & \frac{\partial x_1}{\partial \zeta} \\ \frac{\partial x_2}{\partial \xi} & \frac{\partial x_2}{\partial \eta} & \frac{\partial x_2}{\partial \zeta} \\ \frac{\partial x_3}{\partial \xi} & \frac{\partial x_3}{\partial \eta} & \frac{\partial x_3}{\partial \zeta} \end{bmatrix}$$

or

$$\frac{\partial N_I}{\partial \xi} = \frac{\partial N_I}{\partial \mathbf{x}} \cdot \mathbf{J}.$$

By using the inverse relation of the above relation, the spatial derivatives of shape functions can be obtained as

$$\frac{\partial N_I}{\partial \mathbf{x}} = \frac{\partial N_I}{\partial \xi} \cdot \mathbf{J}^{-1}. \quad (1.138)$$

As seen from the above equation, the derivative of the shape function cannot be obtained if the Jacobian is zero anywhere in the element. In fact, the mapping relation between (x_1, x_2, x_3) and (ξ, η, ζ) is not valid if the Jacobian is zero or negative anywhere in the element ($-1 \leq \xi, \eta, \zeta \leq 1$). The Jacobian plays an important role in evaluating the validity of mapping as well as the quality of the quadrilateral element. The fundamental requirement is that every point in the reference element should be mapped into the interior of the physical element and vice versa. When an interior point in the (ξ, η, ζ) coordinates is mapped into an exterior point in the (x_1, x_2, x_3) coordinates, the Jacobian becomes negative. If multiple points in (ξ, η, ζ) coordinates are mapped into a single point in (x_1, x_2, x_3) coordinates, the Jacobian becomes zero at that point. Thus, it is important to

maintain the element shape so that the Jacobian is positive everywhere in the element.

Displacement–strain relation: Using the spatial derivatives of shape functions, the strain vector can be obtained in the following form:

$$\boldsymbol{\epsilon}(\mathbf{u}) = \sum_{I=1}^8 \mathbf{B}_I \mathbf{u}_I, \quad (1.139)$$

where

$$\mathbf{B}_I = \begin{bmatrix} N_{I,1} & 0 & 0 \\ 0 & N_{I,2} & 0 \\ 0 & 0 & N_{I,3} \\ N_{I,2} & N_{I,1} & 0 \\ 0 & N_{I,3} & N_{I,2} \\ N_{I,3} & 0 & N_{I,1} \end{bmatrix} \quad (1.140)$$

is the discrete displacement–strain matrix of a solid element. In the above equation, $N_{I,1}$ is the spatial derivative of the shape function in Eq. (1.138). The approximation of the strain variation, $\boldsymbol{\epsilon}(\bar{\mathbf{u}})$, can be obtained in a similar way by replacing the nodal displacements with nodal virtual displacements.

Element stiffness matrix and element force vector: Note that all variables in the physical element are mapped into the reference element. Thus, it would be helpful if the integration over element domain, $\Omega^{(e)}$, can be converted into integration over the reference element, which can be achieved using the following relation:

$$\iiint_{\Omega^{(e)}} d\Omega^{(e)} = \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 |\mathbf{J}| d\xi d\eta d\zeta.$$

In order to derive the element stiffness matrix, it is assumed that the entire structural domain is discretized by a set of finite elements. The energy-bilinear form in Eq. (1.100) is constructed for each element, and then, they will be assembled to construct the global stiffness matrix. The energy-bilinear form of the element in Eq. (1.100) can be approximated as

$$a(\mathbf{u}, \mathbf{u}) = \sum_{I=1}^8 \sum_{J=1}^8 \bar{\mathbf{u}}_I^T \left[\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 \mathbf{B}_I^T \mathbf{D} \mathbf{B}_J |\mathbf{J}| d\xi d\eta d\zeta \right] \mathbf{u}_J \equiv \{\bar{\mathbf{d}}\}^T [\mathbf{k}] \{\mathbf{d}\}, \quad (1.141)$$

where $\{\mathbf{d}\} = [u_{11}, u_{12}, u_{13}, u_{21}, u_{22}, u_{23}, \dots, u_{81}, u_{82}, u_{83}]^T$ is the vector of nodal displacements, $\{\bar{\mathbf{d}}\}$ is the vector of nodal virtual displacements, and $[\mathbf{k}]$ is the 24×24 element stiffness matrix. Instead of having the summation over all eight nodes in the above equation, it is possible for the displacement–strain matrix to be augmented for all nodes by $[\mathbf{B}] = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_8]$. Then, the above equation can

be written without having the summation. The MATLAB code in the following section will use this convention.

The load form in Eq. (1.101) contains the body force term and the surface traction term. For simplicity, we only consider the case with the body force. The load form is discretized without the traction force by

$$\ell(\bar{\mathbf{u}}) = \sum_{I=1}^8 \bar{\mathbf{u}}_I^T \int_{-1}^1 \int_{-1}^1 \int_{-1}^1 N_I(\xi) \mathbf{f}^b |\mathbf{J}| d\xi d\eta d\zeta \equiv \{\bar{\mathbf{d}}\}^T \{\mathbf{f}\}. \quad (1.142)$$

By equating the above two equations, the discrete variational equation of a solid element can be obtained as

$$\{\bar{\mathbf{d}}\}^T [\mathbf{k}] \{\mathbf{d}\} = \{\bar{\mathbf{d}}\}^T \{\mathbf{f}\}, \quad \forall \{\bar{\mathbf{d}}\} \in \mathbb{Z}_h, \quad (1.143)$$

where $\mathbb{Z}_h \subset R^{24}$ is the discrete space of kinematically admissible displacements.

Numerical integration: The finite element formulation requires integration over the domain or over the boundary during the construction of the element stiffness matrix and element force vector. Analytical integration is limited to simple one-dimensional problems. Most integrals cannot be evaluated explicitly, and it is often faster to integrate them numerically rather than to evaluate them analytically. Among many numerical integration methods that have been proposed, a Gauss integration rule is commonly used in the finite element formulation due to its simplicity and accuracy. In this section, a brief introduction to the Gauss integration rule is provided. A rigorous study of numerical integration, including error estimates, can be found in Chap. 5 of Atkinson [17].

Consider one-dimensional integration of a function $f(\xi)$ over the interval $[-1, 1]$. Although the integration interval can be arbitrary, the interval $[-1, 1]$, is used without the loss of generality because it is convenient to apply the interval to the reference element in the finite element formulation. A general form of Gauss integration can be written as

$$\int_{-1}^1 f(\xi) d\xi \approx \sum_{i=1}^{NG} \omega_i f(\xi_i), \quad (1.144)$$

where NG is the number of integration points, ξ_i is the integration point, and ω_i is the nonnegative integration weight. The integration points and weights are chosen such that the right side of Eq. (1.144) approximates the left side polynomials, $f(\xi)$, as accurately as possible. In general, an NG -point Gauss integration method integrates $(2NG - 1)$ -order polynomials, exactly. This method is extremely accurate in most cases, and is the one that is the most frequently used in modern finite element formulations. Table 1.4 summarizes the integration points and weights for Gauss integration.

Table 1.4 Gauss integration points and weights

NG	Integration points (ξ_i)	Weights (ω_i)
1	0.0	2.0
2	± 0.5773502692	1.0
3	± 0.7745966692	0.5555555556
	0.0	0.8888888889
4	± 0.8611363116	0.3478546451
	± 0.3399810436	0.6521451549
5	± 0.9061798459	0.2369268851
	± 0.5384693101	0.4786286705
	0.0	0.5688888889

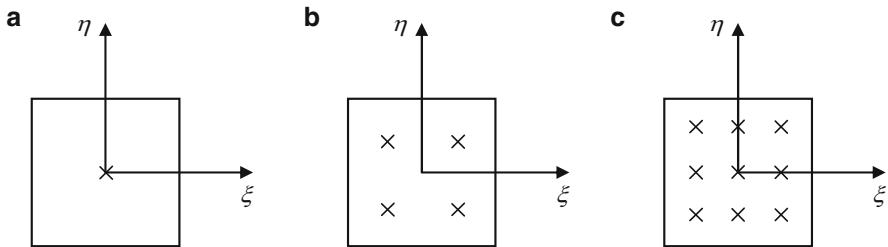


Fig. 1.23 Gauss integration points in two-dimensional reference elements. (a) 1×1 . (b) 2×2 . (c) 3×3

A multidimensional integration can be constructed by employing the one-dimensional integration rule on each coordinate separately. In two- and three-dimensional domains, the Gauss integration rule can be written as

$$\int_{-1}^1 \int_{-1}^1 f(\xi, \eta) d\xi d\eta = \sum_{i=1}^{NG} \sum_{j=1}^{NG} \omega_i \omega_j f(\xi_i, \eta_j) \quad (1.145)$$

and

$$\int_{-1}^1 \int_{-1}^1 \int_{-1}^1 f(\xi, \eta, \zeta) d\xi d\eta d\zeta = \sum_{i=1}^{NG} \sum_{j=1}^{NG} \sum_{k=1}^{NG} \omega_i \omega_j \omega_k f(\xi_i, \eta_j, \zeta_k), \quad (1.146)$$

respectively. Figure 1.23 illustrates the integration points in two-dimensional reference elements. The computational cost of Gauss integration is proportional to $(NG)^2$ for two-dimensional problems and $(NG)^3$ for three-dimensional problems.

1.5.4 A MATLAB Code for Finite Element Analysis

In Chap. 2, a MATLAB code, NLFEA.m, will be introduced, which solves for nonlinear problems using eight-node hexahedral elements. Detailed usage of NLFEA.m will be explained in Chap. 2. In general, a nonlinear finite element analysis code can also solve for linear problems. In this section, MATLAB codes, SHAPEL.m and ELAST3D.m, are introduced that can solve for linear structural problems using NLFEA.m. The first code, SHAPEL.m, calculates the shape functions, their derivatives, and determinant of Jacobian of an eight-node hexahedral element. It takes two input variables, XI and ELXY . The vector, XI , of the reference coordinates, (ξ, η, ζ) , is the location where the shape functions and their derivatives are calculated, and ELXY is the 3×8 matrix that contains the nodal coordinates of eight nodes of the element. Since shape functions are normally calculated at Gauss integration points, XI often contains the (ξ, η, ζ) coordinates of the Gauss integration points. The SHAPEL function returns with three variables: SF array, (8×1) , contains shape functions; GDSF array, (3×8) , contains the derivative of shape functions; and DET is the Jacobian of the mapping. The user can check the validity of the mapping using the Jacobian. In the code, GJ array stores the 3×3 Jacobian matrix, and XNODE array stores the reference coordinates (ξ_I, η_I, ζ_I) of the eight corner nodes that are used in Eq. (1.136).

The second code, ELAST3D.m, assembles the element stiffness matrices to the global stiffness matrix. It has several input variables, whose meanings are summarized in Table 1.5. Note that the two logical variables, UPDATE and LTAN, are introduced to make the program more efficient. When UPDATE is true, then ELAST3D.m will calculate stresses at each integration point and store them into the global array, SIGMA. This process is necessary when stress values are printed out by NLFEA.m. When LTAN is true, then the global stiffness matrix is assembled. In linear analysis, LTAN should be true all the time. ELAST3D.m calls SHAPEL.m to calculate the shape functions and their derivatives at each integration point. As explained in Eq. (1.139), the strain is calculated by multiplying nodal displacements with the derivatives of shape functions as

Table 1.5 Input variables for ELAST3D.m

Variable	Array size	Meaning
ETAN	(6,6)	Elastic stiffness matrix Eq. (1.81)
UPDATE	Logical variable	If true, save stress values
LTAN	Logical variable	If true, calculate the global stiffness matrix
NE	Integer	Total number of elements
NDOF	Integer	Dimension of problem (3)
XYZ	(3,NNODE)	Coordinates of all nodes
LE	(8,NE)	Element connectivity

```
% Strain
DEPS=DSP*SHPD';
DDEPS=[DEPS(1,1) DEPS(2,2) DEPS(3,3) ...
DEPS(1,2)+DEPS(2,1) DEPS(2,3)+DEPS(3,2) DEPS(1,3)+DEPS(3,1)]';
```

Then, stress can be obtained by multiplying the elastic stiffness matrix with the strain as

```
% Stress
STRESS = ETAN*DDEPS;
```

One thing that is uncommon for linear analysis is the residual force array, FORCE, which is different from the external applied force. The FORCE array is required because the linear problem is solved as a nonlinear problem. A detailed explanation of residual force will be provided in Chap. 2. In order to assemble the local stiffness matrix into the global stiffness matrix, the IDOF array is used to store the location of the global DOFs corresponding to the local 24 DOFs. The XG and WGT arrays store one-dimensional integration points and corresponding weights, as in Table 1.4. In this implementation, only two-point integration is used for each coordinate direction.

```
function [SF, GDSF, DET] = SHAPEL(XI, ELXY)
%*****
% Compute shape function, derivatives, and determinant of hexahedral
element
%*****
%%
XNODE=[-1 1 1 -1 -1 1 1 -1;
        -1 -1 1 1 -1 -1 1 1;
        -1 -1 -1 -1 1 1 1 1];
QUAR = 0.125;
SF=zeros(8,1);
DSF=zeros(3,8);
for I=1:8
    XP = XNODE(1,I);
    YP = XNODE(2,I);
    ZP = XNODE(3,I);
    %
    XI0 = [1+XI(1)*XP 1+XI(2)*YP 1+XI(3)*ZP];
    %
    SF(I) = QUAR*XIO(1)*XIO(2)*XIO(3);
    DSF(1,I) = QUAR*XP*XIO(2)*XIO(3);
    DSF(2,I) = QUAR*YP*XIO(1)*XIO(3);
    DSF(3,I) = QUAR*ZP*XIO(1)*XIO(2);
end
```

```

GJ = DSF*ELXY;
DET = det(GJ);
GJINV=inv(GJ);
GDSF=GJINV*DSF;
end


---


function ELAST3D(ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE)
%*****
% Main program computing global stiffness matrix and residual force for
% linear elastic material model.
%*****
%%
global DISPTD FORCE GKF SIGMA
%
% Integration points and weights (2-point integration)
XG=[-0.57735026918963D0, 0.57735026918963D0];
WGT=[1.00000000000000D0, 1.00000000000000D0];
%
% Stress storage index (No. of integration points)
INTN=0;
%
% Loop over elements, this is main loop for computing K and F
for IE=1:NE
    % Element nodal coordinates
    ELXY=XYZ(LE(IE,:));
    % Local to global mapping
    IDOF=zeros(1,24);
    for I=1:8
        II=(I-1)*NDOF+1;
        IDOF(II:II+2)=(LE(IE,I)-1)*NDOF+1:(LE(IE,I)-1)*NDOF+3;
    end
    DSP=DISPTD(IDOF);
    DSP=reshape(DSP,NDOF,8);
    %
    % Loop over integration points
    for LX=1:2, for LY=1:2, for LZ=1:2
        E1=XG(LX); E2=XG(LY); E3=XG(LZ);
        INTN = INTN + 1;
        %
        % Determinant and shape function derivatives
        [~, SHPD, DET] = SHAPEL([E1 E2 E3], ELXY);
        FAC=WGT(LX)*WGT(LY)*WGT(LZ)*DET;
        %
        % Strain
        DEPS=DSP*SHPD';
        DDEPS=[DEPS(1,1) DEPS(2,2) DEPS(3,3) ...
            DEPS(1,2)+DEPS(2,1) DEPS(2,3)+DEPS(3,2) DEPS(1,3)+DEPS(3,1)]';
        %
    end
end

```

```

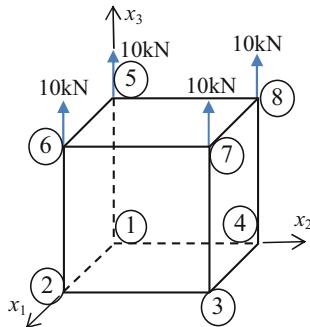
% Stress
STRESS = ETAN*DDEPS;
%
% Update stress (Store stress)
if UPDATE
    SIGMA(:,INTN)=STRESS;
    continue;
end
%
% Add residual force and stiffness matrix
BM=zeros(6,24);
for I=1:8
    COL=(I-1)*3+1:(I-1)*3+3;
    BM(:,COL)=[SHPD(1,I) 0 0;
                0 SHPD(2,I) 0;
                0 0 SHPD(3,I);
                SHPD(2,I) SHPD(1,I) 0;
                0 SHPD(3,I) SHPD(2,I);
                SHPD(3,I) 0 SHPD(1,I)];
end
%
% Residual forces
FORCE(IDOF) = FORCE(IDOF) - FAC*BM'*STRESS;
%
% Tangent stiffness
if LTAN
    EKF = BM'*ETAN*BM;
    GKF(IDOF, IDOF)=GKF(IDOF, IDOF)+FAC*EKF;
end
end, end, end
end
end

```

Example 1.16 (Uniaxial tension of a cube) Using NLFEA.m in Chap. 2, calculate displacement and stress of a three-dimensional brick element under uniaxial tension as shown in Fig. 1.24. Assume an isotropic material with the two Lame's constants of $\lambda = 110.7$ GPa and $\mu = 80.2$ GPa.

Solution NLFEA.m can be called with appropriate model definitions as presented in Chap. 2. A nodal force of 10 kN is applied at the four nodes on the top, while the bottom four nodes are fixed in such a way that the uniaxial tension condition can be met, that is, $u_{1x} = u_{1y} = u_{1z} = u_{2x} = u_{2y} = u_{2z} = u_{3x} = u_{4x} = 0$. Nodal coordinates are defined in **XYZ** array, and element connectivity is in **LE** array. **EXTFORCE** stores externally applied force, and **SDISPT** stores prescribed displacements. These two arrays are given in the format such that each row includes [Node, DOF, Value] format. For linear elastic material, **MID = 0** is used with two material constants in

Fig. 1.24 A brick element under uniaxial tension



PROP array, that is, two Lame's constants **PROP** = [**LAMBDA**, **MU**]. Load increment array, **TIMS**, and program parameters are mainly designed for nonlinear analysis; therefore, they will be discussed in Chap. 2.

```
%  
% One element example  
%  
% Nodal coordinates  
XYZ=[0 0 0;1 0 0;1 1 0;0 1 0;0 0 1;1 0 1;1 1 1;0 1 1];  
%  
% Element connectivity  
LE=[1 2 3 4 5 6 7 8];  
%  
% External forces [Node, DOF, Value]  
EXTFORCE=[5 3 10.0E3; 6 3 10.0E3; 7 3 10.0E3; 8 3 10.0E3];  
%  
% Prescribed displacements [Node, DOF, Value]  
SDISPT=[1 1 0;1 2 0;1 3 0;2 2 0;2 3 0;3 3 0;4 1 0;4 3 0];  
%  
% Material properties  
% MID:0 (Linear elastic) PROP=[LAMBDA NU]  
MID=0;  
PROP=[110.747E3 80.1938E3];  
%  
% Load increments [Start End Increment InitialFactor FinalFactor]  
TIMS=[0.0 1.0 1.0 0.0 1.0]';  
%  
% Set program parameters  
ITRA=30; ATOL=1.0E5; NTOL=6; TOL=1E-6;  
%  
% Calling main function  
NOUT = fopen('output.txt','w');  
NLFEA(ITRA, TOL, ATOL, NTOL, TIMS, NOUT, MID, PROP, EXTFORCE, SDISPT, XYZ, LE);  
fclose(NOUT);
```

Even if the problem is linear, NLFEA.m will solve it as if it is a nonlinear problem. However, due to linear relationship between the applied load and displacement, the Newton–Raphson iteration will converge in the first iteration. The following shows the command-line output after calling NLFEA.m.

```
Time Time step Iter Residual
1.00000 1.000e+00 2 5.45697e-12
```

The details of **Time** and **Time step** will be discussed in Chap. 2. It shows **Iter = 2** because NLFEA.m calculates the internal and external forces at **Iter = 1**, and then checks the convergence at **Iter = 2**. Therefore, even if it shows **Iter = 2**, the actual Newton–Raphson iteration is equal to 1. **Residual** is the maximum norm of the difference between the internal and external loads. The convergence is determined based on the magnitude of the residual.

NLFEA.m stores analysis results in output.txt file, which includes nodal displacements and element stress at all integration points. Below is the contents of output.txt file:

```
TIME = 1.000e+00
Nodal Displacements
Node U1 U2 U3
1 0.000e+00 0.000e+00 0.000e+00
2 -5.607e-08 0.000e+00 0.000e+00
3 -5.607e-08 -5.607e-08 0.000e+00
4 0.000e+00 -5.607e-08 0.000e+00
5 -5.494e-23 1.830e-23 1.933e-07
6 -5.607e-08 4.061e-23 1.933e-07
7 -5.607e-08 -5.607e-08 1.933e-07
8 -8.032e-23 -5.607e-08 1.933e-07

Element Stress
S11 S22 S33 S12 S23 S13
Element 1
0.000e+00 1.091e-11 4.000e+04 -2.322e-13 6.633e-13 -3.317e-12
0.000e+00 0.000e+00 4.000e+04 -3.980e-13 1.327e-13 -9.287e-13
-3.638e-12 7.276e-12 4.000e+04 -1.592e-12 -2.123e-12 -3.317e-12
0.000e+00 0.000e+00 4.000e+04 2.653e-13 -2.123e-12 5.307e-13
0.000e+00 0.000e+00 4.000e+04 5.638e-13 3.449e-12 -1.327e-12
0.000e+00 0.000e+00 4.000e+04 -1.194e-12 4.776e-12 1.061e-12
0.000e+00 0.000e+00 4.000e+04 -7.960e-13 2.919e-12 -3.449e-12
3.638e-12 3.638e-12 4.000e+04 -5.307e-13 3.715e-12 1.061e-12
*** Successful end of program ***
```

Since the applied load and boundary conditions are such that the stress of the cube is in the uniaxial tension in the z-coordinate direction. As expected, the cube is

extended in z -coordinate direction ($u_{5z} = u_{6z} = u_{7z} = u_{8z} = 1.933e - 07$) based on nodal displacements. Due to Poisson's effect, there is lateral contraction ($u_x = u_y = -5.607e - 08$). It is noted that the displacements of those degrees-of-freedom whose displacement is prescribed in **SDISPT** array exactly satisfy the prescribed displacements. However, for other degrees-of-freedom whose values are not prescribed but supposed to be zero have a very small value, such as $u_{5y} = 1.830e - 23$ due to numerical error. These values should be interpreted as zero. Among the six components of stress, only **S33 = 4.000e+04N** is the only nonzero stress component; all other components are effectively zero within numerical error. Note that this value of stress is expected as the applied load is 40 kN and the area is 1. ■

1.6 Exercises

- P1.1 Using Cartesian bases, show that $(\mathbf{u} \otimes \mathbf{v}) \cdot (\mathbf{w} \otimes \mathbf{x}) = (\mathbf{v} \cdot \mathbf{w})\mathbf{u} \otimes \mathbf{x}$ where \mathbf{u} , \mathbf{v} , \mathbf{w} , and \mathbf{x} are rank-1 tensors.
- P1.2 Any rank-2 tensor \mathbf{T} can be decomposed by $\mathbf{T} = \mathbf{S} + \mathbf{W}$, where \mathbf{S} is the symmetric part of \mathbf{T} and \mathbf{W} is the skew part of \mathbf{T} . Let \mathbf{A} be a symmetric rank-2 tensor. Show $\mathbf{A} : \mathbf{W} = 0$ and $\mathbf{A} : \mathbf{T} = \mathbf{A} : \mathbf{S}$.
- P1.3 For a symmetric rank-2 tensor, \mathbf{E} , using the index notation, show that $\mathbf{I} : \mathbf{E} = \mathbf{E}$, where $\mathbf{I} = \frac{1}{2}[\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}]$ is a symmetric unit tensor of rank-4.
- P1.4 The deviator of a symmetric rank-2 tensor is defined as $\mathbf{A}_{\text{dev}} = \mathbf{A} - A^m \mathbf{1}$, where $A^m = \frac{1}{3}(A_{11} + A_{22} + A_{33})$. Find the rank-4 deviatoric identity tensor, \mathbf{I}_{dev} , that satisfies $\mathbf{A}_{\text{dev}} = \mathbf{I}_{\text{dev}} : \mathbf{A}$.
- P1.5 The norm of a rank-2 tensor is defined as $\|\mathbf{A}\| = \sqrt{\mathbf{A} : \mathbf{A}}$. Calculate the following derivative, $\partial \|\mathbf{A}\| / \partial \mathbf{A}$. What is the rank of the derivative?
- P1.6 A rank-2 unit tensor in the direction of rank-2 tensor \mathbf{A} can be defined as $\mathbf{N} = \mathbf{A} / \|\mathbf{A}\|$. Show that $\partial \mathbf{N} / \partial \mathbf{A} = [\mathbf{I} - \mathbf{N} \otimes \mathbf{N}] / \|\mathbf{A}\|$.
- P1.7 Through direct calculation of a rank-2 tensor, show that the following identity $e_{rst} \det[\mathbf{A}] = e_{ijk} A_{ir} A_{js} A_{kt}$ is true.
- P1.8 For a vector, $\mathbf{r} = x_1 \mathbf{e}_1 + x_2 \mathbf{e}_2 + x_3 \mathbf{e}_3$, and its norm, $r = \|\mathbf{r}\|$, prove $\nabla \cdot (r \mathbf{r}) = 4r$.
- P1.9 A velocity gradient is decomposed into symmetric and skew parts, $\nabla \mathbf{v} = \mathbf{d} + \boldsymbol{\omega}$, where

$$d_{ij} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right), \quad \omega_{ij} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} - \frac{\partial v_j}{\partial x_i} \right).$$

Show that

(a) For a symmetric stress tensor, $\boldsymbol{\sigma} : \nabla \mathbf{v} = \boldsymbol{\sigma} : \mathbf{d}$.

$$(b) w_{ij} = \frac{1}{2} e_{ijk} e_{mnk} \frac{\partial v_m}{\partial x_n}.$$

P1.10 A symmetric rank-4 tensor is defined by $\mathbf{D} = \lambda \mathbf{1} \otimes \mathbf{1} + 2\mu \mathbf{I}$, where $\mathbf{1} = [\delta_{ij}]$ is a unit tensor of rank-2 and $\mathbf{I} = \frac{1}{2} [\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}]$ is a symmetric unit tensor of rank-4. When \mathbf{E} is an arbitrary symmetric rank-2 tensor, calculate $\mathbf{S} = \mathbf{D} : \mathbf{E}$ in terms of \mathbf{E} .

P1.11 Using integration-by-parts, calculate $I = \int x \cos(x) dx$.

P1.12 Using integration-by-parts, calculate $I = \int e^x \cos(x) dx$.

P1.13 Calculate the surface integral of the vector function, $\mathbf{F} = x\mathbf{e}_1 + y\mathbf{e}_2$, over the portion of the surface of the unit sphere, $S: x^2 + y^2 + z^2 = 1$, above the xy -plane, i.e., $z \geq 0$.

$$\int_S \mathbf{F} \cdot \mathbf{n} dS.$$

P1.14 Evaluate the surface integral of a vector, $\mathbf{F} = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$, over the closed surface of the cube bounded by the planes, $x = \pm 1, y = \pm 1, z = \pm 1$, using the divergence theorem.

$$\int_S \mathbf{F} \cdot \mathbf{n} dS.$$

P1.15 Consider a unit-depth (in z -axis) infinitesimal element as shown in the figure. Using force equilibrium, derive the governing differential equation in two dimensions (equilibrium in x - and y -directions). Assume that the uniform body forces, $\mathbf{f}^B = [f_1^B, f_2^B]$, are applied to the infinitesimal element.

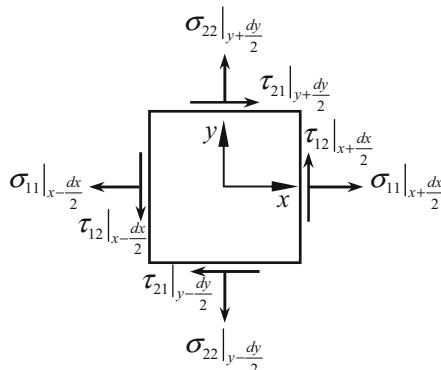


Fig. P1.15

P1.16 In the above unit-depth (in z -axis), infinitesimal element, show that the stress tensor is symmetric using moment equilibrium.

P1.17 The principal stresses at a point in a body are given by $\sigma_1 = 4$, $\sigma_2 = 2$ and $\sigma_3 = 1$, and the principal directions of the first two principal stresses are given by $\mathbf{n}^{(1)} = \frac{1}{\sqrt{2}}(0, 1, -1)$ and $\mathbf{n}^{(2)} = \frac{1}{\sqrt{2}}(0, 1, 1)$. Determine the state of stress at the point, i.e., the six components of a stress tensor.

P1.18 Find the principal stresses and the corresponding principal stress directions for the following cases of plane stress:

- (a) $\sigma_{11} = 40$ MPa, $\sigma_{22} = 0$ MPa, $\sigma_{12} = 80$ MPa.
- (b) $\sigma_{11} = 140$ MPa, $\sigma_{22} = 20$ MPa, $\sigma_{12} = -60$ MPa.
- (c) $\sigma_{11} = -120$ MPa, $\sigma_{22} = 50$ MPa, $\sigma_{12} = 100$ MPa.

P1.19 Determine the principal stresses and their associated directions, when the stress matrix at a point is given by

$$[\boldsymbol{\sigma}] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 1 \end{bmatrix} \text{ MPa.}$$

P1.20 Let the $x'y'z'$ coordinate system be defined using the three principal directions obtained from Problem P1.19. Determine the transformed stress matrix, $[\boldsymbol{\sigma}]_{x'y'z'}$, in the new coordinate system.

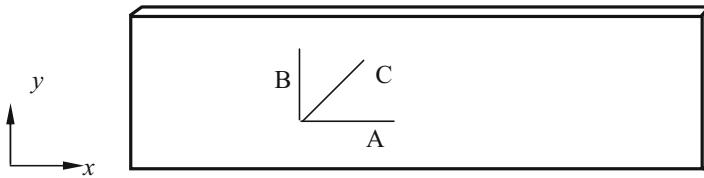
P1.21 The stress-strain relationship for a three-dimensional isotropic solid is given as $\sigma_{ij} = [K\delta_{ij}\delta_{kl} + 2\mu(\delta_{ik}\delta_{jl} - \frac{1}{3}\delta_{ij}\delta_{kl})]\varepsilon_{kl}$, where K is the bulk modulus and μ is the shear modulus. In practice, stress and strain are written in the vector forms such that $\{\boldsymbol{\sigma}\} = \{\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{13}\}^T$ and $\{\boldsymbol{\varepsilon}\} = \{\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \gamma_{12}, \gamma_{23}, \gamma_{13}\}^T$. Then, the stress-strain can be written as $\{\boldsymbol{\sigma}\} = [\mathbf{D}]\{\boldsymbol{\varepsilon}\}$. Write the expression of a 6×6 elasticity matrix, $[\mathbf{D}]$, in terms of K and μ .

P1.22 For steel, the following material data are applicable: Young's modulus, $E = 207$ GPa, and shear modulus, $G = 80$ GPa. For the strain matrix at a point, shown below, determine the symmetric 3×3 stress matrix.

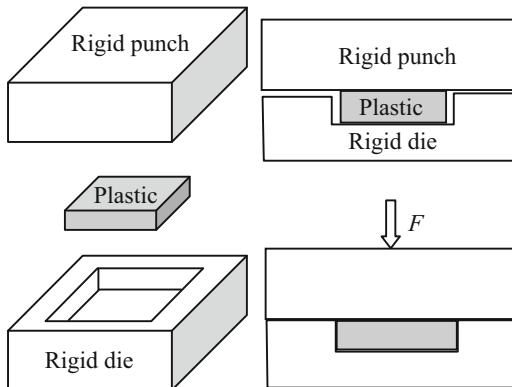
$$[\boldsymbol{\varepsilon}] = \begin{bmatrix} 0.003 & 0 & -0.006 \\ 0 & -0.001 & 0.003 \\ -0.006 & 0.003 & 0.0015 \end{bmatrix}.$$

P1.23 A strain rosette consisting of three strain gages was used to measure the strains at a point in a thin-walled plate. The measured strains in the three gages are $\varepsilon_A = 0.001$, $\varepsilon_B = -0.0006$, and $\varepsilon_C = 0.0007$. Note that gage C is at 45° with respect to the x -axis.

- (a) Determine the complete state of strains and stresses (all six components) at that point. Assume that $E = 70$ GPa, and $\nu = 0.3$.
- (b) What are the principal strains and their directions?
- (c) What are the principal stresses and their directions?
- (d) Show that the principal strains and stresses satisfy the stress-strain relations.

**Fig. P1.23**

P1.24 A rectangular plastic specimen of size $100 \times 100 \times 10 \text{ mm}^3$ is placed in a rectangular metal cavity. The dimensions of the cavity are $101 \times 101 \times 9 \text{ mm}^3$. The plastic is compressed by a rigid punch until it is completely inside the cavity. Due to the Poisson's effect, the plastic also expands in the *x*- and *y*-directions and fills the cavity. Calculate all stress and strain components and the force exerted by the punch. Assume that there is no friction between all contacting surfaces. The metal cavity is rigid. Elastic constants of the plastic are $E = 10 \text{ GPa}$ and $\nu = 0.3$.

**Fig. P1.24**

P1.25 Repeat Problem P1.24 with the elastic constants of the plastic defined as $E = 10 \text{ GPa}$ and $\nu = 0.485$.

P1.26 The strain energy and work done by applied loads are given in the following equations. When the solution is expressed by $u(x) = c_1x + c_2x^2$, calculate the solution using the principle of minimum potential energy.

$$U = \frac{1}{2} \int_0^1 (u')^2 dx, \quad W = \int_0^1 u dx + u(1).$$

P1.27 The governing differential equation for the bar component in the figure is given as

$$\begin{aligned} -(EA(x)u_{,1})_{,1} &= f(x), \quad x \in (0, l), \\ u(0) &= 0, \\ u_{,1}(l) &= 0, \end{aligned}$$

where the subscripted comma denotes the differentiation with respect to the spatial coordinate, i.e., $u_{,1} = du/dx$. Derive the weak form using the principle of virtual work.

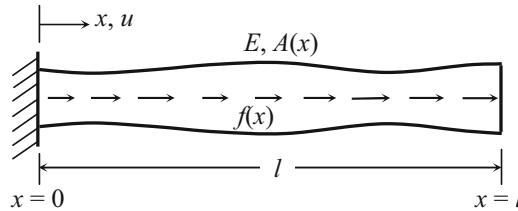


Fig. P1.27

P1.28 Derive the weak form of a two-dimensional, steady-state heat transfer problem.

P1.29 Derive the weak form of a simply supported beam problem.

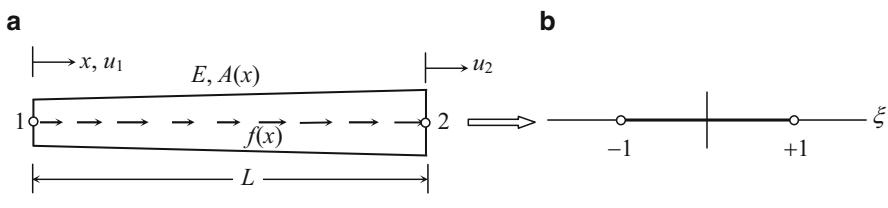
P1.30 When the potential energy of P1.29 is given, derive the variational equation using the principle of minimum potential energy.

$$\Pi = \int_0^L \left(\frac{1}{2} EI(v_{,11})^2 - fv \right) dx.$$

P1.31 Derive the principle of virtual work for the simply supported Kirchhoff plate element from the governing equation:

$$[D(u_{,11} + \nu u_{,22})]_{,11} + [D(u_{,22} + \nu u_{,11})]_{,22} + 2(1 - \nu)[Du_{,12}]_{,12} = f.$$

P1.32 Consider a bar element as shown in the figure. The cross-sectional areas are A_1 and A_2 at nodes 1 and 2, respectively, and vary linearly. In addition, the gravitational acceleration is applied along the axial direction of the bar, such that the distributed load per unit length is $f(x) = \rho g A(x)$, where ρ is the density and g is the gravitational acceleration. Construct the discrete variational equation for the element.



Finite Element

Reference Element

Fig. P1.32

- P1.33 For the Euler beam element shown in the figure, derive the interpolation functions, $N_i(\xi)$, stiffness matrix, \mathbf{k} , and nodal force vector, \mathbf{f} . Assume that the uniformly distributed load is $f(x) = f$. Note that the reference element is defined in the domain, $\xi = [-1, 1]$.

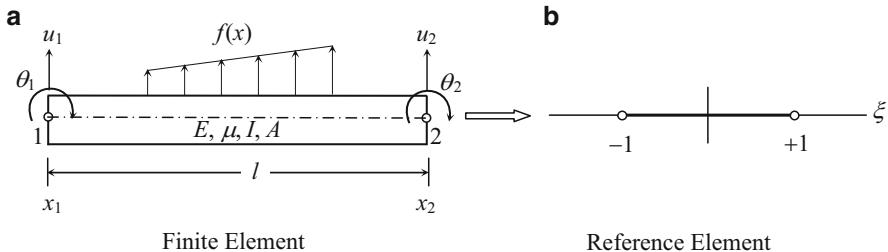


Fig. P1.33

- P1.34 Below is the governing differential equation of one-dimensional bar under uniformly distributed load. Using one bar element, calculate the displacement at $x = L$ and $x = \frac{L}{2}$. Compare these displacements with that of the exact calculation. (Note: the exact solution can be calculated by integrating the differential equation twice.)

$$\begin{aligned} -EAu_{,11} &= f, \quad x \in (0, L), \\ u(0) &= 0, \\ u_{,1}(L) &= 0. \end{aligned}$$

- P1.35 An Euler beam element shown in the figure is under a uniformly distributed couple, C . Calculate the equivalent nodal forces. Using a simply supported beam under a uniform couple, show that the reaction forces are equal and opposite in directions with the equivalent nodal forces

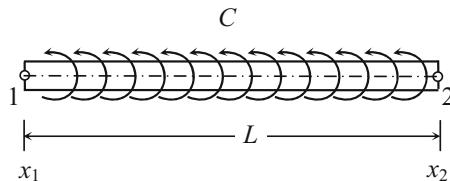


Fig. P1.35

- P1.36 Integrate the following function using one-point and two-point numerical integration (Gauss quadrature). Explain how to integrate it. The exact integral is equal to 2. Compare the accuracy of the numerical integration with the exact one.

$$I = \int_0^{\pi} \sin(x) \, dx.$$

References

1. Kreyszig E. Advanced engineering mathematics. 5th ed. New York: Wiley; 1983.
2. Strang G. Linear algebra and its applications. 2nd ed. New York: Academic; 1980.
3. Hildebrand FB. Advanced calculus for applications. 2nd ed. Englewood Cliffs: Prentice-Hall; 1976.
4. Hooke R. *De Potentia Restitutiva*. London: John Martin; 1678.
5. Atanackovic TM, Guran A. Theory of elasticity for scientists and engineers. New York: Springer; 2000. p. 1–46.
6. Malvern LE. Introduction to the mechanics of a continuous medium. Englewood Cliffs: Prentice-Hall; 1969.
7. Boresi AP, et al. Advanced mechanics of materials. New York: Wiley; 2003.
8. Timoshenko SP. History of strength of materials. New York: Dover; 1983.
9. Aubin J-P. Applied functional analysis. New York: Wiley; 1979.
10. Fichera G. Existence theorems in elasticity. In: Flugge S, editor. *Handbuch de Physik*. Berlin: Springer; 1972. p. 347–89.
11. Fichera G. Existence theorem in elasticity. In: Flugge S, editor. *Handbuch de Physik*. Berlin: Springer; 1972. p. 347–89.
12. Bathe K-J. Finite element procedures. Upper Saddle River: Prentice-Hall; 1996.
13. Hughes TJR. The finite element method: linear static and dynamics finite element analysis. Englewood Cliffs: Prentice-Hall; 1987.
14. Zienkiewicz OC. The finite element method. New York: McGraw-Hill; 1977.
15. Hughes TJR. The finite element method. Englewood Cliffs: Prentice-Hall; 1987.
16. Bathe K-J. Finite element procedures in engineering analysis. Englewood Cliffs: Prentice-Hall; 1996.
17. Atkinson KE. An introduction to numerical analysis. 2nd ed. New York: Wiley; 1989.

Chapter 2

Nonlinear Finite Element Analysis Procedure

2.1 Introduction to Nonlinear Systems in Solid Mechanics

In order to explain nonlinear systems, it is necessary to define linear systems first. A linear system is defined such that the relationship between input and output is linear. Specifically, in structural mechanics, the relationship between applied loads (input) and displacements (output) is linear. When an applied load is doubled, the displacement will also be doubled. Thus, it is unnecessary to solve the linear system again when a different magnitude of load is applied. This property makes it possible to use the method of superposition. Mathematically, linearity can be explained using a linear operator. A general operator, A , is called linear when it satisfies $A(\alpha u + \beta w) = \alpha A(u) + \beta A(w)$ for any u and w in D_A and for any scalars α and β . Even if it is abstract, nonlinear systems are defined as everything else that is not linear. Therefore, it is important to understand characteristics of linear systems in order to understand that of nonlinear systems.

Figure 2.1 shows a linear relationship between input x and output y . In structural mechanics, input x represents applied loads or applied heat, while output y symbolizes displacements, stresses, or temperatures. For example, let x_1 and x_2 be transverse loads applied at two different locations of a beam, and let y be the reaction moment at the wall. Let the reaction moment at the wall be y_1 when only x_1 is applied, vice versa, y_2 is the reaction when only x_2 is applied. Then, when a combined load $2x_1 + 3x_2$ is applied to the beam, there is no need to solve the system again. Because of linearity, the reaction moment under the combined load becomes $2y_1 + 3y_2$, which is basically the principle of superposition. This is very useful, especially when the magnitude of load varies frequently.

In order to understand linear structural systems further, consider the diagram in Fig. 2.2, which illustrates the flow of physical quantities in structural systems. First, when loads are applied to the system, it generates local stresses in order to equilibrate against the globally applied loads. In an elastic system, stresses are generated by deforming its shape, during which strains are generated. Strains at

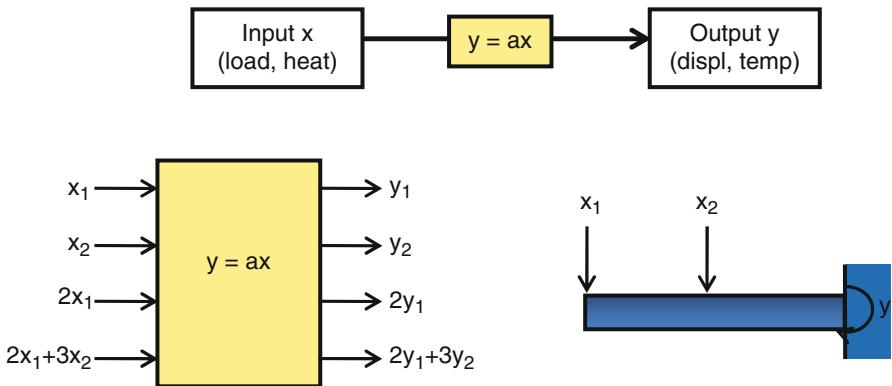


Fig. 2.1 Illustration of linear systems

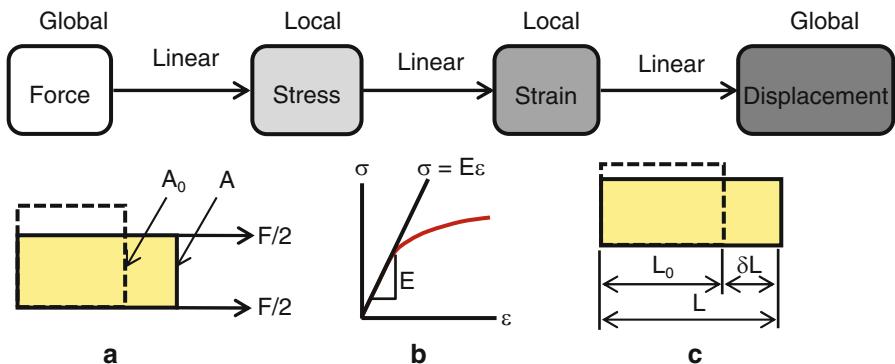


Fig. 2.2 Linearity in structural systems

every point are accumulated (or integrated) to yield displacements in the global level. In such a case, the structural system is called linear when all relationships among loads, stresses, strains, and displacements are linear. If any of them is not linear, then the structural system becomes nonlinear.

Then, let us consider a simple example of uniaxial tension of a bar in order to understand the linear relationship among the abovementioned physical quantities. First, when a load, F , is applied as shown in Fig. 2.2a, the bar elongates. In addition, because of Poisson's effect, the original cross-sectional area, A_0 , of the bar shrinks to A . Then, the stress generated by the load, F , can be calculated by dividing the load by the cross-sectional area, A , i.e., $\sigma = F/A$. However, the cross-sectional area depends on the load; as the load increases, the area decreases. Therefore, the relationship, $\sigma = F/A(F)$, is nonlinear between σ and F . However, if the load is small enough so that the difference between A_0 and A is ignorable, then it is possible to approximate the stress as $\sigma = F/A_0$. Based on this approximation, now the relationship between **load and stress becomes linear**. This approximation becomes

invalid if the elongation increases significantly such that the change in cross section cannot be ignorable.

Next, consider the relationship between stress and strain. For general metallic materials, such as aluminum or steel, Fig. 2.2b illustrates the stress-strain curve that can normally be obtained from uniaxial tension tests. Initially, the stress is increased linearly proportional to the strain. In this region, the stress and strain relationship is linear and reversible; that is, if the stress varies, the strain also varies along the straight line. The slope of this straight line corresponds to Young's modulus. Therefore the relation between stress and strain is linear, i.e., $\sigma = Ee$. When the stress reaches a threshold, called the yield strength, the relationship becomes nonlinear and its behavior is irreversible. Therefore, in order to be a linear relationship between stress and strain, the stress must be less than the yield strength.

Lastly, the relationship between strain and displacement must be linear. Consider the elongation of the bar, again, in Fig. 2.2c. The original length, L_0 , of the bar is increased by δL and ends up as the final length of L . In this case, δL is called the displacements or deformation. Then, the strain is defined as the change in length, i.e., $\epsilon = \delta L/L$. However, since the deformed length, L , already includes the displacement, δL , the relationship becomes nonlinear. Similar to the case of force and stress, if the displacement is small, then the definition of strain can be approximated by $\epsilon = \delta L/L_0$ so that the relationship between displacement and strain can be linear. This approximation is only valid when the displacement is small compared to the length of the bar.

As discussed above, many phenomena in physics show nonlinear behaviors, and linear systems are approximation of nonlinear systems under limited conditions. For example, the relation between the deflection of a beam and applied load at its tip is linear when the deflection is small. This includes small strain, small displacement, and small rotation in solid mechanics. However, as the deflection becomes large, the relation becomes nonlinear. In this sense, a linear system is an approximation of a nonlinear one. Many engineering applications can be solved by considering them as linear. For example, it is not expected to have large deflections in bridges or buildings. In such cases, linear analysis works well for estimating deflections and stresses. In fact, the same system can be solved using nonlinear analysis, but the result will not be much different.

In addition, solving linear systems has several advantages compared to solving nonlinear ones. First, linear systems are easier to solve. All the linear systems in the previous chapter can be solved using the system of linear equations in the form of $[\mathbf{K}]\{\mathbf{Q}\} = \{\mathbf{F}\}$. Nonlinear systems, on the other hand, cannot be solved in such a simple form. In fact, nonlinear systems are often solved using a sequence of linear analyses. Thus, the computational cost of a linear analysis is usually much less than that of a nonlinear analysis. Second, once the problem is well posed, the solution of a linear system always exists and it is unique. However, there is no guarantee that a nonlinear system has a unique solution. In addition, as described before, solutions from linear systems can be superimposed onto each other to produce a solution to other linear systems.

If a structural system is solved using the linearity assumption, the results may end up physically erroneous. For example, consider a cantilevered beam under a couple at the tip as shown in Fig. 2.3. The magnitude of the couple is big enough so that the beam undergoes a large deformation. In reality, it is not difficult to imagine that the beam will go through deformation as shown in Fig. 2.3a. In this case, the length of the neutral axis remains constant even if the beam goes through a large deformation. However, if a linear assumption is used, then the beam will deform as shown in Fig. 2.3b, which elongates the length of the beam, significantly. This happens because the assumption of linearity ignores the effect of bending moments on the rotation of the neutral axis. Because of that, in linear systems, the length of the beam is always measured in the undeformed geometry. In such a case, the assumption of linearity is obviously not valid, and linear analysis leads to erroneous results.

The assumption of linearity can also cause a difficulty that should not happen in practice. For example, consider two trusses that are connected through pin-joints as shown in Fig. 2.4. Since a truss is a two-force member, it can only support an axial force. When a vertical load, F , is applied at the center joint, the two trusses will rotate until they find equilibrium against the load. In that case, the vertical component of axial forces in the two trusses is in equilibrium with the vertical load. Due to the assumption of linearity, however, linear analysis uses the undeformed geometry as a reference, and these two-force members cannot support the vertical load. Therefore, linear analysis will fail to solve the system.

Although linear systems are easy to solve, many engineering applications cannot be modeled as a linear system. In solid mechanics, such a situation usually occurs when the deformation is large, material response is complex, boundary conditions

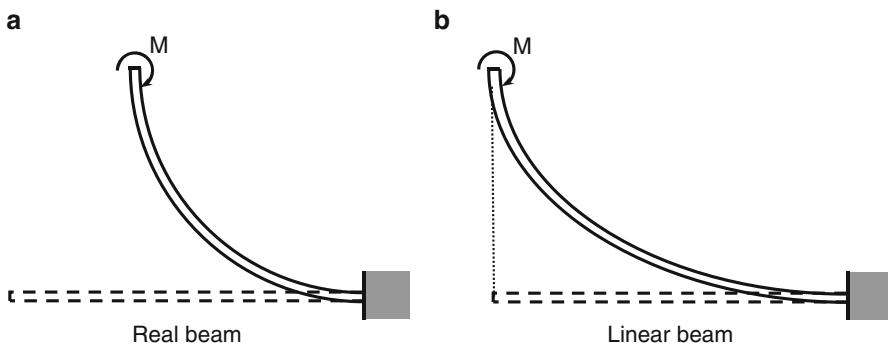


Fig. 2.3 Deformation of a beam under a couple

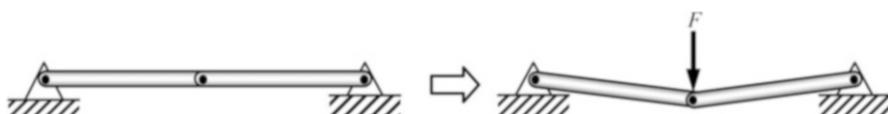


Fig. 2.4 Deformation of trusses under a vertical load

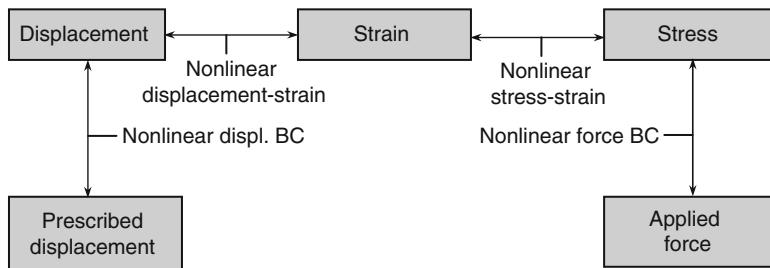


Fig. 2.5 Nonlinearities in solid mechanics

vary, etc. For example, a stamping process of sheet metals involves a large deformation of the blank, complex contact conditions between the blank and die, and permanent deformation of the material. It is clear that such a complex problem cannot be solved using linear analysis. Unfortunately, there is no easy criterion for when a problem can be modeled as linear or nonlinear. The choice of linear or nonlinear analysis often depends on the purpose of the analysis and the level of allowable errors. An important objective of this text is to address when an engineering problem should be modeled as a nonlinear system.

Although there are many different ways of categorizing different nonlinearities, it is generally accepted that four different sources of nonlinearity exist in solid mechanics. Figure 2.5 illustrates the occurrence of these nonlinearities in their relations among applied loads, stresses, strains, displacements, and boundary conditions.

2.1.1 Geometric Nonlinearity

Geometric nonlinearities, in general, represent the cases when the relations among kinematic quantities (i.e., displacement, rotation, and strains) are nonlinear. Such nonlinearities often occur when deformation is large. Figure 2.6a shows an example of geometric nonlinearity when a couple is applied at the tip of a cantilevered beam. Due to the large rotation, linear analysis cannot be used to accurately represent the deformation. It is clear that the relation between the applied couple and the tip displacement is nonlinear, as shown in Fig. 2.6b.

For the linear problems in the previous chapter, the relation between strain and displacement is linear. For example, in the case of a one-dimensional bar element, this relation can be written as

$$\varepsilon(x) = \frac{du(x)}{dx}. \quad (2.1)$$

If the displacement is doubled, the strain will also be doubled, which is a fundamental property of linear problems. Note that the above relation is valid only when displacement and its gradient are infinitesimal. As these quantities become large,

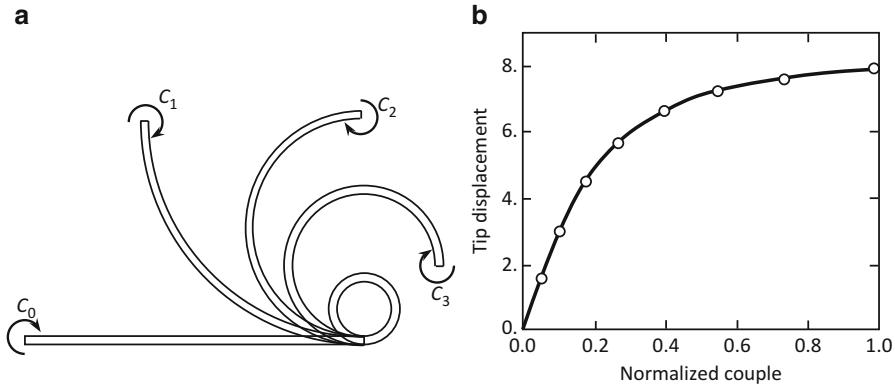


Fig. 2.6 Bending of a cantilever beam under large deformation

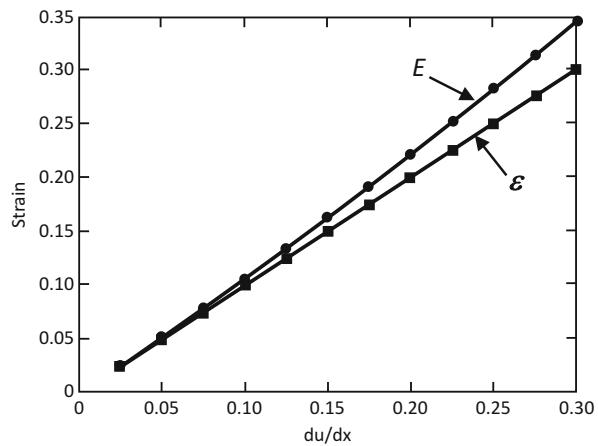


Fig. 2.7 Linear and nonlinear strains

the above relation is not accurate, and the following definition of strain needs to be introduced:

$$E(x) = \frac{du}{dx} + \frac{1}{2} \left(\frac{du}{dx} \right)^2. \quad (2.2)$$

Note that a higher-order term exists in the definition of strain. Due to this higher-order term, the relation between displacement and strain becomes nonlinear. It can be easily observed that when $(du/dx) \ll 1$, the two strains become identical, i.e., $\epsilon(x) \approx E(x)$. In fact, Eq. (2.1) is an approximation of Eq. (2.2) under the condition of infinitesimal deformation. As shown in Fig. 2.7, however, the difference between the two strains becomes larger as the magnitude of strain increases. At 5 % strain, for example, the error between the two strains is 2.5 %, while the error becomes 15 % at 30 % strain.

In general, structural equilibrium is written in the form of an integral equation. Although it is not clearly mentioned in the previous chapter, one of the most important assumptions of linear systems is that, due to infinitesimal deformation, the difference between the deformed and undeformed domains is ignored. Thus, the integration is performed over the undeformed domain in linear systems. In the precise sense, however, this equilibrium should be satisfied in the deformed domain—the structure is in equilibrium after deformation. Then, a dilemma occurs when the **deformed domain is used for integration**. The equilibrium equation, written in integral form over the deformed domain, solves for an unknown displacement, and this displacement determines the deformed domain. Such a dependency between the displacement and the deformed domain is an important criterion to identify geometric nonlinearities. In Chap. 3, detailed discussions of how to consider the effect of a deformed domain will be presented.

2.1.2 Material Nonlinearity

Material nonlinearity represents the case when the relation between stress and strain is not linear. This relation is often referred to as the constitutive relation. In linear systems, this relation is written as

$$\{\boldsymbol{\sigma}\} = [\mathbf{D}]\{\boldsymbol{\epsilon}\}, \quad (2.3)$$

where $[\mathbf{D}]$ is the elastic modulus matrix. Since $[\mathbf{D}]$ is constant, the relation between stress and strain is linear—if the strain is doubled, the stress will also be doubled. This relation represents a general behavior of elastic materials under an infinitesimal deformation. When the stress-strain relation cannot be represented by a constant matrix, $[\mathbf{D}]$, it is called a nonlinear material. In such a material, the elastic modulus matrix **depends on the current status of deformation**. In some cases, it also depends on the past history of deformation. In Chap. 3, **nonlinear elastic and hyperelastic materials** are discussed. These materials are fundamentally elastic—deformation disappears upon removing applied loads. Thus, there exists a strain energy density that depends on deformation. Figure 2.8a illustrates the linear and nonlinear elastic responses using a one-dimensional spring device. Assuming that both the length and cross-sectional area of the device are unitary, displacement becomes strain and applied load becomes stress. In the case of linear springs, the strain energy density is a quadratic form of strain, defined as¹

¹ Here the symbol E is used for the elastic modulus of a material, while $E(x)$ in Eq. (2.2) represents nonlinear strain.

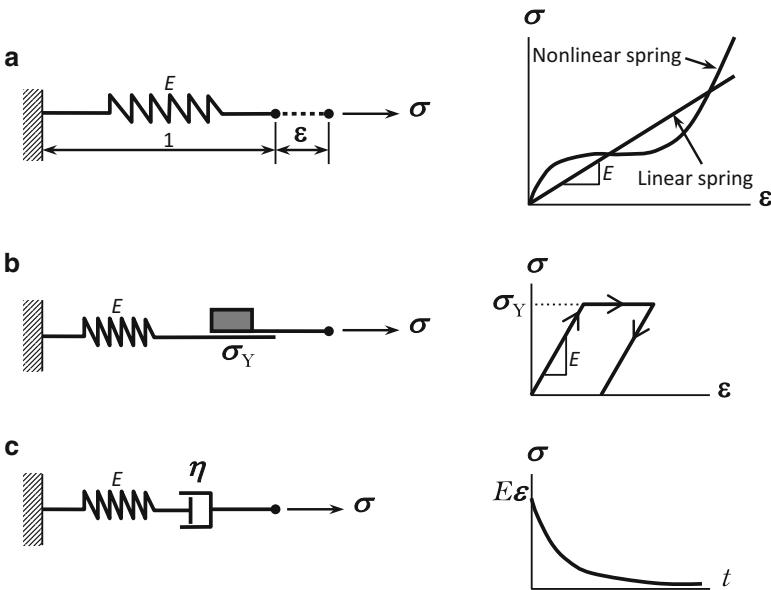


Fig. 2.8 Material nonlinearity models. (a) Linear and nonlinear elastic spring models. (b) Elasto-plastic spring model. (c) Visco-elastic spring model

$$U = \frac{1}{2}E\epsilon^2 \quad (2.4)$$

and the stress-strain relation can be defined by differentiating the strain energy density as

$$\sigma = \frac{dU}{d\epsilon} = E\epsilon. \quad (2.5)$$

In the case of nonlinear springs, the strain energy density is a more complex form than that of Eq. (2.4), and the stress-strain relation becomes nonlinear.

Another important type of material nonlinearity is the **plastic behavior of materials**. This is a common behavior of metal-type materials in which the material deforms elastically up to a certain limit. After that, the material shows permanent deformation, which remains upon removing applied loads. Figure 2.8b illustrates the plastic behavior of a material using a one-dimensional spring and a friction device. The friction device does not slip until the stress reaches a limit value, called the yield stress, σ_Y . When the stress is less than the yield strength, the displacement increases in the same way as a linear spring with slope, E . When the stress reaches the yield strength, the displacement increases without requiring a further increase in stress. The device cannot support stress higher than the yield strength.

Upon removal of the stress, the displacement reduces with the same slope, E . As can be found in Fig. 2.8b, a permanent deformation remains after completely removing the applied stress. As is clear in the above explanation, the stress-strain relation depends on the past history of deformation. In addition, it is difficult to express the stress-strain relation in a simple form. Rather, the relation is determined at an instance of the deformation path. Thus, in general, the stress-strain relation is given in the form of stress rate vs. strain rate. This material behavior is called elastoplasticity. In Chap. 4, detailed discussions on elastoplastic material will be presented.

Another popular material nonlinearity is viscoelasticity, described in Fig. 2.8c. Mathematically it can be modeled using a spring and a dash pot. This material shows a time-dependent behavior. For example, when strain ϵ is applied instantly at time $t = 0$ and remains constant, the stress responds as a linear spring initially and gradually decreases as a function of time. This behavior is common for human tissues, polymers, glasses, etc.

2.1.3 Kinematic Nonlinearity

Kinematic nonlinearity is also called boundary nonlinearity, as the displacement boundary conditions depend on the deformations of the structure. In general, structural equations solve for unknown displacements in the domain with given applied loads and prescribed displacement boundary conditions. When the boundary conditions change as a function of displacements, both the displacements and boundary conditions are unknown. In such a case, it is difficult to solve the structural equations as both sides of $[K]\{Q\} = \{F\}$ have unknown terms. In general, there are two possible cases for kinematic nonlinearity. The first one is when the location on the boundary where boundary conditions are applied is known, but the values are unknown. Diffusion in porous media is an example in which the amount of diffusion on the boundary is a part of the solution. The determination of the boundary conditions is a key part of the solution process. The second case is that both the location on the boundary where boundary conditions are applied and the values on the boundary conditions are unknown. The most common example is the contact constraint between two bodies. As two bodies are in contact, the displacements on the contact boundary are limited such that they cannot penetrate each other. At the same time, it is usually unknown which part of the boundary will be in contact. This kind of problem is more difficult than the first type. Figure 2.9 shows the deformation of a rubber cylinder through contact with a rigid wall. Initially, the contact occurs at a point. As the cylinder deforms, however, the size of the contact boundary increases. As expected, the relation between vertical displacement and applied force is also nonlinear. In Chap. 5, detailed discussions on contact problems are presented.

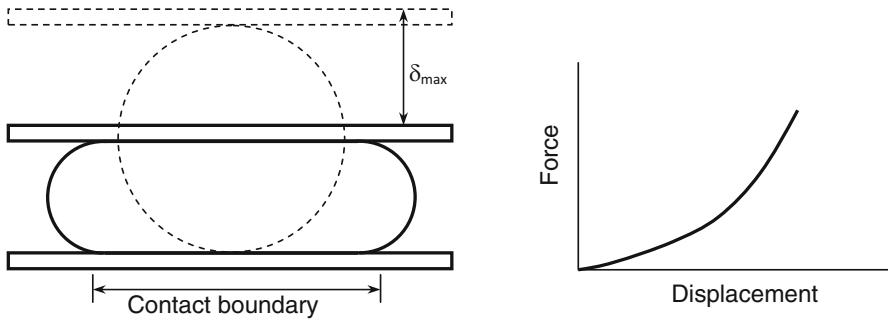


Fig. 2.9 Deformation of a rubber cylinder through a contact with rigid walls

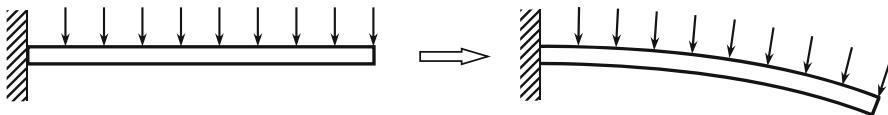


Fig. 2.10 Follow-up pressure load of a beam under large deformation

2.1.4 Force Nonlinearity

Similar to kinematic nonlinearity, force nonlinearity occurs when the **applied forces depend on deformation**. Since force is a vector, its magnitude and/or direction can change according to the deformation of a structure. Force nonlinearity is often accompanied by geometric nonlinearity. The most common example in solid mechanics is pressure loads of fluids. In the deployment of an airbag, for example, the direction and magnitude of pressure loads vary according to the deployment shape of the airbag. Although the contact condition is considered as boundary nonlinearity, the contact force can also be considered as force nonlinearity. As contact boundary varies, the contact force on the boundary also varies. Thus, in the contact problem, both the contact boundary and contact forces are unknown. Figure 2.10 shows a cantilevered beam under a pressure load that follows the deformation of the beam. In this case, the magnitude of the pressure load remains constant, but its direction changes according to the deflection of the beam. This type of load is called a follower load.

The most general case is when all four nonlinearities are present in a single problem. However, this may result in a very complex formulation, and the computational cost could be prohibitive. In practical problems, usually only one or two types of nonlinearities are considered at the same time. In the following chapters, different nonlinearities will be discussed.

Example 2.1 (Cup-drawing process) For the cup-drawing process as shown in Fig. 2.11, identify all nonlinearities.

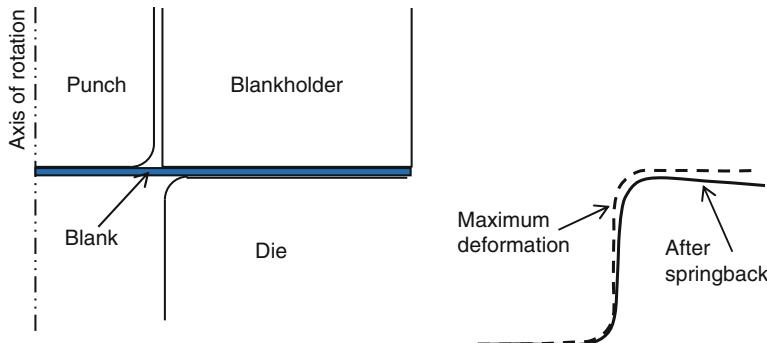


Fig. 2.11 Illustration of cup-drawing process

Solution The cup-drawing process in Fig. 2.11 starts from a circular blank made of a thin metallic plate, such as aluminum or steel. The blank is placed between a die and a blank-holder. The die is fixed and a vertical load is applied to the blank-holder. After that, a punch moves down to make a cup shape out of the blank. During this process, the blank will go through plastic deformation and have a permanent shape change. Once the punch moves down to the maximum depth, the punch and blank-holder are removed. At this point, the blank will recover some part of the plastic deformation through the process called springback. The process objective is to produce a specific shape after springback by controlling the applied load at the blank-holder and fillet radii of punch and die.

First, since the blank will go through a permanent deformation, material nonlinearity exists, similar to the elastoplastic material in Fig. 2.8b. In this process, it is important to control the maximum plastic strain so that it is less than the limit plastic strain in order to prevent tearing of material. Second, the geometry of the deformed blank will be significantly different from that of the initial one. Therefore, geometric nonlinearity exists in the process. In this particular process, the blank will experience not only a large strain but also a large rotation. Therefore, it is important to distinguish the difference between deformed and undeformed geometries. Lastly, kinematic nonlinearity exists between the blank and other parts, such as the punch, the die, and the blank-holder. Since the blank will gradually slide on the die and blank-holder, the contact region will gradually change. In addition, the contact between the punch and the blank is the main driver for the drawing process. Since three nonlinearities simultaneously exist in a single analysis, the cup-drawing process is particularly difficult to solve. ■

2.2 Solution Procedures for Nonlinear Algebraic Equations

Before discussing the finite element formulation for nonlinear problems, it is important to understand some of the solution procedures that are commonly employed to solve the system of nonlinear equations. The solution procedure may even influence the formulation of the problem.

Consider the following system of nonlinear equations:

$$\mathbf{P}(\mathbf{u}) = \mathbf{f}, \quad (2.6)$$

where $\mathbf{u} = \{u_1, u_2, \dots, u_n\}^T$ is a vector of unknowns, $\mathbf{f} = \{f_1, f_2, \dots, f_n\}^T$ is a vector of known quantities, and $\mathbf{P}(\mathbf{u}) = \{P_1(\mathbf{u}), P_2(\mathbf{u}), \dots, P_n(\mathbf{u})\}^T$ is a vector of nonlinear functions of \mathbf{u} . In structural applications, \mathbf{u} is often the displacement vector, \mathbf{f} is the applied force vector, and $\mathbf{P}(\mathbf{u})$ is the internal force vector. Thus, Eq. (2.6) is the **equilibrium between internal and applied forces**. In the linear problems in Chap. 1, the internal force vector is a linear function of \mathbf{u} such that $\mathbf{P}(\mathbf{u}) = \mathbf{K} \cdot \mathbf{u}$ with \mathbf{K} being a constant stiffness matrix. Then, solving a system of linear equations is equivalent to calculating the inverse matrix of \mathbf{K} and multiplying it with the vector, \mathbf{f} . In practice, instead of calculating the inverse matrix, different matrix solution techniques are used, such as LU-decomposition [1].

Since $\mathbf{P}(\mathbf{u})$ is a nonlinear function of \mathbf{u} , nonlinear analysis focuses on how to solve Eq. (2.6) accurately and effectively. The solution methods applicable to general nonlinear functions are all iterative. Starting from an initial estimate, \mathbf{u}^0 , the increment, $\Delta\mathbf{u}$, of the solution is obtained by solving a system of linear equations. Linearization is involved in this process. After obtaining the increment, the solution is iteratively updated until a specified convergence criterion is satisfied. Different methods are available according to the way to calculate the increment, $\Delta\mathbf{u}$; several of these will be discussed in the following subsections.

Example 2.2 (System of nonlinear springs) Consider two serially connected nonlinear springs, as shown in Fig. 2.12. The stiffness of both springs depends on the elongation of springs such that $k_1 = 50 + 500u$ [N/m] and $k_2 = 100 + 200u$ [N/m] with u being the elongation of the spring. The equation for a spring element is

$$\begin{bmatrix} k & -k \\ -k & k \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} f_1 \\ f_2 \end{Bmatrix},$$

where u_1 and u_2 are nodal displacements at the two nodes. When a force of $F = 100$ N is applied at the tip, construct the system of nonlinear equations in the form of Eq. (2.6) using the two spring elements.

Solution Since spring 1 is fixed on the wall, its elongation is equivalent to u_1 , while for spring 2, the elongation is $u_2 - u_1$. In the normal assembly process, the wall is considered as an additional DOF and it is deleted when the displacement boundary condition is applied. However, it is also possible that the fixed DOF is deleted

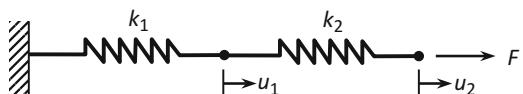


Fig. 2.12 Two nonlinear springs

before assembly and only free DOFs are used for assembly. Using the given stiffness of springs, the assembled matrix equation becomes

$$\begin{bmatrix} 50 + 500u_1 + 100 + 200(u_2 - u_1) & -100 - 200(u_2 - u_1) \\ -100 - 200(u_2 - u_1) & 100 + 200(u_2 - u_1) \end{bmatrix} \begin{Bmatrix} u_1 \\ u_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ F \end{Bmatrix}.$$

Note that the above equations are not linear as the stiffness matrix contains unknown variables. After multiplying the stiffness matrix and the vector of unknowns, the following system of nonlinear equations is obtained:

$$\begin{cases} 300u_1^2 + 400u_1u_2 - 200u_2^2 + 150u_1 - 100u_2 = 0 \\ 200u_1^2 - 400u_1u_2 + 200u_2^2 - 100u_1 + 100u_2 = 100 \end{cases}. \quad (2.7)$$

Figure 2.13 shows the two nonlinear functions along with contour lines whose values correspond to zero. The constant value on the right-hand side of the second equation is moved to the left-hand side in the figure. Then, the solution of the system of nonlinear equations becomes the intersection point of these contour lines. Since multiple contour lines exist, it is possible that multiple solutions may exist. In addition, it is also possible that no solution exists in a certain situation. In general, there is no analytical way of finding the solution of a system of nonlinear equations unless the equations are very simple. In the following subsections, several methods of solving the system of nonlinear equations are discussed. ■

2.2.1 Newton–Raphson Method

This method is popular in numerical analysis to find the roots of nonlinear equations. Basically, most numerical methods for solving a system of nonlinear equations assume an initial estimate, \mathbf{u}^0 , and find its increment, $\Delta\mathbf{u}$, so that the new estimate, $\mathbf{u}^0 + \Delta\mathbf{u}$, is close to the solution to Eq. (2.6). In order to find the

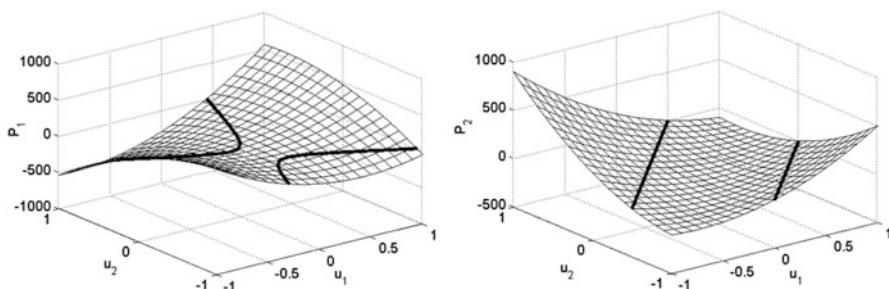


Fig. 2.13 Surface plots of the system of nonlinear equations with zero-level contour

increment, the nonlinear equations are locally approximated by linear ones. This process is repeated until the original nonlinear equations are satisfied. Suppose an approximate solution at the i th iteration is known and is designated by \mathbf{u}^i . The solution at the next iteration can be approximated using the first-order Taylor series as follows:

$$\mathbf{P}(\mathbf{u}^{i+1}) \approx \mathbf{P}(\mathbf{u}^i) + \mathbf{K}_T^i(\mathbf{u}^i) \cdot \Delta\mathbf{u}^i = \mathbf{f}, \quad (2.8)$$

where $\mathbf{K}_T^i(\mathbf{u}^i) \equiv (\partial\mathbf{P}/\partial\mathbf{u})^i$ is the Jacobian matrix at the i th iteration, commonly known as the **tangent stiffness matrix** in structural applications and $\Delta\mathbf{u}^i$ is the solution increment. The goal is to calculate $\Delta\mathbf{u}^i$ and iteratively update the solution, \mathbf{u}^{i+1} . After rearranging the terms, the system of linearized equations can be obtained as

$$\mathbf{K}_T^i \Delta\mathbf{u}^i = \mathbf{f} - \mathbf{P}(\mathbf{u}^i). \quad (2.9)$$

Equation (2.9) is similar to the matrix equation of linear systems, except that (1) the coefficient matrix, $\mathbf{K}_T^i(\mathbf{u}^i)$, is not constant, but a function of \mathbf{u}^i ; (2) the equation solves for the increment, $\Delta\mathbf{u}^i$, not the total solution, \mathbf{u} ; and (3) the right-hand side is not the applied force, but rather the difference between the applied force and internal force. This difference is often referred to as a residual. After solving for the displacement increment, $\Delta\mathbf{u}^i$, a new approximate solution is obtained as follows:

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \Delta\mathbf{u}^i. \quad (2.10)$$

In general, this solution will not satisfy the system of nonlinear equations exactly and there will be some residual or unbalance force defined as follows:

$$\mathbf{R}^{i+1} = \mathbf{f} - \mathbf{P}(\mathbf{u}^{i+1}). \quad (2.11)$$

If the unbalance force is smaller than a given tolerance, the solution, \mathbf{u}^{i+1} , can be accepted as the accurate solution, and the process stops. Otherwise, the process is repeated until this residual becomes very small. The termination criterion is expressed in the **normalized form** as follows:

$$\text{conv} = \frac{\sum_{j=1}^n (R_j^{i+1})^2}{1 + \sum_{j=1}^n (f_j)^2}. \quad (2.12)$$

A constant, 1, is added to the denominator to avoid division by zero when there are no applied loads. The iterations are terminated when the convergence parameter, conv, becomes less than a given tolerance (say 0.01). Sometimes, different criteria can be applied to determine the convergence of the iterative procedure. One is

based on the solution increment. When the increment of the solution is much smaller than the initial increment, then it is assumed that the solution is converged. The solution-based termination criterion becomes

$$\text{conv} = \frac{\sum_{j=1}^n (\Delta u_j^{i+1})^2}{1 + \sum_{j=1}^n (\Delta u_j^0)^2}. \quad (2.13)$$

However, this criterion can be plausible when the convergence is merely slow. Since stresses or forces are derivatives from displacements in structural problems, it is easy to make the displacement converge rather than the force. In practice, many commercial programs monitor both criteria to determine whether the solution has converged or not. More detailed convergence criteria will be discussed later. Instead of the sum of squares in the above two convergence criteria, it is also possible to use the maximum absolute value.

The algorithm of Newton–Raphson method is as follows:

1. Set tolerance = 0.001, $k = 0$, max_iter = 20, and initial estimate $\mathbf{u} = \mathbf{u}^0$
2. Calculate residual $\mathbf{R} = \mathbf{f} - \mathbf{P}(\mathbf{u})$
3. Calculate conv in Eq. (2.12). If $\text{conv} \leq \text{tolerance}$, stop
4. If $k > \text{max_iter}$, stop with error message
5. Calculate Jacobian matrix \mathbf{K}_T in Eq. (2.8)
6. If the determinant of \mathbf{K}_T is zero, stop with error message
7. Calculate solution increment $\Delta\mathbf{u}$ by solving Eq. (2.9)
8. Update solution by $\mathbf{u} = \mathbf{u} + \Delta\mathbf{u}$
9. Set $k = k + 1$
10. Go to Step 2

Two iterations of the procedure for a system with a single DOF are illustrated graphically in Fig. 2.14. In the case of a single DOF, the Jacobian matrix becomes

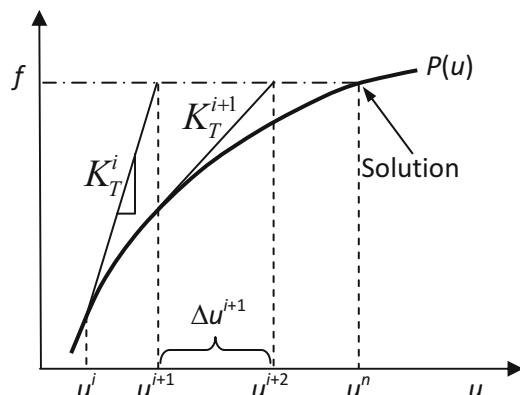


Fig. 2.14 Newton–Raphson method for nonlinear equation $P(u) = f$

the slope of the nonlinear function, $P(u)$. The solution converges rapidly when the starting point is close to the solution. When the current iteration is close to the solution, this method shows a quadratic convergence. Let u_{exact} be the exact solution, and u_n and u_{n+1} be the two consecutive approximations of the solution from the Newton–Raphson method. Then, the method converges quadratically when there exists a constant $c > 0$ such that

$$|u_{\text{exact}} - u_{n+1}| \leq c |u_{\text{exact}} - u_n|^2. \quad (2.14)$$

Since the left-hand side is the error at the $(n+1)$ th iteration and the right-hand side is the square of the error at the n th iteration, errors in the Newton–Raphson method reduce very quickly. In practice, since the exact solution is usually unknown, the solution at the converged iteration is often considered as u_{exact} . In order to show that a numerical algorithm has a quadratic convergence, it is required to show that the following ratio approaches a constant value c :

$$\lim_{n \rightarrow \infty} \frac{|u_{\text{exact}} - u_{n+1}|}{|u_{\text{exact}} - u_n|^2} = c. \quad (2.15)$$

In practice, it is often enough to show that the convergence criterion in Eq. (2.12) reduces quadratically at each iteration.

Example 2.3 (Roots of a system of nonlinear equations) Find the two nodal displacements of the nonlinear springs in Example 2.2 using the Newton–Raphson method. Use the convergence tolerance of 1×10^{-5} , and the initial estimate, $\mathbf{u}^0 = \{0, 0\}^T$. Also, estimate the convergence rate.

Solution In order to solve the system of nonlinear equations in Eq. (2.7), it is necessary to calculate the Jacobian matrix first. By differentiating Eq. (2.7), the Jacobian matrix can be written as

$$\mathbf{K}_T = \begin{bmatrix} \partial P_1 / \partial u_1 & \partial P_1 / \partial u_2 \\ \partial P_2 / \partial u_1 & \partial P_2 / \partial u_2 \end{bmatrix} = \begin{bmatrix} 600u_1 + 400u_2 + 150 & 400(u_1 - u_2) - 100 \\ 400(u_1 - u_2) - 100 & 400(u_2 - u_1) + 100 \end{bmatrix}.$$

It can be shown that the Jacobian matrix is positive definite when the two nodal displacements satisfy the following relation:

$$u_2 - u_1 + 0.25 \geq 0.$$

Physically, if a positive force is applied, u_2 will always be larger than u_1 , and the system should be stable. Below is the MATLAB program that solves the nonlinear spring problem.

```

%
% Example 2.3 Two nonlinear springs (Newton-Raphson method)
%
tol = 1.0e-5; iter = 0; c = 0;
u = [0; 0];
uold = u;
f = [0; 100];
P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2)
      200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];
R = f - P;
conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);
fprintf('\n iter   u1   u2   conv   c');
fprintf('\n %3d %7.5f %7.5f %12.3e %7.5f',iter,u(1),u(2),conv,c);
while conv > tol && iter < 20
Kt = [600*u(1)+400*u(2)+150 400*(u(1)-u(2))-100
      400*(u(1)-u(2))-100 400*u(2)-400*u(1)+100];
delu = Kt\R;
u = uold + delu;
P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2)
      200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];
R = f - P;
conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);
c = abs(0.9-u(2))/abs(0.9-uold(2))^2;
uold = u;
iter = iter + 1;
fprintf('\n %3d %7.5f %7.5f %12.3e %7.5f',iter,u(1),u(2),conv,c);
end

```

Table 2.1 shows the convergence iteration history from the Newton–Raphson method. Note that the algorithm converges at the sixth iteration at which the convergence criterion in Eq. (2.12) becomes smaller than the tolerance. Since the initial slope of the Jacobian matrix is small, the initially predicted displacements are much larger than the actual displacements. As the MATLAB program iterates, the displacements converge to the accurate values, which is $\mathbf{u}_{\text{exact}} = \{0.4, 0.9\}^T$. The last column of Table 2.1 shows the constant, c , in Eq. (2.15), which converges to the value of 1.1. Thus, the algorithm has a quadratic convergence rate. Note that the residual reduction is also approximately quadratic. Figure 2.15 shows the force–displacement curves of the nonlinear springs. The stiffness of the springs gradually increases as the displacements increase. ■

Table 2.1 Convergence history of two nonlinear springs using the Newton–Raphson method

Iteration	u_1	u_2	conv	c
0	0.0000	0.0000	9.999E-01	—
1	2.0000	3.0000	3.280E+02	—
2	1.0244	1.6244	1.981E+01	0.164
3	0.5814	1.0873	9.282E-01	0.357
4	0.4261	0.9261	1.455E-02	0.744
5	0.4007	0.9007	1.033E-05	1.048
6	0.4000	0.9000	6.462E-12	1.109

Fig. 2.15 Force–displacement curves for two nonlinear springs

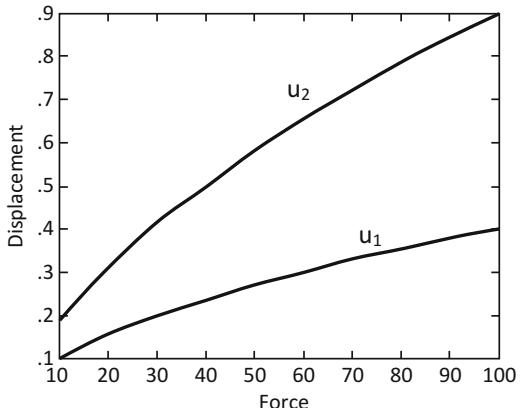
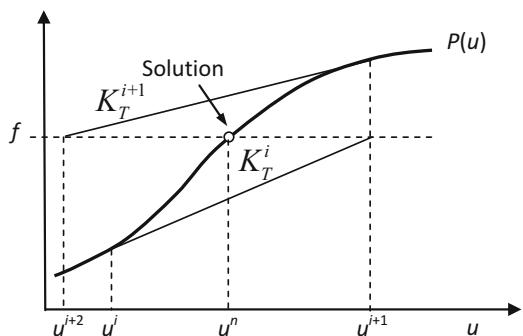


Fig. 2.16 Convergence difficulty in Newton–Raphson iteration



The Newton–Raphson method does not always guarantee convergence to the accurate solution. First, it assumes that the solution increment in Eq. (2.9) is relatively small. As the number of iterations increases, $\Delta\mathbf{u}$ becomes smaller and eventually approaches zero at the accurate solution. However, this assumption is violated when the Jacobian matrix becomes singular, or the determinant of matrix \mathbf{K}_T is zero. In such a case, $\Delta\mathbf{u}$ becomes infinite and the solution diverges (see Step 6 in the algorithm). This means, in a single DOF system, that the slope of $P(u)$ becomes zero and the residual cannot be reduced. Numerically, similar behavior can be observed when the matrix is nearly singular.

Second, as shown in Fig. 2.16, the method may diverge or oscillate between two points if the starting point is too far away from the exact solution. This also happens when the curvature of the $P(u)$ curve changes its sign between two consecutive iterations. In such a case, it is possible that the Newton–Raphson algorithm may result in an infinite loop. In order to prevent an infinite loop, the maximum number of iterations is set and the algorithm stops with an error message when the number of iterations reaches the maximum number of iterations (see Step 4 in the algorithm).

Example 2.4 (Divergence of the Newton–Raphson method) Find a root of the following nonlinear equation using the Newton–Raphson method:

$$P(u) = u + \tan^{-1}(5u). \quad (2.16)$$

Use the convergence tolerance of 1×10^{-5} and the initial estimate $u^0 = 0.5$.

Solution Since the problem has a single variable, the problem becomes a nonlinear algebraic equation. It is trivial that the exact solution will be $u = 0$. The derivative of $P(u)$ with respect to u becomes

$$\frac{dP}{du} = 1 + 5 \cos^2(\tan^{-1}(5u)).$$

Below is the list of MATLAB program that solves the above nonlinear equation for up to 20 iterations.

```
%  
% Example 2.4 Divergence of the Newton-Raphson method.  
%  
xdata=zeros(40,1);  
ydata=zeros(40,1);  
tol = 1.0e-5;  
iter = 0;  
u = 0.5;  
uold = u;  
c=0;  
P = u+atan(5*u);  
R = -P;  
conv= R^2;  
xdata(1)=u;  
ydata(1)=P;  
while conv > tol && iter < 20  
    Kt = 1+5*(cos(atan(5*u)))^2;  
    delu = R/Kt;  
    u = uold + delu;  
    P = u+atan(5*u);  
    R = -P;  
    conv= R^2;  
    uold = u;  
    iter = iter + 1;  
    xdata(2*iter)=u;    ydata(2*iter)=0;  
    xdata(2*iter+1)=u; ydata(2*iter+1)=P;  
end  
%  
plot(xdata,ydata);  
hold on;  
x=[-1:0.1:1];  
y=x+atan(5*x);  
plot(x,y)
```

Fig. 2.17 Divergence of the Newton–Raphson method

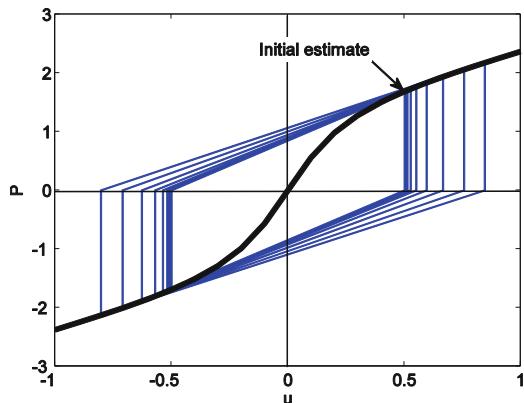


Figure 2.17 shows the convergence history of the Newton–Raphson method for up to 20 iterations. In this case, the approximate solution moves away from the exact one as the number of iterations increases, and eventually, the method diverges. The initial estimate of $u^0 = \pm 0.4990807536$ will make the algorithm oscillate between the two points. The method will converge quickly if the initial estimate is less than that. ■

When the system of nonlinear equations has multiple solutions, this method may converge to different solutions depending on the initial estimate. This does not occur often in structural problems because, in most cases, the starting point is $\mathbf{u} = 0$.

The Jacobian matrix of structural mechanics problems is normally positive definite. Physically, this means that in order to increase the displacement, the applied force should increase. However, in some cases, the displacement may increase without the applied force increasing. Or, displacement continuously increases while the applied force decreases. This causes structural instability. Common examples are bifurcation and snap-through behaviors. Figure 2.18 shows the snap-through behavior of elastic inclined slender beams. Both ends are clamped. The plot shows the relation between the applied force and the vertical displacement at the force application point. Initially, the applied force increases along with the tip displacement (region AC). In this region, the beams are stable, and the Jacobian matrix is positive definite. When they reach point C, the Jacobian matrix becomes singular, and the force cannot increase beyond F_C . Between points C and E, the tip displacement continuously increases while the force decreases. The system is unstable in this region. Beyond E, the beams become stable again with a positive definite Jacobian matrix. When the Newton–Raphson method is used, it can only solve for the response in the region, AC. For example, for a given force, F_B , it always yields point B, not point D. Special techniques are required to follow the force–displacement curve.

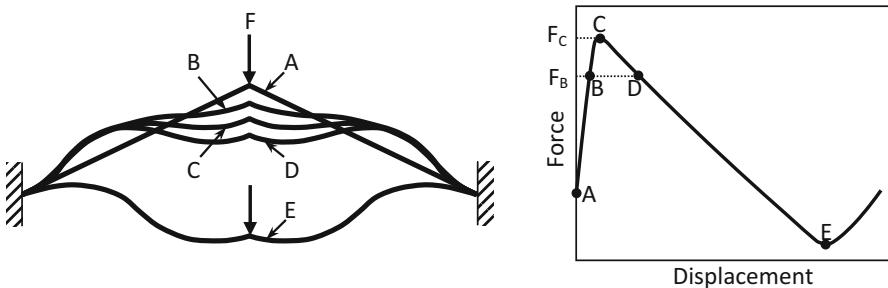


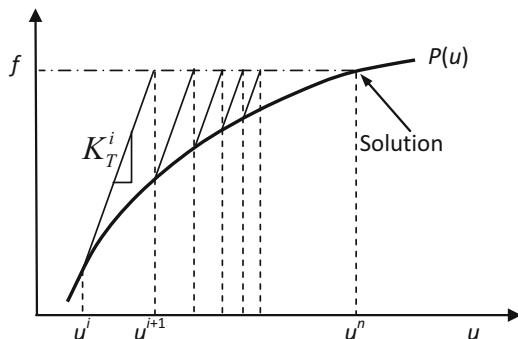
Fig. 2.18 Snap-through behavior of inclined slender beams

2.2.2 Modified Newton–Raphson Method

The Newton–Raphson method requires that at each iteration, the Jacobian matrix should be formed and the system of linearized equations should be solved for the increment of the solution. Computationally, these are expensive tasks. In the finite element framework, building the tangent stiffness matrix and solving the matrix equation are the two most computationally intensive procedures. The modified Newton–Raphson method is an attempt to make these procedures less expensive. Instead of formulating a new tangent stiffness matrix at each iteration, the initial tangent stiffness matrix is repeatedly used for all iterations. This obviously avoids the need to reformulate the tangent stiffness matrix at each iteration. In addition, this can also reduce the computational time required for solving the matrix equation. In solving a matrix equation, the matrix is first decomposed into lower- and upper-triangular forms (LU-decomposition). After that, the vector on the right-hand side is used to solve for the solution (forward and backward substitutions). The LU-decomposition procedure is computationally expensive, while the forward and backward substitutions are relatively inexpensive. For example, if the dimension of the matrix is $N \times N$, the computational cost for the LU-decomposition procedure is proportional to N^2 , while the forward and backward substitutions are proportional to N . When the modified Newton–Raphson method is used, the LU-decomposed matrix is kept and only the forward and backward substitutions are used with different residuals at each iteration. As illustrated in Fig. 2.19, the method usually requires a greater number of iterations for convergence than that of the regular Newton–Raphson method. However, the overall computational cost to obtain the solution can be made less because each iteration is much faster than that of the regular Newton–Raphson method. The method is also a little more stable and is not prone to divergence.

To improve convergence, it is possible to develop a hybrid scheme in which a few iterations are performed with the initial tangent stiffness matrix, after which a new tangent stiffness is formed. The only drawback of this scheme is that it is difficult to decide how many constant tangent stiffness iterations to perform before reformulating a new tangent stiffness matrix.

Fig. 2.19 Modified Newton–Raphson method



Example 2.5 (Nonlinear springs [modified Newton–Raphson method]) Using the modified Newton–Raphson method, solve the displacements of the two nonlinear springs in Example 2.2. Use the initial estimate, $\mathbf{u}^0 = \{0.3 \ 0.6\}^T$. Compare the number of iterations with that of Example 2.3. Also, check the convergence rate.

Solution Below is the MATLAB program for solving the problem. Note that the tangent stiffness matrix is calculated only once before the convergence loop.

```

tol = 1.0e-5;
iter = 0;
u = [0.3; 0.6];
uold = u;
c = 0;
f = [0; 100];
P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2)
      200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];
R = f - P;
conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);
fprintf('\n niter   u1     u2     conv    c');
fprintf('\n %3d %7.5f %7.5f %12.3e %7.5f', iter, u(1), u(2), conv, c);
Kt = [600*u(1)+400*u(2)+150 -400*u(2)+400*u(1)-100
      400*u(1)-400*u(2)-100 400*u(2)-400*u(1)+100];
while conv > tol && iter < 20
    delu = Kt\R;
    u = uold + delu;
    P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2)
          200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];
    R = f - P;
    conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);
    c = abs(0.9-u(2))/abs(0.9-uold(2))^2;
    uold = u;
    iter = iter + 1;
    fprintf('\n %3d %7.5f %7.5f %12.3e %7.5f', iter, u(1), u(2), conv, c);
end

```

Table 2.2 Convergence history of two nonlinear springs using the modified Newton–Raphson method

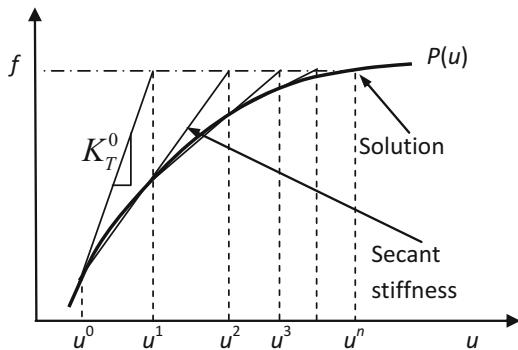
Iteration	u_1	u_2	conv	c
0	0.3000	0.6000	2.848E-01	—
1	0.4143	0.9507	1.464E-02	—
2	0.3956	0.8812	2.378E-03	0.563
3	0.4012	0.9063	3.260E-04	7.328
4	0.3997	0.8978	4.711E-05	17.77
5	0.4001	0.9008	6.561E-06	158.03

Since the tangent stiffness matrix changes significantly as the displacements vary, the algorithm will diverge with the initial estimate of $\mathbf{u}^0 = \{0, 0\}^T$. However, that does not mean that the modified Newton–Raphson method is less stable than the regular Newton–Raphson method. Table 2.2 shows the convergence history of the modified Newton–Raphson algorithm. Although it converges in the sixth iteration, this happens because the initial estimate is close to the exact solution. In fact, the convergence criterion, conv, reduces slower than that of the regular Newton–Raphson method. The last column shows the constant in Eq. (2.15). It is clear that the method does not provide a convergent constant, which indicates that the algorithm does not have a quadratic convergence. ■

2.2.3 Incremental Secant Method

In the regular Newton–Raphson method, the tangent stiffness matrix is calculated at every iteration, while the modified Newton–Raphson method requires calculating it once or after a certain number of iterations. As discussed before, constructing this matrix and solving the matrix equation are the two main sources of computational cost. Although the modified Newton–Raphson method is computationally efficient, it can cause problems as it uses the fixed tangent stiffness matrix (e.g., refer to Example 2.5 with initial estimate, $\mathbf{u}^0 = \{0, 0\}^T$). The main purpose of the incremental secant method is to remove these two tasks so that the computational cost can be reduced, while achieving a certain level of convergence rate that is greater than one. The role of the tangent stiffness matrix is to make the equation converge quickly, while that of the residual is to monitor the accuracy. The algorithms iterate until the residual vanishes, which means that the system of nonlinear equations is satisfied within the range of the tolerance. If the tangent stiffness matrix is not accurate, then the algorithm converges slower. As long as the solution increments are in the right direction, the algorithm will eventually converge to the right solution after performing more iterations. The idea of the incremental secant method is to approximate the tangent stiffness matrix without high computational costs. This is achieved by progressively updating the tangent stiffness matrix using the secant direction between two consecutive solutions. An important aspect, while approximating the tangent stiffness matrix, is that it maintains the positive definite

Fig. 2.20 Incremental secant method



property. In addition, computational cost can be further reduced if the inverse of the tangent stiffness matrix is approximated directly.

The main idea of the incremental secant method can be explained clearly using a single variable example. In Eq. (2.9), the Jacobian matrix is defined as the derivative of the nonlinear function, $P(u)$, with respect to the unknown variable, u . The secant matrix can be obtained by using the finite difference method in approximating the Jacobian matrix as

$$K_s^i = \frac{P(u^i) - P(u^{i-1})}{u^i - u^{i-1}}. \quad (2.17)$$

Note that as u^{i-1} approaches u^i , the secant stiffness approaches the tangent stiffness of the Newton–Raphson method. For the first iteration, the secant method uses the same tangent stiffness matrix with the Newton–Raphson method. After the first iteration, secant stiffness is used in the subsequent iterations instead of the tangent stiffness. The procedure is illustrated in Fig. 2.20. In the case of a single variable problem, the secant direction is the one that connects the two consecutive solutions. The solution increment for the i th iteration is expressed as follows:

$$\Delta u^i = \frac{u^i - u^{i-1}}{P(u^i) - P(u^{i-1})} (f - P(u^i)). \quad (2.18)$$

The convergence rate of the secant method is 1.618, which is the golden ratio. Considering that the Newton–Raphson method has a quadratic convergence, it is faster than the secant method. However, the Newton–Raphson method requires the evaluation of both $P(u)$ and its derivative, $K_T(u)$, at every iteration, while the secant method only requires the evaluation of $P(u)$. Thus, each iteration of the secant method is much faster than that of the Newton–Raphson method. Those methods that approximate the Jacobian matrix are called quasi-Newton methods. They are less expensive than the Newton–Raphson method as they do not require calculating the Jacobian, but they converge slower than the Newton–Raphson method.

Example 2.6 (Nonlinear algebraic equation [secant method]) Using the secant method, find a root of the nonlinear equation in Example 2.4. Use the initial estimate, $u^0 = 2.0$, and convergence tolerance of 1×10^{-5} . Check the convergence rate.

Solution Below is the MATLAB program for solving the nonlinear equation. Note that the exact Jacobian is used at the first iteration and the secant Jacobian for the following iterations:

```
%  
% Example 2.6 Nonlinear algebraic equation (secant method)  
%  
tol = 1.0e-5; iter = 0; c = 0;  
u = 2.0; uold = u;  
P = u+atan(5*u); Pold = P;  
R = -P; conv = R^2;  
fprintf('\niter u conv c');  
fprintf('\n %3d %7.5f %12.3e %7.5f', iter, u, conv, c);  
Ks = 1+5*(cos(atan(5*u)))^2;  
while conv > tol && iter < 20  
    delu = R/Ks;  
    u = uold + delu;  
    P = u+atan(5*u);  
    R = -P;  
    conv = R^2;  
    c = abs(u)/abs(uold)^2;  
    Ks = (P - Pold)/(u - uold);  
    uold = u;  
    Pold = P;  
    iter = iter + 1;  
    fprintf('\n %3d %7.5f %12.3e %7.5f', iter, u, conv, c);  
end
```

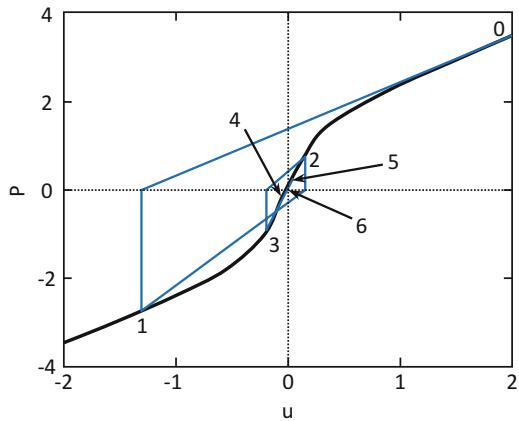
Using the Newton–Raphson method, the nonlinear algebraic equation in Eq. (2.16) diverges when the initial estimate is larger than 0.5. However, the secant method is able to converge even if the initial estimate is 2.0. When the initial estimate of 0.5 is used, the method converges in two iterations. Table 2.3 shows the convergence history of the secant method. It is clear that the secant method is more stable than the Newton–Raphson method and does not diverge because the secant stiffness at the first iteration is adjusted toward the exact solution. The method does not show a quadratic convergence because the ratio in Eq. (2.15) does not approach a constant value (refer to the last column of Table 2.3). However, the convergence criterion reduces faster than that of the modified Newton–Raphson method. Figure 2.21 shows the convergence history of the secant method. ■

Although it is clear how the secant matrix is constructed in the one variable case, multivariable cases are less straightforward. One of the first methods in solving for multivariable nonlinear equations is the one proposed by Broyden [2]. The idea is that the Jacobian matrix is calculated only at the first iteration, and after that, it is

Table 2.3 Convergence history of nonlinear algebraic equation using the secant method

Iteration	u_1	conv	c
0	2.0000	1.205E+01	—
1	-1.3074	7.433E+00	0.327
2	0.1476	6.136E-01	0.086
3	-0.1771	8.136E-01	8.133
4	-0.0033	4.025E-04	0.107
5	0.0006	1.338E-05	54.511
6	0.0000	5.393E-14	0.104

Fig. 2.21 Convergence iteration of the secant method



updated at every iteration using the rank-one update. The solution increment for the i th iteration is expressed as follows:

$$\mathbf{K}_s^i \Delta \mathbf{u}^i = -\mathbf{R}^i, \quad (2.19)$$

where \mathbf{K}_s^i is the secant stiffness matrix and $\mathbf{R}^i = \mathbf{P}(\mathbf{u}^i) - \mathbf{f}$ is the vector of the residual at the i th iteration. Note that the sign in the definition of the residual is intentionally changed from that in Eq. (2.11) in order to make the following algorithm simpler. Using the solutions at two consecutive iterations, \mathbf{u}^{i-1} and \mathbf{u}^i , the secant stiffness matrix is updated. The updated matrix then needs to satisfy the following secant equation:

$$\mathbf{K}_s^i \cdot (\mathbf{u}^i - \mathbf{u}^{i-1}) = \mathbf{R}(\mathbf{u}^i) - \mathbf{R}(\mathbf{u}^{i-1}). \quad (2.20)$$

Thus, the objective of the secant method is to update the secant stiffness matrix with the known increments in the solution and the known terms on the right-hand side. Unfortunately, this process is not unique and many different matrices satisfy the relationship. Broyden initially suggested updating the stiffness matrix by taking the

solution to the secant equation that modifies the matrix minimally. The Broyden's method results in the following form of a rank-one update:

$$\mathbf{K}_s^i = \mathbf{K}_s^{i-1} + \frac{\Delta \mathbf{R} - \mathbf{K}_s^{i-1} \Delta \mathbf{u}}{\|\Delta \mathbf{u}\|^2} \Delta \mathbf{u}^T, \quad (2.21)$$

where $\Delta \mathbf{R} = \mathbf{R}(\mathbf{u}^i) - \mathbf{R}(\mathbf{u}^{i-1})$ and $\Delta \mathbf{u} = \mathbf{u}^i - \mathbf{u}^{i-1}$. Once the secant stiffness matrix is updated, Eq. (2.19) is used to solve for the new increment. Then, the new approximate solution is updated according to

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \Delta \mathbf{u}^i. \quad (2.22)$$

Now, the process moves to the next iteration, and it is repeated until the residual satisfies the convergence criterion in Eq. (2.12).

The above updating formula can save computational time on calculating the stiffness matrix at every iteration, while Eq. (2.19) still needs to be solved at each iteration. Instead of updating the secant stiffness matrix, it is possible to update the inverse of the secant stiffness matrix directly to save computational cost in solving the matrix equation. For example, Eq. (2.19) can be rewritten as

$$\Delta \mathbf{u}^i = -[\mathbf{K}_s^i]^{-1} \mathbf{R}^i \equiv -\mathbf{H}_s^i \mathbf{R}^i. \quad (2.23)$$

Thus, the inverse matrix, \mathbf{H}_s^i , is updated directly, starting from the initial inverse of the stiffness matrix. Broyden used the Sherman–Morrison formula to update the inverse of the secant stiffness matrix as

$$\mathbf{H}_s^i = \mathbf{H}_s^{i-1} + \frac{\Delta \mathbf{u}^i - \mathbf{H}_s^{i-1} \Delta \mathbf{R}}{(\Delta \mathbf{u}^i)^T \mathbf{H}_s^{i-1} \Delta \mathbf{R}} \left((\Delta \mathbf{u}^i)^T \mathbf{H}_s^{i-1} \right). \quad (2.24)$$

In general, the stiffness matrix for solid mechanics is symmetric and positive definite. However, the updated secant matrix in Eq. (2.21) and its inverse in Eq. (2.24) are not symmetric. In order to make the updated matrix symmetric and positive definite, additional constraints are required. The BFGS (Broyden, Fletcher, Goldfarb, and Shanno) method [3] satisfies these properties and is the most widely and successfully used for unconstrained optimization and very useful for finite element analysis. The main drawback of this method is that it may become unstable when the number of iterations increases. In practice, the secant stiffness matrix is reset to the stiffness matrix of the Newton–Raphson method after a certain number of iterations. The procedure can also be implemented carefully to maintain the sparsity of the stiffness matrix.

Example 2.7 (Nonlinear springs [secant method]) Using the Broyden's method, solve the displacements of the two nonlinear springs in Example 2.2. Use the initial

estimate, $\mathbf{u}^0 = \{0.1, 0.1\}^T$. Compare the number of iterations with that of Example 2.3. Also, check the convergence rate.

Solution Below is the MATLAB program for solving the nonlinear equation. The initial estimate is set to $\mathbf{u}^0 = \{0.1, 0.1\}^T$ because if it starts with $\{0, 0\}^T$, it converges to a different solution.

```
%  
% Example 2.7 Two nonlinear springs (Secant method)  
%  
tol = 1.0e-5; iter = 0; c = 0;  
u = [0.1; 0.1]; uold = u;  
f = [0; 100];  
P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2)  
     200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];  
R = P - f; Rold = R;  
conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);  
fprintf('\niter   u1      u2      conv      c');  
fprintf('\n%3d %7.5f %7.5f %12.3e %7.5f', iter, u(1), u(2), conv, c);  
Ks = [600*u(1)+400*u(2)+150 -400*u(2)+400*u(1)-100  
      400*u(1)-400*u(2)-100 400*u(2)-400*u(1)+100];  
while conv > tol && iter < 20  
    delu = -Ks\R;  
    u = uold + delu;  
    P = [300*u(1)^2+400*u(1)*u(2)-200*u(2)^2+150*u(1)-100*u(2);  
         200*u(1)^2-400*u(1)*u(2)+200*u(2)^2-100*u(1)+100*u(2)];  
    R = P - f;  
    conv= (R(1)^2+R(2)^2)/(1+f(1)^2+f(2)^2);  
    c = abs(0.9-u(2))/abs(0.9-uold(2))^2;  
    delR = R - Rold;  
    Ks = Ks + (delR-Ks*delu)*delu'/norm(delu)^2;  
    uold = u; Rold = R;  
    iter = iter + 1;  
    fprintf('\n%3d %7.5f %7.5f %12.3e %7.5f', iter, u(1), u(2), conv, c);  
end
```

Table 2.4 shows the convergence history of the Broyden's method. The algorithm converges in the fifth iteration. Note that the algorithm does not provide a converging constant, c , in Eq. (2.12). Thus, the convergence rate is less than two. But, it can be easily verified that the convergence rate is greater than one. This is common for most quasi-Newton methods in which the convergence rate is between one and two. ■

Table 2.4 Convergence history of nonlinear springs using the Broyden's method

Iteration	u_1	u_2	conv	c
0	0.1000	0.1000	1.010E+00	—
1	0.7000	1.7000	4.040E+00	1.250
2	0.3000	0.6333	1.995E-01	0.417
3	0.3727	0.8273	1.766E-02	1.023
4	0.4035	0.9094	3.170E-04	1.779
5	0.3999	0.8997	3.046E-07	3.307

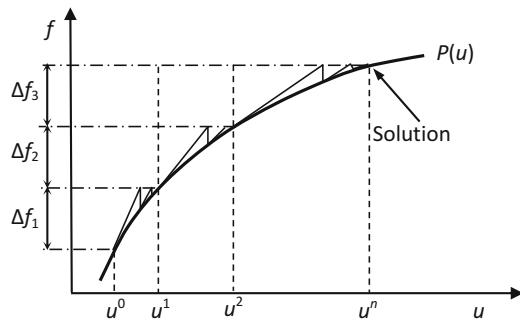
2.2.4 Incremental Force Method

Among practically available numerical methods for solving systems of nonlinear equations, the Newton–Raphson method is the fastest with a quadratic convergence, which is achieved when the initial estimate is close to the solution. Most other methods also have a similar trend. Thus, choosing an initial estimate close to the solution is an important strategy for helping the methods to converge faster.

In solid mechanics problems, the initial estimate is usually set to the undeformed shape of the structure; i.e., all displacements are initially zero. In a stable system, the magnitude of displacement is proportional to the applied load. In linear structures, for example, when the applied load is doubled, displacements are also doubled. In nonlinear structures, this proportionality is generally true even if the relation between the applied load and displacement is nonlinear; i.e., a small magnitude of displacement is expected when the applied load is small. Since the initial estimate usually starts from zero displacement, the Newton–Raphson method converges quickly to the solution when the applied load is small. The convergence difficulty occurs with large applied loads that cause a large magnitude of displacement.

The idea of the incremental force method is to apply the load in increments. Within each load increment, the procedure is the same as the standard Newton–Raphson method. The next load increment is applied after the solution corresponding to the previous load increment has converged. The converged solution at each increment is then used as an initial estimate of the next increment. Figure 2.22 illustrates the procedure for a single degree-of-freedom case. In the first increment, the nonlinear equation is solved assuming that the applied load is Δf_1 , starting from the initial estimate of $u^0 = 0$. The magnitude of this increment is chosen such that the numerical method can converge quickly to the solution, u^1 . In the second increment, the applied load increases to $\Delta f_1 + \Delta f_2$, and the initial estimate of u^1 is used, i.e., the converged solution from the previous increment. Again, the magnitude of the increment, Δf_2 , is chosen such that the numerical method can converge quickly to the solution, u^2 . The above procedure is repeated until the applied load increment reaches the full magnitude. Note that the solutions at the end of each load increment are all valid ones; they are the response of the system at the given level of load.

Fig. 2.22 Incremental force method



The load increments do not have to be uniform, especially when the nonlinearity of the system is not uniform throughout the entire load increments. Many systems are often mildly nonlinear at most load increments, but are highly nonlinear at others. For example, in the elastoplastic material, the response is mildly nonlinear when the material state is either elastic or plastic, but it becomes highly nonlinear when the material state changes from elastic to plastic. In such a case, the highly nonlinear portion will control the size of the load increments if a uniform increment is used. This is an unnecessary and wasteful use of computational resources. It is possible to divide the load into three regions—elastic, elastic-plastic, and plastic regions—and a relatively large load increment is used for the first and last regions, while a small load increment is used for the elastic-plastic region.

Even if the solution at the last load increment is the goal, it is often important to calculate the solutions in the intermediate load increments. First, the history of the response can provide insight into the problem, such as the relation between the applied load and displacement. In addition, when a structure has instability before reaching the final load step, such as bifurcation or snap through, the solutions in the intermediate load steps play an important role in estimating the bifurcation point or the critical load. Path dependence is another important reason to divide the entire load by a number of load steps. In the path-dependent problem, load increments greatly affect the accuracy of the results. For example, in plasticity, excessively large load steps may allow the stress to stay out of the yield surface or may not catch the change of the material state from elastic to plastic.

2.2.4.1 Load Increment in Commercial Software

In commercial finite element programs, the load increment is often referred to as a load step or time step. The term “time step” is sometimes used because most commercial software uses the same program for solving both the static and dynamic problems. In solving a static problem, the term time should be understood as “pseudo-time,” not physical time. In order to solve nonlinear structural problems using a finite element method, it is necessary to specify the starting time, ending

time, and time increment. In static analysis, the time interval controls how the applied load increases. At the starting time (T_{start}), the applied load is zero, and it increases proportionally until the ending time (T_{end}) at which the full magnitude is applied. Thus, at the n th increment, the applied load is calculated by

$$F^n = \frac{T^n - T_{\text{start}}}{T_{\text{end}} - T_{\text{start}}} F, \quad n = 1, \dots, N, \quad (2.25)$$

where F is the full magnitude of the applied load, and T^n is the time at the n th increment, which is calculated by

$$T^n = n \times \Delta T \leq T_{\text{end}}, \quad (2.26)$$

where ΔT is the time increment. The last time step N is determined such that T_N is less than or equal to T_{end} . If T_N is less than T_{end} , an additional time increment is performed at T_{end} . In most cases, the time starts at zero, i.e., $T_{\text{start}} = 0$, except for the case where multiple loads are applied in a sequence. For example, in order to calculate a permanent deformation of a bar, a force that can cause plastic deformation is applied in the first load, and then it is reduced to zero in the second load. In such a case, the starting time of the second load is the ending time of the first load.

2.2.4.2 Automatic Time Stepping

In many cases, it is not trivial to estimate appropriate time steps. Time steps that are too small can help convergence, but it will take quite an amount of computational cost to finish solving the entire load increment. On the other hand, if the time step is too large, the numerical method may not converge, and the iteration will stop when it reaches the maximum allowed iterations. There is no good guideline of how to choose an appropriate time step. It depends on the level of nonlinearity of the system. The best way of checking if the load step used is too small or too large is to count the number of iterations. When the standard Newton–Raphson method is used, the load step is considered to be appropriate if the solution converges in the fifth or sixth iteration. If the solution converges faster than that, the load step can be considered too small and can be increased without reducing the convergence much. On the other hand, if the convergence occurs beyond the tenth iteration, the load step is too large and it would be better if a smaller load step is used. Many commercial programs have the capability of adaptively adjusting the size of time steps by monitoring the number of convergence iterations.

Although adaptive time stepping is a useful tool to gradually control the size of time steps, it is possible that the iteration may not converge if nonlinearity is suddenly introduced. For example, in a contact problem, two bodies are disconnected in one load increment and then in contact in the following increment. Thus, the two bodies suddenly experience a contact force in the interface. The time step

needs to be small enough in order to capture this sudden change in contact force. Let us assume that the current time step is large enough so that the iteration does not converge. In such a case, it is possible to go back to the previously converged time and to reduce the size of the time step by half. If the iteration does not converge with the new time step, it is further reduced by half. This reduction can be repeated until the iteration converges. Once the iteration converges with the reduced time step, a regular adaptive time stepping can be resumed in the following increments. If the formulation or physics of the problem has a fundamental difficulty, the solution may not converge, no matter how many reductions are done. In order to prevent the situation when the time step approaches zero, most commercial programs have the maximum allowed number of reductions and the programs stop with an error message when this number is reached.

2.2.4.3 Force Control vs. Displacement Control

So far, the solution of nonlinear equations for structural problems is explained as equilibrium under applied loads. The objective of the nonlinear spring examples is to find the displacements of the springs for the load at the end. Referring to Fig. 2.23a, the force-controlled solution procedure finds the displacements, u_1, u_2, \dots, u_n , when the force increases to F_1, F_2, \dots, F_n . Since a one-to-one relation exists between the force and displacement, the opposite procedure also works well: finding the reaction forces F_1, F_2, \dots, F_n , when displacement increases to u_1, u_2, \dots, u_n . This is called a displacement-controlled solution procedure. Mathematically, these two procedures are equivalent, but practically, the displacement-controlled procedure can be more stable than the force-controlled one. Consider the load-displacement curve in Fig. 2.23b. The load starts reducing after it reaches the maximum point at C. This type of softening behavior occurs in elastoplastic material due to necking. If the force-controlled procedure is employed, it is not easy to reach point D in the curve as the structure reaches equilibrium at B with the given load F_B . However, in the

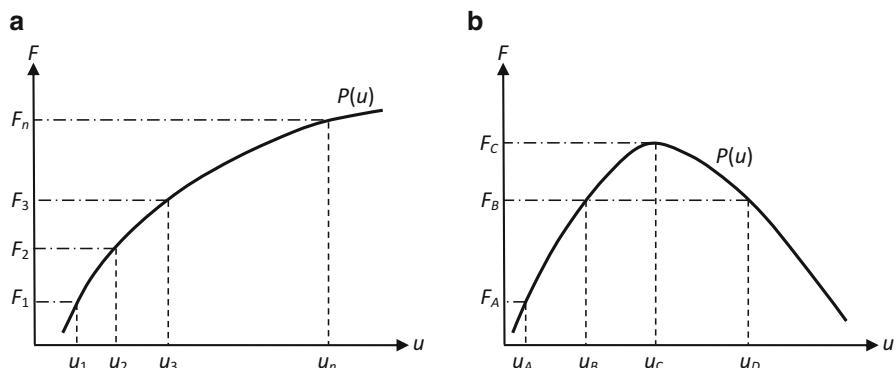


Fig. 2.23 Displacement controlled solution procedure

displacement-controlled procedure, it is possible to reach point D by gradually increasing the displacement and finding the reaction force. Another important aspect of nonlinear problems is that the behavior of the system is unknown in advance. Thus, if a load that is greater than F_C is applied, then the iteration will not converge as no solution exists. However, in the displacement-controlled procedure, the solution can be converged in a broader range of displacements.

Example 2.8 (Displacement-controlled solution procedure) For the nonlinear springs in Example 2.2, plot the force–displacement curve by increasing the displacement, u_2 , from zero to 0.9 with nine increments.

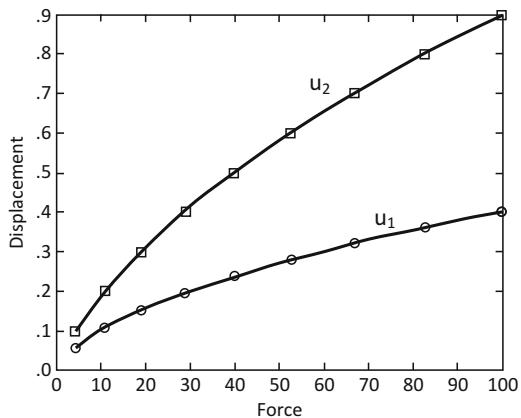
Solution The system of nonlinear equations for the two springs is written below

$$\begin{cases} 300u_1^2 + 400u_1u_2 - 200u_2^2 + 150u_1 - 100u_2 = 0 \\ 200u_1^2 - 400u_1u_2 + 200u_2^2 - 100u_1 + 100u_2 = F \end{cases}. \quad (2.27)$$

Since the displacement, u_2 , is controlled, the applied force, F , and displacement, u_1 , are unknown now. In such a case, it is possible to solve the first equation for u_1 and then, to use the second equation to solve for F . Thus, the problem becomes a nonlinear algebraic equation. Below is the MATLAB program that solves for the nonlinear equation.

```
%  
% Example 2.8 Displacement controlled procedure  
%  
tol = 1.0e-5; conv = 0; u1 = 0; u1old = u1;  
fprintf('\n step u1 u2 F');  
% Displacement increment loop  
for i=1:9  
    u2 = 0.1*i;  
    P = 300*u1^2+400*u1*u2-200*u2^2+150*u1-100*u2;  
    R = -P;  
    conv = R^2;  
    % Convergence loop  
    iter = 0;  
    while conv > tol && iter < 20  
        Kt = 600*u1+400*u2+150;  
        delu1 = R/Kt;  
        u1 = u1old + delu1;  
        P = 300*u1^2+400*u1*u2-200*u2^2+150*u1-100*u2;  
        R = -P;  
        conv= R^2;  
        u1old = u1;  
        iter = iter + 1;  
    end  
    F = 200*u1^2-400*u1*u2+200*u2^2-100*u1+100*u2;  
    fprintf('\n %3d %7.5f %7.5f %7.3f', i, u1, u2, F);  
end
```

Fig. 2.24 Force–displacement curves for two nonlinear springs



The program has two loops: the outer loop is for load increments and the inner loop is for convergence iterations. Each increment requires about two or three iterations. Once the iteration is converged, the reaction force is calculated from the second equation in Eq. (2.27). Figure 2.24 shows the force–displacement curve for the nonlinear springs. It is identical with Fig. 2.15 that is created using the force-controlled procedure. ■

2.3 Steps in the Solution of Nonlinear Finite Element Analysis

As discussed in the previous sections, there are several aspects in which the solution procedure of nonlinear problems is different from that of linear problems. Although different procedures are required for different types of nonlinear problems, the basic steps for nonlinear static problems are outlined in this section.

In structural finite element analysis, unknown variables are usually nodal displacements. In iterative algorithms, the displacement increments are calculated at each iteration and the total displacements are updated using the increments until they converge. Thus, the most important step in the solution process is calculating the incremental displacements. In the following, it is assumed that the $k - 1$ th iteration is completed, which means that all states at the $k - 1$ th iteration are available and the displacement vector, \mathbf{d}^k , is given.

2.3.1 State Determination

For the given displacement vector, \mathbf{d}^k , it is necessary to calculate the current states of the system, such as strains and stresses for structural problems. In order to

simplify the following explanations, the entire structure is modeled by one finite element, and the solution of the structure is approximated by a vector, $\mathbf{d}^k = \{d_1, d_2, \dots, d_n\}^T$, of the nodal displacements. Accordingly, it is also assumed that there is a suitable vector of interpolation functions, $\mathbf{N}(\mathbf{x}) = \{N_1, N_2, \dots, N_n\}^T$. Then, the displacement at a point \mathbf{x} in the structure can be approximated by

$$\mathbf{u}^k(\mathbf{x}) = \mathbf{N}(\mathbf{x}) \cdot \mathbf{d}^k \quad (2.28)$$

It is noted that the interpolation function, $\mathbf{N}(\mathbf{x})$, is often given in the reference coordinate (see Sect. 1.5). The element strain vector can be computed by appropriate differentiation as follows:

$$\boldsymbol{\epsilon}^k = \mathbf{B} \cdot \mathbf{d}^k, \quad (2.29)$$

where \mathbf{B} is the strain–displacement matrix. In general, stress is a function of strain. For nonlinear problems, this relation can be written in the following form:

$$\boldsymbol{\sigma}^k = f(\boldsymbol{\epsilon}^k). \quad (2.30)$$

When the material is linear elastic, the above relation is equivalent to Eq. (2.3). In some materials, the above stress calculation involves the entire history of deformation.

2.3.2 Residual Calculation

Once the state of the structure is determined, the next step is to check if the structure is in equilibrium or not. If it is in equilibrium, the nodal forces due to internal stresses must be equal and opposite in direction to the applied nodal forces. More specifically, the weak form of structural equilibrium can be written as

$$\iiint_{\Omega} \boldsymbol{\epsilon}(\bar{\mathbf{u}})^T \boldsymbol{\sigma} d\Omega = \iint_{\Gamma_s} \bar{\mathbf{u}}^T \mathbf{t} d\Gamma + \iiint_{\Omega} \bar{\mathbf{u}}^T \mathbf{f}^b d\Omega, \quad (2.31)$$

which must be satisfied for all virtual displacements, $\bar{\mathbf{u}}$, that satisfy the essential boundary conditions. In the Galerkin approximation method, the virtual displacement is interpolated using a similar form as in Eq. (2.28), and thus, the virtual strain, $\boldsymbol{\epsilon}(\bar{\mathbf{u}})$, as in Eq. (2.29). By substituting these into the weak form, we have

$$\bar{\mathbf{d}}^T \left(\iiint_{\Omega} \mathbf{B}^T \boldsymbol{\sigma} d\Omega = \iint_{\Gamma_s} \mathbf{N}^T \mathbf{t} d\Gamma + \iiint_{\Omega} \mathbf{N}^T \mathbf{f}^b d\Omega \right). \quad (2.32)$$

Since (2.32) must be satisfied for arbitrary virtual displacements, $\bar{\mathbf{d}}$, it is equivalent to the following equation:

$$\iiint_{\Omega} \mathbf{B}^T \boldsymbol{\sigma} d\Omega = \iint_{\Gamma_s} \mathbf{N}^T \mathbf{t} d\Gamma + \iiint_{\Omega} \mathbf{N}^T \mathbf{f}^b d\Omega. \quad (2.33)$$

The left-hand side represents the equivalent nodal forces due to internal stresses, while the right-hand side represents equivalent nodal forces due to applied forces. If the above equation is satisfied for the given displacements, then the structure is in equilibrium. However, when the structure is not in equilibrium, the difference between them is defined as a vector of residuals as

$$\mathbf{R}^k = \iint_{\Gamma_s} \mathbf{N}^T \mathbf{t} d\Gamma + \iiint_{\Omega} \mathbf{N}^T \mathbf{f}^b d\Omega - \iiint_{\Omega} \mathbf{B}^T \boldsymbol{\sigma}^k d\Omega. \quad (2.34)$$

The superscript, k , is used to denote that the structure is not in equilibrium at the k th iteration. In practice, the integrals in the residual are calculated using numerical integration, such as Gauss quadrature in Sect. 1.5.

2.3.3 Convergence Check

The purpose of nonlinear finite element analysis is to satisfy the equilibrium equation, such as the one in Eq. (2.31), which is equivalent to making the vector of residuals in Eq. (2.34) to vanish. The iteration stops when the magnitude of the residual vector is less than a specific tolerance. In that case, the iteration converges and the solution is the current displacement. However, the iterations may not converge in some cases no matter how many iterations are conducted. In order to prevent an infinite loop of the convergence iteration, the program usually stops when the iteration counter reaches the maximum allowed number of iterations. In that case, the algorithm stops with an error message. In order to prevent stopping the algorithm with errors, the force can be halved and the convergence is tried again, which is called the bisection method. The bisection method can be repeated until either the iteration converges or the maximum allowed number of bisections is reached.

2.3.4 Linearization

Linearization is one of the most important steps in solving nonlinear equations. In this step, the Jacobian matrix is calculated for the Newton–Raphson method, or

the previous Jacobian matrix is updated in the secant method. In structural problems, the Jacobian matrix is often called the tangent stiffness matrix because it is a tangent line of the force–displacement curve in a single DOF case. For practical applications, the structural equations from the finite element method are not given in the form of simple polynomials as in Example 2.2. Thus, linearization is the most complicated procedure, theoretically and computationally. In the following chapters, linearization of different nonlinearities will be discussed in detail. It is noted that errors in the tangent stiffness can cause slow convergence or sometimes divergence. The accuracy of the solution is controlled by residuals.

2.3.5 Solution

Once the tangent stiffness matrix and the vector of residuals are calculated, the following system of linear equations is solved for incremental displacement:

$$\mathbf{K}_T^k \cdot \Delta \mathbf{d}^k = \mathbf{R}^k, \quad (2.35)$$

where $\Delta \mathbf{d}^k$ is the vector of incremental displacements. In order to have a unique solution to the above equation, the tangent stiffness matrix must be positive definite. Then, the total displacement is updated by

$$\mathbf{d}^{k+1} = \mathbf{d}^k + \Delta \mathbf{d}^k. \quad (2.36)$$

Now, the k th iteration is completed and the procedure is repeated for the $(k+1)$ th iteration.

Example 2.9 (Nonlinear bar) A rubber bar of length $L = 1\text{ m}$ is under an axial force, $F = 10\text{ kN}$, as shown in Fig. 2.25. The material has the following nonlinear stress–strain relation: $\sigma = E \cdot \tan^{-1}(m\epsilon)$, where the material constants $E = 100\text{ MPa}$

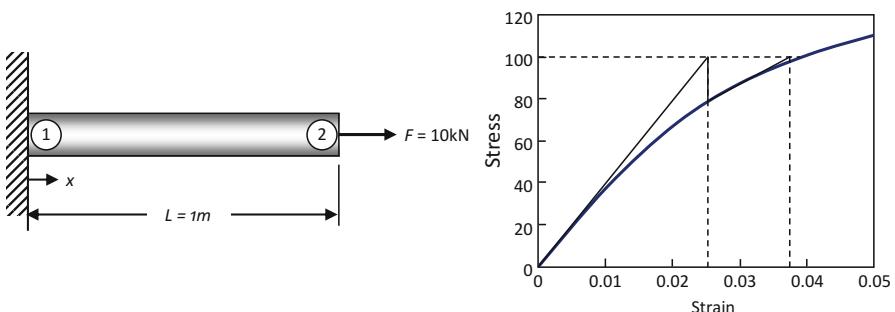


Fig. 2.25 Newton–Raphson iteration of a nonlinear bar

and $m=40$. A uniform stress distribution is assumed over the cross section, $A = 10^{-4} \text{ m}^2$, and the axial force is also assumed uniformly distributed. Assuming an infinitesimal deformation, perform the first two Newton–Raphson iterations to find the strain, stress, and displacement at the tip (node 2) of the element.

Solution The discrete weak form of the bar element can be written as

$$\bar{\mathbf{d}}^T \int_0^L \mathbf{B}^T \sigma A dx = \bar{\mathbf{d}}^T \mathbf{F},$$

where $\mathbf{d} = [d_1, d_2]^T$ is the vector of nodal displacements, $\bar{\mathbf{d}}$ is the vector of virtual nodal displacements, $\mathbf{B}^T = [-1, 1]/L$ is the displacement–strain matrix, and $\mathbf{F} = [F_1, F_2]^T$ is the vector of applied forces. In order to simplify the following steps, the essential boundary condition can be applied in advance, i.e., $d_1 = \bar{d}_1 = 0$. For simplicity of notation, $d = d_2$ and $F = F_2$ will be used in the following derivations. Then, the above discrete weak form becomes a scalar equation. The residual now becomes

$$\begin{aligned} R &= F - \int_0^L \frac{\sigma A}{L} dx \\ \Rightarrow R &= F - \sigma(d)A. \end{aligned}$$

Note that the residual is nothing but the equilibrium between external and internal forces: $P(d) = F$. Due to the nonlinear stress–strain relation, the Newton–Raphson method is used to find the displacement, d , to eliminate the residual. The Jacobian becomes

$$\frac{dP}{dd} = \frac{d\sigma(d)}{dd} A = \frac{d\sigma}{de} \frac{de}{dd} A.$$

The first derivative on the right-hand side can be calculated by differentiating the stress–strain relation and the second derivative from the displacement–strain relation. Using these relations, we have

$$\frac{dP}{dd} = \frac{1}{L} m A E \cos^2\left(\frac{\sigma}{E}\right).$$

Then, the Newton–Raphson equation at the k th iteration becomes

$$\left[\frac{1}{L} m A E \cos^2\left(\frac{\sigma^k}{E}\right) \right] \Delta d^k = F - \sigma^k A.$$

Iteration 1: $d^0 = \epsilon^0 = \sigma^0 = 0$. Then, the above equation becomes

$$\frac{mAE}{L} \Delta d^0 = F,$$

which yields $\Delta d^0 = 0.025$. Then, all variables are updated at the new configuration, as

$$\begin{aligned} d^1 &= d^0 + \Delta d^0 = 0.025 \text{ m}, \\ \epsilon^1 &= d^1/L = 0.025, \\ \sigma^1 &= E \tan^{-1}(m\epsilon^1) = 78.5 \text{ MPa}. \end{aligned}$$

Iteration 2: The Newton–Raphson equation becomes

$$\left[\frac{mAE}{L} \cos^2 \left(\frac{\sigma^1}{E} \right) \right] \Delta d^1 = F - \sigma^1 A,$$

which yields $\Delta d^1 = 0.0107$. Then, all variables are updated at the new configuration, as

$$\begin{aligned} d^2 &= d^1 + \Delta d^1 = 0.0357 \text{ m}, \\ \epsilon^2 &= d^2/L = 0.0357, \\ \sigma^2 &= E \tan^{-1}(m\epsilon^2) = 96 \text{ MPa}. \end{aligned}$$

Iteration 3 will yield a stress value of $\sigma = 99.7$ MPa, which is only 0.3 % different from the exact solution. ■

2.4 MATLAB Code for a Nonlinear Finite Element Analysis Procedure

Although the previous section summarized five important steps in a nonlinear finite element analysis procedure, it is important to understand how these steps are executed in sequence. Figure 2.26 shows a flowchart for a nonlinear finite element analysis procedure. A MATLAB program, NLFEA, is also listed as an example. The procedure assumes the incremental force method with the Newton–Raphson method. The bisection method is used when the Newton–Raphson method failed to converge. The modified Newton–Raphson method and automatic time stepping method can easily be implemented by modifying the current implementation.

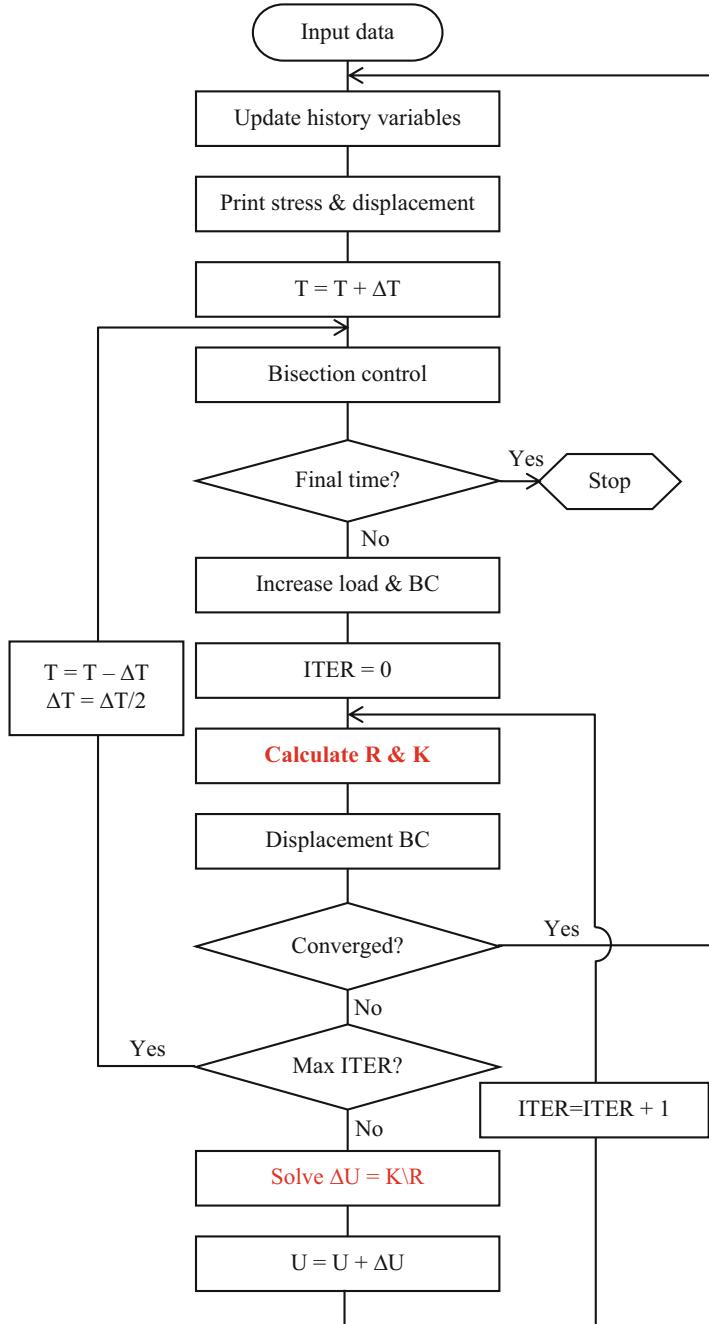


Fig. 2.26 Flowchart for nonlinear finite element procedure

```

function NLFEA(ITRA,TOL,ATOL,NTOL,TIMS,NOUT,MID,PROP,EXTFORCE,SDISPT,
XYZ,LE)
%*****
% MAIN PROGRAM FOR HYPERELASTIC/ELASTOPLASTIC ANALYSIS
%*****
%%%
% Global DISPDD DISPTD FORCE GKF %Global variables
%
[NUMNP, NDOF] = size(XYZ); % Analysis parameters
NE = size(LE,1);
NEQ = NDOF*NUMNP;
%
DISPTD=zeros(NEQ,1); DISPDD=zeros(NEQ,1); % Nodal displacement &
                                             increment
if MID >= 0, ETAN=PLSET(PROP, MID, NE); end % Initialize material
                                               properties
%
ITGZONE(XYZ, LE, NOUT); % Check element connectivity
%
% Load increments [Start End Increment InitialLoad FinalLoad]
NLOAD=size(TIMS,2);
ILOAD=1; % First load increment
TIMEF=TIMS(1,ILOAD); % Starting time
TIMEI=TIMS(2,ILOAD); % Ending time
DELTA=TIMS(3,ILOAD); % Time increment
CUR1=TIMS(4,ILOAD); % Starting load factor
CUR2=TIMS(5,ILOAD); % Ending load factor
DELTAO = DELTA; % Saved time increment
TIME = TIMEF; % Starting time
TDELTA = TIMEI - TIMEF; % Time interval for load step
ITOL = 1; % Bisection level
TARY=zeros(NTOL,1); % Time stamps for bisections
%
% Load increment loop
%-----
ISTEP = -1; FLAG10 = 1;
while(FLAG10 == 1) % Solution has been converged
  FLAG10 = 0; FLAG11 = 1; FLAG20 = 1;
  %
  CDISP = DISPTD; % Store converged
                   displacement
  %
  if(ITOL==1) % No bisection
    DELTA = DELTA0;
    TARY(ITOL) = TIME + DELTA;
  else % Recover previous bisection
    ITOL = ITOL-1; % Reduce the bisection level
    DELTA = TARY(ITOL)-TARY(ITOL+1); % New time increment
  end
end

```

```

TARY(ITOL+1) = 0;                                % Empty converged bisection
                                                level
ISTEP = ISTEP - 1;                                % Decrease load increment
end
TIME0 = TIME;                                     % Save the current time
%
% Update stresses and history variables
UPDATE=true; LTAN=false;
if MID ==0, ELAST3D(ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE);
elseif MID > 0, PLAST3D(MID, PROP, ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE);
elseif MID < 0, HYPER3D(PROP, UPDATE, LTAN, NE, NDOF, XYZ, LE);
else fprintf(NOUT, '\t\t *** Wrong material ID ***\n'); return;
end
%
% Print results
if (ISTEP>=0), PROUT(NOUT, TIME, NUMNP, NE, NDOF); end
%
TIME = TIME + DELTA;                             % Increase time
ISTEP = ISTEP + 1;
%
% Check time and control bisection
while (FLAG11 == 1)                            % Bisection loop start
    FLAG11 = 0;
    if ((TIME-TIMEI)>1E-10)                    % Time passed the end time
        if ((TIMEI+DELTA-TIME)>1E-10)          % One more at the end time
            DELTA=TIMEI+DELTA-TIME;             % Time increment to the end
            DELTA0=DELTA;                      % Saved time increment
            TIME=TIMEI;                      % Current time is the end
        else
            ILOAD=ILOAD+1;                   % Progress to next load step
            if (ILOAD>NLOAD)                % Finished final load step
                FLAG10 = 0;                 % Stop the program
                break;
            else                           % Next load step
                TIME=TIME-DELTA;
                DELTA=TIMS(3, ILOAD);
                DELTA0=DELTA;
                TIME = TIME + DELTA;
                TIMEF = TIMS(1, ILOAD);
                TIMEI = TIMS(2, ILOAD);
                TDELTIA = TIMEI - TIMEF;
                CUR1 = TIMS(4, ILOAD);
                CUR2 = TIMS(5, ILOAD);
            end
        end
    end
%
% Load factor and prescribed displacements

```

```
FACTOR = CUR1 + (TIME-TIMEF)/TDELTA*(CUR2-CUR1);
SDISP = DELTA*SDISPT(:,3)/TDELTA*(CUR2-CUR1);
%
% Start convergence iteration
%-----
ITER = 0;
DISPDD = zeros(NEQ,1);
while(FLAG20 == 1)
    FLAG20 = 0;
    ITER = ITER + 1;
    % Check max iteration
    if(ITER>ITRA), error('Iteration limit exceeds'); end
    %
    % Initialize global stiffness K and residual vector F
    GKF = sparse(NEQ,NEQ);
    FORCE = sparse(NEQ,1);
    %
    % Assemble K and F
    UPDATE=false; LTAN=true;
    if MID ==0, ELAST3D(ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE);
    elseif MID > 0, PLAST3D(MID, PROP, ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE);
    elseif MID < 0, HYPER3D(PROP, UPDATE, LTAN, NE, NDOF, XYZ, LE);
    end
    %
    % Increase external force
    if size(EXTFORCE,1)>0
        LOC = NDOF*(EXTFORCE(:,1)-1)+EXTFORCE(:,2);
        FORCE(LOC) = FORCE(LOC) + FACTOR*EXTFORCE(:,3);
    end
    %
    % Prescribed displacement BC
    NDISP=size(SDISPT,1);
    if NDISP~=0
        FIXEDDOF=NDOF*(SDISPT(:,1)-1)+SDISPT(:,2);
        GKF(FIXEDDOF,:)=zeros(NDISP,NEQ);
        GKF(FIXEDDOF,FIXEDDOF)=PROP(1)*eye(NDISP);
        %
        FORCE(FIXEDDOF)=0;
        if ITER==1, FORCE(FIXEDDOF) = PROP(1)*SDISP(:,); end
    end
    % Check convergence
    if(ITER>1)
        FIXEDDOF=NDOF*(SDISPT(:,1)-1)+SDISPT(:,2);
        ALLDOF=1:NEQ;
        FREEDOF=setdiff(ALLDOF,FIXEDDOF);
        RESN=max(abs(FORCE(FREEDOF)));
        OUTPUT(1, ITER, RESN, TIME, DELTA)
    %
```

```

if (RESN<TOL)
  FLAG10 = 1;
  break;
end
%
if ( (RESN>ATOL) || (ITER>=ITRA) )           % Start bisection
  ITOL = ITOL + 1;
  if (ITOL<NTOL)
    DELTA = 0.5*DELTA;
    TIME = TIME0 + DELTA;
    TARY(ITOL) = TIME;
    DISPTD=CDISP;
    fprintf(1,'Not converged. Bisecting load increment %3d\n',ITOL);
  else
    error('Max No. of bisection');
  end
  FLAG11 = 1;
  FLAG20 = 1;
  break;
end
%
% Solve the system equation
if (FLAG11 == 0)
  SOLN = GKF\FORCE;
  DISPDD = DISPDD + SOLN;
  DISPTD = DISPTD + SOLN;
  FLAG20 = 1;
else
  FLAG20 = 0;
end
if (FLAG10 == 1), break; end
end                               %20 Convergence iteration
end                               %11 Bisection
end                               %10 Load increment
%
% Successful end of program
fprintf(NOUT,' \t\t *** Successful end of program ***\n');
end

```

```

function OUTPUT(FLG, ITER, RESN, TIME, DELTA)
%*****
% Print convergence iteration history
%*****
%%
if FLG == 1
  if ITER>2
    fprintf(1,'%27d %14.5e \n',ITER,full(RESN));
  else

```

```

fprintf(1,' \n \t Time Time step Iter \t Residual \n');
fprintf(1,'%10.5f %10.3e %5d %14.5e \n',TIME,DELTA,ITER,full(RESN));
end
end
end



---


function PROUT(NOUT, TIME, NUMNP, NE, NDOF)
%*****
% Print converged displacements and stresses
%*****
%%%
global SIGMA DISPTD
%
fprintf(NOUT, '\r\n\r\nTIME = %11.3e\r\n\r\nNodal Displacements\r\n',
TIME);
fprintf(NOUT, '\r\nNode      U1      U2      U3');
for I=1:NUMNP
    II=NDOF*(I-1);
    fprintf(NOUT, '\r\n%5d %11.3e %11.3e %11.3e',I,DISPTD(II+1:II+3));
end
fprintf(NOUT, '\r\n\r\nElement Stress\r\n');
fprintf(NOUT, '\r\nS11      S22      S33      S12      S23      S13');
for I=1:NE
    fprintf(NOUT, '\r\nElement %5d',I);
    II=(I-1)*8;
    fprintf(NOUT, '\r\n%11.3e %11.3e %11.3e %11.3e %11.3e %11.3e',SIGMA
    (1:6,II+1:II+8));
end
fprintf(NOUT, '\r\n\r\n');
end



---


function ETAN=PLSET(PROP, MID, NE)
%*****
% Initialize history variables and elastic stiffness matrix
% XQ      : 1-6 = Back stress alpha, 7 = Effective plastic strain
% SIGMA   : Stress for rate-form plasticity
%          : Left Cauchy-Green tensor XB for multiplicative plasticity
% ETAN    : Elastic stiffness matrix
%*****
%%%
global SIGMA XQ
%
LAM=PROP(1);
MU=PROP(2);
%
N = 8*NE;
%
if MID > 30
    SIGMA=zeros(12,N);

```

```

XQ=zeros(4,N);
SIGMA(7:9,:)=1;
ETAN=[LAM+2*MU LAM LAM ;
      LAM LAM+2*MU LAM ;
      LAM LAM LAM+2*MU];
else
  SIGMA=zeros(6,N);
  XQ=zeros(7,N);
  ETAN=[LAM+2*MU LAM LAM 0 0 0;
        LAM LAM+2*MU LAM 0 0 0;
        LAM LAM LAM+2*MU 0 0 0;
        0 0 0 MU 0 0;
        0 0 0 0 MU 0;
        0 0 0 0 0 MU];
end
end

```

```

function VOLUME = ITGZONE(XYZ, LE, NOUT)
%*****
% Check element connectivity and calculate volume
%*****
%%
EPS=1E-7;
NE = size(LE,1);
VOLUME=0;
for I=1:NE
  ELXY=XYZ(LE(I,:),:);
  [~,~,DET] = SHAPEL([0 0 0], ELXY);
  DVOL = 8*DET;
  if DVOL < EPS
    fprintf(NOUT, '\n??? Negative Jacobian ???\nElement connectivity\n');
    fprintf(NOUT, '%5d', LE(I,:));
    fprintf(NOUT, '\nNodal Coordinates\n');
    fprintf(NOUT, '%10.3e %10.3e %10.3e\n', ELXY');
    error('Negative Jacobian');
  end
  VOLUME = VOLUME + DVOL;
end
end

```

The analysis procedure is composed of three nested loops. The first one (loop 10) is the loop for load steps and load increments. In general, the entire solution procedure is composed of **NLOAD** load steps. Multiple load steps are useful when cyclic loadings are applied. Each load step is composed of several load increments.

This corresponds to the incremental force method in Sect. 2.2.4, in which the total applied load is divided by a number of increments.

The second one (loop 11) is for bisection. As discussed before, the Newton–Raphson method can have a difficulty in converging when the starting point is far from the solution. Therefore, the convergence can be improved by reducing the amount of the load increment. Whenever the convergence iteration fails to converge, the load increment is halved, and then, loop 11 is repeated starting from the previously converged point. Since the solution may not converge no matter how small the load increment is, the bisection process stops after the maximum number of bisections is reached. For the purpose of bisection, it is necessary to store the previously converged displacement (**CDISP**).

The third, innermost loop (loop 20) is for convergence iteration. This corresponds to the Newton–Raphson iteration. The major part of this loop is devoted to calculating the residual vector, **FORCE**, and the tangent matrix, **GKF**. If the residual becomes less than a threshold, it is considered that the iteration has been converged. In such a case, the loop ends, and the procedure moves to the next load increment. If iterations do not converge, then bisection is invoked by reducing the load increment by half.

Since MATLAB copies variables when they are sent as an argument of a function, it takes a lot of computer memory and time. Therefore, it is better to define them as a global variable when the size of a variable is large. In the current implementation, several variables are defined as global variables, as summarized in Table 2.5.

During nonlinear analysis, NLFEA calls for four functions: **OUTPUT**, **PROUT**, **ETAN**, and **ITGZONE**. **OUTPUT** function is used to print out iteration history to the MATLAB screen, while **PROUT** prints out analysis results (displacements and stresses) to the output file. Both **ETAN** and **ITGZONE** are called only once before Newton–Raphson iteration starts. **ETAN** is to calculate the initial elastic stiffness for linear elastic and elastoplastic materials, i.e., when **MID** ≥ 0 . **ITGZONE** checks the determinant of element Jacobian matrix and stops NLFEA if the determinant is negative, which indicates that either element has a negative volume or the element connectivity is not correct.

The input data include nodal coordinates, element connectivity, force and displacement boundary conditions, material parameters, and control parameters for the solution procedure. The following sample input data are for one element under z-directional extension. Note that at the end of input data file, NLFEA is called.

Table 2.5 Global arrays for NLFEA.m program

Name	Dimension	Contents
GKF	$NEQ \times NEQ$	Tangent matrix
FORCE	$NEQ \times 1$	Residual vector
DISPTD	$NEQ \times 1$	Displacement vector
DISPDD	$NEQ \times 1$	Displacement increment
SIGMA	$6 \times 8 \times NE$	Stress at each integration point
XQ	$7 \times 8 \times NE$	History variable at each integration point

```

%
% Extension of single element example
%
% Nodal coordinates
XYZ=[0 0 0;1 0 0;1 1 0;0 1 0;0 0 1;1 0 1;1 1 1;0 1 1];
%
% Element connectivity
LE=[1 2 3 4 5 6 7 8];
%
% External forces [Node, DOF, Value]
EXTFORCE=[5 3 10.0; 6 3 10.0; 7 3 10.0; 8 3 10.0];
%
% Prescribed displacements [Node, DOF, Value]
SDISPT=[1 1 0;1 2 0;1 3 0;2 2 0;2 3 0;3 3 0;4 1 0;4 3 0];
%
% Load increments [Start End Increment InitialFactor FinalFactor]
TIMS=[0.0 0.5 0.1 0.0 0.5; 0.5 1.0 0.0 1.0 5 1.0]';
%
% Material properties
%PROP=[LAMBDA MU BETA H Y0]
MID=1;
PROP=[110.747, 80.1938, 0.0, 5., 35.0];
%
% Set program parameters
ITRA=20; ATOL=1.0E5; NTOL=5; TOL=1E-6;
%
% Calling main function
NOUT = fopen('output.txt','w');
NLFEA(ITRA, TOL, ATOL, NTOL, TIMS, NOUT, MID, PROP, EXTFORCE, SDISPT, XYZ, LE);
fclose(NOUT);

```

Nodal coordinates and element connectivity: The current implementation assumes that the node numbers are in sequence. Then, the nodal coordinates are defined using $NNODE \times 3$ matrix **XYZ**. Since the current implementation only supports eight-node hexahedral elements, element connectivity is defined using $NELEN \times 8$ matrix **LE**.

Applied forces and prescribed displacements: Both applied forces and prescribed displacements are given in the format of [node, DOF, value]. DOF is the coordinate direction: 1, 2, or 3. The external force, **EXTFORCE**, may not be required, but the prescribed displacements, **SDISPT**, must be defined in order to remove the rigid-body motion error in static problems.

Load steps and increments: The **TIMS** array is used to define load steps and load increments. Each row of **TIMS** array represents a load step. Each load step has the start time, end time, and increment, which is used for load increment. The initial and final factors are the load factors that will vary during the current load step.

For example, the following **TIMS** array has 10 increments during which the load increases from 0 to 50 % of the total load.

```
TIMS = [0.0, 1.0, 0.1, 0.0, 0.5]
```

The end time of the previous load step and the start time of the following load step must be the same. Otherwise, the program will not function properly. The same is true for the load factors.

Material properties: The current implementation supports three types of nonlinear materials: linear elastic material (**MID**=0), Mooney-Rivlin hyperelasticity (**MID**=−1), and infinitesimal elastoplasticity (**MID**=1).² For a linear elastic material, two Lame's constants are enough to define the material properties: **PROP**=[**LAMBDA**, **MU**]. For hyperelasticity, material properties are composed of two material constants, A10 and A01, and a bulk modulus, D: **PROP**=[**A10**, **A01**, **D**]. For elastoplasticity, the required material properties are two Lame's constants (λ and μ), hardening type (β), plastic modulus (H), and initial yield strength (Y_0). The meaning of these properties will be discussed in Chaps. 3 and 4.

Analysis control parameters: There are several parameters that control the analysis procedure. **ITRA** is the maximum number of convergence iterations in the Newton–Raphson method. If the number of iterations reaches **ITRA**, it is considered that the analysis cannot converge, and the bisection is invoked. During the convergence iteration, if the residual increases larger than **ATOL**, then it is considered that the solution is diverging, and the bisection process is invoked. The total number of bisections is limited by **NTOL**. That is, if the convergence iterations do not converge after **NTOL** consecutive bisections, the program stops with an error message. The convergence iteration is considered converged when the norm of the residual is less than **TOL**. Once the solution is converged at each load increment, nodal displacements and stresses at integration points are printed to an output file designated by **NOUT**.

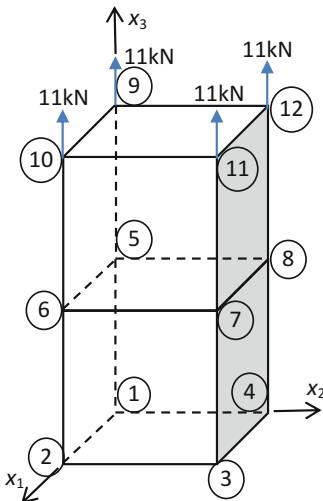
Example 2.10 (Tension of an elastoplastic bar) An elastoplastic bar in the dimension of 1 cm × 1 cm × 2 cm is under axial load as shown in Fig. 2.27. Using two eight-node finite elements, solve for displacements and stresses for both elements. Assume material properties of $\lambda=110.7$ GPa, $\mu=80.2$ GPa, $\sigma_Y=400$ MPa, and $H=100$ MPa.

Solution Since the total applied load is 44 kN, it can be expected that the material will be beyond its yield strength of 400 MPa. In order to capture the transition of material states, the load steps are divided into two parts. In **TIMS** array,

```
TIMS = [0.0 0.8 0.4 0.0 0.8;
        0.8 1.1 0.1 0.8 1.1]';
```

² In Chap. 4, different plasticity models can be used by changing **MID**=1, 2, and 31.

Fig. 2.27 Two brick elements under uniaxial tension



The first load step is for elastic portion. Assuming that the exact load at yield is unknown, 8 kN at each node is applied at the first load step with 4 kN increment. Therefore, only two increments are performed in the first load step. In the second load step, the applied load is increased between 8 and 11 kN with the increment of 1 kN. In order to show how the load factor is working, the input file intentionally applied 10 kN at each node, and the load factor is increased by 1.1, so that the total applied load is 11 kN at the end.

A nodal force of 10 kN is applied at the four nodes on the top, while the bottom four nodes are fixed in such a way that the uniaxial tension condition can be met; that is, $u_{1x} = u_{1y} = u_{1z} = u_{2y} = u_{2z} = u_{3z} = u_{4z} = 0$. Nodal coordinates are defined in **XYZ** array, and element connectivity is in **LE** array. **EXTFORCE** stores externally applied force, and **SDISPT** stores prescribed displacements. These two arrays are given in the format such that each row includes [Node, DOF, Value] format. For elastoplastic material, **MID = 1** is used with five material constants in **PROP** array; they are two Lame's constants, **LAMBDA** and **MU**; hardening parameters, **BETA** and **H**; and initial yield strength, **Y0**. Detailed explanation of elastoplastic material properties will be presented in Chap. 4.

PROP = [110.7E9 80.2E9 0.0 1.E8 4.0E8]

The problem definition and calling NLFEA are listed as follows:

```
%  
% Two-element example  
%  
% Nodal coordinates  
XYZ=[0 0 0; 1 0 0; 1 1 0; 0 1 0;  
     0 0 1; 1 0 1; 1 1 1; 0 1 1;  
     0 0 2; 1 0 2; 1 1 2; 0 1 2]*0.01;
```

```

%
% Element connectivity
LE=[1 2 3 4 5 6 7 8;
   5 6 7 8 9 10 11 12];
%
% External forces [Node, DOF, Value]
EXTFORCE=[9 3 10.0E3; 10 3 10.0E3; 11 3 10.0E3; 12 3 10.0E3];
%
% Prescribed displacements [Node, DOF, Value]
SDISPT=[1 1 0;1 2 0;1 3 0;2 2 0;2 3 0;3 3 0;4 1 0;4 3 0];
%
% Load increments [Start End Increment InitialFactor FinalFactor]
TIMS=[0.0 0.8 0.4 0.0 0.8; 0.8 1.1 0.1 0.8 1.1]';
%
% Material properties PROP=[LAMDA MU BETA H Y0]
MID=1;
PROP=[110.747E9 80.1938E9 0.0 1.E8 4.0E8];
%
% Set program parameters
ITRA=70; ATOL=1.0E5; NTOL=6; TOL=1E-6;
%
% Calling main function
NOUT = fopen('output.txt','w');
NLFEA(ITRA, TOL, ATOL, NTOL, TIMS, NOUT, MID, PROP, EXTFORCE, SDISPT, XYZ, LE);
fclose(NOUT);

```

The Newton–Raphson iteration will be performed at each load increment of each load step. The following output shows the iteration history:

Time	Time step	Iter	Residual
0.40000	4.000e-01	2	3.80851e-12
Time	Time step	Iter	Residual
0.80000	4.000e-01	2	4.32010e-12
Time	Time step	Iter	Residual
0.90000	1.000e-01	2	3.97904e-12
Time	Time step	Iter	Residual
1.00000	1.000e-01	2	3.63798e-12
Time	Time step	Iter	Residual
1.10000	1.000e-01	2	6.66390e+02
		3	1.67060e-09

During the first load step, since the stress is less than the yield strength, the material behavior is identical to a linear elastic material. Therefore, the Newton–Raphson iteration converges at the first iteration at Time = 0.4 and 0.8. During the

Table 2.6 Displacements and stresses of elastoplastic bar

Load factor	u_{5z}	u_{0z}	S_{33} Elem1 (MPa)	S_{33} Elem2 (MPa)	State
0.4	7.73×10^{-6}	1.55×10^{-5}	160	160	Elastic
0.8	1.55×10^{-5}	3.09×10^{-5}	320	320	Elastic
0.9	1.74×10^{-5}	3.48×10^{-5}	360	360	Elastic
1.0	1.93×10^{-5}	3.87×10^{-5}	400	400	Elastic
1.1	4.02×10^{-3}	8.04×10^{-3}	440	440	Plastic

second load step, the material is still elastic at Time = 0.9 and 1.0, where they also converge at the first iteration. At Time = 1.1, however, the material becomes plastic and the iteration converges at the second iteration. The iteration converges very fast in this case because both elements are under the uniform stress condition and the plastic hardening model is linear.

The nodal displacements and element stresses are saved in output.txt file. These results are stored at every load increment. Table 2.6 summarizes displacements and stresses at each load increment. Since the two elements are in the same loading condition, it is expected that both elements have the same constant stresses. Also, the material is elastic until the load factor = 1.0, at which the material is in the initial yield state. At load factor = 1.1, the material deforms plastically until it can support the stress of 440 MPa. The stress after yielding continuously increases due to plastic hardening modulus, H , albeit the slope is much smaller than the initial stress-strain curve. That is why the displacements dramatically increase between load factors 1.0 and 1.1. ■

2.5 Nonlinear Solution Controls Using Commercial Finite Element Programs

Although many commercial finite element analysis programs are available, only three popular programs (Abaqus, ANSYS, and NEiNastran) are discussed in this section. All three programs provide both a graphical user interface (GUI) and text input file for defining the solution controls. Figure 2.28 shows the steps of defining nonlinearity in the flow of nonlinear modeling and analysis using commercial software. It is obvious that material nonlinearity is defined at the stage of defining material properties and force nonlinearity at load conditions. Kinematic nonlinearity, such as contact conditions, is defined at the stage of defining displacement boundary conditions. However, many commercial programs consider contact conditions separately from displacement boundary conditions, and they can be defined separately. It is difficult to see how to define geometric nonlinearities at the stage of defining a load case, but it is common for users to select the large deformation option when a load case is defined. The detailed description of these individual nonlinearities will be discussed in the following chapters.

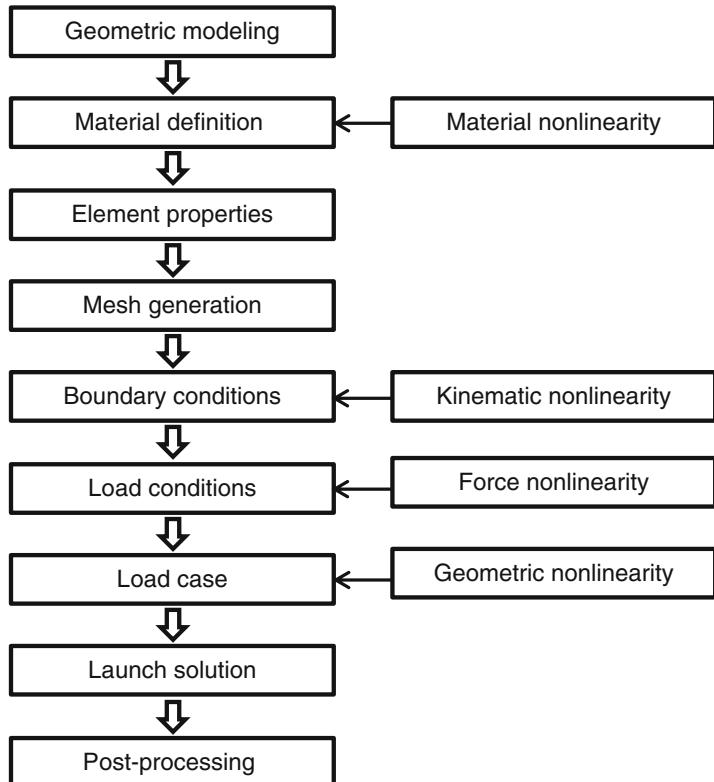


Fig. 2.28 Nonlinearity definition in the analysis flow

Since the input data through the GUI are eventually converted into a text input file, the latter is used in the following explanations. The purpose is not to provide complete usage instructions for the programs. Rather, it focuses on how to apply the solution control methods in the previous sections. For detailed usage, the readers are referred to the manual of the particular program.

2.5.1 Abaqus

The input file of Abaqus consists of keywords. For example, the keyword, *NODE, defines nodal coordinates, and *ELEMENT defines element type and connectivity. The keyword, *STEP, is used for the solution control of the current load. The definition of *STEP ends with the *END STEP keyword. In *STEP, users can specify analysis type, boundary conditions, applied loads, and output controls. This is different from the time steps that are used in this textbook. It is similar to the load case in Nastran.

```
*STEP, INC = 100 (default)
TITLE
```

This keyword starts a new step. By default, the maximum allowable number of increments is 100. The *STEP keyword is followed by the keyword defining the analysis procedure.

```
*STATIC, DIRECT
DT, TEND, DTMIN, DTMAX
```

The *STATIC keyword specifies that the analysis type is nonlinear static. If the parameter, DIRECT, is given, then a fixed time increment, DT, is used until the ending time, TEND. If this parameter is omitted, automatic time stepping will be used starting from the initial time increment DT. When the second line is omitted, DT and TEND will be set to one, i.e., a single load increment. When the automatic time stepping scheme is used, the time increment can reduce until DTMIN and can increase up to DTMAX.

```
*END STEP
```

This command finishes defining the current analysis step. Abaqus can apply multiple loads in sequence by defining the corresponding steps sequentially.

2.5.2 ANSYS

The input file of ANSYS consists of three sections for static analysis: preprocessing (/PREP7), solution (/SOLU), and postprocessing (/POST1). The solution controls are defined in the /SOLU phase.

```
SOLCONTROL, ON(default)/OFF
```

This command activates (ON) or deactivates (OFF) optimized defaults for a set of commands applicable to nonlinear solutions. It is recommended to use the default value (ON) for reliable and efficient default solution settings.

```
AUTOTS, ON(default)/OFF
```

This command determines if the time step is determined automatically by ANSYS, or a fixed time step that is given in DETIM or NSUBST command is used.

Default: ANSYS determines time stepping when SOLCONTROL is ON. No automatic time stepping occurs when SOLCONTROL is OFF.

```
TIME, TIME
```

This command specifies the time at the end of the current load. The starting time is either zero (for the first load) or the end time of the previous load. Since it is not a

physical time for static analysis, it is recommended that each load has a unit time; i.e., the first load has TIME = 1, and the second load has TIME = 2.

DELTIM, DTIME, DTMIN, DTMAX

This command specifies the load increments (time steps) for the current load. If both AUTOTS and SOLCONTROL are OFF, the fixed time step size DTIME is used. The advantage is that the users can obtain the analysis results at the specific load increments. However, the analysis may stop due to non-convergence if the nonlinearity is high at any load increment section.

If AUTOTS is ON, the program uses DTIME as an initial time step, and the following time steps are calculated according to automatic time stepping procedure, which can reduce possible non-convergence problems. When SOLCONTROL is ON, the program automatically sets up the minimum and maximum size of the time steps. When it is OFF, users need to provide these values in DTMIN and DTMAX.

NSUBST, NSBstp, NSBMX, NSBMN

This command plays the same role as the DELTIM command. Instead of specifying the time step size, it provides the number of load increments in NSBstp. If TIME = 1, then the corresponding time step size can be calculated from 1/NSBstp. The same time stepping procedure is applied when AUTOTS = ON and/or SOLCONTROL = ON. NSBMX and NSBMN correspond to DTMAX and DTMIN.

NEQIT, NEQIT

Specifies the maximum number of equilibrium iterations for nonlinear analyses. If the number of iterations becomes NEQIT, the program will either stop with an error message or cutback time step if AUTOTS is ON.

CNVTOL, Lab, VALUE, TOLER, NORM, MINREF

This command specifies convergence criteria for nonlinear analyses. For structural problems, Lab = U (displacements) or F (forces) are frequently used. VALUE is a typical value of displacements or forces, and TOLER is the tolerance to consider the nonlinear iteration convergence. The default value of TOLER is 0.05 (5 %) for Lab = U, and 0.005 (0.5 %) for Lab = F. Since the displacements and forces are vectors, their magnitudes are calculated using NORM and compared with TOLER. The square root of the square sum (L_2 -norm) is used when NORM = 2 (default), while the sum of absolute values is used when NORM = 1 (L_1 -norm). ANSYS monitors this convergence criterion and plots a graph during each iteration.

SOLVE

Once all solution controls are set, this command starts solving the system of nonlinear equations.

2.5.3 NEiNastran

There are several different variations of NASTRAN programs. In this textbook, NEiNastran from NEi software is used. The model input file of NASTRAN consists of two sections: case control and bulk data. The input file starts with the case control section, which ends with the BEGIN BULK entry. The bulk data section starts with the BEGIN BULK entry and ends with the ENDDATA entry. Definitions of nodes and elements, boundary conditions, applied loads, and material properties are given in the bulk data section. The case control section specifies how the loads and boundary conditions will be used for a particular load case. It also controls outputs. The bulk data entry is given in a fixed column format, where each parameter is specified in 8 columns.

SOL NLSTATIC

This command specifies that the solution type is nonlinear static. This is in the case control section and usually the first entry in the input file.

NLPARM = 1

This case control entry specifies that the NLPARM entry in the bulk data section is used for the current load. It is possible that the users can define multiple NLPARM entries with different identification numbers and use a particular one for the current load.

BEGIN BULK

This entry shows that the case control section is over and the bulk data section starts.

NLPARAM

This entry controls the solution procedure of nonlinear analysis. The parameters of the NLPARM entry are as follows:

NLPARM	ID	NINC		KMETHOD	KSTEP	MAXITER	CONV	INTOUT	
	EPSU	EPSP	EPSW						

ID is a unique identification number that can be activated by the NLPARM = ID entry in the case control section. The entire load is divided by NINC increments. The parameter, KMETHOD (AUTO, SEMI, ITER), determines how often the stiffness matrix should be calculated in the modified Newton–Raphson method. AUTO: the program decides when to calculate stiffness matrix, SEMI: the stiffness matrix is always calculated at the first iteration and follows the same procedure as AUTO, and ITER: calculate the stiffness matrix at every KSTEP number of

iterations. In each load increment, the number of iterations is limited by MAXITER. The convergence criteria are specified in the CONV parameter using any combination of U (displacement), P (force), and W (work). The tolerances of these criteria are given in EPSU, EPSP, and EPSW. If INTOUT is YES, the program will output intermediate results.

ENDDATA

This entry shows that the bulk data section is over. This is usually the last entry of the input file.

2.6 Summary

In this chapter, different types of nonlinearities in solid mechanics are introduced, including geometric, material, kinematic, and force nonlinearities. Geometric nonlinearity usually occurs when deformation is large, so that the undeformed and deformed states are significantly different. Material nonlinearity occurs in the constitutive relation, i.e., stress–strain relation. Hyperelastic, elastoplastic, or viscoelastic materials are examples of nonlinear materials. Kinematic nonlinearity usually occurs on the boundary of a structure by constraining deformation. The contact constraint is the most common example of kinematic nonlinearity. Force nonlinearity occurs when the applied force depends on deformation. A common example is when a pressure load is applied to a surface that undergoes a large deformation.

General procedures for solving a system of nonlinear equations are introduced in Sect. 2.2. The key concept is to reduce the magnitude of the residuals at every iteration using the Jacobian matrix. The Newton–Raphson, modified Newton–Raphson, and secant and incremental force methods are introduced. The Newton–Raphson method shows a quadratic convergence when the initial estimate is close to the solution. The convergence rates of all other methods are between one and two. However, the Newton–Raphson method has the highest computational costs because it calculates the Jacobian matrix at every iteration. The Newton–Raphson method may have difficulty in convergence, or the solution may diverge if the initial estimate is too far from the solution. In order to improve the convergence, the incremental force method gradually increases the applied force and uses the previously converged solution as an initial estimate for the following load increment. In commercial finite element programs, this incremental force method is further improved by automatically adjusting the size of load increments such that the Newton–Raphson iteration converges quickly. It is also discussed that the displacement-controlled method is more stable than the force-controlled method for nonlinear structural systems. The solution control commands for three different commercial finite element programs are briefly introduced in Sect. 2.4.

2.7 Exercises

- P2.1. Find the roots of the following nonlinear vector equations using the Newton–Raphson method:

$$\mathbf{P}(\mathbf{u}) \equiv \begin{Bmatrix} u_1 + u_2 \\ u_1^2 + u_2^2 \end{Bmatrix} = \begin{Bmatrix} 3 \\ 9 \end{Bmatrix} \equiv \mathbf{f}.$$

Use the initial estimate $\mathbf{u}^0 = \{1, 5\}^T$ and convergence tolerance = 10^{-5} . Discuss the convergence rate.

- P2.2. Using the modified Newton–Raphson method, solve the nonlinear equations in P2.1. Compare the convergence rate with the Newton–Raphson method.
- P2.3. Using the Broyden's method, solve the nonlinear equations in P2.1. Compare the convergence rate with the Newton–Raphson method.
- P2.4. Using the incremental force method, solve the equations in P2.1. Use five equal-interval load steps.
- P2.5. Consider a uniform bar with a constant Young's modulus, $E = 100$ MPa; cross-sectional area, $A = 2 \times 10^{-4}$ m²; and a unit length, $L = 1$ m. The applied force $F = 10$ kN is large enough such that the relation between displacement and strain is nonlinear:

$$\epsilon(u) = \frac{du}{dx} + \frac{1}{2} \left(\frac{du}{dx} \right)^2.$$

Using a single two-node bar element, calculate the displacement at the tip and strain of the element. Use an increment force method with 10 equal force increments.

Hint: The virtual strain can be obtained through variation of the strain as

$$\epsilon(\bar{u}) = \frac{d\bar{u}}{dx} + \frac{du}{dx} \frac{d\bar{u}}{dx}.$$

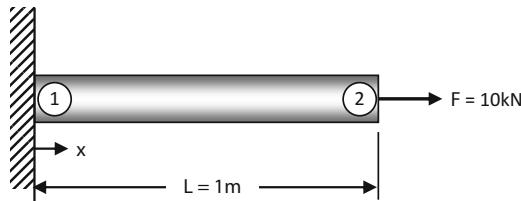


Fig. P2.5

P2.6. Solve Problem P2.5 using the secant method. Do not use the incremental force method. Discuss about the convergence rate.

P2.7. Consider a uniform bar with a cross-sectional area, $A = 2 \times 10^{-4} \text{ m}^2$, and a unit length, $L = 1 \text{ m}$. The bar shows elastoplastic material behavior, as depicted in the figure. The plastic deformation starts at yield stress $\sigma_Y = 400 \text{ MPa}$. In the elastic region, the Young's modulus is $E = 200 \text{ GPa}$, while in the plastic region, the tangent stiffness is $E_T = 20 \text{ GPa}$. When a force, $F = 50 \text{ kN}$, is applied at the end, calculate the tip displacement and stress of the element using one bar element. Use 10 equal-interval force increments. Plot the force-displacement curve. Assume the displacement-strain relation is linear.

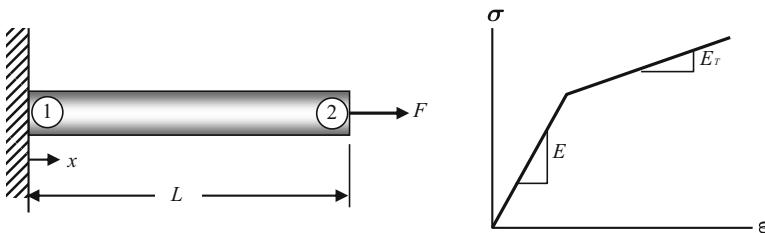


Fig. P2.7

P2.8. Consider the three nonlinear springs in the figure. The stiffness of each spring is given by $k_1 = 500 + 50u$, $k_2 = 200 + 100u$, and $k_3 = 500 + 100u$, where u is the elongation of the spring. Solve the displacements at nodes 1 and 2 using the Newton-Raphson method when $F = 100$.

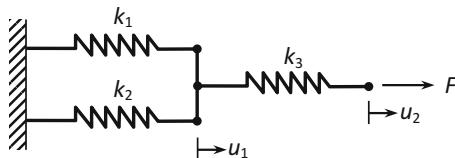


Fig. P2.8

P2.9. Consider a uniform bar in the figure. The stress-strain relation and displacement-strain relation are linear. However, the Young's modulus of the material varies according to the strain.

$$\sigma = E(u)\varepsilon(u), \quad \varepsilon(u) = \frac{du}{dx}, \quad E(u) = E_0 \left(1 - \frac{du}{dx}\right).$$

When one element is used to model the bar, formulate the nonlinear equation with the tip displacement being unknown. Solve the tip displacement using the incremental force method with 10 equal-interval increments. Use

$E_0 = 1.0 \text{ GPa}$, $A = 10^{-4} \text{ m}^2$, and $F = 25 \text{ kN}$. Plot the force–displacement curve. Test what happens when $F = 30 \text{ kN}$, and explain why.

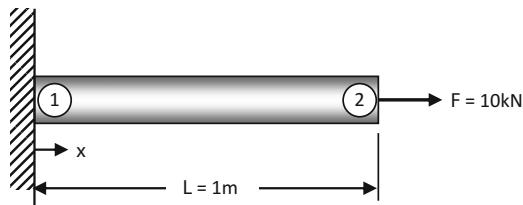


Fig. P2.9

References

1. Press WH et al. Numerical recipes. The art of scientific computing. 3rd ed. New York: Cambridge University Press; 2007.
2. Broyden CG. A class of methods for solving nonlinear simultaneous equations. *Math Comp*. 1965;19(92):577–93.
3. Broyden CG. The convergence of a class of double-rank minimization algorithms. *IMA J Appl Math*. 1970;6:76–90.

Chapter 3

Finite Element Analysis for Nonlinear Elastic Systems

3.1 Introduction

In the previous chapter, different nonlinearities are briefly discussed along with solution procedures. In this and the following chapters, specific nonlinearities will be discussed in detail. In general, nonlinear systems in solid mechanics can be categorized by mild and rough nonlinearities. **Mild nonlinearity** has smooth, path-independent nonlinear relations between stress and strain. Nonlinear elasticity, geometric nonlinearity, and deformation-dependent loads belong to this category. On the other hand, **rough nonlinearity** includes equality and/or inequality constraints in the constitutive relation or kinematic conditions. Elastoplasticity and contact problems belong to this category. In this chapter, finite element formulations for mild nonlinear systems are developed. Rough nonlinearity will be discussed in the following chapters.

An important aspect of the problems in this chapter is that they undergo large deformation, which includes large strain, displacement, and rotation. In linear structural systems, it is assumed that the magnitude of deformation is infinitesimal such that there is no significant difference between the deformed and undeformed shapes. Thus, stress and strain are defined in the undeformed shape, and also the weak form is integrated over the same shape. **Under this assumption, the relation between displacement and strain becomes linear.** However, for large deformation, the difference between the deformed and undeformed shapes is large enough that they cannot be treated the same. Thus, the previous definitions of stress and strain should be modified from the assumption of small deformation. In addition, in linear structural systems, the relation between stress and strain is assumed linear when strain is small. However, this relation becomes nonlinear as deformation increases.

Even though several methods for solving nonlinear equations were introduced in the previous chapter, most were based on the Newton–Raphson method, except for the secant method. These solution procedures require calculating the residual and tangent stiffness matrix at each iteration. The residual is calculated from the

weak form of structural systems, and the tangent stiffness matrix is calculated by differentiating it. This chapter will focus on how to calculate this residual and tangent stiffness matrix for a given nonlinear elasticity model. Since the main purpose of this chapter is to introduce the basic procedure of nonlinear elastic structural systems, only parts of nonlinear elastic and hyperelastic models will be introduced in this book. More detailed discussions on the topic can be found in the books by Belytschko et al. [1] and Wriggers [2].

In general, the structural equilibrium equation—the weak form—is written based on a frame of reference. The structural geometry at any stage of deformation can be a frame of reference, but due to simplicity, either the initial, undeformed geometry or current deformed geometry is often used as a frame of reference. In addition, proper definitions of stress and strain must be used according to the frame of reference. When the equilibrium equation is written at the undeformed geometry, it is called the total Lagrangian formulation. On the other hand, when it is written at the deformed geometry, it is called the updated Lagrangian formulation. Although these two formulations refer to different frames of reference, they represent the same structural equilibrium. Thus, these two formulations are theoretically identical [3]. However, the numerical implementation of these two formulations becomes different as they use different measures of stress and strain as well as different integration domains.

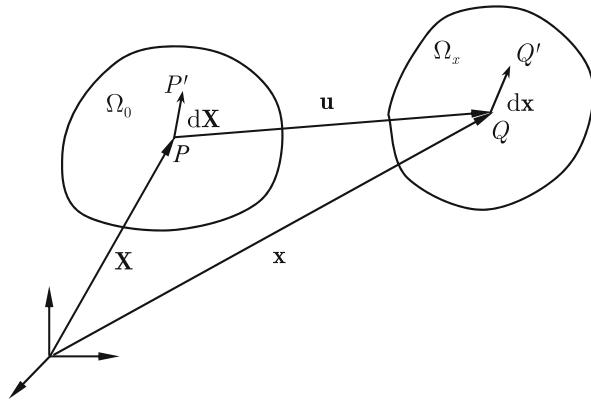
3.2 Stress and Strain Measures in Large Deformation

When deformation is infinitesimal, there is no noticeable difference between the undeformed and deformed geometries. All quantities such as stresses, strains, and displacements are referred to at the original undeformed geometry. However, in reality, the structure is in equilibrium after deformation. Thus, to be precise, the equilibrium equation must be written at the deformed geometry, but this difference is ignored by the infinitesimal deformation assumption. In the case of large deformation, however, the difference between undeformed and deformed geometries is significant, and thus, is not ignorable. Thus, it is important to understand how to represent a large deformation of a material and how to define stress and strain in such a case. The behavior of material under deformation is studied in continuum mechanics. This section presents a brief review of important continuum mechanics concepts related to large displacement problems. For detailed discussions on continuum mechanics, the readers are referred to the book by Malvern [4].

3.2.1 Deformation Gradient

Consider a general solid that is subjected to some forces and displacements so that its geometry changes from the initial (or undeformed) to the current (or deformed) state as shown in Fig. 3.1. This deformation is denoted by a mapping $\Phi(\mathbf{X}, t)$ where $\mathbf{X} = [X_1, X_2, X_3]$ is a material point in the undeformed geometry and t denotes the

Fig. 3.1 Undeformed (Ω_0) and deformed (Ω_x) geometries of a body



deformation process, with $t = 0$ being the undeformed geometry. In continuum, this mapping is one-to-one, and Φ and Φ^{-1} are continuously differentiable. The points in the initial geometry are denoted by vectors with uppercase letters and those in the current geometry by lowercase letters. In particular consider a point P , identified by vector \mathbf{X} , in the initial undeformed geometry that is mapped to a point Q , identified by vector $\mathbf{x} = [x_1, x_2, x_3]$, in the current deformed geometry. The description of the mapping from initial to current geometry is symbolically written as follows:

$$\begin{aligned} x_1 &= x_1(X_1, X_2, X_3) \\ x_2 &= x_2(X_1, X_2, X_3), \\ x_3 &= x_3(X_1, X_2, X_3) \end{aligned}$$

or

$$\mathbf{x} = \Phi(\mathbf{X}, t). \quad (3.1)$$

The above equation says that for a given point P in the undeformed geometry, a unique point Q exists in the deformed geometry. Referring to Fig. 3.1, the above mapping relation can be written as

$$\mathbf{x} = \Phi(\mathbf{X}, t) = \mathbf{X} + \mathbf{u}(\mathbf{X}, t), \quad (3.2)$$

where $\mathbf{u}(\mathbf{X}, t)$ is the displacement of point P .

In Fig. 3.1, neighboring points P' and Q' at infinitesimal distances from P and Q are denoted by vectors $d\mathbf{X}$ and $d\mathbf{x}$, respectively, in the two geometries. The vector $d\mathbf{X}$ deforms to $d\mathbf{x}$. Assuming continuous mapping, the relationship between differential elements $d\mathbf{X}$ and $d\mathbf{x}$ can be expressed as follows:

$$d\mathbf{x} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} d\mathbf{X} \Rightarrow d\mathbf{x} = \mathbf{F} d\mathbf{X}, \quad (3.3)$$

where \mathbf{F} is known as the deformation gradient and is written explicitly as follows:

$$F_{ij} = \frac{\partial x_i}{\partial X_j}. \quad (3.4)$$

Using the relation in Eq. (3.2), the deformation gradient can be written as

$$\mathbf{F} = \mathbf{1} + \frac{\partial \mathbf{u}}{\partial \mathbf{X}} = \mathbf{1} + \nabla_0 \mathbf{u}, \quad (3.5)$$

where the term $\partial \mathbf{u} / \partial \mathbf{X}$ is called the displacement gradient. For the notational simplicity, $\nabla_0 = \partial / \partial \mathbf{X}$ represents the **gradient operator at the undeformed geometry**, such that $(\nabla_0 \mathbf{u})_{ij} = \partial u_i / \partial X_j$. If $\mathbf{F} = \mathbf{1}$, then $d\mathbf{x} = d\mathbf{X}$, which means that there is no deformation. Even if an infinitesimal volume in the undeformed geometry can increase or decrease its size, it cannot shrink to a point, i.e., a zero volume. Mathematically, this means that the **determinant of deformation gradient** must be positive:

$$\det \mathbf{F} \equiv J > 0. \quad (3.6)$$

This property is important to make a valid mapping of $\Phi(\mathbf{X}, t)$ during large deformation.

Example 3.1 (Uniform extension) Consider a cube undergoing uniform extensions in all three directions, so that

$$x_1 = \lambda_1 X_1, \quad x_2 = \lambda_2 X_2, \quad x_3 = \lambda_3 X_3, \quad (3.7)$$

where λ_i are the principal stretches. What is the condition of λ s to be a valid mapping? Calculate the deformation gradient \mathbf{F} . In addition, calculate the deformed volume of an infinitesimal cube that has an initial volume of $dV_0 = dX_1 dX_2 dX_3$. What is the requirement of preserving the volume?

Solution From the **requirement of continuity** in mapping, all stretches must be positive, i.e., $\lambda_i > 0$. If λ_i are constants or functions of time t only, then the deformation is uniform. The deformation gradient may be found from Eq. (3.4) as

$$\mathbf{F} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

If $\lambda_1 = \lambda_2 = \lambda_3$, the cube undergoes a uniform expansion or contraction in all directions and is referred to as a uniform dilatation. The volume of the cube is initially $dV_0 = dX_1 dX_2 dX_3$ and is now $dV_x = dx_1 dx_2 dx_3 = \lambda_1 \lambda_2 \lambda_3 dX_1 dX_2 dX_3 = \lambda_1 \lambda_2 \lambda_3 dV_0$. When there is no volume change, $\lambda_1 \lambda_2 \lambda_3 = 1$. ■

3.2.2 Lagrangian and Eulerian Strains

Since undeformed and deformed geometries are different, the engineering strain from infinitesimal deformation cannot be used for large deformation. In addition, since the definition of strain includes derivative of displacement with respect to the coordinates of reference frame, either undeformed or deformed geometry must be used as a reference. Two different definitions of strain will be discussed in the following.

3.2.2.1 Lagrangian Strain

Lagrangian strain uses undeformed geometry as a reference. Consider the two differential elements, $d\mathbf{x}$ and $d\mathbf{X}$, in Fig. 3.1. The vector $d\mathbf{X}$ is deformed to $d\mathbf{x}$. The change in squares of length of these two vectors can be expressed as follows:

$$\begin{aligned}\|d\mathbf{x}\|^2 - \|d\mathbf{X}\|^2 &= d\mathbf{x}^T d\mathbf{x} - d\mathbf{X}^T d\mathbf{X} \\ &= d\mathbf{X}^T \mathbf{F}^T d\mathbf{X} - d\mathbf{X}^T d\mathbf{X} \\ &= d\mathbf{X}^T (\mathbf{F}^T \mathbf{F} - \mathbf{1}) d\mathbf{X},\end{aligned}\quad (3.8)$$

where $\mathbf{1}$ is a 3×3 identity tensor. Since $d\mathbf{X}^T d\mathbf{X}$ is the square of the length of vector $d\mathbf{X}$, the quantity in the parentheses, $(\mathbf{F}^T \mathbf{F} - \mathbf{1})$, measures the change in squared lengths with respect to the square of the initial length. The term $\mathbf{F}^T \mathbf{F}$ is an important quantity and is defined as a **right Cauchy–Green deformation tensor**:

$$\mathbf{C} = \mathbf{F}^T \mathbf{F}. \quad (3.9)$$

From the last relation in Eq. (3.8), the Lagrangian strain can be defined as

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{1}). \quad (3.10)$$

The factor $1/2$ is used to make the definition identical to the engineering strains in case of infinitesimal strains. When there is no deformation, $\mathbf{F} = \mathbf{1}$ and thus, $\mathbf{E} = \mathbf{0}$. In terms of displacement gradient, the Lagrangian strain tensor can be written as

$$\mathbf{E} = \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{X}} + \frac{\partial \mathbf{u}^T}{\partial \mathbf{X}} + \frac{\partial \mathbf{u}^T}{\partial \mathbf{X}} \frac{\partial \mathbf{u}}{\partial \mathbf{X}} \right) = \frac{1}{2} (\nabla_0 \mathbf{u} + \nabla_0 \mathbf{u}^T + \nabla_0 \mathbf{u}^T \nabla_0 \mathbf{u}). \quad (3.11)$$

It is obvious from the definition that the Lagrangian strain \mathbf{E} is symmetric. In addition, when the displacement gradient is small, then it approaches the following **infinitesimal strain tensor**:

$$\boldsymbol{\epsilon} = \frac{1}{2} (\nabla_0 \mathbf{u} + \nabla_0 \mathbf{u}^T). \quad (3.12)$$

However, physically, $\boldsymbol{\epsilon}$ cannot be an exact measure of deformation because it does not remain constant in rigid-body rotations.

Example 3.2 (Strains in a rigid-body rotation) In a counterclockwise rigid-body rotation through an angle α about the X_3 -axis, the transformation relation can be given by

$$\begin{aligned}x_1 &= X_1 \cos \alpha - X_2 \sin \alpha \\x_2 &= X_1 \sin \alpha + X_2 \cos \alpha, \\x_3 &= X_3\end{aligned}\quad (3.13)$$

Compare the infinitesimal strain $\boldsymbol{\epsilon}$ with the Lagrangian strain \mathbf{E} .

Solution From the given transformation relation, the displacements can be calculated by

$$\begin{aligned}u_1 &= X_1 \cos \alpha - X_2 \sin \alpha - X_1 \\u_2 &= X_1 \sin \alpha + X_2 \cos \alpha - X_2, \\u_3 &= 0.\end{aligned}$$

The infinitesimal strain in Eq. (3.12) can be written as

$$\boldsymbol{\epsilon} = \frac{1}{2} \left[\frac{\partial u_i}{\partial X_j} + \frac{\partial u_j}{\partial X_i} \right] = \begin{bmatrix} \cos \alpha - 1 & 0 & 0 \\ 0 & \cos \alpha - 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In order to calculate the Lagrangian strain, the deformation gradient is calculated first by differentiating the transformation relation as

$$\mathbf{F} = \begin{bmatrix} \frac{\partial x_i}{\partial X_j} \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha & 0 \\ -\sin \alpha & \cos \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The Lagrangian strain in Eq. (3.10) can be written as

$$\mathbf{E} = \frac{1}{2} (\mathbf{F}^T \mathbf{F} - \mathbf{I}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Note that the Lagrangian strain is not affected by the rigid-body rotation, but the infinitesimal strain varies. ■

In the above example, ϵ_{11} and ϵ_{22} are not zero. However, if α is small, then these quantities are small and may be neglected. Although the infinitesimal strain tensor is not an exact measure of deformation, it is convenient for use in the applications involving small strains. An advantage of this strain tensor is its linear relation with

respect to the displacement gradient. This allows for the application of the techniques of linear analysis in solving boundary-value problems in the linear theory of elasticity and helps keep the equations of the theory of plasticity simple. Nevertheless, using Eq. (3.12), it is necessary to keep in mind that the rigid-body rotation has to be small.

3.2.2.2 Eulerian Strain

In Eq. (3.8), the change in length is expressed with respect to the undeformed differential element $d\mathbf{X}$. If the deformed differential element $d\mathbf{x}$ is used as a reference, the change in squares of length of these two vectors can be expressed as follows:

$$\begin{aligned}\|\mathbf{dx}\|^2 - \|\mathbf{dX}\|^2 &= \mathbf{dx}^T \mathbf{dx} - \mathbf{dX}^T \mathbf{dX} \\ &= \mathbf{dx}^T \mathbf{dx} - \mathbf{dx}^T \mathbf{F}^{-T} \mathbf{F}^{-1} \mathbf{dx} \\ &= \mathbf{dx}^T (\mathbf{1} - \mathbf{F}^{-T} \mathbf{F}^{-1}) \mathbf{dx} \\ &= \mathbf{dx}^T (\mathbf{1} - \mathbf{b}^{-1}) \mathbf{dx},\end{aligned}\tag{3.14}$$

where the **left Cauchy–Green deformation tensor** \mathbf{b} is defined as

$$\mathbf{b} = \mathbf{FF}^T.\tag{3.15}$$

Using the left Cauchy–Green deformation tensor, the Eulerian strain tensor can be defined as

$$\mathbf{e} = \frac{1}{2}(\mathbf{1} - \mathbf{b}^{-1}).\tag{3.16}$$

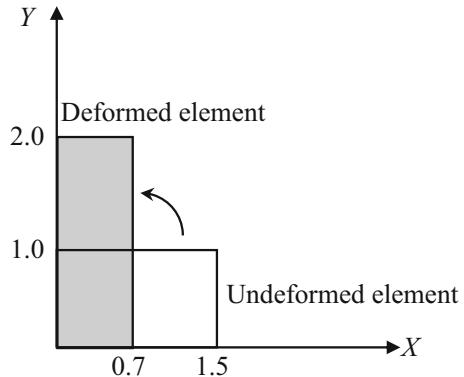
From a similar approach, it can be shown that

$$\mathbf{e} = \frac{1}{2} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{u}^T}{\partial \mathbf{x}} - \frac{\partial \mathbf{u}^T}{\partial \mathbf{x}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \right) = \frac{1}{2} (\nabla_x \mathbf{u} + \nabla_x \mathbf{u}^T - \nabla_x \mathbf{u}^T \nabla_x \mathbf{u}),\tag{3.17}$$

where $\nabla_x = \partial/\partial \mathbf{x}$ represents the **gradient operator at the current geometry**, such that $(\nabla_x \mathbf{u})_{ij} = \partial u_i / \partial x_j$. As with the Lagrangian strain, the Eulerian strain is also symmetric and **approaches the infinitesimal strain** when the displacement gradient is small.

Example 3.3 (Large displacement and rotation) A four-node element undergoes large displacement and rotation in the XY plane, as shown in Fig. 3.2. The element is rotated counterclockwise by 90° , its length is stretched to 2, and width is reduced to 0.7. Calculate the deformation gradient, Lagrangian strain, Eulerian strain, and engineering strain.

Fig. 3.2 Finite element under large deformation and rotation



Solution Using the bilinear shape function in the reference coordinate (s,t) , the mapping relation for the undeformed element can be written as

$$\begin{cases} X = \sum_{I=1}^4 N_I(s,t)X_I = \frac{3}{4}(s+1) \\ Y = \sum_{I=1}^4 N_I(s,t)Y_I = \frac{1}{2}(t+1) \end{cases},$$

and the mapping relation for the deformed element can be written as

$$\begin{cases} x(s,t) = \sum_{I=1}^4 N_I(s,t)x_I = 0.35(1-t) \\ y(s,t) = \sum_{I=1}^4 N_I(s,t)y_I = s+1 \end{cases}.$$

For convenience of notation, let us define the following vectors: $\mathbf{X} = \{X, Y\}^T$, $\mathbf{x} = \{x, y\}^T$, $\mathbf{s} = \{s, t\}^T$. Then, using the chain rule of differentiation, we can calculate the deformation gradient as

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \frac{\partial \mathbf{x}}{\partial \mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{X}} = \begin{bmatrix} 0 & -0.35 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 4/3 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 0 & -0.7 \\ 4/3 & 0 \end{bmatrix}.$$

It can be easily verified that the above deformation gradient transforms a vector $\{1.5, 0\}^T$ into $\{0, 2\}^T$ and $\{0, 1\}^T$ into $\{-0.7, 0\}^T$.

From the definition of the Lagrangian strain in Eq. (3.10), \mathbf{E} can be defined as

$$\mathbf{E} = \frac{1}{2}(\mathbf{F}^T \mathbf{F} - \mathbf{1}) = \begin{bmatrix} 0.389 & 0 \\ 0 & -0.255 \end{bmatrix}.$$

Since the Lagrangian strain has only nonzero diagonal components, the original rectangular shape is maintained. However, the length in the original X -direction increases, whereas it decreases in the original Y -direction.

The Eulerian strain can be calculated from Eq. (3.16):

$$\boldsymbol{\epsilon} = \frac{1}{2}(\mathbf{1} - \mathbf{F}^{-T}\mathbf{F}^{-1}) = \begin{bmatrix} -0.52 & 0 \\ 0 & 0.22 \end{bmatrix}.$$

Note that since the Eulerian strain also has nonzero diagonal components, the original rectangular shape maintains. However, since the frame of reference is different from that of the Lagrangian strain, the magnitudes of diagonal components are different.

In order to calculate engineering strain, it is first necessary to calculate the displacements at each node. By subtracting the nodal coordinates between the deformed and undeformed geometries, the following nodal displacement can be obtained:

$$\begin{aligned} u_1 &= 0.7 & u_2 &= -0.8 & u_3 &= -1.5 & u_4 &= 0.0 \\ v_1 &= 0.0 & v_2 &= 2.0 & v_3 &= 1.0 & v_4 &= -1.0 \end{aligned}.$$

Then, the displacements of the element can be calculated using the shape functions as

$$\begin{aligned} u &= \sum_{I=1}^4 N_I u_I = \frac{1}{4}(-1.6 - 3s - 1.4t) \\ v &= \sum_{I=1}^4 N_I v_I = \frac{1}{4}(2 + 4s - 2t). \end{aligned}$$

The displacement gradient can be calculated from the chain rule of differentiation as

$$\nabla_0 \mathbf{u} = \frac{\partial \mathbf{u}}{\partial \mathbf{s}} \frac{\partial \mathbf{s}}{\partial \mathbf{X}} = \frac{1}{4} \begin{bmatrix} -3 & -1.4 \\ 4 & -2 \end{bmatrix} \begin{bmatrix} \frac{4}{3} & 0 \\ 0 & 2 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} -4 & -2.8 \\ \frac{16}{3} & -4 \end{bmatrix}.$$

The engineering strain can be calculated from the definition in Eq. (3.12)

$$\boldsymbol{\epsilon} = \frac{1}{2}(\nabla_0 \mathbf{u} + \nabla_0 \mathbf{u}^T) = \begin{bmatrix} -1 & 0.32 \\ 0.32 & -1 \end{bmatrix}.$$

Since the engineering strain cannot handle rigid-body rotation, the **shear strain term exists even if the deformed shape remains a rectangle**. In addition, both normal components show compression, even though the actual deformation is tension in one side and compression in the other side. Thus, engineering strain is not appropriate for large deformation and rigid-body rotation. ■

3.2.3 Polar Decomposition

If the deformation gradient \mathbf{F} is nonsingular, there exists a unique orthogonal tensor \mathbf{Q} and unique positive-definite symmetric tensors \mathbf{U} and \mathbf{V} such that

$$\mathbf{F} = \mathbf{Q} \cdot \mathbf{U} = \mathbf{V} \cdot \mathbf{Q}, \quad (3.18)$$

where \mathbf{Q} is a rotation tensor (rigid-body rotation) and \mathbf{U} and \mathbf{V} are right and left stretch tensors, respectively. Note that \mathbf{U} and \mathbf{V} have the same eigenvalues (principal stretches) but different eigenvectors (principal axes of deformation). From continuum mechanics, it can be further verified that the eigenvectors of \mathbf{U} is the same with those of \mathbf{C} . In fact, the relationship between \mathbf{U} and \mathbf{C} is

$$\mathbf{U}^2 = \mathbf{C}, \quad \mathbf{U} = \sqrt{\mathbf{C}}. \quad (3.19)$$

This can be easily shown from the definition of $\mathbf{C} = \mathbf{F}^T \mathbf{F} = \mathbf{U} \mathbf{Q}^T \mathbf{Q} \mathbf{U} = \mathbf{U}^2$. Thus, \mathbf{U} will have the same eigenvectors with \mathbf{C} , and its eigenvalues are square roots of the eigenvalues of \mathbf{C} . In practice, \mathbf{U} can be calculated from eigenvectors and eigenvalues of \mathbf{C} . Let \mathbf{E}_1 , \mathbf{E}_2 , and \mathbf{E}_3 be eigenvectors and λ_1^2 , λ_2^2 , and λ_3^2 be eigenvalues of \mathbf{C} , respectively. Then, the following two matrices can be constructed:

$$\Phi = [\mathbf{E}_1 \quad \mathbf{E}_2 \quad \mathbf{E}_3], \quad \Lambda = \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix}. \quad (3.20)$$

From the spectral decomposition, the right Cauchy–Green deformation tensor can be written as $\mathbf{C} = \Phi \Lambda \Phi^T$, and the left stretch tensor can be calculated by

$$\mathbf{U} = \Phi \sqrt{\Lambda} \Phi^T, \quad \sqrt{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \quad (3.21)$$

If a new coordinate system is established using the three eigenvectors—the eigenvectors are mutually orthogonal—then Λ becomes the right Cauchy–Green deformation tensor in that coordinate system.

In order to explain the polar decomposition physically, let us consider a general deformation denoted by $d\mathbf{x} = \mathbf{F} \cdot d\mathbf{X} + \mathbf{c}$, where \mathbf{c} represents the rigid-body translation. We can decompose the deformation into

$$d\mathbf{x} = \mathbf{F} \cdot d\mathbf{X} + \mathbf{c} = \mathbf{Q} \cdot \mathbf{U} \cdot d\mathbf{X} + \mathbf{c} \quad (3.22)$$

Equation (3.22) shows that the current configuration can be obtained by (1) stretching in the principal directions, (2) rigid-body rotation, and then (3) translation. Whereas, if the left stretch tensor is used, \mathbf{dx} can be decomposed by

$$\mathbf{dx} = \mathbf{F} \cdot d\mathbf{X} + \mathbf{c} = \mathbf{V} \cdot \mathbf{Q} \cdot d\mathbf{X} + \mathbf{c} \quad (3.23)$$

Equation (3.23) shows that the deformed geometry can be obtained by (1) rigid-body rotation, (2) stretching in the principal directions, and then (3) translation.

The above decomposition of deformation can further be explained using eigenvectors and eigenvalues. Let \mathbf{E}_i with $i = 1, 2, 3$ be the three eigenvectors of \mathbf{U} and λ_i with $i = 1, 2, 3$ be the corresponding eigenvalues. In addition, let \mathbf{e}_i with $i = 1, 2, 3$ be the three eigenvectors of \mathbf{V} . The deformation process is illustrated in Fig. 3.3. The first path is to rotate the eigenvector \mathbf{E}_i to \mathbf{e}_i and then to stretch to λ_i in the direction of \mathbf{e}_i . This process is equivalent to Eq. (3.23) with $\mathbf{c} = 0$. The second path is to stretch to λ_i in the direction of \mathbf{E}_i and then to rotate the eigenvectors \mathbf{E}_i to \mathbf{e}_i . This process corresponds to Eq. (3.22) with $\mathbf{c} = 0$. It is clear that two process yields the same deformation.

Example 3.4 (Simple shear deformation) Consider a simple shear problem defined by

$$x_1 = X_1 + kX_2, \quad x_2 = X_2, \quad x_3 = X_3 \quad (3.24)$$

where $k = 2/\sqrt{3}$. (a) Find \mathbf{F} , \mathbf{U} , \mathbf{V} , and \mathbf{Q} . (b) Draw pictures showing the deformed states of the initially rectangular element at each stage of deformation. Compare the results obtained from the right stretch and left stretch tensors. (c) Study the deformation of a diagonal of the initially rectangular element using $X_2 = X_1 \tan \theta$

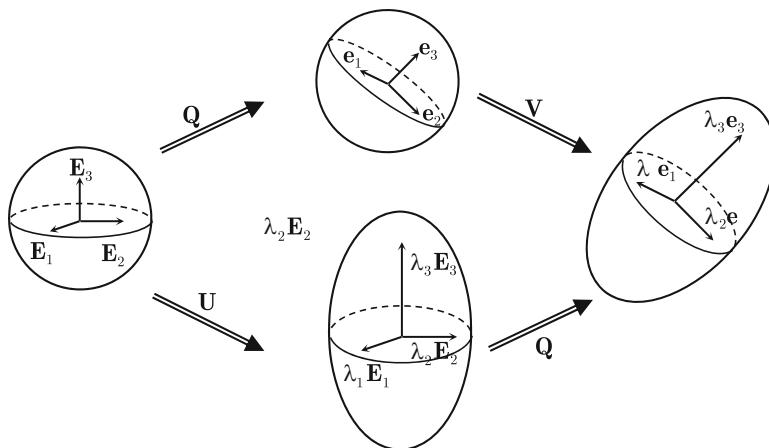


Fig. 3.3 Polar decomposition of deformation

with $\theta = 45^\circ$. (d) Study the deformation of a circle in the initially rectangular element using $X_1^2 + X_2^2 = a^2$.

Solution

- (a) For a given deformation in Eq. (3.24), the deformation gradient and the right Cauchy–Green deformation tensor can be calculated as

$$\mathbf{F} = \begin{bmatrix} 1 & \frac{2}{\sqrt{3}} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{C} = \mathbf{F}^T \mathbf{F} = \begin{bmatrix} 1 & \frac{2}{\sqrt{3}} & 0 \\ \frac{2}{\sqrt{3}} & \frac{7}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

After calculating eigenvalues and eigenvectors of \mathbf{C} , the two matrices in Eq. (3.20) can be calculated as

$$\Phi = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 3 & 0 & 0 \\ 0 & \frac{1}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (3.25)$$

It is easy to check that $\mathbf{C} = \Phi \Lambda \Phi^T$. Now, the right stretch tensor can be calculated by

$$\mathbf{U} = \Phi \sqrt{\Lambda} \Phi^T = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{5}{2\sqrt{3}} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The rotation tensor and the left stretch tensor can be calculated from Eq. (3.18) as

$$\mathbf{Q} = \mathbf{F} \mathbf{U}^{-1} = \begin{bmatrix} \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{V} = \mathbf{F} \mathbf{Q}^T = \begin{bmatrix} \frac{5\sqrt{3}}{6} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- (b) The deformation stages are plotted using a square element, as shown in Fig. 3.4. In the \mathbf{QU} decomposition, the square is first stretched in its eigenvector directions and then rotated by 30° in the clockwise direction. In the \mathbf{VQ} decomposition, the square is first rotated by 30° in the clockwise direction

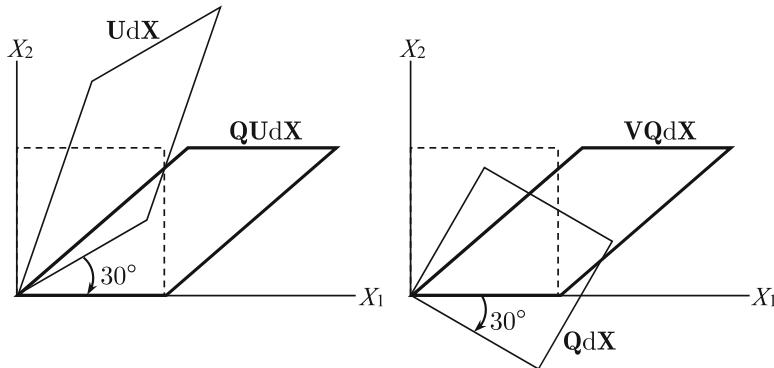


Fig. 3.4 Polar decomposition of simple shear deformation

and then stretched in the eigenvector directions. However, both decompositions yield the same final deformed geometry.

- (c) Consider a straight line $X_2 = X_1 \tan \theta$. From the deformation in Eq. (3.24), X_1 and X_2 can be written in terms of deformed coordinates: $X_1 = x_1 - kx_2$, $X_2 = x_2$. Thus, the initially straight line deforms to

$$x_2 = (x_1 - kx_2) \tan \theta \quad \Rightarrow \quad x_1 = \left(\frac{1}{\tan \theta} + \frac{2}{\sqrt{3}} \right) x_2$$

Note that the initially straight line deforms to another straight line. For example, let $\theta = 45^\circ$, and then this line will deform to another line with angle α :

$$\tan \alpha = \frac{x_2}{x_1} = \frac{1}{1+k} \quad \Rightarrow \quad \alpha \approx 24.9^\circ$$

- (d) The original equation of circle can be written in terms of deformed coordinates as

$$X_1^2 + X_2^2 = a^2 \quad \Rightarrow \quad (x_1 - kx_2)^2 + x_2^2 = a^2$$

$$x_1^2 - 2kx_1x_2 + (1+k^2)x_2^2 = a^2$$

Note that the original circle deforms to an ellipse. ■

3.2.4 Deformation of Surface and Volume

3.2.4.1 Volume Changes

The change in volume cannot be ignorable in large deformation. In addition, many materials show different behaviors between volume-changing deformation (dilatation) and volume-preserving deformation (distortion). Thus, it is important to express the **change in volume in terms of deformation**. Consider an infinitesimal volume element that is composed of three vectors in the undeformed geometry $dV_0 = d\mathbf{X}^1 \cdot (d\mathbf{X}^2 \times d\mathbf{X}^3)$, which is deformed to $dV_x = d\mathbf{x}^1 \cdot (d\mathbf{x}^2 \times d\mathbf{x}^3)$. Using the definition of deformation gradient, i.e., $d\mathbf{x} = \mathbf{F}d\mathbf{X}$, the relation between dV_0 and dV_x can be obtained as

$$dV_x = J dV_0, \quad (3.26)$$

where $J = |\mathbf{F}| = \det(\mathbf{F})$ is the determinant of deformation gradient (see Problem P3.2). Using the relationship in Eq. (3.26), we can calculate the volumetric strain by

$$\frac{dV_x - dV_0}{dV_0} = J - 1. \quad (3.27)$$

Note that if a material is incompressible, then $J = 1$. From the above relation, it is clear that J must be positive because it is impossible for a deformed volume to be zero or negative. In addition, the above relation can provide an important transformation in the integral of weak form. For example, if a function f is to be integrated over the deformed domain, then using Eq. (3.26), the integral domain can be changed to the undeformed geometry as

$$\iiint_{\Omega_x} f d\Omega = \iiint_{\Omega_0} f J d\Omega. \quad (3.28)$$

The above relation yields a very convenient way of solving nonlinear equations. Since the deformed geometry Ω_x is unknown, it is difficult to perform integration on the left-hand side. However, the integral on the right-hand side is performed over the known initial geometry Ω_0 . The determinant J contains all the effects of changing geometry.

3.2.4.2 Area Changes

Similar to the change in volume, the **change in the surface area can also be expressed in terms of deformation**. The change in surface area is especially important when a pressure load is applied on the surface and is dependent on the surface area. Let \mathbf{N} be a unit normal vector on infinitesimal area dS_0 of the parallelogram shown in Fig. 3.5, with two edges ($d\mathbf{X}^1$ and $d\mathbf{X}^2$) on an undeformed surface S_0 . Let

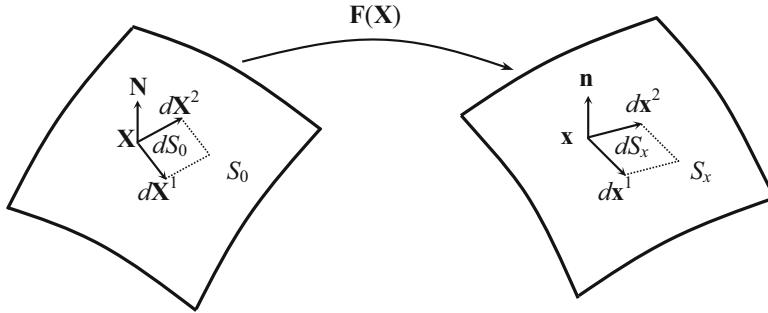


Fig. 3.5 Differential surfaces in the undeformed and deformed configurations

\mathbf{n} be a unit normal vector on an infinitesimal area dS_x of deformed surface S_x , with edges dx^1 and dx^2 . The objective is to find a relationship between dS_0 and dS_x when both surfaces are smooth.

The edges dx^1 and dx^2 can be represented by using the deformation gradient and the edges on the initial boundary as

$$\begin{aligned} dx^1 &= \mathbf{F}d\mathbf{X}^1 \\ dx^2 &= \mathbf{F}d\mathbf{X}^2 \end{aligned} \quad (3.29)$$

Since the mapping between dS_0 and dS_x are one-to-one, the inverse mapping \mathbf{F}^{-1} exists. Using \mathbf{F}^{-1} , the inverse relations can be obtained as

$$\begin{aligned} d\mathbf{X}^1 &= \mathbf{F}^{-1}dx^1 \\ d\mathbf{X}^2 &= \mathbf{F}^{-1}dx^2 \end{aligned} \quad (3.30)$$

Then, the infinitesimal areas of two boundaries can be denoted by using a vector product as

$$\begin{aligned} \mathbf{N}dS_0 &= d\mathbf{X}^1 \times d\mathbf{X}^2 \\ \mathbf{n}dS_x &= dx^1 \times dx^2 \end{aligned} \quad (3.31)$$

The above vector notation can be represented in Cartesian rectangular components as

$$\begin{aligned} N_i dS_0 &= e_{ijk} dX_j^1 dX_k^1 \\ n_r dS_x &= e_{rst} dx_s^1 dx_t^2 \end{aligned} \quad (3.32)$$

where e_{ijk} is a permutation symbol, defined as

$$e_{ijk} = \begin{cases} 0 & \text{when any two indices are equal} \\ +1 & \text{when } i,j,k \text{ are an even permutation of } 1, 2, 3. \\ -1 & \text{when } i,j,k \text{ are an odd permutation of } 1, 2, 3 \end{cases} \quad (3.33)$$

From the first equation of Eq. (3.32), and by using Eq. (3.30),

$$N_i dS_0 = e_{ijk} \frac{\partial X_j}{\partial x_s} \frac{\partial X_k}{\partial x_t} dx_s^1 dx_t^2. \quad (3.34)$$

Multiplying both sides of Eq. (3.34) by $\partial X_i / \partial x_r$ and summing on i , the following relation can be obtained:

$$\frac{\partial X_i}{\partial x_r} N_i dS_0 = e_{ijk} \frac{\partial X_i}{\partial x_r} \frac{\partial X_j}{\partial x_s} \frac{\partial X_k}{\partial x_t} dx_s^1 dx_t^2. \quad (3.35)$$

For any 3×3 matrix with elements a_{mn} , the following identity can be proved by direct calculation:

$$e_{rst} \det[a_{mn}] = e_{ijk} a_{ir} a_{js} a_{kt}. \quad (3.36)$$

Since the deformation gradient \mathbf{F} has $\partial \mathbf{x} / \partial \mathbf{X}$ as elements, the following relations hold:

$$\begin{aligned} e_{ijk} |\mathbf{F}| &= e_{rst} \frac{\partial x_r}{\partial X_i} \frac{\partial x_s}{\partial X_j} \frac{\partial x_t}{\partial X_k} \\ e_{rst} |\mathbf{F}^{-1}| &= e_{ijk} \frac{\partial X_i}{\partial x_r} \frac{\partial X_j}{\partial x_s} \frac{\partial X_k}{\partial x_t}. \end{aligned} \quad (3.37)$$

By substituting the second part of Eq. (3.37) into Eq. (3.35) and by recalling that $|\mathbf{F}^{-1}| = |\mathbf{F}|^{-1}$, the following simplified form can be obtained:

$$\frac{\partial X_i}{\partial x_r} N_i dS_0 = |\mathbf{F}|^{-1} e_{rst} dx_s^1 dx_t^1, \quad (3.38)$$

which can be rewritten using Eq. (3.32) as

$$\mathbf{n} dS_x = J \mathbf{F}^{-T} \cdot \mathbf{N} dS_0. \quad (3.39)$$

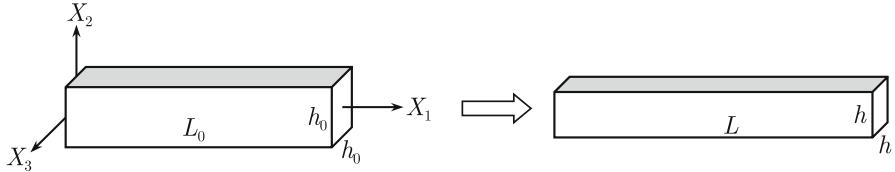


Fig. 3.6 Extension of an incompressible bar

Thus, \mathbf{n} is parallel to $\mathbf{F}^{-T} \cdot \mathbf{N}$. The explicit form of \mathbf{n} can be obtained by normalizing the right side of Eq. (3.39) as

$$\mathbf{n} = \frac{\mathbf{F}(\mathbf{x})^{-T} \mathbf{N}(\mathbf{X})}{\|\mathbf{F}(\mathbf{x})^{-T} \mathbf{N}(\mathbf{X})\|}, \quad (3.40)$$

where $\|\mathbf{a}\| = (\mathbf{a}^T \mathbf{a})^{1/2}$ is the Euclidean norm. By applying Eq. (3.40) to Eq. (3.39), we finally obtain the desired relation between $d\Gamma_0$ and $d\Gamma_x$ as

$$dS_x = J \|\mathbf{F}(\mathbf{x})^{-T} \mathbf{N}(\mathbf{X})\| dS_0. \quad (3.41)$$

Note that the deformed surface dS_x depends on not only J but also the unit normal vector of the undeformed surface.

Example 3.5 (Extension of an incompressible bar) Consider a bar under uniaxial tension, as shown in Fig. 3.6. The initial length and cross-sectional area are L_0 and $A_0 = h_0 \times h_0$, respectively. A force is applied at the tip such that the deformed length and cross-sectional area of the bar become L and $A = h \times h$, respectively. When the material is incompressible, calculate the deformed cross-sectional dimension h and area A in terms of L , L_0 , and h_0 .

Solution Since the bar will maintain its rectangular shape, there is no shear deformation. In addition, since both X_2 and X_3 directions are unconstrained, and the cross section remains a square; i.e., the principal stretches in these two directions will be the same. Thus, the relation between undeformed and deformed geometries can be written as

$$x_1 = \lambda_1 X_1, \quad x_2 = \lambda_2 X_2, \quad x_3 = \lambda_3 X_3.$$

From the given deformation, the principal stretches can be written as

$$\lambda_1 = \frac{L}{L_0}, \quad \lambda_2 = \lambda_3 = \frac{h}{h_0},$$

and the deformation gradient can be obtained as

$$\mathbf{F} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}.$$

The determinant of the deformation gradient becomes

$$J = |\mathbf{F}| = \lambda_1 \lambda_2 \lambda_3 = \frac{L}{L_0} \left(\frac{h}{h_0} \right)^2 = \frac{LA}{L_0 A_0},$$

Thus, from the incompressible condition

$$J = 1 \quad \Rightarrow \quad h = h_0 \sqrt{\frac{L_0}{L}}, \quad A = A_0 \frac{L_0}{L}. \quad \blacksquare$$

3.2.5 Cauchy and Piola-Kirchhoff Stresses

Stress is one of the most important quantities in solid mechanics. It is used in the equilibrium of a structure and also to determine the failure of a material. Similar to strain, stress also depends on the frame of reference. Different stresses can be defined based on the frame of reference used. In general, stress is defined by force acting on an infinitesimal area. In linear analysis, it was unnecessary to distinguish the deformed area from the undeformed area because of the infinitesimal deformation assumption. However, when deformation is large, it is important to clarify what area is used in defining stress. In fact, depending on the area used, the definition of stress changes. Since the undeformed and deformed geometries are used as frames of reference, the areas from these two geometries will be used in defining stresses.

Referring to Fig. 3.7, the stress vector at point Q in the current deformed geometry can be written using the area of the differential element ΔS_x , the force $\Delta \mathbf{f}$ acting on it, and the unit normal \mathbf{n} of the area as

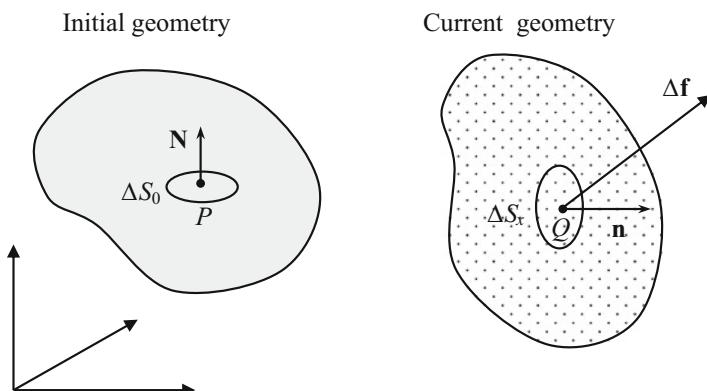


Fig. 3.7 Stress vectors in the initial and deformed geometries

$$\mathbf{t} = \lim_{\Delta S_x \rightarrow 0} \frac{\Delta \mathbf{f}}{\Delta S_x} = \boldsymbol{\sigma} \mathbf{n}, \quad (3.42)$$

where $\boldsymbol{\sigma}$ is known as the Cauchy stress tensor with a dimension of 3×3 . The Cauchy stress tensor refers to the current deformed geometry as a reference for both the force and area, and therefore it is often called the true stress.

A different stress vector can be defined by considering the same force $\Delta \mathbf{f}$ but the differential area ΔS_0 and the unit normal \mathbf{N} in the undeformed geometry as

$$\mathbf{T} = \lim_{\Delta S_0 \rightarrow 0} \frac{\Delta \mathbf{f}}{\Delta S_0} = \mathbf{P}^T \mathbf{N}, \quad (3.43)$$

where \mathbf{P} is known as the first Piola-Kirchhoff stress tensor with a dimension of 3×3 . Different from the Cauchy stress, the first P–K stress \mathbf{P} is not symmetric. In fact, $\boldsymbol{\sigma}$ refers to the current geometry for both force and area, whereas \mathbf{P} refers to the force in the current geometry and the area in the initial geometry.

Since a differential surface area in the current geometry with unit normal \mathbf{n} is related to its counterpart in the initial geometry through Eq. (3.39), the first Piola-Kirchhoff stress is also related to the Cauchy stress. In order to develop this relationship, the infinitesimal force is written in terms of two stresses as

$$d\mathbf{f} = \boldsymbol{\sigma} n dS_x = \mathbf{P}^T \mathbf{N} dS_0. \quad (3.44)$$

Using the relation in Eq. (3.39), the following relationship between \mathbf{P} and $\boldsymbol{\sigma}$ can be obtained:

$$\mathbf{P} = J \mathbf{F}^{-1} \boldsymbol{\sigma}. \quad (3.45)$$

The first Piola-Kirchhoff stress tensor has one undesirable property: it is not symmetric. By post-multiplying \mathbf{P} with the transpose of the inverse of the deformation gradient, a symmetric tensor can be obtained. This pseudo stress tensor is called the second Piola-Kirchhoff stress tensor and will be denoted by \mathbf{S} . Thus, from definition,

$$\mathbf{S} = \mathbf{P} \mathbf{F}^{-T} = J \mathbf{F}^{-1} \boldsymbol{\sigma} \mathbf{F}^{-T}. \quad (3.46)$$

If the second Piola-Kirchhoff stress tensor is known, the Cauchy stress tensor can be obtained by inverting the relationship as follows:

$$\boldsymbol{\sigma} = \frac{1}{J} \mathbf{F} \mathbf{S} \mathbf{F}^T. \quad (3.47)$$

In Eq. (3.47), the denominator J is related to the volume change between the undeformed and deformed geometries. It is inconvenient to calculate stress with the effect of the determinant because it also depends on the deformation. Thus, a new stress measure can be defined that has a similar transformation relationship with

Eq. (3.47), except for the determinant. This new stress is called Kirchhoff stress and is defined as

$$\boldsymbol{\tau} = J\boldsymbol{\sigma} = \mathbf{F}\mathbf{S}\mathbf{F}^T. \quad (3.48)$$

Referring to Eq. (3.28), the convenience of the Kirchhoff stress can be obvious. The weak form of a structural system is in fact a balance between the internal virtual work and the external virtual work. In large deformation systems, the internal virtual work can be defined as

$$\iiint_{\Omega_e} \boldsymbol{\sigma} : \bar{\boldsymbol{\epsilon}} d\Omega = \iiint_{\Omega_0} \boldsymbol{\tau} : \bar{\boldsymbol{\epsilon}} d\Omega. \quad (3.49)$$

Thus, by using the Kirchhoff stress, the integral can be performed in the undeformed geometry. The Kirchhoff stress tensor is exactly the same as the Cauchy stress tensor, except that it refers to the undeformed domain.

It should be emphasized that the Piola-Kirchhoff stresses are just convenient mathematical quantities. They are not directly related to the surface tractions in the deformed geometry as are the Cauchy's stresses. In linear analysis, the displacements are assumed to be small and no distinction is made in the initial and deformed geometries, and hence Piola-Kirchhoff and Cauchy stresses become identical, i.e., $\boldsymbol{\sigma} \approx \boldsymbol{\tau} \approx \mathbf{P} \approx \mathbf{S}$.

The stresses produced in a body are related to material straining. Rigid-body rotations and translations obviously do not cause any stresses regardless of their magnitudes. Thus, the constitutive equations are valid for any stress and strain measures which are invariant under rigid-body motions. It can be shown that the second Piola-Kirchhoff stress and Lagrangian strain are invariant under large rigid-body rotations and translations. Thus, the relationship between them can be written using constitutive equations. In particular, if the material is assumed to remain elastic, then the generalized Hooke's law can be used to relate the stress and strain, which will be discussed in the following section.

Example 3.6 (Cauchy stress and the second Piola-Kirchhoff stress) Consider the uniaxial tension of a bar in Example 3.5. When the applied load at the end of the bar is F , calculate the Cauchy stress and the first and second Piola-Kirchhoff stresses. Assume that the applied force is uniformly distributed over the cross section.

Solution Since the applied force F is uniformly distributed over the cross section, the Cauchy stress becomes

$$\boldsymbol{\sigma} = \begin{bmatrix} F/A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since all components are zero except for σ_{11} , only the first component will be discussed in the following calculations. From the relation in Eq. (3.45), the first Piola-Kirchhoff stress can be calculated by

$$P_{11} = (J\mathbf{F}^{-1}\boldsymbol{\sigma})_{11} = \frac{F 1}{A\lambda_1} = \frac{FA}{AA_0} = \frac{F}{A_0}.$$

Note that the first Piola-Kirchhoff stress is defined with respect to the undeformed cross-sectional area A_0 . Using Eq. (3.47), the second Piola-Kirchhoff stress can be calculated as

$$S_{11} = (J\mathbf{F}^{-1} \cdot \boldsymbol{\sigma} \cdot \mathbf{F}^{-T})_{11} = \frac{F 1}{A\lambda_1^2} = \frac{FA^2}{AA_0^2} = \frac{FA}{A_0^2} = \frac{F}{A_0\lambda_1}.$$

Note that the second Piola-Kirchhoff stress does not have clear physical meaning. In the case of extension, i.e., $\lambda > 1$, the magnitude of the three stresses is such that $\sigma_{11} > P_{11} > S_{11}$. In the case of compression, i.e., $\lambda < 1$, they are $\sigma_{11} < P_{11} < S_{11}$. ■

3.3 Nonlinear Elastic Analysis

As a first step toward formulating nonlinear structural systems, a simple elastic system is introduced in this section. Among different nonlinearities in the previous chapter, only geometric nonlinearity will be discussed. In other words, the structure may experience large deformation, but the stress-strain relation is still linear. Of course, the linear stress-strain relation does not indicate a linear system because different measures of stress and strain are used for large deformation systems. Out of various possible combinations, the second Piola-Kirchhoff stress and Lagrangian strain will be used in the following derivations of nonlinear elastic systems.

One of the unique properties of elasticity is that a potential exists such that the structure is in equilibrium under deformation at which the potential has a minimum value. In such a case, the principle of minimum potential energy can be used to derive the equilibrium equation of nonlinear elastic structures, and the strain energy density can be differentiated with respect to strain in order to obtain stress.

The solution procedure usually requires a (incremental) linearization procedure, such as the Newton–Raphson method in Chap. 2. Based on the reference frame used for the linearization, two formulations will be introduced: the total Lagrangian (material) and the updated Lagrangian (spatial). Through consistent linearization and transformation, it will be shown that the two are equivalent.

3.3.1 Nonlinear Static Analysis: Total Lagrangian Formulation

In this section, a structural equilibrium equation and its linearization will be developed for nonlinear elastic systems using undeformed geometry as a frame of reference. Figure 3.8 shows the initial and deformed geometries of a structure. Even if the goal is to find the final equilibrium geometry when the total force is applied to the structure, the applied force is often incrementally increased and solved for intermediate equilibrium. In Chap. 2, this incremental force was explained using a load step or a time step. Let us consider a static system that is composed of N load steps. Before reaching the final load step, the current load step is denoted by n . We will denote the n th load as t_n . For variables that depend on load steps, we will use a left superscript to denote the variable at a specific load step. For example, ${}^0\Omega(\mathbf{X})$ and ${}^n\Omega(\mathbf{x})$ represent the initial and current domains, respectively. In many cases, for notational convenience, we will omit left superscript n unless needed for clarity.

In starting the n th load step, the applied load is increased and structural equilibrium is sought using iterative methods, such as the Newton–Raphson method. It is further assumed that up to the k th iteration has been finished. The objective is to find the incremental displacement at the $(k+1)$ th iteration so that the residual force vanishes. With this status at hand, a nonlinear equilibrium equation and incremental solution procedure are developed in the following subsections.

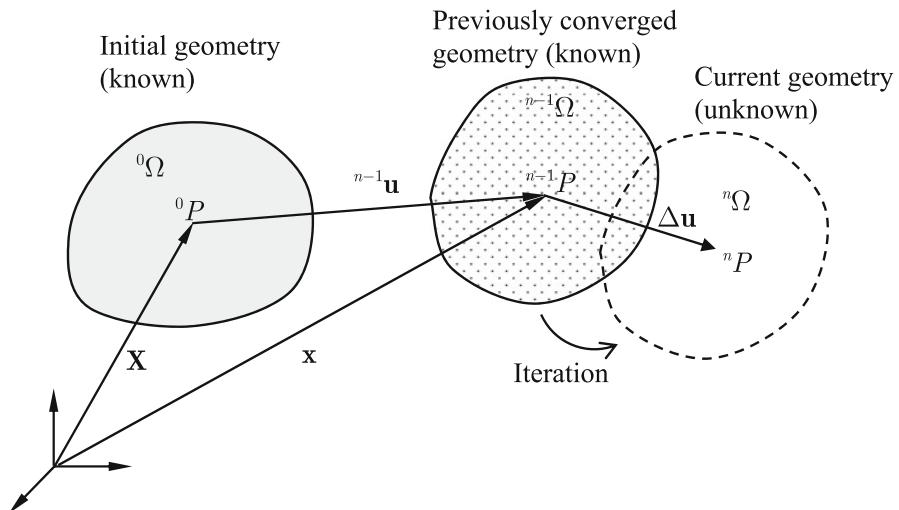


Fig. 3.8 Configuration change during deformation

3.3.1.1 Constitutive Relation

The constitutive theory, or stress–strain relation, describes the macroscopic behavior of a material between deformation (strain) and internal force (stress) caused by deformation. Since the behavior of real materials is in general very complex, it is approximated using physical observations of the material’s response. This approximation can be done separately for different material responses (e.g., elastic, plastic, or viscoplastic).

A structural material is called elastic when a **strain energy density** W exists such that the stress can be obtained by differentiating W with respect to strain. For this approach, appropriate stress and strain measures must be used. For example, when engineering strain is used in defining W , the Cauchy stress must be used as a stress measure. In the same sense, the **second Piola-Kirchhoff stress** must be used when the Lagrangian strain is used. Since the reference frame of the total Lagrangian formulation is undeformed geometry, it is convenient to use the Lagrangian strain in defining the strain energy density. Then, the second Piola-Kirchhoff stress can be obtained by differentiating the strain energy density with respect to the Lagrangian strain.

Even if complex material responses can be introduced, such as hyperelastic material models in Sect. 3.5, a simple form of constitutive relation is first introduced. To take a simple example using **St. Venant–Kirchhoff nonlinear elastic material** [5], consider the following form of the strain energy density:

$$W(\mathbf{E}) = \frac{1}{2} \mathbf{E} : \mathbf{D} : \mathbf{E}, \quad (3.50)$$

where the notation “ $:$ ” is the contraction operator of tensors, such that $\mathbf{a}:\mathbf{b} = a_{ij}b_{ij}$, with summation in repeated indices, and \mathbf{D} is the fourth-order constitutive tensor for isotropic materials, defined by

$$\begin{aligned} \mathbf{D} &= \lambda \mathbf{1} \otimes \mathbf{1} + 2\mu \mathbf{I} \\ D_{ijkl} &= \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) \end{aligned} \quad (3.51)$$

This is basically the same as linear elastic materials. In Eq. (3.51), λ and μ are Lame’s constants for isotropic materials, $\mathbf{1}$ is the second-order unit tensor, \otimes is the symbol for the tensor product, and \mathbf{I} is the fourth-order unit symmetric tensor defined as $I_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$. Two Lame’s constants in the above equation can be expressed using regular elastic constants for the isotropic material as

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{2(1+\nu)}, \quad (3.52)$$

where E is Young’s modulus and ν is Poisson’s ratio. Among four constants, only two are independent for an isotropic material.

The constitutive relation can be obtained by differentiating Eq. (3.50) with respect to the Lagrangian strain \mathbf{E} , to obtain

$$\mathbf{S} \equiv \frac{\partial W(\mathbf{E})}{\partial \mathbf{E}} = \mathbf{D} : \mathbf{E} = \lambda \text{tr}(\mathbf{E})\mathbf{1} + 2\mu\mathbf{E}, \quad (3.53)$$

where \mathbf{S} is the second Piola-Kirchhoff stress. Note that the relation between stress \mathbf{S} and strain \mathbf{E} is linear. In practice, not many materials will show the linear stress-strain relation as in Eq. (3.53) when the strain is large. Thus, the constitutive relation in Eq. (3.53) is restricted to a small strain. As has been seen in Example 3.3, because the Lagrangian strain is not affected by rigid-body motions, this constitutive relation can accurately represent deformation with rigid-body motions. In fact, using polar decomposition, the deformation gradient can be decomposed into $\mathbf{F} = \mathbf{Q} \cdot \mathbf{U}$ where \mathbf{Q} is the rotational tensor and \mathbf{U} is the stretch tensor. Then, only the stretch tensor should affect the strain. In fact, the Lagrangian strain can be written in terms of the stretch tensor by

$$\mathbf{E} = \frac{1}{2}(\mathbf{U}^2 - \mathbf{1}). \quad (3.54)$$

Example 3.7 (St. Venant–Kirchhoff material) For the simple shear deformation in Example 3.4, calculate the first and second Piola-Kirchhoff stresses and Cauchy stress as a function of shear parameter k . Plot the shear stress components from the three stresses as a function of parameter k . Assume isotropic St. Venant–Kirchhoff material between the second Piola-Kirchhoff stress and the Lagrangian strain. The two constants are $E = 100$ MPa and $\nu = 0.25$.

Solution From the given deformation in Example 3.4, the Lagrangian strain can be calculated as

$$\mathbf{E} = \frac{1}{2}(\mathbf{C} - \mathbf{1}) = \frac{1}{2} \begin{bmatrix} 0 & k & 0 \\ k & k^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Convert the two material constants to the Lame's constants because they are convenient to use:

$$\lambda = \frac{\nu E}{(1 + \nu)(1 - 2\nu)} = 40 \text{ MPa}, \quad \mu = \frac{E}{2(1 + \nu)} = 40 \text{ MPa}.$$

Thus, from Eq. (3.53),

$$\mathbf{S} = \mathbf{D} : \mathbf{E} = \lambda \text{tr}(\mathbf{E})\mathbf{1} + 2\mu\mathbf{E} = 20 \begin{bmatrix} k^2 & 2k & 0 \\ 2k & 3k^2 & 0 \\ 0 & 0 & k^2 \end{bmatrix} \text{ MPa}.$$

Note that in addition to the shear component S_{12} , normal components also exist. These normal components are all second order. Thus, if the shear deformation is small, they will approach zero quickly and recover the infinitesimal deformation theory. The first Piola-Kirchhoff and Cauchy stresses can be calculated from the relation in Eqs. (3.45) and (3.47) as

$$\mathbf{P} = \mathbf{SF}^T = 20 \begin{bmatrix} 3k^2 & 2k & 0 \\ 2k + 3k^3 & 3k^2 & 0 \\ 0 & 0 & k^2 \end{bmatrix} \text{ MPa},$$

$$\boldsymbol{\sigma} = \frac{1}{J} \mathbf{FSF}^T = 20 \begin{bmatrix} 5k^2 + 3k^4 & 2k + 3k^3 & 0 \\ 2k + 3k^3 & 3k^2 & 0 \\ 0 & 0 & k^2 \end{bmatrix} \text{ MPa}.$$

Note that all three stresses have the same linear shear stress component and all other terms are higher orders. Since the higher-order terms will decrease quickly for small k values, they all recover the infinitesimal simple shear deformation. The first Piola-Kirchhoff stress is not symmetric, and the Cauchy stress is different from the stress in linear analysis. Indeed, the Cauchy stress here is the current stress at the deformed geometry, while the stress in linear analysis is from small deformation assumption. Figure 3.9 shows the plot of shear stress components. Note that in the St. Venant–Kirchhoff material, the second Piola-Kirchhoff stress is linear, while other stresses are nonlinear. ■

3.3.1.2 Boundary Conditions

To obtain a well-defined mathematical problem, boundary conditions must be added to the equilibrium equation, which prescribes the displacement field \mathbf{u} on Γ^g and the surface tractions \mathbf{f}^S on Γ^S in the following forms:

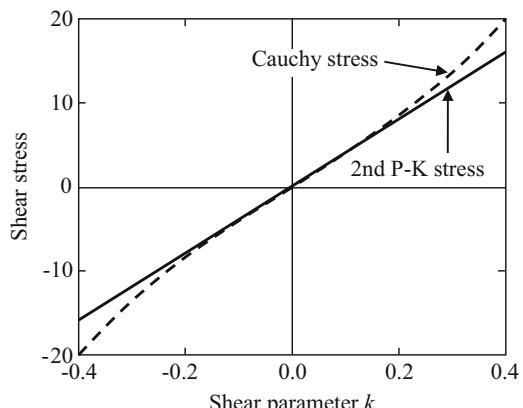


Fig. 3.9 Shear component of the second P–K and Cauchy stresses

$$\begin{aligned}\mathbf{u} &= \mathbf{g} && \text{on } \Gamma^h \\ \mathbf{t} &= \mathbf{P}^T \mathbf{N} && \text{on } \Gamma^s.\end{aligned}\quad (3.55)$$

Mathematically, the first is called the essential boundary condition and the second the natural boundary condition. Note that even if the second Piola-Kirchhoff stress is used as a stress measure, the first Piola-Kirchhoff stress is used in the traction boundary condition as the latter is based on physical forces. Using the displacement boundary conditions, the solution space \mathbb{V} and the space \mathbb{Z} of kinematically admissible displacements are defined as

$$\mathbb{V} = \left\{ \mathbf{u} \mid \mathbf{u} \in [H^1(\Omega)]^N, \mathbf{u}|_{\Gamma^h} = \mathbf{g} \right\}, \quad (3.56)$$

and

$$\mathbb{Z} = \left\{ \bar{\mathbf{u}} \mid \bar{\mathbf{u}} \in [H^1(\Omega)]^N, \bar{\mathbf{u}}|_{\Gamma^h} = 0 \right\}, \quad (3.57)$$

where $H^1(\Omega)$ is the space of functions whose first-order derivatives are bounded in the **energy norm**. The space \mathbb{Z} is the same as solution space \mathbb{V} , except that it satisfies the homogeneous essential boundary conditions.

3.3.1.3 Principle of Minimum Potential Energy

The weak form of a nonlinear elastic system can be obtained from the principle of minimum potential energy. The potential energy of an elastic system is the difference between the stored strain energy Π^{int} and the work done by external forces Π^{ext} . The strain energy can be obtained by integrating the strain energy density function in Eq. (3.50) over the undeformed initial geometry. The work done by applied forces can be obtained by multiplying displacement with the applied forces. For simplicity, it is assumed that the applied forces are conservative, which means that the applied load is independent of deformation. Therefore, these forces can be transformed to the undeformed geometry.

Using the strain energy and work done by applied forces, the potential energy of an elastic system can be obtained as

$$\begin{aligned}\Pi(\mathbf{u}) &= \Pi^{\text{int}}(\mathbf{u}) - \Pi^{\text{ext}}(\mathbf{u}) \\ &= \iint_{\Omega} W(\mathbf{E}) \, d\Omega - \iint_{\Omega} \mathbf{u}^T \mathbf{f}^b \, d\Omega - \int_{\Gamma^s} \mathbf{u}^T \mathbf{t} \, d\Gamma,\end{aligned}\quad (3.58)$$

where \mathbf{f}^b is the body force, \mathbf{t} is the surface traction force on the boundary ${}^0\Gamma^s$, and \mathbf{E} is the Lagrangian strain, defined as

$$\mathbf{E} = \frac{1}{2}(\mathbf{F}^T \mathbf{F} - \mathbf{1}) = \frac{1}{2}(\nabla_0 \mathbf{u}^T + \nabla_0 \mathbf{u} + \nabla_0 \mathbf{u}^T \nabla_0 \mathbf{u}). \quad (3.59)$$

The principle of minimum potential energy holds that displacement field $\mathbf{u} \in \mathbb{V}$ in the equilibrium minimizes Eq. (3.58). In order to find the displacement at the minimum potential energy, a perturbation method is often used. Let us assume that the displacement field \mathbf{u} is perturbed in the direction of $\bar{\mathbf{u}}$ (arbitrary) and τ is the parameter that controls the perturbation size. The perturbed displacement is denoted by

$$\mathbf{u}_\tau = \mathbf{u} + \tau \bar{\mathbf{u}}. \quad (3.60)$$

Note that $\bar{\mathbf{u}}$ corresponds to the virtual displacement in the principle of virtual work. In the above equation, the perturbed solution \mathbf{u}_τ also belongs to the solution space \mathbb{V} . Accordingly, the variation $\bar{\mathbf{u}}$ must satisfy the homogeneous essential boundary condition, i.e., $\bar{\mathbf{u}} \in \mathbb{Z}$.

Then, the first variation of the potential energy can be obtained by taking the first-order variation of $\Pi(\mathbf{u})$ in the direction of $\bar{\mathbf{u}}$, as

$$\bar{\Pi}(\mathbf{u}, \bar{\mathbf{u}}) \equiv \frac{d}{d\tau} \Pi(\mathbf{u} + \tau \bar{\mathbf{u}}) \Big|_{\tau=0}, \quad (3.61)$$

where the overhead bar symbol represents the first-order variation of a function. The process of variation is similar to the differentiation of a function. Note that $\Pi(\mathbf{u})$ only depends on the displacement \mathbf{u} , whereas $\bar{\Pi}(\mathbf{u}, \bar{\mathbf{u}})$ depends on both \mathbf{u} and its variation $\bar{\mathbf{u}}$. Using the potential energy in Eq. (3.58) and equating the first variation to zero, the following variational equation can be obtained:

$$\bar{\Pi}(\mathbf{u}, \bar{\mathbf{u}}) = \iint_{{}^0\Omega} \frac{\partial W(\mathbf{E})}{\partial \mathbf{E}} : \bar{\mathbf{E}} \, d\Omega - \iint_{{}^0\Omega} \bar{\mathbf{u}}^T \mathbf{f}^b \, d\Omega - \int_{{}^0\Gamma^s} \bar{\mathbf{u}}^T \mathbf{t} \, d\Gamma = 0. \quad (3.62)$$

In the above equation, the variation of the work done by applied loads is straightforward as it is linear with respect to the displacement \mathbf{u} . For the variation of the strain energy, using the chain rule of differentiation, the strain energy density is differentiated with respect to the Lagrangian strain, and then the variation of the Lagrangian strain is taken from its definition in Eq. (3.59) as

$$\begin{aligned} \bar{\mathbf{E}}(\mathbf{u}, \bar{\mathbf{u}}) &= \frac{d}{d\tau} \mathbf{E}(\mathbf{u} + \tau \bar{\mathbf{u}}) \Big|_{\tau=0} \\ &= \frac{1}{2}(\nabla_0 \bar{\mathbf{u}} + \nabla_0 \bar{\mathbf{u}}^T + \nabla_0 \bar{\mathbf{u}}^T \nabla_0 \mathbf{u} + \nabla_0 \mathbf{u}^T \nabla_0 \bar{\mathbf{u}}), \\ &= \text{sym}(\nabla_0 \bar{\mathbf{u}}^T + \nabla_0 \bar{\mathbf{u}}^T \nabla_0 \mathbf{u}) \\ &= \text{sym}(\nabla_0 \bar{\mathbf{u}}^T \mathbf{F}) \end{aligned} \quad (3.63)$$

where $\text{sym}(\bullet)$ denotes the symmetric part of a tensor. Note that $\bar{\mathbf{E}}(\mathbf{u}, \bar{\mathbf{u}})$ is a bilinear function of \mathbf{u} and $\bar{\mathbf{u}}$, even if $\mathbf{E}(\mathbf{u})$ is nonlinear. The principle of minimum potential energy says that if the system is in equilibrium, the variation in Eq. (3.62) must vanish for all $\bar{\mathbf{u}}$ that belongs to the space \mathbb{Z} of kinematically admissible displacements. This is similar to the idea that a function has its minimum value when its slope becomes zero.

Since the variational equation (3.62) is similar to that of linear systems, the same notation will be used here. Thus, the variational equation for the nonlinear elastic system can be written as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (3.64)$$

where $a(\mathbf{u}, \bar{\mathbf{u}})$ is the energy form and $\ell(\bar{\mathbf{u}})$ is the load form, defined as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \iint_{\Omega} \mathbf{S}(\mathbf{u}) : \bar{\mathbf{E}}(\mathbf{u}, \bar{\mathbf{u}}) \, d\Omega, \quad (3.65)$$

and

$$\ell(\bar{\mathbf{u}}) = \iint_{\Omega} \bar{\mathbf{u}}^T \mathbf{f}^B \, d\Omega + \int_{\Gamma^S} \bar{\mathbf{u}}^T \mathbf{t} \, d\Gamma. \quad (3.66)$$

The only difference is that the energy form $a(\mathbf{u}, \bar{\mathbf{u}})$ is nonlinear with respect to its arguments. Note that the derivative of strain energy density with respect to Lagrangian strain becomes the second Piola-Kirchhoff stress from Eq. (3.53).

The variational equation in Eq. (3.64) is indeed the weak form of nonlinear elastic systems. It is called the material description or the total Lagrangian formulation, since the stress \mathbf{S} and the strain \mathbf{E} use the initial undeformed geometry as a reference. Note that $a(\mathbf{u}, \bar{\mathbf{u}})$ and $\ell(\bar{\mathbf{u}})$ are linear with respect to $\bar{\mathbf{u}}$ but are nonlinear with respect to displacement \mathbf{u} . Nonlinearity comes from the fact that the stress and strain implicitly depend on \mathbf{u} .

3.3.1.4 Linearization (Tangent Stiffness)

The nonlinear variational equation (3.64) cannot be solved easily due to the nonlinearity involved in the displacement-strain relation. A general nonlinear equation can be solved using a Newton-Raphson iterative method through a sequence of linearization. Let us assume that equilibrium in Eq. (3.64) is not satisfied. Then, the difference between the left- and right-hand sides is defined as a residual,

$$R = a(\mathbf{u}, \bar{\mathbf{u}}) - \ell(\bar{\mathbf{u}}). \quad (3.67)$$

In the Newton-Raphson method, the Jacobian of the residual is required in each iteration. Since the Jacobian in a one-dimensional problem is nothing but a tangent

line at the current solution, it is often called a tangent stiffness, and the process is called linearization. Let the linearization of a function $f(\mathbf{x})$ in the direction $\Delta \mathbf{u}$ be denoted as

$$L[f] \equiv \frac{d}{d\omega} f(\mathbf{x} + \omega \Delta \mathbf{u}) \Big|_{\omega=0} = \frac{\partial f^T}{\partial \mathbf{x}} \Delta \mathbf{u}. \quad (3.68)$$

Note that this process is similar to the variation of a function in Eq. (3.61) by substituting $\Delta \mathbf{u}$ for $\bar{\mathbf{u}}$. If right superscript k denotes the iteration counter, then the linear incremental solution procedure of the nonlinear equation $f(\mathbf{x}^{k+1}) = 0$ becomes

$$\begin{aligned} \frac{\partial f^T}{\partial \mathbf{x}^k} \Delta \mathbf{u}^k &= -f(\mathbf{x}^k) \\ \mathbf{u}^{k+1} &= \mathbf{u}^k + \Delta \mathbf{u}^k \\ \mathbf{x}^{k+1} &= \mathbf{X} + \mathbf{u}^{k+1}. \end{aligned} \quad (3.69)$$

Thus, for a nonlinear elastic system, Eq. (3.67) is solved iteratively until the residual term vanishes. Figure 3.10 illustrates a one-dimensional example of the Newton–Raphson iterative method.

The nonlinear equation (3.64) can be linearized following the same procedure explained in Eq. (3.69). Since the load form in Eq. (3.66) is independent of displacement, it is unnecessary to linearize it. Linearization of the energy form in Eq. (3.65) can be written as

$$L[a(\mathbf{u}, \bar{\mathbf{u}})] = \iint_{\Omega} [\Delta \mathbf{S} : \bar{\mathbf{E}} + \mathbf{S} : \Delta \bar{\mathbf{E}}] d\Omega, \quad (3.70)$$

where $\Delta \mathbf{S}$ is the stress increment and $\Delta \bar{\mathbf{E}}$ is the increment of strain variation. For the St. Venant–Kirchhoff material, the stress–strain relation is linear, and thus, the increment of stress can be written as

$$\Delta \mathbf{S} = \frac{\partial \mathbf{S}}{\partial \mathbf{E}} : \Delta \mathbf{E} = \mathbf{D} : \Delta \mathbf{E}, \quad (3.71)$$

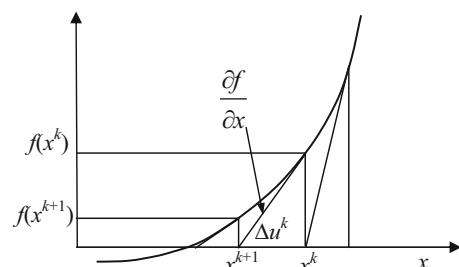


Fig. 3.10 Newton–Raphson method for nonlinear equation $f=0$

where \mathbf{D} is the fourth-order constitutive tensor in Eq. (3.51) and $\Delta\mathbf{E}$ is the increment of Lagrangian strain. By noting that the increment of deformation gradient is $\Delta\mathbf{F} = \nabla_0 \Delta\mathbf{u}$ from Eq. (3.5), the increment of Lagrangian strain and its variation can be obtained as

$$\Delta\mathbf{E}(\mathbf{u}, \Delta\mathbf{u}) = \text{sym}(\nabla_0 \Delta\mathbf{u}^T \mathbf{F}) \quad (3.72)$$

and

$$\Delta\bar{\mathbf{E}}(\Delta\mathbf{u}, \bar{\mathbf{u}}) = \text{sym}(\nabla_0 \bar{\mathbf{u}}^T \nabla_0 \Delta\mathbf{u}). \quad (3.73)$$

Thus, the linearization of the energy form in (3.70) can be explicitly derived with respect to displacement and its variation as

$$L[a(\mathbf{u}, \bar{\mathbf{u}})] = \iint_{\Omega} [\bar{\mathbf{E}} : \mathbf{D} : \Delta\mathbf{E} + \mathbf{S} : \Delta\bar{\mathbf{E}}] \, d\Omega \equiv a^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}). \quad (3.74)$$

The notation $a^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}})$ is used such that the form implicitly depends on the total displacement \mathbf{u} and is bilinear with respect to $\Delta\mathbf{u}$ and $\bar{\mathbf{u}}$. The first integrand of $a^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}})$ in Eq. (3.74) depends on the stress-strain relation. Since it is similar to the stiffness term in linear systems, it is called the tangent stiffness. On the other hand, the second integrand does not exist in linear systems. It only appears in geometric nonlinear problems. Since it has the stress term, it is called the initial stress stiffness.

Let the current load step be t_n and let the current iteration counter be k . Assuming that the applied loads are independent of displacement, the linearized incremental equation of Eq. (3.64) is obtained as

$$a^*({}^n\mathbf{u}^k; \Delta\mathbf{u}^k, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}) - a({}^n\mathbf{u}^k, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (3.75)$$

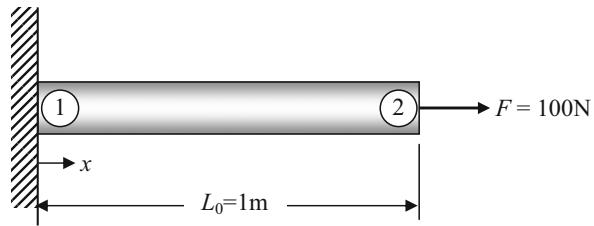
and the total displacement is updated using

$${}^n\mathbf{u}^{k+1} = {}^n\mathbf{u}^k + \Delta\mathbf{u}^k \quad (3.76)$$

Note that incremental equation (3.75) is in the form of $[{}^n\mathbf{K}^k] \cdot \{\Delta\mathbf{u}^k\} = \{{}^n\mathbf{R}^k\}$ after discretization using finite elements, which will be discussed in Sect. 3.6. Equation (3.75) is solved iteratively until the residual vanishes, which means that the original nonlinear equation (3.64) is satisfied.

Example 3.8 (Uniaxial bar: total Lagrangian formulation) Using the total Lagrangian formulation, solve displacement at the tip, stress, and strain of the uniaxial bar in Fig. 3.11 under tip force $F = 100$ N. Use a two-node bar element. Assume St. Venant–Kirchhoff material with $E = 200$ Pa and cross-sectional area $A = 1.0 \text{ m}^2$.

Fig. 3.11 Uniaxial bar analysis using the total Lagrangian formulation



Solution Since Node 1 of the bar element is fixed, u_2 is the only free degree of freedom. In addition, since the length of the element is a unit, the approximation of displacement gradient and its variation become

$$\frac{du}{dX} = u_2, \quad \frac{d\bar{u}}{dX} = \bar{u}_2.$$

Since the problem is one-dimensional, it is sufficient to consider S_{11} for stress and E_{11} for strain. First, the strain energy density of the St. Venant–Kirchhoff material is given as

$$W(E_{11}) = \frac{1}{2}E \cdot (E_{11})^2$$

where E is Young's modulus and the Lagrangian strain E_{11} is defined as

$$E_{11} = \frac{du}{dX} + \frac{1}{2} \left(\frac{du}{dX} \right)^2 = u_2 + \frac{1}{2}(u_2)^2.$$

Note that the strain is a nonlinear function of displacement. For the elastic material, the second Piola-Kirchhoff stress can be calculated by

$$S_{11} = \frac{\partial W}{\partial E_{11}} = E \cdot E_{11} = E \left(u_2 + \frac{1}{2}(u_2)^2 \right).$$

The variation of the Lagrangian strain becomes

$$\bar{E}_{11} = \frac{d\bar{u}}{dX} + \frac{du}{dX} \frac{d\bar{u}}{dX} = \bar{u}_2(1 + u_2).$$

Thus, the energy form can be obtained as

$$a(u, \bar{u}) = \int_0^{L_0} S_{11} \bar{E}_{11} A dX = S_{11} A L_0 (1 + u_2) \bar{u}_2.$$

Note that both stress and strain are constant within the element. Since the applied load is a concentrated force at the tip, no integration is required. The load form simply becomes

$$\ell(\bar{u}) = \bar{u}_2 F.$$

The difference between the energy and load forms can be defined as a residual R , and the nonlinear variational equation satisfies when the residual becomes zero:

$$R = \bar{u}_2(S_{11}AL_0(1 + u_2) - F) = 0$$

for any arbitrary \bar{u}_2 .

In order to solve the nonlinear variational equation using the Newton–Raphson method, the following increments of stress and strain variation are required:

$$\begin{aligned}\Delta S_{11} &= E\Delta E_{11} = E(1 + u_2)\Delta u_2 \\ \Delta \bar{E}_{11} &= \bar{u}_2\Delta u_2\end{aligned}$$

Thus, the linearization of the energy form becomes

$$\begin{aligned}a^*(u; \Delta u, \bar{u}) &= \int_0^{L_0} (\bar{E}_{11} \cdot E \cdot \Delta E_{11} + S_{11} \cdot \Delta \bar{E}_{11}) A dX \\ &= EAL_0(1 + u_2)^2 \bar{u}_2 \Delta u_2 + S_{11}AL_0 \bar{u}_2 \Delta u_2\end{aligned}$$

Let the current iteration counter be k . Since the linearized variational equation must satisfy for all \bar{u}_2 , the coefficients of \bar{u}_2 must be equal to zero:

$$\left[E(1 + u_2^k)^2 + S_{11}^k \right] AL_0 \Delta u_2^k = F - S_{11}^k(1 + u_2^k)AL_0$$

After solving for the incremental displacement, the total displacement is updated by

$$u_2^{k+1} = u_2^k + \Delta u_2^k.$$

The iteration continues until the convergence criterion discussed in Chap. 2 is satisfied. Below is a MATLAB program that solves for the uniaxial bar using the total Lagrangian formulation. Table 3.1a shows the convergence history using the Newton–Raphson method. Note that the converged stress is 75.6 Pa, which is smaller than the infinitesimal assumption stress of 100 Pa. This happens due to the nonlinear displacement–strain relation.

It is interesting to note that without the initial stiffness term, $S_{11}AL\bar{u}_2\Delta u_2$, the Newton–Raphson method converges in the fifth iteration, as shown in Table 3.1b. However, it does not show a quadratic convergence. As discussed before, if there is an error in the Jacobian matrix, the algorithm may still converge but with a lower convergence rate.

Table 3.1 Convergence history of uniaxial bar using the total Lagrangian formulation

Iteration	<i>u</i>	Strain	Stress	conv
(a) With initial stiffness				
0	0.0000	0.0000	0.000	9.999E – 01
1	0.5000	0.6250	125.000	7.655E – 01
2	0.3478	0.4083	81.664	1.014E – 02
3	0.3252	0.3781	75.616	4.236E – 06
(b) Without initial stiffness				
0	0.0000	0.0000	0.000	9.999E – 01
1	0.5000	0.6250	125.000	7.655E – 01
2	0.3056	0.3252	70.448	6.442E – 03
3	0.3291	0.3833	76.651	3.524E – 04
4	0.3238	0.3762	75.242	1.568E – 05
5	0.3250	0.3770	75.541	7.314E – 07

```
%  
% Example 3.8 Uniaxial bar-total Lagrangian formulation  
%  
tol = 1.0e-5; iter = 0; E = 200;  
u = 0; uold = u; f = 100;  
strain = u + 0.5*u^2;  
stress = E*strain;  
P = stress*(1+u);  
R = f - P;  
conv= R^2/(1+f^2);  
fprintf('\n iter u1 E11 S11 conv');  
fprintf('\n %3d %7.5f %7.5f %8.3f %12.3e %7.5f',iter,u,strain,stress,  
conv);  
while conv > tol && iter < 20  
Kt = E*(1+u)^2 + stress;  
delu = R/Kt;  
u = uold + delu;  
strain = u + 0.5*u^2;  
stress = E*strain;  
P = stress*(1+u);  
R = f - P;  
conv= R^2/(1+f^2);  
uold = u;  
iter = iter + 1;  
fprintf('\n %3d %7.5f %7.5f %8.3f %12.3e %7.5f',iter,u,strain,stress,  
conv);  
end
```

3.3.2 Nonlinear Static Analysis: Updated Lagrangian Formulation

A spatial description or an updated Lagrangian formulation uses stress and strain measures, such as the Cauchy stress σ and the engineering strain ϵ , defined at the current geometry. Although the term engineering strain is used, it is different from the one in linear systems. In the updated Lagrangian formulation, the engineering strain is defined in the deformed geometry, which is under large deformation. Thus, the strain is not a linear function of displacement in this case. From the assumption that the applied loads are independent of deformation, the load form $\ell(\bar{\mathbf{u}})$ is the same as that of the total Lagrangian formulation in Eq. (3.66). Thus, only the energy form will be discussed in the following derivation. The updated Lagrangian formulation requires the relation between Cauchy stress and engineering strain. However, the original constitutive relation is given in terms of the second Piola-Kirchhoff stress and Lagrangian stress as in Eq. (3.53). The linear relation between \mathbf{S} and \mathbf{E} does not mean a linear relation between σ and ϵ (refer to Fig. 3.9). Thus, instead of developing a nonlinear relation between σ and ϵ , the linear relation between \mathbf{S} and \mathbf{E} will be used in the following derivations.

The relation between material tensors $(\mathbf{S}, \bar{\mathbf{E}})$ and spatial tensors $(\sigma, \bar{\epsilon})$ can be obtained through transformation as

$$\mathbf{S} = J\mathbf{F}^{-1}\sigma\mathbf{F}^{-T} \quad (3.77)$$

$$\bar{\mathbf{E}} = \mathbf{F}^T\epsilon(\bar{\mathbf{u}})\mathbf{F}, \quad (3.78)$$

where $\epsilon(\bar{\mathbf{u}})$ is the variation of the engineering strain at the current geometry. From the definition of the variation of Lagrangian strain in Eq. (3.63), the variation of engineering strain can be calculated by transforming the Lagrangian strain variation to the current geometry as¹

$$\begin{aligned} \epsilon(\bar{\mathbf{u}}) &= \mathbf{F}^{-T}\bar{\mathbf{E}}\mathbf{F}^{-1} \\ &= \mathbf{F}^{-T}\frac{1}{2}\left(\frac{\partial\bar{\mathbf{u}}^T}{\partial\mathbf{x}}\mathbf{F} + \mathbf{F}^T\frac{\partial\bar{\mathbf{u}}}{\partial\mathbf{x}}\right)\mathbf{F}^{-1} \\ &= \frac{1}{2}\left(\frac{\partial\bar{\mathbf{u}}^T}{\partial\mathbf{x}} + \frac{\partial\bar{\mathbf{u}}}{\partial\mathbf{x}}\right) \\ &\equiv \text{sym}(\nabla_n\bar{\mathbf{u}}) \end{aligned} \quad (3.79)$$

where $\nabla_n = \partial/\partial\mathbf{x}$ is the gradient operator at the current deformed geometry. Note that Eq. (3.79) is different from the variation of the engineering strain in linear systems with small deformations because it is defined in the current geometry under the assumption of large deformation. Even if $\bar{\mathbf{u}}$ is independent of displacement, the

¹This transformation is often called a “push-forward,” while the reverse transformation is called a “pull-back.”

denominator \mathbf{x} depends on displacement. In that sense, the above strain variation is a nonlinear function of displacement.

Although it is possible to derive the energy form for the updated Lagrangian formulation starting from a strain energy density that is defined in terms of engineering strain, it would be difficult to have equivalent strain energy densities that are in a quadratic form for both the Eulerian and engineering strains. In fact, it will be shown that for the same material, the constant constitutive tensor for the total Lagrangian formulation is not constant anymore in the updated Lagrangian formulation. Thus, instead of deriving an equivalent strain energy density for the updated Lagrangian formulation, the one for the total Lagrangian formulation is used in the following derivations. Using Eqs. (3.77) and (3.78), the energy form $a(\mathbf{u}, \bar{\mathbf{u}})$ in Eq. (3.65) can be expressed in terms of the spatial description as

$$\begin{aligned} a(\mathbf{u}, \bar{\mathbf{u}}) &= \iint_{^0\Omega} \mathbf{S} : \bar{\mathbf{E}} \, d\Omega = \iint_{^0\Omega} (\mathbf{J}\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-T}) : (\mathbf{F}^T\boldsymbol{\epsilon}(\bar{\mathbf{u}})\mathbf{F}) \, d\Omega \\ &= \iint_{^n\Omega} \boldsymbol{\sigma} : \boldsymbol{\epsilon}(\bar{\mathbf{u}}) \, d\Omega. \end{aligned} \quad (3.80)$$

In Eq. (3.80), the property $d^n\Omega = Jd^0\Omega$ has been used to change the integral domain from ${}^0\Omega$ to ${}^n\Omega$. Using the above definition of the energy form, the nonlinear variational equation for the updated Lagrangian formulation has the same form as Eq. (3.64), namely,

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (3.81)$$

However, the expression of $a(\mathbf{u}, \bar{\mathbf{u}})$ is different from that of Eq. (3.65).

The linearization of Eq. (3.81) involves Cauchy stress, which is not easy to linearize because the Cauchy stress is defined on current geometry that changes according to deformation. In addition, the integral domain also depends on deformation. These difficulties can be overcome by transforming the linearization in the material description, given in Eq. (3.74), to the spatial description. The integrands of Eq. (3.74) are transformed into the deformed geometry using the same “push-forward” method described in Eq. (3.80) as

$$\begin{aligned} \mathbf{S} : \Delta\bar{\mathbf{E}} &= J(\mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-T}) : \text{sym}(\nabla_0\bar{\mathbf{u}}^T\nabla_0\Delta\mathbf{u}) \\ &= J\boldsymbol{\sigma} : \text{sym}(\nabla_n\bar{\mathbf{u}}^T\nabla_n\Delta\mathbf{u}) \\ &\equiv J\boldsymbol{\sigma} : \boldsymbol{\eta}(\Delta\mathbf{u}, \bar{\mathbf{u}}) \end{aligned} \quad (3.82)$$

and

$$(\bar{\mathbf{E}} : \mathbf{D} : \Delta\mathbf{E}) = (\mathbf{F}^T\boldsymbol{\epsilon}(\bar{\mathbf{u}})\mathbf{F}) : \mathbf{D} : (\mathbf{F}^T\boldsymbol{\epsilon}(\Delta\mathbf{u})\mathbf{F}), \quad (3.83)$$

$$= J\boldsymbol{\epsilon}(\bar{\mathbf{u}}) : \mathbf{c} : \boldsymbol{\epsilon}(\Delta\mathbf{u}),$$

where \mathbf{c} is the fourth-order spatial constitutive tensor. The relation between the material and spatial constitutive tensors is given by

$$c_{ijkl} = \frac{1}{J} F_{ir} F_{js} F_{km} F_{ln} D_{rsmn}. \quad (3.84)$$

In the case of St. Venant–Kirchhoff nonlinear elastic material from Eq. (3.51), by using the left Cauchy–Green deformation tensor $\mathbf{G} = \mathbf{F} \cdot \mathbf{F}^T$, the spatial constitutive tensor becomes

$$c_{ijkl} = \frac{1}{J} [\lambda G_{ij} G_{kl} + \mu (G_{ik} G_{jl} + G_{il} G_{jk})]. \quad (3.85)$$

Note that \mathbf{c} is not constant for the updated Lagrangian formulation and that in addition $\boldsymbol{\sigma} \neq \mathbf{c} : \boldsymbol{\epsilon}$. Linearization of the energy form in the updated Lagrangian formulation can be obtained from Eqs. (3.82) and (3.83) as

$$L[a(\mathbf{u}, \bar{\mathbf{u}})] = \iint_{\Omega} [\boldsymbol{\epsilon}(\bar{\mathbf{u}}) : \mathbf{c} : \boldsymbol{\epsilon}(\Delta \mathbf{u}) + \boldsymbol{\sigma} : \boldsymbol{\eta}(\Delta \mathbf{u}, \bar{\mathbf{u}})] \, d\Omega \equiv a^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}). \quad (3.86)$$

The same notation $a^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}})$ is used with the total Lagrangian formulation so that the configuration implicitly depends on total displacement \mathbf{u} and $a^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}})$ is linear with respect to $\Delta \mathbf{u}$ and $\bar{\mathbf{u}}$.

If the current time step is t_n , the current iteration counter is k , and the external force is independent of displacement, then the linearized equation (3.81) can be obtained as

$$a^*({}^n \mathbf{u}^k; \Delta \mathbf{u}^k, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}) - a({}^n \mathbf{u}^k, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (3.87)$$

Note that when the above terms are calculated, the integration is performed over the domain based on deformation at the k th iteration. It is not the current geometry in a precise sense. However, as the iteration converges, the difference between two consecutive iterations becomes ignorable. Thus, at converged, the iteration satisfies the updated Lagrangian description.

Provided that appropriate constitutive relations are used, as in Eq. (3.84), the two linear formulations, Eqs. (3.75) and (3.87), are theoretically equivalent but with different expressions. The choice of method should depend on how effective the numerical implementation is and how convenient it is to generate the constitutive relation. For example, the strain measure of a total Lagrangian formulation is more complicated than that of the updated Lagrangian formulation. However, the constitutive relation in Eq. (3.51) can easily be used in the total Lagrangian formulation without transforming the constitutive relation into the deformed geometry, as in Eq. (3.84). In the case of elastoplasticity, the plastic evolution process always occurs at the current geometry. Because it is difficult to express plastic evolution in terms of material stress measures, the updated Lagrangian formulation is a more attractive option for this case.

With a small deformation problem, it is possible to approximate \mathbf{F} for 1. From this approximation and from their definitions, the two stress measures (\mathbf{S} and $\boldsymbol{\sigma}$) become identical, and the same is true for the two strain measures (\mathbf{E} and $\boldsymbol{\epsilon}$).

Example 3.9 (Uniaxial bar: updated Lagrangian formulation) Solve the uniaxial bar in Example 3.8 using the updated Lagrangian formulation. Assume that the change in cross-sectional area is ignored in one-dimensional bar.

Solution In the updated Lagrangian formulation, the current length $L = L_0 + u_2$ of the bar is used to define the finite element. The approximation of displacement gradient and its variation then become

$$\frac{du}{dx} = \frac{u_2}{1 + u_2}, \quad \frac{d\bar{u}}{dx} = \frac{\bar{u}_2}{1 + u_2}$$

For a one-dimensional bar, the deformation gradient and its determinant become a scalar and can be calculated by

$$F_{11} = \frac{dx}{dX} = 1 + u_2, \quad J = 1 + u_2$$

The second Piola-Kirchhoff stress S_{11} and the variation of Lagrangian strain E_{11} are used to calculate Cauchy stress and the variation of engineering strain as

$$\begin{aligned}\sigma_{11} &= \frac{1}{J} F_{11} S_{11} F_{11} = E \left(u_2 + \frac{1}{2} u_2^2 \right) (1 + u_2) \\ \varepsilon_{11}(\bar{u}) &= F_{11}^{-T} \bar{E}_{11} F_{11}^{-1} = \frac{\bar{u}_2}{1 + u_2}\end{aligned}$$

Using the above two expressions, the energy form in the updated Lagrangian formulation becomes

$$a(u, \bar{u}) = \int_0^L \sigma_{11} \varepsilon_{11}(\bar{u}) A dx = \sigma_{11} A \bar{u}_2$$

Note that the relation $L = 1 + u_2$ is used in the above integration. Although the above energy seems different from that of the total Lagrangian formulation in Example 3.8, it is identical because $\sigma_{11} = S_{11}(1 + u_2)$. Since the applied load is a concentrated force at the tip, no integration is required. The load form simply becomes

$$\ell(\bar{u}) = \bar{u}_2 F$$

The difference between the energy and load forms can be defined as a residual R , and the nonlinear variational equation satisfies when the residual becomes zero:

$$R = \bar{u}_2 (\sigma_{11} A - F) = 0$$

for any arbitrary \bar{u}_2 .

In order to linearize the energy form, the transformation relations in Eqs. (3.82) and (3.83) are used. First, the spatial constitutive tensor \mathbf{c} has a single component as

$$c_{1111} = \frac{1}{J} F_{11} F_{11} F_{11} F_{11} E = (1 + u_2)^3 E$$

Thus,

$$\int_0^L \varepsilon_{11}(\bar{u}) c_{1111} \varepsilon_{11}(\Delta u) A dx = EA(1 + u_2)^2 \bar{u}_2 \Delta u_2$$

$$\int_0^L \sigma_{11} \eta_{11}(\Delta u, \bar{u}) A dx = \frac{\sigma_{11} A}{1 + u_2} \bar{u}_2 \Delta u_2$$

Thus, the linearization of the energy form becomes

$$a^*(u; \Delta u, \bar{u}) = \int_0^L (\varepsilon_{11}(\bar{u}) c_{1111} \varepsilon_{11}(\Delta u) + \sigma_{11} \eta(\Delta u, \bar{u})) A dx$$

$$= EA(1 + u_2)^2 \bar{u}_2 \Delta u_2 + \frac{\sigma_{11}}{1 + u_2} A \bar{u}_2 \Delta u_2$$

Again, the above linearization of the energy form is identical to that of the total Lagrangian formulation by considering the relation between σ_{11} and S_{11} . The Newton–Raphson iteration scheme is identical to that of the total Lagrangian formulation. Below is the list of MATLAB program that solves for the uniaxial bar using the updated Lagrangian formulation. Table 3.2 shows the convergence history using the Newton–Raphson method. Note that the history of displacement is identical to that of the total Lagrangian formulation because the same constitutive relation is used for both formulations. However, the histories of strain and stress are different because different measures of stress and strain are used. Note that the stress converges at 100 Pa, which is the true stress ($F = 100$ N and $A = 1.0$ m²). The engineering strain is smaller than the Lagrangian strain as it is defined with respect to the current length, and the latter has a higher-order term.

Table 3.2 Convergence history of uniaxial bar using the updated Lagrangian formulation

Iteration	u	Strain	Stress	conv
0	0.0000	0.0000	0.000	9.999E – 01
1	0.5000	0.3333	187.500	7.655E – 01
2	0.3478	0.2581	110.068	1.014E – 02
3	0.3252	0.2454	100.206	4.236E – 06

```

%
% Example 3.9 Uniaxial bar-updated Lagrangian formulation
%
tol = 1.0e-5; iter = 0; E = 200;
u = 0; uold = u; f = 100;
strain = u/(1+u);
stress = E*(u+.5*u^2)*(1+u);
P = stress;
R = f - P;
conv= R^2/(1+f^2);
fprintf('\n niter   u1    E11    S11      conv');
fprintf('\n %3d  %7.5f  %7.5f %8.3f %12.3e %7.5f', iter, u, strain, stress,
conv);
while conv > tol && iter < 20
Kt = E*(1+u)^2 + stress/(1+u);
delu = R/Kt;
u = uold + delu;
strain = u/(1+u);
stress = E*(u+.5*u^2)*(1+u);
P = stress;
R = f - P;
conv= R^2/(1+f^2);
uold = u;
iter = iter + 1;
fprintf('\n %3d  %7.5f  %7.5f %8.3f %12.3e %7.5f', iter, u, strain, stress,
conv);
end

```

3.4 Critical Load Analysis

As shown in Fig. 3.18, structural instability occurs when the load–displacement curve has a negative slope; i.e., displacement rapidly increases without increase of the applied loads. The force-controlled method cannot find a solution because either the structure cannot support the loads beyond a certain level or the next stable state is too far from the current state. Even if the displacement-controlled method can help convergence, it cannot find an appropriate solution in many cases especially when the system has many degrees of freedom. In practice, since instability is not a preferred state, it is not important for the analysis to follow through this state. Instead, it is often more important to predict when the instability initiates, i.e., what maximum load the structure can support before becoming unstable. This maximum load is called a critical load.

From the one-dimensional load–displacement curve, such as the one in Fig. 3.18, the slope of the tangent stiffness becomes zero at the critical load. For systems with many degrees of freedom, the tangent stiffness matrix becomes singular. Thus, based on the incremental equilibrium equation in the previous section, the variational equation of a linear eigenvalue problem can be formulated for stability

analysis in nonlinear structural systems to evaluate the stability status. [6, 7] These problems differ in the assumption made between a critical load factor and the estimated critical load. The critical load factor of a nonlinear structural system can be evaluated by solving a linear eigenvalue problem at any pre-buckling equilibrium configuration. The formulation of a stability analysis may include the effect of large displacements, large rotations, large strains, and material nonlinearities with appropriate kinematic and constitutive descriptions. The critical load can be found by using the property that at least two adjacent states exist at the critical load. [8] The mathematical basis for such a situation immediately follows from incremental equilibrium equations; that is, the left-hand side of Eq. (3.75) or Eq. (3.87) vanishes at the critical limit point. For a single-variable problem, the slope of the displacement–load curve becomes zero. To make it easier to follow the derivations, the linearized energy form in Eq. (3.74) can be divided into two parts:

$$A(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) = \iint_{\Omega} (\bar{\mathbf{E}} : \mathbf{D} : \Delta\mathbf{E}) \, d\Omega \quad (3.88)$$

and

$$G(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) = - \iint_{\Omega} (\mathbf{S} : \Delta\bar{\mathbf{E}}) \, d\Omega. \quad (3.89)$$

With the critical displacement $^{cr}\mathbf{u}$ at the critical limit point $t_n = t^{cr}$, the stability equation becomes

$$a^*(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) \equiv A(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) - G(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) = 0, \quad \forall \bar{\mathbf{y}} \in \mathbb{Z} \quad (3.90)$$

Note that eigen-function \mathbf{y} and its variation $\bar{\mathbf{y}}$ are used instead of incremental displacement $\Delta\mathbf{u}$ and its variation. Since the right-hand side is zero, if the tangent stiffness matrix after discretizing $a^*(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}})$ is positive definite, a trivial solution $\mathbf{y} = \mathbf{0}$ is expected. Nontrivial eigen-function \mathbf{y} is expected only when the tangent stiffness matrix becomes singular, which serves to identify a point of instability; that is, if the final equilibrium state is at the critical limit point, then solution \mathbf{y} must be nontrivial.

Note that Eq. (3.90) cannot be solved from the start because it is based on the displacement $^{cr}\mathbf{u}$, which is unknown. In this case, the incremental force method is useful because the applied loads gradually increase until the instability occurs. However, it does not guarantee that the critical load and the corresponding displacement can be found. It is assumed that the incremental force method is performed up to load step t_n at which the corresponding load is $^n p$. Since the incremental force method cannot converge beyond the critical load, it is clear that t_n is pre-buckling state, i.e., $^n p \leq p_{cr}$. In order to find the critical load, it is first represented by a scalar multiple of the pre-buckling load $^n p$, i.e., $p_{cr} = \zeta^n p$, where the scalar ζ is called the critical load factor. The left-hand side of Eq. (3.90) is then evaluated at the critical limit point using the information that is available at the final

pre-buckling equilibrium state at load step t_n . Linear extrapolation can be used to approximate the left-hand side of Eq. (3.90) to form an eigenvalue problem. After solving the eigenvalue problem, the lowest eigenvalue is considered to be an estimate ${}^n\zeta$ of the critical load factor ζ . Assuming a proportional conservative static loading, the estimated critical load vector ${}^n p_{cr}$ can be expressed with the given load vector ${}^n p$ and the critical load factor ${}^n\zeta$. Two commonly used approaches, one- and two-point linear eigenvalue approaches, are formulated in the variational form, and expressions of the corresponding estimated critical load are presented in the following subsections.

3.4.1 One-Point Approach

The goal of the one-point approach is to approximate the critical state using the state at load t_n . Utilizing the information at the equilibrium state at load step t_n , Eq. (3.90) can be rewritten as an eigenvalue problem. By linearizing the nonlinear relation between energy form G and the additional load increment at the critical limit point, G is approximated using the critical load factor ${}^n\zeta$ at load t_n :

$$G({}^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) = {}^n\zeta G({}^n\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}). \quad (3.91)$$

Referring to Eq. (3.89), the stress ${}^{cr}\mathbf{S}$ at the critical state is approximated by ${}^n\zeta {}^n\mathbf{S}$. In addition, by neglecting the variations of energy form A due to the loading change, $A({}^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}})$ can be written as $A({}^n\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}})$. Then, Eq. (3.90) becomes an eigenvalue problem, which can be called a one-point linear eigenvalue problem, in the form of

$$A({}^n\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) - {}^n\zeta G({}^n\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) = 0, \quad \forall \bar{\mathbf{y}} \in \mathbb{Z}. \quad (3.92)$$

Solving this eigenvalue problem at a given load level ${}^n p$ that is lower than the true critical load p_{cr} leads to the following estimated critical load:

$${}^n p_{cr} = {}^n\zeta {}^n p. \quad (3.93)$$

Note that the above estimate will be accurate if the load ${}^n p$ is close to the actual critical load. Thus, it is important to make the load increment small so that the final pre-buckling load is close enough. However, since p_{cr} is unknown, small load increment will be computationally expensive.

3.4.2 Two-Point Approach

Utilizing the information at the last two states, at load step t_{n-1} and t_n where t_n is the load step at the final equilibrium state, Eq. (3.90) can be rewritten as an eigenvalue problem for the two-point approach. With the assumption that from load step t_{n-1}

onward the energy form $(A - G)$ changes linearly up to one additional load increment but with the same ratio between the two states at t_{n-1} and t_n , the energy form can be written as

$$A(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) - G(^{cr}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) = B(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) + \zeta E(^{n-1}\mathbf{u}^n \mathbf{z}; \mathbf{y}, \bar{\mathbf{y}}), \quad (3.94)$$

where the energy forms B and E are defined as

$$B(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) \equiv A(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) - G(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) \quad (3.95)$$

$$E(^{n-1}\mathbf{u}^n \mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) \equiv B(^n\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) - B(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}), \quad (3.96)$$

and ζ is the critical load factor at time t_n . Consequently, Eq. (3.94) becomes a two-point linear eigenvalue problem, written as

$$B(^{n-1}\mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) + \zeta E(^{n-1}\mathbf{u}^n \mathbf{u}; \mathbf{y}, \bar{\mathbf{y}}) = 0, \quad \forall \bar{\mathbf{y}} \in \mathbb{Z}. \quad (3.97)$$

Solving this problem at a given load level ${}^n p$ that is lower than the true critical load p_{cr} leads to an estimated critical load

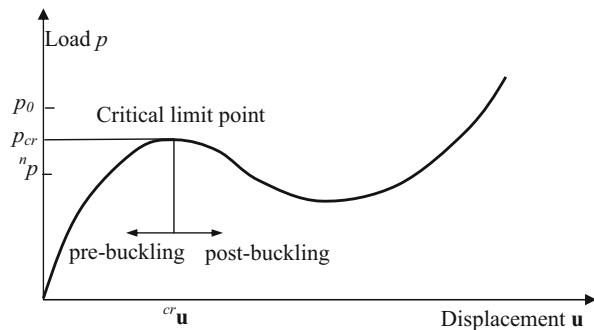
$$p_{cr} = {}^{n-1}p + \zeta \Delta p. \quad (3.98)$$

In Eqs. (3.92) and (3.97), $\zeta \geq 1$ is the smallest eigenvalue, and \mathbf{y} is the corresponding eigen-function. Note that if the load step t_n is at the critical limit point, then $\zeta = 1$ in Eqs. (3.92) and (3.97), and these two equations become identical to Eq. (3.90). The stability analysis of Eq. (3.92) or Eq. (3.97) can be applied to any pre-buckling configurations, and the estimated critical load becomes more accurate as the final equilibrium configuration approaches the critical limit point. The estimated critical loads for both approaches are not conservative, that is, they are larger than the true critical load ${}^n p_{cr} \geq p_{cr}$. A stability analysis equation for linear structural systems can be obtained as a special case of the nonlinear stability equation, with the assumption of linearly elastic material and a small displacement.

3.4.3 Stability Equation with Actual Critical Load Factor

Consider a structural system with the equilibrium path shown in Fig. 3.12. The critical limit point in Fig. 3.12 is a relative maximum point in the nonlinear load-displacement curve and defines the boundary between the pre-buckling and the post-buckling equilibrium paths. Assume that the magnitude of the total applied load vector p_0 is larger than the magnitude of the critical load vector p_{cr} and that unlike the previous cases, the critical load vector p_{cr} occurs at the final pre-buckling equilibrium state at load step $t_n = t_{cr}$, i.e., $p_{cr} = {}^n p$. Note that the load vectors p_{cr} , p_0 , and ${}^n p$ have the same directions, since they are assumed to be proportional loadings. The magnitude of the critical load is unknown before the system is analyzed.

Fig. 3.12 Equilibrium path of nonlinear structural system



With the energy form $a(\mathbf{u}, \bar{\mathbf{u}})$ in the final pre-buckling equilibrium configuration at time $t_n = t_{cr}$ and the load form $\ell(\bar{\mathbf{u}})$ which is the virtual work done by the total applied load vector p_0 , the equilibrium equation (3.64) can be rewritten as

$$a(^n\mathbf{u}, \bar{\mathbf{u}}) = \beta \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (3.99)$$

where the actual critical load factor β is defined as the ratio of the magnitude of the critical load vector $p_{cr} = ^n p$ to the magnitude of the total applied load vector p_0 , that is,

$$p_{cr} = \beta p_0. \quad (3.100)$$

The actual critical load factor $\beta \leq 1.0$ can be evaluated only after the critical load is known. When the total applied load vector p_0 is equal to the critical load vector $p_{cr} = ^n p$, $\beta = 1$ and the equilibrium equation (3.99) is the same as the equilibrium equation (3.64). Note that the actual critical load factor β does not depend on the configuration at t_n and is not related to the estimated critical load factor $^n \zeta$ which varies with configuration t_n .

3.5 Hyperelastic Materials

As has been shown in Sect. 3.3, material behavior is described by its constitutive relation when subjected to deformation or deformation history. Different constitutive relations can represent different material behaviors. The St. Venant–Kirchhoff material in Sect. 3.3 provides a linear relation between stress and strain, which is a simple extension of the one used for linear elastic materials. Unfortunately, this model provides meaningful results only when the strains are small because most materials show a nonlinear relation for large deformation. It is important to employ a constitutive model that is appropriate for the material, the structure, and the finite element formulation.

When the material status can completely be describable with a given total strain, the constitutive relation is called hyperelasticity. In such a material, a strain energy

density exists as a function of strain, and stress can be obtained by differentiating the strain energy density with respect to strain. This material model is independent of deformation history; i.e., the same deformation is expected if the final load is the same. Rubberlike materials or human tissues belong in this category. On the other hand, when the constitutive relation is given in terms of the stress and strain rates, it is called hypoelasticity, which is often used in describing the plastic behavior of materials. In such a material, the deformation history must be followed to calculate stress because two states that have the same strain may have different stresses depending on the loading history. Hyperelastic materials will be discussed in this section, while hypoelastic materials will be discussed in Chap. 4. In this section, the static response of hyperelastic materials is formulated based on the material description, i.e., the total Lagrangian formulation. Different hyperelastic materials will be introduced, but the main development will be explained using the Mooney-Rivlin material, which is the most popular material model. In general, hyperelastic materials exhibit the property of being incompressible during finite deformation; i.e., the volume of the material is preserved. This is a common behavior of many elastic materials in finite deformation. Numerically, a constraint, $J = 1$, needs to be imposed to make the material incompressible. However, this causes numerical difficulties called volumetric locking. Due to incompressibility, the hydrostatic portion of stress cannot be calculated from the volumetric strain. The penalty method [9], the selective reduced integration method [10], and the mixed formulation method [11] have been successfully used for incompressible and nearly incompressible materials.

If a strain energy density exists, such that stress can be obtained from the derivative of the strain energy with respect to strain, the system is called path independent. Thus, it is theoretically possible to solve the nonlinear equilibrium equation for the given total magnitude of applied load. However, this equation is still solved by using the incremental force method with a number of load steps to finally reach the total applied load for computational purposes. The hyperelasticity problem contains both material nonlinearity from constitutive relations and geometric nonlinearity from kinematics.

3.5.1 Strain Energy Density

Modeling engineering materials at large deformations is still an active research area. Without elaborating details of material modeling procedures, a method that can describe the behavior of isotropic elastic materials which undergo finite deformation is presented. In hyperelasticity, the stored strain energy density is used to compute stress. For isotropic materials, the constitutive relation has to be independent of the coordinate frame selected because the material has the same property for all directions. For example, the strain component E_{11} cannot be used for the constitutive relation because its value depends on the coordinate system. Thus, it is natural that the strain energy density is defined using invariants of strain or

alternatively that of the deformation tensor. When the undeformed state is used as the frame of reference, the three invariants of the right Cauchy–Green deformation tensor \mathbf{C} in Eq. (3.9) are given as

$$I_1 = \text{tr}(\mathbf{C}) = \lambda_1^2 + \lambda_2^2 + \lambda_3^2, \quad (3.101)$$

$$I_2 = \frac{1}{2} \left[(\text{tr} \mathbf{C})^2 - \text{tr}(\mathbf{C}^2) \right] = \lambda_1^2 \lambda_2^2 + \lambda_2^2 \lambda_3^2 + \lambda_3^2 \lambda_1^2, \quad (3.102)$$

and

$$I_3 = \det \mathbf{C} = \lambda_1^2 \lambda_2^2 \lambda_3^2, \quad (3.103)$$

where λ_1^2 , λ_2^2 , and λ_3^2 are three eigenvalues of the left Cauchy–Green deformation tensor \mathbf{C} . From polar decomposition, it has been shown that λ_1 , λ_2 , and λ_3 are three eigenvalues of the left stretch tensor \mathbf{U} —also called the principal stretches. The above three invariants will remain unchanged for different coordinate systems. In order to be a valid deformation, the three invariants must be positive (refer to Example 3.1). The square root of I_3 in Eq. (3.103) measures the volume change of the material. If the material is incompressible, it is clear that $I_3 = 1$. The three invariants are identical for both the left and right Cauchy–Green deformation tensors. When there is no deformation, i.e., $\lambda_1 = \lambda_2 = \lambda_3 = 1$, $I_1 = I_2 = 3$, and $I_3 = 1$.

Example 3.10 (Invariants) Show that the three invariants of the left Cauchy–Green deformation tensor \mathbf{G} are equal to those of \mathbf{C} when the three eigenvalues of the deformation gradient are λ_1 , λ_2 , and λ_3 .

Solution The three invariants will remain constant for different coordinate systems. Thus, it is possible to choose the three principal directions of the deformation gradient as basis vectors for the new coordinate system $X'Y'Z'$ so that the deformation gradient will only have diagonal components:

$$\mathbf{F}_{X'Y'Z'} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Then the right and left Cauchy–Green deformation tensors become identical

$$\mathbf{C}_{X'Y'Z'} = (\mathbf{F}^T \mathbf{F})_{X'Y'Z'} = \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix}, \quad \mathbf{G}_{X'Y'Z'} = (\mathbf{F} \mathbf{F}^T)_{X'Y'Z'} = \begin{bmatrix} \lambda_1^2 & 0 & 0 \\ 0 & \lambda_2^2 & 0 \\ 0 & 0 & \lambda_3^2 \end{bmatrix},$$

and the three invariants of the two tensors are identical. ■

Using the three invariants, a general form of strain energy density can be defined in the following polynomials:

$$W(I_1, I_2, I_3) = \sum_{m+n+k=1}^{\infty} A_{mnk} (I_1 - 3)^m (I_2 - 3)^n (I_3 - 1)^k, \quad (3.104)$$

where A_{mnk} are coefficients of polynomials. In general, deformation of a material can be decomposed into volumetric and distortional parts. If the material is incompressible, i.e., $I_3 = 1$, then the volumetric part of the strain energy density is eliminated, and only the first two terms contribute to the strain energy density. This part of the stored energy is called the distortional strain energy density and is defined as

$$W_1(I_1, I_2) = \sum_{m+n=1}^{\infty} A_{mn} (I_1 - 3)^m (I_2 - 3)^n. \quad (3.105)$$

Note that Eq. (3.105) does not impose the incompressibility condition. A separate constraint must be used to make the material incompressible. All the models listed above account for nonconstant shear modulus. However, caution needs to be exercised on inclusion of higher-order terms to fit the data, since this may result in unstable energy functions, yielding nonphysical results outside the range of the experimental data. Section 3.9 will discuss about how to find hyperelastic material parameters by fitting experimental data. Various hyperelastic material models are proposed using Eq. (3.105). Some examples are as follows.

Neo-Hookean model: This model has only one nonzero parameter, A_{10} , and all other parameters are zero. Using the undeformed state as a frame of reference, the strain energy density can be defined as

$$W_1(I_1) = A_{10}(I_1 - 3). \quad (3.106)$$

In order to be equivalent to the linear elastic material in small deformation, the parameter A_{10} is related to the shear modulus by $A_{10} = \mu/2$. The stress-strain relation becomes linear with a proportional constant of $2A_{10} = \mu$. However, this model will show a nonlinear relationship when the deformation becomes larger due to the nonlinear displacement-strain relation. This model gives a good correlation with the experimental data up to 40 % strain in uniaxial tension and up to 90 % strains in simple shear. This model is often used to describe the behavior of cross-linked polymers.

Mooney-Rivlin model: This model is an extension of the Neo-Hookean model by including the effect of the second invariant. The expression of the strain energy density has two nonzero parameters as

$$W_1(I_1, I_2) = A_{10}(I_1 - 3) + A_{01}(I_2 - 3) \quad (3.107)$$

This model is the most popular in finite element analysis of hyperelastic materials, not because of accuracy but because of its simplicity. The Mooney-Rivlin model is

good up to 100 % strain of tensile test, but has some difficulty in describing the compression mode of deformation. Moreover, the Mooney-Rivlin model fails to account for the hardening of the material at large strains.

Yeoh model: This model only uses I_1 and the experimental results are fitted using a cubic function as

$$W_1(I_1) = A_{10}(I_1 - 3) + A_{20}(I_1 - 3)^2 + A_{30}(I_1 - 3)^3. \quad (3.108)$$

This model corresponds well with experiments for large strain. Since only the first invariant is used, the Yeoh model is often called the reduced polynomial model.

Ogden model: This model uses the principal stretches (the three eigenvalues of the deformation gradient) to define the distortional strain energy density as

$$W_1(\lambda_1, \lambda_2, \lambda_3) = \sum_{i=1}^N \frac{\mu_i}{\alpha_i} (\lambda_1^{\alpha_i} + \lambda_2^{\alpha_i} + \lambda_3^{\alpha_i} - 3) \quad (3.109)$$

where N , μ_i , and α_i are material parameters. N usually goes up to three. Note that when the material is incompressible, the three principal stretches are not independent, i.e., $\lambda_1\lambda_2\lambda_3 = 1$. By comparing with the linear elastic material, the initial shear modulus can be obtained by

$$\mu = \frac{1}{2} \sum_{i=1}^N \alpha_i \mu_i \quad (3.110)$$

When $N = 1$ and $\alpha_1 = 1$, it becomes the Neo-Hookean material. When $N = 2$, $\alpha_1 = 2$, and $\alpha_2 = -2$, it becomes the Mooney-Rivlin material. The model gives a good correlation with test data in simple tension up to 700 %. The model accommodates nonconstant shear modulus and slightly compressible material behavior. As strain increases, the material shows hardening when $\alpha > 2$, while softening when $\alpha < 2$.

Example 3.11 (Stress-strain relationship for Neo-Hookean model) Plot the nominal stress-strain relationship for a Neo-Hookean model under uniaxial tension and compression and compare it with linear elastic material with the same modulus. Assume material parameter $A_{10} = 10 \text{ MPa}$ and incompressibility.

Solution Let us suppose that a uniaxial load is stretched so that $\lambda_1 = \lambda$ where λ is an arbitrary stretch along the rod's length. From the assumption of incompressibility, $\lambda_1\lambda_2\lambda_3 = 1$ and $\lambda_2 = \lambda_3 = 1/\sqrt{\lambda}$. Therefore, $\lambda_2 = \lambda_3 = 1/\sqrt{\lambda}$. From Eq. (3.106), the strain energy density of the Neo-Hookean material model becomes

$$W = A_{10}(I_1 - 3) = A_{10}(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3) = A_{10}\left(\lambda^2 + \frac{2}{\lambda} - 3\right)$$

The nominal stress (first Piola-Kirchhoff stress) in the direction of stretch can be obtained by differentiating the strain energy density with respect to the principal stretch as

$$P = \frac{\partial W}{\partial \lambda} = 2A_{10} \left(\lambda - \frac{1}{\lambda^2} \right) = \mu \left(1 + \varepsilon - \frac{1}{(1 + \varepsilon)^2} \right)$$

Figure 3.13 shows the stress-strain curve for the Neo-Hookean material, along with a linear elastic material. Since Poisson's ratio for the incompressible material is 0.5, Young's modulus will be $E = 3\mu$. Both curves share the same tangent line at $\varepsilon = 0$, but the error increases as the strain increases. One thing to note is that the Neo-Hookean model shows a quite different behavior in compression from the linear elastic material behavior. ■

Example 3.12 (Relation between Mooney-Rivlin and Ogden models) Write the material parameters μ_1 and μ_2 in the Ogden model in terms of A_{10} and A_{01} in the Mooney-Rivlin model. Use $N = 2$, $\alpha_1 = 2$, and $\alpha_2 = -2$ for the Ogden model.

Solution From Eq. (3.109), the Ogden model with two terms can be rewritten as

$$W_1(\lambda_1, \lambda_2, \lambda_3) = \frac{\mu_1}{2} (\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3) - \frac{\mu_2}{2} (\lambda_1^{-2} + \lambda_2^{-2} + \lambda_3^{-2} - 3)$$

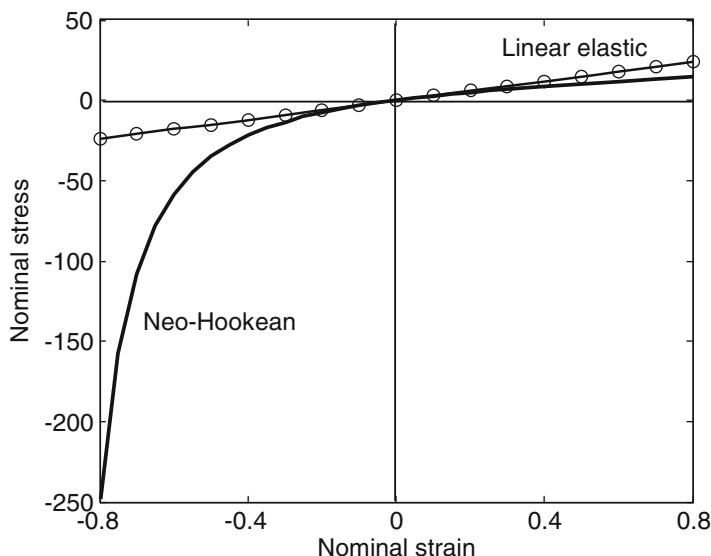


Fig. 3.13 Stress-strain relationship for Neo-Hookean material

From the definition of the invariants, the first term is identical to the first term of the Mooney-Rivlin material. Thus, $\mu_1 = 2A_{10}$. For the second term, using the incompressibility condition $\lambda_1\lambda_2\lambda_3 = 1$, it can be rewritten as

$$\lambda_1^{-2} + \lambda_2^{-2} + \lambda_3^{-2} = \lambda_2^2\lambda_3^2 + \lambda_1^2\lambda_3^2 + \lambda_1^2\lambda_2^2 = I_2.$$

Thus, the second term is also identical to the second term of the Mooney-Rivlin material. Therefore, $\mu_2 = -2A_{01}$. From Eq. (3.110), the equivalent shear modulus for small deformation can be calculated by

$$\mu = \frac{1}{2}(\alpha_1\mu_1 + \alpha_2\mu_2) = \mu_1 - \mu_2 = 2(A_{10} + A_{01}).$$

■

Example 3.13 (Strain energy density for St. Venant–Kirchhoff material) Show that St. Venant–Kirchhoff material has the following strain energy density:

$$W(\mathbf{E}) = \frac{\lambda}{2}[\text{tr}(\mathbf{E})]^2 + \mu[\text{tr}(\mathbf{E}^2)] \quad (3.111)$$

by deriving the material constitutive tensor $\mathbf{D} = \lambda\mathbf{1} \otimes \mathbf{1} + 2\mu\mathbf{I}$.

Solution First, the second Piola-Kirchhoff stress is calculated by differentiating the strain energy density with respect to the Lagrangian strain:

$$\mathbf{S} = \frac{\partial W(\mathbf{E})}{\partial \mathbf{E}} = \lambda \text{tr}(\mathbf{E}) \frac{\partial \text{tr}(\mathbf{E})}{\partial \mathbf{E}} + \mu \frac{\partial \text{tr}(\mathbf{E}^2)}{\partial \mathbf{E}}$$

For the first term on the right-hand side, the following properties are used: $\text{tr}(\mathbf{E}) = \mathbf{1} : \mathbf{E}$ and

$$\frac{\partial \text{tr}(\mathbf{E})}{\partial \mathbf{E}} = \mathbf{1} \quad \Rightarrow \quad \lambda \text{tr}(\mathbf{E}) \frac{\partial \text{tr}(\mathbf{E})}{\partial \mathbf{E}} = \lambda \mathbf{1} (\mathbf{1} : \mathbf{E}) = \lambda (\mathbf{1} \otimes \mathbf{1}) : \mathbf{E}.$$

For the second term, it is convenient to use index notation to derive the following relation:

$$\begin{aligned} \frac{\partial E_{ij}E_{ji}}{\partial E_{kl}} &= \delta_{ik}\delta_{jl}E_{ji} + E_{ij}\delta_{jk}\delta_{il} = E_{lk} + E_{lk} = 2E_{lk} \\ \Rightarrow \quad \frac{\partial \text{tr}(\mathbf{E}^2)}{\partial \mathbf{E}} &= 2\mathbf{E} = 2\mathbf{I} : \mathbf{E} \end{aligned}$$

Since \mathbf{I} is the fourth-order unit tensor, contraction with any second-order tensor will yield the tensor itself. Thus, the second Piola-Kirchhoff stress can be written as

$$\begin{aligned}\mathbf{S} &= \lambda \text{tr}(\mathbf{E}) \frac{\partial \text{tr}(\mathbf{E})}{\partial \mathbf{E}} + \mu \frac{\partial \text{tr}(\mathbf{E}^2)}{\partial \mathbf{E}} \\ &= \lambda(\mathbf{1} \otimes \mathbf{1}) : \mathbf{E} + 2\mu\mathbf{E} \\ &= [\lambda(\mathbf{1} \otimes \mathbf{1}) + 2\mu\mathbf{I}] : \mathbf{E} \\ &= \mathbf{D} : \mathbf{E}.\end{aligned}$$

Note that the material constitutive tensor \mathbf{D} is identical to that in Eq. (3.51). ■

3.5.2 *Nearly Incompressible Hyperelasticity*

Incompressibility of a material can cause many difficulties in the constitutive relation, especially when it is combined with nonlinearities such as large displacements, large strains, and contact. Perfect incompressibility, which corresponds to Poisson's ratio of one-half, is an idealization to make modeling more amenable for obtaining closed-form solutions. In the real world, natural as well as filled rubbers are slightly compressible, thereby, facilitating development of algorithms for nearly incompressible behavior of elastomers. "Near-incompressibility" means that Poisson's ratio is not exactly one-half, but close to it. For example, 0.49 or higher values are often used for the nearly incompressible behavior of elastomers.

As discussed previously, the hydrostatic pressure portion of stress causes volume change (dilatation). However, if the material is incompressible, the volume remains constant for different values of pressure. In other words, stress cannot be obtained by differentiating the strain energy density because the hydrostatic pressure portion of stress cannot be determined from deformation.

It has been observed from experiments that many rubberlike materials show nearly incompressible properties. It means that only a small volume change occurs under a large hydrostatic pressure. In such materials, the near-incompressibility can be imposed by using a large bulk modulus, which relates hydrostatic pressure to volumetric strain. Since the material is stiff in dilatation and soft in distortion, it is necessary to separate these two parts in order to reduce numerical difficulties associated with a large difference in stiffness. This has to be done in the level of strain energy density.

In the previous section, it is discussed that the third invariant I_3 is related to dilatation, while the other two invariants, I_1 and I_2 , are related to distortion. However, I_1 and I_2 do not remain constant during dilatation (see Example 3.13).

In order to separate the distortion part from dilatation, it is necessary to introduce the so-called reduced invariants, J_1 , J_2 , and J_3 , defined by

$$J_1 = I_1 I_3^{-1/3}, \quad J_2 = I_2 I_3^{-2/3}, \quad J_3 = I_3^{1/2}, \quad (3.112)$$

where I_1 , I_2 , and I_3 are the three invariants of the right Cauchy–Green deformation tensor \mathbf{C} . It can be easily verified that J_1 and J_2 are constant under pure dilatation; they are only related to distortion, while J_3 is related to dilatation. Of course, when the material is purely incompressible, the reduced invariants are the same with the invariants of \mathbf{C} .

Using the reduced invariants, it is possible to separate the distortion effect from dilatation in defining the strain energy density, as

$$W(J_1, J_2, J_3) = W_1(J_1, J_2) + W_2(J_3), \quad (3.113)$$

where $W_1(J_1, J_2)$ is the distortional strain energy density and $W_2(J_3)$ is the dilatational strain energy density. The distortional energy density can be defined using Eq. (3.105) by substituting the reduced invariants for the original invariants. An example of the dilatational energy density is related to the bulk modulus of the material as

$$W_2(J_3) = \frac{K}{2}(J_3 - 1)^2, \quad (3.114)$$

where K is the bulk modulus. The relationship between the bulk modulus and Lame's constants for an isotropic material can be written as

$$K = \lambda + \frac{2}{3}\mu. \quad (3.115)$$

The above relation is valid for linear elastic materials. For general nearly incompressible materials, a large value of the bulk modulus is used—two or three orders of magnitude larger than material parameters in the distortional part. The material becomes incompressible as the bulk modulus approaches infinity.

Example 3.14 (Dilatation) A dilatation status is defined by a constant deformation for every direction. In terms of deformation, this status can be represented by

$$x_1 = \lambda X_1, \quad x_2 = \lambda X_2, \quad x_3 = \lambda X_3,$$

where $\lambda > 0$ is the stretch ratio. It is clear that all deformation is volumetric and there is no shear deformation. Show that $W_1(J_1, J_2)$ in the form of Eq. (3.105) vanishes in such a volumetric deformation.

Solution First, the deformation gradient and the right Cauchy–Green deformation tensor are defined by

$$\mathbf{F} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}, \quad \mathbf{C} = \mathbf{F}^T \mathbf{F} = \begin{bmatrix} \lambda^2 & 0 & 0 \\ 0 & \lambda^2 & 0 \\ 0 & 0 & \lambda^2 \end{bmatrix}.$$

It is clear that the three principle stretches are all the same and have the value of λ^2 . Thus, the three invariants of the deformation tensor can be calculated as

$$\begin{aligned} I_1 &= \lambda^2 + \lambda^2 + \lambda^2 = 3\lambda^2 \\ I_2 &= 3\lambda^4 \\ I_3 &= \lambda^6. \end{aligned}$$

Note that all three invariants are not constant under dilatation. Using the formula in Eq. (3.112), the reduced invariants becomes

$$J_1 = \frac{3\lambda^2}{\lambda^2} = 3, \quad J_2 = \frac{3\lambda^4}{\lambda^4} = 3, \quad J_3 = \lambda^3.$$

Note that J_1 and J_2 are constant under pure dilatation. Thus, the distortional part of the strain energy density for volumetric change becomes

$$W_1(J_1, J_2) = 0,$$

■

3.5.2.1 Mooney-Rivlin Material Model

The Mooney-Rivlin model is one of the most popular models that are often used for modeling rubberlike materials. Although a very simple material model is discussed here, the method can be extended to general hyperelastic material models. With the near-incompressibility constraint, the strain energy density is defined using the reduced invariants as

$$\begin{aligned} W(J_1, J_2, J_3) &= A_{10}(J_1 - 3) + A_{01}(J_2 - 3) + \frac{K}{2}(J_3 - 1)^2 \\ &\equiv \quad W_1(J_1, J_2) \quad + \quad W_2(J_3), \end{aligned} \tag{3.116}$$

where A_{10} and A_{01} are the material constants and K is the bulk modulus. For a small strain, $2(A_{10} + A_{01})$ is equivalent to the shear modulus, and $6(A_{10} + A_{01})$ to Young's modulus for a three-dimensional solid, and $8(A_{10} + A_{01})$ is equivalent to the Young's modulus for a two-dimensional solid.

Since W_1 is independent of dilatation, volumetric deformation is only related to W_2 . The hydrostatic pressure is defined as the derivative of W_2 with respect to J_3 , as

$$p \equiv \frac{\partial W(J_1, J_2, J_3)}{\partial J_3} = \frac{\partial W_2(J_3)}{\partial J_3} = K(J_3 - 1). \tag{3.117}$$

Note that the term $J_3 - 1$ is equivalent to volumetric strain in small deformation (see Problem P3.10). Thus, the above relation recovers the linear elastic relation for small deformation. Since a large value of K is used to impose near-

incompressibility, numerical instability can result in computing the pressure from the displacement.

For the constitutive relation, stress can be obtained by differentiating the strain energy density in Eq. (3.116) with respect to strain. Since the material description is used, Eq. (3.116) is differentiated with respect to the Lagrangian strain to obtain the second Piola-Kirchhoff stress as

$$\mathbf{S} = \frac{\partial W}{\partial \mathbf{E}} = \frac{\partial W_1}{\partial J_1} \frac{\partial J_1}{\partial \mathbf{E}} + \frac{\partial W_1}{\partial J_2} \frac{\partial J_2}{\partial \mathbf{E}} + \frac{\partial W_2}{\partial J_3} \frac{\partial J_3}{\partial \mathbf{E}} \quad (3.118)$$

From the definition of the strain energy density, the second Piola-Kirchhoff stress can be rewritten as

$$\mathbf{S} = A_{10}J_{1,\mathbf{E}} + A_{01}J_{2,\mathbf{E}} + K(J_3 - 1)J_{3,\mathbf{E}}, \quad (3.119)$$

where the subscribed comma denotes derivative, i.e., $J_{1,\mathbf{E}} = \partial J_1 / \partial \mathbf{E}$. Thus, the next task is to obtain the expression of $J_{1,\mathbf{E}}$, $J_{2,\mathbf{E}}$, and $J_{3,\mathbf{E}}$ in terms of the Lagrangian strain. To this end, the derivatives of the reduced invariants with respect to Lagrangian strain can be written as

$$\begin{aligned} J_{1,\mathbf{E}} &= I_{1,\mathbf{E}}(I_3)^{-1/3} - \frac{1}{3}I_1(I_3)^{-4/3}I_{3,\mathbf{E}} \\ J_{2,\mathbf{E}} &= I_{2,\mathbf{E}}(I_3)^{-2/3} - \frac{2}{3}I_2(I_3)^{-5/3}I_{3,\mathbf{E}} \\ J_{3,\mathbf{E}} &= \frac{1}{2}(I_3)^{-1/2}I_{3,\mathbf{E}}, \end{aligned} \quad (3.120)$$

where the derivatives of the invariants can be obtained from

$$\begin{aligned} I_{1,\mathbf{E}} &= 2\mathbf{1} \\ I_{2,\mathbf{E}} &= 2(I_1\mathbf{1} - \mathbf{C}) \\ I_{3,\mathbf{E}} &= 2I_3\mathbf{C}^{-1}. \end{aligned} \quad (3.121)$$

The formulation in this section is called the penalty method in imposing the near-incompressibility constraints because the penalty parameter (i.e., the bulk modulus K) is used. As discussed earlier, this formulation will experience numerical instability when the hydrostatic pressure is calculated from displacement, which is called volumetric locking. Such instability stems from the fact that a small change in displacement can cause a large change in pressure due to the large magnitude of the bulk modulus. Even if various numerical techniques are available to eliminate/reduce volumetric locking, selective reduced integration, mixed formulation, and a perturbed Lagrangian formulation will be discussed in the following subsection.

3.5.2.2 Selective Reduced Integration

The strain energy density in Eq. (3.116) is composed of two parts: distortion and dilatation. After finite element discretization, let \mathbf{K}_s and \mathbf{K}_p be the distortion and dilatation stiffness matrices, respectively. Due to a large magnitude of bulk modulus in the penalty method, the stiffness of the dilatation part is much larger than that of the distortion part. Thus, if \mathbf{K}_p is not singular, the incremental equation $(\mathbf{K}_s + \mathbf{K}_p)\Delta\mathbf{u} = \mathbf{R}$ will yield $\Delta\mathbf{u} = \mathbf{0}$. Thus, the residual on the right-hand side is unable to be reduced, and the effect of \mathbf{K}_s is ignored due to the large difference in stiffness values. In this case, the stiffness is called over-constrained, and this phenomenon is called “volumetric locking.”

The selective reduced integration method is the most convenient way of reducing volumetric locking in the near-incompressibility constraint. The basic idea is to make the dilatation stiffness \mathbf{K}_p singular so that the distortion stiffness \mathbf{K}_s can affect the calculation of incremental displacement. Stiffness matrices are nonsingular (after applying appropriate boundary conditions) if a regular integration order is used. However, when integration is performed with one order less than the regular one, it is found that the stiffness matrix becomes singular and the finite element equation yields acceptable incremental solutions. For example, when a two-dimensional quadrilateral element is used for a Mooney-Rivlin material, \mathbf{K}_s is calculated using the 2×2 Gauss quadrature rule, while \mathbf{K}_p is integrated using the 1×1 quadrature rule. The same reduced integration scheme should be used for stress calculation: the distortion part ($\mathbf{S} = A_{10}J_{1,E} + A_{01}J_{2,E}$) with 2×2 and the dilatation part ($\mathbf{S} = K(J_3 - 1)J_{3,E}$) with 1×1 quadrature rule.

3.5.2.3 Mixed Formulation

In the mixed formulation, the hydrostatic pressure is treated as an independent variable instead of calculating it from Eq. (3.117). Thus, the strain energy density also depends on the hydrostatic pressure, and it is unnecessary to define the bulk modulus. This hydrostatic pressure is in fact a Lagrange multiplier in the mixed formulation. In the mixed formulation, the dilatational strain energy density is defined as

$$W_2(J_3, p) = p(J_3 - 1). \quad (3.122)$$

The advantage of the mixed formulation is that the pressure is not dependent on the displacement so there is no numerical instability involved. However, the constitutive relation becomes positive semi-definite and a special treatment is required in solving the matrix equation.

3.5.2.4 Perturbed Lagrangian Formulation

As explained in the mixed formulation section, the pressure plays the role of Lagrange multiplier in imposing the constraint. In the perturbed Lagrangian formulation, the product of a small constant with the sum of the squares of the Lagrange multipliers is added to the mixed formulation. The inverse of the bulk modulus is often used for the small constant. In the perturbed Lagrangian formulation, the dilatational strain energy density function is defined as

$$W_2(J_3, p) = p(J_3 - 1) - \frac{1}{2K}p^2. \quad (3.123)$$

If the Lagrange multiplier is removed in the element level through static condensation, then the perturbed Lagrangian formulation becomes identical to the penalty method. However, a reduced integration scheme is often employed for the pressure terms. In such a case, the perturbed Lagrangian formulation is the same as the selected reduced integration method.

Example 3.15 (Stress calculation in the perturbed Lagrangian formulation) When the dilatational strain energy density function is defined as Eq. (3.123), write the expression of stress as in Eq. (3.119) for the perturbed Lagrangian formulation. Also, show that the perturbed Lagrangian stress becomes identical to that of the penalty method when the pressure variable is eliminated in the element level.

Solution Using Eq. (3.123), the strain energy density for the perturbed Lagrangian formulation can be written as

$$W(J_1, J_2, J_3, p) = A_{10}(J_1 - 3) + A_{01}(J_2 - 3) + p(J_3 - 1) - \frac{1}{2K}p^2$$

Differentiating it with respect to the Lagrangian strain yields the second Piola-Kirchhoff stress as

$$\mathbf{S} = \frac{\partial W}{\partial \mathbf{E}} = A_{10}J_{1,\mathbf{E}} + A_{01}J_{2,\mathbf{E}} + pJ_{3,\mathbf{E}}$$

Note that the hydrostatic pressure p is an independent variable. In the variational equation with displacement and pressure variables, an additional equation is required for the pressure variable, which can be obtained by differentiating W with respect to p and equating the terms to zero:

$$\frac{\partial W}{\partial p} = \left(J_3 - 1 - \frac{p}{K} \right) = 0$$

If the above equation is used to eliminate the pressure variable in the element level, it simply becomes $p = K(J_3 - 1)$, which is equivalent to the pressure term in the penalty method. ■

3.5.2.5 Algorithm for Stress Calculation

In computer programming, it is convenient to use vector and matrix notation rather than tensor notation. Thus, in the following algorithm, vector and matrix notation will be used. Below is the procedure of stress calculation for the Mooney-Rivlin hyperelastic material. The inputs are the Lagrangian strain and material parameters, and the outputs are the six components of the second Piola-Kirchhoff stress:

1. For given strain $\{\mathbf{E}\} = \{E_{11}, E_{22}, E_{33}, E_{12}, E_{23}, E_{13}\}^T$ and given material constants (A_{10} , A_{01} , and K) for the penalty method, or material constants (A_{10} and A_{01}) and the hydrostatic pressure p for the mixed formulation method, perform the following calculation.
2. Set $\{\mathbf{1}\} = \{1, 1, 1, 0, 0, 0\}^T$ and $\{\mathbf{C}\} = 2 \times \{\mathbf{E}\} + \{\mathbf{1}\}$.
3. Calculate the three invariants:

$$I_1 = C_1 + C_2 + C_3.$$

$$I_2 = C_1 \times C_2 + C_1 \times C_3 + C_2 \times C_3 - C_4 \times C_4 - C_5 \times C_5 - C_6 \times C_6.$$

$$I_3 = (C_1 \times C_2 - C_4 \times C_4) \times C_3 + (C_4 \times C_6 - C_1 \times C_5) \times C_5 + (C_4 \times C_5 - C_2 \times C_6) \times C_6.$$

4. Calculate the derivatives of invariants with respect to the Lagrangian strain:

$$\{I_{1,\mathbf{E}}\} = 2 \times \{\mathbf{1}\}.$$

$$\{I_{2,\mathbf{E}}\} = 2 \times \{C_2 + C_3, C_3 + C_1, C_1 + C_2, -C_4, -C_5, -C_6\}^T.$$

$$\{I_{3,\mathbf{E}}\} = 2 \times \{C_2 C_3 - C_5 C_5, C_3 C_1 - C_6 C_6, C_1 C_2 - C_4 C_4,$$

$$C_5 C_6 - C_3 C_4, C_6 C_4 - C_1 C_5, C_4 C_5 - C_2 C_6\}^T.$$

5. Calculate the derivatives of the reduced invariants.

$$\{J_{1,\mathbf{E}}\} = I_3^{-1/3} \{I_{1,\mathbf{E}}\} - \frac{1}{3} I_1 I_3^{-4/3} \{I_{3,\mathbf{E}}\}$$

$$\{J_{2,\mathbf{E}}\} = I_3^{-2/3} \{I_{2,\mathbf{E}}\} - \frac{2}{3} I_2 I_3^{-5/3} \{I_{3,\mathbf{E}}\}$$

$$\{J_{3,\mathbf{E}}\} = \frac{1}{2} I_3^{-1/2} \{I_{3,\mathbf{E}}\},$$

6. Calculate the second Piola-Kirchhoff stress from Eq. (3.119):

$$\{\mathbf{S}\} = A_{10} \{J_{1,\mathbf{E}}\} + A_{01} \{J_{2,\mathbf{E}}\} + p \{J_{3,\mathbf{E}}\}.$$

When the penalty method is used, then use $K(J_3 - 1)$ instead of p .

3.5.3 Variational Equation and Linearization

Once the stress is calculated from the strain energy density, the energy form of the variational equation can readily be obtained using the nonlinear elasticity formulation developed in Sect. 3.3, which is rewritten here as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (3.124)$$

where the energy and load forms are defined as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \iint_{\Omega} \mathbf{S}(\mathbf{u}) : \bar{\mathbf{E}}(\mathbf{u}, \bar{\mathbf{u}}) \, d\Omega \quad (3.125)$$

and

$$\ell(\bar{\mathbf{u}}) = \iint_{\Omega} \bar{\mathbf{u}}^T \mathbf{f}^b \, d\Omega + \int_{\Gamma^s} \bar{\mathbf{u}}^T \mathbf{t} \, d\Gamma. \quad (3.126)$$

The definitions of the above two forms are identical to those of the nonlinear elastic equation. However, the second Piola-Kirchhoff stress is calculated from Eq. (3.119) for hyperelastic materials. Note that the undeformed state is used as a frame of reference.

As discussed before, the load form is independent of deformation when it is conservative. The energy form is nonlinear through the constitutive relation and strain tensor. Linearization of the second Piola-Kirchhoff stress can be expressed in terms of the displacement increment as

$$\Delta \mathbf{S} = \mathbf{W}_{,\mathbf{E},\mathbf{E}} : \Delta \mathbf{E} = \mathbf{D} : \Delta \mathbf{E} \quad (3.127)$$

where \mathbf{D} is the fourth-order constitutive tensor at the current load step, referring to the undeformed state, and $\Delta \mathbf{E}$ is the Lagrangian strain increment. Note that the constitutive tensor \mathbf{D} for the St. Venant-Kirchhoff material was constant, as in Eq. (3.51). However, it is now a function of deformation for the hyperelastic material.

The constitutive tensor \mathbf{D} can be obtained by differentiating the second Piola-Kirchhoff stress in Eq. (3.119) to obtain

$$\mathbf{D} = \frac{\partial \mathbf{S}}{\partial \mathbf{E}} = A_{10} J_{1,\mathbf{EE}} + A_{01} J_{2,\mathbf{EE}} + K(J_3 - 1) J_{3,\mathbf{EE}} + K J_{3,\mathbf{E}} \otimes J_{3,\mathbf{E}} \quad (3.128)$$

where $J_{1,\mathbf{EE}} = \partial^2 J_1 / \partial \mathbf{E} \partial \mathbf{E}$ is the second-order derivative of the reduced invariant with respect to the Lagrangian strain. The same notation is used for $J_{2,\mathbf{EE}}$ and $J_{3,\mathbf{EE}}$. The last term, $J_{3,\mathbf{E}} \otimes J_{3,\mathbf{E}}$, is a tensor product using $J_{3,\mathbf{E}}$ twice. Note that the constitutive tensor has major and minor symmetries, i.e., $D_{ijkl} = D_{klji}$ and

$D_{ijl} = D_{jik}$. Unfortunately, the expressions of the second-order derivatives are lengthy. From Eq. (3.120),

$$\begin{aligned} J_{1,\mathbf{EE}} &= I_{1,\mathbf{EE}} I_3^{-\frac{1}{3}} - \frac{1}{3} I_3^{-\frac{4}{3}} (I_{1,\mathbf{E}} \otimes I_{3,\mathbf{E}} + I_{3,\mathbf{E}} \otimes I_{1,\mathbf{E}}) + \frac{4}{9} I_1 I_3^{-\frac{2}{3}} I_{3,\mathbf{E}} \otimes I_{3,\mathbf{E}} - \frac{1}{3} I_1 I_3^{-\frac{4}{3}} I_{3,\mathbf{EE}} \\ J_{2,\mathbf{EE}} &= I_{2,\mathbf{EE}} I_3^{-\frac{2}{3}} - \frac{2}{3} I_3^{-\frac{5}{3}} (I_{2,\mathbf{E}} \otimes I_{3,\mathbf{E}} + I_{3,\mathbf{E}} \otimes I_{2,\mathbf{E}}) + \frac{10}{9} I_2 I_3^{-\frac{8}{3}} I_{3,\mathbf{E}} \otimes I_{3,\mathbf{E}} - \frac{2}{3} I_2 I_3^{-\frac{5}{3}} I_{3,\mathbf{EE}} \\ J_{3,\mathbf{EE}} &= -\frac{1}{4} I_3^{-\frac{3}{2}} I_{3,\mathbf{E}} \otimes I_{3,\mathbf{E}} + \frac{1}{2} I_3^{-\frac{1}{2}} I_{3,\mathbf{EE}} \end{aligned} \quad (3.129)$$

where the second-order derivatives of the invariants can be obtained from

$$\begin{aligned} I_{1,\mathbf{EE}} &= \mathbf{0} \\ I_{2,\mathbf{EE}} &= 4\mathbf{1} \otimes \mathbf{1} - \mathbf{I} \\ I_{3,\mathbf{EE}} &= 4I_3 \mathbf{C}^{-1} \otimes \mathbf{C}^{-1} - I_3 \mathbf{C}^{-1} \mathbf{I} \mathbf{C}^{-1} \end{aligned} \quad (3.130)$$

where $I_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$ is a symmetric fourth-order identity tensor.

Once the constitutive tensor is calculated, linearization of the energy form yields the same expression as in Eq. (3.74), which is rewritten here as

$$a^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) \equiv \iint_{\Omega} (\bar{\mathbf{E}} : \mathbf{D} : \Delta\mathbf{E} + \mathbf{S} : \Delta\bar{\mathbf{E}}) \, d\Omega. \quad (3.131)$$

Compared to the St. Venant–Kirchhoff material, the only differences are the expressions of the constitutive tensor and the second Piola–Kirchhoff stress. Thus, the same Newton–Raphson iterative method can be used for solving hyperelasticity problems.

The stress in Eq. (3.119) and material stiffness in Eq. (3.128) can easily be implemented in computer programs. Below is the MATLAB programs, Mooney, that calculates the second Piola–Kirchhoff stress and material stiffness for a given deformation gradient. The mixed and perturbed Lagrangian formulations require slight modification of the program.

PROGRAM Mooney

```
%  
% 2nd PK stress and material stiffness for Mooney–Rivlin material  
%  
function [Stress D] = Mooney(F, A10, A01, K, ltan)  
% Inputs:  
% F = Deformation gradient [3x3]  
% A10, A01, K = Material constants  
% ltan = 0 Calculate stress alone; 1 Calculate stress and material stiffness  
% Outputs:  
% Stress = 2nd PK stress [S11, S22, S33, S12, S23, S13];  
% D = Material stiffness [6x6]
```

```

%
X12 = 1/2; X13 = 1/3; X23 = 2/3; X43 = 4/3; X53 = 5/3; X89 = 8/9;
%
C = F' *F;
C1=C(1,1); C2=C(2,2); C3=C(3,3); C4=C(1,2); C5=C(2,3); C6=C(1,3);
I1 = C1+C2+C3;
I2 = C1*C2+C1*C3+C2*C3-C4^2-C5^2-C6^2;
I3 = det(C);
J3 = sqrt(I3);
J3M1 = J3 - 1.0D+00;
%
I1E = 2*[1 1 1 0 0 0]';
I2E = 2*[C2+C3, C3+C1, C1+C2, -C4, -C5, -C6]';
I3E = 2*[C2*C3-C5^2, C3*C1-C6^2, C1*C2-C4^2, ...
           C5*C6-C3*C4, C6*C4-C1*C5, C4*C5-C2*C6]';
%
W1 = I3^(-X13); W2 = X13*I1*I3^(-X43); W3 = I3^(-X23);
W4 = X23*I2*I3^(-X53); W5 = X12*I3^(-X12);
%
J1E = W1*I1E - W2*I3E;
J2E = W3*I2E - W4*I3E;
J3E = W5*I3E;
%
Stress = A10*J1E + A01*J2E + K*J3M1*J3E;
%
% Material stiffness
%
D = zeros(6);
if ltan == 1
%
I2EE = [0 4 4 0 0 0; 4 0 4 0 0 0; 4 4 0 0 0 0;
         0 0 0 -2 0 0; 0 0 0 0 -2 0; 0 0 0 0 0 -2];
I3EE = [ 0      4*C3  4*C2   0      -4*C5  0;
          4*C3  0      4*C1   0      0      -4*C6;
          4*C2  4*C1   0      -4*C4  0      0;
          0      0      -4*C4 -2*C3  2*C6  2*C5;
         -4*C5  0      0      2*C6 -2*C1  2*C4;
          0      -4*C6  0      2*C5  2*C4 -2*C2];
%
W1 = X23*I3^(-X12);     W2 = X89*I1*I3^(-X43); W3 = X13*I1*I3^(-X43);
W4 = X43*I3^(-X12);     W5 = X89*I2*I3^(-X53); W6 = I3^(-X23);
W7 = X23*I2*I3^(-X53); W8 = I3^(-X12);           W9 = X12*I3^(-X12);
%
J1EE = -W1*(J1E*J3E' + J3E*J1E') + W2*(J3E*J3E') - W3*I3EE;
J2EE = -W4*(J2E*J3E' + J3E*J2E') + W5*(J3E*J3E') + W6*I2EE - W7*I3EE;
J3EE = -W8*(J3E*J3E') + W9*I3EE;
%
D = A10*J1EE + A01*J2EE + K*(J3E*J3E') + K*J3M1*J3EE;
end
return;

```

3.6 Finite Element Formulation for Nonlinear Elasticity

So far, formulations and solution procedures of nonlinear elastic and hyperelastic problems have been discussed in the continuum domain. In practice, the structure is discretized by finite elements and the equilibrium equations are applied to these elements. In this section, finite element discretization is discussed using a four-node, plane-strain, quadrilateral solid element. Different types of elements have different schemes of interpolation, which will only affect the displacement–strain relation. The same algorithm for stress calculation and constitutive tensor can be used for different types of elements.

Figure 3.14 shows a quadrilateral element defined in the undeformed state. Even if a real structure is composed of many elements, for the simplicity of explanation, it is assumed that the structure is modeled by one element. In the total Lagrangian formulation, all interpolation functions are calculated in the undeformed geometry. In the computer implementation of the nonlinear finite element program, matrix–vector notation is more convenient than tensor notation. In matrix–vector notation, a second-order symmetric tensor (e.g., stress and strain) is expressed using a vector, while a fourth-order symmetric tensor (e.g., constitutive tensor) is expressed using a matrix. For example, the stress and strain vectors are defined as

$$\{\mathbf{S}\} = \{S_{11} \quad S_{22} \quad S_{12}\}^T \quad (3.132)$$

and

$$\{\mathbf{E}\} = \{E_{11} \quad E_{22} \quad 2E_{12}\}^T. \quad (3.133)$$

In the above definitions, the symmetric property of the tensor is used.

In the displacement-based implementation of finite elements, the displacement vector $\mathbf{u} = \{u_1, u_2\}^T$ is given for each node of an element. Subscript I will be used to denote the node such that \mathbf{u}_I will be the displacement vector at node I . The displacement within the element can be calculated using the following interpolation scheme:

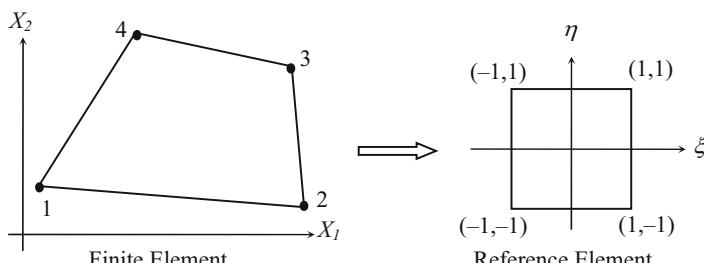


Fig. 3.14 Quadrilateral plane solid element

$$\mathbf{u} = \sum_{I=1}^{N_e} N_I(\xi) \mathbf{u}_I, \quad (3.134)$$

where N_e is the number of nodes of the element, $\xi = \{\xi, \eta\}^T$ is the natural coordinate vector at the reference element [see Fig. 3.14], and $N_I(\xi)$ is the interpolation or shape function. In the iso-parametric mapping method, the material coordinates within the element are also interpolated using the same interpolation function. In the material description, the reference coordinate $\mathbf{X} = \{X_1, X_2\}^T$ is interpolated using

$$\mathbf{X} = \sum_{I=1}^{N_e} N_I(\xi) \mathbf{X}_I \quad (3.135)$$

where $\mathbf{X}_I = \{X_{I1}, X_{I2}\}^T$ is the nodal coordinate of node I at the undeformed geometry.

In order to express the nonlinear equation in terms of nodal displacements, it is first necessary to interpolate the displacement gradient vector. Using the interpolation scheme in Eq. (3.134), the displacement gradient can be expressed as

$$\nabla_0 \mathbf{u} = \frac{\partial \mathbf{u}}{\partial \mathbf{X}} = \sum_{I=1}^{N_e} \frac{\partial N_I(\xi)}{\partial \mathbf{X}} \mathbf{u}_I, \quad (3.136)$$

where the derivatives of the interpolation function can be obtained using the chain rule of differentiation and the Jacobian relation, as explained in Eq. (1.139) in Chap. 1. The only difference is that the undeformed coordinate \mathbf{X} should be used. Using index notation, the components of displacement gradient can be written as

$$u_{i,j} = \sum_{I=1}^{N_e} N_{I,j}(\xi) u_{Ii} \quad (3.137)$$

where $N_{I,j}$ is the derivative of N_I with respect to X_i and u_{Ii} is the component of \mathbf{u}_I . In the following, a subscribed comma will be used for differentiation with respect to X_i .

From the displacement gradient, the deformation gradient can be calculated using Eq. (3.5). The following vector form of the displacement gradient is defined first:

$$\nabla_0 \mathbf{u} = \{ u_{1,1} \quad u_{1,2} \quad u_{2,1} \quad u_{2,2} \}^T \quad (3.138)$$

Then, the deformation gradient can be written as

$$\{\mathbf{F}\} = \{ F_{11} \quad F_{12} \quad F_{21} \quad F_{22} \}^T = \{ 1 + u_{1,1} \quad u_{1,2} \quad u_{2,1} \quad 1 + u_{2,2} \}^T. \quad (3.139)$$

For a given displacement gradient, the Lagrangian strain can be calculated by

$$\{\mathbf{E}\} = \begin{Bmatrix} E_{11} \\ E_{22} \\ 2E_{12} \end{Bmatrix} = \begin{Bmatrix} u_{1,1} + \frac{1}{2}(u_{1,1}u_{1,1} + u_{2,1}u_{2,1}) \\ u_{2,2} + \frac{1}{2}(u_{1,2}u_{2,1} + u_{2,2}u_{2,2}) \\ u_{1,2} + u_{2,1} + u_{1,2}u_{1,1} + u_{2,1}u_{2,2} \end{Bmatrix}. \quad (3.140)$$

Using the Lagrangian strain, the second Piola-Kirchhoff stress can be obtained by differentiating the strain energy density function with respect to \mathbf{E} . For example, in the case of the St. Venant–Kirchhoff material, the second Piola-Kirchhoff stress can be calculated from Eq. (3.53), while the hyperelastic material can be calculated from Eq. (3.119).

Next, the variation of Lagrangian strain $\bar{\mathbf{E}}$ in Eq. (3.63) can be written in vector notation as

$$\{\bar{\mathbf{E}}\} = [\mathbf{B}_N]\{\bar{\mathbf{d}}\} \quad (3.141)$$

where $\{\bar{\mathbf{d}}\} = \{\bar{d}_{11} \ \bar{d}_{12} \ \bar{d}_{21} \ \bar{d}_{22} \ \dots \ \bar{d}_{42}\}^T$ is the variation of nodal displacements and $[\mathbf{B}_N]$ is the nonlinear displacement-strain matrix defined as

$$[\mathbf{B}_N] = \begin{bmatrix} F_{11}N_{1,1} & F_{21}N_{1,1} & F_{11}N_{2,1} & F_{21}N_{2,1} & \dots & F_{11}N_{4,1} & F_{21}N_{4,1} \\ F_{12}N_{1,2} & F_{22}N_{1,2} & F_{12}N_{2,2} & F_{22}N_{2,2} & \dots & F_{12}N_{4,2} & F_{22}N_{4,2} \\ F_{11}N_{1,2} & F_{21}N_{1,2} & F_{11}N_{2,2} & F_{21}N_{2,2} & \dots & F_{11}N_{4,2} & F_{21}N_{4,2} \\ +F_{12}N_{1,1} & +F_{22}N_{1,1} & +F_{12}N_{2,1} & +F_{22}N_{2,1} & \dots & +F_{12}N_{4,1} & +F_{22}N_{4,1} \end{bmatrix}. \quad (3.142)$$

Note that the nonlinear displacement-strain matrix $[\mathbf{B}_N]$ is clearly different from $[\mathbf{B}]$ in linear systems in Chap. 1. The latter remains unchanged for a given element and given integration point, while the former changes according to displacement as it contains the components of deformation gradient.

Using Eq. (3.141) and the second Piola-Kirchhoff stress in Eq. (3.132), the discrete version of the energy form can be derived as

$$a(\mathbf{u}, \bar{\mathbf{u}}) = \iint_{\Omega} \mathbf{S} : \bar{\mathbf{E}} d\Omega = \{\bar{\mathbf{d}}\}^T \iint_{\Omega} [\mathbf{B}_N]^T \{\mathbf{S}\} d\Omega \equiv \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{int}\}, \quad (3.143)$$

where $\{\mathbf{f}^{int}\}$ is the discrete internal force vector. Note that both the tensor and matrix–vector notations are used in the equation.

In addition, the discrete external force vector can be derived from the definition of the load form as

$$\begin{aligned}\ell(\bar{\mathbf{u}}) &= \iint_{\Omega} \bar{\mathbf{u}}^T \mathbf{f}^b d\Omega + \int_{\Gamma^s} \bar{\mathbf{u}}^T \mathbf{t} d\Gamma \\ &= \sum_{I=1}^{N_e} \bar{\mathbf{u}}_I^T \left\{ \iint_{\Omega} N_I(\xi) \mathbf{f}^b d\Omega + \int_{\Gamma^s} N_I(\xi) \mathbf{t} d\Gamma \right\} \equiv \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{ext}\}. \end{aligned} \quad (3.144)$$

When concentrated nodal forces are applied, they can directly be added to the corresponding locations in $\{\mathbf{f}^{ext}\}$. Since the applied loads are assumed to be independent of deformation, the external force $\{\mathbf{f}^{ext}\}$ is a fixed vector. Thus, the discrete version of solving the nonlinear variational equation is to find the internal force that has the same value as the external force, i.e.,

$$\{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{int}(\mathbf{d})\} = \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{ext}\}, \quad \forall \{\bar{\mathbf{d}}\} \in \mathbb{Z}_h, \quad (3.145)$$

where \mathbb{Z}_h is the discrete counter part of space \mathbb{Z} . Since the displacement variation is zero at the nodes where displacements are prescribed, Eq. (3.145) satisfies $\{\mathbf{f}^{int}(\mathbf{d})\} = \{\mathbf{f}^{ext}\}$ for all nodes whose displacements are not prescribed.

Since the internal force is a nonlinear function of deformation, Eq. (3.145) needs to be solved using an iterative method, such as the Newton–Raphson method, which requires the Jacobian matrix or, equivalently, the tangent stiffness matrix. In the total Lagrangian formulation, the tangent stiffness matrix corresponds to discretization of the linearized energy form in Eq. (3.74).

As we discussed in Sect. 3.2.1, the incremental Lagrangian strain has a similar expression as

$$\{\Delta \mathbf{E}\} = [\mathbf{B}_N] \{\Delta \mathbf{d}\}. \quad (3.146)$$

Then, the first term in the structural energy form can be written as

$$\iint_{\Omega} \bar{\mathbf{E}} : \mathbf{D} : \Delta \mathbf{E} d\Omega = \{\bar{\mathbf{d}}\}^T \left[\iint_{\Omega} [\mathbf{B}_N]^T [\mathbf{D}] [\mathbf{B}_N] d\Omega \right] \{\Delta \mathbf{d}\}, \quad (3.147)$$

where the 3×3 matrix $[\mathbf{D}]$ is the matrix version of the constitutive tensor \mathbf{D} in Eq. (3.51). In the case of the St. Venant–Kirchhoff material, the matrix $[\mathbf{D}]$ becomes

$$[\mathbf{D}] = \begin{bmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{bmatrix} \quad (3.148)$$

In the case of the Mooney–Rivlin material, the matrix $[\mathbf{D}]$ can be calculated from the matrix version of Eq. (3.128) (see Problem P3.27).

The second term, the initial stiffness, of the linearized energy form can be written as

$$\iint_{\Omega} \mathbf{S} : \Delta \bar{\mathbf{E}} d\Omega = \{\bar{\mathbf{d}}\}^T \left[\iint_{\Omega} [\mathbf{B}_G]^T [\Sigma] [\mathbf{B}_G] d\Omega \right] \{\Delta \mathbf{d}\} \quad (3.149)$$

where

$$[\Sigma] = \begin{bmatrix} S_{11} & S_{12} & 0 & 0 \\ S_{12} & S_{22} & 0 & 0 \\ 0 & 0 & S_{11} & S_{12} \\ 0 & 0 & S_{12} & S_{22} \end{bmatrix} \quad (3.150)$$

$$[\mathbf{B}_G] = \begin{bmatrix} N_{1,1} & 0 & N_{2,1} & 0 & N_{3,1} & 0 & N_{4,1} & 0 \\ N_{1,2} & 0 & N_{2,2} & 0 & N_{3,2} & 0 & N_{4,2} & 0 \\ 0 & N_{1,1} & 0 & N_{2,1} & 0 & N_{3,1} & 0 & N_{4,1} \\ 0 & N_{2,1} & 0 & N_{2,2} & 0 & N_{3,2} & 0 & N_{4,2} \end{bmatrix} \quad (3.151)$$

Different from $[\mathbf{B}_N]$, $[\mathbf{B}_G]$ provides a linear relation between displacement and strain. This is expected because its counterpart in continuum, $\Delta \bar{\mathbf{E}}(\Delta \mathbf{u}, \bar{\mathbf{u}})$, is bilinear with respect to displacement variation and displacement increment.

Using Eqs. (3.147) and (3.149), the tangent stiffness matrix can be calculated as

$$[\mathbf{K}_T] = \iint_{\Omega} \left[[\mathbf{B}_N]^T [\mathbf{D}] [\mathbf{B}_N] + [\mathbf{B}_G]^T [\Sigma] [\mathbf{B}_G] \right] d\Omega \quad (3.152)$$

In general, the above integration as well as the one in the internal force in Eq. (3.143) are evaluated using the Gauss quadrature rule. Normally, 2×2 integration points are used for a quadrilateral element. In the case of a hyperelastic material with near-incompressibility constraint, however, the dilatation part may cause volumetric locking if 2×2 integration points are used. In order to relieve volumetric locking, in practice, $2 \times 2 + 1$ integration points are used in which an additional point is added at the center of the element such that the distortion part is calculated at 2×2 points and the dilatation part at the element center.

The discretized version of incremental equation in Eq. (3.75) can now be written in the form of finite element matrix equation as

$$\{\bar{\mathbf{d}}\}^T [\mathbf{K}_T] \{\Delta \mathbf{d}\} = \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{ext} - \mathbf{f}^{int}\}, \quad \forall \{\bar{\mathbf{d}}\} \in \mathbb{Z}_h \quad (3.153)$$

The above linear system of equations needs to be solved iteratively until the residual force (right-hand side) vanishes. Different methods of solving nonlinear equations in Chap. 2 can be used. For example, in the case of the modified Newton–Raphson method, the tangent stiffness matrix $[\mathbf{K}_T]$ at the first iteration is repeatedly used. In the case of the incremental force method, the external force vector $\{\mathbf{f}^{ext}\}$ is divided by the number of increments, and the Newton–Raphson method is employed at each load increment.

3.7 MATLAB Code for Hyperelastic Material Model

In this section, a MATLAB code, HYPER3D.m, is introduced that can solve for nonlinear elastic problems with the hyperelastic material model using the total Lagrangian formulation. The code can be called from NLFEA.m in Chap. 2. As explained in Fig. 2.26, the main function of HYPER3D.m is to build the tangent stiffness matrix, $[K]$, and the residual force vector, $\{R\}$. Then, NLFEA.m will solve for the displacement increment as a part of the Newton–Raphson iteration.

HYPER3D.m shares most of its input variables with that of ELAST3D.m in Chap. 1, which was explained in Table 1.5. Only difference is that `MID` and `PROP` are used instead of `ETAN`. The current implementation does not use `MID` inside of ELAST3D.m, which is an integer for the material identification number. In order to use HYPER3D.m, `MID` should be a negative integer, such as -1. The array `PROP` stores hyperelastic material constants. The current implementation uses Mooney–Rivlin material, which uses three material properties, $\text{PROP} = [\mathbf{A}_{10}, \mathbf{A}_{01}, k]$. As with ELAST3D.m in Chap. 1, the logical variable, `UPDATE`, is used to store the calculated stresses in the global array `SIGMA`, and the logical variable, `LTAN`, is used to calculate the tangent stiffness matrices and store them in the global array `GKF`. The residual force, `FORCE`, will always be calculated.

In order to assemble the local stiffness matrix into the global stiffness matrix, the `IDOF` array is used to store the location of the global DOFs corresponding to the local 24 DOFs. The `xG` and `wGT` arrays store one-dimensional integration points and corresponding weights, as in Table 1.4. In this implementation, only two-point integration is used for each coordinate direction.

At each integration point of an element, the derivatives of finite element shape functions are calculated by calling SHAPEL.m. Since the total Lagrangian formulation is used, the derivatives are evaluated with respect to the undeformed geometry. Using the derivatives of shape functions, the deformation gradient, \mathbf{F} , is calculated using Eq. (3.5). Then, using the deformation gradient, \mathbf{F} , and hyperelastic material properties in `PROP`, the second Piola-Kirchhoff stress and tangent stiffness matrix are calculated by calling the Mooney.m function.

In the total Lagrangian formulation, the stress–strain relation is nonlinear, and the nonlinear displacement–strain matrix, \mathbf{B}_N , in Eq. (3.142) and the linear displacement–strain matrix, \mathbf{B}_G , in Eq. (3.151) are stored in `BN(6,24)` and `BG(9,24)` arrays, respectively. In order to save computational time, the residual force array, `FORCE`, is always calculated, while the tangent stiffness array, `GKF`, is calculated only when the logical variable, `LTAN`, is true. This functionality can be used when the modified Newton–Raphson iteration is used. Since the total Lagrangian formulation uses the Lagrangian strain and the 2nd Piola-Kirchhoff stress, function CAUCHY is used to convert the 2nd Piola-Kirchhoff stress to Cauchy stress, which is used in printouts.

```

function HYPER3D(PROP, UPDATE, LTAN, NE, NDOF, XYZ, LE)
%*****
% MAIN PROGRAM COMPUTING GLOBAL STIFFNESS MATRIX RESIDUAL FORCE FOR
% HYPERELASTIC MATERIAL MODELS
%*****
%%
global DISPTD FORCE GKF SIGMA
%
% Integration points and weights
XG=[-0.57735026918963D0, 0.57735026918963D0];
WGT=[1.00000000000000D0, 1.00000000000000D0];
%
% Index for history variables (each integration pt)
INTN=0;
%
%LOOP OVER ELEMENTS , THIS IS MAIN LOOP TO COMPUTE K AND F
for IE=1:NE
    % Nodal coordinates and incremental displacements
    ELXY=XYZ(LE(IE,:));
    % Local to global mapping
    IDOF=zeros(1,24);
    for I=1:8
        II=(I-1)*NDOF+1;
        IDOF(II:II+2)=(LE(IE,I)-1)*NDOF+1:(LE(IE,I)-1)*NDOF+3;
    end
    DSP=DISPTD(IDOF);
    DSP=reshape(DSP,NDOF,8);
    %
    %LOOP OVER INTEGRATION POINTS
    for LX=1:2, for LY=1:2, for LZ=1:2
        E1=XG(LX); E2=XG(LY); E3=XG(LZ);
        INTN = INTN + 1;
        %
        % Determinant and shape function derivatives
        [~, SHPD, DET] = SHAPEL([E1 E2 E3], ELXY);
        FAC=WGT(LX)*WGT(LY)*WGT(LZ)*DET;
        %
        % Deformation gradient
        F=DSP*SHPD' + eye(3);
        %
        % Computer stress and tangent stiffness
        [STRESS, DTAN] = Mooney(F, PROP(1), PROP(2), PROP(3), LTAN);
        %
        % Update plastic variables
    end
end

```

```

if UPDATE
    STRESS=CAUCHY(F, STRESS);
    SIGMA(:,INTN)=STRESS;
    continue;
end
%
% Add residual force and tangent stiffness matrix
BN=zeros(6,24);
BG=zeros(9,24);
for I=1:8
    COL=(I-1)*3+1:(I-1)*3+3;
    BN(:,COL)=[SHPD(1,I)*F(1,1) SHPD(1,I)*F(2,1) SHPD(1,I)*F(3,1);
                SHPD(2,I)*F(1,2) SHPD(2,I)*F(2,2) SHPD(2,I)*F(3,2);
                SHPD(3,I)*F(1,3) SHPD(3,I)*F(2,3) SHPD(3,I)*F(3,3);
                SHPD(1,I)*F(1,2)+SHPD(2,I)*F(1,1)
SHPD(1,I)*F(2,2)+SHPD(2,I)*F(2,1) SHPD(1,I)*F(3,2)+SHPD(2,I)*F(3,1);
                SHPD(2,I)*F(1,3)+SHPD(3,I)*F(1,2)
SHPD(2,I)*F(2,3)+SHPD(3,I)*F(2,2) SHPD(2,I)*F(3,3)+SHPD(3,I)*F(3,2);
                SHPD(1,I)*F(1,3)+SHPD(3,I)*F(1,1)
SHPD(1,I)*F(2,3)+SHPD(3,I)*F(2,1) SHPD(1,I)*F(3,3)+SHPD(3,I)*F(3,1)];
    %
    BG(:,COL)=[SHPD(1,I) 0 0;
                SHPD(2,I) 0 0;
                SHPD(3,I) 0 0;
                0 SHPD(1,I) 0;
                0 SHPD(2,I) 0;
                0 SHPD(3,I) 0;
                0 0 SHPD(1,I);
                0 0 SHPD(2,I);
                0 0 SHPD(3,I)];
end
%
% Residual forces
FORCE(IDOF) = FORCE(IDOF) - FAC*BN'*STRESS;
%
% Tangent stiffness
if LTAN
    SIG=[STRESS(1) STRESS(4) STRESS(6);
          STRESS(4) STRESS(2) STRESS(5);
          STRESS(6) STRESS(5) STRESS(3)];
    SHEAD=kron(eye(3),SIG);
    %
    EKF = BN'*DTAN*BN + BG'*SHEAD*BG;
    GKF(IDOF, IDOF)=GKF(IDOF, IDOF)+FAC*EKF;
end
end; end; end;
end

```

```

function STRESS=CAUCHY(F, S)
%*****
% CONVERT 2ND PK STRESS INTO CAUCHY STRESS
%*****
%%
PK=[S(1) S(4) S(6);S(4) S(2) S(5);S(6) S(5) S(3)];
DETF = det(F);
PKF = PK*F';
ST = F*PKF/DETF;
STRESS=[ST(1,1) ST(2,2) ST(3,3) ST(1,2) ST(2,3) ST(1,3)]';
end

```

Example 3.16 (Hyperelastic analysis using MATLAB) Consider a unit cube as shown in Fig. 3.15. An eight-node solid element is used to model the cube. The positive X_1 face (Face 4) is extended with a stretch ratio $\lambda = 6.0$. The following boundary conditions are given: $u_1 = 0$ at Face 6, $u_2 = 0$ at Face 3, and $u_3 = 0$ at Face 1. Using NLFEA, calculate the relation between the stretch ratio and Cauchy stress. Use the incompressible Mooney-Rivlin hyperelastic material with $A_{10} = 80$ MPa, $A_{01} = 20$ MPa, and $K = 10^7$ MPa. Use 20 load increments. Compare the results with the analytical solution.

Solution The following MATLAB script defines an eight-node solid element with boundary conditions. Since this is a displacement-controlled problem, **EXTFORCE** array is empty. Instead, **SDISPT** array has nonzero prescribed displacements for those nodes at Face 4. In order to make stretch ratio = 6, the displacement should be 5.0. The total load is divided by 20 increments in **TIMS** array.

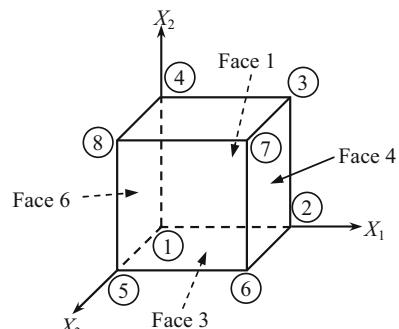


Fig. 3.15 Extension of an incompressible unit cube

```

%
% Ex 3-16 Hyperelastic tension example
%
% Nodal coordinates
XYZ=[0 0 0;1 0 0;1 1 0;0 1 0;0 0 1;1 0 1;1 1 1;0 1 1];
%
% Element connectivity
LE=[1 2 3 4 5 6 7 8];
%
% No external force
EXTFORCE=[];
%
% Prescribed displacements [Node, DOF, Value]
SDISPT=[1 1 0;4 1 0;5 1 0;8 1 0; % u1=0 for Face 6
         1 2 0;2 2 0;5 2 0;6 2 0; % u2=0 for Face 3
         1 3 0;2 3 0;3 3 0;4 3 0; % u3=0 for Face 1
         2 1 5;3 1 5;6 1 5;7 1 5]; % u1=5 for Face 4
%
% Load increments [Start End Increment InitialFactor FinalFactor]
TIMS=[0.0 1.0 0.05 0.0 1.0]';
%
% Material properties
MID=-1;
PROP=[80 20 1E7];
%
% Set program parameters
ITRA=30; ATOL=1.0E5; NTOL=6; TOL=1E-6;
%
% Calling main function
NOUT = fopen('output.txt','w');
NLFEA(ITRA, TOL, ATOL, NTOL, TIMS, NOUT, MID, PROP, EXTFORCE, SDISPT, XYZ, LE);
fclose(NOUT);

```

The convergence history shows an important capability of NLFEA. In the first increment (**Time = 0.05**), the residual is 1.175E5, which is larger than **ATOL = 1.0E5**. Therefore, NLFEA considers that the residual is too high, as which the load increment is bisected. After reducing the load increment to 0.025, the iteration converges in four iterations. Once the iteration converges, the bisection recovers the initial time increment. Therefore, after **Time = 0.05**, the time increment is recovered to 0.05. The residual during the iteration shows a nice quadratic convergence as the residual changes by several orders of magnitude in each iteration. The convergence iteration is considered to be converged when the residual is less than **TOL = 1E-6**.

Time	Time step	Iter	Residual
0.05000	5.000e-02	2	1.17493e+05
Not converged. Bisecting load increment 2			

```

Time  Time step Iter      Residual
0.02500 2.500e-02    2 2.96114e+04
                  3 2.55611e+02
                  4 1.84747e-02
                  5 1.51867e-10

Time  Time step Iter      Residual
0.05000 2.500e-02    2 2.48106e+04
                  3 1.69171e+02
                  4 7.67766e-03
                  5 2.39898e-10

Time  Time step Iter      Residual
0.10000 5.000e-02    2 8.45251e+04
                  3 1.88898e+03
                  4 8.72537e-01
                  5 1.86783e-07

...
Time  Time step Iter      Residual
1.00000 5.000e-02    2 8.55549e+03
                  3 8.98726e+00
                  4 9.88176e-06
                  5 1.66042e-09

```

Figure 3.16 shows the relationship between the stretch ratio and Cauchy stress. Stress increases nonlinearly. It is noted that due to incompressibility, the volume of stretched cube can be calculated by $6 \times (1-0.5917) \times (1-0.5917) = 1.00025$, which is almost identical to the initial volume. ■

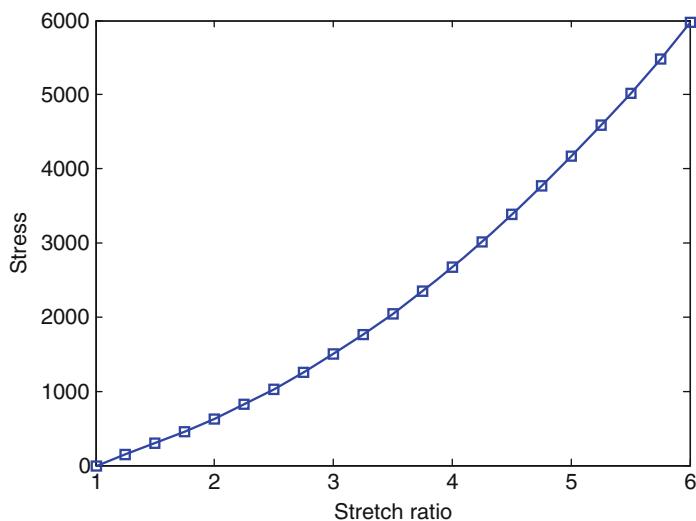


Fig. 3.16 Stress–extension ration graph for the extension of a unit cube

3.8 Nonlinear Elastic Analysis Using Commercial Finite Element Programs

In this section, nonlinear elastic analysis procedures using three commercial finite element programs are discussed. Since controlling load steps was already discussed in Chap. 2, defining nonlinear materials (including hyperelastic materials) will be discussed in this section.

3.8.1 Usage of Commercial Programs

3.8.1.1 Abaqus

***STEP, NLGEOM=YES**

The ***STEP** keyword was discussed in Chap. 2. In addition to specifying the maximum allowable load increments, the option can also be used to include geometrically nonlinear effects. By setting the **NLGEOM** parameter to YES, Abaqus considers the effects of large deformation and rotation.

Abaqus does not support the St. Venant–Kirchhoff material. Instead, if the material is defined as linear elastic with Young’s modulus and Poisson’s ratio, and **NLGEOM** parameter, it is considered as a nonlinear elastic material with a constant relation between Cauchy stress and engineering strain at the deformed state (a hypoelastic constitutive relation).

```
*MATERIAL, NAME=MOONEY
*HYPERELASTIC, MOONEY-RIVLIN
A10, A01,
```

In order to define a hyperelastic material in Abaqus, the ***HYPERELASTIC** keyword is used with the material definition. This keyword defines the material type and parameters. Since Abaqus uses a mixed formulation for incompressibility, there is no need to define the bulk modulus. Abaqus supports the Arruda–Boyce, Marlow, Mooney–Rivlin, Neo–Hookean, Ogden, polynomial, Yeoh, and user-defined hyperelastic material models.

3.8.1.2 ANSYS

NLGEOM, ON

This command activates geometric nonlinear analysis. This affects the integration domain, stiffness matrix, and stress and strain calculations. The integration

domain is updated to the current deformed domain. The stiffness matrix changes according to the current displacement. Large strain effects and the nonlinear stress–strain relation are used.

ANSYS does not support the St. Venant–Kirchhoff material. Instead, if the material is defined as a linear elastic with Young’s modulus and Poisson’s ratio, and NLGEOM is ON, it is considered as a nonlinear elastic material with a constant relation between Cauchy stress and engineering strain at the deformed state.

```
TB, HYPER, 1, , 2, MOONEY
TBDATA, 1, A10
TBDATA, 2, A01
TBDATA, 3, 2/K
```

The first command, **TB**, defines a Mooney-Rivlin material with two parameters. In fact, ANSYS allows 2, 3, 5, and 9 parameter models based on Eq. (3.105). The following three **TB DATA** commands provide the values of the parameters. Note that the number of **TB DATA** is three because the last parameter is the incompressibility parameter, which corresponds to 2/K with K being the bulk modulus. ANSYS also provides polynomial, Neo–Hookean, Ogden, Arruda–Boyce, Gent, Yeoh, Blatz–Ko foam, Ogden compressible foam, and user-defined hyperelastic models.

3.8.1.3 NEiNastran

```
PARAM, LGDISP, 1
```

This BULKDATA command activates geometric nonlinear analysis. This affects the integration domain, stiffness matrix, and stress and strain calculations. The integration domain is updated to the current deformed domain. The stiffness matrix changes according to the current displacement. Large strain effects and nonlinear stress–strain relation are used.

```
MATHP
```

This entry defines hyperelastic material parameters. The parameters of the MATHP entry are as follows:

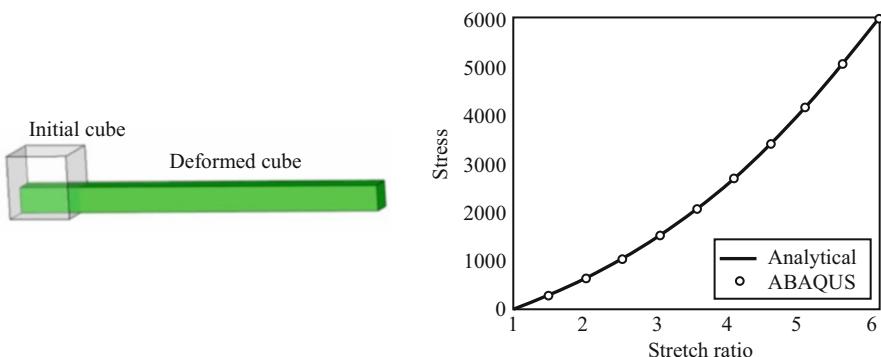
MATHP	MID	A10	A01	D1				
		NA	ND					
	A20	A11	A02	D2				
	A30	A21	A12	A03	D3			
	A40	A31	A22	A13	A04	D4		
	A50	A41	A32	A23	A14	A05	D5	
	TAB1	TAB2	TAB3	TAB4				TABD

MID is a unique identification number of the material. By default, the model assumes the two-parameter Mooney-Rivlin material with **A10** and **A01**. **D1** is the half of the bulk modulus. When more than two parameters are used, it is necessary to define **NA** and **ND**, which specify the order of distortional and dilatational strain energy polynomials. Depending on **NA** and **ND**, additional coefficients **A20**, **A11**, etc., should be defined. When test data are available, **DTAB1**, **TAB2**, **TAB3**, **TAB4**, and **TABD** can be used to input test data so that NEiNastran can calculate coefficients using a regression method.

Example 3.17 (Hyperelastic analysis using Abaqus) Consider a unit cube as shown in Fig. 3.15. Using Abaqus, calculate the relation between the stretch ratio and Cauchy stress. Use the incompressible Mooney-Rivlin hyperelastic material with $A_{10} = 80 \text{ MPa}$ and $A_{01} = 20 \text{ MPa}$. Use 20 load increments. Compare the results with the analytical solution.

Solution Below is the list of Abaqus commands used to solve the uniform extension of a cube. C3D8RH in Abaqus is a hybrid eight-node linear brick element with reduced integration and hourglass control. The element has the hydrostatic pressure as an independent variable. The ***SOLID SECTION** keyword assigns the material (named MOONEY in this example) to the element via the **ELSET** parameter. In order to apply the stretch ratio of 6.0, the displacement at Face 4 needs to increase by 5.0 m.

The following figure shows the initial and deformed geometry of the cube. The stress–stretch ratio curve shows that the numerical results agree well with the analytical results.



```

*HEADING
- Mooney-Rivlin Uniaxial tension
*NODE, NSET=ALL
1,
2,1.
3,1.,1.,
4,0.,1.,
*HYPERELASTIC, MOONEY-RIVLIN
80., 20.,
*STEP, NLGEOM, INC=20
UNIAXIAL TENSION
*STATIC, DIRECT
1.,20.
*BOUNDARY, OP=NEW

```

```

5,0.,0.,1.                      FACE1,3
6,1.,0.,1.                      FACE3,2
7,1.,1.,1.                      FACE6,1
8,0.,1.,1.                      FACE4,1,1,5.
*NSET,NSET=FACE1
1,2,3,4
*NSET,NSET=FACE3
1,2,5,6
*NSET,NSET=FACE4
2,3,6,7
*NSET,NSET=FACE6
4,1,8,5
*ELEMENT,TYPE=C3D8RH,ELSET=ONE
1,1,2,3,4,5,6,7,8
*SOLID SECTION,ELSET=ONE,MATERIAL=MOONEY *END STEP
*MATERIAL,NAME=MOONEY

```



3.8.2 Modeling Examples of Nonlinear Elastic Materials

In this section, several analysis problems are used to discuss about modeling issues as well as verifying the accuracy of analysis results with that of literature.

Hyperelastic thick cylinder under internal pressure [12]: An infinitely long cylinder in Fig. 3.17a is made of Mooney-Rivlin material with $A_{10}=80$ psi and $A_{01}=20$ psi. The material is nearly incompressible where the compressibility is

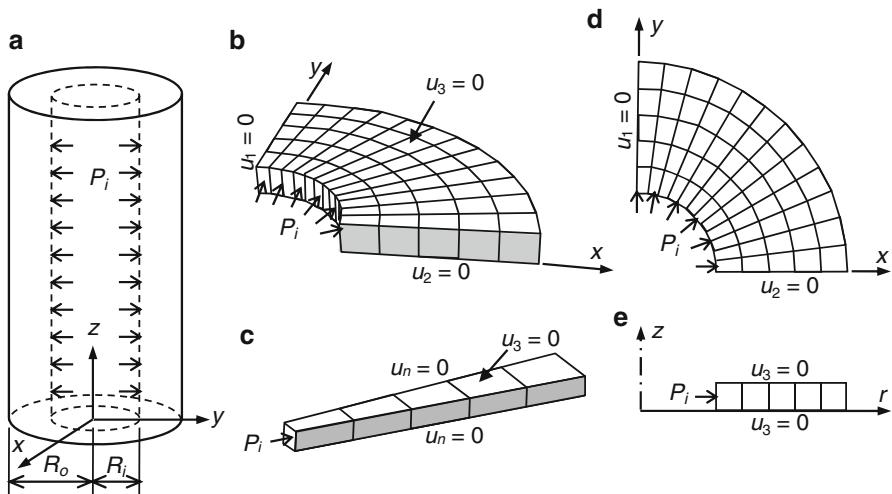


Fig. 3.17 Hyperelastic thick cylinder under internal pressure

Table 3.3 Target results for hyperelastic thick cylinder under internal pressure

Criterion	Target value
Displacement at the inner radius (in)	7.180
Stress at the first element (psi)	-122.0

equivalent to Poisson's ratio of 0.49. An internal pressure of $P_i = 150$ psi is applied. The inner and outer radii of the cylinder are, respectively, $R_i = 7.0$ in and $R_o = 18.625$ in. The objective is to find the radial displacement at the inner radius and the radial stress at radius $R = 8.16$ in (center of 1st element) and to compare the results with the target values in Table 3.3.

Since the cylinder is infinitely long, it is impossible to model the entire cylinder in the axial direction. Therefore, an approximation must be adopted in this direction. From the fact that the cylinder is infinitely long, it is reasonable to assume that there is no deformation in the axial direction, which corresponds to the plane-strain condition. Therefore, it is expected that the results will be independent of axial location z . In the circular cross section, it is also expected that the results will be only a function of radius and independent of angular position θ . This type of problem is called axisymmetric. Therefore, it would be unnecessary to model the entire circular region. In addition, if the entire circular region is modeled, it would be difficult to apply to displacement boundary conditions. All rigid-body motions must be removed in static analysis.

The problem statement does not provide the bulk modulus for incompressibility. Since Poisson's ratio is given, this information needs to be used to estimate the bulk modulus. The results are not particularly sensitive to this value because the material is unconfined. First, the Young's modulus can be approximated by $E = 6(A_{10} + A_{01})$. Then, using the definition of bulk modulus, it can be calculated from

$$K = \frac{E}{3(1 - 2\nu)} = \frac{600}{0.06} = 10,000 \text{ psi}$$

The problem can be solved by different modeling strategies. In the following, the problem can be modeled using 3D solid element, 2D plane-strain elements, and axisymmetric elements. It is suggested to try with different element types and compare the results with each other.

- (a) **3D solid element:** The advantage of using 3D solid elements is that it closely models the real geometry. 8-node or 20-node hexahedron or 10-node tetrahedron elements can be used to model the cylindrical geometry. Since the results do not vary along the z -axis, it would be enough to have a single element in that direction. In order to apply the plane-strain condition, the z -directional displacements are fixed. Since there is no systematic way to apply the axisymmetric condition for 3D solid element, the first suggestion is to model the first quadrant of the circular cross section (see Fig. 3.17b). The advantage of this model is that it is straightforward to apply boundary conditions. In order to have the same effect with the full circular geometry,

u_1 is fixed for those nodes on the yz -plane, and u_2 is fixed for those nodes on the xz -plane. It would be better to arrange elements in such a way that element boundaries are aligned with lines with a constant angle θ , although this will make small elements near the inner radius and large elements near the outer radius. The other possibility is to further reduce the number of elements along the angular direction because the results are independent of θ (Fig. 3.17c). In such a case, axisymmetric condition can be imposed by fixing the displacements in the normal direction to the cut plane. Since these planes may not be parallel to any global coordinate direction, it is necessary to establish local coordinates and apply displacement boundary conditions using them. The number of elements along the radial direction can be determined through the convergence analysis, but about five elements should be reasonable. The stress evaluation point $R = 8.16\text{in}$ is chosen as the center of the first element when five equal-length elements are used in the radial direction. Lastly, it is important to understand that when linear elements are used, the circular boundaries of inner and outer circumferences are approximated by piecewise straight line segments, which may cause an error in the applied pressure in the inner surface.

- (b) **Plane-strain elements:** The plane-strain elements only require modeling the cross-sectional area (xy plane) and assume there is no deformation in the axial direction. For the purpose of geometric modeling, either top or bottom portion of the 3D solid elements in the previous section is required (see Fig. 3.17d). 8-node or 4-node quadrilateral or 6-node triangular elements can be used to model the circular cross section. Similar to the 3D solid elements, either the first quadrant or a portion of small angle θ can be modeled using proper symmetric boundary conditions. Since the plane-strain elements do not have a degree of freedom in the z -direction, there is no need to fix u_3 displacement. For the standard plane-strain element, no thickness information is required. In such a case, it is equivalent to assume a unit thickness.
- (c) **Axisymmetric elements:** An axisymmetric problem models a plane geometry and rotates it with respect to an axis of rotation to build a 3D geometry. Therefore, the cylinder model is cut by xz -plane, and the positive side of the cross section is used to make axisymmetric elements (see Fig. 3.17e). The same types of element geometry with the plane-strain elements can be used. This is why many finite element programs use the same elements for plane strain and axisymmetry and use different properties to distinguish them. The conventional xyz coordinates are interpreted as $r\theta z$ coordinates in axisymmetric problems. For boundary conditions, the axisymmetric model cannot move in the radial and angular directions. Therefore, only u_3 at the top and bottom edges need to be fixed. In axisymmetric problems, u_1 displacement is interpreted as the radial displacement u_r , and the circumferential displacement u_2 corresponds to u_θ .

Hyperelastic circular plate [13]: A flat circular membrane made of a rubber material is subjected to uniform pressure on the bottom surface (see Fig. 3.18a). The

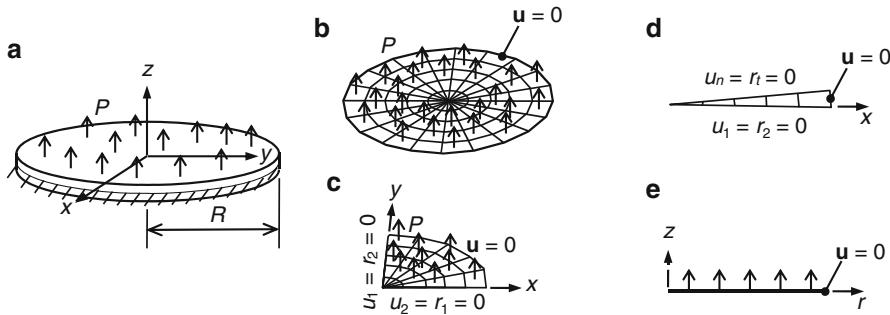


Fig. 3.18 Hyperelastic circular plate

Table 3.4 Target results for hyperelastic circular plate

Pressure (psi)	Displacement u_3 at the center (in)
4.0	2.250
24.0	6.200
38.0	10.900

radius of membrane is $R = 7.5\text{ in}$ and thickness $t = 0.5 \text{ in}$. The edge of membrane is fixed. The Mooney-Rivlin hyperelastic material is used with material constants $A_{10} = 80 \text{ psi}$ and $A_{01} = 20 \text{ psi}$. The objective is to determine the response as pressure is increased to 50 psi and to compare the results with the target values in Table 3.4.

Since the plate is thin compared to the radius, 3D solid elements will not perform well. This is especially true when a thin plate goes through bending deformation. At least four or five elements are required through the thickness direction in order to capture the bending behavior. Then in order to maintain a good aspect ratio, elements in xy plane should also be in the similar size. That means a huge number of elements must be used to model the membrane structures using 3D solid element. This is generally true for modeling structures made of sheet metals or membranes. Therefore, it would be better to model the membrane using shell elements. Shell elements need to be created on the neutral plane (midplane) of the membrane. The thickness of membrane is given as a property in the shell element. In addition to the three nodal displacements at each node, a shell element has two or three rotational degrees of freedom, depending on the theory used for implementation. These rotational degrees of freedom will be denoted by r_1 and r_2 in addition to displacements u_1 , u_2 , and u_3 for shell elements.

Similar to the pressurized cylinder problem, the problem is axisymmetric. The easiest way of modeling is to make the circular plate, to fix the entire boundary, and to apply a uniform pressure load (see Fig. 3.18b). However, in order to make the axisymmetric behavior, elements should be laid out in the pattern of rotational symmetry. Note that the shape of element at the center is triangular and all others are quadrilateral. Since the results will not be a function of angular position θ , it is possible to reduce the entire circle into the first quadrant with appropriate symmetric boundary conditions. Since the shell element has additional rotational degree of

freedom, caution is required to apply for the symmetric boundary condition. First, for the edge along the y -axis, $u_1 = r_2 = 0$, where r_2 is the rotational degree of freedom along y -axis. Similarly, for the edge along the x -axis, $u_2 = r_1 = 0$ (see Fig. 3.18c). That is, the rotation along a symmetry line must vanish. Since the results are independent of angular positions, the number of elements can be further reduced by one in θ direction (see Fig. 3.18d). In that case, local coordinates are required to provide the symmetric boundary condition. For the inclined edge, the displacement normal to the edge and the rotation tangential to the edge must be fixed for symmetric boundary condition.

Instead of 3D shell elements, axisymmetric shell elements can be used to model the circular plate (see Fig. 3.18e). If a line is drawn from the center to the edge along x -axis, and if the line is rotated by 360° degree with respect to z -axis, then it becomes the circular plate. The shape of axisymmetric shell element is a line for a linear element and a curve for a quadratic element. In the axisymmetric shell element, there is no need to provide symmetric boundary conditions. The x - and z -directional displacements at the node at the edge need to be fixed, and x -directional displacement at the node at the center needs to be fixed. Note that x -axis should be interpreted as r -axis.

Figure 3.19 shows the deformed shape of the membrane at different levels of pressure. It shows that the thickness of the membrane gradually decreases as the deformation increases. Since the membrane will go through a large deformation, the surface area that the pressure is applied continuously varies; i.e., the applied load depends on deformation. This is the case of force nonlinearity. Depending on the relationship between force and displacement, it is possible that the deformation-dependent force can cause instability. Therefore, it is beneficial to start with small force increments and to plot the force–displacement graph to predict any possible instability.

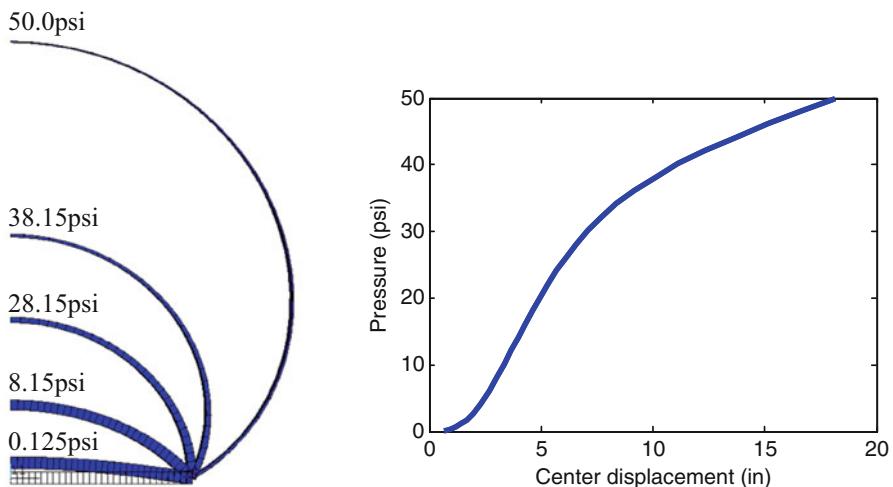


Fig. 3.19 Deformed geometry and pressure–displacement curve for the hyperelastic circular plate

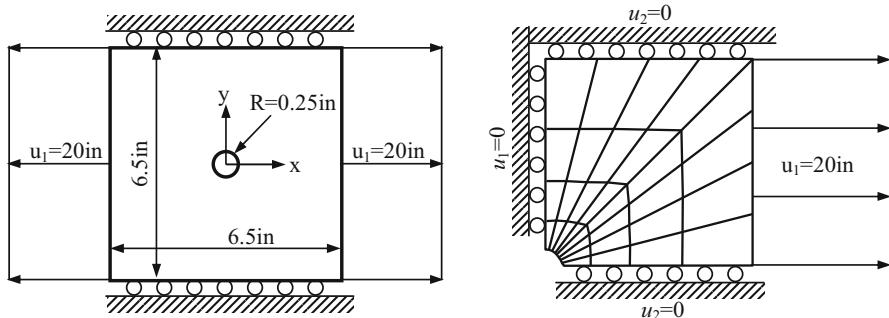


Fig. 3.20 Stretching of a sheet with a hole

Stretching of a sheet with a hole [13, 14]: The objective of this example is to verify the results of hyperelastic materials in plane stress using the uniform large stretching of a thin, initially square sheet containing a centrally located circular hole. The results are compared with those provided in Oden (1972) for different forms of the strain energy function using the experimental results of Treloar (1944). The geometry and the mesh for a quarter-sheet are shown in Fig. 3.20. The undeformed square sheet is 2 mm (0.079 in.) thick and is 165 mm (6.5 in.) on each side. It has a centrally located internal hole of radius 6.35 mm (0.25 in.). The sheet is confined in y -direction at top and bottom and uniformly extended by 20 in in both positive and negative x -directions. This is a very large deformation as the imposed displacement is seven times larger than the dimension of undeformed geometry. The force–displacement curve can be obtained by calculating the reaction force at the edge of imposing displacement.

For the modeling purpose, it is possible to make a mesh to the entire sheet. However, since the deformation is symmetric with respect to both x - and y -axis, it is better to use the symmetric modeling, in which only the first quadrant can be used for generating finite elements. The body is modeled with 32 second-order plane stress elements: CPS8R for Abaqus, PLANE183 for ANSYS, and CQUAD8 for NEiNastran. For the cut edges, $u_2 = 0$ symmetric condition is applied to the edge parallel to x -axis, while $u_1 = 0$ is applied to the edge parallel to y -axis.

The experimental data of Treloar (1944) composed of uniaxial, biaxial, and planar tension data are applied to these models. The first hyperelastic material model is the Mooney-Rivlin model that is discussed in this chapter, whose strain energy density is given in the following form:

$$U = A_{10}(I_1 - 3) + A_{01}(I_2 - 3)$$

with the incompressibility condition. The second model is similar to the Mooney-Rivlin model, but with more parameters. The Biderman model uses four parameters:

$$U = A_{10}(I_1 - 3) + A_{01}(I_2 - 3) + A_{20}(I_1 - 3)^2 + A_{30}(I_1 - 3)^3$$

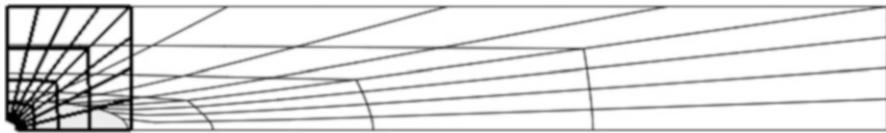


Fig. 3.21 Stretched geometry of a rectangular sheet with a hole

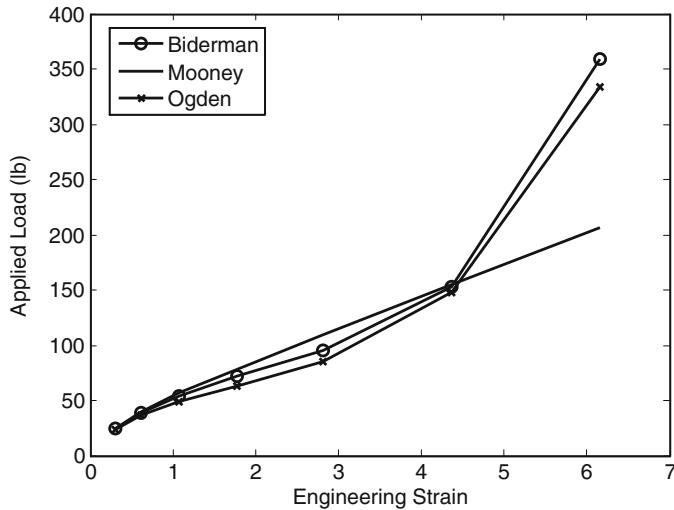


Fig. 3.22 Load responses of a sheet with a hole

The constants used by Oden (1972) are $A_{10} = 0.1863$ MPa (27.02 psi), $A_{01} = 0.00979$ MPa (1.42 psi), and, for the Biderman model, $A_{20} = -0.00186$ MPa (-0.27 psi) and $A_{30} = 0.0000451$ MPa (0.00654 psi). It is noted that as the magnitudes of model parameters are gradually decreasing as the order of polynomial increases. The third material model is the Ogden hyperelasticity model, which is defined using the principal stretches, λ_1 , λ_2 , and λ_3 , as

$$U = \sum_{i=1}^3 \frac{2\mu_i}{\alpha_i^2} (\bar{\lambda}_1^{\alpha_i} + \bar{\lambda}_2^{\alpha_i} + \bar{\lambda}_3^{\alpha_i} - 3)$$

where material parameters μ_i and α_i can be calculated by fitting the experimental data. Many finite element programs provide the capability of calculating model parameters by fitting experimental data.

The final displaced configuration is shown in Fig. 3.21; and the load responses are shown in Fig. 3.22, where the load is plotted as a function of the overall nominal strain of the sheet in the x-direction. The results of Biderman and Ogden models are seen to agree closely with Oden's. The Mooney-Rivlin strain energy function (with and as the only nonzero terms) cannot predict the "locking" of the response at higher strains that is predicted by the Biderman and Ogden strain energy functions.

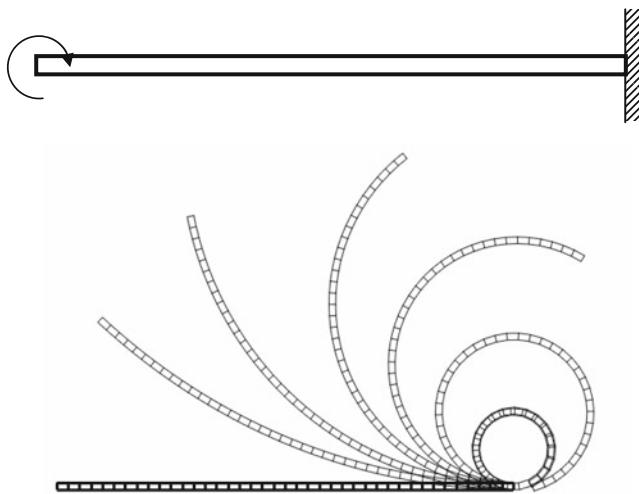


Fig. 3.23 Large deformation of a beam under tip moment

Large deformation of a cantilever beam [15, 16]: A cantilever beam shown in Fig. 3.23 is under bending moment at the tip. The beam is 10 m long with a rectangular cross section of 100 mm x 147.8 mm. Since the length is hundred times larger than the height of cross section, the beam can be considered as a slender member. For material properties, Young's modulus for 100 MPa is assumed. The beam is modeled by 40 CPS4I elements in Abaqus. Since the beam is modeled by plane solid elements, the moment is applied through a distributing coupling constraint. The distributing coupling constraint is used to couple the nodes at the cantilever tip to a reference node placed at the tip. The moment of 3384.78 N-m is applied to this reference node, resulting in a force-couple at the bottom and top nodes of the cantilever tip. The value of tip moment is selected so that the beam can wind twice, based on analytical relation of $ML/EI = 2\pi n$ with $n = 2$. Figure 3.23 shows the deformed plots of beam during nonlinear analysis. As expected, the beam winds twice when the total bending moment is applied.

3.9 Fitting Hyperelastic Material Parameters from Test Data

Although the hyperelastic materials in the previous sections can represent complex behaviors of elastomers, it is difficult to obtain material parameters from experiments. Compared to well-established test procedures for metallic materials [17], the appropriate experiments are not yet clearly defined by national or international standard organizations. This is partly because the complex mathematical models are required to define the nonlinear and the nearly incompressible attributes of

elastomers. This section explains how to determine hyperelastic material parameters from experiments. For simplicity, we will only present the method using a Mooney-Rivlin material. The material parameters of other types of hyperelastic models can be determined in a similar way.

It is important to note that the procedure of material parameter determination should be independent of the finite element model. That is, it would be inappropriate to calculate material parameters by comparing finite element analysis results with experimental data. This is because the finite element results may have numerical errors in calculation. Therefore, experiments should achieve “pure” states of strain such that the stress–strain curve represents the elastomer behavior only in the desired state. In addition, since experiments are not failure oriented, strain or stress is gradually increased in the working range, and data are collected at various points.

3.9.1 Elastomer Test Procedures

Constitutive models for hyperelastic materials are developed from strain energy functions and require nominal stress vs. nominal strain data to fit most models available. In general, it is desirable to represent the three major strain states which are uniaxial tension, uniaxial compression, and pure shear. If compressibility is a concern, then bulk compressibility information is also recommended. For incompressible elastomers, the basic strain states are simple tension, simple compression, equi-biaxial tension, simple shear, pure shear, and volumetric compression. It will be shown later that equi-biaxial tension is equivalent to simple compression. The volumetric compression test is used to determine nearly incompressible attributes of the elastomer through the dilatational parameter D . The other four tests are used to determine the distortional constant A_{10} and A_{01} . Most hyperelastic material models share common test data input requirements. In general, engineering stress and engineering strain² data sets are obtained by stretching the elastomer in several modes of deformation. The engineering stress is the current force divided by the original area, and the engineering strain is the change in length divided by the original length. All test data presented and discussed herein will use engineering stress and engineering strain measures. These data sets are fitted to determine the parameters in material models through least-square method. Figure 3.24 shows a typical set of three stress–strain curves appropriate for input into fitting routines.

Simple tension test: Simple tension test is very popular for elastomers and is similar to the conventional tension test in metals. The most significant difference from the standardized test methods is that in order to achieve a state of pure tensile strain, the specimen should be much longer in the direction of stretching than in the width and thickness dimensions. In such a case, the specimen is at least 10 times longer than the

² Since engineering stress and strain are often considered in infinitesimal deformation, it would be appropriate to call them as nominal stress and nominal strain.

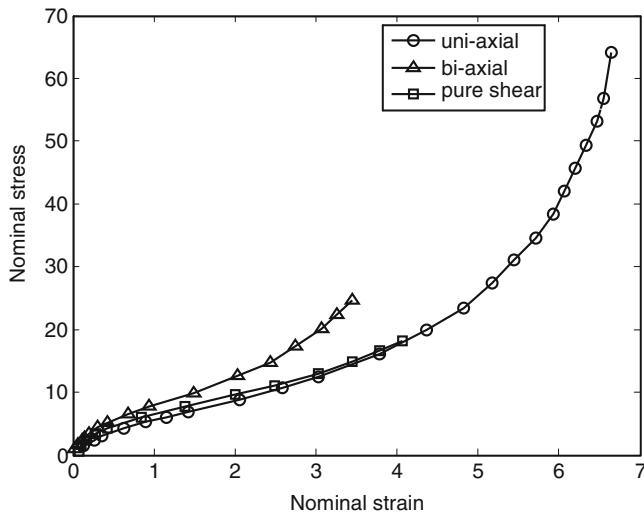


Fig. 3.24 Typical test data for elastomers [19]

width or thickness, so that there is no lateral constraint to specimen thinning. Also, the cross section of the specimen is rectangular, as it is cut from a sheet. Since the instrument clamps can produce complex stress and strain states, the specimen strain must be measured on the specimen, but away from the clamp, where a pure tension strain state is occurring. A non-contacting strain measuring device such as a video extensometer or laser extensometer is often used for this purpose.

Pure shear test: A state of pure shear can be obtained by twisting a circular shaft. Due to flexibility of elastomer, however, it is difficult to conduct a torsional test. Instead, a test similar to simple tension test is used to generate a pure shear stress state. Because the material is nearly incompressible, a state of pure shear exists in the specimen at a 45° angle to the stretching direction. The specimen is perfectly constrained in the lateral direction such that all specimen thinning occurs in the thickness direction. The specimen must be at least 10 times wider than the length in the stretching direction.

Simple compression test: Since most elastomers show quite different behaviors between tension and compression, it is often required to provide compression test data by compressing a specimen using two platens. The compression test specimen is in the shape of a thick button. Due to extra thickness, the fabrication process can be different from other test specimen. Also, it is difficult to achieve a pure state of strain during the compression test because of the friction between the specimen and instrument. As the cross section increases during compression, the specimen cannot freely expand in the contacting surface, which generates a shear stress. Even a small friction coefficient such as 0.1 can cause substantial shear strain; often the maximum shear strain exceeds the maximum compression strain.

Equi-biaxial tension test: For incompressible or nearly incompressible materials, equi-biaxial extension of a specimen creates a state of strain equivalent to pure compression. Although the actual experiment is more complex than the simple compression experiment, a pure state of strain can be achieved which results in a more accurate material model. The equi-biaxial strain state may be achieved by radially stretching a circular disc. As an elastomer is radially strained in all directions in a single plane, the state of strain in the material is the same as that in simple compression. The measured experimental parameters are radial strain and the radial stress. These biaxial strains and biaxial stresses can be converted directly to compression strains and compression stresses as follows:

$$\sigma_C = \sigma_b(1 + \epsilon_b)^3$$

$$\epsilon_C = \frac{1}{(\epsilon_b + 1)^2} - 1$$

where σ_C is nominal engineering compression stress, σ_b is nominal biaxial extension stress, ϵ_C is nominal engineering compression strain, and ϵ_b is nominal biaxial extension strain. The biaxial stress can be obtained by $\sigma_b = F/(\pi^*D^*t)$, where D and t are, respectively, the diameter and thickness of the specimen.

Volumetric compression test: Volumetric compression is an experiment where the near-incompressibility of the material is examined. In this experiment, a cylindrical specimen is constrained in a fixture and compressed. The actual displacement during compression is very small, and great care must be taken to measure only the specimen compliance and not the stiffness of the instrument itself. The initial slope of the resulting stress-strain function is the bulk modulus. This value is typically 2–3 orders of magnitude greater than the shear modulus for elastomers.

3.9.2 Data Preparation

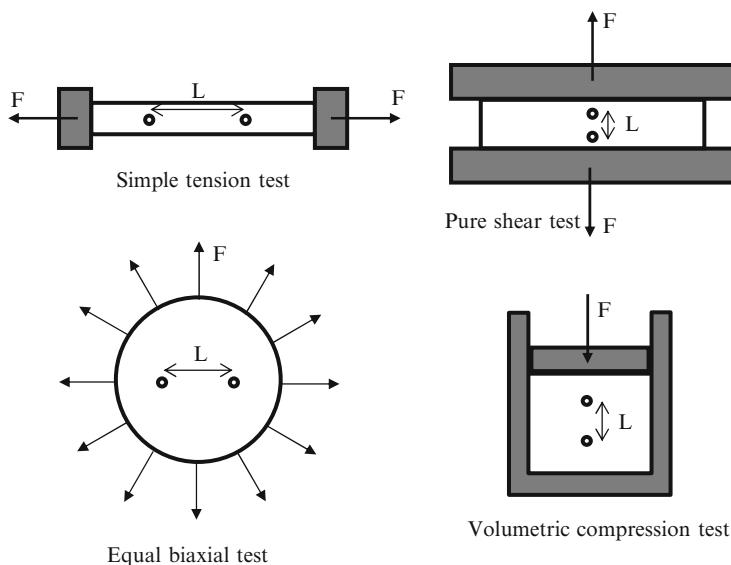
It is possible to use any combinations of the above experiments for finding material parameters. However, it is important that enough number of independent experimental data must be provided so that the curve-fitting algorithm does not have rank deficiency. Also, multiple test types are recommended in order to capture different behaviors of the material.

All experimental data must be converted into stress–strain pairs. As mentioned before, nominal stress vs. nominal stretch will be used for this purpose. Since all tests measure principal stress and principle stretch, the following discussions will also be in the principle values. Table 3.5 summarizes types of data used in curve fitting. It is noted that all tests measure the principle stretches between two points as shown in Fig. 3.25 and all principal stresses are based on the force divided by the initial area.

For Mooney-Rivlin material, it is necessary to determine the three material parameters: A_{10} , A_{01} , and K . In practical point of view, since volumetric

Table 3.5 Measuring stress and strain for elastomer characterization tests

Experiment type	Stretch	Stress
Uniaxial tension	Stretch ratio $\lambda = L/L_0$	Nominal stress $T^E = F/A_0$
Equi-biaxial tension	Stretch ratio $\lambda = L/L_0$ in y-direction	Nominal stress $T^E = F/A_0$ in y-direction
Pure shear test	Stretch ratio $\lambda = L/L_0$	Nominal stress $T^E = F/A_0$
Volumetric test	Compression ratio $\lambda = L/L_0$	Pressure $T^E = F/A_0$

**Fig. 3.25** Types of elastomer tests

compression test is difficult, K is often assumed based on a desired level of incompressibility. For example, it is possible to calculate equivalent K corresponding to Poisson's ratio of 0.499. In general, most hyperelastic materials show incompressible or nearly incompressible behavior. However, in the process of curve fitting, it is assumed that the material is completely incompressible to determine the distortional coefficients A_{10} and A_{01} .

Since all tests yield a simple stress state, it is convenient to present in the coordinates parallel to the principle directions. Then, the stretch ratios that are measured in Table 3.5 are nothing but the principle stretches. The deformation gradient is then defined in terms of the principle stretches as

$$\mathbf{F} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

where λ_1 , λ_2 , and λ_3 are the principle stretches. If nominal strains are measures, then the principle stretches can easily be calculated by $\lambda_i = 1 + \epsilon_i$.

Because of incompressibility, $\det(\mathbf{F}) = \lambda_1\lambda_2\lambda_3 = 1$. In such a case, the deviatoric invariants can be written in terms of the principle stretches as

$$\begin{aligned} J_1 &= \lambda_1^2 + \lambda_2^2 + \lambda_3^2 \\ J_2 &= \lambda_1^{-2} + \lambda_2^{-2} + \lambda_3^{-2}. \end{aligned}$$

Using the incompressibility condition, all three principle stretches can be identified once one of them is measured.

From the assumption of incompressibility, the strain energy density function can be written in terms of the two invariants as

$$U = A_{10}(J_1 - 3) + A_{01}(J_2 - 3) \quad (3.154)$$

Therefore, the objective here is to analytically calculate the nominal stress as a function of principle stretches. Since different tests have different principle stretches, individual tests are treated separately.

Uniaxial test: The three principle stretches for uniaxial tension test can be written in terms of the measured principle stretch, $\lambda = \lambda_1$, as

$$\lambda_1 = \lambda, \quad \lambda_2 = \lambda_3 = 1/\sqrt{\lambda}$$

The principle stress can be obtained by differentiating the strain energy density with respect to the principal strain. Since the principle strain and the principle stretch have a relation of $\lambda_i = 1 + \varepsilon_i$, the principle stress can also be obtained by differentiating with respect to the principle stretch, that is,

$$T = \frac{\partial U}{\partial \lambda} = 2(1 - \lambda^{-3})(A_{10}\lambda + A_{01}) \quad (3.155)$$

Note that the nominal stress is a linear function of material parameters. Therefore, the linear least-square method can be used to find these coefficients. In order to use the least-square method, the above expression is rewritten as

$$T(A_{10}, A_{01}, \lambda) = \{\mathbf{x}\}^T \{\mathbf{b}\} = [2(\lambda - \lambda^{-2}) \quad 2(1 - \lambda^{-3})] \begin{Bmatrix} A_{10} \\ A_{01} \end{Bmatrix} \quad (3.156)$$

where $\{\mathbf{x}\}^T$ is a row vector, which will be used in the linear regression process, and $\{\mathbf{b}\}$ is the vector of unknown coefficients.

Equi-biaxial test: The three principle stretches for equi-biaxial test can be written in terms of the measured principle stretch, $\lambda = \lambda_1$, as

$$\lambda_1 = \lambda_2 = \lambda, \quad \lambda_3 = 1/\lambda^2$$

Since the two principle stretches are applied, the nominal principle stress can be calculated by

$$T = \frac{1}{2} \frac{\partial U}{\partial \lambda} = 2(\lambda - \lambda^{-5})(A_{10} + \lambda^2 A_{01}) \quad (3.157)$$

In the case of equi-biaxial test, the row vector \mathbf{x}^T is defined as $[2(\lambda - \lambda^{-5}), 2(\lambda^3 - \lambda^{-3})]$.

Pure shear test (planar test): The three principle stretches for pure shear test can be written in terms of the measured principle stretch, $\lambda = \lambda_1$, as

$$\lambda_1 = \lambda, \quad \lambda_2 = 1, \quad \lambda_3 = 1/\lambda$$

where λ_1 is the principle stretch in the loading direction. In order to show that this test is equivalent to the pure shear state, the following logarithmic strain can be calculated:

$$\varepsilon_1 = \ln \lambda_1 = -\ln \lambda_3 = -\varepsilon_3, \quad \varepsilon_2 = \ln \lambda_2 = 0$$

which corresponds to a pure shear state at an angle of 45° to the loading direction. The nominal principle stress can be obtained by

$$T = \frac{\partial U}{\partial \lambda} = 2(\lambda - \lambda^{-3})(A_{10} + A_{01}) \quad (3.158)$$

In the case of pure shear test, the row vector \mathbf{x}^T is defined as $[2(\lambda - \lambda^{-3}), 2(\lambda - \lambda^{-3})]$.

Volumetric test: The three principle stretches for volumetric compression test can be written in terms of the measured principle stretch, $\lambda = \lambda_1$, as

$$\lambda_1 = \lambda_2 = \lambda_3 = \lambda$$

Therefore, $J_1 = J_2 = 3$, and $J_3 = \lambda^3 = V/V_0$ (volume ratio). For the Mooney-Rivlin material model, the hydrostatic pressure can be written as

$$p = K(\lambda^3 - 1) \quad (3.159)$$

Similar to uniaxial compression test, the volumetric test will also experience shear deformation between the elastomer and instrument. However, since the magnitude of shear stress is orders of magnitude smaller than the hydrostatic pressure, its effect can be negligible.

3.9.3 Curve Fitting

In general, the curve fitting can be performed in two stages. If the volumetric compression data is available, it is used to determine K first. If the volumetric compression data is not available, the bulk modulus can be estimated from Poisson's ratio or can be assumed. After that, all other test data are used simultaneously to determine constants A_{10} and A_{01} .

Let us assume that there are an NDT number of experimental data. For example, when 10 uniaxial tension test data and 5 pure shear test data are available, NDT = 15. Then, the following array is created that contains test types, stretches, and stresses for all data:

Type	1	1	1	...	4	4	...	4
λ	λ_1	λ_2	λ_3	...	λ_i	λ_{i+1}	...	λ_{NDT}
T^E	T_1^E	T_2^E	T_3^E	...	T_i^E	T_{i+1}^E	...	T_{NDT}^E

The superscript “ E ” in the stress is used because it is experimental data. The objective of curve fitting is to find material constants, A_{10} and A_{01} , such that the difference between measured stress and calculated stress is minimized.

For the assumed material constants A_{10} and A_{01} , the nominal stress can be calculated using the principle stretches in Eqs. (3.155), (3.157), and (3.158). Since there are NDT stretches, the nominal stress is calculated at each of these stretches, which are called $T(A_{10}, A_{01}, \lambda_k)$, $k = 1, \dots, NDT$. Then, the curve-fitting process is to find the material constants, A_{10} and A_{01} , to minimize the difference between experimentally obtained stress and calculated stress using the material constants:

$$\underset{A_{10}, A_{01}}{\text{minimize}} \quad \sum_{k=1}^{NDT} (T_k^E - T(A_{10}, A_{01}, \lambda_k))^2 \quad (3.160)$$

The above equation can be solved using an optimization algorithm and/or regression method. In the case of Mooney-Rivlin material, $T(A_{10}, A_{01}, \lambda_k)$ is linear with respect to A_{10} and A_{01} . Therefore, a simple linear least-square regression can be used for curve fitting. However, other materials, such as Ogden material, need an optimization algorithm. In the following, the linear regression method is explained.

In order to apply the least-square method, nominal stresses from test data and from model prediction are compared at the same value of principle stretch. Since there are NDT test data, the same number of model predictions is combined together in the following matrix form:

$$\{\mathbf{T}\} = \begin{Bmatrix} T_1 \\ T_2 \\ \vdots \\ T_{NDT} \end{Bmatrix} = \begin{bmatrix} \mathbf{x}(\lambda_1)^T \\ \mathbf{x}(\lambda_2)^T \\ \vdots \\ \mathbf{x}(\lambda_{NDT})^T \end{bmatrix} \{\mathbf{b}\} = [\mathbf{X}] \{\mathbf{b}\}$$

where $\{\mathbf{x}(\lambda_k)\}^T$ is the row vector defined at each test type and $\{\mathbf{b}\}$ is the vector of the unknown material parameters. Note that the dimension of coefficient matrix $[\mathbf{X}]$ is $NDT \times 2$. Also, the nominal stresses from all tests are combined together to build the following vector:

$$\{\mathbf{T}^E\} = \begin{Bmatrix} T_1^E \\ T_2^E \\ \vdots \\ T_{NDT}^E \end{Bmatrix}$$

Then, the objective function of the optimization problem in Eq. (3.160) can be rewritten in terms of vector notation as

$$\begin{aligned} \{\mathbf{e}\}^T \{\mathbf{e}\} &= \{\mathbf{T}^E - \mathbf{T}\}^T \{\mathbf{T}^E - \mathbf{T}\} \\ &= \{\mathbf{T}^E - \mathbf{Xb}\}^T \{\mathbf{T}^E - \mathbf{Xb}\} \\ &= \{\mathbf{T}^E\}^T \{\mathbf{T}^E\} - 2\{\mathbf{b}\}^T [\mathbf{X}]^T \{\mathbf{T}^E\} + \{\mathbf{b}\}^T [\mathbf{X}]^T [\mathbf{X}] \{\mathbf{b}\} \end{aligned}$$

The above equation is nothing but the sum of squares of errors between test and model prediction. The minimum of the above error that can be obtained by setting the derivative of the above expression with respect to $\{\mathbf{b}\}$ is equal to zero. Therefore, the following form of least-square equation can be obtained:

$$[\mathbf{X}]^T [\mathbf{X}] \{\mathbf{b}\} = [\mathbf{X}]^T \{\mathbf{T}^E\} \quad (3.161)$$

Since the dimension of the coefficient matrix $[\mathbf{X}]^T [\mathbf{X}]$ is 2x2, the above equation can be solved easily.

3.9.4 Stability of Constitutive Model

In uniaxial tension of a linear elastic material, the slope in the stress-strain curve is always positive. It means that in order to extend the material, the force must be increased; this is a fundamental requirement for a stable material. If the slope is negative, it is possible that application of a load to a material point can lead to arbitrary deformations. Such a requirement of stable material is called Drucker stability [18]. Material stability is not an issue for linear elastic material because the material will always be stable with a positive Young's modulus and Poisson's ratio. In the case of nonlinear material, however, it is possible that the slope becomes negative locally, especially when the material parameters are identified by fitting test data. Therefore, it is important to check material stability after material parameters are determined from test data.

In the case of Mooney-Rivlin hyperelastic material, the stability requirement can be defined in terms of the incremental work done by arbitrary change in deformation, as

$$d\sigma : d\epsilon > 0 \quad (3.162)$$

Using the tangent stiffness, the above equation can be rewritten in terms of arbitrary strain increment as

$$d\epsilon : \mathbf{D} : d\epsilon > 0$$

The above requirement is identical to the positive-definite property of the material tensor. It is clear that since \mathbf{D} is positive definite for a linear elastic material, the material satisfies the stability requirement. However, in the case of nonlinear materials, the material tensor \mathbf{D} is a function of deformation, and it is impractical to check all possible deformations. Therefore, the stability check is normally performed at several specified deformations, such as uniaxial tension and compression, equi-biaxial tension and compression, and planar tension and compression. Also, the range of the stretch ratio over which the stability is checked can be chosen from 0.1 to 10.

Since the above deformations for checking stability are all in the principle directions, the stability requirement can be written in the principle components as

$$d\sigma_1 d\epsilon_1 + d\sigma_2 d\epsilon_2 + d\sigma_3 d\epsilon_3 > 0$$

In addition, in the case of incompressible materials, the hydrostatic pressure $p = (\sigma_1 + \sigma_2 + \sigma_3)/3$ cannot cause any deformation. Therefore, it is possible to choose any arbitrary hydrostatic pressure value. For convenience, p is chosen such that $\sigma_3 = d\sigma_3 = 0$. Then, the above stability requirement can further be simplified by

$$d\sigma_1 d\epsilon_1 + d\sigma_2 d\epsilon_2 > 0$$

By using the incremental stress-strain relation in the principle components, the above stability requirement can be written as

$$\begin{Bmatrix} d\epsilon_1 & d\epsilon_2 \end{Bmatrix} \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \begin{Bmatrix} d\epsilon_1 \\ d\epsilon_2 \end{Bmatrix} > 0$$

where D_{11} , D_{12} , D_{21} , and D_{22} are components of material tensor for a hyperelastic materials. For example, they can be calculated from Eq. (3.128) by considering the three coordinate directions are principle directions. For Mooney-Rivlin materials, the components of material tensor can be calculated by

$$\begin{aligned} D_{11} &= 4(\lambda_1^2 + \lambda_3^2)(A_{10} + \lambda_1^2 A_{01}) \\ D_{22} &= 4(\lambda_2^2 + \lambda_3^2)(A_{10} + \lambda_2^2 A_{01}) \\ D_{12} = D_{21} &= 4\lambda_3^2 A_{10} + 4\lambda_3^{-2} A_{01} \end{aligned}$$

In order to satisfy the stability requirement, the material tensor must be positive definite, which is equivalent to the following requirements:

$$\begin{aligned} D_{11} + D_{22} &> 0 \\ D_{11}D_{22} - D_{12}D_{21} &> 0 \end{aligned} \tag{3.163}$$

Note that these requirements must satisfy for all possible deformations.

3.10 Summary

In this chapter, finite element formulations for nonlinear elastic problems are discussed. Among four different nonlinearities, geometric and material nonlinearities are considered. All nonlinearities in the chapter are considered to be mild because they are path independent and the status of a material does not change abruptly. For large deformation, the deformation gradient plays an important role in connecting undeformed and deformed states. The left and right Cauchy–Green deformation tensors are used to describe large deformation. It is shown that every deformation can uniquely be decomposed into a principal stretch and a rigid-body rotation. Depending on the frame of reference, different stress and strain measures need to be used in large deformation. Engineering, Lagrangian, and Eulerian strains are defined. It has been shown that Lagrangian and Eulerian strains are independent of rigid-body motion, while engineering strain is not. The definition of stress also depends on the frame of reference. Cauchy stress and the first and second Piola-Kirchhoff stresses are defined. These stresses are related to each other through the deformation gradient.

Nonlinear elastic systems can be modeled using either the total or updated Lagrangian formulation. Since the total Lagrangian formulation uses the undeformed state as a frame of reference, the Lagrangian strain and the second Piola-Kirchhoff stress are used to describe the status of a material. The principle of minimum potential energy is used to derive the nonlinear variational equation, and the tangent stiffness is calculated through linearization for the Newton–Raphson method. Since the updated Lagrangian formulation uses the deformed state as a frame of reference, the engineering strain and the Cauchy stress are used to describe the status of a material. Instead of deriving the nonlinear variational equation and its linearization, the expressions in the total Lagrangian formulation are transformed to the deformed state using the deformation gradient. It has been shown that the two formulations are mathematically identical. Selection between two methods should be based on the given constitutive relation and convenience of computer implementation.

In nonlinear analysis, load–displacement curves sometimes do not show a monotonic trend due to instability of the system, such as bifurcation or buckling. The curve starts decreasing after it reaches the maximum load, which is called a critical load. Normally nonlinear static analysis can be performed successfully for the applied load that is less than the critical load. Then, the critical load can be estimated using either the one-point or two-point method. The basic idea is to find the state in which the tangent stiffness becomes singular. More advanced methods such as the arclength method can be used to find the actual critical load as well as the behavior of the system (post-buckling analysis).

Hyperelasticity includes both geometric and material nonlinearities. It is accustomed to large deformation, and the stress–strain relation is nonlinear, which can be obtained by differentiating the strain energy density. Several hyperelastic material models are introduced, but the derivation for the nonlinear variational equation is explained for the Mooney-Rivlin material. Hyperelastic materials, and also most nonlinear elastic materials, show incompressibility or near-incompressibility. When the penalty method with a large value of bulk modulus is used, numerical instability,

which is called volumetric locking, occurs. In order to eliminate or reduce volumetric locking, several approaches are introduced, including selective reduced integration, the mixed formulation, and the perturbed Lagrangian formulation. The first one is the most convenient because only the integration scheme is changed. The other two methods require introducing the hydrostatic pressure as an independent variable.

3.11 Exercises

- P3.1 Derive the expression of the Eulerian strain in Eq. (3.17).
- P3.2 Derive the relation in volume change in Eq. (3.26) for an infinitesimal hexahedron whose edges are initially parallel to the coordinate directions.
- P3.3 Consider a square block under oscillating simple shear deformation. The relation between undeformed and deformed geometry is given as

$$x_1 = X_1 + aX_2 \sin \omega t, \quad x_2 = X_2, \quad x_3 = X_3$$

Calculate the deformation gradient and the change in volume.

- P3.4 Many materials often show very different behaviors between volume-changing deformation and volume-preserving deformation. The former is called dilatation, while the latter is called distortion. In such a case, it is necessary to separate the dilatational and distortional parts from the deformation gradient. For example, the deformation gradient can be decomposed into $\mathbf{F} = \mathbf{F}_v \cdot \mathbf{F}_d$, where \mathbf{F}_v is the dilatational part and \mathbf{F}_d is the distortional part. Calculate \mathbf{F}_v and \mathbf{F}_d using the third invariant of the deformation tensor.
- P3.5 Repeat Problem P3.4 for the Cauchy–Green deformation tensor; i.e., decompose $\mathbf{C} = \mathbf{C}_v \cdot \mathbf{C}_d$.
- P3.6 Consider a bar with a square cross section in the figure under uniaxial tension loading. The principal stretch in the X_1 direction is given by $\lambda > 1$. When material is incompressible, compare the X_1 component of normal strain using Lagrangian, Eulerian, and engineering strains.

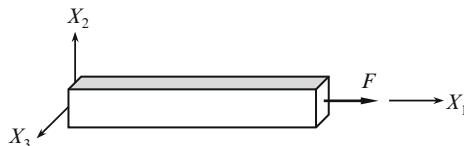
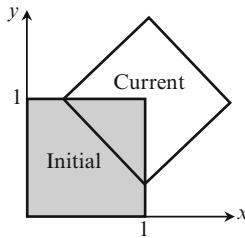


Fig. P3.6

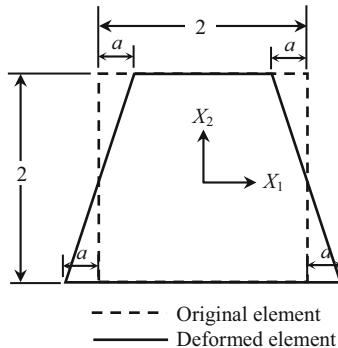
- P3.7 A four-node square element undergoes large displacement and rotation in the XY plane, as shown in the figure. The node initially at the origin is moved to $(1, 1 - \sin\pi/4)$ and the element is rotated by 45° . Calculate the deformation gradient. Compute the Lagrangian strains and demonstrate that no strain occurs during rigid-body motion.

**Fig. P3.7**

P3.8 A square plane-strain element is deformed as shown in the figure. The relationship between deformed and undeformed coordinates is given as

$$x_1 = X_1 - aX_1X_2, \quad x_2 = X_2, \quad x_3 = X_3$$

Compare the engineering strain and Lagrangian strain. Show that the two strain measures become identical as “ a ” approaches zero.

**Fig. P3.8**

P3.9 The relationship between deformed and undeformed coordinates for the pure bending of a plane-strain solid is given as

$$x_1 = X_1 - aX_1X_2, \quad x_2 = X_2 + \frac{1}{2}aX_1^2, \quad x_3 = X_3$$

Compare the engineering strain and Lagrangian strain. Show that the two strain measures become identical as “ a ” approaches zero.

P3.10 In the small deformation theory, the volumetric strain $(dV_x - dV_0)/dV_0$ is approximated by $\epsilon_{11} + \epsilon_{22} + \epsilon_{33}$, while in the large deformation theory, it is given by $J - 1$. Show that when the deformation is small, the latter can be approximated by the former.

P3.11 An initially straight beam AB is bent into a circular arc A'B' as shown in the figure. The deformation is specified as

$$x_1 = g(X_2) \cos \frac{\pi(1 - X_1)}{2}, \quad x_2 = g(X_2) \sin \frac{\pi(1 - X_1)}{2}, \quad x_3 = X_3$$

where $g(X_2)$ is a simple function of X_2 . (a) Find the deformation gradient in terms of $g(X_2)$. (b) If the volume of the beam does not change, find $g(X_2)$. (c) Using $g(X_2)$ in (b), find \mathbf{U} , \mathbf{Q} , and \mathbf{V} .

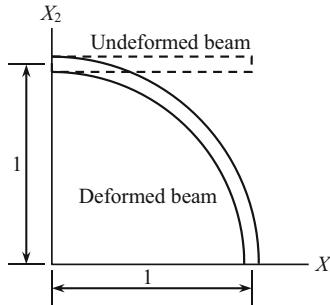


Fig. P3.11

P3.12 Consider a square element under pure shear deformation as shown in the figure. The relation between deformed and undeformed coordinates becomes

$$x_1 = X_1 + kX_2, \quad x_2 = kX_1 + X_2$$

(a) Calculate deformation gradient \mathbf{F} , Lagrangian strain \mathbf{E} , Eulerian strain \mathbf{e} , and engineering strain $\boldsymbol{\epsilon}$. (b) Calculate principal stretch tensors \mathbf{U} and \mathbf{V} and rotation tensor \mathbf{Q} .

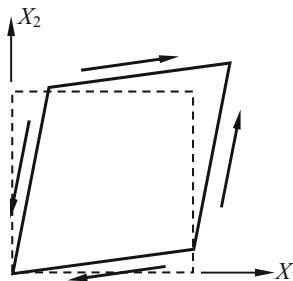


Fig. P3.12

P3.13 A square block of surface area A on all sides is under pure shear deformation due to the uniformly distributed load F on the top surface, as shown in the figure. The deformation of the block is such that the deformed coordinates can be written as

$$x_1 = X_1 + aX_2, \quad x_2 = X_2, \quad x_3 = X_3$$

Calculate Cauchy stress and the first and second Piola-Kirchhoff stresses.

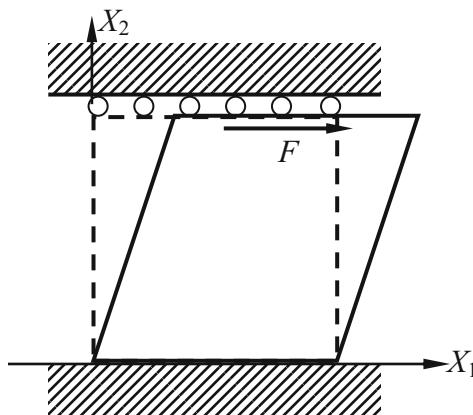


Fig. P3.13

- P3.14 A force F is applied at the tip of the uniform bar element shown in the figure. The initial length and the cross-sectional area of the bar are, respectively, A_0 and L_0 . The elastic modulus of the material is E . Calculate the tip displacement by solving the total Lagrangian variational equation with the St. Venant–Kirchhoff nonlinear elastic material model. Assume the following numerical values: $E = 700$ MPa, $A_0 = 1.0 \times 10^{-4}$ m 2 , $L_0 = 1.0$ m, and $R = 10$ kN. Compare the tip displacement with that from the linear elastic model when (a) $E = 700$ MPa and (b) $E = 70$ GPa.

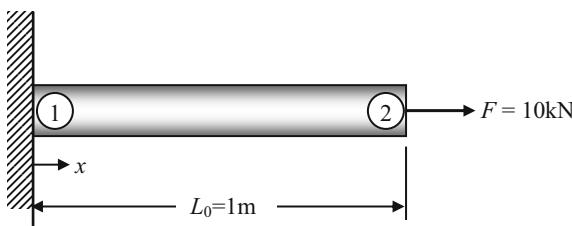


Fig. P3.14

- P3.15 Solve Problem P3.14 using force equilibrium; i.e., internal force caused by stress is equal to external force.

- P3.16 Consider a plane-strain square element with unit depth as shown in the figure. Use the St. Venant–Kirchhoff isotropic material model with two Lame's constants λ and μ . A uniformly distributed force T_x (force per area) is horizontally applied at the top surface. Assuming it is a simple shear problem, the deformation of the element can be written as

$$\begin{cases} x_1 = X_1 + kX_2 \\ x_2 = X_2 \end{cases}$$

- (a) Find the relation between k and T_x , (b) find the reaction force in the X_2 direction at the top surface, and (c) compare the results with that of the linear elastic model.

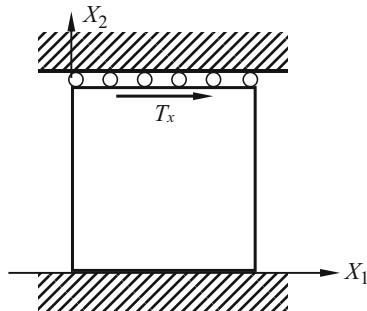


Fig. P3.16

- P3.17 Consider a deformation of a rectangular bar whose deformed geometry is given as

$$x_1 = \alpha X_1, \quad x_2 = \beta X_2, \quad x_3 = \beta X_3$$

When the material is incompressible and St. Venant-Kirchhoff material properties are given as $E = 600$ MPa and $\nu = 0.49$, write the expression of the S_{11} component of the second Piola-Kirchhoff stress as a function of α . In addition, write the expression of σ_{11} of the Cauchy stress as a function of α . Plot S_{11} and σ_{11} in the range of $\alpha = [0.7 \text{ } 1.5]$.

- P3.18 Consider a simple shear deformation of a square whose deformed geometry is given as

$$x_1 = X_1 + \alpha X_2, \quad x_2 = X_2, \quad x_3 = X_3$$

When the material is incompressible and St. Venant-Kirchhoff material properties are given as $E = 600$ MPa and $\nu = 0.49$, write the expression of the S_{12} component of the second Piola-Kirchhoff stress as a function of α . In addition, write the expression of σ_{12} of the Cauchy stress as a function of α . Plot S_{12} and σ_{12} in the range of $\alpha = [0.0 \text{ } 1.5]$.

- P3.19 Consider the following deformation with $|\alpha| \leq 1$:

$$x_1 = X_1 + \alpha X_2, \quad x_2 = \sqrt{1 - \alpha^2} X_2, \quad x_3 = X_3$$

Assume St. Venant-Kirchhoff material with two material parameters λ and μ . (a) Show that the above deformation is a pure shear deformation in terms

of the Lagrangian strain. (b) Calculate the second Piola-Kirchhoff stress and Cauchy stress in terms of α , λ , and μ .

- P3.20 A force F is applied at the tip of the uniform bar shown in the figure. The displacement of the bar is given as $u(X) = \lambda X$ where λ is the principal stretch. The initial length and the cross-sectional area of the bar are, respectively, A_0 and L_0 . The elastic modulus of the material is E . Calculate the tip displacement by solving for the principal stretch using the total Lagrangian formulation with the St. Venant-Kirchhoff material model. Assume the following numerical values: $E = 700$ MPa, $A_0 = 1.0 \times 10^{-4}$ m 2 , $L_0 = 1.0$ m, and $F = 10$ kN. Compare the tip displacement with that of the linear elastic model when (a) $E = 700$ MPa and (b) $E = 70$ GPa.

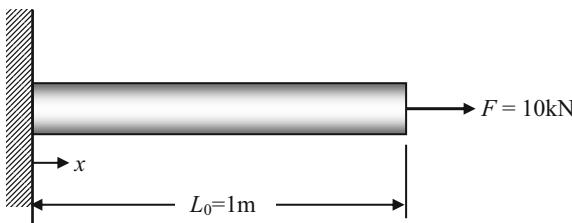


Fig. P3.20

- P3.21 Solve Problem P3.20 using force equilibrium; i.e., internal force caused by stress is equal to external force.

- P3.22 Consider two bar elements under a force at the tip. Using the displacement-controlled method, plot the load-displacement curve (F vs. u_2 and u_3). Increase the tip displacement u_3 up to 1.0 m by ten equal increments. Assume St. Venant-Kirchhoff material with $E = 100$ MPa and cross-sectional areas of $A^{(1)} = 1.0 \times 10^{-4}$ m 2 and $A^{(2)} = 0.5 \times 10^{-4}$ m 2 .

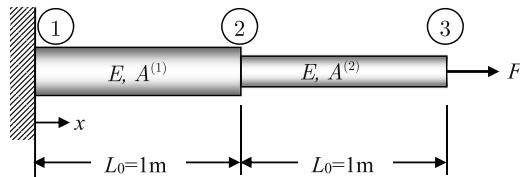


Fig. P3.22

- P3.23 Consider a nonlinear elastic uniaxial bar element under tip force $F = 100$ N shown in Fig. 3.11. The stress-strain relation is given in terms of Cauchy stress and engineering strain in the deformed geometry: $\sigma_{11} = E\varepsilon_{11}$. Using the updated Lagrangian formulation, solve for displacement at the tip and the stress and strain of the uniaxial bar. Assume $E = 200$ Pa and the cross-sectional area $A = 1.0$ m 2 .

- P3.24 Consider a deformation of a rectangular bar whose deformed geometry is given as

$$x_1 = \alpha X_1, \quad x_2 = \beta X_2, \quad x_3 = \beta X_3$$

When the material is an incompressible Mooney-Rivlin hyperelastic material with $A_{10} = 80$ MPa and $A_{01} = 20$ MPa, write the expression of the S_{11} component of the second Piola-Kirchhoff stress as a function of α . In addition, write the expression of σ_{11} of the Cauchy stress as a function of α . Plot S_{11} and σ_{11} in the range of $\alpha = [0.7, 1.5]$.

- P3.25 Consider a simple shear deformation of a square whose deformed geometry is given as

$$x_1 = X_1 + \alpha X_2, \quad x_2 = X_2, \quad x_3 = X_3$$

When the material is incompressible Mooney-Rivlin hyperelastic material with $A_{10} = 80$ MPa and $A_{01} = 20$ MPa, write the expression of the S_{12} component of the second Piola-Kirchhoff stress as a function of α . In addition, write the expression of σ_{12} of the Cauchy stress as a function of α . Plot S_{12} and σ_{12} in the range of $\alpha = [0.0, 1.5]$.

- P3.26 Derive the energy form and its linearization of a Mooney-Rivlin hyperelastic material using the perturbed Lagrangian method. Use a mixed variable $\mathbf{r} = [\mathbf{u}^T, p]^T$.

- P3.27 Derive the 3×3 $[\mathbf{D}]$ matrix in Eq. (3.147) for a two-dimensional Mooney-Rivlin material with three material parameters (A_{10} , A_{01} , and K). Use the penalty method for near-incompressibility.

- P3.28 Derive the 3×3 $[\mathbf{D}]$ matrix in Eq. (3.147) for a two-dimensional Mooney-Rivlin material with three material parameters (A_{10} , A_{01} , and K). Use the perturbed Lagrangian method for near-incompressibility.

- P3.29 A nearly incompressible rubber block is confined between two frictionless rigid walls as shown in the figure. When uniform pressure P is applied to the right end, the length of the block is changed by $x_1 = (1 - \alpha)X_1$. When $\alpha = 0.1$, (a) calculate the value of the strain energy density and (b) the magnitude of applied pressure P on the right end. Assume a plane-strain problem and use a Mooney-Rivlin material with $A_{10} = 80$ MPa, $A_{01} = 20$ MPa, and $K = 1,000$ MPa.

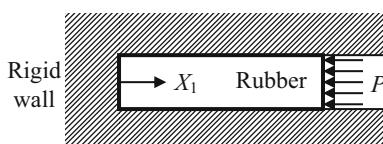


Fig. P3.29

P3.30 Consider a unit cube shown in Fig. 3.15. Using an eight-node solid element, perform biaxial extension analysis using Abaqus. Apply uniform extensions in both X_1 and X_2 directions so that the deformed shape will be $5 \times 5 \times t_3$. Plot stress σ_{11} and thickness t_3 as a function of stretch.

References

1. Belytschko T, Moran B, Liu WK. Nonlinear finite elements for continua and structures. New York, NY: Wiley; 2000.
2. Wriggers P. Nonlinear finite element methods. Berlin: Springer; 2008.
3. Bathe K-J. Finite element procedures in engineering analysis. Englewood Cliffs, NJ: Prentice-Hall; 1996.
4. Malvern LE. Introduction to the mechanics of a continuous medium. Englewood Cliffs, NJ: Prentice-Hall; 1969.
5. Ciarlet PG. Lectures on three-dimensional elasticity. Berlin: Springer; 1983.
6. Brendel B, Ramm E. Linear and nonlinear stability analysis of cylindrical shells. Comput Struct. 1980;12:549–58.
7. Bathe K-J, Dvorkin EN. On the automatic solution of nonlinear finite element equations. Comput Struct. 1983;17:871–9.
8. Brush OB, Almroth BO. Buckling of bars, plates, and shells. New York, NY: McGraw-Hill; 1975.
9. Oden JT, Kikuchi N. Finite element methods for constrained problems in elasticity. Int J Numer Methods Eng. 1982;18:701–25.
10. Malkas DS, Hughes TJR. Mixed finite element methods-reduced and selective integration techniques: a unification of concept. Comput Methods Appl Mech Eng. 1978;15:63–81.
11. Sussman T, Bathe KJ. A finite element formulation for nonlinear incompressible elastic and inelastic analysis. Comput Struct. 1987;26:357–409.
12. Oden JT. Finite elements of nonlinear continua. New York, NY: McGraw-Hill Book Co., Inc.; 1972. p. 325–31.
13. Oden JT. Finite Elements of Nonlinear Continua. New York, NY: McGraw-Hill Book Co., Inc.; 1972. p. 318–21.
14. Treloar LRG. Stress-strain data for vulcanized rubber under various types of deformation. Trans Faraday Soc. 1944;40:59–70.
15. Bissopp RE, Drucker DC. Large deflection of cantilever beams. Q Appl Math. 1945;3 (1):272–5.
16. Love AEH. A treatise on the mathematical theory of elasticity. New York, NY: Dover Publications; 1944.
17. Czichos H. Springer handbook of materials measurement methods. Berlin: Springer; 2006. p. 303–4.
18. Drucker DC. A definition of a stable inelastic material. ASME J Appl Mech. 1959;26:101–95.
19. Treloar LRG. Stress-strain data for vulcanised rubber under various types of deformation. Trans Faraday Soc. 1940;40:59–70.

Chapter 4

Finite Element Analysis for Elastoplastic Problems

4.1 Introduction

The unique property of nonlinear elastic materials in the previous chapter is that a **strain energy density** exists and stress is defined by differentiating it with respect to the appropriate strain. This property of elastic materials is called *history independent*. For example, let the current strain of a nonlinear elastic bar be ε . This state can be reached either by gradually stretching the bar until the strain becomes ε or by gradually compressing it after stretching it beyond the strain ε . Even if these two load histories are different, the bar will have the same value of stress because the current strain is the same. Using the history-independent property, it is easy to conclude that when stress disappears, strain does too. Thus, when the applied load is removed, the structure will always come back to its initial geometry (reversible). No permanent deformation will remain.

Different from elastic materials, some materials, such as steels or aluminum alloys, show permanent deformation when a force larger than a certain limit (elastic limit) is applied and removed. A simple example is bending a paper clip. If a small force is applied and removed, the paper clip comes back to its initial geometry, but when the force is larger than the elastic limit (irreversible), it does not. In contrast to elasticity, this behavior of materials is called plasticity. Since these materials are initially elastic and then become plastic, this behavior of materials is called *elastoplasticity*, which is the main topic of this chapter.

Elastoplasticity, along with **hyperelasticity** in the previous chapter, belongs to material nonlinearity, which comes from the constitutive relation, i.e., from the stress-strain relation. Unlike hyperelasticity, there is no one-to-one relationship between stress and strain for elastoplasticity. For example, let the bar in the previous example be made of an elastoplastic material. Then, for a given value ε of strain, stress can have different values depending on whether the bar is stretching or compressing. In fact, depending on the history of loads, stress can have any value less than the elastic limit. Thus, the **stress-strain relation** cannot be given in terms of

total strain. Instead, the constitutive relation is given in terms of the rates of stress and strain. This type of constitutive relation is called *hypoelasticity*, against hyperelasticity. The term “rate” does not mean the time rate. Rather, it should be understood as an increment in static analysis. Since the relation is given in the rate form, stress can only be calculated by integrating the stress rate over the past load history. Thus, stress calculation is history or path dependent.

A key step in elastoplastic analysis is to **separate elastic and plastic strains from total strain**. Once elastic strain is calculated, stress can easily be calculated from it; plastic strain does not contribute to stress. In fact, various constitutive relations in Chap. 3 can be used by considering all strains to be elastic. When total strain is small (infinitesimal deformation), it is possible to assume that the total strain can be additively decomposed into elastic and plastic strains. In this case, no geometric nonlinearity is considered; i.e., displacement–strain relation is linear and integration is performed over the initial undeformed geometry. Considering that the plastic deformation of metals normally occurs at 0.2 % strain, metal plasticity often satisfies small strain conditions. Sections 4.2 and 4.3 are based on infinitesimal elastoplasticity. In a large structure, even if strain is small, the structure may undergo a large rigid-body motion due to accumulated deformation. In such a case, it is possible to modify infinitesimal elastoplasticity to accommodate stress calculation with the effect of rigid-body motion. Since the rate of Cauchy stress is not independent of rigid-body motion, different types of rates, called objective stress rates, are used in the constitutive relation, which is discussed in Sect. 4.4. When deformation is large, the assumption of additive decomposition of elastic and plastic strains is not valid anymore. A hyperelasticity-based elastoplasticity is discussed in Sect. 4.5 in which the deformation gradient is multiplicatively decomposed into elastic and plastic parts and the stress–strain relation is given in the principal directions. This model can represent both geometric and material nonlinearities during large elastoplastic deformation.

4.2 One-Dimensional Elastoplasticity

Elastoplasticity occurs when a material experiences both elastic and plastic deformation. In this section, the basic concepts of elastoplastic finite element analysis are introduced first using a one-dimensional bar. Generalization to multidimensional elastoplasticity will be presented in Sect. 4.3. It is assumed that deformation is small so that geometric nonlinear effects can be negligible. Therefore, no distinction between different stress measures will be necessary. Elastoplasticity with large deformation will be considered in Sects. 4.4 and 4.5.

4.2.1 Elastoplastic Material Behavior

In a **one-dimensional tension test**, once a material deforms beyond the elastic limit, it shows a **complex stress–strain relation**. For example, metals show that initially

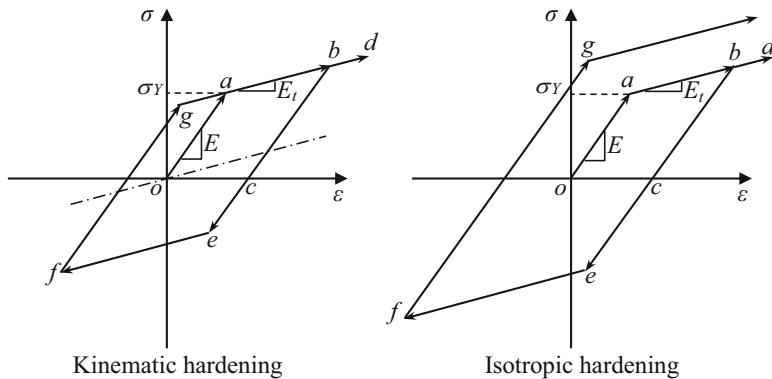


Fig. 4.1 Hardening models for elastoplasticity

stress increases proportional to strain (elastic). After it reaches the elastic limit, also called yield stress, the material starts deforming plastically (see Fig. 1.11 in Chap. 1). At the first stage of plastic deformation, stress further increases proportional to strain but with a much smaller slope (strain-hardening) until it reaches an ultimate strength. After that, stress starts gradually decreasing (strain-softening) until the material fractures. In addition, if the applied load is reduced (unloading) after the material becomes plastic, it does not follow the previous stress–strain curve; the material becomes elastic immediately. If cyclic loads are applied, then the material behavior becomes more and more complicated.

Modeling the behavior of a material depends on the purpose of analysis. For example, if the objective is to find the material behavior until it fractures, it is necessary to model all stages of the stress–strain responses in detail. However, when the objective is to **find the material's response under small deformation**, it is possible to simplify the material behavior by regarding the **elastic and strain-hardening parts only**. Of course, this model is not appropriate to predict fracture of the material. An idealized elastoplastic stress–strain behavior from a uniaxial tension/compression test is shown in Fig. 4.1. When a tension load is applied, the behavior is initially elastic until yield stress, σ_Y , is reached (line $o-a$). The *elastic modulus* is the slope of the line and denoted by E . If the applied load is removed from this region, the stress–strain relation follows the same curve (line $a-o$). Point a is called a yield point, and the material is not elastic anymore beyond this point.

After yielding, the plastic phase begins and stress further increases with a slope of E_t , known as the **tangent modulus** (line $a-b$). During this phase, strain is composed of elastic and plastic parts. **In this simplified model, it is assumed that strain-hardening is linear.** If the load is reduced after the material undergoes a plastic state, it becomes elastic again, and stress decreases linearly with the same initial elastic slope, E (line $b-c$). If the applied load is completely removed, a permanent plastic strain remains (strain at point c). If the load increases again, then

the material follows line $c-b$ and has a new yield stress (stress at point b). After that, the stress increases by following the same slope as the tangent modulus (line $b-d$). In the elastoplastic material, the yield stress changes due to the strain-hardening effect. On the other hand, if a compressive load is applied at point c , which is a stress-free state, the negative value of stress is developed with the slope of E . If the compressive load continues, the material will eventually yield again in compression (point e).

If the applied load is proportional, i.e., the load increases monotonically, the strain-hardening is simply described by the tangent modulus, E_t . However, when combined cyclic loadings and unloadings are applied, it can be complicated depending on when yielding occurs in the opposite direction. Several different hardening models have been proposed for determining the yield stress in cyclic loading situations. Figure 4.1 shows two of the most commonly used models, known as the *kinematic* and the *isotropic hardening*. The kinematic hardening model assumes that the elastic range (twice the initial yield stress) remains constant. The center of the elastic region moves along the dashed line through the origin, parallel to the strain-hardening line. Thus, line segments $b-e$ and $f-g$ are both equal and twice the length $o-a$. In the isotropic hardening model, the magnitude of yield stress for the reversed loading is equal to that of the previous yield stress. That is, the magnitude of stress is the same at points b and e . Thus, the elastic range grows in this model.

In nonlinear finite element analysis, the Newton–Raphson iterative method solves for the displacement increment that can reduce the residual of nonlinear equation. In addition, using the finite element interpolation scheme, it is easy to calculate the strain increment from the given displacement increment. Thus, an important task of elastoplastic analysis is to find stress increment from the given strain increment. However, it is unknown that, out of the given strain increment, how much is elastic and how much is plastic. Once this decomposition is done, the stress increment can be calculated using the elastic strain increment. Note that plastic strain does not contribute to the increase of stress. When the material is elastic, whether it is in the initial elastic phase or becomes elastic due to unloading, the strain increment is purely elastic and no plastic strain increment is present. Then the stress increment can be obtained by multiplying the strain increment with the elastic modulus. Since this procedure is basically identical to linear elastic systems, only the case when the material is in the plastic phase will be discussed.

Figure 4.2 shows the stress–strain curve for a one-dimensional elastoplastic material. When the applied load is proportional, both isotropic and kinematic hardenings provide the same result. At the previous load increment, it is assumed that the material is already in the plastic phase. At the current load increment, the strain increment is given from the Newton–Raphson method. Since the material is in the plastic phase and the load continuously increases, strain increment $\Delta\epsilon$ can be decomposed into elastic and plastic strains. That is,

$$\Delta\epsilon = \Delta\epsilon_e + \Delta\epsilon_p, \quad (4.1)$$

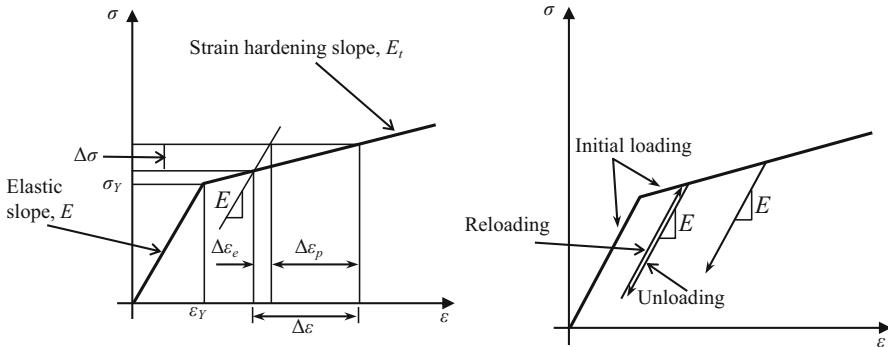


Fig. 4.2 One-dimensional elastoplasticity with strain-hardening

where subscripts e and p denote elastic and plastic, respectively. The elastic portion of strain will be removed when the applied load decreases or changes its direction (unloading), whereas the plastic portion remains constant during elastic unloading and increases again when the material yields. The plastic strain can only increase even if the material yields in opposite directions; it never decreases as it is an accumulation of plastic deformation. From the assumption of small deformation, Eq. (4.1) can be used to yield

$$\varepsilon = \varepsilon_e + \varepsilon_p, \quad (4.2)$$

where ε_e and ε_p are the sum of elastic and plastic strain increments, respectively, for all previous load increments.

Equation (4.2) provides a very important difference between nonlinear elastic material and elastoplastic material. In the former, a unique stress is determined from the given magnitude of total strain, which is in fact total elastic strain. In the latter, however, it is possible to have infinitely different values of elastic strain for a given total strain by changing the value of plastic strain. Thus, **in elastoplasticity, it is impossible to determine the amount of stress for a given total strain**. Since the plastic strain is accumulated every time the material yields, it is necessary to follow every load increment to calculate the plastic strain. This property of elastoplasticity is called **path dependent or history dependent**. In order to determine stress, complete history of the load must be considered. The load history is taken into account by the accumulated plastic strain, ε_p . In addition, the yield stress of the material is determined by the magnitude of the plastic strain.

Although the objective is to separate the elastic strain increment from the plastic strain increment, it is assumed for the moment that the elastic strain increment is known. The stress increment, $\Delta\sigma$, can then be calculated using the elastic strain increment as

$$\Delta\sigma = E\Delta\varepsilon_e. \quad (4.3)$$

In addition to the elastic and tangent moduli, a *plastic modulus*, H , is defined, which is the slope of the strain-hardening portion of the stress-strain curve after removal of elastic strain components. Thus,

$$H = \frac{\Delta\sigma}{\Delta\epsilon_p}. \quad (4.4)$$

Even if the stress-strain curve in Fig. 4.2 is given in terms of E and E_t , it is common for the material properties to be given in terms of E and H . Since the stress-strain curve for the proportional loading is same for both isotropic and kinematic hardening, the plastic modulus is also same for both hardening models.

The three moduli (elastic, plastic, and tangent) are related to each other. The stress increment during the plastic phase can be written using any of the three moduli as

$$\Delta\sigma = E\Delta\epsilon_e = H\Delta\epsilon_p = E_t\Delta\epsilon. \quad (4.5)$$

The relation, $\Delta\sigma = H\Delta\epsilon_p$, may mislead that the plastic strain increases stress, but this is due to the stain-hardening effect. By substituting the relation in Eq. (4.5) into the strain increments in Eq. (4.1), we have

$$\frac{\Delta\sigma}{E_t} = \frac{\Delta\sigma}{E} + \frac{\Delta\sigma}{H} \Rightarrow \frac{1}{E_t} = \frac{1}{E} + \frac{1}{H}. \quad (4.6)$$

Thus, for the given E and E_t , the plastic modulus H can be determined as follows:

$$H = \frac{EE_t}{E - E_t}. \quad (4.7)$$

On the other hand, if E and H are specified, then E_t can be computed as follows:

$$E_t = \frac{EH}{E + H} = E \left(1 - \frac{E}{E + H} \right). \quad (4.8)$$

When the material is under the elastoplastic strain (along line $a-b$ in Fig. 4.1), the plastic portion can be determined for the given total strain increment, as

$$\begin{aligned} \Delta\epsilon = \Delta\epsilon_e + \Delta\epsilon_p &= \frac{\Delta\sigma}{E} + \Delta\epsilon_p = \frac{H\Delta\epsilon_p}{E} + \Delta\epsilon_p = \Delta\epsilon_p \left(\frac{H}{E} + 1 \right) \\ &\Rightarrow \Delta\epsilon_p = \frac{\Delta\epsilon}{1 + H/E} \end{aligned} \quad (4.9)$$

Thus, for the given amount of strain increment, the plastic strain increment can be calculated using the ratio between plastic and elastic moduli. As a special case, when no strain-hardening exists, i.e., stress remains at the constant yield stress during

plastic deformation, the strain increment becomes purely plastic. Note that the formula in Eq. (4.9) is not working when the material is initially in the elastic state.

Example 4.1 (Elastic and Plastic Strain) A force 12 kN is gradually applied at the end of an elastoplastic bar. When the yield stress of the material is 100 MPa, calculate the elastic and plastic strains. Use the following material properties: $E = 100 \text{ GPa}$ and $H = 10 \text{ GPa}$. The cross-sectional area of the bar is $A = 1.0 \times 10^{-4} \text{ m}^2$.

Solution In a one-dimensional bar, it is assumed that the force is uniformly distributed over the cross section. Since the total stress, $\sigma = F/A = 120 \text{ MPa}$, is larger than the yield stress, it can be concluded that the material is under plastic deformation. In addition, since the load is proportionally applied, there is no need to distinguish isotropic and kinematic hardening. It is convenient to divide the entire deformation into elastic and elastoplastic phases. The material is initially elastic until it reaches yield stress. Thus, until stress reaches yield stress $\sigma^{(1)} = \sigma_Y = 100 \text{ MPa}$, strain is purely elastic:

$$\Delta\epsilon_e^{(1)} = \frac{\sigma_Y}{E} = 0.001.$$

After yielding, the remaining stress increment, $\Delta\sigma = 20 \text{ MPa}$, is in the elastoplastic phase. The elastic and plastic strain increments can be calculated from

$$\begin{aligned}\Delta\epsilon_e^{(2)} &= \frac{\Delta\sigma^{(2)}}{E} = 0.0002, \\ \Delta\epsilon_p^{(2)} &= \frac{\Delta\sigma^{(2)}}{H} = 0.002.\end{aligned}$$

Thus, the total elastic and plastic strains become

$$\begin{aligned}\epsilon_e &= \Delta\epsilon_e^{(1)} + \Delta\epsilon_e^{(2)} = 0.0012 \\ \epsilon_p &= \Delta\epsilon_p^{(2)} = 0.002\end{aligned}$$

■

4.2.2 Finite Element Formulation for Elastoplasticity

In this section, a finite element formulation for an elastoplastic bar element is presented based on “the steps in the solution of nonlinear finite element analysis” in Sect. 2.3. Due to the small deformation assumption, only material nonlinearity will be considered.

In order to be more general, the incremental force method in Sect. 2.2.4 is considered, where the applied load is first divided by N load increments, which are denoted by $[t_1, t_2, \dots, t_N]$. Even if the analysis procedure is static, these *load*

increments are also called *time increments*. It is also assumed that the analysis procedure has been completed up to load increment t_n , and a new solution at load increment t_{n+1} is sought using the Newton–Raphson method. In the following explanation, the left superscript n or $n+1$ is used to denote the load increment, while the right superscript, k or $k+1$, is used for the iteration counter. Since all variables at t_n are already converged, iteration counter only appears for those variables at t_{n+1} . When a variable does not show a load increment, it should be considered a variable at t_{n+1} or a constant value. It is also assumed that iteration k has been finished and the current iteration is $k+1$. The iteration count will be omitted whenever possible unless it can cause confusions. For most cases, variables at t_{n+1} represent the variables at $(k+1)$ th iteration, i.e., the current iteration.

Referring to Fig. 4.3, it is assumed that the entire structure is modeled by one bar element, and the solution of the element is approximated by a vector, $\mathbf{d} = \{u_1, u_2\}^T$, of nodal displacements and its increment. In the view of load increment and iteration counter, it is possible to consider two different definitions of displacement increments, as

$$\begin{aligned}\Delta \mathbf{d}^k &= {}^{n+1}\mathbf{d}^k - {}^n\mathbf{d} \\ \delta \mathbf{d}^k &= {}^{n+1}\mathbf{d}^{k+1} - {}^{n+1}\mathbf{d}^k,\end{aligned}\quad (4.10)$$

where $\Delta \mathbf{d}$ is the increment from the last converged load increment to the previous iteration, while $\delta \mathbf{d}$ is the increment from the previous iteration. The former is used to calculate the stress increment, while the latter is the displacement increment calculated from the Newton–Raphson iteration. Therefore, $\delta \mathbf{d}$ is accumulated into $\Delta \mathbf{d}$ during the Newton–Raphson iteration, and $\Delta \mathbf{d}$ is set to 0 before starting a new load increment.

As discussed in Chap. 1, there is a suitable vector of interpolation functions, $\mathbf{N}(\mathbf{x}) = [N_1, N_2]$, for the bar element so that the displacement increment can be interpolated by

$$\Delta u(x) = [N_1 N_2] \left\{ \begin{array}{c} \Delta u_1 \\ \Delta u_2 \end{array} \right\} = \mathbf{N} \cdot \Delta \mathbf{d}. \quad (4.11)$$

The corresponding strain increment can also be calculated by differentiating the above displacement by

$$\Delta \boldsymbol{\varepsilon} = \frac{d}{dx}(\Delta u) = \left[-\frac{1}{L} \quad \frac{1}{L} \right] \left\{ \begin{array}{c} \Delta u_1 \\ \Delta u_2 \end{array} \right\} = \mathbf{B} \cdot \Delta \mathbf{d}, \quad (4.12)$$

where \mathbf{B} is the displacement–strain matrix—in the case of a bar element, it is a row vector. Again, the strain increment is from the previous converged load increment

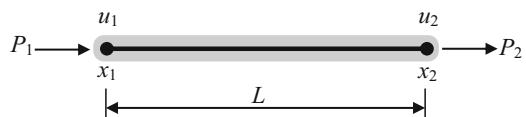


Fig. 4.3 One-dimensional elastoplastic bar element

to the previous iteration, which is known. Using the same interpolation scheme, the increment from the previous iteration can also be calculated by $\delta u(x) = \mathbf{N} \cdot \delta \mathbf{d}$ and $\delta \epsilon = \mathbf{B} \cdot \delta \mathbf{d}$.

If the bar element in Fig. 4.3 is in equilibrium, the nodal forces due to internal stresses must be equal and opposite in direction to the applied nodal forces. More specifically, the weak form of structural equilibrium can be written as

$$\bar{\mathbf{d}}^T \int_0^L \mathbf{B}^{Tn+1} \sigma^{k+1} A dx = \bar{\mathbf{d}}^T n^{+1} \mathbf{F}, \quad \forall \bar{\mathbf{d}} \in R^2, \quad (4.13)$$

where $\bar{\mathbf{d}} = [\bar{u}_1, \bar{u}_2]^T$ is the vector of virtual nodal displacements, A is the cross-sectional area, and $n^{+1} \mathbf{F} = [n^{+1} F_1, n^{+1} F_2]^T$ is the vector of applied forces, which is assumed to be given.

Since the relationship between stress and strain is nonlinear, the above variational equation is nonlinear in terms of strain or, equivalently, displacement. In order to solve the nonlinear equation, the Newton–Raphson method in Chap. 2 can be employed. Assuming that the applied load is prescribed or independent of displacement, only the left-hand side of Eq. (4.13) needs to be linearized. Since only material nonlinearity is considered, the stress is linearized using the first-order Taylor series expansion, as

$$n^{+1} \sigma^{k+1} \approx n^{+1} \sigma^k + \frac{\partial \sigma}{\partial \epsilon} \delta \epsilon = n^{+1} \sigma^k + D^{ep} \delta \epsilon, \quad (4.14)$$

where D^{ep} is the elastoplastic tangent modulus. From the stress–strain curve in Fig. 4.1, the slope, D^{ep} , can be determined by

$$D^{ep} = \begin{cases} E & \text{if elastic} \\ E_t & \text{if plastic} \end{cases}. \quad (4.15)$$

By substituting Eq. (4.14) into Eq. (4.13) and moving known terms to the right-hand side, the following linearized variational equation can be obtained:

$$\bar{\mathbf{d}}^T \left[\int_0^L \mathbf{B}^T D^{ep} \mathbf{B} A dx \right] \delta \mathbf{d} = \bar{\mathbf{d}}^T \mathbf{F} - \bar{\mathbf{d}}^T \int_0^L \mathbf{B}^{Tn+1} \sigma^k A dx, \quad (4.16)$$

where the terms in the bracket is called the *tangent stiffness matrix* for the bar element and can be written as

$$\mathbf{k}_T = \frac{AD^{ep}}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

and the right-hand side of Eq. (4.16) is called the *residual*, \mathbf{R} , which is defined as

$$^{n+1}\mathbf{R}^k = ^{n+1}\mathbf{F} - \int_0^L \mathbf{B}^{Tn+1} \sigma^k A \, dx = \begin{Bmatrix} ^{n+1}F_1 + ^{n+1}\sigma^k A \\ ^{n+1}F_2 - ^{n+1}\sigma^k A \end{Bmatrix}. \quad (4.17)$$

Note that if the residual in the above equation becomes 0, then it means that the original nonlinear Eq. (4.13) is satisfied.¹ Therefore, the Newton–Raphson iteration stops and moves to the next load increment. In order to calculate the residual, it is necessary to calculate stress, $^{n+1}\sigma^k$. The stress calculations are more involved because of nonlinearity in the stress–strain relation and dependency on the prior stress history. For an elastoplastic material, the stress is a function of previous state, plastic variables, and strain increment. Therefore, it can be written as

$$^{n+1}\sigma^k = f(^n\sigma, ^n\varepsilon_p, \Delta\varepsilon^k, \dots). \quad (4.18)$$

In stress calculation, the following situations must be considered: whether the material is in the elastic or strain-hardening portions of the curve, loading or unloading, and changes in the stress at which yielding takes place based on the hardening rule employed. This process is called **state determination** and will be explained in detail in the following section.

If the residual does not vanish, another iteration is required based on Eq. (4.16). Since Eq. (4.16) must satisfy for arbitrary $\bar{\mathbf{d}} \in R^2$, it is equivalent to solving the following incremental matrix equation:

$$\mathbf{k}_T \cdot \delta\mathbf{d}^k = ^{n+1}\mathbf{R}^k. \quad (4.19)$$

Therefore, the Newton–Raphson iteration solves for the incremental displacement from the previous iteration using the residual at the last iteration. When the system is composed of many bar elements, individual element equations are assembled in the usual manner as in Chap. 1 and solved for displacement increments, after applying boundary conditions. After obtaining the nodal displacement increments by solving Eq. (4.19), the displacement increment, $\Delta\mathbf{d}$, is updated by $\Delta\mathbf{d}^{k+1} = \Delta\mathbf{d}^k + \delta\mathbf{d}^k$, and the process repeated until the residual vanishes.

4.2.3 Determination of Stress State

As discussed in the previous section, the relationship between stress and strain is nonlinear when plastic deformation occurs. In this section, the procedure of determining stress will be discussed when strain and its increment are available. For the

¹ Although the iteration counter is different, it does not matter when the residual vanishes.

purpose of explanation, isotropic and kinematic hardenings are treated separately, but later a combined procedure will be presented.

4.2.3.1 Isotropic Hardening Model

During the Newton–Raphson iterative method, it is assumed that the following variables are given: the strain increment ($\Delta\epsilon$) from the last load increment to the previous iteration, plastic strain ($^n\epsilon_p$), and stress ($^n\sigma$) at load increment n . The basic steps in computing stress are as follows.

- Compute the current yield stress.** This depends on the accumulated plastic strain, $^n\epsilon_p$, and the plastic modulus, H .

$$^n\sigma_Y = {}^0\sigma_Y + H^n\epsilon_p, \quad (4.20)$$

where ${}^0\sigma_Y$ is the initial yield stress. Due to strain-hardening, the yield stress increases according to plastic strain. Note that the yield stress is the same for both tension and compression and it increases along with the plastic strain.

- Elastic predictor.** Assume an elastic behavior during the strain increment and calculate stress increment and trial stress.

$$\Delta^{tr}\sigma = E\Delta\epsilon, \quad (4.21)$$

$$^{tr}\sigma = {}^n\sigma + \Delta^{tr}\sigma. \quad (4.22)$$

- Check yield status.** Check if the trial stress satisfies the yield condition; that is, the stress must be lower than the yield stress. For that purpose, the following **trial yield function** is defined:

$$^{tr}f = |^{tr}\sigma| - {}^n\sigma_Y. \quad (4.23)$$

If $^{tr}f \leq 0$, then the material stays elastic. As illustrated in Fig. 4.4, the material is either on the initial loading curve below the yield stress or on the unloading/reloading curve. In either case, the stress based on the elastic assumption is correct. Set $^{n+1}\sigma = {}^{tr}\sigma$ and move on to the next load increment. In this case, the

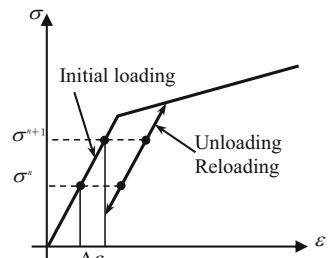
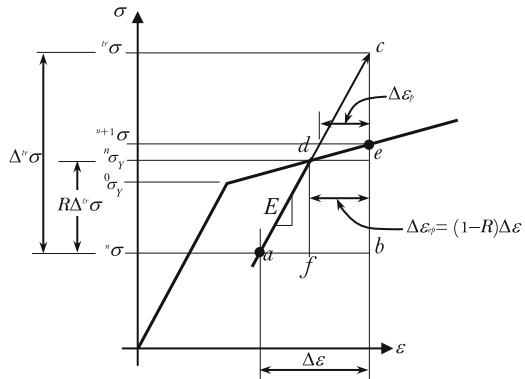


Fig. 4.4 Stress state under elastic region

Fig. 4.5 Elastic to plastic transition



strain increment is purely elastic, and the plastic strain does not change, i.e., ${}^{n+1}\varepsilon_e = {}^n\varepsilon_e + \Delta\varepsilon$ and ${}^{n+1}\varepsilon_p = {}^n\varepsilon_p$.

- 4. Plastic corrector.** If ${}^n\varepsilon > 0$, then the material yields during this increment. As illustrated in Fig. 4.5, at load increment t_n , it is assumed that the material is at point a (elastic). This elastic state is reached by unloading from a plastic state (e.g., point d). Thus, the current yield stress is ${}^n\sigma_Y$. At load increment t_{n+1} , a strain increment $\Delta\varepsilon$ is given, and the corresponding updated state of stress is sought. If the material is continuously elastic, the updated stress will be at point c (trial stress ${}^n\sigma$). However, it is impossible for the state of stress to be above the stress-strain curve in Fig. 4.5. Thus, a transition from elastic to plastic state occurs in this step and the material moves up to point e (plastic). Considering that the plastic strain increment does not contribute to the stress increment, the trial stress is updated by subtracting the portion of plastic strain increment, as

$${}^{n+1}\sigma = {}^n\sigma - \text{sgn}({}^n\sigma) E \Delta\varepsilon_p, \quad (4.24)$$

where $\text{sgn}()$ is a sign function, which takes “+1” when its argument is positive and “−1” when negative. This function is added because the material can also yield in compression. Since Eq. (4.24) corrects the trial stress by plastic strain increment, this process is often called a **plastic corrector**. In addition, since it makes the trial stress back to the yield stress, it is also called **return-mapping**. As can be seen in Fig. 4.5, an **algorithmic challenge** is that while the trial stress returns, the yield stress also increases (from point d to e). In Sect. 4.3, return-mapping for multidimensional stress will be discussed.

- 5. Plastic consistency condition.** In the plastic correction formula in Eq. (4.24), the plastic strain increment is still unknown. In order to calculate it, the **plastic consistency condition** is used, where the corrected stress must be on the yield surface during the loading process. This condition can be written as

$${}^{n+1}f = |{}^{n+1}\sigma| - {}^{n+1}\sigma_Y = 0. \quad (4.25)$$

Therefore, the goal is to find the plastic strain increment that satisfies the above condition. Note that not only the plastic strain increment reduces the corrected stress, ${}^{n+1}\sigma$, but also it increases the yield stress, ${}^{n+1}\sigma_Y$. In order to calculate the plastic strain increment, Eq. (4.25) can be expanded as

$$\begin{aligned} |{}^{\text{tr}}\sigma - \text{sgn}({}^{\text{tr}}\sigma)E\Delta\epsilon_p| - ({}^n\sigma_Y + H\Delta\epsilon_p) &= 0 \\ \Rightarrow |{}^{\text{tr}}\sigma| - {}^n\sigma_Y - (E + H)\Delta\epsilon_p &= 0. \end{aligned}$$

Note that the above formula works for both positive and negative trial stress. Therefore, the plastic strain increment can be obtained as

$$\Delta\epsilon_p = \frac{|{}^{\text{tr}}\sigma| - {}^n\sigma_Y}{E + H} = \frac{{}^{\text{tr}}f}{E + H}. \quad (4.26)$$

Since ${}^{\text{tr}}f > 0$, the plastic strain increment is always positive. With this plastic strain increment, the stress in Eq. (4.24) can be updated, which concludes the state determination. The plastic strain increment reduces the trial stress and increases the yield stress so that the plastic consistency condition satisfies.

Since a linear hardening model with a constant plastic modulus, H , is used, the plastic consistency condition in Eq. (4.25) can explicitly be solved in terms of the plastic strain increment. If a nonlinear hardening model is used, however, then the plastic consistency condition must be solved iteratively using a method like the Newton–Raphson method.

In order to express the plastic strain increment in terms of the total strain increment, a purely elastic fraction of strain increment, denoted by R in Fig. 4.5, is calculated from similar triangles abc and dec as

$$\frac{|\Delta\epsilon|}{|\Delta{}^{\text{tr}}\sigma|} = \frac{(1 - R)|\Delta\epsilon|}{|{}^{\text{tr}}\sigma| - {}^n\sigma_Y} \Rightarrow R = 1 - \frac{{}^{\text{tr}}f}{|\Delta{}^{\text{tr}}\sigma|}. \quad (4.27)$$

It is mentioned that R is the interpolating factor between the elastic and tangent moduli. When R is equal to one, the material is purely elastic. When R is equal to 0, the initial material status is plastic and the strain increment has both elastic and plastic portions. Using the relation of ${}^{\text{tr}}f = (1 - R)E|\Delta\epsilon|$, the plastic strain increment in Eq. (4.26) can be written as

$$\Delta\epsilon_p = \frac{(1 - R)E}{E + H} |\Delta\epsilon|. \quad (4.28)$$

Note that when $R = 0$, the relation becomes identical to that in Eq. (4.9). In a similar way, the updated stress in Eq. (4.24) can be expressed as

$${}^{n+1}\sigma = {}^{\text{tr}}\sigma - \text{sgn}({}^{\text{tr}}\sigma) \frac{(1 - R)E^2}{E + H} |\Delta\epsilon|. \quad (4.29)$$

The first term on the right-hand side is from elastic trial, and the second term is from plastic correction. Therefore, the algorithm for determining the state of stress in elastoplastic is divided into (a) elastic trial and (b) plastic correction.

6. **Algorithmic tangent stiffness.** The elastoplastic tangent modulus D^{ep} in Eq. (4.15) represents the relation between the stress increment and strain increment (hypoelasticity), which is nothing but the slope of the stress–strain curve in Fig. 4.2. It would be beneficial to compare the tangent modulus of the state determination algorithm in Eq. (4.24) with the slope of the stress–strain curve. In the literature, the former is referred to the *algorithmic tangent modulus*, while the latter the *continuum tangent modulus*. When the material is elastic or in the elastic state of unloading/reloading process, the plastic strain increment becomes 0 and the trial stress becomes the updated stress. Therefore, the algorithmic tangent modulus becomes identical to the elastic modulus. When the material is in the plastic state, the algorithmic tangent modulus can be obtained by differentiating the updated stress increment with respect to the strain increment, as

$$D^{\text{alg}} = \frac{\partial \Delta\sigma}{\partial \Delta\varepsilon} = \frac{\partial^{tr}\sigma}{\partial \Delta\varepsilon} - \text{sgn}({}^{tr}\sigma) E \frac{\partial \Delta\varepsilon_p}{\partial \Delta\varepsilon}. \quad (4.30)$$

Since the trial stress is elastic, it is straightforward that the first term on the right-hand side is nothing but the elastic modulus. The derivative of the plastic strain increment can be obtained by differentiating Eq. (4.26) with respect to the strain increment

$$\frac{\partial \Delta\varepsilon_p}{\partial \Delta\varepsilon} = \frac{1}{E+H} \frac{\partial^{tr}f}{\partial \Delta\varepsilon} = \text{sgn}({}^{tr}\sigma) \frac{E}{E+H}.$$

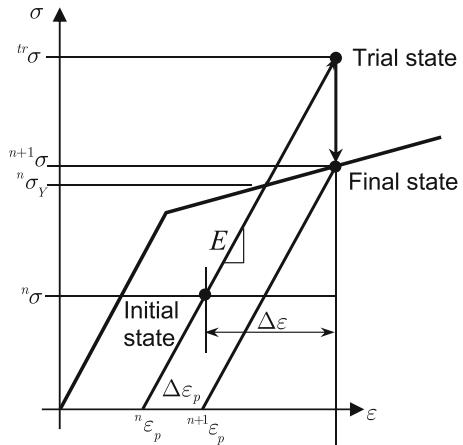
After substituting into Eq. (4.30), the algorithmic tangent modulus can be obtained as

$$D^{\text{alg}} = \begin{cases} E & \text{if elastic} \\ E_t & \text{if plastic} \end{cases}. \quad (4.31)$$

It is interesting to note that the continuum tangent modulus, D^{ep} , in Eq. (4.15) is identical to the algorithmic counterpart in Eq. (4.31); that is, the state determination algorithm is consistent with the stress–strain curve. In multidimensional elastoplasticity in the next section, however, it will be shown that the two tangent moduli are different and show a quite different convergence behavior during Newton–Raphson iterations. In addition, the two tangent moduli will be different when a nonlinear hardening model is employed.

Example 4.2 (Elastoplastic Bar (Isotropic Hardening)) An elastoplastic bar is under variable load history. At load step t_n , the stress and plastic strain are ${}^n\sigma = 150$ MPa and ${}^n\varepsilon_p = 1.0 \times 10^{-4}$, respectively. When strain increment is $\Delta\varepsilon = 0.002$, calculate stress and plastic strain. Assume isotropic hardening with $E = 200$ GPa, $H = 25$ GPa, and ${}^0\sigma_Y = 250$ MPa.

Fig. 4.6 Elastoplastic bar
(not scaled)



Solution At a given plastic strain ${}^n\varepsilon_p = 1.0 \times 10^{-4}$, the yield stress is

$${}^n\sigma_Y = {}^0\sigma_Y + H{}^n\varepsilon_p = 252.5 \text{ MPa.}$$

Since ${}^n\sigma < {}^n\sigma_Y$, the material is in the elastic state at load step t_n (Refer to the initial state in Fig. 4.6). For given strain increment, the trial stress can be obtained as

$$\Delta{}^{\text{tr}}\sigma = E\Delta\varepsilon = 400 \text{ MPa, } {}^{\text{tr}}\sigma = {}^n\sigma + \Delta{}^{\text{tr}}\sigma = 550 \text{ MPa.}$$

Since ${}^{\text{tr}}f = |{}^{\text{tr}}\sigma| - {}^n\sigma_Y = 297.5 \text{ MPa} > 0$, the material yields during the current load increment. The plastic strain increment can be calculated from the plastic consistency condition as

$$\Delta\varepsilon_p = \frac{{}^{\text{tr}}f}{E + H} = 1.322 \times 10^{-3}.$$

Therefore, stress and plastic strain are updated as

$$\begin{aligned} {}^{n+1}\sigma &= {}^{\text{tr}}\sigma - \text{sgn}({}^{\text{tr}}\sigma)E\Delta\varepsilon_p = 285.6 \text{ MPa,} \\ {}^{n+1}\varepsilon_p &= {}^n\varepsilon_p + \Delta\varepsilon_p = 1.422 \times 10^{-3}. \end{aligned}$$

Note that the updated stress is identical to the new yield stress ${}^{n+1}\sigma_Y$. ■

4.2.3.2 Kinematic Hardening Model

The main difference between the kinematic and isotropic hardening models is the evolution of yield surface. The elastic range (twice of yield stress) increases in isotropic hardening, while the range remains constant in kinematic hardening.

Instead, the center of elastic range moves parallel to the hardening curve as plastic strain increases. In order to model this effect, the following **shifted stress** is defined:

$$\eta = \sigma - \alpha, \quad (4.32)$$

where α is called the **back stress**, which represents the center of elastic range. In kinematic hardening, the shifted stress is used instead of σ in determining the material status. Therefore, the back stress is considered as a plastic variable and must be stored and updated at each load increment. The load history information is stored in the back stress.

In the incremental formulation, it is assumed that the following variables are given: the **strain increment** ($\Delta\epsilon$), **stress** (${}^n\sigma$), and **back stress** (${}^n\alpha$) at load increment t_n . The basic steps in computing stress are as follows.

- Elastic predictor.** Assume an elastic behavior during the strain increment and calculate stress increment and trial stress.

$$\Delta^{tr}\sigma = E\Delta\epsilon. \quad (4.33)$$

$${}^{tr}\sigma = {}^n\sigma + \Delta^{tr}\sigma. \quad (4.34)$$

During the elastic predictor, the plastic variable remains constant as

$${}^{tr}\alpha = {}^n\alpha. \quad (4.35)$$

Accordingly, the elastic predictor of the shifted stress is given as

$${}^{tr}\eta = {}^{tr}\sigma - {}^{tr}\alpha. \quad (4.36)$$

- Check yield status.** Check if the trial stress satisfies the yield condition; that is, the stress must be lower than the yield stress. For that purpose, the following trial yield function is defined:

$${}^{tr}f = |{}^{tr}\eta| - {}^0\sigma_Y. \quad (4.37)$$

Note that in kinematic hardening, the initial yield stress remains constant; that is, the elastic range does not increase. If ${}^{tr}f \leq 0$, then the material stays elastic. Set ${}^{n+1}\sigma = {}^{tr}\sigma$ and move on to the next load increment. In this case, the incremental strain is purely elastic, and the back stress does not change, i.e., ${}^{n+1}\alpha = {}^n\alpha$, and all strain increment is elastic, i.e., ${}^{n+1}\epsilon_e = {}^n\epsilon_e + \Delta\epsilon$.

- Plastic corrector.** If ${}^{tr}f > 0$, then the material yields during this increment. When the material experiences plastic deformation, both the trial stress and back stress are updated as

$${}^{n+1}\sigma = {}^{tr}\sigma - \text{sgn}({}^{tr}\eta)E\Delta\epsilon_p, \quad (4.38)$$

$${}^{n+1}\alpha = {}^n\alpha + \text{sgn}({}^{tr}\eta)H\Delta\epsilon_p. \quad (4.39)$$

Note that the trial stress is reduced by the plastic strain increment proportional to the elastic modulus, while the back stress is increased proportional to the plastic modulus. The stress update formula is almost identical to that of isotropic hardening, except that the shifted stress is used in sgn function. The back stress is updated in a similar way of yield stress in isotropic hardening in Eq. (4.20) because the stress-strain curve in a proportional loading is identical for both hardening models. Therefore, the same plastic modulus is used. Even if the plastic strain increment is always positive, the back stress can be negative depending on the stress history; that is, the back stress increases if the yielding occurs in tension, while decreases in compression.

4. **Plastic consistency condition.** In the plastic correction formulas in Eqs. (4.38) and (4.39), the unknown plastic strain increment is calculated from the plastic consistency condition as

$${}^{n+1}f = |{}^{n+1}\eta| - {}^0\sigma_Y = 0. \quad (4.40)$$

The above consistency condition can be expanded in terms of plastic strain increment as

$$\begin{aligned} & |{}^{tr}\sigma - \text{sgn}({}^{tr}\eta)E\Delta\epsilon_p - {}^{tr}\alpha - \text{sgn}({}^{tr}\eta)H\Delta\epsilon_p| - {}^0\sigma_Y = 0 \\ & \Rightarrow |{}^{tr}\sigma - {}^{tr}\alpha| - {}^0\sigma_Y - (E + H)\Delta\epsilon_p = 0. \end{aligned}$$

Note that the above formula works for both positive and negative trial stress. Therefore, the plastic strain increment can be obtained as

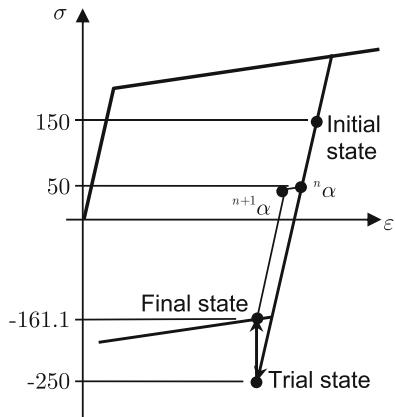
$$\Delta\epsilon_p = \frac{{}^{tr}f}{E + H}. \quad (4.41)$$

Note that the formula for plastic strain increment is identical to that of isotropic hardening in Eq. (4.26). With this plastic strain increment, the stress and back stress are updated according to Eqs. (4.38) and (4.39), which concludes the state determination.

In the case of kinematic hardening, the plastic strain increment does not have to be stored, as the load history information is stored in the back stress. Since the plastic strain increment and stress update formula are identical for both isotropic and kinematic hardening models, the same algorithmic tangent stiffness can be used for both models.

Example 4.3 (Elastoplastic Bar (Kinematic Hardening)) An elastoplastic bar is under variable load history. At load step t_n , the stress and back stress are ${}^n\sigma = 150$ MPa and ${}^n\alpha = 50$ MPa, respectively. When strain increment is $\Delta\epsilon = -0.002$, calculate stress and back stress. Assume kinematic hardening with $E = 200$ GPa, $H = 25$ GPa, and ${}^0\sigma_Y = 200$ MPa.

Fig. 4.7 Elastoplastic bar in kinematic hardening (not scaled)



Solution Since the shifted stress ${}^n\eta = {}^n\sigma - {}^n\alpha = 100 \text{ MPa} < {}^0\sigma_Y$, the material is in the elastic state at load step t_n (Refer to Fig. 4.7). For given strain increment, the trial stresses can be obtained as

$$\begin{aligned}\Delta {}^{tr}\sigma &= E\Delta\epsilon = -400 \text{ MPa}, & {}^{tr}\sigma &= {}^n\sigma + \Delta {}^{tr}\sigma = -250 \text{ MPa} \\ {}^{tr}\alpha &= {}^n\alpha = 50 \text{ MPa}, & {}^{tr}\eta &= {}^{tr}\sigma - {}^{tr}\alpha = -300 \text{ MPa}\end{aligned}$$

Since ${}^{tr}f = |{}^{tr}\eta| - {}^0\sigma_Y = 100 \text{ MPa} > 0$, the material yields in compression during the current load increment. The plastic strain increment can be calculated from the plastic consistency condition as

$$\Delta\epsilon_p = \frac{{}^{tr}f}{E + H} = 0.444 \times 10^{-3}.$$

Therefore, stress and back stress are updated as

$$\begin{aligned}{}^{n+1}\sigma &= {}^{tr}\sigma - \text{sgn}({}^{tr}\eta)E\Delta\epsilon_p = -161.1 \text{ MPa}, \\ {}^{n+1}\alpha &= {}^{tr}\alpha + \text{sgn}({}^{tr}\eta)H\Delta\epsilon_p = 38.9 \text{ MPa}.\end{aligned}$$

Note that the updated back stress is reduced due to the yielding in compression. ■

4.2.3.3 Combined Isotropic/Kinematic Hardening Model

Many practical materials show a combined effect of isotropic and kinematic hardenings, especially for polycrystalline metals. In such a case, the yield stress initially increases due to plastic hardening, but it decreases when the direction of strain changes. This phenomenon is related to the dislocation structure in the cold worked metal. As deformation occurs, the dislocation accumulates at barriers and

produces dislocation pileups and tangles. This effect is often referred to as *Bauschinger effect*.

In order to model the combined effect of kinematic and isotropic hardening, a new parameter, β , is introduced. The parameter varies between 0 and 1. From the fact that isotropic hardening changes the yield stress and the kinematic hardening changes the back stress, the parameter interpolates between the two hardening models as

$$^{n+1}\sigma_Y = ^n\sigma_Y + (1 - \beta)H\Delta\epsilon_p, \quad (4.42)$$

$$^{n+1}\alpha = ^n\alpha + \text{sgn}({}^T\eta)\beta H\Delta\epsilon_p. \quad (4.43)$$

When $\beta=0$, it becomes isotropic hardening, while when $\beta=1$, it becomes kinematic hardening. In the combined hardening model, both the plastic strain and back stress are plastic variables and need to be updated and stored at each load increment. It would be a good practice to show that the plastic strain increment that is obtained from the plastic consistency condition is identical to the case of isotropic hardening in Eq. (4.26) and of kinematic hardening in Eq. (4.41).

The above algorithms can easily be implemented in computer programs. Below is a MATLAB program, **combHard1D**, that calculates stress, back stress, and plastic strain for a given strain increment. The program also requires stress, back stress, and plastic strain from the previous load step, as well as material properties.

```
PROGRAM combHard1D
%
% 1D Linear combined isotropic/kinamtic hardening model
%
function [stress, alpha, ep]=combHard1D(mp, deps, stressN, alphaN, epN)
% Inputs:
% mp = [E, beta, H, Y0];
% deps = strain increment
% stressN = stress at load step N
% alphaN = back stress at load step N
% epN = plastic strain at load step N
%
E=mp(1); beta=mp(2); H=mp(3); Y0=mp(4); %material properties
ftol = Y0*1E-6; %tolerance for yield
stresstr = stressN + E*deps; %trial stress
etatr = stresstr - alphaN; %trial shifted stress
fyld = abs(etatr) - (Y0+(1-beta)*H*epN); %trial yield function
if fyld < ftol %yield test
    stress = stresstr; alpha = alphaN; ep = epN; %trial states are final
    return;
else
    dep = fyld/(E+H); %plastic strain increment
end
stress = stresstr - sign(etatr)*E*dep; %updated stress
alpha = alphaN + sign(etatr)*beta*H*dep; %updated back stress
ep = epN + dep; %updated plastic strain
return;
```

Fig. 4.8 Two-bar assembly subjected to axial load



Example 4.4 (Two Bars in Parallel) An assembly of two bars with different material and section properties is subjected to an axial load as shown in Fig. 4.8. The initial length of the two bars is 100. Determine axial displacement, stresses, and strains when $P = 15$ is applied at the tip. Assume that the elongations of both bars are the same. Assume the following properties for the bars:

- Bar 1: $A = 0.75$, $E = 10,000$, $E_t = 1,000$, $^0\sigma_Y = 5$, kinematic hardening
- Bar 2: $A = 1.25$, $E = 5,000$, $E_t = 500$, $^0\sigma_Y = 7.5$, isotropic hardening

Solution The two bars can be modeled using two nodes and two elements. Both elements are connected to the same nodes. Since the node on the wall is fixed, its displacement is 0 and can be ignored. Then, the finite element matrix equation becomes a scalar equation with the node at the tip as a single degree of freedom. Since a single load increment will be used, the index for load increment will not be used.

Iteration 1: Initially, both elements are in the elastic state. Thus, the incremental finite element equation becomes

$$\left(\frac{E_1 A_1}{L_1} + \frac{E_2 A_2}{L_2} \right) \Delta u^1 = P - (\sigma_1^0 A_1 + \sigma_2^0 A_2).$$

Note that at the first iteration, the stresses in both elements are 0. Thus, the incremental displacement and incremental strain can be obtained from

$$\Delta u^1 = 0.1091, \quad \Delta \epsilon^1 = 0.001091.$$

Note that the incremental strain is identical for both elements. Now, it needs to calculate stress and plastic variables for both elements.

Element 1: Kinematic hardening

$$\begin{aligned} \Delta^{tr} \sigma_1^1 &= E_1 \Delta \epsilon^1 = 10.91, \quad {}^{tr} \sigma_1^1 = \sigma_1^0 + \Delta^{tr} \sigma_1^1 = 10.91, \\ {}^{tr} \alpha_1^1 &= \alpha_1^0 = 0, \quad {}^{tr} \eta_1^1 = {}^{tr} \sigma_1^1 - {}^{tr} \alpha_1^1 = 10.91, \\ {}^{tr} f_1 &= |{}^{tr} \eta_1^1| - \sigma_{Y1}^0 = 5.91. \end{aligned}$$

Since ${}^tr f_1 > 0$, the material is in the yield state. The plastic strain increment and stress and back stress are updated as

$$\Delta \varepsilon_{p1}^1 = \frac{{}^tr f_1}{E_1 + H_1} = 5.3182 \times 10^{-4},$$

$$\sigma_1^1 = {}^tr \sigma_1^1 - \text{sgn}({}^tr \eta_1^1) E_1 \Delta \varepsilon_{p1}^1 = 5.5909,$$

$$\alpha_1^1 = {}^tr \alpha_1^1 + \text{sgn}({}^tr \eta_1^1) H_1 \Delta \varepsilon_{p1}^1 = 0.5909,$$

$$\varepsilon_{p1}^1 = \varepsilon_{p1}^0 + \Delta \varepsilon_{p1}^1 = 5.3182 \times 10^{-4}.$$

Element 2: Isotropic hardening

$$\begin{aligned} {}^tr \sigma_2^1 &= E_2 \Delta \varepsilon^1 = 5.4545, & {}^tr \sigma_2^1 &= \sigma_2^0 + \Delta {}^tr \sigma_2^1 = 5.4545, \\ {}^tr f_2^1 &= |{}^tr \sigma_2^1| - \sigma_{Y2}^0 = -2.0455. \end{aligned}$$

Since ${}^tr f_2 < 0$, the material remains elastic, and the trial state is the final state:

$$\sigma_2^1 = 5.4545, \quad \varepsilon_{p2}^1 = 0.$$

Residual check: Residual = $P - (\sigma_1^1 A_1 + \sigma_2^1 A_2) = 3.9886$.

Since the residual is not equal to 0, it is not yet converged. Move to the next iteration.

Iteration 2: Since Element 1 changes from elastic to plastic, the elastoplastic tangent modulus $E_{ep1} = E_{t1}$ can be used in calculating displacement increment, while Element 2 uses the elastic modulus. Thus, the incremental equation becomes

$$\left(\frac{E_{t1} A_1}{L_1} + \frac{E_2 A_2}{L_2} \right) \Delta u^2 = P - (\sigma_1^1 A_1 + \sigma_2^1 A_2).$$

The displacement increment and the strain increment can be obtained from

$$\Delta u^2 = 0.0570, \quad \Delta \varepsilon^2 = 5.6981 \times 10^{-4}, \quad u^2 = u^1 + \Delta u^2 = 0.1661.$$

Element 1:

$$\begin{aligned} {}^tr \sigma_1^2 &= E_1 \Delta \varepsilon^2 = 5.6981, & {}^tr \sigma_1^2 &= \sigma_1^1 + \Delta {}^tr \sigma_1^2 = 11.2890, \\ {}^tr \alpha_1^2 &= \alpha_1^1 = 0.5909, & {}^tr \eta_1^2 &= {}^tr \sigma_1^2 - {}^tr \alpha_1^2 = 10.6981, \\ {}^tr f_1^2 &= |{}^tr \eta_1^2| - \sigma_{Y1}^0 = 5.6981. \end{aligned}$$

Since ${}^T f_1 > 0$, the material is continuously yielding. Then, the plastic strain increment, stress, and back stress can be updated by

$$\Delta \varepsilon_{p1}^2 = \frac{{}^T f_1^2}{E_1 + H_1} = 5.1280 \times 10^{-4},$$

$$\sigma_1^2 = {}^T \sigma_1^2 - \text{sgn}({}^T \eta_1^2) E_1 \Delta \varepsilon_{p1}^2 = 6.1607,$$

$$\alpha_1^2 = {}^T \alpha_1^2 + \text{sgn}({}^T \eta_1^2) H_1 \Delta \varepsilon_{p1}^2 = 1.1607,$$

$$\varepsilon_{p1}^2 = \varepsilon_{p1}^1 + \Delta \varepsilon_{p1}^2 = 1.0446 \times 10^{-3}.$$

Element 2:

$$\begin{aligned} \Delta {}^T \sigma_2^2 &= E_2 \Delta \varepsilon^2 = 2.8490, & {}^T \sigma_2^2 &= \sigma_2^1 + \Delta {}^T \sigma_2^2 = 8.3036, \\ {}^T f_2^2 &= |{}^T \sigma_2^2| - \sigma_{Y2}^0 = 0.8036. \end{aligned}$$

Since ${}^T f_2 > 0$, the material is in the yield state. The plastic strain increment and stress can be updated as

$$\Delta \varepsilon_{p2}^2 = \frac{{}^T f_2^2}{E_2 + H_2} = 1.446 \times 10^{-4},$$

$$\sigma_2^2 = {}^T \sigma_2^2 + \text{sgn}({}^T \eta_2^2) E_2 \Delta \varepsilon_{p2}^2 = 7.5804,$$

$$\varepsilon_{p2}^2 = \varepsilon_{p2}^1 + \Delta \varepsilon_{p2}^2 = 1.4464 \times 10^{-4},$$

$$\sigma_{Y2}^2 = \sigma_{Y2}^0 + H_2 \Delta \varepsilon_{p2}^2 = 7.5804.$$

Residual check: Residual = $P - (\sigma_1^2 A_1 + \sigma_2^2 A_2)$ = 0.9040.

Since the residual is not equal to 0, it is not yet converged. Move to the next iteration.

Iteration 3: Since both elements are in the plastic state, the tangent moduli are used in the increment equation, as

$$\left(\frac{E_{t1} A_1}{L_1} + \frac{E_{t2} A_2}{L_2} \right) \Delta u^3 = P - (\sigma_1^2 A_1 + \sigma_2^2 A_2).$$

The displacement increment and the strain increment can be obtained from

$$\Delta u^3 = 0.0657, \quad \Delta \varepsilon^3 = 6.5747 \times 10^{-4}, \quad u^3 = u^2 + \Delta u^3 = 0.2318.$$

Element 1:

$$\begin{aligned}\Delta^{tr}\sigma_1^3 &= E_1\Delta\epsilon^3 = 6.5747, \quad {}^{tr}\sigma_1^3 = \sigma_1^2 + \Delta^{tr}\sigma_1^3 = 12.7354, \\ {}^{tr}\alpha_1^3 &= \alpha_1^2 = 1.1607, \quad {}^{tr}\eta_1^3 = {}^{tr}\sigma_1^3 - {}^{tr}\alpha_1^3 = 11.5747, \\ {}^{tr}f_1^3 &= |{}^{tr}\eta_1^3| - \sigma_{Y1}^0 = 6.5747.\end{aligned}$$

Since ${}^{tr}f_1 > 0$, the material is continuously yielding. The plastic strain increment, stress, and back stress can be updated by

$$\begin{aligned}\Delta\epsilon_{p1}^3 &= \frac{{}^{tr}f_1^3}{E_1 + H_1} = 5.9180 \times 10^{-4}, \\ \sigma_1^3 &= {}^{tr}\sigma_1^3 - \text{sgn}({}^{tr}\eta_1^3)E_1\Delta\epsilon_{p1}^3 = 6.8182, \\ \alpha_1^3 &= {}^{tr}\alpha_1^3 + \text{sgn}({}^{tr}\eta_1^3)H_1\Delta\epsilon_{p1}^3 = 1.8182, \\ \epsilon_{p1}^3 &= \epsilon_{p1}^2 + \Delta\epsilon_{p1}^3 = 1.6364 \times 10^{-3}.\end{aligned}$$

Element 2:

$$\begin{aligned}\Delta^{tr}\sigma_2^3 &= E_2\Delta\epsilon^3 = 3.2873, \quad {}^{tr}\sigma_2^3 = \sigma_2^2 + \Delta^{tr}\sigma_2^3 = 10.8677, \\ {}^{tr}f_2^3 &= |{}^{tr}\sigma_2^3| - \sigma_{Y2}^2 = 3.2873.\end{aligned}$$

Since ${}^{tr}f_2 > 0$, the material is continuously yielding. The plastic strain and stress can be updated by

$$\begin{aligned}\Delta\epsilon_{p2}^3 &= \frac{{}^{tr}f_2^3}{E_2 + H_2} = 5.9180 \times 10^{-4}, \\ \sigma_2^3 &= {}^{tr}\sigma_2^3 + \text{sgn}({}^{tr}\eta_2^3)E_2\Delta\epsilon_{p2}^3 = 7.9091, \\ \epsilon_{p2}^3 &= \epsilon_{p2}^2 + \Delta\epsilon_{p2}^3 = 7.3636 \times 10^{-4}.\end{aligned}$$

Residual check: Residual = $P - (\sigma_1^3 A_1 + \sigma_2^3 A_2) = 0.0$.

Since the residual is equal to 0, the iteration converges. ■

Example 4.5 (Two Bars in Parallel) Solve the two bar problem in Example 4.4 using MATLAB programs.

Solution Below is the list of MATLAB programs that calculate the two bars in parallel in Example 4.4. Since only the node at the right end is allowed to move, a single nonlinear finite element equation is solved using the Newton–Raphson method. The program converges in the third iteration with 0 residual. Table 4.1 shows the history of convergence iteration.

Table 4.1 Convergence history of two elastoplastic bars using the Newton–Raphson method

Iteration	u	σ_1	σ_2	ϵ_{p1}	ϵ_{p2}	Residual
0	0.0000	0.000	0.000	0.000000	0.000000	1.50E + 1
1	0.1091	5.591	5.455	0.000532	0.000000	3.99E + 0
2	0.1661	6.161	7.580	0.001045	0.000145	9.04E - 1
3	0.2318	6.818	7.909	0.001636	0.000736	0.00E + 0

```

%
% Example 4.5 Two elastoplastic bars in parallel
%
E1=10000; Et1=1000; sYield1=5;
E2=5000; Et2=500; sYield2=7.5;
mp1 = [E1, 1, E1*Et1/(E1-Et1), sYield1];
mp2 = [E2, 0, E2*Et2/(E2-Et2), sYield2];
nS1 = 0; nA1 = 0; nep1 = 0;
nS2 = 0; nA2 = 0; nep2 = 0;
A1 = 0.75; L1 = 100;
A2 = 1.25; L2 = 100;
tol = 1.0E-5; u = 0; P = 15; iter = 0;
Res = P - nS1*A1 - nS2*A2;
Dep1 = E1; Dep2 = E2;
conv = Res^2/(1+P^2);
fprintf('\niter      u      S1      S2      A1      A2');
fprintf('      ep1      ep2      Residual');
fprintf('\n %3d %7.4f %7.3f %7.3f %7.3f %8.6f %8.6f %10.3e',...
       iter,u,nS1,nS2,nA1,nA2,nep1,nep2,Res);
while conv > tol && iter < 20
    delu = Res / (Dep1*A1/L1 + Dep2*A2/L2);
    u = u + delu;
    delE = delu / L1;
    [Snew1, Anew1, epnew1]=combHard1D(mp1,delE,nS1,nA1,nep1);
    [Snew2, Anew2, epnew2]=combHard1D(mp2,delE,nS2,nA2,nep2);
    Res = P - Snew1*A1 - Snew2*A2;
    conv = Res^2/(1+P^2);
    iter = iter + 1;
    Dep1 = E1; if epnew1 > nep1; Dep1 = Et1; end
    Dep2 = E2; if epnew2 > nep2; Dep2 = Et2; end
    nS1 = Snew1; nA1 = Anew1; nep1 = epnew1;
    nS2 = Snew2; nA2 = Anew2; nep2 = epnew2;
    fprintf('\n %3d %7.4f %7.3f %7.3f %7.3f %8.6f %8.6f %10.3e',...
           iter,u,nS1,nS2,nA1,nA2,nep1,nep2,Res);
end
■

```

4.3 Multidimensional Elastoplasticity

The basic concepts presented in the previous section for one-dimensional systems can be generalized for multidimensional systems. However, in the one-dimensional case, it is relatively straightforward to obtain the required stress–strain relation because experiments are usually performed in the uniaxial tension test. The relationship is given in terms of axial stress and strain. In addition, strain-hardening models are also relatively straightforward because either the magnitude of yields stress increases (isotropic hardening) or maintains the same range between tension and compression (kinematic hardening). These strain-hardening behaviors can also be obtained using uniaxial tension/compression tests. However, one-dimensional elastoplasticity can only be used for very limited applications, such as bars and trusses.

When a structural system is in two or three dimensions, it is more difficult to apply the theory of elastoplasticity from the previous section because now stress is not a scalar quantity, but a tensor with up to six components. If a procedure similar to the previous section is going to be used, then material tests with different combinations of stress components must be performed. This is practically impossible because there are infinite numbers of possible combinations. For example, let us consider biaxial loading of a plane stress structure, which yields one of the simplest stress states beyond the uniaxial case. Only nonzero stress components are σ_{11} and σ_{22} . One possible method of stress combination is such that σ_{11} is fixed at 100 MPa (let's say) and σ_{22} gradually increases beyond yielding of the material. Now, different combinations are possible by fixing σ_{11} at different values and gradually increasing σ_{22} . All these possible combinations of stress components must be tested in order to obtain the stress–strain relationship for biaxial loadings. Thus, it is practically impossible to obtain stress–strain relationships for multidimensional systems except for very limited cases.

Instead of developing stress–strain relationships for all possible combinations, a key concept in multidimensional elastoplasticity is to use a physics-based model that can represent all possible cases. As mentioned in the above example, the possible number of combinations is infinite even if only two stress components are involved. Thus, a scalar measure of stress and strain that can represent the multidimensional stress status should be used for practicality. In addition, this measure should be independent of coordinate systems used for an isotropic material, i.e., invariant, because the material shows identical behavior for all directions. For this purpose, an equivalent stress and an effective strain are introduced in the first section. Since the multidimensional model should also satisfy the one-dimensional case, the stress–strain relationship can be obtained from uniaxial tension/compression tests. The key ingredients are a certain form of yield criterion with a hardening rule and the elastoplastic stress–strain law relating incremental stress to strain in the plastic region. Among many different yield criteria, the von Mises yield criterion is widely used for the isotropic metal plasticity, which will be discussed in the second section. As discussed in Sect. 4.2, the kinematic and isotropic hardening models are discussed in the multidimensional system. A general formulation for the incremental elastoplastic stress–strain relationship is presented next. The remaining sections give computational details for a few specific forms commonly used in practice.

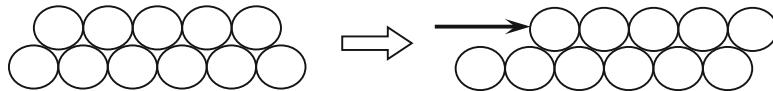


Fig. 4.9 Material failures due to relative sliding

4.3.1 Yield Functions and Yield Criteria

Material yielding occurs due to relative sliding of the material's molecules within its lattice structure, which is similar to shear deformation. The material will not come back to its original shape after the applied load is removed. As illustrated in Fig. 4.9, such a sliding deformation preserves the volume of the material. If the intermolecular distance is changed, then the volume is also changed. However, it is very difficult to have a permanent deformation of a material by changing intermolecular distance. Thus, it is commonly accepted that the material failure is related to the shear deformation.

4.3.1.1 Maximum Shear Stress Criterion

One of the simplest failure criteria for a ductile material is the maximum shear stress criterion, proposed by Tresca (1864). This criterion uses the maximum shear stress, τ_{\max} , as an equivalent stress. The maximum shear stress is the radius of the largest of Mohr's circles. Let σ_1 , σ_2 , and σ_3 be the three principal stresses, ordered by $\sigma_1 \geq \sigma_2 \geq \sigma_3$. Then, the maximum shear stress can be defined as

$$\tau_{\max} = \frac{\sigma_1 - \sigma_3}{2}. \quad (4.44)$$

Note that τ_{\max} is independent of the coordinate system used because the principal stresses are independent of the coordinate system used. This criterion assumes that material failure occurs when τ_{\max} is equal to the shear stress in a tensile specimen at yield, τ_Y . Note that in the tensile test at yield, $\sigma_1 = \sigma_Y$ and $\sigma_2 = \sigma_3 = 0$. Thus, from Mohr's circles, it is easy to find that τ_Y is a half of σ_Y . When τ_{\max} is less than τ_Y , the material is elastic. It is impossible that τ_{\max} is greater than τ_Y . In this criterion, the elastic range of the material is defined when

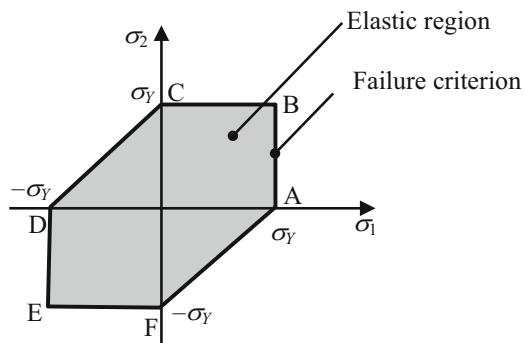
$$\tau_{\max} \leq \tau_Y \left(= \frac{1}{2} \sigma_Y \right). \quad (4.45)$$

The yield criterion is the boundary of the elastic range, i.e.,

$$f(\boldsymbol{\sigma}) = \tau_{\max} - \tau_Y = 0, \quad (4.46)$$

where $f(\boldsymbol{\sigma})$ is the yield function and $f(\boldsymbol{\sigma}) = 0$ is the yield criterion. Although the above equation has a simple form, the actual combination of multidimensional stress states

Fig. 4.10 Maximum shear stress criterion



can be complicated. Figure 4.10 shows the hexagonal failure envelope for the two-dimensional maximum shear stress theory, i.e., $\sigma_3 = 0$. From the uniaxial tension test, the material fails when $\sigma_1 = \sigma_Y$ and $\sigma_2 = 0$ (point A). If σ_2 starts increasing from point A, it does not affect the yield criterion until it increases to σ_Y (point B) because τ_{\max} is determined by σ_1 and σ_3 . Along line BC, τ_{\max} is determined by σ_2 and σ_3 . Along line CD, τ_{\max} is determined by σ_1 and σ_2 , and as σ_1 becomes more negative, σ_2 must decrease. The elastic range in Eq. (4.45) corresponds to the interior of the hexagon, in which the material is elastic. The yield criterion in Eq. (4.46) corresponds to the boundary, which is often called the **yield surface**.

Example 4.5 (Maximum Shear Stress Criterion) Consider a thin square plate in the xy plane under biaxial tension. When $\sigma_{xx} = -200$ MPa, determine $\sigma_{yy} > 0$ that makes the material yield. Use the maximum shear stress criterion with $\sigma_Y = 500$ MPa.

Solution Since σ_{xx} and σ_{yy} are only nonzero stress components, $\sigma_1 = \sigma_{yy}$, $\sigma_2 = 0$, and $\sigma_3 = \sigma_{xx}$. Thus, the maximum shear stress becomes

$$\tau_{\max} = \frac{\sigma_1 - \sigma_2}{2} = \frac{\sigma_1 + 200}{2}.$$

Since the shear stress at yield is $\tau_Y = \sigma_Y/2 = 250$ MPa, from the yield criterion

$$\tau_{\max} = \frac{\sigma_1 - 200}{2} = 250 \text{ MPa.}$$

Thus, the material yields when $\sigma_{yy} = 300$ MPa. ■

4.3.1.2 Distortion Energy Criterion

The maximum shear stress criterion is based on shear stress only. In general, a deformation can be divided into volume-changing (dilatation) and volume-preserving (distortion) parts. For example, for a given strain tensor $\boldsymbol{\epsilon}$, the *volumetric strain* in a small deformation is defined as

$$\varepsilon_V = \text{tr}(\boldsymbol{\varepsilon}) = \varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}. \quad (4.47)$$

Then, it would be more general to say that plastic deformation is related to the volume-preserving part of the strain, which is often called the **deviatoric strain**. In the case of pure shear deformation, for example, $\varepsilon_{12} = \varepsilon_{21}$ are the only nonzero strain components. Thus, the volumetric strain becomes 0, and the shear deformation is identical to the deviatoric strain. However, in the case of more complex deformations, the deviatoric strain may have nonzero normal strain components, e.g., $\varepsilon_{11} = -(\varepsilon_{22} + \varepsilon_{33})$. The difficulty in using the deviatoric strain or stress for the yield criterion is that it is not a scalar; it has six components in general and depends on the coordinate system used. Thus, in order to use the deviatoric strain or stress, it is necessary to have a scalar quantity that is defined from deviatoric strain or stress.

The next and more important criterion is the **distortion energy criterion**, which is the main focus of this chapter. As stress gradually increases, so does the **strain energy density** of the material. Since these two quantities are correlated, it is possible to use the strain energy density for a failure criterion. The advantage of using the strain energy density is that it is always a scalar even if all six stress components are nonzero. As previously discussed, **since volumetric deformation would not contribute to the material failure**, this part of deformation must be removed from the strain energy density before using it for the failure criterion. The strain energy density after removing the volumetric part is called the **distortion strain energy density**. The concept of the distortion energy criterion is to compare the distortion energy of a multidimensional stress state to that of a tensile test at yield. The material is considered to have failed when the distortion energy from multidimensional stress has the same value with that of the tensile test at yield. When the distortion energy is less than that of the tensile test at yield, the material is considered to be elastic.

The distortion strain energy is defined as the difference of the strain energy from its volumetric part. In order to calculate the distortion energy, the volumetric parts of stress (hydrostatic pressure) and strain (mean strain) are first defined as

$$\sigma_m = \frac{1}{3}\text{tr}(\boldsymbol{\sigma}) = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \quad (4.48)$$

and

$$\varepsilon_m = \frac{1}{3}\text{tr}(\boldsymbol{\varepsilon}) = \frac{1}{3}(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}), \quad (4.49)$$

where $\text{tr}(\bullet)$ is a trace operator such that $\text{tr}(\boldsymbol{\sigma}) = \sigma_{kk}$. Note that the volumetric strain differs from the mean strain by a factor of three. In order to calculate the distortion energy, the **deviatoric stress and strain tensors** are defined as

$$\mathbf{s} \equiv \boldsymbol{\sigma} - \sigma_m \mathbf{1} = \mathbf{I}_{\text{dev}} : \boldsymbol{\sigma} \quad (4.50)$$

and

$$\mathbf{e} \equiv \boldsymbol{\varepsilon} - \varepsilon_m \mathbf{1} = \mathbf{I}_{\text{dev}} : \boldsymbol{\varepsilon}, \quad (4.51)$$

where $\mathbf{1} = [\delta_{ij}]$ is the second-order unit tensor, \mathbf{I} is the fourth-order unit symmetric tensor defined as $I_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$, $\mathbf{I}_{\text{dev}} = \mathbf{I} - \frac{1}{3}\mathbf{1} \otimes \mathbf{1}$ is the fourth-order unit deviatoric tensor, \otimes is the tensor product, and “ $:$ ” is the double contraction operator of tensors. Note that $\text{tr}(\mathbf{s}) = 0$ and $\text{tr}(\mathbf{e}) = 0$.

Example 4.6 (Fourth-Order Unit Symmetric Tensor) Show that $\mathbf{e} = \mathbf{I} : \mathbf{e}$ using the definition of $I_{ijkl} = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$.

Solution From the property that the Kronecker delta symbol changes index, i.e., $\delta_{ij}\delta_{jk} = \delta_{ik}$, the double contraction between the fourth-order unit symmetric tensor with strain tensor becomes

$$I_{ijkl}\varepsilon_{kl} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\varepsilon_{kl} = \frac{1}{2}(\varepsilon_{ij} + \varepsilon_{ji}) = \varepsilon_{ij}.$$

The symmetric property of the strain tensor is used in the last equality. ■

For isotropic, linear elastic materials, the constitutive relation between the stress and strain can be written as

$$\boldsymbol{\sigma} = [\lambda\mathbf{1} \otimes \mathbf{1} + 2\mu\mathbf{I}] : \mathbf{e} \equiv \mathbf{D} : \mathbf{e}, \quad (4.52)$$

where λ and μ are Lamé's constants, and \mathbf{D} is the fourth-order constitutive tensor. Using the property of $\mathbf{1} : \mathbf{e} = \text{tr}(\mathbf{e}) = 3\varepsilon_m$, it is possible to decompose the constitutive relation in Eq. (4.52) into volumetric and deviatoric parts, as

$$\begin{aligned} \boldsymbol{\sigma} &= \lambda(3\varepsilon_m)\mathbf{1} + 2\mu(\mathbf{e} + \varepsilon_m\mathbf{1}) \\ &= (3\lambda + 2\mu)\varepsilon_m\mathbf{1} + 2\mu\mathbf{e}. \end{aligned} \quad (4.53)$$

If the above equation is compared with Eq. (4.50), the first part on the right-hand side is the volumetric part and the second is the deviatoric part of the stress. Thus, the following decomposed constitutive relation can be obtained:

$$\mathbf{s} = 2\mu\mathbf{e} \quad (4.54)$$

and

$$\sigma_m = (3\lambda + 2\mu)\varepsilon_m. \quad (4.55)$$

It is straightforward to show that the bulk modulus (the constant that relates σ_m with ε_V) is defined as $K = (3\lambda + 2\mu)/3$. It is interesting to note that only the shear modulus μ appears in the relation between deviatoric stress and strain.

Using Eqs. (4.50) and (4.51), the distortion energy density can be defined as

$$U_d = \frac{1}{2}\mathbf{s} : \mathbf{e}. \quad (4.56)$$

Using the relation $\mathbf{s} = 2\mu\mathbf{e}$, the deviatoric strain energy density can be rewritten in terms of stress as

$$U_d = \frac{1}{4\mu} \mathbf{s} : \mathbf{s}. \quad (4.57)$$

In the case of a one-dimensional tensile test, the material fails when $\sigma_{11} = \sigma_Y$ and all other stress components are 0. Then, the deviatoric stress becomes

$$\mathbf{s} = \begin{bmatrix} \frac{2}{3}\sigma_Y & 0 & 0 \\ 0 & -\frac{1}{3}\sigma_Y & 0 \\ 0 & 0 & -\frac{1}{3}\sigma_Y \end{bmatrix}. \quad (4.58)$$

Thus, the distortion energy density at the status of material failure in the tensile test becomes

$$U_d|_{1D} = \frac{1}{6\mu} \sigma_Y^2. \quad (4.59)$$

A material in multidimensional stress status yields when the distortion energy density becomes equal to that of the tensile test at the yield point. By equating Eq. (4.57) with Eq. (4.59), it can be concluded that the material in the multidimensional stress status yields when the following stress measure equals to the one-dimensional yield stress:

$$\sigma_e \equiv \sqrt{\frac{3}{2} \mathbf{s} : \mathbf{s}} = \sigma_Y. \quad (4.60)$$

In the above equation, σ_e is called the equivalent stress or the von Mises stress. The material is considered to be elastic when the equivalent stress is less than the one-dimensional yield stress. Thus, a similar state determination as in one-dimensional stress can be used for multidimensional cases using the equivalent stress. Note that even if the equivalent stress is used to determine the state of material failure, it comes from the criterion based on distortion energy.

The counterpart of equivalent stress is the effective strain, which is defined using the definition of distortion strain energy density as

$$U_d = \frac{1}{2} \mathbf{s} : \mathbf{e} = \frac{1}{2} \sigma_e e_e, \quad (4.61)$$

where e_e is the *effective strain*. Using Eqs. (4.57) and (4.60), the distortion energy U_d can be expressed in terms of equivalent stress, as

$$U_d = \frac{1}{4\mu} \left(\frac{2}{3} \sigma_e^2 \right) = \frac{1}{6\mu} \sigma_e^2. \quad (4.62)$$

By comparing Eq. (4.61) with Eq. (4.62), the effective strain can be written in terms of equivalent stress, as

$$e_e = \frac{1}{3\mu} \sigma_e = \frac{1}{3\mu} \sqrt{\frac{3}{2} \mathbf{s} : \mathbf{s}}. \quad (4.63)$$

By using the relation $\mathbf{s} = 2\mu\mathbf{e}$, the effective strain can be expressed in terms of deviatoric strain as

$$e_e = \sqrt{\frac{2}{3} \mathbf{e} : \mathbf{e}}. \quad (4.64)$$

Note that the equivalent stress in Eq. (4.60) and the effective strain in Eq. (4.64) have similar definitions from deviatoric stress and strain, respectively, except for the coefficients. **The usage of effective strain is related to the plastic deformation.** The scalar plastic strain in one-dimensional plasticity becomes a tensor in multidimensional plasticity.

Example 4.7 (One-Dimensional Equivalent Strain) A bar is under axial stress σ and axial strain ϵ . Calculate the equivalent stress and effective strain in terms of σ , ϵ , and Poisson's ratio ν .

Solution For a uniaxial tension problem, $\sigma_{11} = \sigma$ is the only nonzero stress component. Thus, the deviatoric stress becomes

$$\mathbf{s} = \begin{bmatrix} \frac{2}{3}\sigma & 0 & 0 \\ 0 & -\frac{1}{3}\sigma & 0 \\ 0 & 0 & -\frac{1}{3}\sigma \end{bmatrix}.$$

Using Eq. (4.60), the equivalent stress can be obtained by

$$\sigma_e = \sqrt{\frac{3}{2} \mathbf{s} : \mathbf{s}} = \sigma.$$

Note that the equivalent stress is identical to the axial stress. Thus, the material will yield when the axial stress reaches the yield stress, which is consistent to the definition of the yield stress.

For a uniaxial tension problem, three nonzero strain components are $\varepsilon_{11} = \varepsilon$, $\varepsilon_{22} = \varepsilon_{33} = -\nu\varepsilon$. All shear strains are equal to 0. From Eq. (4.49), the mean strain becomes $\varepsilon_m = (1 - 2\nu)\varepsilon/3$. The deviatoric strain is then obtained from Eq. (4.51), as

$$\mathbf{e} = \frac{(1 + \nu)\varepsilon}{3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

The effective strain can be obtained from Eq. (4.64), as

$$\mathbf{e} : \mathbf{e} = 6 \left(\frac{(1 + \nu)\varepsilon}{3} \right)^2,$$

$$e_e = \sqrt{\frac{2}{3} \mathbf{e} : \mathbf{e}} = \frac{2(1 + \nu)}{3} \varepsilon.$$

■

4.3.2 Von Mises Yield Criterion

The distortion energy theory in the previous section is also called the **von Mises yield criterion**, which states that yielding occurs when the equivalent stress reaches the yield stress of the material in uniaxial tension. The equivalent stress σ_e in Eq. (4.60) can be expressed as follows:

$$\sigma_e = \sqrt{\frac{3}{2} \mathbf{s} : \mathbf{s}} \equiv \sqrt{3J_2}, \quad (4.65)$$

$J_2 = -I_2$

where I_2 is the second invariant of the deviatoric stress.² It can be expressed in several alternative forms as follows. In terms of stress components, it can be written as

$$J_2 = \frac{1}{6} [(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2] + \tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2. \quad (4.66)$$

Or, in terms of principal stresses,

$$J_2 = \frac{1}{6} [(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2]. \quad (4.67)$$

² The same symbol was used for the reduced invariant of the Cauchy–Green deformation tensor in Chap. 3. Since this symbol is widely used in the literature, it is kept here.

Using the equivalent stress, the **yield function** and corresponding yield criterion can be defined as

$$f'(\sigma) \equiv \sigma_e^2 - \sigma_Y^2 = 3J_2 - \sigma_Y^2 = 0, \quad (4.68)$$

where σ_Y is the yield stress from the tensile test. Thus, even if the status of stress is multidimensional, one-dimensional experimental data can still be used by considering the equivalent stress as a tensile stress.

From the definition of J_2 in terms of principal stresses, it can easily be seen that the **von Mises yield function represents an ellipse in two dimensions**. As shown in Fig. 4.11, any point inside the ellipse ($f' < 0$) represents an elastic stress state. The inside of the ellipse is called the elastic domain of the material. Points on the yield surface ($f' = 0$) correspond to the stress state that causes the material to yield. It is impossible for a stress state to reside outside of the yield surface.

Figure 4.11 also plots the maximum shear stress criterion. Note that the two criteria meet at the six vertices, but the maximum shear stress criterion is inside of the von Mises criterion, which means that the former is more conservative than the latter. For example, in the situation of uniaxial tension, which corresponds to the stress state along the σ_1 axis, both criteria predict the same yield point at $\sigma_1 = \sigma_Y$. However, in the situation of pure shear stress, along the line OA in Fig. 4.11, the maximum shear stress criterion predicts the material failure earlier than the von Mises criterion. It is also noted that the von Mises yield surface is smooth, while the maximum shear stress criterion has six vertices. **The smoothness of the yield surface helps to find the yield point numerically.**

A plastic deformation can be physically explained by atomic dislocation. An elastic deformation corresponds to the variation in the intermolecular distance without causing atomic dislocation, while a plastic deformation implies relative

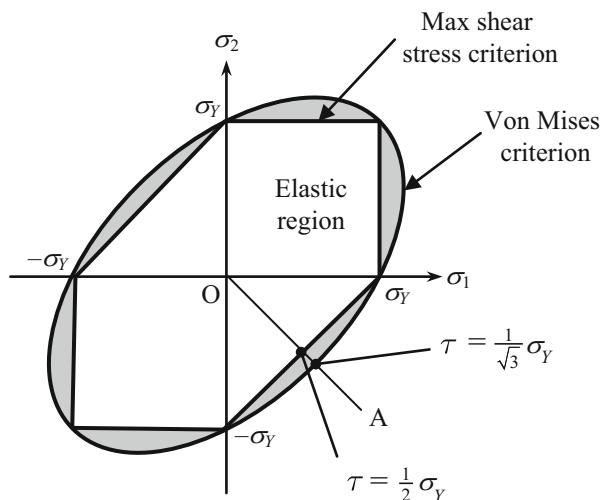


Fig. 4.11 Von Mises yield criterion

sliding of the atomic layers and a permanent shape change without changing the structural volume. Thus, plastic behavior can be efficiently described by the deviator of a tensor, which preserves the volumetric components. As previously discussed in Eq. (4.65), J_2 is the second invariant of the deviatoric stress. Specifically, consider a stress tensor $\sigma = [\sigma_{ij}]$, ($i, j = 1, 2, 3$). The second invariant of the deviatoric stress can be written as

$$J_2 = \frac{1}{2}[\mathbf{s} : \mathbf{s} - \text{tr}(\mathbf{s})^2] = \frac{1}{2}\mathbf{s} : \mathbf{s}. \quad (4.69)$$

Note that $\text{tr}(\mathbf{s}) = 0$ because the trace part of the stress tensor is moved to σ_m .

The yield function in Eq. (4.68) is given as a square of effective stress. Since both the effective stress and the yield stress are always positive, it is unnecessary to define the yield criterion using square terms. It is also more convenient to define the yield criterion without having squares so that its unit is the same as that of stress. Thus, the von Mises yield criterion can be rewritten as

$$f(\sigma) = ||\mathbf{s}|| - \sqrt{\frac{2}{3}}\sigma_Y = 0, \quad (4.70)$$

where $||\mathbf{s}|| = (\mathbf{s} : \mathbf{s})^{1/2}$ is the norm of the deviatoric stress. The ellipse in Fig. 4.11 becomes a circle if it is plotted in the principal deviatoric stresses. It is then possible to consider that $\sqrt{(2/3)}\sigma_Y$ is the radius of the yield circle in deviatoric stress space.

Example 4.8 (Pure Shear Deformation) Consider a pure shear deformation of a plane strain square block, as shown in Fig. 4.12 where a constant shear stress τ is applied. The yield stress of the material is σ_Y . Calculate the shear stress τ when the material yields using the yield function formula in Eq. (4.70).

Solution In the pure shear problem, the stress tensor is identical to the deviatoric stress because the volumetric part of the stress vanishes:

$$\sigma = \begin{bmatrix} 0 & \tau & 0 \\ \tau & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{s}.$$

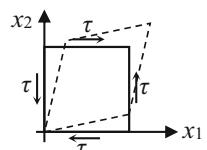


Fig. 4.12 Pure shear deformation

Then, the von Mises yield function can be written as

$$f(\sigma) = \|\mathbf{s}\| - \sqrt{\frac{3}{2}\sigma_Y} = \sqrt{2}\tau - \sqrt{\frac{2}{3}\sigma_Y} = 0.$$

Thus, the material yields when the shear stress becomes

$$\tau = \frac{1}{\sqrt{3}}\sigma_Y.$$

Note that this value of shear stress is different from that of the maximum shear stress criterion, $\tau = \sigma_Y/2$. In Fig. 4.10, line OA represents the case when a shear stress gradually increases in a pure shear problem. Based on the maximum shear stress criterion, the material yields at $\tau = \sigma_Y/2$, while $\tau = \sigma_Y/\sqrt{3}$ from the von Mises criterion. The maximum shear stress criterion provides more conservative yield point. ■

Example 4.9 (Uniaxial Tensile Test) An axial stress σ is applied to a uniaxial bar. The yield stress of the material is σ_Y . Calculate the tensile stress σ when the material yields using the yield function formula in Eq. (4.70).

Solution In the uniaxial tension problem, the stress and its deviator can be written as

$$\sigma = \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{s} = \frac{\sigma}{3} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Then, the norm of the deviatoric stress becomes

$$\|\mathbf{s}\| = \frac{\sigma}{3} \sqrt{2^2 + 1^2 + 1^2} = \sqrt{\frac{2}{3}\sigma}.$$

Thus, the von Mises yield criterion can be written as

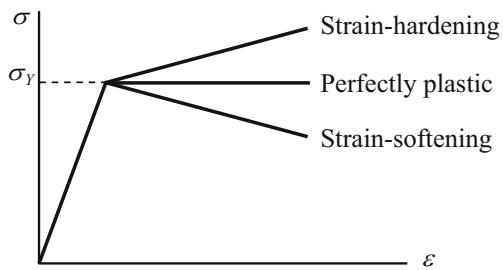
$$f(\sigma) = \sqrt{\frac{2}{3}\sigma} - \sqrt{\frac{2}{3}\sigma_Y} = 0 \Rightarrow \sigma = \sigma_Y.$$

Thus, the von Mises yield criterion is consistent with the determination of yield stress in tension test. ■

4.3.3 Hardening Models

The yield criterion in the previous section determines whether a material yields or not, based on the given yield stress of the material. In Eq. (4.70), it is assumed that the

Fig. 4.13 Post-plastic behaviors of materials



yield stress of the material remains constant. However, as discussed in Sect. 4.2, the yield stress itself varies according to the plastic deformation, which is called strain-hardening. In general, there are three classes of materials in terms of the yield stress varies due to plastic deformation (see Fig. 4.13): (1) yield stress increases proportional to plastic deformation (strain-hardening), (2) yield stress remains constant (perfectly plastic), and (3) yield stress decreases as plastic deformation increases (strain-softening). Generally, metals are strain-hardening materials, and geotechnical materials may exhibit strain-softening under certain conditions. The strain-hardening material is regarded as stable and will be considered in this section.

In Fig. 4.1, two different modes of strain-hardening were discussed: isotropic and kinematic hardenings. The elastic range of the isotropic hardening model continuously grows due to plastic deformation, while it remains constant for the kinematic hardening model but moves parallel to the strain-hardening line. These definitions of strain-hardening models can be extended to multidimensional plasticity. The yield criterion in Eq. (4.70) can be considered as an equation of a circle with the center at the origin and a radius of $\sqrt{(2/3)\sigma_Y}$. In the isotropic hardening model, the location of the center is fixed, and the radius increases uniformly, i.e., σ_Y increases. On the other hand, for the kinematic hardening model, the radius is fixed, and the location of the center moves in the stress space; i.e., the norm of deviatoric stress changes to $\|\mathbf{s} - \boldsymbol{\alpha}\|$ where $\boldsymbol{\alpha}$ is the location of center of the yield surface. Figure 4.14 illustrates these two hardening models. Since tensile tests are to be used for describing the material behavior beyond yielding, the same hardening parameters that are used in one-dimensional plasticity should be used for multidimensional hardenings. In the following, these two hardening models will be discussed in detail.

4.3.3.1 Isotropic Hardening

In the isotropic hardening model, the subsequent yielding depends on the accumulated effective plastic strain e_p . For the linear isotropic hardening model, the yield stress increases according to the effective plastic strain as

$$\sigma_Y = \sigma_Y^0 + H e_p, \quad (4.71)$$

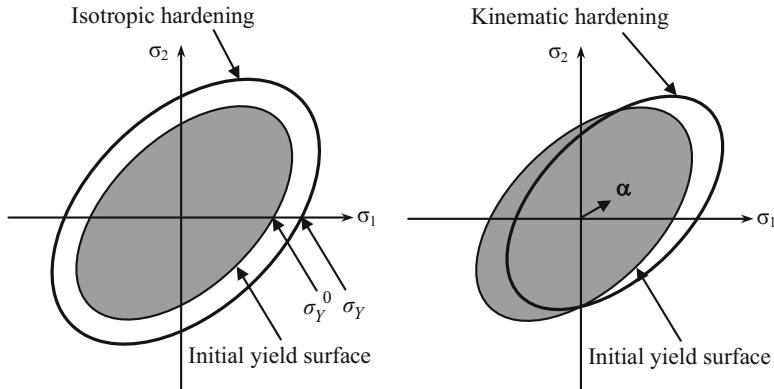


Fig. 4.14 Hardening models in two dimension

where the plastic modulus H is obtained from the uniaxial stress–strain relationship as follows:

$$H = \frac{\Delta\sigma}{\Delta e_p}. \quad (4.72)$$

The radius of the yield surface increases according to the effective plastic strain proportional to the plastic modulus. The above definition of plastic modulus can be applicable for general nonlinear hardening because it is defined as a rate form. In such a case, H will be the slope of stress–plastic strain curve at a given total plastic strain. In the case of linear hardening, H will be a constant.

When $H=0$, the material is called elasto-perfectly-plastic. The class of perfectly plastic material is an idealization with the purpose of keeping the constitutive equation simple. This idealization is reasonable for materials that do not show significant strain-hardening. The adequacy of this idealization depends on the purpose and requirement of specific applications. If only monotonic loading is of interest and does not call for a refined solution, then this idealization may lead to a satisfactory solution. However, due to progress in industries which give rise to problems that are subjected to complex loading conditions and impose stricter requirements, this idealization is no longer adequate in many applications and strain-hardening should be considered.

4.3.3.2 Kinematic Hardening

In the kinematic hardening, the subsequent yield surfaces are shifted in the stress space (see Fig. 4.14). Thus, the equation for a subsequent yield surface can be obtained from the one used for initial yield surface by introducing a shift in stress. This shift in the center of the yield surface is called back stress and denoted by α .

Then, the distance from the center of the yield surface to the yield surface can be measured by the difference of \mathbf{s} from $\boldsymbol{\alpha}$. It would be convenient to define the following shifted stress first:

$$\boldsymbol{\eta} = \mathbf{s} - \boldsymbol{\alpha} \quad (4.73)$$

Note that both the back stress and shifted stress are deviatoric stresses. Next, the equation for subsequent yield surface is defined as follows:

$$\|\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}\sigma_Y = 0. \quad (4.74)$$

In kinematic hardening, σ_Y is the initial yield stress and remains constant. The back stress depends on the current stress and the accumulated effective plastic strain e_p . According to [Ziegler's rule](#), the increment in back stress of the linear kinematic hardening model is written as

$$\Delta\boldsymbol{\alpha} = \sqrt{\frac{2}{3}}H\Delta e_p \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|}. \quad (4.75)$$

As with the isotropic hardening, the effective plastic strain plays an important role in determining the evolution of the back stress. In addition, the back stress increases in the parallel direction with the shifted stress. Since $\boldsymbol{\eta} = \mathbf{s} - \boldsymbol{\alpha}$ is the radial direction of the yield surface, the increment is always in the radial direction.

4.3.3.3 Combined Hardening

The difference between isotropic and kinematic hardening is clear. The former increases the radius of the yield surface, while the latter moves the center of the yield surface. However, many materials show a combined behavior of both models; i.e., the yield stress increases due to plastic deformation, but the material yields earlier in the opposite direction. This is caused by dislocation pileups and tangles (back stress). When strain direction is changed, this makes the dislocation easy to move. A combined linear isotropic/kinematic hardening model uses a parameter $\beta \in [0,1]$ to consider this combined effect, called the [Bauschinger effect](#). In this model, the yield surface is defined as

$$\|\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}[\sigma_Y^0 + (1 - \beta)He_p] = 0 \quad (4.76)$$

and the increment of the back stress is determined by

$$\Delta\boldsymbol{\alpha} = \sqrt{\frac{2}{3}}\beta H\Delta e_p \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|}. \quad (4.77)$$

This model is general enough to represent both isotropic and kinematic hardening as special cases: β equals one for kinematic hardening and 0 for isotropic hardening.

Example 4.10 (Isotropic/Kinematic Hardenings) A uniaxial bar is under proportional loading with axial stress σ . When the effective plastic strain is $e_p = 0.1$, calculate the value of axial stress. Consider three different hardening models: (a) isotropic, (b) kinematic, and (c) combined hardening with $\beta = 0.5$. Assume that the initial yield stress is 400 MPa and the plastic modulus is $H = 200$ MPa.

Solution Since the applied stress is proportional loading, it is expected that the material is in the plastic phase, and all three models provide the same stress value. The difference occurs only when the direction of loading changes. In the case of uniaxial tension, the stress and deviatoric stress become

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{s} = \begin{bmatrix} \frac{2}{3}\sigma & 0 & 0 \\ 0 & -\frac{1}{3}\sigma & 0 \\ 0 & 0 & -\frac{1}{3}\sigma \end{bmatrix}.$$

Thus, the norm of the deviatoric stress becomes

$$\|\mathbf{s}\| = \sqrt{\frac{2}{3}}\sigma.$$

(a) Isotropic hardening: from the definition of the yield function,

$$\|\mathbf{s}\| - \sqrt{\frac{2}{3}}(\sigma_Y^0 + He_p) = \sqrt{\frac{2}{3}}\sigma - \sqrt{\frac{2}{3}}(400 + 200 \times 0.1) = 0, \\ \sigma = 420 \text{ MPa.}$$

(b) Kinematic hardening: from the definition of yield function,

$$\|\mathbf{s} - \boldsymbol{\alpha}\| - \sqrt{\frac{2}{3}}\sigma_Y^0 = 0.$$

Note that $\Delta\boldsymbol{\alpha}$ is parallel to \mathbf{n} and the loading direction remains fixed. Thus, $\boldsymbol{\alpha}$ is parallel to \mathbf{s} , in which the norm of the shifted stress can be written as $\|\mathbf{s} - \boldsymbol{\alpha}\| = \|\mathbf{s}\| - \|\boldsymbol{\alpha}\|$. Thus, the yield function can be rewritten as

$$\|\mathbf{s} - \boldsymbol{\alpha}\| - \sqrt{\frac{2}{3}}\sigma_Y^0 = \|\mathbf{s}\| - \|\boldsymbol{\alpha}\| - \sqrt{\frac{2}{3}}\sigma_Y^0 = \sqrt{\frac{2}{3}}\sigma - \sqrt{\frac{2}{3}}He_p - \sqrt{\frac{2}{3}}\sigma_Y^0 = 0,$$

$$\sigma = \sigma_Y^0 + H e_p = (400 + 200 \times 0.1) = 420 \text{ MPa.}$$

- (c) Combined hardening: Similar to the kinematic hardening model, α is parallel to s . Thus, the yield function can be written as

$$\begin{aligned} ||s - \alpha|| - \sqrt{\frac{2}{3}}[\sigma_Y^0 + (1 - \beta)H e_p] \\ = ||s|| - ||\alpha|| - \sqrt{\frac{2}{3}}[\sigma_Y^0 + (1 - \beta)H e_p] \\ = \sqrt{\frac{2}{3}}\sigma - \sqrt{\frac{2}{3}}\beta H e_p - \sqrt{\frac{2}{3}}\sigma_Y^0 - \sqrt{\frac{2}{3}}(1 - \beta)H e_p \\ = 0 \end{aligned} .$$

Thus, the applied stress can be solved:

$$\sigma = \sigma_Y^0 + H e_p = (400 + 200 \times 0.1) = 420 \text{ MPa.}$$

Note that all three models provide the same stress value. ■

4.3.4 Classical Elastoplasticity Model

In the previous section, the yield function and strain-hardening models for multidimensional plasticity were discussed. With these models, the objective of this section is to determine the current stress state and evolution of plastic variables. Based on the von Mises yield criterion and isotropic/kinematic hardenings, the plastic variables are effective plastic strain, e_p , and back stress, α . The former is a scalar, while the latter is a second-order tensor. First, all relations will be derived in the rate form, and then written in the incremental form for the purpose of numerical integration in the following section. For example, the relation between strain rate $\dot{\epsilon}$ and strain increment $\Delta\epsilon$ is that $\Delta\epsilon = \Delta t \dot{\epsilon}$ where Δt is the time increment.

1. Additive Decomposition

The fundamental assumption in small deformation elastoplasticity is that the elastic and plastic parts can be decomposed additively. This assumption is a fundamental difference compared to the finite deformation elastoplasticity. Assuming a small elastic strain, the strain and its rate can be additively decomposed into elastic and plastic parts as

$$\epsilon = \epsilon^e + \epsilon^p, \quad \dot{\epsilon} = \dot{\epsilon}^e + \dot{\epsilon}^p, \quad (4.78)$$

where superscripts e and p denote elastic and plastic parts, respectively. The superposed “dot” denotes the rate of a quantity. It is reminded that in static problems, the rate is equivalent to the load increment. From the assumption that

the plastic deformation only occurs in the deviatoric space, the plastic strain $\boldsymbol{\epsilon}^p$ and its rate are deviatoric tensors, i.e., $\text{tr}(\dot{\boldsymbol{\epsilon}}^p) = 0$. As with one-dimensional plasticity, the total strain and its rate are given. However, it is unknown how much of them are elastic or plastic. Thus, the objective is to find the elastic or plastic strain for a given total strain and its rate. The elastic strain produces stress in the material, while the plastic strain is independent of stress. However, the plastic strain affects the yield stress in the strain-hardening model.

2. Strain Energy Density

It is usually assumed that a strain energy density exists for the elastic part, such that the stress can be determined by taking a derivative of the strain energy density with respect to the elastic strain. The elastic part of the elastoplasticity model is the same as a linear elastic material. Since the relationship between stress and elastic strain is linear, this function takes a quadratic form. Thus, the following form of strain energy density can be considered:

$$W(\boldsymbol{\epsilon}^e) = \frac{1}{2}\boldsymbol{\epsilon}^e : \mathbf{D} : \boldsymbol{\epsilon}^e = \frac{1}{2}(\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^p) : \mathbf{D} : (\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^p), \quad (4.79)$$

where \mathbf{D} is a fourth-order constitutive tensor. The elastic part of the strain is usually unknown until the plastic behavior of the material is identified. By differentiating the above definition, stress can be related to the elastic strain as

$$\boldsymbol{\sigma} = \frac{\partial W(\boldsymbol{\epsilon}^e)}{\partial \boldsymbol{\epsilon}^e} = \mathbf{D} : \boldsymbol{\epsilon}^e = \mathbf{D} : (\boldsymbol{\epsilon} - \boldsymbol{\epsilon}^p). \quad (4.80)$$

Now, it is clear that \mathbf{D} is the constitutive tensor introduced in Eq. (4.52) when the strain is interpreted as elastic. Due to the assumption in Eq. (4.78), the rate form of the above equation can be written as

$$\dot{\boldsymbol{\sigma}} = \mathbf{D} : (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p), \quad (4.81)$$

where $\mathbf{D} = (\lambda + (2/3)\mu)\mathbf{1} \otimes \mathbf{1} + 2\mu\mathbf{I}_{\text{dev}}$ is the fourth-order isotropic constitutive tensor. Using the decompositions in Eqs. (4.54) and (4.55), the relationship in Eq. (4.81) can be further decomposed into volumetric and deviatoric parts as

$$\dot{\sigma}_m = \frac{1}{3}\text{tr}(\dot{\boldsymbol{\sigma}}) = \left(\lambda + \frac{2}{3}\mu\right)\text{tr}(\dot{\boldsymbol{\epsilon}}) = (3\lambda + 2\mu)\dot{\epsilon}_m \quad (4.82)$$

and

$$\dot{\mathbf{s}} = 2\mu(\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p), \quad (4.83)$$

respectively. In Eq. (4.82), the property that $\text{tr}(\dot{\boldsymbol{\epsilon}}^p) = 0$ is used. As can be found in Eq. (4.82), the volumetric stress (hydrostatic pressure) is independent of plastic deformation, which is consistent with the fact that the yield function is defined using the deviatoric stress alone.

3. Yield Function

For metal plasticity, the von Mises yield criterion with the associative flow rule is commonly used to describe material behavior after elastic deformation. Accordingly, the yield criterion or yield function is formulated as

$$f(\boldsymbol{\eta}, e_p) \equiv \|\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}\kappa(e_p) \leq 0, \quad (4.84)$$

where $\boldsymbol{\eta} = \mathbf{s} - \boldsymbol{\alpha}$ is the shifted stress; $\boldsymbol{\alpha}$ is the back stress, which is the center of the yield surface (the elastic domain), and is determined by the kinematic hardening model; $\kappa(e_p)$ is the radius of the elastic domain determined by the isotropic hardening model; and e_p is the effective plastic strain. The combined isotropic/kinematic hardening model is used in Eq. (4.84). The elastic domain generated by the yield function in Eq. (4.84) forms a convex set as

$$E = \{(\boldsymbol{\eta}, e_p) | f(\boldsymbol{\eta}, e_p) \leq 0\}. \quad (4.85)$$

In general, the yield surface defined by f in the above equation is smooth and convex. In mathematical terms, the plasticity can be thought of as a projection of the stress onto the yield surface. If the material is assumed to be purely elastic, the stress will be much higher than that of the elastoplastic material for a given strain. Then, elastoplasticity projects this stress onto the yield surface because it is impossible for the stress to be outside of the elastic domain. Since the yield surface is convex, the projection becomes a contraction mapping, which guarantees the existence of a unique projection. The same concept has been applied in the case of one-dimensional plasticity in Sect. 4.2. Initially the stress increment is assumed to be purely elastic. If the estimated stress becomes larger than the yield stress, this stress is brought back to the yield stress. This procedure would be relatively easy if the yield stress is fixed. However, while the stress is brought back to the yield surface, the yield surface itself changes according to the hardening model. Thus, it is necessary to identify how the plastic deformation modifies the yield function.

4. Associative Flow Rule

The flow rule determines the evolution of the plastic strain $\boldsymbol{\epsilon}^p$. In the case of one-dimensional plasticity, the plastic strain is a scalar and its value only increases. For multidimensional plasticity, since $\boldsymbol{\epsilon}^p$ is a tensor, it is necessary to determine its magnitude as well as its direction. Thus, a general form of the flow rule can be written as

$$\dot{\boldsymbol{\epsilon}}^p = \gamma \mathbf{r}(\boldsymbol{\sigma}, \boldsymbol{\xi}), \quad (4.86)$$

where $\boldsymbol{\xi} = (\boldsymbol{\alpha}, e_p)$ represents the plastic variables, and γ is called a plastic consistency parameter. In general, $\gamma \geq 0$ where there is no plastic deformation at 0. This is consistent to the fact that the plastic strain only increases in one-dimensional plasticity.

Advanced elastoplasticity shows that the flow rule can be obtained from the constrained optimization theory in which Eq. (4.84) is an inequality constraint. Referring to the principle of minimum potential energy in Chap. 3, the structural equilibrium of an elastic material can be obtained by minimizing the potential energy, which is the sum of strain energy and potential of applied loads. Thus, the equilibrium equation is obtained from the optimality condition; i.e., the first-order derivatives of the potential energy become 0. In elastoplasticity, this optimization problem is modified such that the stress must stay within the elastic domain in Eq. (4.84). This condition can be considered as a constraint to the optimization problem. If the calculated stress from minimizing potential energy stays inside of the elastic domain, no constraint is required. However, if the calculated stress is outside of the elastic domain, it needs to be brought back to the boundary of the elastic domain. In that case, the plastic consistency parameter becomes a Lagrange multiplier to impose the constraint, and it is always nonnegative.

The expression of $\mathbf{r}(\boldsymbol{\sigma}, \xi)$ depends on the plasticity model. It is often assumed that there is a flow potential (or plastic potential), g , such that the plastic strain evolves in the direction normal to the flow potential. That is,

$$\dot{\epsilon}^p = \gamma \frac{\partial g(\boldsymbol{\sigma}, \xi)}{\partial \boldsymbol{\sigma}}, \quad (4.87)$$

where $g(\boldsymbol{\sigma}, \xi)$ is the flow potential. When the flow potential is the same as the yield function, the plastic model is called associative. Thus,

$$\dot{\epsilon}^p = \gamma \frac{\partial f(\boldsymbol{\eta}, e_p)}{\partial \boldsymbol{\eta}} = \gamma \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|} = \gamma \mathbf{N}, \quad (4.88)$$

where \mathbf{N} is a unit deviatoric tensor normal to the yield surface, and γ is a plastic consistency parameter, which is nonnegative. If the material status is elastic, γ must be 0, but if it is plastic, then γ must be positive. Thus, the plastic strain increases in the direction normal to the yield surface and has the magnitude of plastic consistency parameter γ .

As the material undergoes plastic deformation, the plastic variables (back stress and effective plastic strain) also change according to the hardening model. A general form of hardening rule can be written as

$$\dot{\xi} = \gamma \mathbf{h}(\boldsymbol{\sigma}, \xi). \quad (4.89)$$

Note that the evolution of the plastic variables is also proportional to the plastic consistency parameter. In particular, the rate of back stress can be determined by the kinematic hardening model as

$$\dot{\alpha} = H_\alpha(e_p) \gamma \frac{\partial f(\boldsymbol{\eta}, e_p)}{\partial \boldsymbol{\eta}} = H_\alpha(e_p) \gamma \mathbf{N}, \quad (4.90)$$

where $H_\alpha(e_p)$ is a nonlinear form of the plastic modulus for kinematic hardening. In the case of linear hardening, it becomes a constant, $H_\alpha(e_p) = H$. The rate of effective strain can be expressed by

$$\dot{e}_p = \sqrt{\frac{2}{3}} |\dot{\epsilon}^p(t)| = \sqrt{\frac{2}{3}} \gamma, \quad (4.91)$$

where $\dot{\epsilon}^p$ is the rate of deviatoric plastic strain.

Although it is not covered in this text, nonlinear hardening models are also available in literature. For example, a nonlinear kinematic hardening can be defined as the following evolution of back stress:

$$\dot{\alpha} = H(e_p) \dot{\epsilon}^p, \quad H(e_p) = H_0 \exp\left(-\frac{e_p}{e_p^\infty}\right), \quad (4.92)$$

where e_p^∞ is the asymptotic limit of the plastic strain, and H_0 is the initial hardening modulus. This is also called the saturated hardening model. In the case of nonlinear isotropic hardening, it is possible to define the following form of the radius of yield surface:

$$\kappa(e_p) = \sigma_Y^0 + (\sigma_Y^\infty - \sigma_Y^0) \left[1 - \exp\left(-\frac{e_p}{e_p^\infty}\right) \right], \quad (4.93)$$

where σ_Y^∞ is the asymptotic limit of the yield stress.

5. Plastic Consistency Parameter

As mentioned before, γ is 0 when the material is elastic ($f < 0$) and positive when plastic ($f = 0$). In optimization, this is called the Kuhn-Tucker condition and can be written as

$$\gamma \geq 0, \quad f \leq 0, \quad \gamma f = 0. \quad (4.94)$$

It is possible to view the nonpositive property of the yield function as a constraint, and the plastic consistency parameter γ can be seen as the Lagrange multiplier corresponding to the inequality constraint. The above Kuhn-Tucker condition satisfies all possible states of the material. For example, when the material is in the elastic state, i.e., the stress is within the elastic domain, it becomes

$$f < 0, \quad \gamma = 0 \Rightarrow \gamma f = 0.$$

When the stress is on the yield surface, i.e., in the plastic state, Eq. (4.94) is satisfied because $f=0$. However, when the state varies, it is possible to have three different cases:

- (a) Elastic unloading $\dot{f} < 0, \gamma = 0 \Rightarrow \gamma\dot{f} = 0$.
- (b) Neutral loading $\dot{f} = 0, \gamma = 0, \Rightarrow \gamma\dot{f} = 0$.
- (c) Plastic loading $\dot{f} = 0, \gamma > 0, \Rightarrow \gamma\dot{f} = 0$.

Thus, $\gamma f = 0$ in Eq. (4.94) is equivalent to $\gamma\dot{f} = 0$ when the stress is on the yield surface. Therefore, the rate form of Kuhn-Tucker condition can be used in calculating the plastic consistency parameter. Among three possible cases, only the last case, plastic loading, is of interest because the remaining cases can be identified with $\gamma = 0$, which would not have any change in the plastic variables. Thus, when the plastic loading state continues,

$$\gamma > 0 \quad \dot{f}(\boldsymbol{\sigma}, \boldsymbol{\xi}) = 0, \quad (4.95)$$

which means that the yield function remains constant during the plastic loading state. From the requirement that the yield surface remains 0, the following condition can be obtained:

$$\dot{f}(\boldsymbol{\sigma}, \boldsymbol{\xi}) = \frac{\partial f}{\partial \boldsymbol{\sigma}} : \dot{\boldsymbol{\sigma}} + \frac{\partial f}{\partial \boldsymbol{\xi}} : \dot{\boldsymbol{\xi}} = 0.$$

By substituting the rates of stress and plastic variables, the above equation can be written as

$$\frac{\partial f}{\partial \boldsymbol{\sigma}} : \mathbf{D} : (\dot{\boldsymbol{\epsilon}} - \dot{\boldsymbol{\epsilon}}^p) + \frac{\partial f}{\partial \boldsymbol{\xi}} \cdot \gamma \mathbf{h} = 0.$$

Since the rate of plastic strain can also be written in terms of the plastic consistency parameter, the above equation can be rewritten as

$$\frac{\partial f}{\partial \boldsymbol{\sigma}} : \mathbf{D} : \dot{\boldsymbol{\epsilon}} - \frac{\partial f}{\partial \boldsymbol{\sigma}} : \mathbf{D} : \gamma \mathbf{r} + \frac{\partial f}{\partial \boldsymbol{\xi}} \cdot \gamma \mathbf{h} = 0.$$

The above equation can be solved for the plastic consistency parameter as

$$\gamma = \frac{\left\langle \frac{\partial f}{\partial \boldsymbol{\sigma}} : \mathbf{D} : \dot{\boldsymbol{\epsilon}} \right\rangle}{\frac{\partial f}{\partial \boldsymbol{\sigma}} : \mathbf{D} : \mathbf{r} - \frac{\partial f}{\partial \boldsymbol{\xi}} \cdot \mathbf{h}}, \quad (4.96)$$

where $\langle x \rangle$ is equal to x if $x > 0$; otherwise, it is 0. From the requirement of $\gamma > 0$, the numerator in the above equation must be nonnegative. The physical meaning of this condition is that the normal direction to the yield surface and the stress

increment rate must have an acute angle when the material is under plastic loading (see Fig. 4.15):

$$\cos \theta = \frac{\frac{\partial f}{\partial \sigma} : \mathbf{D} : \dot{\epsilon}}{\left\| \frac{\partial f}{\partial \sigma} \right\| \left\| \mathbf{D} : \dot{\epsilon} \right\|}. \quad (4.97)$$

If $\theta < 90^\circ$, the material is in plastic loading; if $\theta = 0^\circ$, it is in neutral loading; and if $\theta > 90^\circ$, it is in elastic unloading.

6. Elastoplastic Tangent Stiffness

In one-dimensional systems, elastoplastic modulus D^{ep} was calculated based on elastic and plastic moduli. In multidimensional systems, the counterpart is called the continuum elastoplastic tangent stiffness. It represents the relation between the rates of stress and strain. By substituting the plastic consistency parameter into Eq. (4.81),

$$\dot{\sigma} = \mathbf{D} : \dot{\epsilon} - \mathbf{D} : \gamma \mathbf{r} = \mathbf{D} : \dot{\epsilon} - \mathbf{D} : \mathbf{r} \frac{\left\langle \frac{\partial f}{\partial \sigma} : \mathbf{D} : \dot{\epsilon} \right\rangle}{\frac{\partial f}{\partial \sigma} : \mathbf{D} : \mathbf{r} - \frac{\partial f}{\partial \xi} \cdot \mathbf{h}}.$$

The above equation can be rewritten in terms of stress and strain rates, as

$$\dot{\sigma} = \left[\mathbf{D} - \frac{\left\langle \mathbf{D} : \mathbf{r} \otimes \frac{\partial f}{\partial \sigma} : \mathbf{D} \right\rangle}{\frac{\partial f}{\partial \sigma} : \mathbf{D} : \mathbf{r} - \frac{\partial f}{\partial \xi} \cdot \mathbf{h}} \right] : \dot{\epsilon} \equiv \mathbf{D}^{ep} : \dot{\epsilon}, \quad (4.98)$$

where \mathbf{D}^{ep} is the *continuum elastoplastic tangent stiffness*. In general, \mathbf{D}^{ep} is not symmetric. However, when the associative flow rule is used, i.e., $\mathbf{r} = \partial f / \partial \sigma$, it becomes symmetric. The explicit expression of \mathbf{D}^{ep} can be obtained when the flow rule and hardening model are specified. Since a part of the strain rate is plastic and therefore does not increase stress, the elastic stiffness \mathbf{D} is reduced by the plastic consistency parameter. A similar observation can be made for the case of the one-dimensional system in Eq. (4.8).

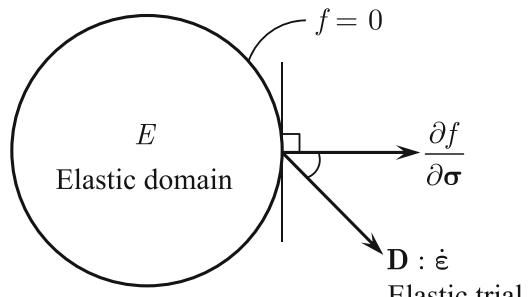


Fig. 4.15 Angle between elastic trial stress and normal to the yield surface

In order to be a stable material, the rate of work due to stress rate must be positive, i.e., $\dot{\sigma} : \dot{\epsilon} > 0$. Equation (4.98) implies that the elastoplastic tangent stiffness \mathbf{D}^{ep} must be positive definite for a stable material. In addition, in order to have a stable hardening behavior, the rate of work during the plastic deformation must be positive, i.e., $\dot{\sigma} : \dot{\epsilon}^p \geq 0$. These two conditions are called Drucker's postulate.

Example 4.11 (Plastic Consistency Parameter and Elastoplastic Tangent Stiffness) Consider the following combined linear isotropic/kinematic hardening model with associative flow rule:

$$\begin{aligned}\kappa(e_p) &= \sigma_Y^0 + (1 - \beta)He_p \\ \dot{\alpha} &= \frac{2}{3}\beta H\dot{\epsilon}^p\end{aligned}. \quad (4.99)$$

Using five material parameters ($\lambda, \mu, \beta, H, \sigma_Y^0$) and the current value of stress (σ) and plastic variables (α, e_p), calculate the plastic consistency parameters, γ , and elastoplastic tangent stiffness, \mathbf{D}_{ep} .

Solution The plastic consistency parameter can be calculated from the rate of change of yield function as

$$\begin{aligned}f(\mathbf{s}, \alpha, e_p) &= \|\mathbf{s} - \alpha\| - \sqrt{\frac{2}{3}}[\sigma_Y^0 + (1 - \beta)He_p] = 0, \\ \dot{f} &= \frac{\partial f}{\partial \mathbf{s}} : \dot{\mathbf{s}} + \frac{\partial f}{\partial \alpha} : \dot{\alpha} + \frac{\partial f}{\partial e_p} \dot{e}_p = \mathbf{N} : \dot{\mathbf{s}} - \mathbf{N} : \dot{\alpha} - \sqrt{\frac{2}{3}}(1 - \beta)H\dot{e}_p = 0.\end{aligned}$$

The purpose is to write the above equation in terms of the plastic consistency parameter. The deviatoric stress rate, back stress rate, and effective plastic strain can be written in terms of the plastic consistency parameter, as

$$\begin{aligned}\dot{\mathbf{s}} &= 2\mu(\dot{\mathbf{e}} - \dot{\mathbf{e}}^p) = 2\mu\dot{\mathbf{e}} - 2\mu\gamma\mathbf{N} \\ \dot{\alpha} &= \frac{2}{3}\beta H\dot{\epsilon}^p = \frac{2}{3}\beta H\gamma\mathbf{N} \\ \dot{e}_p &= \sqrt{\frac{2}{3}}\gamma\end{aligned}.$$

By substituting these relations in the rate of yield function,

$$\dot{f} = 2\mu\mathbf{N} : \dot{\mathbf{e}} - 2\mu\gamma\mathbf{N} : \mathbf{N} - \frac{2}{3}\beta H\gamma\mathbf{N} : \mathbf{N} - \frac{2}{3}(1 - \beta)H\gamma = 0.$$

Note that $\mathbf{N} : \mathbf{N} = 1$ and $\mathbf{N} : \dot{\mathbf{e}} = \mathbf{N} : \dot{\mathbf{e}}$ are used. The above equation is linear with respect to the plastic consistency parameter and can be solved for it, to yield

$$\gamma = \frac{2\mu \mathbf{N} : \dot{\mathbf{e}}}{2\mu + \frac{2}{3}H}. \quad (4.100)$$

For elastoplastic tangent stiffness, consider the following constitutive relation in a rate form:

$$\dot{\sigma} = \mathbf{D} : \dot{\epsilon} - \mathbf{D} : \dot{\epsilon}^p = \mathbf{D} : \dot{\epsilon} - \dot{\gamma} \mathbf{D} : \mathbf{N},$$

where $\mathbf{D} = (\lambda + (2/3)\mu)\mathbf{1} \otimes \mathbf{1} + 2\mu\mathbf{I}_{\text{dev}}$. Since \mathbf{N} is a unit deviatoric tensor, $\mathbf{D} : \mathbf{N} = 2\mu\mathbf{N}$. Thus,

$$\dot{\sigma} = \mathbf{D} : \dot{\epsilon} - 2\mu\mathbf{N} \frac{2\mu\mathbf{N} : \dot{\epsilon}}{2\mu + \frac{2}{3}H} = \left[\mathbf{D} - \frac{4\mu^2}{2\mu + \frac{2}{3}H} \mathbf{N} \otimes \mathbf{N} \right] : \dot{\epsilon}.$$

Thus, the elastoplastic tangent stiffness can be obtained by

$$\mathbf{D}^{\text{ep}} = \mathbf{D} - \frac{4\mu^2}{2\mu + \frac{2}{3}H} \mathbf{N} \otimes \mathbf{N}. \quad (4.101)$$

Note that the first term on the right-hand side of the expression of elastoplastic tangent stiffness \mathbf{D}^{ep} is the elastic constitutive tensor \mathbf{D} in Eq. (4.81). Thus, the effect of plastic deformation appears in the second term through the plastic modulus H and unit deviatoric tensor \mathbf{N} , which depends on the current stress and back stress. ■

Example 4.12 (Plastic Deformation of a Bar) Consider a bar under a uniaxial tension load. At load step t_n , the axial stress $\sigma_{11} = 300$ MPa, and the material is purely elastic before t_n . At load step t_{n+1} , a strain increment $\Delta\epsilon_{11} = \dot{\epsilon}\Delta t = 0.1$ is given, determine stress and plastic variables. The material is combined linear isotropic/kinematic hardening, and the material parameters are given in Table 4.2.

Solution Although it is a uniaxial problem, a three-dimensional stress state will be considered. At load step t_n , the stress and its deviator can be written as

$$\sigma = \begin{bmatrix} 300 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ MPa}, \quad \mathbf{s} = \begin{bmatrix} 200 & 0 & 0 \\ 0 & -100 & 0 \\ 0 & 0 & -100 \end{bmatrix} \text{ MPa}.$$

Table 4.2 Elastoplastic material parameters

E	μ	ν	σ_Y	H	β
2.4 GPa	1.0 GPa	0.2	300 MPa	100 MPa	0.3

Due to the effect of Poisson's ratio, the incremental strain and its deviator can be written as

$$\Delta\boldsymbol{\epsilon} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & -0.02 & 0 \\ 0 & 0 & -0.02 \end{bmatrix}, \quad \Delta\boldsymbol{\epsilon} = \begin{bmatrix} 0.08 & 0 & 0 \\ 0 & -0.04 & 0 \\ 0 & 0 & -0.04 \end{bmatrix}.$$

Since the material is purely elastic, the plastic variables are all 0, i.e., ${}^n\boldsymbol{\alpha} = \mathbf{0}$, ${}^n\boldsymbol{e}_p = \mathbf{0}$. Accordingly, ${}^n\boldsymbol{\eta} = {}^n\boldsymbol{s} - {}^n\boldsymbol{\alpha} = {}^n\boldsymbol{s}$. The trial state can be obtained by assuming the strain increment is elastic:

$${}^{tr}\boldsymbol{\eta} = {}^{tr}\boldsymbol{s} = {}^n\boldsymbol{s} + 2\mu\Delta\boldsymbol{\epsilon} = \begin{bmatrix} 360 & 0 & 0 \\ 0 & -180 & 0 \\ 0 & 0 & -180 \end{bmatrix} \text{ MPa.}$$

In order to calculate the unit deviatoric tensor \mathbf{N} , the trial shifted stress is normalized

$$\|{}^{tr}\boldsymbol{\eta}\| = \sqrt{360^2 + 180^2 + 180^2} = 180\sqrt{6} \text{ MPa,}$$

$$\mathbf{N} = \frac{{}^{tr}\boldsymbol{\eta}}{\|{}^{tr}\boldsymbol{\eta}\|} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

It is necessary to check if the trial state satisfies the yield criterion or not. The yield function with the trial state becomes

$$f({}^{tr}\boldsymbol{\eta}, {}^{tr}\boldsymbol{e}_p) = \|{}^{tr}\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}\kappa({}^n\boldsymbol{e}_p) = 180\sqrt{6} - 300\sqrt{\frac{2}{3}} = 80\sqrt{6} > 0.$$

Thus, the trial state stays outside the yield surface, and the material will go through plastic deformation. The plastic consistency parameter can be obtained by

$$\gamma = \frac{2\mu\mathbf{N} : \Delta\boldsymbol{\epsilon}}{2\mu + \frac{2}{3}H} = 0.0948.$$

Using the plastic consistency parameter, the stress and plastic variables are updated by

$${}^{n+1}\boldsymbol{\sigma} = {}^n\boldsymbol{\sigma} + \mathbf{D} : \Delta\boldsymbol{\epsilon} - 2\mu\gamma\mathbf{N} = \begin{bmatrix} 385.2 & 0 & 0 \\ 0 & 77.4 & 0 \\ 0 & 0 & 77.4 \end{bmatrix} \text{ MPa,}$$

$${}^{n+1}\boldsymbol{\alpha} = {}^n\boldsymbol{\alpha} + \frac{2}{3}\beta H\gamma \mathbf{N} = \begin{bmatrix} 1.54 & 0 & 0 \\ 0 & -0.77 & 0 \\ 0 & 0 & -0.77 \end{bmatrix} \text{ MPa},$$

$${}^{n+1}e_p = {}^n e_p + \sqrt{\frac{2}{3}}\gamma = 0.0774.$$

Note that the stress is not uniaxial anymore. Due to plastic deformation, both σ_{22} and σ_{33} exist. ■

4.3.5 Numerical Integration

Since constitutive relations and evolution of plastic variables are in the form of rates in the elastoplastic model, they need to be integrated over time (or load) increments. In static problems, the time increment should be understood as a load increment (refer to Chap. 3). The full magnitude of load is first divided by N increments, and the structural equilibrium at each increment is sought with the incremental force method. It is assumed that the solutions and the status of material at time t_n are known, which includes stress and plastic variables. Then, at time t_{n+1} , the Newton–Raphson method solves for the incremental displacements during the convergence iteration. Thus, the objective is to update stress and plastic variables from time t_n to t_{n+1} using the given displacement increments or equivalently using the given strain increments.

Note that the structural equilibrium is only satisfied at the discrete set of time increments. Thus, it is possible that there might be discretization error in time, especially when the status of material changes within a time increment. If smaller size of time increment is used, the error will be reduced. This is different from the nonlinear elastic systems in which the size of time increment is determined in order to help convergence. In the case of elastoplastic systems, it may affect the accuracy of analysis.

Although there are many integration methods for solving differential equations, it is important that the method should provide accurate and robust results. The backward Euler time integration method, which will be used in the following derivations, has been popular because it is simple and provides unconditional stability. It is well known that the return-mapping algorithm, with the radial return method as a special case, is an effective and robust method for plasticity [1]. Thus, time integration of elastoplasticity model using return-mapping algorithm will be discussed in this section. In the return-mapping algorithm, a two-step method is often used. First, the elastic trial status is computed in which all strain increments are purely elastic. If the trial stress resides outside the elastic domain, then the trial stress is projected onto the yield surface, which is a convex set. This step is called the return-mapping to the yield surface. During the return-mapping step, the yield surface itself changes due to the evolution of plastic variables (strain-hardening). Thus, it is challenging to find the return-mapping point on the yield surface, while the radius and center location of the yield surface are changing.

1. Return-Mapping Algorithm

For associative plasticity, it is well known that the backward Euler method produces the closest point projection. Since the displacement increment at time t_{n+1} is known, the strain increment at time t_{n+1} can be computed from the definition of strain. The first step is called the elastic predictor and uses this incremental strain. The stress and hardening parameters are predicted elastically as

$${}^{\text{tr}}\mathbf{s} = {}^n\mathbf{s} + 2\mu\Delta\mathbf{e}, \quad {}^{\text{tr}}\boldsymbol{\alpha} = {}^n\boldsymbol{\alpha}, \quad {}^{\text{tr}}e_p = {}^n e_p, \quad (4.102)$$

$${}^{\text{tr}}\boldsymbol{\eta} = {}^{\text{tr}}\mathbf{s} - {}^{\text{tr}}\boldsymbol{\alpha}, \quad (4.103)$$

where the left superscript n denotes the time t_n , and “tr” denotes the trial status. In the elastic predictor step, all strain increments are considered to be elastic and thus, all plastic variables are fixed. Thus, there is no change in plastic variables. Although both volumetric and deviatoric parts of stress change, only the change in deviatoric stress is considered above because the hydrostatic stress does not affect plasticity.

If the trial stress ${}^{\text{tr}}\boldsymbol{\eta}$ is within the elastic domain, i.e., $f({}^{\text{tr}}\boldsymbol{\eta}, {}^{\text{tr}}e_p) \leq 0$, then the status of the material is elastic, and the stress and plastic variables are updated using the trial predictors as

$${}^{n+1}\mathbf{s} = {}^{\text{tr}}\mathbf{s}, \quad {}^{n+1}\boldsymbol{\alpha} = {}^{\text{tr}}\boldsymbol{\alpha}, \quad {}^{n+1}e_p = {}^{\text{tr}}e_p. \quad (4.104)$$

This is considered as the end of time integration when the status of the material is elastic.

If the trial stress ${}^{\text{tr}}\boldsymbol{\eta}$ is outside the elastic domain, i.e., $f({}^{\text{tr}}\boldsymbol{\eta}, {}^{\text{tr}}e_p) > 0$, then the status of the material becomes plastic, and the plastic correction step needs to be carried out to find the plastic status of the material. The stress and plastic variables are corrected by considering plastic deformation. Figure 4.16 illustrates the process of elastic prediction and plastic correction steps. First, because the plastic strain does not contribute to the stress, the trial stress is reduced proportional to the plastic strain increment as

$${}^{n+1}\mathbf{s} = {}^{\text{tr}}\mathbf{s} - 2\mu\Delta\mathbf{e}^p = {}^{\text{tr}}\mathbf{s} - 2\mu\hat{\gamma}\mathbf{N}. \quad (4.105)$$

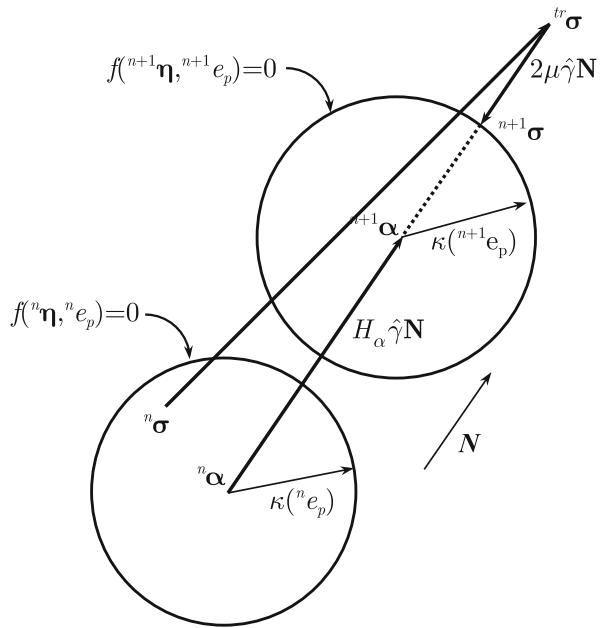
The plastic variables are also updated simultaneously with the stress, according to the flow rule as

$${}^{n+1}\boldsymbol{\alpha} = {}^{\text{tr}}\boldsymbol{\alpha} + H_\alpha\hat{\gamma}\mathbf{N}, \quad (4.106)$$

$${}^{n+1}e_p = {}^n e_p + \sqrt{\frac{2}{3}}\hat{\gamma}, \quad (4.107)$$

where $\hat{\gamma} = \gamma\Delta t$ is the plastic consistency parameter, and $\mathbf{N} = {}^{n+1}\boldsymbol{\eta}/\|{}^{n+1}\boldsymbol{\eta}\|$ is a unit deviatoric tensor, normal to the yield surface at time t_{n+1} . Note that stress and back stress are corrected in the parallel direction to \mathbf{N} ; the trial stress reduces,

Fig. 4.16 Return-mapping of isotropic elastoplasticity



while the back stress increases. This fact makes it convenient for finding the updated stress on the updated yield surface.

Note that the plastic strain increment $\Delta\epsilon^P$, or equivalently, $\hat{\gamma}N$ is unknown yet. In order to simplify the following calculations, consider the shifted stress at time t_{n+1} :

$${}^{n+1}\boldsymbol{\eta} = {}^{n+1}\mathbf{s} - {}^{n+1}\boldsymbol{\alpha} = {}^{tr}\boldsymbol{\eta} - (2\mu + H_\alpha)\hat{\gamma}\mathbf{N}. \quad (4.108)$$

Since ${}^{n+1}\boldsymbol{\eta}$ is parallel to \mathbf{N} , ${}^{tr}\boldsymbol{\eta}$ must also be parallel to \mathbf{N} , which means that the final updated stress moves in the same direction as the trial stress. Thus, the unit normal tensor to the yield surface can be computed from the trial stress by

$$\mathbf{N} = \frac{{}^{tr}\boldsymbol{\eta}}{\|{}^{tr}\boldsymbol{\eta}\|}, \quad (4.109)$$

which is known from the elastic predictor step. Thus, the plastic correction step condenses to determine the plastic consistency parameter $\hat{\gamma}$, from which the plastic strain increment can be obtained. The basic idea is to make the yield

function satisfy the yield condition at the updated state. Thus, at the return-mapped point, the following yield condition must be satisfied:

$$\begin{aligned} f^{(n+1)} \mathbf{\eta}, {}^{n+1} e_p) &= ||^{n+1} \mathbf{\eta}|| - \sqrt{\frac{2}{3}} \kappa({}^{n+1} e_p) \\ &= ||^{\text{tr}} \mathbf{\eta}|| - (2\mu + H_\alpha({}^{n+1} e_p)) \hat{\gamma} - \sqrt{\frac{2}{3}} \kappa({}^{n+1} e_p) = 0, \end{aligned} \quad (4.110)$$

which is a nonlinear equation in terms of $\hat{\gamma}$. Equation (4.110) can be solved for $\hat{\gamma}$ using the local Newton–Raphson method. Note that this is different from the Newton–Raphson method for the convergence iteration in which equilibrium between internal and external forces is sought (e.g., see Eq. (4.19)). This is a local iteration to find the stress point on the yield surface. Thus, the solution procedure for elastoplastic systems has double iteration loops. The inside, local iteration loop usually converges quickly within five or six iterations, but it needs to be performed at every integration point that has plastic deformation. Below is the flowchart for the local Newton–Raphson iteration to find the stress point on the yield surface:

1. Initialize variables

$$k = 0, \quad e_p^k = {}^n e_p, \quad \gamma^k = 0, \quad f_{\text{TOL}} = \sigma_Y^0 \times 10^{-7}, \quad k_{\text{MAX}} = 20.$$

2. Yield function

$$f^k = ||^{\text{tr}} \mathbf{\eta}|| - (2\mu + H_\alpha(e_p^k)) \gamma^k - \sqrt{\frac{2}{3}} \kappa(e_p^k).$$

3. Jacobian relation

$$\frac{\partial f}{\partial \gamma} = 2\mu + H_\alpha + \sqrt{\frac{2}{3}} H_{\alpha, e_p} \gamma^k + \frac{2}{3} \kappa_{, e_p}.$$

4. Update the plastic consistency parameter and effective plastic strain

$$\gamma^{k+1} = \gamma^k + \frac{f^k}{\partial f / \partial \gamma}, \quad e_p^{k+1} = {}^n e_p + \sqrt{\frac{2}{3}} \gamma^{k+1}.$$

5. Check convergence

If $(|f^k| > f_{\text{TOL}})$ $k = k + 1$ and go to Step 2.

If $(k > k_{\text{MAX}})$ stop with error message.

If isotropic/kinematic hardening is a linear function of $\hat{\gamma}$, or of the effective plastic strain, then only one iteration is required to compute the return map point because the equation becomes linear.

2. Updating Stress and Plastic Variables

After $\hat{\gamma}$ is found, the deviatoric stress can be updated at time t_{n+1} by

$${}^{n+1}\mathbf{s} = {}^n\mathbf{s} + 2\mu\Delta\mathbf{e} - 2\mu\hat{\gamma}\mathbf{N}. \quad (4.111)$$

However, once the return-mapped point is found, the deviatoric stress is no longer necessary. Rather, the stress itself can be updated by

$${}^{n+1}\boldsymbol{\sigma} = {}^n\boldsymbol{\sigma} + \Delta\boldsymbol{\sigma}, \quad (4.112)$$

where the stress increment can be calculated from

$$\Delta\boldsymbol{\sigma} = \mathbf{D} : \Delta\mathbf{e} - 2\mu\hat{\gamma}\mathbf{N}. \quad (4.113)$$

In addition, the back stress and the effective plastic strain are updated by

$${}^{n+1}\boldsymbol{\alpha} = {}^n\boldsymbol{\alpha} + H_{\boldsymbol{\alpha}}\hat{\gamma}\mathbf{N}, \quad (4.114)$$

$${}^{n+1}e_p = {}^n e_p + \sqrt{\frac{2}{3}}\hat{\gamma}. \quad (4.115)$$

Note that the stress and back stress increments corresponding to the plastic correction component in Eqs. (4.113) and (4.114) are in the same direction as \mathbf{N} , which is a radial direction of the yield surface, as shown in Fig. 4.15.

Example 4.13 (Plastic Consistency Parameter) Consider the combined linear isotropic/kinematic hardening model in Example 4.11. Calculate the plastic consistency parameter $\hat{\gamma}$ during time integration, and compare it with the rate form γ in Eq. (4.100).

Solution In the case of linear hardening, the yield criterion in Eq. (4.110) becomes a linear function of the plastic consistency parameter. For the given combined linear isotropic/kinematic hardening model, the yield function at t_{n+1} can be written as

$$f({}^{n+1}\mathbf{n}, {}^{n+1}e_p) = \left\| {}^{tr}\mathbf{n} - \left(2\mu + \frac{2}{3}\beta H\right)\hat{\gamma}\mathbf{N}\right\| - \sqrt{\frac{2}{3}}k({}^n e_p) - \frac{2}{3}(1-\beta)H\hat{\gamma} = 0. \quad (4.116)$$

Since ${}^{tr}\mathbf{n}$ and \mathbf{N} are parallel, the first term on the right-hand side (norm) can be split into two individual norms, to yield

$$\left\| {}^{tr}\mathbf{n} - \left(2\mu + \frac{2}{3}\beta H\right)\hat{\gamma}\mathbf{N}\right\| = \left\| {}^{tr}\mathbf{n} \right\| - \left(2\mu + \frac{2}{3}\beta H\right)\hat{\gamma}.$$

Then, it is clear that Eq. (4.116) is a linear function of the plastic consistency parameter, which can be solved using

$$\hat{\gamma} = \frac{\|{}^t\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}\kappa({}^n\boldsymbol{e}_p)}{2\mu + \frac{2}{3}H}. \quad (4.117)$$

The above equation provides a convenient implementation strategy for checking plastic loading and calculating the plastic consistency parameter. The return-mapping step starts when the trial stress stays outside the yield surface, as

$$f({}^t\boldsymbol{\eta}, {}^t\boldsymbol{e}_p) = \|{}^t\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}\kappa({}^n\boldsymbol{e}_p) > 0.$$

In fact, this trial yield function is identical to the numerator in Eq. (4.117). Thus, the plastic consistency parameter can be calculated by

$$\hat{\gamma} = \frac{f({}^t\boldsymbol{\eta}, {}^t\boldsymbol{e}_p)}{2\mu + \frac{2}{3}H}.$$

It is noted that both the rate form γ and the incremental form $\hat{\gamma}$ have the same denominator. In order to compare the numerators, the trial yield function is explicitly written in terms of stress and plastic variables:

$$f({}^t\boldsymbol{\eta}, {}^t\boldsymbol{e}_p) = \|{}^n\boldsymbol{\eta} + 2\mu\Delta\boldsymbol{e}\| - \sqrt{\frac{2}{3}}\kappa({}^n\boldsymbol{e}_p) > 0.$$

The physical meaning of $f({}^t\boldsymbol{\eta}, {}^t\boldsymbol{e}_p)$ is the radial distance from the yield surface to ${}^t\boldsymbol{s}$ in stress space, as shown in Fig. 4.17. In order to plot the numerator in Eq. (4.100), $2\mu\mathbf{N} : \dot{\boldsymbol{\epsilon}}$, it is converted to the incremental counterpart as $2\mu\mathbf{N} : \Delta\boldsymbol{e}$. Since \mathbf{N} is a unit deviatoric tensor, $\mathbf{N} : \Delta\boldsymbol{\epsilon} = \mathbf{N} : \Delta\boldsymbol{e}$, which is a projection of $\Delta\boldsymbol{e}$ to \mathbf{N} . As shown in Fig. 4.17, the two formulations become equivalent when (a) the material is in the plastic state at t_n , and (b) $\Delta\boldsymbol{e}$ is parallel to the ${}^n\boldsymbol{\eta}$. In general, these two requirements are satisfied when the size of time increment is very small. ■

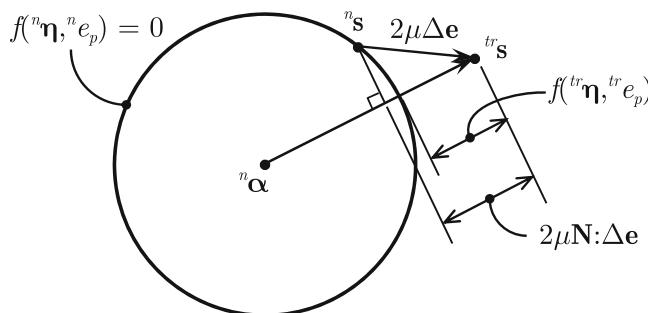


Fig. 4.17 Difference in the plastic consistency parameters from rate form and incremental form

Example 4.14 (Plastic Deformation of a Bar) Determine stress and plastic variables in Example 4.12 using incremental time integration. The axial incremental strain is given as $\Delta e_{11} = 0.1$. Assume the same material properties and hardening parameters with that of Example 4.12.

Solution The two formulations, rate and incremental forms, are identical until the determination of plastic consistency parameter, which can be obtained using the trial yield function in the incremental integration by

$$\gamma = \frac{f({}^t\boldsymbol{\eta}, {}^t\boldsymbol{e}_p)}{2\mu + \frac{2}{3}H} = 0.0948.$$

Note that the above plastic consistency parameter is identical with that of Example 4.12. This happens because (a) the material was initially plastic and (b) $\Delta\boldsymbol{e}$ is parallel to the ${}^n\boldsymbol{\eta}$. Thus, the updated stress and plastic variables are supposed to be identical, too:

$$\begin{aligned} {}^{n+1}\boldsymbol{\sigma} &= {}^n\boldsymbol{\sigma} + \mathbf{D} : \Delta\boldsymbol{\epsilon} - 2\mu\gamma\mathbf{N} = \begin{bmatrix} 385.2 & 0 & 0 \\ 0 & 77.4 & 0 \\ 0 & 0 & 77.4 \end{bmatrix} \text{ MPa,} \\ {}^{n+1}\boldsymbol{\alpha} &= {}^n\boldsymbol{\alpha} + \frac{2}{3}\beta H\gamma\mathbf{N} = \begin{bmatrix} 1.54 & 0 & 0 \\ 0 & -0.77 & 0 \\ 0 & 0 & -0.77 \end{bmatrix} \text{ MPa,} \\ {}^{n+1}\boldsymbol{e}_p &= {}^n\boldsymbol{e}_p + \sqrt{\frac{2}{3}}\gamma = 0.0774. \end{aligned}$$

3. Consistent Tangent Stiffness

As discussed in Chap. 2, if the Jacobian matrix (or tangent stiffness here) is accurate, the Newton–Raphson method shows a quadratic convergence. In structural analysis, since the residual force is related to stress, the Jacobian matrix requires the derivative of stress with respect to strain, which is called the tangent stiffness. The continuum elastoplastic tangent stiffness, \mathbf{D}^{ep} , in Eq. (4.98) can be used to this purpose, but numerical tests show that the Newton–Raphson iteration does not show a quadratic convergence when \mathbf{D}^{ep} is used. Simo and Taylor [2] showed that this happens because \mathbf{D}^{ep} is not consistent with the time integration algorithm. \mathbf{D}^{ep} is tangent stiffness between stress and strain rates, while the time integration algorithm uses a finite size of time increment. The tangent stiffness must be consistent with the time integration algorithm to achieve quadratic convergence during Newton–Raphson iteration. In this section, the tangent stiffness that is consistent with the time integration algorithm is derived. The incremental stress in Eq. (4.113) is differentiated with respect to the incremental strain, which produces a consistent constitutive relation with the return-mapping algorithm as

$$\mathbf{D}^{\text{alg}} = \frac{\partial \Delta \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}} = \mathbf{D} - 2\mu \mathbf{N} \otimes \frac{\partial \hat{\gamma}}{\partial \Delta \boldsymbol{\epsilon}} - 2\mu \hat{\gamma} \frac{\partial \mathbf{N}}{\partial \Delta \boldsymbol{\epsilon}}, \quad (4.118)$$

where \mathbf{D}^{alg} stands for consistent (or algorithmic) tangent stiffness. The above relation requires derivatives of the plastic consistency parameter and unit deviatoric tensor.

Since the consistency condition in Eq. (4.110) must be satisfied for all strain states between load steps t_n and t_{n+1} , the differential of f with respect to $\Delta \boldsymbol{\epsilon}$ must vanish, from which the relation between $\hat{\gamma}$ and $\Delta \boldsymbol{\epsilon}$ can be obtained. In order to differentiate Eq. (4.110), the following relation can help:

$$\frac{\partial \|\mathbf{n}^{\text{tr}}\|}{\partial \Delta \boldsymbol{\epsilon}} = 2\mu \frac{\mathbf{n}^{\text{tr}}}{\|\mathbf{n}^{\text{tr}}\|} = 2\mu \mathbf{N},$$

$$\frac{\partial H_\alpha(n+1 e_p)}{\partial \Delta \boldsymbol{\epsilon}} = \sqrt{\frac{2}{3}} \frac{\partial H_\alpha}{\partial e_p} \frac{\partial \hat{\gamma}}{\partial \Delta \boldsymbol{\epsilon}}.$$

In the derivation of the above equation, the property that \mathbf{N} is a deviatoric tensor is used. In addition, the relation of $\Delta e_p = \sqrt{(2/3)}\hat{\gamma}$ is used. The yield function in Eq. (4.110) can then be differentiated with respect to the strain increment to obtain

$$\frac{\partial f}{\partial \Delta \boldsymbol{\epsilon}} = 2\mu \mathbf{N} - \left(2\mu + H_\alpha + \sqrt{\frac{2}{3}} H_{\alpha, e_p} \hat{\gamma} + \frac{2}{3} \kappa_{e_p} \right) \frac{\partial \hat{\gamma}}{\partial \Delta \boldsymbol{\epsilon}} = 0,$$

where $H_{\alpha, e_p} = \partial H_\alpha / \partial e_p$ and $\kappa_{e_p} = \partial \kappa / \partial e_p$. Thus, the derivative of the plastic consistency parameter with respect to the strain increment can be obtained as

$$\frac{\partial \hat{\gamma}}{\partial \Delta \boldsymbol{\epsilon}} = \frac{2\mu \mathbf{N}}{(2\mu + H_\alpha + \sqrt{\frac{2}{3}} H_{\alpha, e_p} \hat{\gamma} + \frac{2}{3} \kappa_{e_p})}. \quad (4.119)$$

Next, the increment of the unit normal tensor to the yield function can also be expressed as

$$\begin{aligned} \frac{\partial \mathbf{N}}{\partial \Delta \boldsymbol{\epsilon}} &= \frac{\partial N}{\partial \mathbf{n}^{\text{tr}}} : \frac{\partial \mathbf{n}^{\text{tr}}}{\partial \Delta \boldsymbol{\epsilon}} \\ &= \left[\frac{\mathbf{I}}{\|\mathbf{n}^{\text{tr}}\|} - \frac{\mathbf{n}^{\text{tr}} \otimes \mathbf{n}^{\text{tr}}}{\|\mathbf{n}^{\text{tr}}\|^3} \right] : 2\mu \mathbf{I}_{\text{dev}}. \\ &= \frac{2\mu}{\|\mathbf{n}^{\text{tr}}\|} [\mathbf{I}_{\text{dev}} - \mathbf{N} \otimes \mathbf{N}] \end{aligned}$$

Thus, from Eq. (4.113) the consistent or algorithmic tangent stiffness becomes

$$\mathbf{D}^{\text{alg}} = \frac{\partial \Delta \boldsymbol{\sigma}}{\partial \Delta \boldsymbol{\epsilon}} = \mathbf{D} - \frac{4\mu^2 \mathbf{N} \otimes \mathbf{N}}{2\mu + H_\alpha + \sqrt{\frac{2}{3}H_{\alpha,e_p}\hat{\gamma} + \frac{2}{3}\kappa_{e_p}}} - \frac{4\mu^2 \hat{\gamma}}{\|\mathbf{I}^{\text{tr}} \mathbf{q}\|} [\mathbf{I}_{\text{dev}} - \mathbf{N} \otimes \mathbf{N}]. \quad (4.120)$$

It is interesting to compare the above tangent stiffness with the continuum elastoplastic tangent stiffness in Eq. (4.98). Since Eq. (4.98) is written in terms of a general hardening model, it is simplified for the case of the isotropic/kinematic hardening model, as

$$\mathbf{D}^{\text{ep}} = \mathbf{D} - \frac{4\mu^2 \mathbf{N} \otimes \mathbf{N}}{2\mu + H_\alpha + \frac{2}{3}\kappa_{e_p}}. \quad (4.121)$$

By comparing the two equations, it is clear that \mathbf{D}^{ep} does not have the third term as in \mathbf{D}^{alg} , which represents the effect of change of \mathbf{N} due to the strain increment. Since the rate form only considers infinitesimal strain increment (strain rate), it does not take into account the change in direction. However, when the strain increment is not small, it may change the direction of shifted stress, and thus, \mathbf{N} . This effect did not appear in the 1D elastoplasticity model in Sect. 4.2 because a scalar stress is used with fixed \mathbf{N} . The other difference is the denominator of the second term on the right-hand side. They are similar, but \mathbf{D}^{ep} does not include the nonlinear hardening effect. Thus, they become identical when the hardening is linear. This happens because, in a sense, the rate form differentiates the hardening model first and then takes increments, while the incremental form differentiates after taking increments. Note that the two tangent stiffnesses become identical when $\hat{\gamma} = 0$.

4. Incremental Equations for Elastoplasticity

For notational convenience, the energy form and its linearization are defined as

$$a(\xi; \mathbf{u}, \bar{\mathbf{u}}) \equiv \iint_{\Omega} \boldsymbol{\epsilon}(\bar{\mathbf{u}}) : \mathbf{N} \boldsymbol{\sigma} \, d\Omega, \quad (4.122)$$

$$a^*(\xi; \mathbf{u}, \delta\mathbf{u}, \bar{\mathbf{u}}) \equiv \iint_{\Omega} \boldsymbol{\epsilon}(\bar{\mathbf{u}}) : \mathbf{D}^{\text{alg}} : \boldsymbol{\epsilon}(\delta\mathbf{u}) \, d\Omega. \quad (4.123)$$

The notation $a^*(\xi, \mathbf{u}; \delta\mathbf{u}, \bar{\mathbf{u}})$ is used such that the form implicitly depends on the plastic variable ξ and the total displacement \mathbf{u} and is bilinear with respect to $\delta\mathbf{u}$ and $\bar{\mathbf{u}}$. The energy form also implicitly depends on the plastic variables. Note that unlike the geometric nonlinear systems in Chap. 3, the initial stiffness term does not appear, since only infinitesimal deformation is being considered. Total and updated Lagrangian formulations become identical for the infinitesimal deformation problem.

Since only material nonlinearity is considered, the weak form of structural equilibrium can be written at load step t_n as

$$a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (4.124)$$

Let the current load step be t_{n+1} and let the current iteration counter be k . Assuming that the applied loads are independent of displacement, the linearized incremental equation of is obtained as

$$a^*(^n\xi; ^{n+1}\mathbf{u}; \delta\mathbf{u}^k, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}) - a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}, \quad (4.125)$$

and the total displacement is updated using

$$^{n+1}\mathbf{u}^{k+1} = ^{n+1}\mathbf{u}^k + \delta\mathbf{u}^k. \quad (4.126)$$

Note that incremental equation (4.125) is in the form of $[^{n+1}\mathbf{K}^k] \cdot \{\delta\mathbf{u}^k\} = \{^{n+1}\mathbf{R}^k\}$ after discretization using finite elements. Equation (4.125) is solved iteratively until the residual vanishes, which means that the original nonlinear equation (4.124) is satisfied. It is emphasized here that the linearized increment Eq. (4.125) solves for displacement increment $\delta\mathbf{u}^k = ^{n+1}\mathbf{u}^{k+1} - ^{n+1}\mathbf{u}^k$ between two consecutive iterations, but the strain increment should be calculated using the displacement increment $\Delta\mathbf{u}^k = \Delta^{n+1}\mathbf{u}^k + ^n\mathbf{u}$ as in Eq. (4.10). This is because the stress and all history variables are updated from the previous converged load increment, not from the previous iteration.

Unlike nonlinear elastic systems, the elastoplastic system requires one more step after the nonlinear equation (4.124) converges at load step t_{n+1} . Since the stress and plastic variables will be used in the next load step, they need to be updated at the end of the current load step. This step is identical to the updating procedures for stress and plastic variables, described in Eqs. (4.112)–(4.115). During iteration, these variables are calculated, but they are not stored because they are not the converged values. Once the nonlinear equation is converged, these values are updated and stored.

4.3.6 Computational Implementation of Elastoplasticity

In this section, implementation of the elastoplasticity with von Mises yield criterion and combined linear isotropic/kinematic hardening model is presented. Even if it is possible to develop the finite element formulation for various element types, an eight-node hexahedral solid element in Chap. 1 will be used for demonstration purposes. Since only the material nonlinearity is considered, it is assumed that the strain as well as the rigid-body rotation is small.

In the computer implementation of finite element programs, matrix-vector notation is more convenient than tensor notation. In matrix-vector notation, a second-order symmetric tensor is expressed using a vector, while a fourth-order symmetric tensor is expressed using a matrix. For example, the Cauchy stress and incremental strain vectors are defined as

$$\{\boldsymbol{\sigma}\} = [\sigma_{11} \quad \sigma_{22} \quad \sigma_{33} \quad \sigma_{12} \quad \sigma_{23} \quad \sigma_{13}]^T$$

and

$$\{\Delta\boldsymbol{\epsilon}\} = [\Delta\epsilon_{11} \quad \Delta\epsilon_{22} \quad \Delta\epsilon_{33} \quad 2\Delta\epsilon_{12} \quad 2\Delta\epsilon_{23} \quad 2\Delta\epsilon_{13}]^T,$$

respectively. In the above definitions, the symmetric property of the tensor is used.

It is assumed that the incremental displacement vector $\Delta\mathbf{d}_I = \{\Delta d_{I1}, \Delta d_{I2}, \Delta d_{I3}\}^T$ is given for each node of the element. The subscript I is used to denote the node such that \mathbf{d}_I will be the displacement vector for node I . In each element, the node numbers are locally defined such that $I = 1, 2, \dots, 8$ for the hexahedral element (see Fig. 4.18). The displacement increment within the element can be calculated using the following interpolation scheme:

$$\Delta\mathbf{u} = \sum_{I=1}^8 N_I(\xi) \Delta\mathbf{d}_I, \quad (4.127)$$

where $\xi = \{\xi, \eta, \zeta\}^T$ is the natural coordinate vector at the reference element, $N_I(\xi)$ is the interpolation or shape function whose expression is given in Eq. (1.136), and $\Delta\mathbf{d}_I$ is the vector of nodal displacement increment. Since stress and plastic variables are calculated at the integration points, the value of natural coordinate is selected at the integration point.

For given displacement increments, strain increments can be calculated in a similar way as with the linear elastic material because the deformation is assumed to be infinitesimal. Thus, the strain increment can be interpolated by

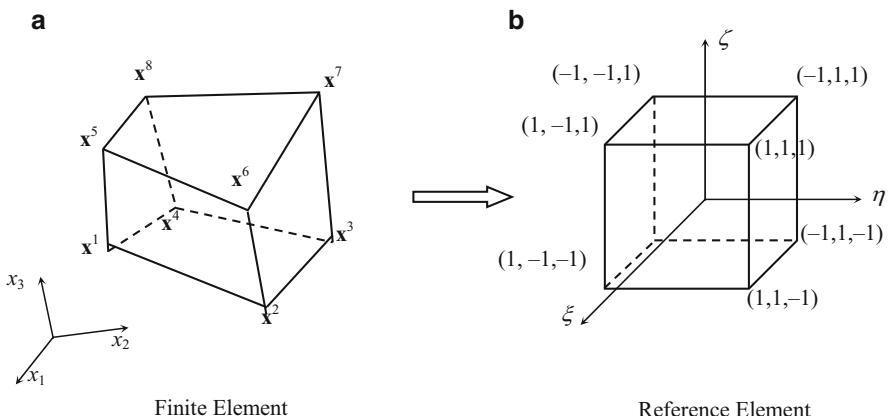


Fig. 4.18 Eight-node three-dimensional isoparametric solid element. (a) Finite Element (b) Reference Element

$$\{\Delta \boldsymbol{\epsilon}\} = \sum_{I=1}^8 \mathbf{B}_I \Delta \mathbf{d}_I,$$

where

$$\mathbf{B}_I = \begin{bmatrix} N_{I,1} & 0 & 0 \\ 0 & N_{I,2} & 0 \\ 0 & 0 & N_{I,3} \\ N_{I,2} & N_{I,1} & 0 \\ 0 & N_{I,3} & N_{I,2} \\ N_{I,3} & 0 & N_{I,1} \end{bmatrix} \quad (4.128)$$

is the discrete displacement-strain matrix of a solid element, and $N_{I,j} = \partial N_I / \partial x_j$ is the derivatives of the shape function N_I with respect to the physical coordinates whose expression is given in Eq. (1.138). Note that since the shape function is defined in the natural coordinates, it is necessary to use Jacobian relation between the physical and natural coordinates. Because of the infinitesimal deformation assumption, the total displacement and the total strain can be obtained by adding all incremental displacements and strains, respectively. Thus, the total displacements and strains at load increment t_{n+1} can be obtained by

$$^{n+1}\mathbf{u} = {}^n\mathbf{u} + \Delta \mathbf{u}$$

and

$$\{{}^{n+1}\boldsymbol{\epsilon}\} = \{{}^n\boldsymbol{\epsilon}\} + \{\Delta \boldsymbol{\epsilon}\}.$$

In addition to the strain increment, the stress and plastic variables at the previous load increment t_n are required. In the case of combined linear isotropic/kinematic hardening model, the following variables are needed: ${}^n\boldsymbol{e}_p$ and

$$\begin{aligned} \{{}^n\boldsymbol{\sigma}\} &= \{{}^n\sigma_{11} \quad {}^n\sigma_{22} \quad {}^n\sigma_{33} \quad {}^n\sigma_{12} \quad {}^n\sigma_{23} \quad {}^n\sigma_{13}\}^T, \\ \{{}^n\boldsymbol{\alpha}\} &= \{{}^n\alpha_{11} \quad {}^n\alpha_{22} \quad {}^n\alpha_{33} \quad {}^n\alpha_{12} \quad {}^n\alpha_{23} \quad {}^n\alpha_{13}\}^T. \end{aligned}$$

Using the above stress and plastic variables and material parameters $(\lambda, \mu, \beta, H, \sigma_Y^0)$, the updated stress and plastic variables are calculated using the return-mapping algorithm.

4.3.6.1 Return-Mapping

At each integration point of the element, the stress and plastic variables are determined using the return-mapping algorithm. In the following flowchart, the parentheses for vectors are not used for notational brevity:

1. Unit tensor $\mathbf{1} = [1 \quad 1 \quad 1 \quad 0 \quad 0 \quad 0]^T$.
2. Trial stress ${}^n\boldsymbol{\sigma} = {}^n\boldsymbol{\sigma} + \mathbf{D} \cdot \Delta \boldsymbol{\epsilon}$.

3. Trace of stress $\text{tr}({}^t\boldsymbol{\sigma}) = {}^t\sigma_{11} + {}^t\sigma_{22} + {}^t\sigma_{33}$.
4. Shifted stress ${}^t\boldsymbol{\eta} = {}^t\boldsymbol{\sigma} - {}^n\boldsymbol{\alpha} - \frac{1}{3}\text{tr}({}^t\boldsymbol{\sigma})\mathbf{1}$.
5. Norm $\|{}^t\boldsymbol{\eta}\| = \sqrt{({}^t\eta_{11})^2 + ({}^t\eta_{22})^2 + ({}^t\eta_{33})^2 + 2[({}^t\eta_{12})^2 + ({}^t\eta_{23})^2 + ({}^t\eta_{13})^2]}$.
6. Yield function $f = \|{}^t\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}}[\sigma_Y^0 + (1 - \beta)H^n e_p]$.
7. Check for the yield status.
IF $f < 0$ THEN
The material is elastic

$$\mathbf{D}^{\text{alg}} = \mathbf{D} = \begin{bmatrix} \lambda + 2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda + 2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda + 2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{bmatrix}.$$

EXIT
ENDIF

8. Consistency parameter $\hat{\gamma} = \frac{f}{2\mu + \frac{2}{3}H}$.
9. Unit deviatoric vector $\mathbf{N} = \frac{{}^t\boldsymbol{\eta}}{\|{}^t\boldsymbol{\eta}\|}$.
10. Update stress ${}^{n+1}\boldsymbol{\sigma} = {}^t\boldsymbol{\sigma} - 2\mu\hat{\gamma}\mathbf{N}$.
11. Update back stress ${}^{n+1}\boldsymbol{\alpha} = {}^n\boldsymbol{\alpha} + (2/3)\beta H\hat{\gamma}\mathbf{N}$.
12. Update eff. plastic strain ${}^{n+1}e_p = {}^n e_p + \sqrt{(2/3)}\hat{\gamma}$.
13. Consistent tangent stiffness

$$c_1 = \frac{4\mu^2}{2\mu + \frac{2}{3}H}, \quad c_2 = \frac{4\mu^2\hat{\gamma}}{\|{}^t\boldsymbol{\eta}\|},$$

$$\mathbf{I}^{\text{dev}} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} & 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \end{bmatrix},$$

$$D_{ij}^{\text{alg}} = D_{ij} - (c_1 - c_2)N_i N_j - c_2 \times I_{ij}^{\text{dev}}.$$

Note that the above calculation must be performed for each integration point of an element. Accordingly, the stress and plastic variables at each integration point must be stored and updated. There is no need to store the stress and plastic variables during un-converged iteration. Once the Newton–Raphson method is converged at load increment t_{n+1} , they are stored, and next load increment starts.

In practical implementation, the stress update procedure is often separated from algorithmic tangent stiffness calculation because the former is used more frequently than the latter. Below are two MATLAB programs, **combHard** and **combHardTan**. Both programs require stress, back stress, and effective plastic strain at the previous load increment as inputs, as well as material parameters. It is expected that stress and back stress are column vectors with dimension of 6×1 . Program **combHard** then returns updates stress, back stress and effective plastic strain, while **combHardTan** returns the algorithmic tangent stiffness matrix.

```
PROGRAM combHard
%
% Linear combined isotropic/kinematic hardening model
%
function [stress, alpha, ep]=combHard(mp,Ddeps,stressN,alphaN,epN)
% Inputs:
% mp = [lambda, mu, beta, H, Y0];
% D = elastic stiffness matrix
% stressN= [s11, s22, s33, t12, t23, t13];
% alphan = [a11, a22, a33, a12, a23, a13];
%
Iden = [1 1 1 0 0 0]';
two3 = 2/3; stwo3=sqrt(two3); %constants
mu=mp(2); beta=mp(3); H=mp(4); Y0=mp(5); %material properties
ftol = Y0*1E-6; %tolerance for yield
stresstr = stressN + D*deps; %trial stress
I1 = sum(stresstr(1:3)); %trace(stresstr)
str = stresstr - I1*Iden/3; %deviatoric stress
eta = str - alphan; %shifted stress
etat = sqrt(eta(1)^2 + eta(2)^2 + eta(3)^2 ...
    + 2*(eta(4)^2 + eta(5)^2 + eta(6)^2)); %norm of eta
fyld = etat - stwo3*(Y0+(1-beta)*H*epN); %trial yield function
if fyld < ftol %yield test
    stress = stresstr; alpha = alphan; ep = epN; %trial states are final
    return;
else
    gamma = fyld/(2*mu + two3*H); %plastic consistency param
    ep = epN + gamma*stwo3; %updated eff. plastic strain
end
N = eta/etat; %unit vector normal to f
stress = stresstr - 2*mu*gamma*N; %updated stress
alpha = alphan + two3*beta*H*gamma*N; %updated back stress
```

```

PROGRAM combHardTan
%
% Tangent stiffness for linear combined isotropic/kinematic hardening
model
%
function [Dtan]=combHardTan(mp,D,deps,stressN,alphaN,epN)
% Inputs:
% mp = [lambda, mu, beta, H, Y0];
% D = elastic stiffness matrix
% stressN = [s11, s22, s33, t12, t23, t13];
% alphaN = [a11, a22, a33, a12, a23, a13];
%
Iden = [1 1 1 0 0 0]';
two3 = 2/3; stwo3=sqrt(two3); %constants
mu=mp(2); beta=mp(3); H=mp(4); Y0=mp(5); %material properties
ftol = Y0*1E-6; %tolerance for yield
stresstr = stressN + D*deps; %trial stress
I1 = sum(stresstr(1:3)); %trace(stresstr)
str = stresstr - I1*Iden/3; %deviatoric stress
eta = str - alphaN; %shifted stress
etat = sqrt(eta(1)^2 + eta(2)^2 + eta(3)^2 ...
    + 2*(eta(4)^2 + eta(5)^2 + eta(6)^2)); %norm of eta
fyld = etat - stwo3*(Y0+(1-beta)*H*epN); %trial yield function
if fyld < ftol %yield test
    Dtan = D; return;
end
gamma = fyld/(2*mu + two3*H); %plastic consistency param
N = eta/etat; %unit vector normal to f
var1 = 4*mu^2/(2*mu+two3*H); %coefficients
var2 = 4*mu^2*gamma/etat;
Dtan = D - (var1-var2)*N*N' + var2*Iden*Iden'/3; %tangent stiffness
Dtan(1,1) = Dtan(1,1) - var2; %contr. from 4th-order I
Dtan(2,2) = Dtan(2,2) - var2;
Dtan(3,3) = Dtan(3,3) - var2;
Dtan(4,4) = Dtan(4,4) - .5*var2;
Dtan(5,5) = Dtan(5,5) - .5*var2;
Dtan(6,6) = Dtan(6,6) - .5*var2;

```

4.3.6.2 Finite Element Procedure for Elastoplasticity

Once stress and plastic variables are determined, they can be used for solving the nonlinear equilibrium equation. First, the variation of strain can be interpolated using the same strain-displacement matrix \mathbf{B}_I as

$$\{\boldsymbol{\epsilon}(\bar{\mathbf{u}})\} = \sum_{I=1}^8 \mathbf{B}_I \bar{\mathbf{d}}_I = [\mathbf{B}]\{\bar{\mathbf{d}}\}, \quad (4.129)$$

where $\bar{\mathbf{d}}_I = \{\bar{d}_{I1}, \bar{d}_{I2}, \bar{d}_{I3}\}^T$ is the displacement variation at node I , while $\bar{\mathbf{d}} = \{\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2, \dots, \bar{\mathbf{d}}_8\}^T$ is the displacement variation of all nodes in the element. Similarly, $[\mathbf{B}] = [\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_8]$.

Using the above equation and the updated stress in Eq. (4.112), the discrete version of the energy form can be derived as

$$\begin{aligned} a({}^n\xi; {}^{n+1}\mathbf{u}, \bar{\mathbf{u}}) &= \iint_{\Omega} \boldsymbol{\epsilon}(\bar{\mathbf{u}})^T \{{}^{n+1}\boldsymbol{\sigma}\} d\Omega = \{\bar{\mathbf{d}}\}^T \iint_{\Omega} [\mathbf{B}]^T \{{}^{n+1}\boldsymbol{\sigma}\} d\Omega \\ &\equiv \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{\text{int}}\}, \end{aligned} \quad (4.130)$$

where $\{\mathbf{f}^{\text{int}}\}$ is the discrete internal force vector. When numerical integration is used, $\{\mathbf{f}^{\text{int}}\}$ can be calculated by

$$\{\mathbf{f}^{\text{int}}\} = \sum_{K=1}^{\text{NG}} \left([\mathbf{B}]^T \{{}^{n+1}\boldsymbol{\sigma}\} |\mathbf{J}| \right)_K \omega_K,$$

where NG is the number of integration points, $|\mathbf{J}|$ is the Jacobian between the physical and reference elements, and ω is integration weight.

In addition, the discrete external force vector can be derived from the definition of the load form as

$$\begin{aligned} \ell(\bar{\mathbf{u}}) &= \iint_{\Omega} \bar{\mathbf{u}}^T \mathbf{f}^B d\Omega + \int_{\Gamma^S} \bar{\mathbf{u}}^T \mathbf{f}^S d\Gamma \\ &= \sum_{I=1}^4 \bar{\mathbf{d}}_I^T \left\{ \iint_{\Omega} N_I(\xi) \mathbf{f}^b d\Omega + \int_{\Gamma^S} N_I(\xi) \mathbf{f}^S d\Gamma \right\}. \\ &\equiv \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{\text{ext}}\} \end{aligned} \quad (4.131)$$

When concentrated nodal forces are applied, they can directly be added to the corresponding locations in $\{\mathbf{f}^{\text{ext}}\}$. Since the applied loads are assumed to be independent of deformation, the external force $\{\mathbf{f}^{\text{ext}}\}$ is a fixed vector. Thus, the discrete version of solving the nonlinear equilibrium equation is to find the internal force that has the same value as the external force, i.e.,

$$\{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{\text{int}}(\mathbf{d})\} = \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{\text{ext}}\}, \quad \forall \{\bar{\mathbf{d}}\} \in \mathbb{Z}_h, \quad (4.132)$$

where \mathbb{Z}_h is the discrete counterpart of space \mathbb{Z} . Since the displacement variation is 0 at the nodes where displacements are prescribed, Eq. (4.132) satisfies $\{\mathbf{f}^{\text{int}}(\mathbf{d})\} = \{\mathbf{f}^{\text{ext}}\}$ for all nodes whose displacements are not prescribed.

Since the internal force is a nonlinear function of deformation, Eq. (4.132) needs to be solved using an iterative method, such as the Newton–Raphson method, which requires the Jacobian matrix, or equivalently, the tangent stiffness matrix. Using the consistent tangent stiffness, the linearized energy form can be discretized by

$$a^*({}^n\xi, {}^{n+1}\mathbf{u}; \delta\mathbf{u}, \bar{\mathbf{u}}) = \{\bar{\mathbf{d}}\}^T \iint_{\Omega} [\mathbf{B}]^T [\mathbf{D}^{\text{alg}}] [\mathbf{B}] d\Omega \{\delta\mathbf{d}\}. \quad (4.133)$$

The integral term in the above equation is called the tangent stiffness matrix. After using numerical integration, it becomes

$$[\mathbf{K}_T] = \sum_{K=1}^{NG} \left([\mathbf{B}]^T [\mathbf{D}^{\text{alg}}] [\mathbf{B}] |\mathbf{J}| \right)_K \omega_K. \quad (4.134)$$

In general, the above integration as well as the one in the internal force in Eq. (4.130) are evaluated using the Gauss quadrature rule. Normally, 2×2 integration points are used for a quadrilateral element.

The discretized version of incremental equation in Eq. (4.125) can now be written in the form of finite element matrix equation as

$$\{\bar{\mathbf{d}}\}^T [\mathbf{K}_T] \{\delta\mathbf{d}\} = \{\bar{\mathbf{d}}\}^T \{\mathbf{f}^{\text{ext}} - \mathbf{f}^{\text{int}}\}, \quad \forall \{\bar{\mathbf{d}}\} \in \mathbb{Z}_h. \quad (4.135)$$

The above linear system of equations needs to be solved iteratively until the residual force (right-hand side) vanishes. Different methods of solving nonlinear equations in Chap. 2 can be used. For example, in the case of the modified Newton–Raphson method, the tangent stiffness matrix $[\mathbf{K}_T]$ at the first iteration is repeatedly used. In the case of the incremental force method, the external force vector $\{\mathbf{f}^{\text{ext}}\}$ is divided by the number of increments, and the Newton–Raphson method is employed at each load increment.

Finally, it is reminded that the above algorithm only works under the assumption of small deformation and rotation. When deformation becomes large, the assumption of additive decomposition, i.e., $\Delta\boldsymbol{\epsilon} = \Delta\boldsymbol{\epsilon}_e + \Delta\boldsymbol{\epsilon}_p$, cannot be valid. When deformation is small but a rigid-body rotation is present, the assumption of additive decomposition can still be used with caution, which will be discussed in the Sect. 4.4.

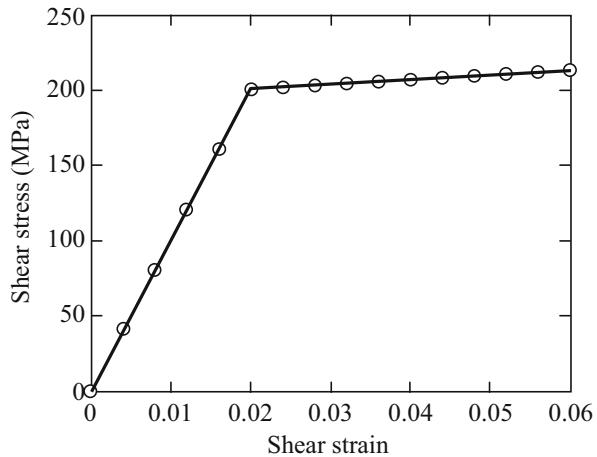
Example 4.15 (Shear Deformation of Elastoplastic Square) A plane strain square undergoes simple shear deformation in the x – y plane. When the shear strain increment is $\Delta\gamma_{12} = 0.004$ at each step, plot shear stress τ_{12} vs. shear strain γ_{12} curve for 15 load increments. Calculate the slopes in the elastic and plastic regions and compare these slopes from theoretical ones. Assume linear isotropic hardening with the following material properties: $E = 24$ GPa, $\nu = 0.2$, $H = 1.0$ GPa, and $\sigma_Y^0 = 200\sqrt{3}$ MPa.

Solution Since the square undergoes simple shear deformation, only nonzero stress and strain components are τ_{12} and γ_{12} , respectively. There is no effect on the deviatoric component because all diagonal components are 0. Below is the list of MATLAB programs that solve for stresses with given strain increments. Note that the input **deps** is the strain increment, not the total strain. Figure 4.19 shows shear stress vs. shear strain curve. As shown in Example 4.8, in pure shear deformation, the material will yield at

$$\tau_{12} = \frac{1}{\sqrt{3}} \sigma_Y^0 = 200 \text{ MPa},$$

which is consistent with the curve in Fig. 4.19.

Fig. 4.19 Stress-strain curve for elastoplastic shear deformation



```
%  
% Example 4.15 - Shear deformation of elastoplastic square  
%  
Young = 24000; nu=0.2; mu=Young/2/(1+nu); lambda=nu*Young/((1+nu)*(1-  
2*nu));  
beta = 0; H = 1000; sY = 200*sqrt(3);  
mp = [lambda mu beta H sY];  
Iden=[1 1 1 0 0 0]';  
D=2*mu*eye(6) + lambda*Iden*Iden';  
stressN=zeros(6,1); deps= zeros(6,1); alphaN = zeros(6,1); epN=0;  
for i=1:15  
    deps(4) = 0.004;  
    [stress, alpha, ep]=combHard(mp,D,deps,stressN,alphaN,epN);  
    X(i) = i*deps(4); Y(i) = stress(4); Z(i)  
    stressN = stress; alphaN = alpha; epN = ep;  
end  
X = [0 X]; Y=[0 Y]; plot(X,Y);
```

The slope of elastic region is 10,000 MPa, which corresponds to μ . The slope in plastic region is 322.5 MPa, which corresponds to tangent stiffness in the plastic region. Since τ_{12} is only nonzero stress component, the trial stress can be written in the scalar form:

$${}^{\text{tr}}\tau_{12} = {}^n\tau_{12} + \mu\Delta\gamma_{12}.$$

For the convenience of discussion, it is assumed that ${}^n\tau_{12}$ is at the initial yield point, i.e., ${}^n\tau_{12} = 200$ MPa. Then, the yield function at the trial stress becomes positive (outside of the yield surface):

$$f = \sqrt{2}{}^{\text{tr}}\tau_{12} - \sqrt{\frac{2}{3}}\sigma_Y^0 = \sqrt{2}{}^n\tau_{12} + \sqrt{2}(\mu\Delta\gamma_{12}) - \sqrt{\frac{2}{3}}\sigma_Y^0 = \sqrt{2}\mu\Delta\gamma_{12} > 0.$$

Thus, the plastic consistency parameter can be calculated by

$$\gamma = \frac{\sqrt{2}\mu}{2\mu + \frac{2}{3}H} \Delta\gamma_{12}.$$

Then, the return-mapping algorithm for the shear stress becomes

$${}^{n+1}\tau_{12} = {}^n\tau_{12} + \mu\Delta\gamma_{12} - \mu\gamma N_{12}.$$

It is straightforward to obtain $N_{12} = 1/\sqrt{2}$ from the definition of \mathbf{N} . After substituting the plastic consistency parameter, the updated shear stress becomes

$${}^{n+1}\tau_{12} = {}^n\tau_{12} + \mu \left[1 - \frac{2\mu}{2\mu + \frac{2}{3}H} \right] \Delta\gamma_{12}.$$

The coefficient of $\Delta\gamma_{12}$ is the slope in the plastic region of shear stress–shear strain curve. ■

4.4 Finite Rotation with Objective Integration

The elastoplasticity with return-mapping algorithm in the previous section can be extended when the structure undergoes a small strain but with a finite rotation. In such a case, if a body-fixed coordinate that rotates with the structure is used, stresses will not be affected by the rigid-body rotation, which is the main idea in this section. However, it can be shown that the rate of Cauchy stress is not invariant under rigid-body rotation. The rate of a tensor that is independent of imposed rigid-body rotation is called an objective rate. Thus, the concept of an objective stress rate will be introduced first. For a finite rotational problem, objective rate tensors have to be used to describe the motion of the structure and to obtain the material frame independent results. Although an extensive amount of research has been done pertaining to objective rates, only co-rotational Cauchy stress rate will be introduced in this section. Other types of objective rates can be used in a similar manner. The same constitutive relation for elastoplasticity in Sect. 4.3 can be used in terms of objective stress rate. The approach is therefore valid for elastoplastic systems with small elastic deformation, relatively large plastic deformation, and large rigid-body rotation. Numerical difficulties associated with the objective rate include the intricate transformation of a stress tensor into a rotation-free configuration, the unsymmetric properties of the tangent stiffness, and the difficulty in obtaining an exact tangent stiffness. Nevertheless, this model has been implemented in a good deal of application software.

4.4.1 Objective Tensor and Objective Rate

1. Objective Tensor

Any tensor that is not affected by superimposed rigid-body translations and rotations of a spatial frame is called an *objective tensor*. It is noted that a rotation of the body is equivalent to the rotation of the reference frame of the same magnitude but in opposite direction. Consider two reference frames $x - y - z$ and $\bar{x} - \bar{y} - \bar{z}$ in Fig. 4.20. The former is translated by $\mathbf{c}(t)$ and rotated by $\mathbf{Q}(t)^T$ to arrive at the latter. The position of a particle P at time t is \mathbf{x} in $x - y - z$ frame and $\bar{\mathbf{x}}$ in $\bar{x} - \bar{y} - \bar{z}$ frame. The former vector \mathbf{x} is seen by an observer sitting in $\bar{x} - \bar{y} - \bar{z}$ frame as $\mathbf{Q}(t) \cdot \mathbf{x}$. In fact, the position vectors are related by

$$\bar{\mathbf{x}} = \mathbf{Q}(t) \cdot \mathbf{x} + \mathbf{c}(t), \quad (4.136)$$

where $\mathbf{c}(t)$ is the position vector of the origin of $x - y - z$ frame and $\mathbf{Q}(t)^T$ is the orthogonal tensor that gives the orientation of $x - y - z$ frame relative to $\bar{x} - \bar{y} - \bar{z}$ frame. Thus, $x - y - z$ frame is different from $\bar{x} - \bar{y} - \bar{z}$ by a rigid-body translation $\mathbf{c}(t)$ and a rotation $\mathbf{Q}(t)^T$.

Objectivity is also known as reference frame indifference. Quantities that depend only on the orientation of the reference frame, which is given by \mathbf{Q} , and not on the other aspects of the motion of the reference frame (e.g., translation, velocity and acceleration, angular velocity and angular acceleration) are said to be indifferent or objective. Therefore, components of a tensor observed by two different observers are different. This difference is due to the different orientations of the observers but not to relative motions between the observers.

Definition Objective tensor. A scalar f , a vector \mathbf{v} , and a second-order tensor \mathbf{T} are objective, if for reference frames $\bar{\mathbf{x}}$ and \mathbf{x} related by

$$\bar{\mathbf{x}} = \mathbf{c}(t) + \mathbf{Q}(t)\mathbf{x}, \quad (4.137)$$

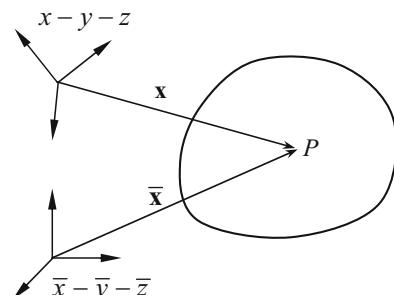


Fig. 4.20 Two frames different by rigid-body translation and rotation

i.e., differing only by a rigid-body motion, the corresponding scalar \bar{f} , vector $\bar{\mathbf{v}}$, and tensor $\bar{\mathbf{T}}$ are related by the following:

$$\begin{aligned}\bar{f} &= f, \\ \bar{\mathbf{v}} &= \mathbf{Q}\mathbf{v}, \\ \bar{\mathbf{T}} &= \mathbf{QTQ}^T.\end{aligned}$$

■

In this context, the change of coordinate system is not considered (each observer is free to choose a coordinate system), but it is concerned with the change of observer's positions and orientations, or the change of reference frame. It is convenient to imagine that the observers are attached to the continua and move with the continua. If two motions are different only by a rigid-body motion, then the two reference frames as seen in the eyes of the two observers are different by a translation of the origin and a rotation of orientation. Therefore, in the discussion of objectivity, transformation of coordinate system does not play any part.

Example 4.16 (Objective Tensor) When the two spatial frames are related by Eq. (4.137), show that how the deformation gradient \mathbf{F} , the right Cauchy-Green deformation tensor $\mathbf{C} = \mathbf{F}^T\mathbf{F}$, and the left Cauchy-Green deformation tensor $\mathbf{b} = \mathbf{FF}^T$ transform.

Solution The deformation gradient in the two different frames can be written as

$$\mathbf{F} = \frac{\partial \mathbf{x}}{\partial \mathbf{X}}, \quad \bar{\mathbf{F}} = \frac{\partial \bar{\mathbf{x}}}{\partial \mathbf{X}}.$$

Note that the material frame, \mathbf{X} , never changes. The discussion in the objectivity is only related to the spatial frames. By substituting the relation between the two frames in Eq. (4.137),

$$\bar{\mathbf{F}} = \frac{\partial \bar{\mathbf{x}}}{\partial \mathbf{X}} = \frac{\partial (\mathbf{c} + \mathbf{Qx})}{\partial \mathbf{X}} = \mathbf{Q} \frac{\partial \mathbf{x}}{\partial \mathbf{X}} = \mathbf{QF}.$$

Note that the deformation gradient is a second-order tensor, but it behaves like a vector. This happens because the deformation gradient depends on two frames: the material frame (\mathbf{X}) and the spatial frame (\mathbf{x}). Since the vector in the material frame remains unchanged, it behaves like a vector.

The right Cauchy-Green deformation tensor in $\bar{x} - \bar{y} - \bar{z}$ frame can be written as

$$\bar{\mathbf{C}} = \bar{\mathbf{F}}^T \bar{\mathbf{F}} = (\mathbf{QF})^T (\mathbf{QF}) = \mathbf{F}^T \mathbf{Q}^T \mathbf{QF} = \mathbf{F}^T \mathbf{F} = \mathbf{C}.$$

Thus, the right Cauchy-Green deformation tensor behaves like a scalar. This is expected because \mathbf{C} is a material tensor.

The left Cauchy-Green deformation tensor in $\bar{x} - \bar{y} - \bar{z}$ frame can be written as

$$\bar{\mathbf{b}} = \bar{\mathbf{F}}\bar{\mathbf{F}}^T = (\mathbf{Q}\mathbf{F})(\mathbf{Q}\mathbf{F})^T = \mathbf{Q}\mathbf{F}\mathbf{F}^T\mathbf{Q}^T = \mathbf{Q}\mathbf{b}\mathbf{Q}^T.$$

Thus, the spatial tensor \mathbf{b} is an objective tensor. ■

As discussed in the above example, only spatial tensors are considered to determine objectivity. Consider the velocity gradient in two different frames:

$$\mathbf{L} = \frac{\partial \mathbf{v}}{\partial \mathbf{x}}, \quad \bar{\mathbf{L}} = \frac{\partial \bar{\mathbf{v}}}{\partial \bar{\mathbf{x}}}, \quad (4.138)$$

where the velocity can be obtained by differentiating the relation $\bar{\mathbf{x}} = \mathbf{Q}\mathbf{x}$ with ignoring the rigid-body translation, as

$$\bar{\mathbf{v}} = \dot{\bar{\mathbf{x}}} = \mathbf{Q}\mathbf{v} + \dot{\mathbf{Q}}\mathbf{x}.$$

Since \mathbf{Q} is an orthogonal tensor, $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and $\mathbf{x} = \mathbf{Q}^{-1}\bar{\mathbf{x}} = \mathbf{Q}^T\bar{\mathbf{x}}$. Using this relation, the above equation becomes

$$\bar{\mathbf{v}} = \mathbf{Q}\mathbf{v} + \dot{\mathbf{Q}}\mathbf{Q}^T\bar{\mathbf{x}}. \quad (4.139)$$

Thus, the spatial velocity vector $\bar{\mathbf{v}}$ is not objective under rigid-body rotation. The velocity gradient in Eq. (4.138) can be obtained by differentiating the relation in Eq. (4.139) as

$$\bar{\mathbf{L}} = \frac{\partial \bar{\mathbf{v}}}{\partial \bar{\mathbf{x}}} = \mathbf{Q} \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \bar{\mathbf{x}}} + \dot{\mathbf{Q}}\mathbf{Q}^T \frac{\partial \bar{\mathbf{x}}}{\partial \bar{\mathbf{x}}} = \mathbf{Q}\mathbf{L}\mathbf{Q}^T + \dot{\mathbf{Q}}\mathbf{Q}^T. \quad (4.140)$$

This shows that the velocity gradient tensor is not an objective tensor because of the presence of the second term. This means that the velocity gradient cannot be used to describe the material behavior.

Example 4.17 (Rate of Deformation Tensor and Spin Tensor) The symmetric part of the velocity gradient \mathbf{L} is called the rate of deformation tensor \mathbf{d} , whereas the skew-symmetric part of \mathbf{L} is called the spin tensor \mathbf{W} . Show that \mathbf{d} is objective and \mathbf{W} is not objective.

Solution From the definition of the velocity gradient, the rate of deformation can be written as

$$\bar{\mathbf{d}} = \frac{1}{2}(\bar{\mathbf{L}} + \bar{\mathbf{L}}^T) = \frac{1}{2}\left(\frac{\partial \bar{\mathbf{v}}}{\partial \bar{\mathbf{x}}} + \left(\frac{\partial \bar{\mathbf{v}}}{\partial \bar{\mathbf{x}}}\right)^T\right) = \frac{1}{2}\left(\frac{\partial \bar{\mathbf{v}}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \bar{\mathbf{x}}} + \left(\frac{\partial \bar{\mathbf{v}}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \bar{\mathbf{x}}}\right)^T\right).$$

Using the relation $\bar{\mathbf{v}} = \mathbf{Q}\mathbf{v} + \dot{\mathbf{Q}}\mathbf{x}$ and $\partial\mathbf{x}/\partial\bar{\mathbf{x}} = \mathbf{Q}^T$, the above relation can be simplified into

$$\bar{\mathbf{d}} = \frac{1}{2}(\mathbf{QLQ}^T + \mathbf{QL}^T\mathbf{Q}^T) + \frac{1}{2}(\dot{\mathbf{QQ}}^T + \mathbf{QQ}^T\dot{\mathbf{Q}}).$$

It is clear that the first term on the right-hand side is the rate of deformation in $x - y - z$ frame. For the second term, consider the following property of orthogonal tensor: $\mathbf{QQ}^T = \mathbf{0}$. Since this relation is satisfied for all time, its time rate should be 0, i.e., $\dot{\mathbf{QQ}}^T + \mathbf{QQ}^T\dot{\mathbf{Q}} = \mathbf{0}$. Thus, the second term in the above equation vanishes, and

$$\bar{\mathbf{d}} = \mathbf{QdQ}^T.$$

Thus, the rate of deformation is an objective rate. On the other hand, the spin tensor becomes

$$\bar{\mathbf{W}} = \mathbf{QWQ}^T + \frac{1}{2}(\dot{\mathbf{QQ}}^T - \mathbf{QQ}^T\dot{\mathbf{Q}}).$$

Due to the second term on the right-hand side, the spin tensor is not an objective tensor. ■

In the above example, it is seen that \mathbf{L} and \mathbf{W} depend on the spin of the rotating system, but \mathbf{d} depends only on the orientation of the reference (spatial) frame. Note that for two chosen coordinate systems, \mathbf{L} transforms like a second-order tensor if $\dot{\mathbf{Q}} = \mathbf{0}$.

2. Objective Rate

Suppose that \mathbf{T} (any symmetric tensor) is an objective tensor. Is it true to say that its rate $\dot{\mathbf{T}}$ is also objective? The concept of the rate of a tensor $\dot{\mathbf{T}}$ is important to find the effect of deformation or loading history on the media. If a tensor is known as some instantaneous moments, then by Taylor series expansion with respect to time, the derivatives of this tensor field are required in order to determine the future response of the media. In the rate form of elastoplasticity, the constitutive relation is given in terms of stress and strain rates. Knowing an objective tensor \mathbf{T} (stress or strain), the question is whether $\dot{\mathbf{T}}$ is objective or not. If it is not objective, it cannot be used to predict the future value of \mathbf{T} .

Let \mathbf{T} be an objective tensor such that it satisfies the following transformation relation:

$$\bar{\mathbf{T}} = \mathbf{QTQ}^T. \quad (4.141)$$

Then the rate of \mathbf{T} can be written as

$$\dot{\bar{\mathbf{T}}} = \dot{\mathbf{QTQ}}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{QT}\dot{\mathbf{Q}}^T.$$

The presence of $\dot{\mathbf{Q}}\mathbf{T}\mathbf{Q}^T$ and $\mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T$ in the above equation shows that $\dot{\mathbf{T}}$ is not an objective rate. To determine an objective derivative, these two terms that include $\dot{\mathbf{Q}}$ and $\dot{\mathbf{Q}}^T$ need to be eliminated from the above expression, since these terms cannot appear in an objective tensor transformation. From the transformation of the velocity gradient in Eq. (4.140), $\dot{\mathbf{Q}}$ can be written as

$$\dot{\mathbf{Q}} = \bar{\mathbf{L}}\mathbf{Q} - \mathbf{Q}\mathbf{L}.$$

By taking the transpose of the above equation,

$$\dot{\mathbf{Q}}^T = \mathbf{Q}^T\bar{\mathbf{L}}^T - \mathbf{L}^T\mathbf{Q}^T.$$

By substituting the above two relations into the expression of $\dot{\bar{\mathbf{T}}}$,

$$\begin{aligned}\dot{\bar{\mathbf{T}}} &= (\bar{\mathbf{L}}\mathbf{Q} - \mathbf{Q}\mathbf{L})\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}(\mathbf{Q}^T\bar{\mathbf{L}}^T - \mathbf{L}^T\mathbf{Q}^T) \\ &= \bar{\mathbf{L}}\mathbf{Q}\mathbf{T}\mathbf{Q}^T - \mathbf{Q}\mathbf{L}\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}\mathbf{Q}^T\bar{\mathbf{L}}^T - \mathbf{Q}\mathbf{L}\mathbf{T}\mathbf{Q}^T \\ &= \bar{\mathbf{L}}\dot{\mathbf{T}} - \mathbf{Q}\mathbf{L}\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \bar{\mathbf{L}}\mathbf{L}^T - \mathbf{Q}\mathbf{L}\mathbf{T}\mathbf{Q}^T\end{aligned}$$

or, after rearranging,

$$\dot{\bar{\mathbf{T}}} - \bar{\mathbf{L}}\dot{\mathbf{T}} - \bar{\mathbf{L}}\mathbf{L}^T = \mathbf{Q}(\dot{\mathbf{T}} - \mathbf{L}\dot{\mathbf{T}} - \mathbf{T}\mathbf{L}^T)\mathbf{Q}^T.$$

Thus, it follows that the following rate is an objective derivative:

$$\dot{\mathbf{T}} - \mathbf{L}\dot{\mathbf{T}} - \mathbf{T}\mathbf{L}^T. \quad (4.142)$$

Note that the above expression includes not only $\dot{\mathbf{T}}$ but also \mathbf{T} itself. The above rate is called the Truesdell rate. It is possible to show that that the following derivatives are also objective:

Co-rotational rate (Jaumann rate):

$$\dot{\mathbf{T}}^J \equiv \dot{\mathbf{T}} - \mathbf{W}\mathbf{T} + \mathbf{T}\mathbf{W}. \quad (4.143)$$

Convective rate:

$$\dot{\mathbf{T}}^C \equiv \dot{\mathbf{T}} + \mathbf{L}\dot{\mathbf{T}} + \mathbf{T}\mathbf{L}^T. \quad (4.144)$$

Any one of the above three derivatives is as good as another, although they are not equal to each other. If \mathbf{T} is stress, these are objective stress rates. In the spatial description, the constitutive model is given between the rate of stress and the rate of

strain. Since the constitutive model has to be independent of the reference frame, it is important to use the objective stress rate in describing the constitutive model.

Example 4.18 (Objective Stress Rates) Using the properties of $\dot{\mathbf{Q}} = \mathbf{Q}\mathbf{W}^T - \overline{\mathbf{W}}^T\mathbf{Q}$ and $\dot{\mathbf{Q}} = \mathbf{Q}\mathbf{L}^T - \overline{\mathbf{L}}^T\mathbf{Q}$, derive (a) co-rotational rate and (b) convective rate.

Solution:

(a) Using the property, $\dot{\mathbf{Q}} = \mathbf{Q}\mathbf{W}^T - \overline{\mathbf{W}}^T\mathbf{Q}$, the rate of $\dot{\mathbf{T}}$ can be written as

$$\begin{aligned}\dot{\mathbf{T}} &= \dot{\mathbf{Q}}\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}\dot{\mathbf{Q}}^T \\ &= (\mathbf{Q}\mathbf{W}^T - \overline{\mathbf{W}}^T\mathbf{Q})\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}(\mathbf{W}\mathbf{Q}^T - \mathbf{Q}^T\overline{\mathbf{W}}) \\ &= \mathbf{Q}(\mathbf{W}^T\mathbf{T})\mathbf{Q}^T - \overline{\mathbf{W}}^T\dot{\mathbf{T}} + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}\mathbf{W}\mathbf{Q}^T - \overline{\mathbf{W}}^T\mathbf{Q}.\end{aligned}$$

After rearranging and using the property of $\mathbf{W}^T = -\mathbf{W}$, the transformation relation becomes

$$\dot{\mathbf{T}} + \overline{\mathbf{W}}^T - \overline{\mathbf{W}}\dot{\mathbf{T}} = \mathbf{Q}(\dot{\mathbf{T}} + \mathbf{T}\mathbf{W} - \mathbf{W}\dot{\mathbf{T}})\mathbf{Q}^T.$$

Thus, $\dot{\mathbf{T}} + \mathbf{T}\mathbf{W} - \mathbf{W}\dot{\mathbf{T}}$ is an objective rate.

(b) Using the property, $\dot{\mathbf{Q}} = \mathbf{Q}\mathbf{L}^T - \overline{\mathbf{L}}^T\mathbf{Q}$, the rate of $\dot{\mathbf{T}}$ can be written as

$$\begin{aligned}\dot{\mathbf{T}} &= \dot{\mathbf{Q}}\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}\dot{\mathbf{Q}}^T \\ &= (\mathbf{Q}\mathbf{L}^T - \overline{\mathbf{L}}^T\mathbf{Q})\mathbf{T}\mathbf{Q}^T + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}(\mathbf{L}\mathbf{Q}^T - \mathbf{Q}^T\overline{\mathbf{L}}) \\ &= \mathbf{Q}(\mathbf{L}^T\mathbf{T})\mathbf{Q}^T - \overline{\mathbf{L}}^T\dot{\mathbf{T}} + \mathbf{Q}\dot{\mathbf{T}}\mathbf{Q}^T + \mathbf{Q}\mathbf{T}\mathbf{L}\mathbf{Q}^T - \overline{\mathbf{L}}^T\mathbf{Q}.\end{aligned}$$

After rearranging, the transformation relation becomes

$$\dot{\mathbf{T}} + \overline{\mathbf{L}}^T\dot{\mathbf{T}} + \overline{\mathbf{L}}\mathbf{Q} = \mathbf{Q}(\dot{\mathbf{T}} + \mathbf{L}^T\mathbf{T} + \mathbf{T}\mathbf{L})\mathbf{Q}^T.$$

Thus, $\dot{\mathbf{T}} + \mathbf{L}^T\mathbf{T} + \mathbf{T}\mathbf{L}$ is an objective rate. ■

4.4.2 Finite Rotation and Objective Rate

1. Jaumann Stress Rate

It is well known that the rate of the Cauchy stress tensor is not objective, even if the Cauchy stress itself is objective. As a body rotates without deformation, the Cauchy stress tensor changes because the direction of the stress tensor has also changed. In this section, instead of rate in the previous section, an increment is used; for

example, the increment of rotation can be written as $\Delta\mathbf{Q} = \dot{\mathbf{Q}}\Delta t$. Let \mathbf{E}_j be a basis in the j th direction of the material frame and let \mathbf{e}_j be the corresponding basis in the spatial frame that is under rigid-body rotation such that they are related by $\mathbf{e}_j = \mathbf{Q}\mathbf{E}_j$. The increment of \mathbf{e}_j can be computed by

$$\Delta\mathbf{e}_j = \mathbf{W}\mathbf{e}_j, \quad (4.145)$$

where $W_{i,j} = (\Delta u_{i,j} - \Delta u_{j,i})/2$ is a component of the spin tensor that represents a rigid-body rotation.

In spatial Cartesian coordinates, the Cauchy stress tensor can be written as

$$\boldsymbol{\sigma} = \sigma_{ij}\mathbf{e}_i \otimes \mathbf{e}_j,$$

where \mathbf{e}_i is a unit basis vector in the i th direction of Cartesian coordinates. In the above equation, the summation rule is used for the repeated indices. The incremental form of the Cauchy stress tensor can be obtained by taking increments of the above equation as

$$\Delta\boldsymbol{\sigma} = \Delta\sigma_{ij}\mathbf{e}_i \otimes \mathbf{e}_j + \sigma_{ij}\Delta\mathbf{e}_i \otimes \mathbf{e}_j + \sigma_{ij}\mathbf{e}_i \otimes \Delta\mathbf{e}_j.$$

In the above equation $\Delta\sigma_{ij}$ is the increment of stress component in the frame that rotates with \mathbf{e}_i , which is the definition of co-rotational rate in Eq. (4.143). Thus, it is indeed the Jaumann stress increment. After using the relation in Eq. (4.145), the stress increment becomes

$$\Delta\boldsymbol{\sigma} = \Delta\sigma_{ij}^J\mathbf{e}_i \otimes \mathbf{e}_j + \sigma_{ij}W_{ik}\mathbf{e}_k \otimes \mathbf{e}_j + \sigma_{ij}\mathbf{e}_i \otimes W_{jk}\mathbf{e}_k.$$

After changing dummy indices, the stress increment can be written in terms of the spatial basis as

$$\Delta\boldsymbol{\sigma} = \left(\Delta\sigma_{ij}^J + \sigma_{kj}W_{ik} - \sigma_{ik}W_{kj} \right) \mathbf{e}_i \otimes \mathbf{e}_j, \quad (4.146)$$

where $\Delta\sigma_{ij}^J$ is the Jaumann or co-rotational Cauchy stress increment, which is the objective rate because it takes an increment of the tensor with respect to the principal axis of the deformation rate tensor. Although the Jaumann stress increment is deficient with large shear strain problems because it produces an artificial oscillation for a simple shear problem, it is the one most frequently used. Due to the small strain assumption, the constitutive relation in the infinitesimal elastoplasticity in Sect. 4.3 can still be used for finite rotational case. The constitutive equation in Eq. (4.120) must be written in the form of the objective rate

$$\Delta\boldsymbol{\sigma}^J = \mathbf{D}^{\text{alg}} : \Delta\boldsymbol{\epsilon}. \quad (4.147)$$

The other two terms in Eq. (4.146) represent the effect of rigid-body rotation.

2. Objective Time Integration

Even if the incremental constitutive relation in Eqs. (4.146) and (4.147) is written using the objective stress rate, it is only accurate for small rigid-body rotations because it is assumed that the spin tensor \mathbf{W} is constant during the increment, which is not true for finite rotation. Hughes and Winget [3] observed an excessive amount of error in the rate form and proposed an algorithm to preserve objectivity for large rotational increments. Their idea is to define the strain increment and spin tensor in the midpoint configuration, which is between load steps t_n and t_{n+1} . In order to rotate the strain increment and spin tensor to the rotation-free midpoint configuration, the following relations have to be used:

$${}^{n+\frac{1}{2}}\mathbf{x} = {}^n\mathbf{x} + \frac{1}{2}\Delta\mathbf{u} = {}^{n+1}\mathbf{x} - \frac{1}{2}\Delta\mathbf{u}.$$

Since the updated Lagrangian formulation will be used for elastoplasticity, it is convenient to use the second relation in the above equation. If the above relation is differentiated with respect to the coordinate direction ${}^{n+1}\mathbf{x}$ at the current configuration,

$$\frac{\partial {}^{n+\frac{1}{2}}\mathbf{x}}{\partial {}^{n+1}\mathbf{x}} = 1 - \frac{1}{2} \frac{\partial \Delta\mathbf{u}}{\partial {}^{n+1}\mathbf{x}} = 1 - \frac{1}{2}\mathbf{L},$$

where $\mathbf{L} = \partial \Delta\mathbf{u} / \partial {}^{n+1}\mathbf{x}$ is the velocity gradient (from the fact that $\Delta\mathbf{u} = \mathbf{v}\Delta t$). The inverse of the above relation can be obtained by

$$\frac{\partial {}^{n+1}\mathbf{x}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} = \left(\mathbf{1} - \frac{1}{2}\mathbf{L} \right)^{-1}.$$

Then, the displacement gradient at the midpoint configuration can be obtained by

$$\frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} = \frac{\partial \Delta\mathbf{u}}{\partial {}^{n+1}\mathbf{x}} \frac{\partial {}^{n+1}\mathbf{x}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} = \mathbf{L} \left(\mathbf{1} - \frac{1}{2}\mathbf{L} \right)^{-1}.$$

Then, the strain increment and spin tensor are defined at the midpoint configuration as

$$\Delta\boldsymbol{\epsilon} = \frac{1}{2} \left(\frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} + \frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}}^T \right), \quad (4.148)$$

$$\mathbf{W} = \frac{1}{2} \left(\frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} - \frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}}^T \right) \quad (4.149)$$

and stress and back stress at the previous load step are updated to the rotation-free configuration by

$${}^n\bar{\boldsymbol{\sigma}} = \mathbf{R}^n \boldsymbol{\sigma} \mathbf{R}^T, \quad {}^n\bar{\boldsymbol{\alpha}} = \mathbf{R}^n \boldsymbol{\alpha} \mathbf{R}^T, \quad (4.150)$$

where $\mathbf{R} = \mathbf{1} + (\mathbf{1} - \frac{1}{2}\mathbf{W})^{-1}\mathbf{W}$, which is orthogonal and incrementally objective, and can be obtained by applying the generalized midpoint rule to $d\mathbf{R}/dt = \mathbf{WR}$. The usual return-mapping algorithm from Sect. 4.3 can be applied to Eq. (4.150) after the transformation.

The following MATLAB program, **rotatedStress**, can be used to update the stress and back stress to the rotation-free configuration. Inputs to the program are the velocity gradient ($\mathbf{L} = \partial \Delta \mathbf{u} / \partial^{n+1} \mathbf{x}$, 3×3 matrix), stress (6×1 vector), and back stress (6×1 vector). Then, the program returns rotated stress and back stress. Once the stress and back stress are rotated, they can be considered as the stress and back stress at the previous load step in the return-mapping procedure.

```
PROGRAM rotatedStress
%
% Rotate stress and back stress to the rotation-free configuration
%
function [stress, alpha] = rotatedStress(L, S, A)
%L = [dui/dxj] velocity gradient
str=[S(1) S(4) S(6);S(4) S(2) S(5);S(6) S(5) S(3)];
alp=[A(1) A(4) A(6);A(4) A(2) A(5);A(6) A(5) A(3)];
factor=0.5;
R = L*inv(eye(3) + factor*L);
W = .5*(R-R');
R = eye(3) + inv(eye(3) - factor*W)*W;
str = R*str*R';
alp = R*alp*R';
stress=[str(1,1) str(2,2) str(3,3) str(1,2) str(2,3) str(1,3)]';
alpha=[alp(1,1) alp(2,2) alp(3,3) alp(1,2) alp(2,3) alp(1,3)]';
```

4.4.3 Incremental Equation for Finite Rotation Elastoplasticity

Even if small strain is assumed in elastoplastic deformation, it is considered to be a large deformation problem due to finite rotations. This is a major difference from the infinitesimal elastoplasticity in Sect. 4.3, in which both small strain and small rotation are assumed. Thus, the deformed geometry must be separated from the undeformed geometry, which is similar to the geometric nonlinear analysis in Chap. 3. Thus, either the total or updated Lagrangian formulation can be used. However, it is inconvenient to express the equilibrium equation of an elastoplastic problem in the total Lagrangian formulation, since the evolution of plastic variables is directly related to the Cauchy stress. Thus, the updated

Lagrangian formulation is a natural choice. It is assumed that the stress and plastic variables as well as deformed geometry up to load step t_n are known, and the same variables and geometry at current load step t_{n+1} are wanted to be computed. In the updated Lagrangian formulation, the energy form can be written at the current geometry as

$$a({}^n\xi; {}^{n+1}\mathbf{u}, \bar{\mathbf{u}}) \equiv \iint_{{}^{n+1}\Omega} \nabla_{{}^{n+1}\mathbf{u}} : {}^{n+1}\boldsymbol{\sigma} d\Omega. \quad (4.151)$$

For a symmetric tensor \mathbf{S} and nonsymmetric tensor \mathbf{L} , it can be shown that $\mathbf{S}:\mathbf{L} = \mathbf{S}:\text{sym}(\mathbf{L})$, where $\text{sym}(\mathbf{L})$ is a symmetric part of \mathbf{L} , i.e., $\text{sym}(\mathbf{L}) = \frac{1}{2}(\mathbf{L} + \mathbf{L}^T)$. Thus, the integrand of the above equation is the same as $\boldsymbol{\epsilon}(\bar{\mathbf{u}}) : {}^{n+1}\boldsymbol{\sigma}$, which is equivalent to the updated Lagrangian formulation in Eq. (4.80). Note that in addition to displacement and stress, the plastic variables from the previous load step are required to calculate energy form at load step t_{n+1} .

Equation (4.151) is a nonlinear function of displacement since the deformed geometry and stress at the current load step are unknown a priori. A linearization is required to solve the nonlinear equation using the Newton–Raphson method iteratively. Let the external load be independent of displacement, that is, conservative. Since the derivatives and integration in Eq. (4.151) are carried out with respect to the current geometry, it is convenient to transform them to the undeformed geometry using the deformation gradient and the Jacobian relation. The energy form can be transformed into an undeformed geometry by

$$a({}^n\xi; {}^{n+1}\mathbf{u}, \bar{\mathbf{u}}) = \iint_{{}^0\Omega} (\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \boldsymbol{\sigma} J d\Omega. \quad (4.152)$$

Note that $\mathbf{F} = \partial^{n+1}\mathbf{x}/\partial\mathbf{X}$, $d^{n+1}\Omega = Jd^0\Omega$, and $J = \det(\mathbf{F})$. The integrand in this equation is the same as $\mathbf{T} : \bar{\mathbf{F}}$, where $\mathbf{T} = J\mathbf{F}^{-1}\boldsymbol{\sigma}$ is the first Piola-Kirchhoff stress tensor and the variation of displacement gradient is $\nabla_0 \bar{\mathbf{u}} = \bar{\mathbf{F}}$. This transformation is temporary as the updated Lagrangian formulation will be recovered at the end of linearization.

Since the constitutive relation is given by the rates of Cauchy stress and engineering strain, the linearization is carried out with respect to Cauchy stress in the undeformed geometry. Note that linearization is equivalent to taking increment. To linearize Eq. (4.152), the increment of the deformation gradient is first written as

$$\Delta\mathbf{F} = \frac{\partial}{\partial\omega} \left[\frac{\partial(\mathbf{x} + \omega\Delta\mathbf{u})}{\partial\mathbf{X}} \right]_{\omega=0} = \frac{\partial\Delta\mathbf{u}}{\partial\mathbf{X}} = \nabla_0\Delta\mathbf{u}.$$

In addition, from the incremental relation of $\mathbf{FF}^{-1} = \mathbf{1}$, the following is given:

$$\Delta(\mathbf{F}^{-1}) = -\mathbf{F}^{-1}\nabla_0\Delta\mathbf{u}\mathbf{F}^{-1} = -\mathbf{F}^{-1}\nabla_{n+1}\Delta\mathbf{u}.$$

The incremental form of the Jacobian of the deformation gradient [4] can be derived by direct linearization of the expression of $|F_{mn}| = (1/6)e_{ijk}e_{rst}F_{ir}F_{js}F_{kt}$ with the identity of $e_{ijk}e_{ijr} = 2\delta_{kr}$:

$$\Delta J = \Delta \det(\mathbf{F}) = J \operatorname{div}(\Delta\mathbf{u}).$$

By using the above two relations, the linearization of the energy form in Eq. (4.152) can be obtained as

$$\begin{aligned} & \Delta[a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}})] \\ &= \Delta \left[\iint_{^0\Omega} (\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \boldsymbol{\sigma} J d\Omega \right] \\ &= \iint_{^0\Omega} [(\nabla_0 \bar{\mathbf{u}} \Delta(\mathbf{F}^{-1})) : \boldsymbol{\sigma} J + (\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \Delta\boldsymbol{\sigma} J + (\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \boldsymbol{\sigma} \Delta J] d\Omega. \end{aligned}.$$

For the notational convenience, the left superscript $n+1$ is omitted for the stress. The first term of the integrands can be simplified as

$$\nabla_0 \bar{\mathbf{u}} \Delta(\mathbf{F}^{-1}) : \boldsymbol{\sigma} J = -\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1} \nabla_{n+1} \Delta\mathbf{u} : \boldsymbol{\sigma} J = -\nabla_{n+1} \bar{\mathbf{u}} \nabla_{n+1} \Delta\mathbf{u} : \boldsymbol{\sigma} J.$$

In order to make the above expression convenient for the following derivations, the following rearrangement can be performed. For second-order tensors, the following relation can be satisfied: $(\mathbf{AB}):\mathbf{S} = \mathbf{A}:\mathbf{SB}^T$. This can be easily shown using the index notation as $A_{ik}B_{kj}S_{ij} = A_{ik}S_{ij}B_{kj}$. Thus, the above term can be rewritten as

$$\nabla_0 \bar{\mathbf{u}} \Delta(\mathbf{F}^{-1}) : \boldsymbol{\sigma} J = -\nabla_{n+1} \bar{\mathbf{u}} : \boldsymbol{\sigma} \nabla_{n+1} \Delta\mathbf{u}^T J.$$

The second term can be simplified using the Jaumann rate as

$$(\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \Delta\boldsymbol{\sigma} J = \nabla_{n+1} \bar{\mathbf{u}} : (\Delta\boldsymbol{\sigma}^J + \mathbf{W}\boldsymbol{\sigma} - \boldsymbol{\sigma}\mathbf{W}) J.$$

Finally, the last terms can be simplified as

$$(\nabla_0 \bar{\mathbf{u}} \mathbf{F}^{-1}) : \boldsymbol{\sigma} \Delta J = \nabla_{n+1} \bar{\mathbf{u}} : \boldsymbol{\sigma} \operatorname{div}(\Delta\mathbf{u}) J.$$

Thus, after combining these three terms, the linearization of the energy form can be written as

$$\begin{aligned} & \Delta[a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}})] \\ &= \iint_{^{n+1}\Omega} \nabla_{n+1} \bar{\mathbf{u}} : [\Delta\boldsymbol{\sigma}^J + \mathbf{W}\boldsymbol{\sigma} - \boldsymbol{\sigma}\mathbf{W} + \boldsymbol{\sigma} \operatorname{div}(\Delta\mathbf{u}) - \boldsymbol{\sigma} (\nabla_{n+1} \Delta\mathbf{u})^T] d\Omega. \end{aligned}$$

In the elastoplastic constitutive relation, the Jaumann rate is given as $\Delta\boldsymbol{\sigma}^J = \mathbf{D}^{\text{alg}}:\Delta\boldsymbol{\epsilon}$, and the spin tensor can be written in terms of incremental

displacement gradient. Thus, the integrands of the above equation can be explicitly written in terms of incremental displacement gradient. For example, the $\mathbf{W}\boldsymbol{\sigma}$ term in the integrand can be expressed in terms of displacement increment, as

$$\begin{aligned} W_{im}\sigma_{mj} &= \frac{1}{2} \left(\frac{\partial \Delta u_i}{\partial x_m} - \frac{\partial \Delta u_m}{\partial x_i} \right) \sigma_{mj} = \frac{1}{2} \sigma_{mj} (\delta_{ik} \delta_{ml} - \delta_{mk} \delta_{il}) \frac{\partial \Delta u_k}{\partial x_l} \\ &= \frac{1}{2} (\sigma_{lj} \delta_{ik} - \sigma_{kj} \delta_{il}) [\nabla_{n+1} \Delta \mathbf{u}]_{kl} \end{aligned}$$

Thus, after rearranging terms, the linearization of the energy form can be written as

$$\begin{aligned} \Delta [a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}})] \\ = \iint_{^{n+1}\Omega} [\nabla_{n+1} \bar{\mathbf{u}} : (\mathbf{D}^{\text{alg}} - \mathbf{D}^*) : \nabla_{n+1} \Delta \mathbf{u} + \boldsymbol{\sigma} : \boldsymbol{\eta}(\bar{\mathbf{u}}, \Delta \mathbf{u})] d\Omega. \\ \equiv a^*(^n\xi, ^{n+1}\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) \end{aligned} \quad (4.153)$$

where the same notation as in Eq. (4.123) is used even if the above form is for the updated Lagrangian formulation. In the above equation, $\boldsymbol{\sigma} : \boldsymbol{\eta}(\Delta \mathbf{u}, \bar{\mathbf{u}})$ is the initial stiffness term with $\boldsymbol{\eta}(\Delta \mathbf{u}, \bar{\mathbf{u}}) = \text{sym}(\nabla_{n+1} \bar{\mathbf{u}}^T \nabla_{n+1} \Delta \mathbf{u})$, and

$$D_{ijkl}^* = -\sigma_{ij} \delta_{kl} + \frac{1}{2} (\sigma_{il} \delta_{jk} + \sigma_{jl} \delta_{ik} + \sigma_{ik} \delta_{jl} + \sigma_{jk} \delta_{il}) \quad (4.154)$$

represents the rotational effect of the Cauchy stress tensor. Note that the linearized energy form in Eq. (4.153) is not symmetric because \mathbf{D}^* is not symmetric. With Eq. (4.153), the incremental equation of the Newton–Raphson method can be obtained. Since the reference configuration is the current, unknown one, the last iteration of the current load step is chosen as the reference configuration. Let the current load step be t_{n+1} and let the current iteration counter be $k+1$. Assuming that the external force is independent of displacement, the incremental equation is obtained as

$$a^*(^n\xi, ^{n+1}\mathbf{u}; \Delta \mathbf{u}^{k+1}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}) - a(^n\xi; ^{n+1}\mathbf{u}, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (4.155)$$

The above iteration is repeated until the residual term on the right-hand side vanishes. After the displacement increment is computed by solving linear systems of equations using Eq. (4.155), the return-mapping procedure is carried out to obtain the status of stress for each integration point, including internal plastic variables.

Note that the linearized energy form in Eq. (4.153) is not an accurate Jacobian relation because it does not include the effect of rotating stress and back stress to the rotation-free midpoint configuration as in Eq. (4.150). For more accurate expression of the linearized energy form, the readers are referred to Fish and someone [5].

4.4.4 Computational Implementation of Finite Rotation

The implementation of elastoplasticity with finite rotation is an extension of the infinitesimal elastoplasticity in Sect. 4.3.6. Thus, instead of explaining the entire process, only different parts will be presented. Let the current load step be t_{n+1} and the current iteration counter be $k+1$. Same with infinitesimal elastoplasticity, the available variables are ${}^n\boldsymbol{\sigma}$, ${}^n\boldsymbol{\alpha}$, ${}^n\mathbf{e}_p$, and displacement increment $\Delta\mathbf{u}^k$.

In the updated Lagrangian formulation with objective time integration, strain is defined with respect to the geometry at $t_{n+\frac{1}{2}}$. This can be achieved by calculating velocity gradient at t_n first and transform it to $t_{n+\frac{1}{2}}$. The Jacobian relation at t_n can be defined as

$${}^n\mathbf{J} = \begin{bmatrix} \frac{\partial {}^n\mathbf{x}_1}{\partial \xi} & \frac{\partial {}^n\mathbf{x}_2}{\partial \xi} \\ \frac{\partial {}^n\mathbf{x}_1}{\partial \eta} & \frac{\partial {}^n\mathbf{x}_2}{\partial \eta} \end{bmatrix},$$

where $({}^n\mathbf{x}_1, {}^n\mathbf{x}_2)$ is the material coordinate at load step t_n , while (ξ, η) is the corresponding coordinate in the reference element. Then, the derivatives of shape functions with respect to $({}^n\mathbf{x}_1, {}^n\mathbf{x}_2)$ can be written in terms of the derivatives with respect to (ξ, η) by

$$\begin{bmatrix} \frac{\partial N_I}{\partial {}^n\mathbf{x}_1} \\ \frac{\partial N_I}{\partial {}^n\mathbf{x}_2} \end{bmatrix} = {}^n\mathbf{J}^{-1} \begin{bmatrix} \frac{\partial N_I}{\partial \xi} \\ \frac{\partial N_I}{\partial \eta} \end{bmatrix}.$$

Using the above derivatives of shape function, the velocity gradient can be interpolated by

$$\left(\frac{\partial \Delta\mathbf{u}}{\partial {}^n\mathbf{x}} \right)_{ij} = \sum_{I=1}^4 \frac{\partial N_I}{\partial {}^n\mathbf{x}_j} \Delta\mathbf{d}_{Ii}.$$

Then, from the relation that ${}^{n+\frac{1}{2}}\mathbf{x} = {}^n\mathbf{x} + (1/2) \Delta\mathbf{u}$, the incremental deformation gradient can be defined as

$$\frac{\partial {}^{n+1/2}\mathbf{x}}{\partial {}^n\mathbf{x}} = \frac{\partial ({}^n\mathbf{x} + \frac{1}{2}\Delta\mathbf{u})}{\partial {}^n\mathbf{x}} = \mathbf{1} + \frac{1}{2} \frac{\partial \Delta\mathbf{u}}{\partial {}^n\mathbf{x}}.$$

Then the velocity gradient at $t_{n+\frac{1}{2}}$ can be obtained by

$$\nabla_{n+\frac{1}{2}}\Delta\mathbf{u} = \frac{\partial \Delta\mathbf{u}}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} = \frac{\partial \Delta\mathbf{u}}{\partial {}^n\mathbf{x}^n} \cdot \frac{\partial {}^n\mathbf{x}^n}{\partial {}^{n+\frac{1}{2}}\mathbf{x}} = \frac{\partial \Delta\mathbf{u}}{\partial {}^n\mathbf{x}^n} \left(1 + \frac{1}{2} \frac{\partial \Delta\mathbf{u}}{\partial {}^n\mathbf{x}^n} \right)^{-1}.$$

Using the above velocity gradient at the midpoint configuration, the following steps will update stress and plastic variables and will solve for incremental displacement:

1. Strain increment and spin tensor at the midpoint

$$\Delta \boldsymbol{\epsilon} = \text{sym}(\nabla_{n+\frac{1}{2}} \Delta \mathbf{u}), \quad \mathbf{W} = \text{skew}(\nabla_{n+\frac{1}{2}} \Delta \mathbf{u}).$$

2. Rotation matrix $\mathbf{R} = \mathbf{1} + (\mathbf{1} - \frac{1}{2}\mathbf{W})^{-1}\mathbf{W}$.
3. Rotated stress and back stress

$$\bar{\boldsymbol{\sigma}}^n = \mathbf{R}\boldsymbol{\sigma}^n\mathbf{R}^T \quad \bar{\boldsymbol{\alpha}}^n = \mathbf{R}\boldsymbol{\alpha}^n\mathbf{R}^T.$$

4. Return-mapping with ${}^n\bar{\boldsymbol{\sigma}}$ and ${}^n\bar{\boldsymbol{\alpha}}$ (use the procedures in Sect. 4.3.6).

$$5. \text{ Internal force } \{\mathbf{f}^{\text{int}}\} = \sum_{K=1}^{\text{NG}} \left([\mathbf{B}]^T \{{}^{n+1}\boldsymbol{\sigma}\} |\mathbf{J}| \right)_K \omega_K.$$

$$6. \text{ Tangent stiffness } [\mathbf{K}_T] = \sum_{K=1}^{\text{NG}} \left([\mathbf{B}]^T [\mathbf{D}^{\text{alg}} - \mathbf{D}^*] [\mathbf{B}] |\mathbf{J}| \right)_K \omega_K.$$

$$\mathbf{D}^* = \begin{bmatrix} -\sigma_{11} & \sigma_{11} & \sigma_{11} & -\sigma_{12} \\ \sigma_{22} & -\sigma_{22} & \sigma_{22} & -\sigma_{12} \\ \sigma_{33} & \sigma_{33} & -\sigma_{33} & 0 \\ -\sigma_{12} & -\sigma_{12} & 0 & -\frac{1}{2}(\sigma_{11} + \sigma_{22}) \end{bmatrix}.$$

$$7. \text{ Initial stiffness } [\mathbf{K}_S] = \sum_{K=1}^{\text{NG}} \left([\mathbf{B}_G]^T [\boldsymbol{\Sigma}] [\mathbf{B}_G] |\mathbf{J}| \right)_K \omega_K.$$

$$[\mathbf{B}_G] = \begin{bmatrix} N_{1,1} & 0 & N_{4,1} & 0 \\ N_{1,2} & 0 & N_{4,2} & 0 \\ 0 & N_{1,1} & 0 & N_{4,1} \\ 0 & N_{1,2} & 0 & N_{4,2} \end{bmatrix} \quad [\boldsymbol{\Sigma}] = \begin{bmatrix} \sigma_{11} & \sigma_{12} & 0 & 0 \\ \sigma_{12} & \sigma_{22} & 0 & 0 \\ 0 & 0 & \sigma_{11} & \sigma_{12} \\ 0 & 0 & \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

8. Solve for incremental displacement

$$[\mathbf{K}_T + \mathbf{K}_S]\{\delta \mathbf{d}^{k+1}\} = \{\mathbf{f}^{\text{ext}}\} - \{\mathbf{f}^{\text{int}}\}.$$

9. Update displacement

$${}^{n+1}\mathbf{d}^{k+1} = {}^{n+1}\mathbf{d}^k + \delta \mathbf{d}^{k+1}$$

$$\Delta \mathbf{d}^{k+1} = \Delta \mathbf{d}^k + \delta \mathbf{d}^{k+1}.$$

If the residual force does not vanish, the iteration counter k is changed to $k + 1$ and the above procedures are repeated. When the iteration converges, updated stress

and plastic variables are stored and n is changed to $n+1$, and the next load step starts. Note that the Newton–Raphson iteration solves for displacement increment between iterations k and $k+1$, $\delta\mathbf{d}^{k+1}$, which is different from the displacement increment between load steps n and $n+1$, $\Delta\mathbf{d}^{k+1}$. Thus, it is important to update both \mathbf{d} and $\Delta\mathbf{d}$ using $\delta\mathbf{d}$.

It is reminded that the elastoplasticity with finite rotation is only valid when the rigid-body rotation is large but the strain is small so that additive decomposition between elastic and plastic strains is possible. As the strain becomes large, it is not accurate to assume the additive decomposition. In addition, the assumption of linear constitutive relation between stress and elastic strain is not valid anymore. Thus, more general model is required for elastoplasticity with finite deformation.

Example 4.19 (Objective Time Integration) A plane strain square undergoes rigid-body rotation and shear deformation such that the velocity gradient at each load step is given as

$$\left[\frac{\partial \Delta \mathbf{u}}{\partial \mathbf{x}} \right] = \begin{bmatrix} 0 & 0.024 & 0 \\ -0.02 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Compare the stresses from infinitesimal deformation and finite rotation elastoplasticity by plotting shear stress τ_{12} vs. shear strain γ_{12} curve for 15 load increments. Compare all stress components at the last load step. Assume linear isotropic hardening with the following material properties: $E = 24$ GPa, $\nu = 0.2$, $H = 1.0$ GPa, and $\sigma_Y^0 = 200\sqrt{3}$ MPa.

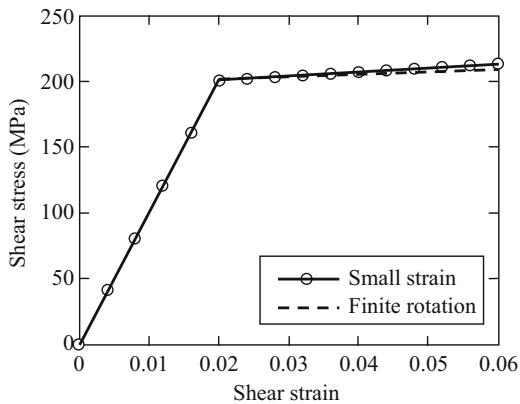
Solution The given velocity gradient includes spin tensor (skew-symmetric part) and rate of deformation part (symmetric part). After removing the rigid-body rotation, the strain increment becomes $\Delta\gamma_{12} = 0.004$. The latter is related to the strain increment, while the former represents the rigid-body rotation. Below is the list of MATLAB program that solves for the shear deformation problem with and without considering the rotational effect. The variable “**stress**” is the updated stress from infinitesimal deformation assumption, while “**stressR**” is the one from finite rotation assumption. At the last load increment, these two stresses are

$$\text{stress} = [0 \ 0 \ 0 \ 212.9 \ 0 \ 0]^T$$

$$\text{stressR} = [43.4 \ -43.4 \ 0 \ 208.2 \ 0 \ 0]^T.$$

The difference in shear stress σ_{12} is relatively small. However, in the finite rotational formulation, the normal stresses are developed due to the rotation of the reference frame. Figure 4.21 show the shear stress vs. shear strain curve.

Fig. 4.21 Stress-strain curve for elastoplastic shear deformation



```

%
% Example 4.19 - Shear deformation of a square (finite rotation)
%
clear;
Young = 24000; nu=0.2; mu=Young/2/(1+nu); lambda=nu*Young/((1+nu)*(1-2*nu));
beta = 0; H = 1000; sY = 200*sqrt(3);
mp = [lambda mu beta H sY];
Iden=[1 1 1 0 0 0]';
D=2*mu*eye(6) + lambda*Iden*Iden';
D(4,4)=mu; D(5,5)=mu; D(6,6)=mu;
L = zeros(3,3);
stressN=[0 0 0 0 0 0]';
deps=[0 0 0 0 0 0]';
alphaN=[0 0 0 0 0 0]';
epN=0;
stressRN=stressN; alphaRN=alphaN; epRN=epN;
for i=1:15
    deps(4) = 0.004; L(1,2) = 0.024; L(2,1) = -0.02;
    [stressRN, alphaRN] = rotatedStress(L, stressRN, alphaRN);
    [stressR, alphaR, epR]=combHard(mp,D,deps,stressRN,alphaRN,epRN);
    [stress, alpha, ep]=combHard(mp,D,deps,stressN,alphaN,epN);
    X(i) = i*deps(4); Y1(i) = stress(4); Y2(i) = stressR(4);
    stressN = stress; alphaN = alpha; epN = ep;
    stressRN = stressR; alphaRN = alphaR; epRN = epR;
end
X = [0 X]; Y1=[0 Y1]; Y2=[0 Y2]; plot(X,Y1,X,Y2);

```

4.5 Finite Deformation Elastoplasticity with Hyperelasticity

Many difficulties associated with finite deformation can be resolved by using a new plasticity model, where the constitutive relation is based on hyperelasticity. Remember that hyperelasticity is a general model for nonlinear elastic relation between stress and strain. In elastoplasticity, the hyperelastic relation can be used between stress and elastic strain because the plastic strain cannot produce stress. However, this requires a complete reformulation of plasticity. Among many elastoplasticity theories with finite deformation, Simo [6] proposed a method that is close to the infinitesimal plasticity in Sect. 4.3. This formulation will be discussed in this section.

4.5.1 Multiplicative Decomposition

The theory of multiplicative plasticity proposed by Lee [7] is used to overcome the assumption of small elastic strain in the theory of classical infinitesimal plasticity, which uses an additive decomposition of the strain and its rate. The computational framework of this theory is proposed by Simo, which preserves the conventional return-mapping algorithm in the principal stress space. Although the new plasticity theory is based on completely different assumptions, at the end, it becomes surprisingly close to the classical infinitesimal elastoplasticity at the implementation level.

In finite deformation, the deformation gradient plays an important role. The deformation gradient $\mathbf{F}(\mathbf{X})$ relates a vector $d\mathbf{X}$ in the undeformed configuration to a vector $d\mathbf{x}$ in the deformed configuration. This theory assumes that there exists an imaginary, intermediate configuration, Ω_p , such that $\mathbf{F}(\mathbf{X})$ can be decomposed into $\mathbf{F}^e(\mathbf{X})$ and $\mathbf{F}^p(\mathbf{X})$, as shown in Fig. 4.22. Since stress is only dependent on

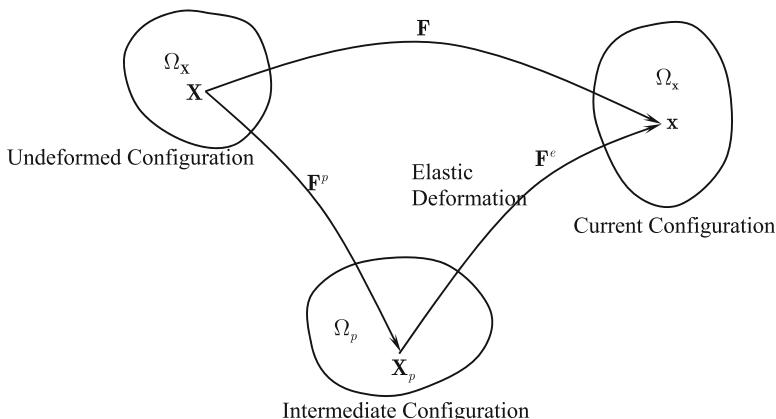


Fig. 4.22 Multiplicative decomposition of deformation

elastic strain, if stress is removed from the current configuration, the intermediate configuration can be obtained. Of course, there is no guarantee that a continuous intermediate configuration can be obtained. However, at least locally at a material point \mathbf{X} , this can be achievable. Thus, $\mathbf{F}(\mathbf{X})$ takes the following form of the local multiplicative decomposition:

$$\mathbf{F}(\mathbf{X}) = \mathbf{F}^e(\mathbf{X})\mathbf{F}^p(\mathbf{X}) \quad (4.156)$$

where $\mathbf{F}^p(\mathbf{X})$ denotes the deformation through the intermediate configuration, which is related to the internal variables, and $\mathbf{F}^{e-1}(\mathbf{X})$ defines the local, stress-free, unloaded process. In this decomposition, it is clear that plastic deformation leads to the intermediate configuration, from which elastic deformation leads to the current configuration. This does not mean that physically plastic deformation occurs first, followed by elastic deformation. It is a mathematical decomposition of the finite deformation elastoplasticity. In fact, plastic deformation simultaneously occurs with elastic deformation.

4.5.2 Finite Deformation Elastoplasticity

In the infinitesimal elastoplasticity, it was discussed that the Cauchy stress needs to be used because the elastoplastic deformation occurs at the current configuration. In the finite deformation, not only the evolution of stress but also that of the geometry needs to be accounted for. Thus, it would be convenient if the following stress measure is used:

$$\boldsymbol{\tau} = J\boldsymbol{\sigma}, \quad (4.157)$$

where $\boldsymbol{\tau}$ is the Kirchhoff stress and $J=|\mathbf{F}|$ is the Jacobian relation between deformed and undeformed configurations. Then the constitutive relation can cover the effects of both stress and change in domain. For example, an integral over the deformed geometry can be converted into an integral over the undeformed geometry by

$$\iint_{n+1}\Omega \boldsymbol{\sigma} d\Omega = \iint_0\Omega \boldsymbol{\tau} d\Omega.$$

However, the Kirchhoff stress is different from the first Piola-Kirchhoff stress \mathbf{P} and they are related by $\boldsymbol{\tau}=\mathbf{F}\mathbf{P}$ (refer to Sect. 3.2.5). Note that when deformation is infinitesimal, the Kirchhoff stress approaches the Cauchy stress. The elastic domain is defined using the Kirchhoff stress as

$$E = \{(\boldsymbol{\tau}, \mathbf{q}) | f(\boldsymbol{\tau}, \mathbf{q}) \leq 0\}, \quad (4.158)$$

where \mathbf{q} is the vector of stress-like plastic variables that characterize the hardening property of the material. The yield function $f(\boldsymbol{\tau}, \mathbf{q})$ in Eq. (4.158) is an isotropic function of $\boldsymbol{\tau}$ due to the principle of objectivity, that is, the yield function does not depend on the orientation of the stress or on plastic variables, such as the von Mises yield function.

In thermodynamics, a free energy is the irreversible energy that is available for doing thermodynamic work. In mechanical systems, it is assumed that the free energy locally depends on $\mathbf{F}^e(\mathbf{X})$ only, since the free energy represents stored energy through elastic deformation. The free energy is independent of the orientation, in the same context of the yield function. Thus, similar to the strain energy density in Chap. 3, it can be defined using either right or left Cauchy-Green deformation tensor. Since the plastic evolution occurs at the current configuration, it would be convenient to define it using the left Cauchy-Green deformation tensor \mathbf{b} . Since the deformation is composed of elastic and plastic parts, the free energy is defined as

$$\psi = \psi(\mathbf{b}^e, \xi), \quad (4.159)$$

where $\mathbf{b}^e \equiv \mathbf{F}^e \mathbf{F}^{eT}$ is the elastic left Cauchy-Green deformation tensor, and ξ is the vector of strain-like plastic variables conjugate to \mathbf{q} in the sense that $\mathbf{q} = -\partial\psi/\partial\xi$. In the viewpoint of classical plasticity, \mathbf{q} can be seen as back stress, and ξ as the effective plastic strain. Here, the general notation \mathbf{q} and ξ is kept until a specific plastic model is introduced.

The stress-strain relation can be obtained by defining a local dissipation function and using the principle of maximum dissipation, which says that the plastic deformation occurs in the direction that maximizes the dissipation function. By ignoring thermoelastic parts, local dissipation function D is defined per unit reference volume as the difference between the rate of stress work and the rate of free energy change as

$$D \equiv \boldsymbol{\tau} : \mathbf{d} - \frac{d}{dt}\psi(\mathbf{b}^e, \xi) \geq 0, \quad (4.160)$$

where $\mathbf{d} = \text{sym}(\mathbf{L})$ denotes the rate of deformation and $\mathbf{L} = \dot{\mathbf{F}}\mathbf{F}^{-1}$ is the velocity gradient. The time rate of the free energy can be obtained using the chain rule and the time rate of $\mathbf{b}^e = \mathbf{F}\mathbf{C}^p \mathbf{F}^T$, with $\mathbf{C}^p = \mathbf{F}^{pT} \mathbf{F}^p$ as

$$\dot{\mathbf{b}}^e = \mathbf{L}\mathbf{b}^e + \mathbf{b}^e\mathbf{L}^T + \mathbf{L}_v(\mathbf{b}^e), \quad (4.161)$$

where \mathbf{C}^p is the plastic, right Cauchy-Green tensor, and $\mathbf{L}_v(\mathbf{b}^e)$ is referred to as the Lie derivative of \mathbf{b}^e , which is obtained by pulling \mathbf{b}^e back to the undeformed configuration and, after taking a time derivative, pushing \mathbf{b}^e forward to the current configuration (see Example 4.20). Considering $\dot{\mathbf{b}}^e$ as the rate of elastic strain, the first two terms on the right-hand side correspond to the total rate of strain, while the

Lie derivative corresponds to the negative of the rate of plastic strain. Thus, the above equation is suitable for the elastic predictor and plastic correction algorithm.

In order to obtain the constitutive and evolution relations, it is necessary to expand the local dissipation function D as

$$\begin{aligned} D &\equiv \boldsymbol{\tau} : \mathbf{d} - \frac{d}{dt}\psi(\mathbf{b}^e, \dot{\boldsymbol{\xi}}) \\ &= \boldsymbol{\tau} : \mathbf{d} - \frac{\partial\psi}{\partial\mathbf{b}^e} : \dot{\mathbf{b}}^e - \frac{\partial\psi}{\partial\dot{\boldsymbol{\xi}}} \cdot \dot{\boldsymbol{\xi}} \\ &= \boldsymbol{\tau} : \mathbf{d} - \frac{\partial\psi}{\partial\mathbf{b}^e} : (\mathbf{L}\mathbf{b}^e + \mathbf{b}^e\mathbf{L}^T + \mathbf{L}_v(\mathbf{b}^e)) + \mathbf{q} \cdot \dot{\boldsymbol{\xi}} \\ &= \boldsymbol{\tau} : \mathbf{d} - 2\frac{\partial\psi}{\partial\mathbf{b}^e}\mathbf{b}^e : \mathbf{L} + \left(2\frac{\partial\psi}{\partial\mathbf{b}^e}\mathbf{b}^e\right) : \left[-\frac{1}{2}\mathbf{L}_v(\mathbf{b}^e)\mathbf{b}^{e-1}\right] + \mathbf{q} \cdot \dot{\boldsymbol{\xi}} \\ &= \left(\boldsymbol{\tau} - 2\frac{\partial\psi}{\partial\mathbf{b}^e}\mathbf{b}^e\right) : \mathbf{d} + \left(2\frac{\partial\psi}{\partial\mathbf{b}^e}\mathbf{b}^e\right) : \left[-\frac{1}{2}\mathbf{L}_v(\mathbf{b}^e)\mathbf{b}^{e-1}\right] + \mathbf{q} \cdot \dot{\boldsymbol{\xi}} \geq 0 \end{aligned} . \quad (4.162)$$

For symmetric matrices, the property $\mathbf{A}:\mathbf{BC} = \mathbf{AC}:\mathbf{B}$ is used, and the skew-symmetric part of \mathbf{L} (the spin tensor) vanishes by multiplying it with the symmetric matrix; i.e., for a symmetric tensor \mathbf{S} , $\mathbf{S}:\mathbf{L} = \mathbf{S}:\mathbf{d}$. Inequality in Eq. (4.162) holds for all admissible stresses and plastic variables. When the material is in the elastic range, the rate of plastic variables becomes 0, i.e., $\mathbf{L}_v(\mathbf{b}^e) = \dot{\boldsymbol{\xi}} = 0$. Then the first term on the right-hand side of the last equation must be 0 because the dissipation function must be nonnegative for arbitrary \mathbf{d} . For example, if the term yields a positive dissipation with \mathbf{d}^1 , then it would yield a negative dissipation for $\mathbf{d}^2 = -\mathbf{d}^1$. Thus, the only possible way is that the coefficient of \mathbf{d} in the above equation must vanish. Thus, the following constitutive relations and reduced form of dissipation inequality can be obtained:

$$\boldsymbol{\tau} = 2\frac{\partial\psi}{\partial\mathbf{b}^e}\mathbf{b}^e, \quad (4.163)$$

$$D = \boldsymbol{\tau} : \left[-\frac{1}{2}\mathbf{L}_v(\mathbf{b}^e)\mathbf{b}^{e-1}\right] + \mathbf{q} \cdot \dot{\boldsymbol{\xi}} \geq 0. \quad (4.164)$$

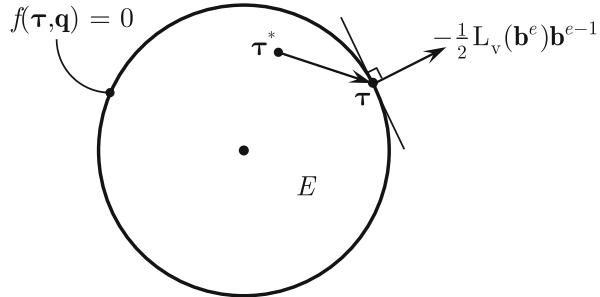
Equation (4.163) provides a stress-strain relation in terms of the Kirchhoff stress and elastic left Cauchy-Green tensor. Note that this relation is given in the form of hyperelasticity, not in the rate form.

In the principle of maximum dissipation, the state variables $\{\boldsymbol{\tau}, \mathbf{q}\}$ maximize the dissipation function D for given rate $\{\mathbf{L}_v(\mathbf{b}^e), \dot{\boldsymbol{\xi}}\}$ of plastic variables. This is equivalent to say that for all possible state variables within the elastic domain, the following inequality holds:

$$D = (\boldsymbol{\tau} - \boldsymbol{\tau}^*) : \left[-\frac{1}{2}\mathbf{L}_v(\mathbf{b}^e)\mathbf{b}^{e-1}\right] + (\mathbf{q} - \mathbf{q}^*) \cdot \dot{\boldsymbol{\xi}} \geq 0, \quad \forall \{\boldsymbol{\tau}^*, \mathbf{q}^*\} \in E. \quad (4.165)$$

The two terms (stress and plastic variables) in the above equation can be considered separately. For the stress term, for example, the Lie derivative term must be in the

Fig. 4.23 Principle of maximum dissipation



normal direction to the elastic domain in order to make the dissipation nonnegative for arbitrary τ^* (see Fig. 4.23). Thus, the above dissipation inequality satisfies if and only if the Lie derivative term is parallel to the normal direction to the elastic domain E . This is similar to the classical plasticity in which the rate of plastic strain is normal to the yield surface. Thus, the evolution equations can be obtained by using the normal property and plastic consistency parameter γ , as follows:

$$-\frac{1}{2}L_v \mathbf{b}^e = \gamma \frac{\partial f(\tau, \mathbf{q})}{\partial \tau} \mathbf{b}^e. \quad (4.166a)$$

$$\dot{\xi} = \gamma \frac{\partial f(\tau, \mathbf{q})}{\partial \mathbf{q}}. \quad (4.166b)$$

$$\gamma \geq 0, \quad f(\tau, \mathbf{q}) \leq 0, \quad \gamma f(\tau, \mathbf{q}) = 0. \quad (4.166c)$$

The first two equations become hardening models for plastic variables. The last equation is the same as the Kuhn-Tucker condition of the classical elastoplasticity problem such that $\gamma=0$ when the material is elastic, while $f=0$ when it is plastic.

Example 4.20 (Time Derivative of the Elastic Left Cauchy-Green Tensor) Derive the time derivative of the elastic left Cauchy-Green tensor in Eq. (4.161). Define an appropriate form of the Lie derivative.

Solution From the relation of $\mathbf{F}^e = \mathbf{FF}^{p-1}$, \mathbf{b}^e can be written as

$$\mathbf{b}^e = (\mathbf{FF}^{p-1})(\mathbf{F}^{p-1}\mathbf{T}^T) = \mathbf{F}(\mathbf{F}^{p-1}\mathbf{F}^p)^{-1}\mathbf{F}^T = \mathbf{FC}^{p-1}\mathbf{F}^T.$$

The time rate of \mathbf{b}^e can then be written as

$$\dot{\mathbf{b}}^e = \dot{\mathbf{F}}\mathbf{C}^{p-1}\mathbf{F}^T + \mathbf{F}\mathbf{C}^{p-1}\dot{\mathbf{F}}^T + \mathbf{F}\frac{d}{dt}(\mathbf{C}^{p-1})\mathbf{F}^T.$$

Using the velocity gradient $\mathbf{L} = \dot{\mathbf{F}}\mathbf{F}^{-1}$, the above equation can be written as

$$\begin{aligned} \dot{\mathbf{b}}^e &= \mathbf{L}\mathbf{FC}^{p-1}\mathbf{F}^T + \mathbf{F}\mathbf{C}^{p-1}\mathbf{F}^T\mathbf{L}^T + \mathbf{F}\frac{d}{dt}(\mathbf{C}^{p-1})\mathbf{F}^T \\ &= \mathbf{L}\mathbf{b}^e + \mathbf{b}^e\mathbf{L}^T + \mathbf{F}\frac{d}{dt}(\mathbf{C}^{p-1})\mathbf{F}^T. \end{aligned}$$

By comparing with Eq. (4.161), it is clear that the Lie derivative can be defined as

$$L_v(\mathbf{b}^e) = \mathbf{F} \frac{d}{dt} (\mathbf{C}^{p-1}) \mathbf{F}^T.$$

From the observation that \mathbf{C}^{p-1} is obtained by transforming the Eulerian tensor \mathbf{b}^e to the undeformed configuration (pull-back operation),

$$\mathbf{C}^{p-1} = \mathbf{F}^{-1} \mathbf{b}^e \mathbf{F}^{-T}.$$

The physical meaning of Lie derivative is (a) pulling back \mathbf{b}^e to the undeformed configuration, (b) differentiating it at the undeformed configuration, and (c) pushing forward it to the current configuration. ■

4.5.3 Time Integration

The rate form elastoplastic evolution in the previous section needs to be integrated to calculate plastic variables at a given load step. Let the system be converged at load step t_n such that all variables, $\{\mathbf{F}, \mathbf{b}^e, \boldsymbol{\xi}\}$, are known. At new load step t_{n+1} , the Newton–Raphson iteration provides displacement increment $\Delta\mathbf{u}$. Then, the objective is to update the variables $\{\mathbf{F}^{n+1}, \mathbf{b}^e, \boldsymbol{\xi}^{n+1}\}$ at load step t_{n+1} . Note that \mathbf{b}^e is a primary variable instead of stress in the classical plasticity model. Once \mathbf{b}^e is known, the stress can be calculated by differentiating the free energy with respect to \mathbf{b}^e . Thus, in this model, strain is updated, rather than stress. In the following derivations, the left superscript $n+1$ is omitted when it is clear in the context.

Deformation between load step t_n and t_{n+1} can be represented using the following relative deformation gradient:

$$\mathbf{f}(\mathbf{x}) = \frac{\partial^{n+1}\mathbf{x}}{\partial^n\mathbf{x}} = \mathbf{1} + \nabla_n \Delta\mathbf{u}, \quad (4.167)$$

which can be obtained by differentiating $^{n+1}\mathbf{x} = ^n\mathbf{x} + \Delta\mathbf{u}$ with respect to $^n\mathbf{x}$. The deformation gradient at time t_{n+1} is then $^{n+1}\mathbf{F}(\mathbf{X}) = \mathbf{f}(\mathbf{x})^n\mathbf{F}(\mathbf{X})$. The first-order system of evolution equations can be obtained by inserting Eq. (4.166) into Eq. (4.161) as follows:

$$\dot{\mathbf{b}}^e = [\mathbf{L}\mathbf{b}^e + \mathbf{b}^e \mathbf{L}^T] - 2\gamma \frac{\partial f(\boldsymbol{\tau}, \mathbf{q})}{\partial \boldsymbol{\tau}} \mathbf{b}^e, \quad (4.168a)$$

$$\dot{\boldsymbol{\xi}} = \gamma \frac{\partial f(\boldsymbol{\tau}, \mathbf{q})}{\partial \mathbf{q}}, \quad (4.168b)$$

$$\gamma \geq 0, \quad f(\boldsymbol{\tau}, \mathbf{q}) \leq 0, \quad \gamma f(\boldsymbol{\tau}, \mathbf{q}) = 0. \quad (4.168c)$$

The above differential equations need to be solved using time integration with the following initial conditions: $\{\mathbf{f}, \mathbf{b}^e, \boldsymbol{\xi}\} = \{\mathbf{1}, ^n\mathbf{b}^e, ^n\boldsymbol{\xi}\}$.

Kirchhoff stress and stress-like plastic variables are derived using the primary variables by the following constitutive law:

$$\boldsymbol{\tau} = 2 \frac{\partial \psi(\mathbf{b}^e, \xi)}{\partial \mathbf{b}^e} \mathbf{b}^e, \quad (4.169)$$

$$\mathbf{q} = -\frac{\partial \psi(\mathbf{b}^e, \xi)}{\partial \xi}, \quad (4.170)$$

This process represents the major difference from the classical plasticity model, where an evolution equation is expressed in terms of stress and back stress.

As in classical infinitesimal plasticity, the evolution equations in Eq. (4.168) can be split into a trial elastic state and a plastic return-mapping. In Eq. (4.168a), the first term, $[\mathbf{L}\mathbf{b}^e + \mathbf{b}^e \mathbf{L}^T]$, corresponds to the trial elastic state in which the displacement increment in \mathbf{L} is used to increase \mathbf{b}^e . In practice, this can be achieved using the relative deformation gradient. From the given displacement increment, the relative deformation gradient can first be calculated using Eq. (4.167). The trial elastic state can then be obtained by eliminating plastic flow and pushing the elastic left Cauchy-Green deformation tensor forward to the current configuration using the relative deformation gradient as

$${}^{tr}\mathbf{b}^e = \mathbf{f}^n \mathbf{b}^e \mathbf{f}^T, \quad {}^{tr}\xi = {}^n\xi. \quad (4.171)$$

In terms of deformation gradient, this process is equivalent to ${}^n\mathbf{F}^p = {}^n\mathbf{F}^p$ and ${}^n\mathbf{F}^e = \mathbf{f}^n \mathbf{F}^e$. Thus, the incremental deformation is assumed to be purely elastic.

If $\boldsymbol{\tau}$ and \mathbf{q} , which are evaluated using the trial state, are within the elastic domain, then the trial stress and the stress-like plastic variable are exact, and time integration is finished with the material being elastic. Otherwise, the material is plastic, and plastic return-mapping is carried out by integrating Eq. (4.168) between load step t_n and t_{n+1} with constraints imposed on the stress through the yield function in Eq. (4.168c). By integrating the differential part of Eq. (4.168), the following return-mapping algorithm can be obtained:

$$\mathbf{b}^e = {}^n\mathbf{b}^e \exp \left[-2\Delta\gamma \frac{\partial f(\boldsymbol{\tau}, \mathbf{q})}{\partial \boldsymbol{\tau}} \right], \quad (4.172a)$$

$$\xi = {}^n\xi + \Delta\gamma \frac{\partial f(\boldsymbol{\tau}, \mathbf{q})}{\partial \mathbf{q}}, \quad (4.172b)$$

$$\Delta\gamma \geq 0, \quad f(\boldsymbol{\tau}, \mathbf{q}) \leq 0, \quad \Delta\gamma f(\boldsymbol{\tau}, \mathbf{q}) = 0. \quad (4.172c)$$

The property that the solution of differential equation $\dot{y} = Ay$ is $y = y_0 \exp(A\Delta t)$ is used in Eq. (4.172a), and $\Delta\gamma = \gamma\Delta t$ is used. This integration algorithm has first-order accuracy and unconditional stability.

Note that the return-mapping procedure in the above equations are different from that of the classical plasticity because it is based on strains (\mathbf{b}^e, ξ), not stresses ($\boldsymbol{\tau}, \mathbf{q}$).

In addition, the return-mapping of \mathbf{b}^e is given in terms of multiplication with an exponential function, which is not straightforward to implement. The idea proposed by Simo was that if logarithm is taken in Eq. (4.172a), the multiplication will be converted into an addition, and the strains in \mathbf{b}^e will be converted into logarithmic strains. Another important concept is that when a material is isotropic, the principal directions of $\boldsymbol{\tau}$ is aligned with that of \mathbf{b}^e by using the isotropic assumption. From this observation, the spectral decompositions of \mathbf{b}^e and $\boldsymbol{\tau}$ can be written as

$$\mathbf{b}^e = \sum_{i=1}^3 \lambda_i^2 \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i \quad (4.173)$$

and

$$\boldsymbol{\tau} = \sum_{i=1}^3 \tau_i^p \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i, \quad (4.174)$$

respectively, where λ_i is the principal stretch, $\boldsymbol{\tau}^p = \{\tau_1^p, \tau_2^p, \tau_3^p\}^T$ is the vector of principal Kirchhoff stresses, and $\tilde{\mathbf{n}}^i$ is the spatial eigenvector corresponding to the material eigenvector $\tilde{\mathbf{N}}^i$. In addition, it is assumed that the plastic return-mapping occurs with the fixed principal directions. Since ${}^T\mathbf{b}^e$ has the same principal directions as \mathbf{b}^e , the principal directions of $\boldsymbol{\tau}$ and \mathbf{b}^e can be computed from the known principal directions of ${}^T\mathbf{b}^e$. Thus, the return-mapping occurs in the principal stretches with fixed principal directions. The counterpart in the classical plasticity is the radial return-mapping in which the trial shifted stress ${}^T\boldsymbol{\eta}$ is parallel to the updated shifted stress $\boldsymbol{\eta}$ (see Fig. 4.15).

The facts that the principal directions of $\boldsymbol{\tau}$ and \mathbf{b}^e are the same and the principal directions are fixed during the return-mapping make it possible to modify the algorithm into stress-based one. In addition, the algorithm becomes similar to the classical plasticity when principal stresses and principal logarithmic stretches are used. For the simplicity of derivations, the vector of logarithmic, elastic principal stretches are defined by $\mathbf{e} = [e_1, e_2, e_3]^T = [\log(\lambda_1), \log(\lambda_2), \log(\lambda_3)]^T$. In the elastoplasticity theory, it can be assumed that the free energy is in the following quadratic form:

$$\psi(\mathbf{e}, \xi) = \frac{1}{2} \lambda [e_1 + e_2 + e_3]^2 + \mu [e_1^2 + e_2^2 + e_3^2] + K(\xi), \quad (4.175)$$

where $K(\xi)$ denotes energy from isotropic hardening law and λ and μ are Lame's constants. Then the hyperelastic constitutive relation in Eq. (4.169) can be reduced to $\boldsymbol{\tau}^p = \partial\psi/\partial\mathbf{e}$ in principal space. Using the above free energy, the relation between principal stress and the logarithmic elastic principal stretch becomes

$$\boldsymbol{\tau}^p = \frac{\partial \psi}{\partial \mathbf{e}} = \mathbf{c}^e \mathbf{e}, \quad (4.176)$$

where $\mathbf{c}^e = (\lambda + \frac{2}{3}\mu)\tilde{\mathbf{1}} \otimes \tilde{\mathbf{1}} + 2\mu\mathbf{1}_{dev}$ is the elastic 3×3 constitutive tensor for an isotropic material, $\tilde{\mathbf{1}} = [1, 1, 1]^T$ is the first-order vector, and $\mathbf{1}_{dev} = \mathbf{1} - \frac{1}{3}(\tilde{\mathbf{1}} \otimes \tilde{\mathbf{1}})$ is the second-order unit deviatoric tensor. These notations can be thought of as a second-order version of the fourth-order notations given in Eq. (4.50). Taking the logarithm of Eq. (4.172a) and pre-multiplying by \mathbf{c}^e yields the following return-mapping algorithm forms in the principal stress space:

$${}^{tr}\boldsymbol{\tau}^p = \mathbf{c}^e \mathbf{e}^{tr}, \quad (4.177a)$$

$$\boldsymbol{\tau}^p = {}^{tr}\boldsymbol{\tau}^p - \Delta\gamma \mathbf{c}^e \frac{\partial \tilde{f}(\boldsymbol{\tau}^p, \mathbf{q})}{\partial \boldsymbol{\tau}^p}, \quad (4.177b)$$

$$\boldsymbol{\xi} = {}^n\boldsymbol{\xi} + \Delta\gamma \frac{\partial \tilde{f}(\boldsymbol{\tau}^p, \mathbf{q})}{\partial \mathbf{q}}, \quad (4.177c)$$

$$\Delta\gamma \geq 0, \quad \tilde{f}(\boldsymbol{\tau}^p, \mathbf{q}) \leq 0, \quad \Delta\gamma \tilde{f}(\boldsymbol{\tau}^p, \mathbf{q}) = 0, \quad (4.177d)$$

where \mathbf{e}^{tr} is the logarithmic principal stretch of ${}^n\mathbf{b}^e$ and $\tilde{f}(\boldsymbol{\tau}^p, \mathbf{q})$ is a different expression of $f(\boldsymbol{\tau}, \mathbf{q})$, whose explicit expression will be presented later. The normal to the yield surface in the stress space can be written in the normal to the principal stress space by $\partial f / \partial \boldsymbol{\tau} = \sum_{i=1}^3 \partial \tilde{f} / \partial \tau_i^n \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i$. This return-mapping algorithm is the same as that of the classical plasticity, with a difference that principal Kirchhoff stress and logarithmic strain are used instead of Cauchy stress and engineering strain.

4.5.4 Return-Mapping Algorithm

Since a plastic behavior can be efficiently described by the deviatoric part of a vector, which preserves the volume change, a deviatoric principal stress is defined by

$$\mathbf{s} = \boldsymbol{\tau}^p - \frac{1}{3}(\boldsymbol{\tau}^p \cdot \tilde{\mathbf{1}})\tilde{\mathbf{1}} = \mathbf{1}_{dev} \cdot \boldsymbol{\tau}^p. \quad (4.178)$$

For plasticity, the von Mises yield criterion and the associative flow rule are commonly used to describe metal-like material behavior after elastic deformation. Accordingly, the yield criterion or yield function is formulated as

$$f(\mathbf{q}, e_p) = ||\mathbf{q}|| - \sqrt{\frac{2}{3}}\kappa(e_p) = 0, \quad (4.179)$$

where $\boldsymbol{\eta}$ is equal to $\mathbf{s} - \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}$ is the back stress, which is the center of the yield surface and is determined by the kinematic hardening law. $\kappa(e_p)$ is the radius of the yield surface and is determined by the isotropic hardening rule. For a general nonlinear hardening rule, $\kappa(e_p)$ can be expressed by

$$\kappa(e_p) = \sigma_Y^0 + K_{e_p}(e_p), \quad (4.180)$$

where σ_Y^0 is the initial yield stress from a uniaxial tension test, and $K_{e_p}(e_p) = \partial K / \partial e_p$ is the isotropic hardening law. The effective plastic strain e_p can be determined by integrating the rate of plastic strain as

$$e_p = \int_0^t \sqrt{\frac{2}{3}} \|\dot{\mathbf{e}}^p(\tau)\| d\tau. \quad (4.181)$$

The plastic variables are reduced to the effective plastic strain and back stress. Note that the above yield function and hardening models are similar to that of the classical plasticity in Sect. 4.3. The only difference is that they are formulated in the principal stress space.

Now, the return-mapping algorithms in Eq. (4.177) can be implemented in the principal Kirchhoff stress space by

$${}^{tr}\boldsymbol{\tau}^p = \mathbf{c}^e \mathbf{e}^{tr}, \quad (4.182a)$$

$$\boldsymbol{\tau}^p = {}^r\boldsymbol{\tau}^p - 2\mu\Delta\gamma \mathbf{N}, \quad (4.182b)$$

$$\boldsymbol{\alpha} = {}^n\boldsymbol{\alpha} + \Delta\gamma H_\alpha \mathbf{N}, \quad (4.182c)$$

$$e_p = {}^n e_p + \sqrt{\frac{2}{3}} \Delta\gamma, \quad (4.182d)$$

where $H_\alpha(e_p)$ is a plastic modulus for kinematic hardening and

$$\mathbf{N} \equiv \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|} = \frac{{}^{tr}\boldsymbol{\eta}}{\|{}^{tr}\boldsymbol{\eta}\|} \quad (4.183)$$

is an outward unit normal vector on the yield surface. The plastic consistency parameter, $\Delta\gamma$, is computed from the fact that the yield function remains 0 during the plastic deformation:

$$\begin{aligned} f(\boldsymbol{\eta}, e_p) &= \|\boldsymbol{\eta}\| - \sqrt{\frac{2}{3}} \kappa(e_p) \\ &= \|{}^{tr}\boldsymbol{\eta}\| - (2\mu + H_\alpha(e_p)) \Delta\gamma - \sqrt{\frac{2}{3}} \kappa(e_p) = 0, \end{aligned} \quad (4.184)$$

which is a nonlinear equation with respect to $\Delta\gamma$. Equation (4.184) can be solved using a local Newton–Raphson method to compute $\Delta\gamma$. As the elastic domain E is smooth and convex, the return-mapping algorithm becomes robust. If the

isotropic/kinematic hardening is a linear function of $\Delta\gamma$ or of the effective plastic strain, then $\Delta\gamma$ can explicitly be calculated without any iteration. When the hardenings are nonlinear, the following derivative of the yield function with respect to $\Delta\gamma$ is required during the iteration:

$$\frac{\partial f}{\partial \Delta\gamma} = 2\mu + H_\alpha + \sqrt{\frac{2}{3}}H_{\alpha,e_p}\Delta\gamma + \frac{2}{3}\kappa_{e_p}.$$

This part of algorithm is the same as the return-mapping algorithm in Sect. 4.3.5. Once $\Delta\gamma$ is obtained, it is used to calculate principal stresses and plastic variables in Eq. (4.182).

The Kirchhoff stress tensor can be obtained from Eq. (4.174) using the principal stress and principal direction as

$$\boldsymbol{\tau} = \sum_{i=1}^3 \tau_i^p \mathbf{m}^i, \quad (4.185)$$

where $\mathbf{m}^i = \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i$ is the matrix of principal directions. The left Cauchy-Green deformation tensor is updated and stored by the formula in Eq. (4.173), which represents the intermediate configuration:

$$\mathbf{b}^e = \sum_{i=1}^3 \exp[2e_i] \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i, \quad (4.186)$$

where $\mathbf{e} = \mathbf{e}^{tr} - \Delta\gamma \mathbf{N}$ is an elastic logarithmic principal strain. Equation (4.186) corresponds to the update of $\mathbf{C}^{p-1} = \mathbf{F}^{-1} \mathbf{b}^e \mathbf{F}^{-T}$.

Example 4.21 (Incompressible Elastic Cube) An incompressible elastic cube undergoes the following deformation:

$$x_1 = \alpha X_1, \quad x_2 = \beta X_2, \quad x_3 = \beta X_3.$$

Using the linear relationship between principal Kirchhoff stress and logarithmic stretch, find the Kirchhoff stress tensor as a function of stretch α . Use the following material properties: λ and μ .

Solution Since the cube is elastic, there is no need to separate elastic and plastic part of deformation. Thus, the superscript “e” will be omitted in the following derivation. For given deformation, the deformation gradient and left Cauchy-Green deformation tensor become

$$\mathbf{F} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \beta \end{bmatrix}, \quad \mathbf{b} = \mathbf{F}\mathbf{F}^T = \begin{bmatrix} \alpha^2 & 0 & 0 \\ 0 & \beta^2 & 0 \\ 0 & 0 & \beta^2 \end{bmatrix}.$$

From the requirement of incompressibility, the relation between α and β can be obtained as $\beta = 1/\sqrt{\alpha}$. Since \mathbf{b} has only diagonal components, the three eigenvalues and eigenvectors become

$$\begin{aligned}\lambda_1 &= \alpha^2, & \mathbf{n}^1 &= [1 \quad 0 \quad 0]^T \\ \lambda_2 &= \alpha^{-1}, & \mathbf{n}^2 &= [0 \quad 1 \quad 0]^T \\ \lambda_3 &= \alpha^{-1}, & \mathbf{n}^3 &= [0 \quad 0 \quad 1]^T\end{aligned}$$

Then, the logarithmic stretch can be obtained by

$$\mathbf{e} = \{2\log\alpha \quad -\log\alpha \quad -\log\alpha\}^T.$$

The stress-strain relation in the principal space, $\boldsymbol{\tau}^p = \mathbf{c} \cdot \mathbf{e}$, can be written as

$$\boldsymbol{\tau}^p = \begin{bmatrix} \lambda + 2\mu & \lambda & \lambda \\ \lambda & \lambda + 2\mu & \lambda \\ \lambda & \lambda & \lambda + 2\mu \end{bmatrix} \begin{Bmatrix} 2\log\alpha \\ -\log\alpha \\ -\log\alpha \end{Bmatrix} = \begin{Bmatrix} 4\mu\log\alpha \\ -2\mu\log\alpha \\ -2\mu\log\alpha \end{Bmatrix}.$$

Then, the Kirchhoff stress can be obtained using

$$\boldsymbol{\tau} = \sum_{i=1}^3 \boldsymbol{\tau}_i^p \mathbf{n}^i \otimes \mathbf{n}^i = 2\mu\log\alpha \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

Note that the relation between stress and logarithmic stretch is linear. ■

4.5.5 Consistent Algorithmic Tangent Operator

The exact tangent operator can be obtained by taking the derivative of the Kirchhoff stress tensor with respect to the strain. This spatial tangent operator has the following relation to the material tangent operator:

$$c_{ijkl} = 2F_{il}F_{jl}F_{kk}F_{ll} \frac{\partial S_{IJ}}{\partial C_{KL}}, \quad (4.187)$$

where C_{KL} is a component of the right Cauchy-Green tensor, and S_{IJ} is a component of the second Piola-Kirchhoff stress defined by $\mathbf{S} = \mathbf{F}^{-1}\boldsymbol{\sigma}\mathbf{F}^{-T}$. Since stress is a function of elastic trial strain and since the intermediate configuration is held fixed in the elastic trial process, all material tensors are referred to the intermediate configuration and linearization is carried out with respect to that configuration. The return-mapping in the previous section gives the following relation:

$$\mathbf{c}^{\text{alg}} \equiv \frac{\partial \boldsymbol{\tau}^p}{\partial \mathbf{e}^{tr}} = \mathbf{c}^e - \frac{4\mu^2 \mathbf{N} \otimes \mathbf{N}}{2\mu + H_\alpha + \sqrt{\frac{2}{3}H_{\alpha, e_p} \Delta\gamma + \frac{2}{3}\kappa_{e_p}}} - \frac{4\mu^2 \Delta\gamma}{\|\boldsymbol{\eta}^{tr}\|} [\mathbf{1}_{\text{dev}} - \mathbf{N} \otimes \mathbf{N}], \quad (4.188)$$

which is a 3×3 symmetric matrix in the principal stress space. Equation (4.188) has the same form as the classical plasticity model in Eq. (4.120), except that principal stress and logarithmic stretch are now used. Using the property in Eq. (4.187), the stress in Eq. (4.185) is differentiated to yield

$$\frac{\partial \boldsymbol{\tau}}{\partial \boldsymbol{\epsilon}} = \sum_{i=1}^3 \left[\frac{\partial \tau_i^p}{\partial e_j^{tr}} \left(2\mathbf{F}^e \frac{\partial e_j^{tr}}{\partial \mathbf{C}^e} \mathbf{F}^{eT} \right) \otimes \mathbf{m}^i + 2\tau_i^p \left(\mathbf{F}^e \frac{\partial \mathbf{m}^i}{\partial \mathbf{C}^e} \mathbf{F}^{eT} \right) \right], \quad (4.189)$$

where e_j^{tr} is a function of total deformation and is independent of the plastic evolution law. The following relation is the differential version of the eigenvalue problem $\mathbf{C}^e \tilde{\mathbf{N}} = \lambda^2 \tilde{\mathbf{N}}$, derived in Sect. 4.6.1:

$$2\mathbf{F}^e \frac{\partial e_j^{tr}}{\partial \mathbf{C}^e} \mathbf{F}^{eT} = \mathbf{m}^j. \quad (4.190)$$

The last term in Eq. (4.189) is independent of plastic flow because plastic evolution is carried out in the fixed, principal direction. In Sect. 4.6.2, it is explicitly shown from finite elasticity that

$$\begin{aligned} \mathbf{F}^e \frac{\partial \mathbf{m}^i}{\partial \mathbf{C}^e} \mathbf{F}^{eT} &= \frac{1}{d_i} \left\{ \mathbf{I}_{b^e} - \mathbf{b}^e \otimes \mathbf{b}^e - I_3 \lambda_i^{-2} [\mathbf{I} - (\mathbf{1} - \mathbf{m}^i) \otimes (\mathbf{1} - \mathbf{m}^i)] \right\} \\ &\quad + \frac{\lambda_i^2}{d_i} \left\{ \mathbf{b}^e \otimes \mathbf{m}^i + \mathbf{m}^i \otimes \mathbf{b}^e + (I_1 - 4\lambda_i^2) \mathbf{m}^i \otimes \mathbf{m}^i \right\} \\ &\equiv \hat{\mathbf{c}}^i \end{aligned}, \quad (4.191)$$

where $d_i = (\lambda_i^2 - \lambda_j^2)(\lambda_i^2 - \lambda_k^2)$ with an even permutation between $i-j-k$, I_1 and I_3 are the first and third invariant of \mathbf{b}^e , and $I_{ijkl} = \frac{1}{2}(b_{ik}b_{jl} + b_{il}b_{jk})$ can be obtained by pushing \mathbf{I} forward to the current configuration. The algorithmic tangent operator in Eq. (4.187) can thus be expressed as

$$\mathbf{c} = \sum_{i=1}^3 \sum_{j=1}^3 c_{ij}^{\text{alg}} \mathbf{m}^i \otimes \mathbf{m}^j + 2 \sum_{i=1}^3 \tau_i^p \hat{\mathbf{c}}^i, \quad (4.192)$$

which contains all symmetric properties between indices.

4.5.6 Variational Principles for Finite Deformation

The energy form at time t_n can be written using Kirchhoff stress and engineering strain at the current configuration as

$$a(^n\xi; \mathbf{u}, \bar{\mathbf{u}}) = \iint_{^n\Omega} \boldsymbol{\tau} : \boldsymbol{\epsilon}(\bar{\mathbf{u}}) d\Omega. \quad (4.193)$$

Since Kirchhoff stress is used, the integration domain is an undeformed configuration. It is assumed that the constitutive equation is given in the total form for elastic material as in Eq. (4.187). The explicit form of c_{ijkl} can be obtained from Eq. (4.192). The updated Lagrangian formulation of the linearized energy form has a specific expression as

$$a^*({}^n\xi, {}^n\mathbf{b}^e, \mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) \equiv \iint_{\Omega} [\boldsymbol{\varepsilon}(\bar{\mathbf{u}}):\mathbf{c} : \boldsymbol{\varepsilon}(\Delta\mathbf{u}) + \boldsymbol{\tau}:\boldsymbol{\eta}(\Delta\mathbf{u}, \bar{\mathbf{u}})] d\Omega. \quad (4.194)$$

If current load step is t_{n+1} and the iteration counter is $k+1$, then the linearized incremental equation becomes

$$a^*({}^n\xi, {}^{n+1}\mathbf{u}^{k+1}; \Delta\mathbf{u}^{k+1}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}) - a({}^n\xi, {}^{n+1}\mathbf{u}, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (4.195)$$

The above equation is solved iteratively until the right-hand side (the residual force) vanishes. After convergence, time is increased and the same analysis procedure described above is repeated until a final configuration is reached. Note that integration of the internal energy term is carried out on the undeformed configuration because Kirchhoff stress is used.

4.5.7 Computer Implementation of Finite Deformation Elastoplasticity

The implementation of the above finite deformation elastoplasticity can be similar to the infinitesimal elastoplasticity if the Kirchhoff stress and logarithmic strain are used in the principal space. Below are the lists of two MATLAB programs: **mulPlast** and **mulPlastTan**. The former performs return-mapping to find stress and plastic variables, while the latter calculates consistent tangent stiffness matrix. Since there is no shear stress, the principal stress and logarithmic strain are stored in 3×1 vectors. The input parameters are velocity gradient at the current load step, and plastic variables from the previous load step along with material properties. First, the incremental deformation gradient is calculated from the velocity gradient, and the elastic left Cauchy-Green deformation tensor, \mathbf{b}^e , is updated using the incremental deformation gradient. Second, the eigenvalues (principal stretches) and eigenvectors (principal directions) of \mathbf{b}^e are calculated. Third, using logarithmic principal stretches, the return-mapping procedure is performed to find principal Kirchhoff stresses. Lastly, stress and plastic variables are updated using the plastic consistency parameters. The Kirchhoff stress is calculated using the principal stresses and principal directions.

The tangent stiffness of the multiplicative plasticity consists of two stages. In the first stage, the consistent tangent stiffness matrix between principal stresses and principal logarithmic stretches is calculated, which is similar to that of the infinitesimal plasticity. The second stage is to calculate the effect of elastic principal stretches and that of the principal directions.

```

PROGRAM mulPlast
%
% Multiplicative plasticity with linear combined hardening
%
function [stress, b, alpha, ep]=mulPlast(mp,D,L,b,alpha,ep)
%mp = [lambda, mu, beta, H, Y0];
%D = elasticity matrix b/w prin stress & log prin stretch (3x3)
%L = [dui/dxj] velocity gradient
%b = elastic left C-G deformation vector (6x1)
%alpha = principal back stress (3x1)
%ep = effective plastic strain
%
Iden = [1 1 1]'; two3 = 2/3; stwo3=sqrt(two3); %constants
mu=mp(2); beta=mp(3); H=mp(4); Y0=mp(5); %material properties
ftol = Y0*1E-6; %tolerance for yield
R = inv(eye(3)-L); %inc. deformation gradient
bm=[b(1) b(4) b(6);b(4) b(2) b(5);b(6) b(5) b(3)];
bm = R*bm*R'; %trial elastic left C-G
[V,P]=eig(bm); %eigenvalues and vectors
b=[bm(1,1) bm(2,2) bm(3,3) bm(1,2) bm(2,3) bm(1,3)]';
M=zeros(6,3); %eigenvector matrices
M(1,:)=V(1,:).^2;
M(2,:)=V(2,:).^2;
M(3,:)=V(3,:).^2;
M(4,:)=V(1,:).*V(2,:);
M(5,:)=V(2,:).*V(3,:);
M(6,:)=V(1,:).*V(3,:);
eigen=[P(1,1) P(2,2) P(3,3)]'; %principal stretch
%if abs(eigen(1)-eigen(2)) < 1E-12; eigen(2) = eigen(2) + 1E-12; end;
%if abs(eigen(2)-eigen(3)) < 1E-12; eigen(2) = eigen(2) + 1E-12; end;
deps = 0.5*log(eigen); %logarithmic
sigtr = D*deps; %trial principal stress
eta = sigtr - alpha - sum(sigtr)*Iden/3; %shifted stress
etat = norm(eta); %norm of eta
%etat=sqrt(eta(1)^2+eta(2)^2+eta(3)^2);
fyld = etat - stwo3*(Y0+(1-beta)*H*ep); %trial yield function
if fyld < ftol %yield test
    sig = sigtr; %trial states are final
    stress = M*sig; %stress (6x1)
else
    gamma = fyld/(2*mu + two3*H); %plastic consistency param
    ep = ep + gamma*stwo3; %updated eff. plastic strain
    N = eta/etat; %unit vector normal to f
    deps = deps - gamma*N; %updated elastic strain
    sig = sigtr - 2*mu*gamma*N; %updated stress
    alpha = alpha + two3*beta*H*gamma*N; %updated back stress
    stress = M*sig; %stress (6x1)
    b = M*exp(2*deps); %updated elastic left C-G
end

```

```

PROGRAM mulPlastTan
%
% Tangent stiffness of multiplicative plasticity with linear hardening
%
function [Dtan]=mulPlastTan(mp,D,L,b,alpha,ep)
%
Iden = [1 1 1]'; two3 = 2/3; stwo3=sqrt(two3); %constants
mu=mp(2); beta=mp(3); H=mp(4); Y0=mp(5); %material properties
ftol = Y0*1E-6; %tolerance for yield
R = inv(eye(3)-L); %inc. deformation gradient
bm=[b(1) b(4) b(6);b(4) b(2) b(5);b(6) b(5) b(3)];
bm = R*bm*R'; %trial elastic left C-G
[V,P]=eig(bm); %eigenvalues and vectors
b=[bm(1,1) bm(2,2) bm(3,3) bm(1,2) bm(2,3) bm(1,3)]';
M=zeros(6,3); %eigenvector matrices
M(1,:)=V(1,:).^2;
M(2,:)=V(2,:).^2;
M(3,:)=V(3,:).^2;
M(4,:)=V(1,:).*V(2,:);
M(5,:)=V(2,:).*V(3,:);
M(6,:)=V(1,:).*V(3,:);
eigen=[P(1,1) P(2,2) P(3,3)]'; %principal stretch
%if abs(eigen(1)-eigen(2)) < 1E-12; eigen(2) = eigen(2) + 1E-12; end;
%if abs(eigen(2)-eigen(3)) < 1E-12; eigen(2) = eigen(2) + 1E-12; end;
deps = 0.5*log(eigen); %logarithmic
sigtr = D*deps; %trial principal stress
eta = sigtr - alpha - sum(sigtr)*Iden/3; %shifted stress
etat = norm(eta); %norm of eta
fyld = etat - stwo3*(Y0+(1-beta)*H*ep); %trial yield function
if fyld >= ftol %yield test
    gamma = fyld/(2*mu + two3*H); %plastic consistency param
    N = eta/etat; %unit vector normal to f
    sig = sigtr - 2*mu*gamma*N; %updated stress
    var1 = 4*mu^2/(2*mu+two3*H);
    var2 = 4*mu^2*gamma/etat; %coefficients
    D = D - (var1-var2)*N*N'+var2*Iden*Iden'/3;%tangent stiffness
    D(1,1) = D(1,1) - var2; %contr. from 4th-order I
    D(2,2) = D(2,2) - var2;
    D(3,3) = D(3,3) - var2;
end
J1 = sum(eigen);
J3 = eigen(1)*eigen(2)*eigen(3);
I2=[1 1 1 0 0 0]';
I4=eye(6);I4(4,4)=.5;I4(5,5)=.5;I4(6,6)=.5;
Ibb=[0,b(4)^2-b(1)*b(2),b(6)^2-b(1)*b(3),0,b(4)*b(6)-b(1)*b(5),0;
b(4)*b(4)-b(1)*b(2),0,b(5)^2-b(2)*b(3),0,0,b(4)*b(5)-b(2)*b(6);
b(6)^2-b(1)*b(3),b(5)^2-b(2)*b(3),0,b(5)*b(6)-b(3)*b(4),0,0;

```

```

0,0,b(5)*b(6)-b(3)*b(4), (b(1)*b(2)-b(4)^2)/2, ...
(b(2)*b(6)-b(4)*b(5))/2, (b(1)*b(5)-b(4)*b(6))/2, ...
b(4)*b(6)-b(1)*b(5), 0,0, (b(2)*b(6)-b(4)*b(5))/2, ...
(b(2)*b(3)-b(5)^2)/2, (b(3)*b(4)-b(5)*b(6))/2, ...
0,b(4)*b(5)-b(2)*b(6), 0, (b(1)*b(5)-b(4)*b(6))/2, ...
(b(3)*b(4)-b(5)*b(6))/2, (b(1)*b(3)-b(6)^2)/2];
%
d1=1/((eigen(2)-eigen(1))*(eigen(3)-eigen(1)));
d2=1/((eigen(3)-eigen(2))*(eigen(1)-eigen(2)));
d3=1/((eigen(1)-eigen(3))*(eigen(2)-eigen(3)));
t11=-J3*d1/eigen(1); t12=-J3*d2/eigen(2); t13=-J3*d3/eigen(3);
t21=d1*eigen(1); t22=d2*eigen(2); t23=d3*eigen(3);
t31=t21*(J1-4*eigen(1)); t32=t22*(J1-4*eigen(2)); t33=t23*(J1-4*eigen
(3));
%
CT1=d1*Ibb+t11*(I4-(I2-b)*(I2-b)')+t21*(b*M(:,1)'+M(:,1)*b')+t31*M
(:,1)*M(:,1)';
CT2=d2*Ibb+t12*(I4-(I2-b)*(I2-b)')+t22*(b*M(:,2)'+M(:,2)*b')+t32*M
(:,2)*M(:,2)';
CT3=d3*Ibb+t13*(I4-(I2-b)*(I2-b)')+t23*(b*M(:,3)'+M(:,3)*b')+t33*M
(:,3)*M(:,3)';
%
Dtan = M*D*M' + 2*(CT1*sig(1)+CT2*sig(2)+CT3*sig(3));

```

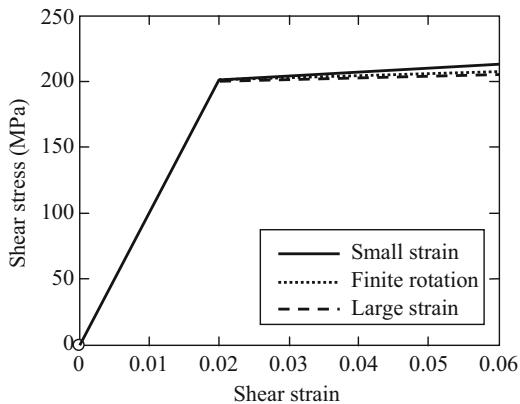
Example 4.22 (Shear Deformation of a Square) Solve the shear deformation of a square in Example 4.19 using multiplicative plasticity. Plot shear stresses from small strain, finite rotation, and large strain as a function of shear stress.

Solution The shear stresses from the small strain and finite rotation are available in Example 4.19. Below is the list of MATLAB program that solves for the shear deformation problem for all three cases. The variable “**stress**” is the updated stress from infinitesimal deformation assumption, “**stressR**” is the one from finite rotation assumption, and “**stressM**” is the one from large strain assumption. At the last load increment, these three stresses are

$$\begin{aligned} \text{stress} &= [0 \quad 0 \quad 0 \quad 212.9 \quad 0 \quad 0]^T \\ \text{stressR} &= [43.4 \quad -43.4 \quad 0 \quad 208.2 \quad 0 \quad 0]^T \\ \text{stressM} &= [-56.9 \quad -152.3 \quad -78.7 \quad 206.7 \quad 0 \quad 0]^T. \end{aligned}$$

The difference in shear stress σ_{12} is relatively small. However, in the finite rotational and large strain formulations, the normal stresses are developed due to the rotation of the reference frame. Figure 4.24 shows the shear stress vs. shear strain curve.

Fig. 4.24 Stress-strain curve for elastoplastic shear deformation



```
%  
% Example 4.22 - Shear deformation of a square  
%  
Young = 24000; nu=0.2; mu=Young/2/(1+nu); lambda=nu*Young/((1+nu)*(1-  
2*nu));  
beta = 0; H = 1000; sY = 200*sqrt(3);  
mp = [lambda mu beta H sY];  
Iden=[1 1 1 0 0 0]';  
D=2*mu*eye(6) + lambda*Iden*Iden';  
D(4,4)=mu; D(5,5)=mu; D(6,6)=mu;  
Iden=[1 1 1]';  
DM=2*mu*eye(3) + lambda*Iden*Iden';  
L = zeros(3,3);  
stressN=[0 0 0 0 0 0]';  
deps=[0 0 0 0 0 0]';  
alphaN=[0 0 0 0 0 0]';  
epN=0;  
stressRN=stressN; alphaRN=alphaN; epRN=epN;  
bMN=[1 1 1 0 0 0]';  
alphaMN = [0 0 0]';  
epMN=0;  
for i=1:15  
    deps(4) = 0.004; L(1,2) = 0.024; L(2,1) = -0.02;  
    [stressRN, alphaRN] = rotatedStress(L, stressRN, alphaRN);  
    [stressR, alphaR, epR]=combHard(mp,D,deps,stressRN,alphaRN,epRN);  
    [stress, alpha, ep]=combHard(mp,D,deps,stressN,alphaN,epN);  
    [stressM, bM, alphaM, epM]=mulPlast(mp,DM,L,bMN,alphaMN,epMN);  
    X(i)=i*deps(4);Y1(i)=stress(4);Y2(i)=stressR(4);Y3(i)=stressM(4);  
    stressN = stress; alphaN = alpha; epN = ep;  
    stressRN = stressR; alphaRN = alphaR; epRN = epR;  
    bMN=bM; alphaMN = alphaM; epMN = epM;  
end  
X = [0 X]; Y1=[0 Y1]; Y2=[0 Y2]; Y3 = [0 Y3]; plot(X,Y1,X,Y2,X,Y3);  
■
```

4.6 Mathematical Formulas from Finite Elasticity

In this section, mathematical formulas used in the finite elasticity are derived in detail.

4.6.1 Linearization of Principal Logarithmic Stretches

Let the reference frame be the intermediate configuration for an elastoplastic problem, which is fixed in the trial state. The principal stretches λ_i are functions of the total deformation and independent of the plastic flow. For simplicity, all variables denote elastic trial status in this section without being given a specific notation. The right and left Cauchy-Green tensors have the same principal values λ_i^2 , and the eigenvalue problem is

$$\mathbf{C}\tilde{\mathbf{N}}^i = \lambda_i^2\tilde{\mathbf{N}}^i, \quad \mathbf{b}\tilde{\mathbf{n}}^i = \lambda_i^2\tilde{\mathbf{n}}^i. \quad (4.196)$$

The relation between principal directions is

$$\mathbf{F}\tilde{\mathbf{N}}^i = \lambda_i\tilde{\mathbf{n}}^i. \quad (4.197)$$

By differentiating Eq. (4.196), the following relation can be obtained:

$$\mathbf{d}\mathbf{C}\tilde{\mathbf{N}}^i + \mathbf{C}d\tilde{\mathbf{N}}^i = 2\lambda_i d\lambda_i \tilde{\mathbf{N}}^i + \lambda_i^2 d\tilde{\mathbf{N}}^i, \quad \text{no sum on } i. \quad (4.198)$$

Then, taking an inner product with $\tilde{\mathbf{N}}$ and using the property that $\tilde{\mathbf{N}} \cdot d\tilde{\mathbf{N}} = 0$, the following can be obtained:

$$2\lambda_i d\lambda_i = \tilde{\mathbf{N}}^i \mathbf{d}\mathbf{C}\tilde{\mathbf{N}}^i = \text{tr}[\mathbf{d}\mathbf{C}(\tilde{\mathbf{N}}^i \otimes \tilde{\mathbf{N}}^i)] \quad (4.199)$$

or

$$\frac{\partial \lambda_i}{\partial \mathbf{C}} = \frac{1}{2\lambda_i} \tilde{\mathbf{N}}^i \otimes \tilde{\mathbf{N}}^i. \quad (4.200)$$

Since the logarithmic strain is defined by the principal stretch,

$$e_i = \log(\lambda_i) \quad (4.201)$$

and

$$2\frac{\partial e_i}{\partial \mathbf{C}} = 2\frac{\partial e_i}{\partial \lambda_i} \frac{\partial \lambda_i}{\partial \mathbf{C}} = \lambda_i^{-2} \tilde{\mathbf{N}}^i \otimes \tilde{\mathbf{N}}^i. \quad (4.202)$$

The push-forward of this result along with the relation in Eq. (4.197) yields

$$2\mathbf{F} \frac{\partial e_i}{\partial \mathbf{C}} \mathbf{F}^T = \tilde{\mathbf{n}}^i \otimes \tilde{\mathbf{n}}^i. \quad (4.203)$$

Since all the relations are transformed to the current configuration, the properties of the intermediate configuration are completely removed.

4.6.2 Linearization of the Eigenvector of the Elastic Trial Left Cauchy-Green Tensor

For simplicity, all superscripts of elastic trial status are ignored. Let $\tilde{\mathbf{n}}^A$ be the principal direction of \mathbf{b} corresponding to the principal value λ_A^2 , and let $\tilde{\mathbf{N}}^A$ be the principal direction of \mathbf{C} . The following relation is thus satisfied:

$$\mathbf{C} = \sum_{A=1}^3 \lambda_A^2 \tilde{\mathbf{N}}^A \otimes \tilde{\mathbf{N}}^A, \quad \mathbf{b} = \sum_{A=1}^3 \lambda_A^2 \tilde{\mathbf{n}}^A \otimes \tilde{\mathbf{n}}^A \quad (4.204)$$

and I_1 , I_2 , and I_3 are the three invariants of \mathbf{C} . The relation of eigenvector bases between material and spatial description is

$$\mathbf{M}^A \equiv \lambda_A^{-2} \tilde{\mathbf{N}}^A \otimes \tilde{\mathbf{N}}^A = \mathbf{F}^{-1} (\tilde{\mathbf{n}}^A \otimes \tilde{\mathbf{n}}^A) \mathbf{F}^{-T} = \mathbf{F}^{-1} \mathbf{m}^A \mathbf{F}^{-T}. \quad (4.205)$$

The explicit form of \mathbf{M}^A can be computed by Serrin's representation theorem, namely,

$$\mathbf{M}^A = \frac{1}{d_A} [\mathbf{C} - (I_1 - \lambda_A^2) \mathbf{1} + I_3 \lambda_A^{-2} \mathbf{C}^{-1}], \quad (4.206)$$

where $d_A = (\lambda_A^2 - \lambda_B^2)(\lambda_A^2 - \lambda_C^2) = 2\lambda_A^4 - I_1\lambda_A^2 + I_3\lambda_A^{-2}$ with A, B , and C having an even permutation. The following properties can be derived by using the chain rule of differentiation and direct computation:

$$\frac{\partial \mathbf{C}}{\partial \mathbf{C}} = \mathbf{I} = \frac{1}{2} (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}), \quad (4.207)$$

$$\frac{\partial \mathbf{C}^{-1}}{\partial \mathbf{C}} = -\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \mathbf{C}} \mathbf{C}^{-1} = -\mathbf{I}_{\mathbf{C}^{-1}} = -\frac{1}{2} (C_{ik}^{-1} C_{jl}^{-1} + C_{il}^{-1} C_{jk}^{-1}), \quad (4.208)$$

$$\frac{\partial \lambda_A^2}{\partial \mathbf{C}} = \lambda_A^2 \mathbf{M}^A, \quad (4.209)$$

$$\frac{\partial I_1}{\partial \mathbf{C}} = \mathbf{1}, \quad \frac{\partial I_3}{\partial \mathbf{C}} = I_3 \mathbf{C}^{-1}, \quad (4.210)$$

$$\frac{\partial d_A}{\partial \mathbf{C}} = (4\lambda_A^4 - I_1\lambda_A^2 - I_3\lambda_A^{-2})\mathbf{M}^A - \lambda_A^2\mathbf{1} + I_3\lambda_A^{-2}\mathbf{C}^{-1}. \quad (4.211)$$

The derivative of Eq. (4.206) with respect to \mathbf{C} becomes

$$\begin{aligned} \frac{\partial \mathbf{M}^A}{\partial \mathbf{C}} &= \frac{1}{d_A} \left[\frac{\partial \mathbf{C}}{\partial \mathbf{C}} - \mathbf{1} \otimes \left(\frac{\partial I_1}{\partial \mathbf{C}} - \frac{\partial \lambda_A^2}{\partial \mathbf{C}} \right) \right] \\ &\quad + \frac{1}{d_A} \left[\lambda_A^{-2} \mathbf{C}^{-1} \otimes \frac{\partial I_3}{\partial \mathbf{C}} - I_3 \lambda_A^{-4} \mathbf{C}^{-1} \otimes \frac{\partial \lambda_A^2}{\partial \mathbf{C}} + I_3 \lambda_A^2 \frac{\partial \mathbf{C}^{-1}}{\partial \mathbf{C}} - \mathbf{M}^A \otimes \frac{\partial d_A}{\partial \mathbf{C}} \right]. \end{aligned} \quad (4.212)$$

By using the property of Eqs. (4.207) through (4.211), the following explicit form can be obtained:

$$\begin{aligned} \frac{\partial \mathbf{M}^A}{\partial \mathbf{C}} &= \frac{1}{d_A} \left[\mathbf{I} - \mathbf{1} \otimes \mathbf{1} - I_3 \lambda_A^{-2} [\mathbf{I}_{\mathbf{C}^{-1}} - (\mathbf{C}^{-1} - \mathbf{M}^A) \otimes (\mathbf{C}^{-1} - \mathbf{M}^A)] \right] \\ &\quad + \frac{\lambda_A^2}{d_A} [(\mathbf{1} \otimes \mathbf{M}^A + \mathbf{M}^A \otimes \mathbf{1}) + (I_1 - 4\lambda_A^{-4}) \mathbf{M}^A \otimes \mathbf{M}^A] \end{aligned} \quad (4.213)$$

This relation was originally derived by Simo and Taylor [8]. The spatial version of Eq. (4.213) can be obtained through a transformation as

$$\begin{aligned} \frac{\partial \mathbf{m}^A}{\partial \mathbf{g}} &= \frac{1}{d_A} [\mathbf{I}_b - \mathbf{b} \otimes \mathbf{b} - I_3 \lambda_A^{-2} [\mathbf{I} - (\mathbf{1} - \mathbf{m}^A) \otimes (\mathbf{1} - \mathbf{m}^A)]] \\ &\quad + \frac{\lambda_A^2}{d_A} [(\mathbf{b} \otimes \mathbf{m}^A + \mathbf{m}^A \otimes \mathbf{b}) + (I_1 - 4\lambda_A^{-4}) \mathbf{m}^A \otimes \mathbf{m}^A], \end{aligned} \quad (4.214)$$

which is equivalent to $\mathbf{F}(\partial \mathbf{m}^A / \partial \mathbf{C}) \mathbf{F}^T$.

4.7 MATLAB Code for Elastoplastic Material Model

In this section, two MATLAB codes, PLSET.m and PLAST3D.m, are introduced that can solve for nonlinear elastoplastic problems with three different options: (1) $\text{MID} = 1$ for elastoplasticity with infinitesimal strain (Sect. 4.4), (2) $\text{MID} = 2$ for elastoplasticity with finite rotation (Sect. 4.4), and (3) $\text{MID} = 31$ for finite deformation elastoplasticity (Sect. 4.5). The codes are called from NLFEA.m in Chap. 2.

PLSET.m initializes plastic variables before starting analysis. Therefore, this function is called only once from NLFEA.m. First, PLSET.m allocates memory for global arrays for history-dependent variables, SIGMA and XQ. These variables are used at each integration point of every element. Since the current implementation uses two-point integration in each coordinate direction, the size of array should be $8 * \text{NE}$, where NE is the number of elements. At each integration point, the size of XQ is either 7 or 4, and that of SIGMA is 6 or 12, depending on plasticity

formulations. PLSET.m also calculates the elastic stiffness matrix ETAN using the two Lame's constants, LAM and MU.

As explained in Fig. 2.25, the main function of PLAST3D.m is to build the tangent stiffness matrix, $[K]$, and the residual force vector, $\{R\}$. Then, NLFEA.m will solve for the displacement increment as a part of the Newton–Raphson iteration. Also, PLAST3D keeps track of history-dependent variables and stores them in global arrays after the solution converges at a given load increment.

PLAST3D.m shares most of its input variables with that of ELAST3D.m in Chap. 1, which was explained in Table 1.5. The only difference is that MID and PROP are used in addition to ETAN. MID should be 1, 2, or 31 for elastoplastic material models. The array PROP stores elastoplastic material constants. The current implementation uses combined linear hardening model using von Mises criterion, which uses five material properties, PROP = [lambda, mu, beta, H, Y0]. The first two variables are Lame's constant, beta is the combined hardening parameter, H is the plastic modulus, and Y0 is the initial yield stress. As with ELAST3D.m in Chap. 1, the logical variable, UPDATE, is used to store the stresses and history variables in the global array SIGMA and XQ, respectively, and the logical variable, LTAN, is used to calculate the tangent stiffness matrices and store them in the global array GKF. The residual force, FORCE, will always be calculated.

In order to assemble the local stiffness matrix into the global stiffness matrix, the IDOF array is used to store the location of the global DOFs corresponding to the local 24 DOFs. The XG and WGT arrays store one-dimensional integration points and corresponding weights, as in Table 1.4. In this implementation, only two-point integration is used for each coordinate direction.

At each integration point of an element, the derivatives of finite element shape functions are calculated by calling SHAPEL.m. It is noted that the derivatives of shape function from SHAPEL.m is with respect to the undeformed configuration; that is, SHPD = $\partial N_i / \partial X_j$. Since the formulations with MID = 2 or 31 use the updated Lagrangian formulation, the material derivatives are converted into the spatial derivatives by multiplying with the inverse of the deformation gradient.

The global array XQ stores both the back stress and the effective plastic strain at the previous converged load increment. When MID = 1 or 2, it stores $[\alpha_{11}, \alpha_{22}, \alpha_{33}, \alpha_{12}, \alpha_{23}, \alpha_{13}, e_p]$ at each integration point, while when MID = 31, it stores $[\alpha_{11}, \alpha_{22}, \alpha_{33}, e_p]$ because this model uses the principal components of back stress. When MID = 1 or 2, the global array SIGMA stores stress components at the previous converged load increment, $[\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{13}]$. When MID = 1, SIGMA stores stress components as well as the elastic left Cauchy-Green deformation vector, $[\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{23}, \sigma_{13}, b_{11}, b_{22}, b_{33}, b_{12}, b_{23}, b_{13}]$.

With a given strain increment, DDEPS, combHard.m and mulPlast.m are used to calculate stress and history variables. For the case of finite rotation, the stress and back stress are rotated to the neutral configuration in rotatedStress.m. The outcomes of these functions are stress, back stress, effective plastic strain, and elastic left Cauchy-Green deformation vector. If the logical variable UPDATE is true, then these variables are stored in the global arrays, SIGMA and XQ. This only occurs when the residual becomes less than the convergence tolerance TOL (TOL is an input variable to NLFEA.m).

Once the stress is calculated, it is used to build the global residual vector, FORCE, which represents $\{\mathbf{f}^{\text{int}}\}$ in Eq. (4.130). The tangent stiffness array, GKF, is calculated only when the logical variable, LTAN, is true. This functionality can be used when the modified Newton–Raphson iteration is used.

```

function ETAN=PLSET (PROP, MID, NE)
%*****
% Initialize history variables and elastic stiffness matrix
% XQ    : Back stress alpha and Effective plastic strain
% SIGMA : Stress for rate-form plasticity
%          : added Left Cauchy-Green tensor B for multiplicative plasticity
% ETAN  : Elastic stiffness matrix
%*****
%%
global SIGMA XQ
%
LAM=PROP(1);
MU=PROP(2);
%
N = 8*NE;
%
if MID > 30
    SIGMA=zeros(12,N);
    XQ=zeros(4,N);
    SIGMA(7:9,:)=1;
    ETAN=[LAM+2*MU LAM      LAM      ;
           LAM      LAM+2*MU LAM      ;
           LAM      LAM      LAM+2*MU];
else
    SIGMA=zeros(6,N);
    XQ=zeros(7,N);
    ETAN=[LAM+2*MU LAM      LAM      0 0 0;
           LAM      LAM+2*MU LAM      0 0 0;
           LAM      LAM      LAM+2*MU 0 0 0;
           0        0        0        MU 0 0;
           0        0        0        0   MU 0;
           0        0        0        0   0   MU];
end
end

function PLAST3D(MID, PROP, ETAN, UPDATE, LTAN, NE, NDOF, XYZ, LE)
%*****
% MAIN PROGRAM COMPUTING GLOBAL STIFFNESS MATRIX RESIDUAL FORCE FOR
% PLASTIC MATERIAL MODELS
%*****
%%
global DISPDD DISPTD FORCE GKF XQ SIGMA
%
% Integration points and weights (2-point integration)

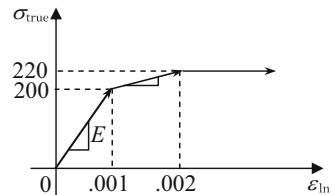
```

```

XG=[-0.57735026918963D0, 0.57735026918963D0];
WGT=[1.00000000000000D0, 1.00000000000000D0];
%
% Index for history variables (each integration pt)
INTN=0;
%
%LOOP OVER ELEMENTS, THIS IS MAIN LOOP TO COMPUTE K AND F
for IE=1:NE
    % Nodal coordinates and incremental displacements
    ELXY=XYZ(LE(IE,:),:);
    % Local to global mapping
    IDOF=zeros(1,24);
    for I=1:8
        II=(I-1)*NDOF+1;
        IDOF(II:II+2)=(LE(IE,I)-1)*NDOF+1:(LE(IE,I)-1)*NDOF+3;
    end
    DSP=DISPTD(IDOF);
    DSPD=DISPDD(IDOF);
    DSP=reshape(DSP,NDOF,8);
    DSPD=reshape(DSPD,NDOF,8);
%
%LOOP OVER INTEGRATION POINTS
for LX=1:2, for LY=1:2, for LZ=1:2
    E1=XG(LX); E2=XG(LY); E3=XG(LZ);
    INTN = INTN + 1;
%
% Determinant and shape function derivatives
[~, SHPD, DET] = SHAPEL([E1 E2 E3], ELXY);
FAC=WGT(LX)*WGT(LY)*WGT(LZ)*DET;
%
% Previous converged history variables
if MID > 30
    NALPHA=3;
    STRESSN=SIGMA(7:12,INTN);
else
    NALPHA=6;
    STRESSN=SIGMA(1:6,INTN);
end
ALPHAN=XQ(1:NALPHA,INTN);
EPN=XQ(NALPHA+1,INTN);
%
% Strain increment
if MID == 2 || MID == 31
    F=DSP*SHPD' + eye(3);
    SHPD=inv(F)/*SHPD;
end
DEPS=DSPD*SHPD';
DDEPS=[DEPS(1,1) DEPS(2,2) DEPS(3,3) ...
        DEPS(1,2)+DEPS(2,1) DEPS(2,3)+DEPS(3,2) DEPS(1,3)+DEPS(3,1)]';
%
```

```
% Computer stress, back stress & effective plastic strain
if MID == 1
    % Infinitesimal plasticity
    [STRESS, ALPHA, EP]=combHard(PROP, ETAN, DDEPS, STRESSN, ALPHAN, EPN) ;
elseif MID == 2
    % Plasticity with finite rotation
    FAC=FAC*det(F) ;
    [STRESSN, ALPHAN] = rotatedStress(DEPS, STRESSN, ALPHAN) ;
    [STRESS, ALPHA, EP]=combHard(PROP, ETAN, DDEPS, STRESSN, ALPHAN, EPN) ;
elseif MID == 31
    [STRESS, B, ALPHA, EP]=mulPlast(PROP, ETAN, DEPS, STRESSN, ALPHAN, EPN) ;
end
%
% Update plastic variables
if UPDATE
    SIGMA(1:6, INTN)=STRESS;
    XQ(:, INTN) = [ALPHA; EP];
    if MID > 30
        SIGMA(7:12, INTN)=B;
    end
    continue;
end
%
% Add residual force and tangent stiffness matrix
BM=zeros(6,24);
BG=zeros(9,24);
for I=1:8
    COL=(I-1)*3+1:(I-1)*3+3;
    BM(:,COL)=[SHPD(1,I) 0 0;
                0 SHPD(2,I) 0;
                0 0 SHPD(3,I);
                SHPD(2,I) SHPD(1,I) 0;
                0 SHPD(3,I) SHPD(2,I);
                SHPD(3,I) 0 SHPD(1,I)];
    %
    BG(:,COL)=[SHPD(1,I) 0 0;
                SHPD(2,I) 0 0;
                SHPD(3,I) 0 0;
                0 SHPD(1,I) 0;
                0 SHPD(2,I) 0;
                0 SHPD(3,I) 0;
                0 0 SHPD(1,I);
                0 0 SHPD(2,I);
                0 0 SHPD(3,I)];
end
%
% Residual forces
FORCE(IDOF) = FORCE(IDOF) - FAC*BM'*STRESS;
%
```

Fig. 4.25 Hardening models for elastoplasticity



```
% Tangent stiffness
if LTAN
    if MID == 1
        DTAN=combHardTan(PROP, ETAN, DDEPS, STRESSN, ALPHAN, EPN);
        EKF = BM' * DTAN * BM;
    elseif MID == 2
        DTAN=combHardTan(PROP, ETAN, DDEPS, STRESSN, ALPHAN, EPN);
        CTAN=[-STRESS(1) STRESS(1) STRESS(1) -STRESS(4) 0 -STRESS(6);
               STRESS(2) -STRESS(2) STRESS(2) -STRESS(4) -STRESS(5) 0;
               STRESS(3) STRESS(3) -STRESS(3) 0 -STRESS(5) -STRESS(6);
               -STRESS(4) -STRESS(4) 0 -0.5*(STRESS(1)+STRESS(2)) -0.5*STRESS(6)
               -0.5*STRESS(5);
               0 -STRESS(5) -STRESS(5) -0.5*STRESS(6) -0.5*(STRESS(2)+STRESS(3)) -
               0.5*STRESS(4);
               -STRESS(6) 0 -STRESS(6) -0.5*STRESS(5) -0.5*STRESS(4) -0.5*(STRESS
               (1)+STRESS(3))];
        SIG=[STRESS(1) STRESS(4) STRESS(6);
              STRESS(4) STRESS(2) STRESS(5);
              STRESS(6) STRESS(5) STRESS(3)];
        SHEAD=kroon(eye(3),SIG);
        EKF = BM' * (DTAN+CTAN)*BM + BG'*SHEAD*BG;
    elseif MID == 31
        DTAN=mulPlastTan(PROP, ETAN, DEPS, STRESSN, ALPHAN, EPN);
        SIG=[STRESS(1) STRESS(4) STRESS(6);
              STRESS(4) STRESS(2) STRESS(5);
              STRESS(6) STRESS(5) STRESS(3)];
        SHEAD=zeros(9);
        SHEAD(1:3,1:3)=SIG;
        SHEAD(4:6,4:6)=SIG;
        SHEAD(7:9,7:9)=SIG;
        %
        EKF = BM' * DTAN*BM + BG' *SHEAD*BG;
    end
    GKF(IDOF, IDOF)=GKF(IDOF, IDOF)+FAC*EKF;
end, end, end
end
end
end
```

4.8 Elastoplasticity Analysis of Using Commercial Finite Element Programs

4.8.1 Usage of Commercial Programs

In this section, elastoplasticity analysis procedures using three commercial finite element programs are discussed.

1. Abaqus

Although elastoplasticity analysis for isotropic material using von Mises criterion is presented in this chapter, Abaqus supports broader range of plasticity including anisotropic plasticity, rate-dependent yield criteria, creep and swelling, porous media plasticity, etc. In this section, the metal plasticity with von Mises criterion in Abaqus will be discussed. Abaqus supports linear and nonlinear isotropic/kinematic hardening models.

The elastic material properties are the same as elastic material. For small elastic strain assumption, the linear elastic material properties can be used for this purpose (keyword ***ELASTIC**). In order to specify hardening properties, Abaqus requires data in the pairs of yield stress and effective plastic strain.

In Abaqus plastic materials, Cauchy stress and logarithmic plastic strain are used. Since material properties in tensile test are calculated from nominal stress and engineering strain, they need to be converted using the following relations:

$$\sigma_{\text{true}} = \sigma_{\text{nom}}(1 + \epsilon_{\text{nom}}), \quad (4.215)$$

$$\epsilon_{\ln}^{\text{pl}} = \epsilon_{\ln} - \epsilon_{\ln}^{\text{el}} = \ln(1 + \epsilon_{\text{nom}}) - \frac{\sigma_{\text{true}}}{E}, \quad (4.216)$$

where σ_{true} and σ_{nom} are respectively the Cauchy stress and nominal stress, ϵ_{nom} is engineering strain, and $\epsilon_{\ln}^{\text{pl}}$ is the logarithmic plastic strain. Consider the Cauchy stress and logarithmic strain curve in Fig. 4.25. The Young's modulus can be calculated from the initial slope of the curve and it becomes $E = 200$ GPa. Let the Poisson's ratio be 0.3. If the stress unit is MPa, then the ***ELASTIC** keyword is defined as

```
*ELASTIC
200.E3,.3
```

The material starts yielding at $\sigma_{\text{true}} = 200$ MPa and is linearly hardened until 220 MPa at logarithmic strain $\epsilon_{\ln} = 0.002$. At the initial yielding the plastic strain is 0, and at $\epsilon_{\ln} = 0.002$, the plastic strain is $\epsilon_{\ln}^{\text{pl}} = 0.0009$. After that the hardening is 0. In this case, the keyword ***PLASTIC** can be used to describe the hardening as

```
*PLASTIC
200.0, 0.0
220.0, 0.0009
```

Fig. 4.26 Extension of an elastoplastic unit cube

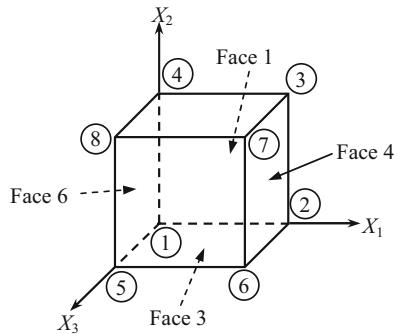
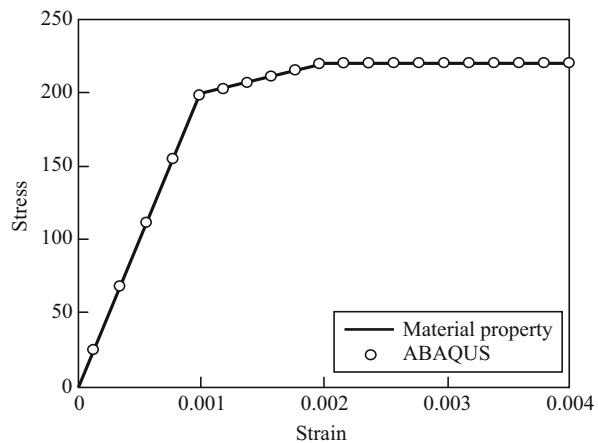


Fig. 4.27 Stress–strain response for extension of a cube



By default, Abaqus assumes the isotropic hardening model. In order to use kinematic or combined hardening model, **HARDENING** option can be used. The following keyword defines elastoplastic material with linear kinematic hardening:

***PLASTIC, HARDENING=KINEMATIC**

Example 4.23 (Uniform Tension of a Cube) Consider a unit cube as shown in Fig. 4.26. An eight-node solid element (C3D8) is used to model the cube. The positive X_1 face (Face 4) is extended with a strain $\epsilon = 0.004$. The following boundary conditions are given: $u_1 = 0$ at Face 6, $u_2 = 0$ at Face 3, and $u_3 = 0$ at Face 1. Using Abaqus, calculate the relation between Cauchy stress and strain. Use elastoplastic with isotropic hardening material. The elastic properties are $E = 200$ GPa and $\nu = 0.3$. The initial yields stress is $\sigma_Y = 200$ MPa and is linearly hardened until 220 MPa at logarithmic strain $\epsilon_{ln} = 0.002$. After that, there is no hardening (see Fig. 4.25). Compare the stress–strain curve from the material definition and the response from Abaqus.

Solution Below is the list of Abaqus commands used to solve the uniform extension of an elastoplastic cube. An eight-node linear brick element, C3D8, in Abaqus

is used. The stress–strain curve in Fig. 4.27 shows that the numerical results agree well with the hardening curve.

```

*HEADING
- Extension of elastoplastic cube
*NODE, NSET=ALL
1,
2,1.
3,1.,1.,
4,0.,1.,
5,0.,0.,1.
6,1.,0.,1.
7,1.,1.,1.
8,0.,1.,1.
*NSET, NSET=FACE1
1,2,3,4
*NSET, NSET=FACE3
1,2,5,6
*NSET, NSET=FACE4
2,3,6,7
*NSET, NSET=FACE6
4,1,8,5
*ELEMENT, TYPE=C3D8, ELSET=ONE
1,1,2,3,4,5,6,7,8
*SOLID SECTION, ELSET=ONE, MATERIAL=ALLE
■

```

2. ANSYS

ANSYS supports bilinear/multilinear/nonlinear isotropic/kinematic hardening models as well as anisotropic plasticity. The term bilinear is the same with linear hardening in this text because ANSYS considers that the material response is linear for both elastic and plastic states. For elastoplastic materials, **TB**, **TBDATA**, and **TBPT** commands are used to define stress–strain behavior. By default, ANSYS assumes infinitesimal deformation. When structures experience large strain, the **NLGEOM** command should be used to include the effect of geometric nonlinearity. For large strain analyses, stress–strain properties must be input in terms of true stress and logarithmic strain.

TB, Lab, MAT, NTEMP, NPTS

The **TB** command activates a data table for nonlinear material properties. The first parameter **Lab** specifies hardening models. **Lab** can take the following options:

BISO: Bilinear isotropic hardening

BKIN: Bilinear kinematic hardening

KINH: Multilinear kinematic hardening (strain–stress data)

MISO: Multilinear isotropic hardening

MKIN: Multilinear kinematic hardening (plastic strain–stress data)

PLASTIC: Nonlinear plasticity with stress vs. plastic strain data

The second parameter, **MAT**, is a material reference number that can be used by **MAT** command. If the temperature varies during the plasticity analysis, **NTEMP** is used to set the number of temperatures for which data will be provided. **NPTS** is the number of data points to be specified for a given temperature. Data points are defined with the **TBDATA** or **TBPT** commands.

TB DATA, STLOC, C1, C2, C3, C4, C5, C6

The **TB DATA** command defines data for the data table. The first parameter **STLOC** is the starting location in table for entering data. For example, if **STLOC** = 1, data input in the **C1** field applies to the first table constant, **C2** applies to the second table constant, etc. If **STLOC** = 5, data input in the **C1** field applies to the fifth table constant, etc. **C1, C2, C3, ..., C6** are data values assigned to six locations starting with **STLOC**.

TBPT, Oper, X, Y

The **TBPT** command adds/deletes a point on a nonlinear data curve. If **Oper** = **DEFI**, it adds a point at (**X**, **Y**). If a point already exists with the same **X** value, it is replaced. If **Oper** = **DELETE**, it deletes a point at point **X**.

The following example defines an elastoplastic material with bilinear kinematic hardening model in the Newton millimeter units. **TB DATA** is used to provide the yield stress and hardening modulus.

```
MP,EX,1,200E3      ! Young's modulus = 200GPa
MP,PRXY,1,0.3      ! Poisson's ratio = 0.3
TB,BKIN,1,1         ! Activate a data table
TBDATA,1,200.,20.E3 ! σy = 200MPa; H = 20GPa
```

The following example defines an elastoplastic material with multilinear kinematic hardening model. Three data points are provided in terms of (strain, stress).

```
TB,KINH,1,1,3      ! Activate a data table
TBPT,,0.001,200.   ! Strain = 0.001, Stress = 200MPa
TBPT,,0.002,220.   ! Strain = 0.002, Stress = 220MPa
TBPT,,0.200,250.   ! Strain = 0.200, Stress = 250MPa
```

3. NEiNastran

NEiNastran supports four different yield criteria using von Mises, Tresca, Mohr-Coulomb, and Drucker-Prager models. In terms of hardening, it supports isotropic, kinematic, and combined hardening models. **MATS1** bulk data card is used to specify elastoplastic material properties.

MATS1

This entry defines elastoplastic hardening models and parameters. It must be associated with the elastic material properties that are defined in **MAT1**, **MAT2**, **MAT8**, **MAT9**, or **MAT12**. The parameters of the **MATS1** entry are as follows:

MATS1	MID	TID	TYPE	H	YF	HG	LIM1	LIM2	
--------------	------------	------------	-------------	----------	-----------	-----------	-------------	-------------	--

The **MATS1** card can define either nonlinear elastic material or elastoplastic material. For the latter, **TYPE = PLASTIC** is used. **MID** is the identification number of **MAT1**, **MAT2**, **MAT8**, **MAT9**, or **MAT12**. The hardening parameters can be provided either using either the table identification **TID** or the work hardening slope **H**, but not both. **H** is the plastic modulus (slope of stress vs. plastic strain) in units of stress. For more than a single slope in the plastic range, the stress-strain data must be supplied on a **TABLES1** entry referenced by **TID**, and this field must be blank. **YF** field specifies the yield function criterion (1 = von Mises, 2 = Tresca, 3 = Mohr-Coulomb, 4 = Drucker-Prager). The hardening rule is specified in **HG** (1 = isotropic, 2 = kinematic, 3 = combined isotropic and kinematic hardening). If **TID** is given, **TABLES1** entries (**xi**, **yi**) of stress-strain data (ϵ_k , σ_k) must conform to the following rules. The curve must be defined in the first quadrant. The first point must be at origin ($\epsilon_1 = 0$, $\sigma_1 = 0$) and the second point (**x2**, **y2**) must be at the initial yield point (σ_Y) specified on the **MATS1** entry. The slope of the line joining the origin to the yield stress must be equal to the valued of E .

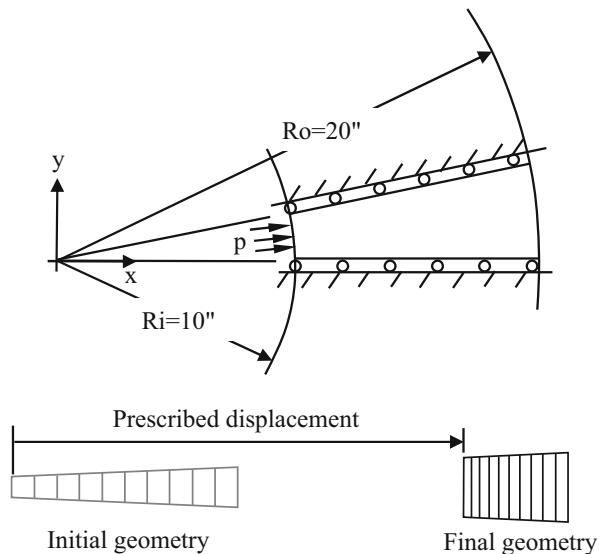
4.8.2 Modeling Examples of Elastoplastic Materials

In this section, several analysis problems are used to discuss about modeling issues as well as verifying the accuracy of analysis results with that of literature.

Elastoplastic Cylinder Under Internal Pressure: An elastoplastic cylinder, subjected to internal pressure, is shown in Fig. 4.28. Assuming that both ends are fixed, a plane strain condition can be applied. The internal pressure is applied such that the inner radius experiences a large change (a factor of three). The geometry of the cylinder is given such that the inner radius is 254 mm (10 in.) and the outer radius of 508 mm (20 in.). It is noted that axisymmetric modeling can also be used as both the geometry and load conditions are identical for a given angle q . Either ten four-node quadrilateral elements (CPE4 for Abaqus or CQUAD4 for NEiNastran) or five eight-node hexahedral elements (CPE8 for Abaqus or CQUAD8 for NEiNastran) can be used to model the cylinder with periodic boundary condition. The cylinder is assumed to be made of an elasto-perfectly-plastic material with the following material properties: Young's modulus $E = 207$ GPa (30,000 ksi), Poisson's ratio $\nu = 0.3$, and yield strength $\sigma_Y = 207$ MPa (30 ksi).

This is a good example to discuss about the difference between the force-controlled and displacement-controlled method. In the case of elasto-perfectly-plastic material, the material cannot support loads beyond its yield strength. Therefore, if a force-controlled method is used, it is possible that the user can apply a load larger than the material can support. Then, the system becomes unstable and the

Fig. 4.28 Elastoplastic cylinder under internal pressure



nonlinear iteration diverges. In fact, at the point of material yield, the deformation continuously increases without further increase in the load. On the other hand, the displacement-controlled method can maintain stability during nonlinear iteration even if the force remains constant or decreases. This difference is explained in Fig. 2.23 of Chap. 2. In the case of elastoplastic cylinder under internal pressure, the inner radius is gradually increased by prescribing the radial displacement at the innermost nodes, and the applied pressure is then calculated as a reaction force to these prescribed displacements. The cylinder is expanded to three times its initial radius in a small number of increments. This requires very large strain increments and would probably be too large for a more complicated problem that involves shear and rotation as well as direct straining. However, large strain increments are suitable for this simple case.

Normally, it is difficult to obtain analytical solution for elastoplastic problems except for very limited cases. In this example, however, since the strains are so large, the results can be compared by the exact, rigid-plastic solution by Prager and Hodge [9]. The stresses from the rigid-plasticity theory are given as

$$\begin{aligned}\sigma_{rr}(R) &= \frac{1}{\sqrt{3}}\sigma_Y \ln \left[\frac{R^2 + r_i^2 - R_i^2}{R_o^2 + r_i^2 - R_i^2} \right] \\ \sigma_{\theta\theta}(R) &= \sigma_{rr}(R) + \frac{2}{\sqrt{3}}\sigma_Y \\ \sigma_{zz}(R) &= \sigma_{rr}(R) + \frac{1}{\sqrt{3}}\sigma_Y\end{aligned},$$

where \$R_i\$ and \$R_o\$ are the initial inner and outer radii, respectively; \$r_i\$ is the current inner radius; and \$R\$ is the radius, in the initial configuration, of the material point at

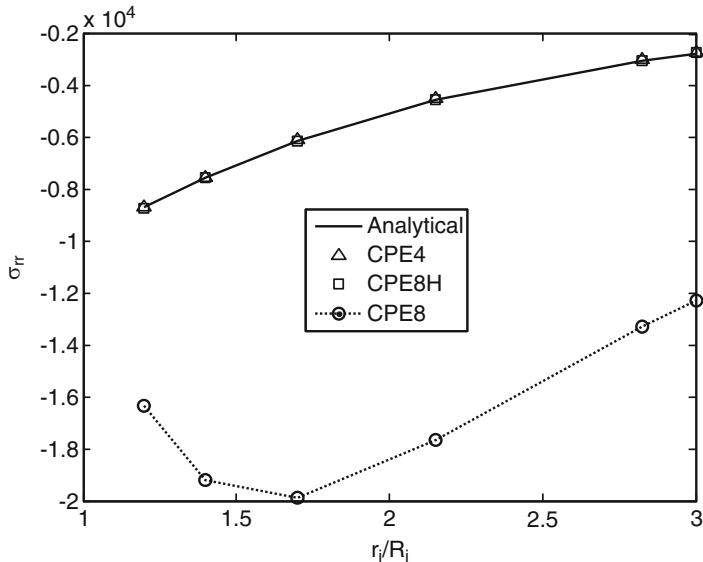


Fig. 4.29 Stress σ_{rr} at $R/R_i = 1.5$

which the stresses are being calculated. Note that since $\sigma_{\theta\theta}$ and σ_{zz} are a function of σ_{rr} , only σ_{rr} needs to be compared.

In order to test the performance of different element formulations, three types of elements are used in Abaqus: CPE8, CPE8H, and CPE4. Figure 4.29 compares the radial stress at $R/R_i = 1.5$ location at different levels of deformation using these three element models to that given by the exact, rigid-plastic solution. Both CPE8H and CPE4 modes agree very closely with the exact solution, but the results from the fully integrated 8-node (CPE8) element are significantly different.

The pure displacement 8-node elements (CPE8) give poor results because the strains are calculated directly from the interpolation functions at each integration point and the incompressibility requirement causes a severe oscillation in the mean pressure stress throughout each element. However, in the hybrid, eight-node elements the mean pressure stress is interpolated independently, so an accurate value is obtained for this variable. In addition, the four-node elements in Abaqus are constant strain/stress elements for this case (because these elements are coded with a constant hoop strain value and use “selective reduced integration,” in which the volume strain is computed at the centroid only) and so also provide accurate pressure stress values.

Results for models using the fully integrated versions of plane strain elements are shown here to caution the user. With rare exceptions the fully integrated 8-node quadrilaterals are not as effective as the reduced integration versions of the same elements; the reduced integration 8-node quadrilaterals are, hence, almost always recommended over their fully integrated counterparts. This particular problem

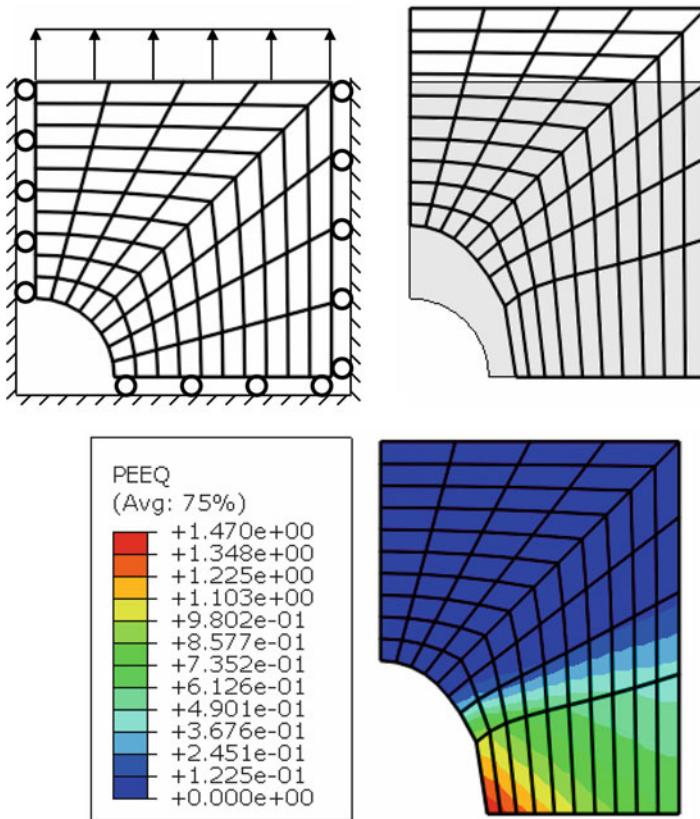


Fig. 4.30 Finite element model and results for a plate with a hole

gives a dramatic illustration of a difficulty encountered with full integration in a problem in which the bulk behavior of the material is very much stiffer than the shear behavior, a type of behavior commonly encountered.

Stretching of a Plate with a Hole: A square 30×30 plate containing a hole of radius 4 is stretched in the y -direction, while displacements in the x -direction are restrained along its outer perimeter. Figure 4.30 shows the initial quarter symmetry models with four-node quadrilateral elements. The elastic material properties of the plate are a Young's modulus of 1×10^9 and a Poisson's ratio of 0.3. The isotropic von Mises plasticity specification uses constant isotropic hardening with an initial yield of 1×10^6 and a hardening modulus of 4×10^5 .

The analysis is performed in Abaqus/Explicit, where the plate is stretched by ramping the velocity at the top nodes to 5 for the first half of the step time and then keeping a constant velocity of 5 at these nodes for the rest of the analysis. At the end of time, the vertical displacement at the top nodes is 2.5.

The contours of the equivalent plastic strain in each of the plates, obtained from the analysis performed exclusively in Abaqus/Explicit, are shown in Fig. 4.30.

Inspection of the deformed shapes and regions of high plastic strain shows that the hole enlarges not only in the stretching direction but also in the lateral direction, where the highest plastic deformation is observed. The contour of equivalent plastic strain at the end of the import analysis is also shown in Fig. 4.30.

4.9 Summary

In this chapter, finite element formulations for elastoplastic problems are discussed, which correspond to material nonlinearity. Elastoplastic problems are considered rough nonlinear because their responses are history dependent and the status of material can change abruptly. Permanent material dislocation during plastic deformation is represented using the evolution of internal plastic variables. For large deformation problems, both material and geometry nonlinearities exist, which makes the problem more difficult to solve.

First, one-dimensional plasticity is introduced with linear hardening models and small deformation assumption in Sect. 4.2. Two different hardening models are discussed: isotropic and kinematic hardenings. The former increases the elastic domain, while the latter maintains the size of elastic domain but moves the center of it. From the small strain assumption, the strain is additively decomposed into elastic and plastic strains in which only elastic strain is related to stress. Plastic deformation depends on load history and it is stored in plastic strain. The state determination of stress is based on (a) elastic trial and (b) plastic return-mapping. In the elastic trial state, the strain increment is assumed to be elastic and stress increases accordingly. If the trial stress is out of elastic range (i.e., beyond the current yield stress), it is returned to the yield stress. The plastic strain increment is identified during this return-mapping process.

In multidimensional stress states, it is impractical to perform tests in all possible stress combinations. One-dimensional tension test data can still be used for determining failure of multidimensional stress using failure theories, which are based on equivalent stress. Since failure criteria should be independent of coordinate systems, they are defined using invariants. The Tresca criterion uses the maximum shear stress, while the von Mises criterion uses the second invariant (J_2) of deviatoric stress. Same as one-dimensional case, the algorithm is composed of elastic trial and plastic return-mapping. In the case of von Mises criterion, the return-mapping occurs in the radial direction of deviatoric stress. While the stress is returning from the trial states, the yield surface varies simultaneously. The final return-mapping point is determined by the plastic consistency condition. It is shown that when linear hardening is used, this return-mapping point can be found explicitly. Otherwise, the local Newton–Raphson method is required to find the return-mapping point. Since the continuum tangent stiffness is inconsistent with finite step size of time integration, the convergence iteration does not show quadratic convergence. In order to guarantee the quadratic convergence of the Newton–Raphson iteration, an algorithmic tangent stiffness that is

consistent with the return-mapping algorithm is obtained by differentiating the time integration algorithm with finite step size.

When the structure experiences small strain but finite rigid-body motion, the classical theory of elastoplasticity with the assumption of infinitesimal deformation needs to be modified to take into account the rigid-body motion. The objective rate and objective time integration play important roles to express the rigid-body motion systematically. An objective stress rate must be used to define the constitutive relation because the material response should be independent of coordinate systems. In addition, the midpoint configuration is used to reduce errors involved in nonuniform rotation and spin. Although a good deal of research has been performed on the objective rate, difficulties still remain concerning numerical integration methods that satisfy all physical requirements. The difficulty in obtaining an exact tangent stiffness operator is another drawback to this approach.

A new method for expressing the kinematics of finite deformation elastoplasticity using the hyperelastic constitutive relation is becoming a desirable approach to isotropic material. This method defines a stress-free intermediate configuration composed of a plastic deformation, and obtains the stress simply by taking a derivative of the strain energy density with respect to the intermediate configuration. The multiplicative decomposition of an elastoplastic deformation is converted into an additive decomposition by defining appropriate stress and strain measures. Even if the final variational equation is represented using the updated Lagrangian formulation, the reference for a constitutive relation is implicitly a stress-free intermediate configuration. By using the constitutive relation between principal stresses and logarithmic stretches, better accuracy is obtained for a large elastic strain problem than with the classical elastoplasticity method. In addition, the same return-mapping algorithm from classical theory can be used in the principal stress space.

4.10 Exercises

- P4.1 A force is gradually applied at the end of an elastoplastic bar such that it is in the plastic phase. When the total magnitude of strain is $\varepsilon = 0.003$, calculate the applied force, axial stress, elastic strain, and plastic strain. Use the following material properties: $E = 100 \text{ GPa}$, $H = 10 \text{ GPa}$, and $\sigma_Y = 100 \text{ MPa}$. The cross-sectional area of the bar is $A = 1.0 \times 10^{-4} \text{ m}^2$.
- P4.2 A force 12 kN is gradually applied and then removed at the end of an elastoplastic bar. When the yield stress of the material is 100 MPa, calculate plastic strains and tip displacement after removing the applied force. Use the following material properties: $E = 100 \text{ GPa}$ and $H = 10 \text{ GPa}$. The cross-sectional area of the bar is $A = 1.0 \times 10^{-4} \text{ m}^2$ and the length of the bar is $L_0 = 1 \text{ m}$.
- P4.3 A uniaxial bar is under tensile force $F = 12 \text{ kN}$ at load step t_n . (a) When the plastic strain is $\varepsilon_p^n = 0.002$, determine the yield status of the material. (b) If

the applied force is increased to $F = 15 \text{ kN}$ at load step t_{n+1} , calculate plastic strain and tip displacement. Assume the initial yield stress $\sigma_Y = 100 \text{ MPa}$, $E = 100 \text{ GPa}$, and $H = 10 \text{ GPa}$. The cross-sectional area of the bar is $A = 1.0 \times 10^{-4} \text{ m}^2$ and the length of the bar is $L_0 = 1 \text{ m}$.

P4.4 An elastoplastic bar is under variable load history. At load step t_n , the stress and plastic strain are $\sigma^n = 200 \text{ MPa}$ and $\epsilon_p^n = 1.0 \times 10^{-4}$, respectively. (a) Is the material in elastic or plastic state? (b) When strain increment is $\Delta\epsilon = -0.003$, calculate stress and plastic strain. Assume isotropic hardening with $E = 200 \text{ GPa}$, $H = 25 \text{ GPa}$, and $\sigma_Y = 250 \text{ MPa}$.

P4.5 Repeat Problem P4.4 using the kinematic hardening model. For back stress, use $\alpha^n = 2.5 \text{ MPa}$.

P4.6 Repeat Problem P4.5 using the combined hardening model with $\beta = 0.5$.

P4.7 For the combined isotropic/kinematic hardening model, derive the expression of plastic strain increment from the plastic consistency condition.

P4.8 An elastoplastic bar is clamped at the left end, and variable loads are applied at the right end, as shown in the table. Plot the stress–strain curve by changing the applied forces by 5 kN increments. Assume the following material properties with isotropic hardening: $E = 70 \text{ GPa}$, $H = 10 \text{ GPa}$, $\sigma_Y = 250 \text{ MPa}$. The length of the bar is $L = 1 \text{ m}$, and the cross-sectional area is $A = 1.0 \times 10^{-4} \text{ m}^2$.

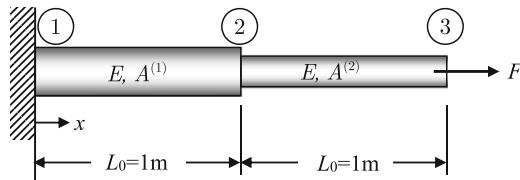
Load step	1	2	3	4
Force (kN)	30	20	35	20

P4.9 An elastoplastic bar is clamped at the left end, and variable displacements are applied at the right end, as shown in the table. Plot the stress–strain curve by changing the tip displacement by 1 mm increments. Assume the following material properties with isotropic hardening: $E = 70 \text{ GPa}$, $H = 10 \text{ GPa}$, $\sigma_Y = 250 \text{ MPa}$. The length of the bar is $L = 1 \text{ m}$, and the cross-sectional area is $A = 1.0 \times 10^{-4} \text{ m}^2$.

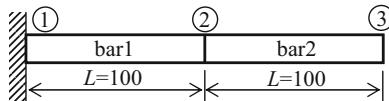
Load step	1	2	3	4
Displacement (mm)	5.0	3.0	7.0	6.0

P4.10 A force of $P = 15$ is applied to the two parallel bars in Example 4.2 and then removed. Using **combHard1D** programs, calculate tip displacement and residual stresses for the two bars after unloading. Use 15 load increments for each loading and unloading cycle. Plot stresses vs. tips displacement in the XY graph.

P4.11 A force 12 kN is gradually applied at the end of an elastoplastic bar. When the yield stress of the material is 100 MPa, calculate displacement at the tip. Use the following material properties: $E = 100 \text{ GPa}$ and $H = 10 \text{ GPa}$. The cross-sectional areas of the bars are $A^{(1)} = 1.0 \times 10^{-4} \text{ m}^2$ and $A^{(2)} = 0.5 \times 10^{-4} \text{ m}^2$.

**Fig. P4.11**

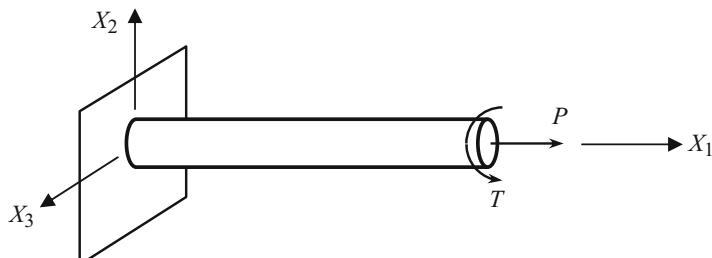
- P4.12 Two one-dimensional bars are connected serially as shown in the figure. At load step n , bar1 was plastic and bar2 was elastic. At load step $n+1$, the increments of nodal displacements are given as $\Delta\mathbf{u} = [\Delta u_1, \Delta u_2, \Delta u_3] = [0.0, -0.01, 0.0]$. Calculate stresses and plastic strains of both bars at load step $n+1$.

**Fig. P4.12**

	bar1	bar2
Young modulus (E)	10,000	5,000
Tangent modulus (E_t)	1,000	500
Previous stress (σ^n)	6.0	7.4
Initial yield stress (σ_Y)	5.0	7.5
Plastic strain (ϵ_p)	$9E-4$	0.0
Yield status	Plastic	Elastic
Hardening	Isotropic	Isotropic

- P4.13 Write the expression of the fourth-order unit symmetric tensor and unit deviatoric tensor in the 6×6 matrix notation.

- P4.14 A solid shaft as shown in the figure is subjected to tensile force P and a torque T . The force and torque are such that the normal stress $\sigma_{xx} = \sigma$ and shear stress $\tau = \sigma$. The shear stress is along the circumference of the shaft. Using the von Mises criterion, determine the values of σ when the material yields first time. The yield stress from the uniaxial tension test is σ_Y .

**Fig. P4.14**

- P4.15 A plane stress plate is under biaxial stress state in which $\sigma_{xx} = -\sigma_{yy} = \sigma$. When the applied load is proportional, determine σ when the material yields first time. The yield stress from the uniaxial tension test is σ_Y .
- P4.16 A square is under proportional loading with shear stress $\tau_{12} = \tau_{21} = \tau$. When the effective plastic strain is $e_p = 0.1$, calculate the value of shear stress. Consider three different hardening models: (a) isotropic, (b) kinematic, and (c) combined hardening with $\beta = 0.5$. Assume that the initial yield stress is 400 MPa and the plastic modulus is $H = 200$ MPa.
- P4.17 A pure shear deformation is applied to the square element as shown in the figure such that $\sigma_{12} = \sigma_{21}$ is only nonzero stress component. At load step n , the stress value was $\sigma_{12} = 50$, and there was no plastic deformation. At load step $n + 1$, incremental strain $\Delta e_{12} = \Delta e_{21} = 0.005$ is applied. Calculate stress components and effective plastic strain at load step $n + 1$. Use the following material properties: shear modulus $\mu = 1,000$, plastic modulus $H = 100$, initial yield stress $\sigma_Y = 100$.

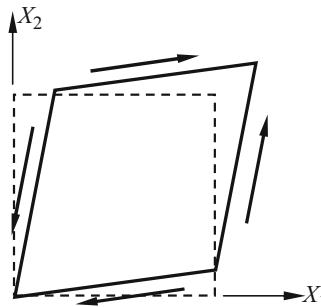


Fig. P4.17

- P4.18 Displacements of a simple shear deformation in the figure can be expressed by $u_1 = kx_2, u_2 = 0$. At load step n , $k = 0.016$ and the material is elastic. At load step $n + 1$, $\Delta k = 0.008$. Calculate stress and plastic strain. Check if the updated state is on the yield function, i.e., $f(\sigma^{n+1}, e_p^{n+1}) = 0$. Use the following material properties: shear modulus $\mu = 100$, plastic modulus $H = 10$, initial yield stress $\sigma_Y = \sqrt{12}$.

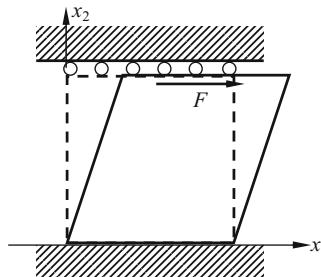


Fig. P4.18

- P4.19 At load step t_n , a unit square is under uniaxial stress state with $\sigma_{11} = 100$ MPa, and all other stress components and plastic variables are 0. At load step t_{n+1} , additional shear stress is applied such that $\Delta\gamma_{12} = 0.002$. Determine stress, back stress, and effective plastic strain. Assume the following material properties: $\lambda = \mu = 100$ GPa, $H = 10$ GPa, $\sigma_Y = 100$ MPa, combined isotropic/kinematic hardening with $\beta = 0.5$.
- P4.20 Using Abaqus perform a uniaxial tension test of a unit cube (C3D8) in x_3 -direction. Assume elastoplastic material with linear isotropic hardening ($E = 2.0E5$, $\nu = 0.3$, $\sigma_Y = 200$, $H = 2.0E4$). Displace at $x_3 = 1$ surface is controlled as shown in the figure with three steps. Use ten increments in each step. Plot stress-strain curve for all 30 increments.

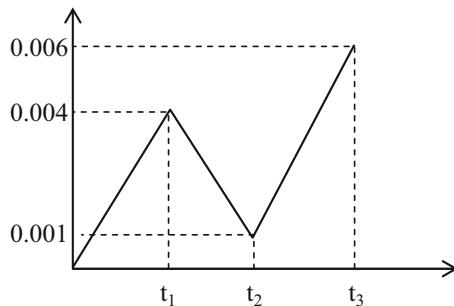


Fig. P4.20

- P4.21 Calculate D_{ep} and D^{alg} for one-dimensional elastoplasticity problem using the von Mises yield criterion and linear combined isotropic/kinematic hardening. Assume material properties: $(E, H, \sigma_Y^0, \beta)$.
- P4.22 In the saturated isotropic hardening model, the yield stress starts from initial value of σ_Y^0 and approaches σ_Y^∞ as the plastic strain increases.

$$\kappa(e_p) = \sigma_Y^0 + (\sigma_Y^\infty - \sigma_Y^0) \left[1 - \exp\left(-\frac{e_p}{e_p^\infty}\right) \right].$$

Since the hardening model is nonlinear, it is required to have a local Newton-Raphson method to find the plastic consistency parameter. Modify MATLAB program **combHard** so that it can solve for the above saturated isotropic hardening model. Test the program by solving the pure shear problem in P4.15. Assume the following material properties: shear modulus $\mu = 1,000$, plastic modulus $H = 100$, initial yield stress $\sigma_Y = 100$, asymptotic yield stress = 200, and asymptotic effective plastic strain = 0.05.

- P4.23 An plane strain square undergoes the following elastic deformation:

$$x_1 = X_1 + kX_2, \quad x_2 = X_2, \quad x_3 = X_3.$$

Using the linear relationship between principal Kirchhoff stress and logarithmic stretch, find the Kirchhoff stress tensor when $k = 0.02$. Use the following material properties: $\lambda = \mu = 100$ GPa.

- P4.24 A history of biaxial loadings is applied to a $1 \text{ mm} \times 1 \text{ mm}$ square, as shown in the figure. The square is constrained in the Y -direction along the bottom edge and in the X -direction along the left edge. The model is displaced in the X and Y directions at the right and top edges by $R = 2.5 \times 10^{-5} \text{ mm}$, respectively. Calculate σ_{xx} , σ_{yy} , σ_{zz} , and von Mises stress at each load step. Use the following material properties: $E = 250 \text{ GPa}$, $\nu = 0.25$, $\sigma_y = 5 \text{ MPa}$, and $E_T = 50 \text{ GPa}$.

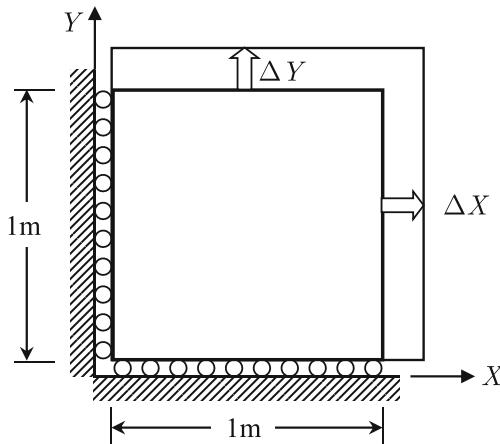


Fig. P4.24

Load step	ΔX	ΔY	Description
1	R	0	First yield
2	R	0	Plastic flow
3	0	R	Elastic unloading
4	0	R	Plastic reloading
5	$-R$	0	Plastic flow
6	$-R$	0	Plastic flow
7	0	$-R$	Elastic unloading
8	0	$-R$	Plastic flow

References

1. Simo JC, Govindjee S. Nonlinear B-stability and symmetric preserving return mapping algorithms for plasticity and viscoplasticity. *Int J Numer Methods Eng.* 1991;31:151–76.
2. Simo JC, Taylor RL. Consistent tangent operator for rate-independent elastoplasticity. *Comput Methods Appl Mech Eng.* 1985;48:101–18.
3. Hughes TJR, Winget J. Finite rotation effects in numerical integration of rate constitutive equations arising in large-deformation analysis. *Int J Numer Methods Eng.* 1980;15:1862–7.

4. Malvern LE. Introduction to mechanics of a continuous medium. Englewood-Cliff: Prentice-Hall; 1969.
5. Fish J, Shek K. Computational aspects of incrementally objective algorithms for large deformation plasticity. *Int J Numer Methods Eng.* 1999;44:839–51.
6. Simo JC. Algorithms for static and dynamic multiplicative plasticity that preserve the classical return mapping schemes of the infinitesimal theory. *Comp Methods Appl Mech Eng.* 1992;99:61–112.
7. Lee EH. Elastic-plastic deformation at finite strains. *J Appl Mech.* 1969;36:1–6.
8. Simo JC, Taylor RL. Quasi-compressible finite elasticity in principal stretches. Continuum basis and numerical algorithms. *Comput Methods Appl Mech Eng.* 1991;85:273–310.
9. Prager W, Hodge PG. Theory of perfectly plastic solids. New York: Wiley; 1951.

Chapter 5

Finite Element Analysis for Contact Problems

5.1 Introduction

When two or more bodies collide, contact occurs between two surfaces of the bodies so that they cannot overlap in space. Metal forming, vehicle crash, projectile penetration, various seal designs, and bushing and gear systems are only a few examples of contact phenomena. During sheet-metal formation, for example, a simple-shaped blank is formed into a desired shape through contact against a punch and die. In such a case, it is important to determine contact locations between a deformable blank and a rigid or deformable punch and die. In a broader sense, contact is a common and important aspect of mechanical systems, where multiple parts are assembled to compose the system. In fact, contact is the main tool to join multiple parts together, which includes screws, bolts, welds, etc.

The objective of contact analysis is to answer the following questions: (a) whether two or more bodies are in contact, (b) if they are, where the location or region of contact is, (c) how much contact force or pressure occurs in the contact interface, and (d) if there is a relative motion after contact in the interface. In this chapter, these questions will be addressed in the continuum and finite element domains.

Contact is categorized as boundary nonlinearity, in contrast to both geometric nonlinearity, which emerges from finite deformation problems, and material nonlinearity, which is a product of nonlinear constitutive relations. The nonlinearity of contact can be explained in two aspects. Firstly, if two separate bodies come into contact, the graph of the contact force vs. displacement looks like a cliff because the contact force stays at zero when two bodies are separate and increases vertically after the bodies come into contact. In such a case, a functional relationship is not available because there is no one-to-one relationship between contact force and displacement. A similar phenomenon happens in the tangential direction under friction where two bodies are stuck together until the tangential force reaches a threshold, after which continuous sliding occurs without further increasing the

tangential force. Such an abrupt change in contact force and slip makes the problem highly nonlinear. Secondly, in order to be a well-posed problem in mechanics, either displacement (kinematics) or force (kinetics), but not both, must be given for every material point. Then, the finite element equation solves for unknown information with given information. On the displacement boundary, for example, if displacement is given, reaction force should be calculated. On the other hand, on the traction boundary, if the applied force is given, the corresponding displacement is to be calculated. Note that these two boundaries are clearly identified in the problem definition stage. In the case of contact, however, both displacement and contact force are unknown, except for very limited cases; that is, the contact boundary is a part of the solution. The user can only identify a candidate of contact boundary before solving the problem. Therefore, the finite element analysis procedure must find (a) whether a material point in the boundary of a body is in contact with the other body, and if it is in contact, (b) the corresponding contact force must be calculated. Since the contact force at a material point can affect the deformation of neighboring points, this process needs to be repeated until finding right states for all points that are possible in contact. Because of this procedural nature, contact nonlinearity is often addressed algorithmically (Fig. 5.1).

For the case of an elastic system, equilibrium can be described as finding a displacement field that minimizes the potential energy. Contact can then be considered as a constraint of the optimization formulation, such that the potential energy is minimized while satisfying the contact constraint; that is, a body cannot penetrate the other body.¹ The constrained optimization problem can be converted into an unconstrained one by using the penalty regularization or Lagrange multiplier methods. Therefore, most contact algorithms are derived based on these two methods. Once understanding that contact can be considered as a constraint to the structural equilibrium, it can be applicable to nonelastic materials, such as elastoplastic material, as it is basically independent of material models used. Therefore, it is possible to treat the contact formulation independent of constitutive models.

Although contact problems can be formulated in a variety of ways, the slave–master concept is commonly used in finite element-based applications. In the slave–master concept, one body is called a slave body, and the other is called a master body. Although the selection of slave and master bodies is arbitrary, some guidelines will be given later in the chapter. The contact constraint is then imposed in such a way that the slave surface cannot penetrate the master surface. Or, in finite elements, the nodes on the slave boundary cannot penetrate the surface elements on the master boundary. It is also possible that the role of slave and master can be changed so that the master surface cannot penetrate the slave surface.

¹ Rigorous discussions on this topic with variational inequality and its equivalence to the constrained optimization can be found in J. Sokolowski and J. P. Zolesio, *Introduction to Shape Optimization*, Springer-Verlag, Berlin, 1991. A brief summary will be presented in Sect. 5.3.

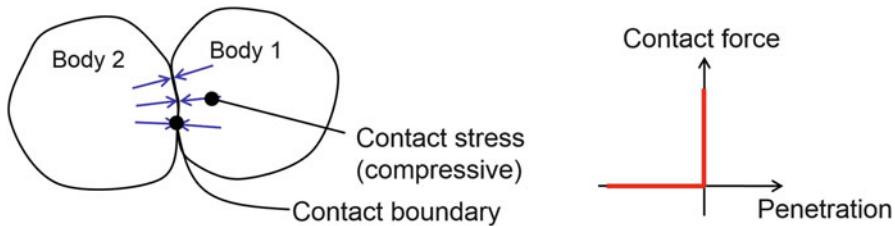


Fig. 5.1 Contact boundary and contact force

This chapter is organized as follows. In Sect. 5.2, simple one-point contact examples are presented in order to show the characteristics of contact phenomena and possible solution strategies. In Sect. 5.3, a general formulation of contact is presented based on the variational formulation similar to previous chapters. To facilitate comprehension, the complexity of formulation is gradually increased by moving from flexible-to-rigid contact to flexible-to-flexible contact, from line-to-line to surface-to-surface contact, and including friction. Section 5.4 focuses on finite element discretization and numerical integration of the contact variational form. Three-dimensional contact formulation is presented in Sect. 5.5. From the finite element point of view, all formulations involve the use of some form of constraint equation. Because of the highly nonlinear and discontinuous nature of contact problems, great care and trial and error are necessary to obtain solutions to practical problems. Section 5.6 presents modeling issues related to contact analysis, such as selecting slave and master bodies, removing rigid-body motions, etc.

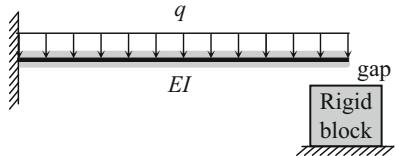
5.2 Examples of Simple One-Point Contact

In order to illustrate key features of a contact problem, simple one-point contact examples are presented in this section. The concepts in this section will be generalized to curve or surface contact problems in Sect. 5.3.

5.2.1 Contact of a Cantilever Beam with a Rigid Block

Consider a cantilever beam subjected to a distributed load. The deflection of the free end of the cantilever is limited by a rigid block. There is a small gap between the end of the beam and the rigid block as shown in Fig. 5.2. The following numerical data are assumed: distributed load $q = 1 \text{ kN/m}$, length of the beam $L = 1 \text{ m}$, flexural rigidity $EI = 10^5 \text{ N m}^2$, and initial gap $\delta = 1 \text{ mm}$.

Fig. 5.2 Cantilever beam supported by a rigid block with a gap



5.2.1.1 Solution Using Trial and Error

In such a simple case, there are two possibilities. If the initial gap is larger than the deflection of the beam, then there will be no contact. On the other hand, if the deflection is larger than the gap, then the gap is closed under the given load. One solution strategy would be first to assume that the gap is large enough so that it will not close under the given load. In that case, the rigid block has no influence on the deformations. From the Euler beam model, the deflection curve and the tip deflection can be given as

$$v_N(x) = \frac{qx^2}{24EI} (x^2 + 6L^2 - 4Lx), \quad v_N(L) = \frac{qL^4}{8EI} = 0.00125 \text{ m.}$$

The solution shows that the tip deflection is larger than the gap, and therefore, the assumption of gap not closing is wrong.

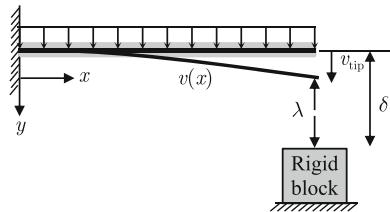
Now, when the gap is closed, contact occurs between the beam and the rigid block. Even if the rigid block has a finite width, it is assumed that the contact only occurs at the tip of the beam, i.e., one-point contact. Since the rigid block prevents the deflection of the beam, its effect can be modeled by applying a force, i.e., a contact force, such that the beam cannot penetrate the rigid block. Since the beam deflection is small, the rule of superposition is used for the effect of the two loads. The deflection curve and the tip deflection of the beam under the force at the tip can be given as

$$v_c(x) = \frac{-\lambda x^2}{6EI} (3L - x), \quad v_c(L) = \frac{-\lambda}{3 \times 10^5}.$$

Here, a negative sign is used for the force because the direction of contact force is opposite to the applied distributed load. At this point, the contact force, λ , is unknown, which can be calculated from the condition that the beam cannot penetrate the rigid block; that is, the deflection of the combined loads is the same with the gap, as

$$v_{\text{tip}} = v_N(L) + v_c(L) = 0.00125 - \frac{\lambda}{3 \times 10^5} = 0.001 = \delta.$$

Fig. 5.3 Deflection of cantilever beam with gap and contact force



From the above relation, the contact force can be calculated to be $\lambda = 75$ N. The deflection curve can then be obtained by combining the two deflection curves, as $v(x) = v_N(x) + v_c(x)$.

5.2.1.2 Solution Using Contact Constraint

The issue in the previous trial-and-error approach is that the solution (deflection of the beam) has to be calculated first in order to determine the status of contact. When contact occurs at multiple points, the procedure can be quite complicated to check all possible combinations of the contact points. A more systematic contact formulation can be developed by considering both the contact force and the gap between the beam and the rigid block as unknowns and adding an additional constraint. The unknown contact force is denoted by λ that acts on the beam and the rigid block in the opposite directions.²

In order to assign consistent directions, one of the two contact points is considered a master and the other a slave. The master is assumed to be fixed while the slave moves to initiate the contact. For a general situation, when both bodies in potential contact are loaded, the choice between a master and a slave may be arbitrary. More details will be discussed in Sect. 5.5 for selecting the master and slave. In this problem, the beam obviously is the slave and the rigid block is the master.

With the downward deflection being positive, Fig. 5.3 shows the positive directions for this contact force on the beam and the rigid block. The contact force is treated as an externally applied load, even if it is unknown. Because of Newton's third law, the contact force acts in an equal and opposite direction to the beam and the block. In this particular example, since the rigid block is fixed, it is unnecessary to consider the equilibrium of the rigid block.

Treating the contact forces as externally applied loads and using the superposition rule of two independent loads, the beam deflection curve can be obtained by

² It will be clear later that the contact force is equivalent to the Lagrange multiplier in the constrained optimization, which is the reason to use the Greek symbol λ .

$$v(x) = \frac{qx^2}{24EI} (x^2 + 6L^2 - 4Lx) - \frac{\lambda x^2}{6EI} (3L - x). \quad (5.1)$$

This deflection curve must be supplemented by a contact constraint, which is defined using the following gap function:

$$g = v_{\text{tip}} - \delta \leq 0. \quad (5.2)$$

The physical requirements of contact are that there should be no penetration, the contact force should be positive, and when the gap is greater than zero, the contact force should be zero and vice versa. These requirements dictate that the solution satisfies the following three conditions:

$$\begin{aligned} \text{No penetration : } & g \leq 0, \\ \text{Positive contact force : } & \lambda \geq 0, \\ \text{Consistency condition : } & \lambda g = 0. \end{aligned} \quad (5.3)$$

The above requirements are exactly the same as those of the Lagrange multiplier in a constrained optimization problem. The consistency condition in the above equation can be used to find the correct contact status as well as the contact force. Since the gap is also a function of contact force, using Eqs. (5.1) and (5.2), the above consistency condition can be written as

$$\lambda g = \lambda \left(0.00025 - \frac{\lambda}{3 \times 10^5} \right) = 0.$$

The above quadratic equation has two solutions: $\lambda = 0$ N or $\lambda = 75$ N. When $\lambda = 0$ N, the gap becomes $g = 0.00025 > 0$, which violates the condition of no penetration. Therefore, this cannot be a possible configuration. On the other hand, when $\lambda = 75$ N, the gap becomes $g = 0$. Since this solution satisfies all requirements, this is the solution. In fact, the solution is consistent with the solution from the direct method.

In the above example, the additional unknown (contact force) is added as a Lagrange multiplier, and the consistency condition is used to determine contact status and contact force. In the penalty method, it is also possible to impose the contact constraint without introducing additional unknowns. In the penalty method, a small amount of penetration is allowed, and the contact force is applied proportional to the amount of penetration. Since the gap in Eq. (5.2) can be both positive and negative, the following penetration function is defined:

$$\phi_N = \frac{1}{2}(|g| + g), \quad (5.4)$$

which is zero when $g \leq 0$ and has the same value with g when $g > 0$. Then, the contact force is defined using the penetration function

Table 5.1 Penetrations and contact forces for different penalty parameters

Penalty parameter	Penetration (m)	Contact force (N)
3×10^5	1.25×10^{-4}	37.50
3×10^6	2.27×10^{-5}	68.18
3×10^7	2.48×10^{-6}	74.26
3×10^8	2.50×10^{-7}	74.92
3×10^9	2.50×10^{-8}	75.00

$$\lambda = K_N \phi_N, \quad (5.5)$$

where K_N is a penalty parameter. The contact force will be zero when the gap is open and proportionally increase with the penetration. The basic concept is that this method allows a small amount of penetration and then penalizes it by applying a large force. A benefit is that the contact force is now related to the gap, albeit the relationship is nonlinear.

The definition of the gap in Eq. (5.2) can be used to calculate the contact status and contact force, as

$$g = 0.00025 - \frac{K_N}{3 \times 10^5} \frac{1}{2}(|g| + g).$$

When $g \leq 0$ is assumed, the above equation is self-conflicting, which means that penetration occurs. When $g > 0$, the above equation can be solved for the gap with a given penalty parameter K_N . Table 5.1 shows the amount of penetration and contact forces for different values of the penalty parameter. It can be observed that as the penalty parameter increases, the penetration decreases and the contact force converges to the accurate value.

Example 5.1. Lagrange multiplier when no contact When the distributed load is 500 N/m, calculate the tip deflection of the beam and determine if contact occurs or not using the Lagrange multiplier method.

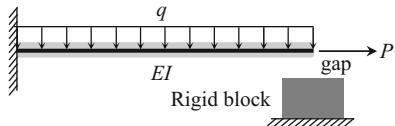
Solution From Eq. (5.1), the tip deflection can be written in terms of the Lagrange multiplier as

$$v_{\text{tip}} = 0.625 \times 10^{-3} - \frac{\lambda}{3 \times 10^5}.$$

Using the gap function in Eq. (5.2), the contact consistency condition in Eq. (5.3) can be written as

$$\lambda g = \lambda \left(-0.375 \times 10^{-3} - \frac{\lambda}{3 \times 10^5} \right) = 0.$$

Fig. 5.4 Cantilever beam supported with a potential frictional contact at the tip with a rigid block



The above equation has two solutions: $\lambda = -112.5 \text{ N}$ and $\lambda = 0$. The former violates the inequality condition of the Lagrange multiplier; therefore, it is an invalid solution.³ The latter yields the gap function of $g = -0.375 \times 10^{-3} < 0$, which satisfies the inequality condition; therefore, it is a valid solution. In fact, it physically means that the gap is not closed and contact does not occur. This can be confirmed by the tip deflection of $v_{\text{tip}} = 0.625 \text{ mm}$, which is smaller than the initial gap. ■

5.2.2 Contact of a Cantilever Beam with Friction

Consider a slightly more complicated problem that involves both a normal contact and friction. A cantilever beam is subjected to a distributed load and an axial load. The free end of the cantilever could potentially contact the block, as shown in Fig. 5.4. Again, it is assumed that the contact can occur only at the tip of the beam. The block surface has a known coefficient of friction μ . The load sequence is such that the transversely distributed load q is applied first, followed by the axial load P . The same numerical data as in Sect. 5.2.1 are used for the beam deflection. For the axial direction, the following data are used: axial load $P = 100 \text{ N}$, axial rigidity $EA = 10^5 \text{ N}$, and friction coefficient $\mu = 0.5$.

5.2.2.1 Solution with No Frictional Resistance

From the assumption of infinitesimal deformation, the transverse behavior of the beam can be decoupled with the axial behavior. Therefore, the beam deflection will be identical to the previous section, and the beam will be in contact with the rigid block with a contact force of 75 N. In the axial direction, the displacement can be modeled using an axially loaded bar. Therefore, the tip displacement due to the axial load becomes

$$u_{\text{tip}}^{\text{no-friction}} = \frac{PL}{EA} = 1.0 \text{ mm}. \quad (5.6)$$

This tip displacement will be compared with the case when friction exists at the contact point.

³ It is interesting to note that the physical interpretation of the negative Lagrange multiplier is the force that is required to apply at the tip of the beam in order to close the gap.

Fig. 5.5 Tangential slip of cantilever beam with friction force



5.2.2.2 Frictional Constraint Function

When friction exists on the interface, the contact point may or may not slip, depending on interface conditions, such as friction coefficient, contact force, and tangential force. In Coulomb's friction model, sliding along the contact surface will take place when the tangential component of the contact force is greater than the frictional resistance. Similar to the normal contact case, the tangential friction force also occurs both in the beam and the rigid block in equal and opposite directions. The positive directions of contact and friction forces are shown in Fig. 5.5. Denoting the normal force at the contact surface as λ and tangential force t , the physical requirements for a frictional constraint are as follows.

$$\begin{aligned} \text{Stick condition : } & t - \mu\lambda < 0, \quad u_{\text{tip}} = 0, \\ \text{Slip condition : } & t - \mu\lambda = 0, \quad u_{\text{tip}} > 0, \\ \text{Consistency condition : } & u_{\text{tip}}(t - \mu\lambda) = 0. \end{aligned} \quad (5.7)$$

When the stick condition occurs, the contact point will not move in the tangential direction, and the tangential force will be determined based on the equilibrium with the externally applied loads, whose magnitude should be less than $\mu\lambda$. When the slip condition occurs, the tangential force will be the same as $\mu\lambda$ and the contact point will continuously move tangentially until the system finds an equilibrium.

The above frictional constraint is similar to the one in Eq. (5.3). Therefore, either the penalty method or Lagrange multiplier method can be applied. The only difference is that now the slip displacement u_{tip} is considered as a Lagrange multiplier, while the friction force is considered as a constraint.

5.2.2.3 Solution Using Trial and Error

In the trial-and-error approach, one condition is assumed first, and then after solving the problem, the other requirements are checked. If all requirements are satisfied,

then the initial assumption is correct and the state is determined. Otherwise, other conditions are assumed until all possible conditions are exhausted.

If the stick condition is assumed first, it means that

$$u_{\text{tip}} = \frac{PL}{EA} - \frac{tL}{EA} = 0 \Rightarrow t = P = 100\text{N}.$$

However, this tangential friction force violates the requirement $t - \mu\lambda < 0$. Therefore, the stick condition is not valid.

In order to check with the slip condition, the friction force is first calculated from $t = \mu\lambda = 37.5\text{ N}$. The friction force will generate the following displacement:

$$u_{\text{tip}} = \frac{PL}{EA} - \frac{tL}{EA} = 0.625\text{ mm}, \quad (5.8)$$

which satisfies the requirement. Therefore, the slip condition is valid. Note that the slip is less than that of the frictionless assumption in Eq. (5.6).

5.2.2.4 Solution Using Frictional Constraint

In the Lagrange multiplier method, the consistency condition in Eq. (5.7) is used to impose the constraint condition. Compared to the case of normal contact, the choice of the Lagrange multiplier is not obvious in this case. Between the tip displacement and frictional force, the tip displacement is chosen as a Lagrange multiplier and the frictional forcing term, $t - \mu\lambda$, is chosen as a constraint. For the case when the Lagrange multiplier and constraint are switched, the readers are referred to Exercise Problem P5.2. Using the tip displacement formula in Eq. (5.8), the tangential force can be written in terms of the tip displacement as

$$t = P - \frac{EA}{L}u_{\text{tip}}.$$

Therefore, the consistency condition can be written as

$$u_{\text{tip}} \left(P - \frac{EA}{L}u_{\text{tip}} - \mu\lambda \right) = 0. \quad (5.9)$$

The above consistency condition has two solutions: $u_{\text{tip}} = 0$ and

$$u_{\text{tip}} = \frac{(P - \mu\lambda)L}{EA}.$$

The first solution, $u_{\text{tip}} = 0$, corresponds to the stick condition and yields $t = P$. However, since $t - \mu\lambda = 62.5\text{ N} > 0$, it violates the stick condition. The second solution,

$$u_{\text{tip}} = \frac{(P - \mu\lambda)L}{EA} = 0.625 \text{ mm} > 0$$

corresponds to the slip condition and yields $t = \mu\lambda$, which satisfies the slip condition. Therefore, this is the valid state.

In the penalty method, the constraints on the frictional force are penalized when it violates the condition, that is, when $t - \mu\lambda > 0$. In the same way with the normal contact case, the following penalty function is defined for the frictional force:

$$\phi_T = \frac{1}{2}(|t - \mu\lambda| + t - \mu\lambda). \quad (5.10)$$

Note that $\phi_T = 0$ when $t - \mu\lambda \leq 0$ and $\phi_T = t - \mu\lambda > 0$ when the constraint is violated. In the penalty method, the relationship between the slip displacement and the frictional force can be established by

$$u_{\text{tip}} = K_T \phi_T, \quad (5.11)$$

where K_T is the penalty parameter for the tangential slip. When $t - \mu\lambda \leq 0$, the above equation represents a stick condition exactly; i.e., $u_{\text{tip}} = 0$. Therefore, no approximation is involved in the case of a stick condition. On the other hand, the above equation shows a slip condition when $t - \mu\lambda > 0$, i.e., when the constraint is violated. However, the slip condition is penalized with a large value of penalty parameter K_T so that the violation remains small.

In order to find the tip displacement using the penalty method, the frictional force is expressed in terms of the tip displacement, as

$$t = P - \frac{EA}{L} u_{\text{tip}}. \quad (5.12)$$

By substituting the above equation into Eq. (5.11), the following tip displacement can be obtained:

$$u_{\text{tip}} = \frac{K_T L (P - \mu\lambda)}{L + K_T EA}.$$

For a large value of K_T , the above equation can be approximated by

$$u_{\text{tip}} \approx \frac{(P - \mu\lambda)L}{EA},$$

and the frictional force becomes

Table 5.2 Tip displacement and frictional forces for different penalty parameters

Penalty parameter	Tip displacement (m)	Frictional force (N)
1×10^{-4}	5.68×10^{-4}	43.18
1×10^{-3}	6.19×10^{-4}	38.12
1×10^{-2}	6.24×10^{-4}	37.56
1×10^{-1}	6.25×10^{-4}	37.50
1×10^0	6.25×10^{-4}	37.50

$$t = P - \frac{EA}{L} u_{\text{tip}} \approx \mu \lambda,$$

which is nothing but the slip condition. Table 5.2 shows the tip displacements and frictional forces for different values of penalty parameter. It can be observed that as the penalty parameter increases, the tip displacement and the frictional force converges to the accurate value of 0.625 mm and 37.5 N, respectively. It is noted that the penalty parameter is relatively small compared to the case of normal contact because the penalty parameter relates the frictional force to the slip displacement.

Example 5.2. Lagrange multiplier for friction When the force at the tip is $P = 25$ N, calculate the tip displacement of the beam and determine if a stick or slip occurs using the Lagrange multiplier method.

Solution From Eq. (5.9), the consistency condition has two solutions: $u_{\text{tip}} = 0$ and $u_{\text{tip}} = (P - \mu \lambda)L/EA$. The first solution, $u_{\text{tip}} = 0$, corresponds to the stick condition and yields $t = P = 25$ N. Since $t - \mu \lambda = -7.5$ N < 0, it satisfies the stick condition. Therefore, the beam is in the stick condition. On the other hand, if the slip condition is checked, the tip displacement

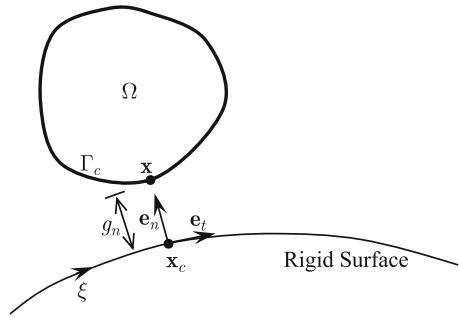
$$u_{\text{tip}} = \frac{(P - \mu \lambda)L}{EA} = -0.075 \text{ mm} < 0$$

becomes negative, which violates the stick condition. Therefore, the stick condition is not a valid state. ■

5.3 General Formulation for Contact Problems

The one-point contact examples in the previous section are limited to a practical point of view, as most contact in engineering applications occurs along a line (one or two dimensional) or a surface (three dimensional). In this section, the basic concepts of one-point contact are extended to two or three dimensions. In order to simplify the presentation, only the penalty method will be discussed.

Fig. 5.6 Contact condition in two dimensions



5.3.1 Contact Condition with Rigid Surface

The general formulation is illustrated with reference to contact between two bodies, as shown in Fig. 5.6. The concepts can easily be generalized to contact involving more than two bodies. Note that each body is assumed to be properly supported such that no rigid-body motion is possible even without the contact. In the case of contact with a rigid body, it is natural that the flexible body is selected as a slave body and the rigid body as a master body.

Contact conditions can be divided into normal impenetrability and tangential slip. The impenetrability condition prevents the slave body from penetrating into the master body, while the tangential slip represents the frictional behavior on the contact surface. Figure 5.6 illustrates a general contact condition with a rigid surface in two dimensions. A part of the slave boundary is denoted by contact boundary, Γ_c . Although the actual contact region is unknown and is a part of the solution, the user specifies the contact boundary such that all possible contacts can only occur within this boundary. It is assumed that a point x on the contact boundary will be in contact with a point x_c on the master surface if the contact actually occurs. In the following, the contact condition will only be discussed with respect to a single slave point x . Since the motion of the rigid surface is prescribed throughout the analysis, a natural coordinate ξ is used to represent the location on a rigid surface. Thus, the coordinates of contact point x_c on the master surface can also be represented using a natural coordinate at the contact point ξ_c by

$$\mathbf{x}_c = \mathbf{x}_c(\xi_c). \quad (5.13)$$

However, the contact point x_c , or equivalently ξ_c , is yet unknown. In the three-dimensional space case, two natural coordinates are required to describe the master surface.

In general, contact analysis is to find the contact point and the contact force at the contact point, including contact pressure and frictional force. In finite element analysis, either displacement or force is known at the boundary and the other

unknown variable is solved through the equilibrium requirements. In contact analysis, however, both the contact point \mathbf{x}_c and the contact force at that point are unknown, which makes the contact problem challenging. Usually, a trial-and-error approach is taken in which the contact point is searched from the current geometry, and the contact constraint is imposed once the point is in contact.

The first step of contact analysis is to find the contact point $\mathbf{x}_c(\xi_c)$ on the master surface corresponding to a slave point \mathbf{x} . It is necessary to identify this point in order to determine if the two points are in contact or not. Mathematically, this is called the orthogonal projection, or the closest point from the slave point \mathbf{x} . When the master boundary is a straight line, the closest point can explicitly be found. For a general nonlinear curve, however, the following nonlinear equation is solved to find the contact point:

$$\varphi(\xi_c) = (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_t(\xi_c) = 0, \quad (5.14)$$

where $\mathbf{e}_t = \mathbf{t}/\|\mathbf{t}\|$ is the unit tangential vector and $\mathbf{t} = \mathbf{x}_{c,\xi} = \partial \mathbf{x}_c / \partial \xi$. Equation (5.14) is called the contact consistency condition, and $\mathbf{x}_c(\xi_c)$ is the closest projection point of $\mathbf{x} \in \Gamma_c$ onto the rigid surface that satisfies Eq. (5.14).

Once the contact point is found, it is necessary to determine if the contact actually occurs, which can be done by measuring the distance between the two points. At the same time, the impenetrability condition can be imposed by using the same distance, as shown in Fig. 5.5. The impenetrability condition can be defined by using the normal gap function g_n , which measures the normal distance, as

$$g_n \equiv (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_n(\xi_c) \geq 0, \quad \mathbf{x} \in \Gamma_c, \quad (5.15)$$

where $\mathbf{e}_n(\xi_c)$ is the unit outward normal vector of the master surface at the contact point.

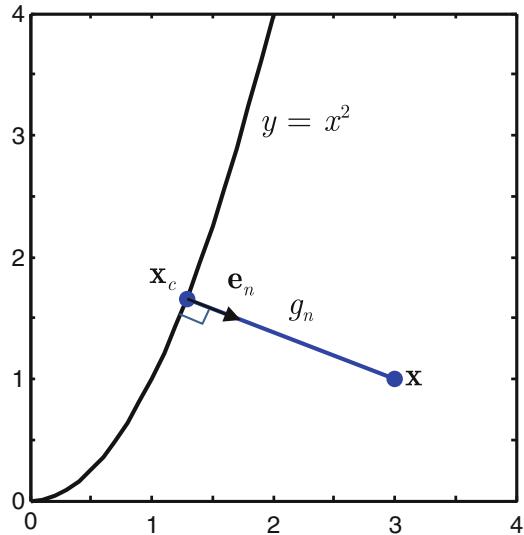
As the contact point moves along the master boundary, a frictional force in a tangential direction to the master boundary resists the tangential relative movement. The tangential slip function g_t is the measure of the relative movement of the contact point along the rigid surface and is defined as

$$g_t \equiv \|\mathbf{t}^0\|(\xi_c - \xi_c^0), \quad (5.16)$$

where both the tangential vector \mathbf{t}^0 and the natural coordinate ξ_c^0 are the values at the previously converged time increment or load increment. The superscript “0” will denote the previous configuration time in the following derivations.

Example 5.3. Projection to a parabola When a rigid boundary is given as $y=x^2$, find a projection point from $\mathbf{x}=\{3, 1\}^T$ using Eq. (5.14) and distance using Eq. (5.15) when $x > 0$.

Fig. 5.7 Projection to a parabola



Solution In order to use the contact consistency condition to find a project point, it is necessary to represent the rigid boundary using a parameter. In the case of the given parabola, the parametric relation can be written as $\mathbf{x}_c = \{\xi, \xi^2\}^T$. The unit tangential vector can be calculated by

$$\mathbf{e}_t = \frac{\mathbf{x}_{c,\xi}}{\|\mathbf{x}_{c,\xi}\|} = \frac{1}{\sqrt{1 + 4\xi^2}} \begin{Bmatrix} 1 \\ 2\xi \end{Bmatrix}.$$

By defining the unit normal vector in the z -coordinate as $\mathbf{k} = \{0 \ 0 \ 1\}^T$, the unit normal vector to the rigid boundary can be defined as

$$\mathbf{e}_n = \mathbf{e}_t \times \mathbf{k} = \frac{1}{\sqrt{1 + 4\xi^2}} \begin{Bmatrix} 2\xi \\ -1 \end{Bmatrix}.$$

Then, the closest project point from $\mathbf{x} = \{3, 1\}^T$ can be found by

$$\varphi(\xi) = (\mathbf{x} - \mathbf{x}_c(\xi))^T \mathbf{e}_t(\xi) = \frac{3 + \xi - 2\xi^3}{\sqrt{1 + 4\xi^2}} = 0.$$

Since the solution is in the first quadrant, $\xi > 0$, the numerator of the above equation can be solved for $\xi_c = 1.29$. Therefore, the contact point on the rigid boundary becomes $\mathbf{x}_c = \{1.29, 1.66\}^T$, as shown in Fig. 5.7.

The distance between the slave point x and contact point \mathbf{x}_c can be obtained using the gap function, as

$$g_n = (\mathbf{x} - \mathbf{x}_c)^T \mathbf{e}_n = \frac{-\xi_c^2 + 6\xi_c - 1}{\sqrt{1 + 4\xi_c^2}} = 1.83.$$

The above result can be verified by using the distance formula between two points as

$$\sqrt{(3 - 1.29)^2 + (1 - 1.66)^2} = 1.83.$$

■

5.3.2 Variational Inequality in Contact Problems

Before deriving a contact variational formula, it is beneficial to discuss the fundamental properties of the contact problem. Although only a linear elastic problem will be considered for simplicity, due to the inequality constraint on the deformation field, the contact problem is nonlinear even in a linear elastic case. The differential equation of the contact problem can be written as follows:

Governing equilibrium equation:

$$\begin{aligned} \sigma_{ij,j} + f_i^B &= 0, \quad \mathbf{x} \in \Omega, \\ u_i(\mathbf{x}) &= 0, \quad \mathbf{x} \in \Gamma^g, \\ \sigma_{ij} n_j &= f_i^S, \quad \mathbf{x} \in \Gamma^S. \end{aligned} \tag{5.17}$$

Contact conditions:

$$\begin{aligned} \mathbf{u}^T \mathbf{e}_n + g_n &\geq 0, \\ \sigma_n &\geq 0, \quad \mathbf{x} \in \Gamma_c, \\ \sigma_n (\mathbf{u}^T \mathbf{e}_n + g_n) &= 0. \end{aligned} \tag{5.18}$$

The first inequality in Eq. (5.18) can be obtained from the incremental form of the impenetrability condition in Eq. (5.15), since a small deformation linear problem is assumed. Note that the expression of contact conditions in Eq. (5.18) is similar to that of Eq. (5.3). Therefore, either the Lagrange multiplier method or the penalty method can be used to impose the contact condition. The inequality contact constraint in Eq. (5.18) can be considered by constructing a closed convex set \mathbb{K} , defined as

$$\mathbb{K} = \left\{ \mathbf{w} \in [H^1(\Omega)]^N \mid \mathbf{w}|_{\Gamma^g} = 0 \quad \text{and} \quad \mathbf{w}^T \mathbf{e}_n + g_n \geq 0 \text{ on } \Gamma_c \right\}. \tag{5.19}$$

The convex set \mathbb{K} satisfies all kinematic constraints (displacement conditions).

If \mathbf{u} is the solution to Eqs. (5.17) and (5.18), then $\mathbf{u} \in \mathbb{K}$. The variational inequality can be derived from the weak formulation of the differential Eq. (5.17). In previous chapters, the weak form is obtained by multiplying the governing differential equation with a virtual displacement $\bar{\mathbf{u}}$, which belongs to the space of kinematically admissible displacements. In order to make the elements in the convex set \mathbb{K} kinematically admissible, the virtual displacement $\bar{\mathbf{u}}$ is substituted by $\mathbf{w} - \mathbf{u}$ for all $\mathbf{w} \in \mathbb{K}$. Therefore, after multiplying $\mathbf{w} - \mathbf{u}$ and integrating by parts, the weak form becomes

$$\begin{aligned} & \int_{\Omega} \sigma_{ij}(\mathbf{u}) \epsilon_{ij}(\mathbf{w} - \mathbf{u}) \, d\Omega \\ &= - \int_{\Omega} \sigma_{ij,j}(w_i - u_i) \, d\Omega + \int_{\Gamma^S \cup \Gamma_c} \sigma_{ij} n_j (w_i - u_i) \, d\Gamma \\ &= \ell(\mathbf{w} - \mathbf{u}) + \int_{\Gamma_c} \sigma_{ij} n_j (w_i - u_i) \, d\Gamma, \end{aligned} \quad (5.20)$$

where the last term in Eq. (5.20), which is not known until the solution is obtained, is always nonnegative, as shown below:

$$\begin{aligned} & \int_{\Gamma_c} \sigma_{ij} n_j (w_i - u_i) \, d\Gamma \\ &= \int_{\Gamma_c} \sigma_n (w_n - u_n) \, d\Gamma \\ &= \int_{\Gamma_c} \sigma_n (w_n + g_n - u_n - g_n) \, d\Gamma \\ &= \int_{\Gamma_c} \sigma_n (w_n + g_n) \, d\Gamma \geq 0, \quad \forall \mathbf{w} \in \mathbb{K}. \end{aligned} \quad (5.21)$$

Thus, variational equation (5.20) becomes a variational inequality as

$$a(\mathbf{u}, \mathbf{w} - \mathbf{u}) \geq \ell(\mathbf{w} - \mathbf{u}), \quad \forall \mathbf{w} \in \mathbb{K}, \quad (5.22)$$

where $\mathbf{u} \in \mathbb{K}$ is the solution.

Figure 5.8 shows the relationship between the solution without contact, $\mathbf{u}' \in \mathbb{Z}$, and the solution with contact, \mathbf{u} . If the solution \mathbf{u}' belongs to the convex set \mathbb{K} , it satisfies the contact condition and is the solution. However, if \mathbf{u}' is out of the convex set, that is, \mathbf{u}' violates the contact condition, then it has to move to \mathbf{u} through the orthogonal project, which belongs to the convex set. This conceptual explanation can be illustrated using the beam deflection problem in Sect. 5.2.1. As shown in Fig. 5.9, v' is the deflection curve when there is no rigid block or contact constraint. Since the distributed load is large enough so that the tip deflection is larger than the initial gap, the contact constraint is violated. That is, v' belongs to the space of kinematically admissible displacements, but not in the convex set. Therefore, v' is projected to v on the boundary of the convex set by applying the contact force. The

Fig. 5.8 Projection of a solution on to a convex set

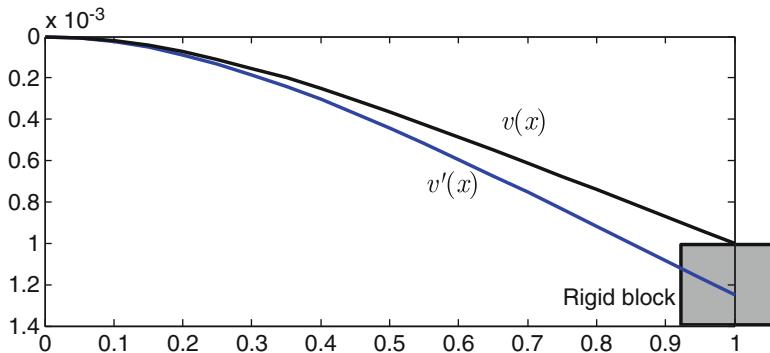
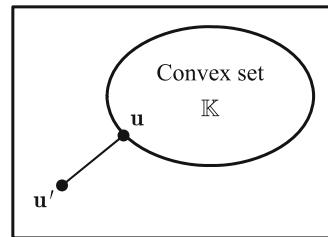


Fig. 5.9 Beam deflection with and without rigid block

physical meaning of the contact force is the force required to project v' onto the convex set, i.e., to satisfy the contact constraint.

The existence and uniqueness of the solution to the variational inequality has been extensively studied for linear elastic material by Duvaut and Lions [1] and Kikuchi and Oden [2]. The existence of a solution to Eq. (5.22) for the nonlinear elastic problem has been proved by Ciarlet [3] for a polyconvex strain energy function.

The same variational inequality in Eq. (5.22) can be used for the nonlinear elastic contact problem with the appropriate structural energy form, as seen in previous chapters. The constraint set of a large deformation problem contains the impenetrability condition in Eq. (5.15) as

$$\mathbb{K} = \left\{ \mathbf{w} \in [H^1(\Omega)]^N \mid \mathbf{w}|_{\Gamma^g} = 0 \quad \text{and} \quad (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_n \geq 0 \quad \text{on } \Gamma_c \right\}. \quad (5.23)$$

From an engineering point of view, however, it is not convenient to solve the variational inequality directly without mentioning the construction of a test function on constraint set \mathbb{K} . The good news is that it is possible to show that the variational inequality is equivalent to the constrained optimization problem of the total potential energy. If the total potential energy is

$$\Pi(\mathbf{u}) = \frac{1}{2}a(\mathbf{u}, \mathbf{u}) - \ell(\mathbf{u}), \quad (5.24)$$

where $a(\mathbf{u}, \mathbf{u})$ is positive definite, then the directional derivative of $\Pi(\mathbf{u})$ in the direction of \mathbf{v} is defined as

$$\langle D\Pi(\mathbf{u}), \mathbf{v} \rangle = a(\mathbf{u}, \mathbf{v}) - \ell(\mathbf{v}), \quad (5.25)$$

where the bilinear property of $a(\cdot, \cdot)$ and the linear property of $\ell(\cdot)$ are used. Using the directional derivative, the variational inequality $a(\mathbf{u}, \mathbf{w} - \mathbf{u}) \geq \ell(\mathbf{w} - \mathbf{u})$ can then be rewritten as

$$\langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle \geq 0. \quad (5.26)$$

To show that Eq. (5.26) is equivalent to the constrained minimization problem, let us consider the following relation. For an arbitrary $\mathbf{w} \in \mathbb{K}$,

$$\Pi(\mathbf{w}) - \Pi(\mathbf{u}) = \langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle + \frac{1}{2}a(\mathbf{w} - \mathbf{u}, \mathbf{w} - \mathbf{u}). \quad (5.27)$$

Since $a(\cdot, \cdot)$ is positive definite, the last term in Eq. (5.27) is always nonnegative; thus,

$$\Pi(\mathbf{w}) \geq \Pi(\mathbf{u}) + \langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u} \rangle, \quad \forall \mathbf{w} \in \mathbb{K}, \quad (5.28)$$

which means

$$\Pi(\mathbf{u}) = \min_{\mathbf{w} \in \mathbb{K}} \Pi(\mathbf{w}) = \min_{\mathbf{w} \in \mathbb{K}} \left[\frac{1}{2}a(\mathbf{w}, \mathbf{w}) - \ell(\mathbf{w}) \right]. \quad (5.29)$$

If $\Pi(\mathbf{w})$ is convex, and set \mathbb{K} is closed and convex, then both the constrained minimization problem in Eq. (5.29) and the variational inequality have a unique solution \mathbf{u} . The variational inequality in Eq. (5.22) can be solved using the constrained minimization problem in Eq. (5.29). Many optimization theories can be used, including mathematical programming, sequential quadratic programming, and active set strategies. For further information on the numerical treatment of contact constraints, the mathematical programming method [4, 5], active set strategies [6], and the sequential quadratic programming method [7] are available.

5.3.3 Penalty Regularization

In the viewpoint of finite element analysis, the constrained optimization problem in Eq. (5.29) is not trivial to solve, because the fundamental idea of finite element analysis is to build test functions that satisfy zero displacement (kinematic)

boundary conditions, which is nothing but the space of kinematically admissible displacements. Using nodal interpolation of finite elements, this condition can easily be obtained by setting the nodal displacement to zero for those nodes that belong to the displacement boundary. However, it is not trivial to build test functions that satisfy the contact constraint because the contact boundary is unknown until the problem is solved.

Instead of solving the constrained optimization, it is easier to convert the constrained optimization problem into an unconstrained optimization problem by using either the penalty method or the Lagrange multiplier method. The former penalizes the potential energy proportional to the amount of constraint violation such that the minimum of the penalized potential energy approximately satisfies the contact constraint. The latter augments the potential energy by a product of the contact constraint and a Lagrange multiplier, which corresponds to the force to impose the contact constraint, such that the minimum of augmented potential energy can satisfy the contact constraint as well as identify the Lagrange multiplier or the contact force. The advantages and disadvantages of the two methods can be found in traditional optimization textbooks [8]. In this section, only the penalty method will be discussed, but a similar approach can be developed for the Lagrange multiplier method.

In order to penalize when $\mathbf{w} \notin \mathbb{K}$ in Eq. (5.29), if a region Γ_c exists that violates the impenetrability condition in Eq. (5.15), then the potential energy is penalized using a penalty function. That is, the potential energy is penalized when penetration occurs. Similarly, the tangential movement of Eq. (5.16) can also be penalized under the stick condition. The contact penalty function must first be defined for the penetrated region by

$$P = \frac{1}{2}\omega_n \int_{\Gamma_c} g_n^2 d\Gamma + \frac{1}{2}\omega_t \int_{\Gamma_c} g_t^2 d\Gamma, \quad (5.30)$$

where ω_n and ω_t are the penalty parameters for normal contact and tangential slip, respectively. The penalty function defined in Eq. (5.30) leads to an exterior penalty method whereby the solution approaches from the infeasible region. This means that the impenetrability condition will be violated, but the amount of violation decreases as the penalty parameter is increased.

The constrained minimization problem in Eq. (5.29) is converted to an unconstrained minimization problem by adding a penalty function to the total potential energy. Thus,

$$\Pi(\mathbf{u}) = \min_{\mathbf{w} \in \mathbb{K}} \Pi(\mathbf{w}) \approx \min_{\mathbf{w} \in \mathbb{Z}} [\Pi(\mathbf{w}) + P(\mathbf{w})]. \quad (5.31)$$

Note that the solution space is changed to \mathbb{Z} from \mathbb{K} because of the penalty function. Therefore, it is much more convenient to build test functions $\mathbf{w} \in \mathbb{Z}$. The variation of Eq. (5.31) contains two contributions that will be examined in this section: one

from the structural potential and the other from the penalty function. The variation of P yields the contact variational form, which is defined by

$$\begin{aligned} b(\mathbf{u}, \bar{\mathbf{u}}) &\equiv \omega_n \int_{\Gamma_c} g_n \bar{g}_n d\Gamma + \omega_t \int_{\Gamma_c} g_t \bar{g}_t d\Gamma \\ &= b_N(\mathbf{u}, \bar{\mathbf{u}}) + b_T(\mathbf{u}, \bar{\mathbf{u}}), \end{aligned} \quad (5.32)$$

where $b_N(\mathbf{u}, \bar{\mathbf{u}})$ and $b_T(\mathbf{u}, \bar{\mathbf{u}})$ are the normal and tangential contact variational forms, respectively. The variable with an over-bar represents the variation of the variable. $b_T(\mathbf{u}, \bar{\mathbf{u}})$ appears only when there is friction in the contact interface. In Eq. (5.32), $\omega_n g_n$ corresponds to the compressive normal contact force, and $\omega_t g_t$ corresponds to the tangential traction force. The latter increases linearly with the tangential slip g_t until it reaches a normal force multiplied by the friction coefficient. The contact variational form in Eq. (5.32) can be expressed in terms of the displacement variation. To make subsequent derivations easier to follow, it is necessary to define several scalar symbols, as follows:

$$\begin{aligned} \alpha &\equiv \mathbf{e}_n^T \mathbf{x}_{c,\xi\xi}, \quad \beta \equiv \mathbf{e}_t^T \mathbf{x}_{c,\xi\xi}, \quad \gamma \equiv \mathbf{e}_n^T \mathbf{x}_{c,\xi\xi\xi} \\ c &\equiv \|\mathbf{t}\|^2 - g_n \alpha, \quad \nu \equiv \|\mathbf{t}\| \|\mathbf{t}^0\| / c. \end{aligned} \quad (5.33)$$

Note that α , β , and γ are related to the higher-order derivatives of the master boundary. If the rigid boundary is approximated by a piecewise linear function, then $\alpha = \beta = \gamma = 0$ and $\nu = 1$.

Example 5.4. Penalty method for beam contact Using the potential energy and the penalty method, calculate the deflection curve for the cantilever beam example in Sect. 5.2.1 with different values of penalty parameter. Assume the following form of beam deflection curve $v(x) = a_2 x^2 + a_3 x^3 + a_4 x^4$ and calculate unknown coefficients.

Solution The potential energy of a cantilever beam under a distributed load can be written as

$$\Pi = \frac{1}{2} \int_0^L EI(v_{,xx})^2 dx - \int_0^L qv dx. \quad (5.34)$$

Since the given form of deflection curve satisfies the essential boundary conditions at $x = 0$, i.e., $v(0) = v_{,x}(0) = 0$, it already belongs to the space of kinematically admissible displacements.

In order to apply the penalty constraint to the region where the impenetrability constraint is violated, the deflection curve of the beam should be calculated first by minimizing the potential energy in Eq. (5.34). If the impenetrability constraint is violated, then the penalty constraint is applied to the violated region. This process makes the problem nonlinear, and the solution can be found through an iterative

procedure. However, to simplify the presentation, it is assumed that the impenetrability constraint is violated only at the tip.

In this particular problem, the contact boundary becomes a point at the tip of the beam. In order to define the penalty function, the following form of gap function is defined first:

$$g_n = \delta - v_{\text{tip}} = \delta - a_2 - a_3 - a_4.$$

The integral form of the penalty function in Eq. (5.30) is defined at a point, $x=L$, as

$$P = \frac{1}{2}\omega_n g_n^2.$$

Therefore, the penalized potential energy becomes

$$\Pi + P = \frac{1}{2} \int_0^L EI(v_{,xx})^2 dx - \int_0^L qv dx + \frac{1}{2}\omega_n g_n^2.$$

Note that the above penalized potential function is a function of unknown coefficients, a_2 , a_3 , and a_4 . The requirement of its minimum is that the potential energy is stationary with respect to these unknown coefficients.⁴ By substituting the expression of v and $v_{,xx}$ into the penalized potential energy, and differentiating with respect to a_2 , a_3 , and a_4 , the following linear system of equations can be obtained:

$$\begin{bmatrix} 4EI + \omega_n & 6EI + \omega_n & 8EI + \omega_n \\ 6EI + \omega_n & 12EI + \omega_n & 18EI + \omega_n \\ 8EI + \omega_n & 18EI + \omega_n & \frac{144}{5}EI + \omega_n \end{bmatrix} \begin{Bmatrix} a_2 \\ a_3 \\ a_4 \end{Bmatrix} = \begin{Bmatrix} \frac{1}{3}q + \omega_n \delta \\ \frac{1}{4}q + \omega_n \delta \\ \frac{1}{5}q + \omega_n \delta \end{Bmatrix}.$$

For the given material, geometric, and load parameters, the three unknown coefficients can be calculated by solving the above matrix equations. For a positive penalty parameter, the coefficient matrix is positive definite. Therefore, a unique solution is expected. Table 5.3 shows the three unknown coefficients, penetration ($a_2 + a_3 + a_4 - \delta$), and the contact force ($-\omega_n g_n$) for different values of the penalty parameter. Similar to the results in Table 5.1, the tip displacement converges to the accurate value as the penalty parameter increases. ■

⁴This is the Rayleigh-Ritz method. For details, readers are referred to N. H. Kim and B. V. Sankar, Introduction to Finite Element Analysis and Design, Wiley & Sons, NY, 2008.

Table 5.3 Coefficients of deflection curve, penetrations, and contact forces for different penalty parameters

Penalty parameter	a_1	a_2	a_3	Penetration (m)	Contact force (N)
3×10^5	2.31×10^{-3}	-1.60×10^{-3}	4.17×10^{-4}	1.25×10^{-4}	37.50
3×10^6	2.16×10^{-3}	-1.55×10^{-3}	4.17×10^{-4}	2.27×10^{-5}	68.18
3×10^7	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.48×10^{-6}	74.26
3×10^8	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.50×10^{-7}	74.92
3×10^9	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.50×10^{-8}	75.00
True value	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	0.0	75.00

5.3.4 Frictionless Contact Formulation

As an ideal case, the contact formulation when there is no friction in the contact interface is addressed first. Computationally, the frictionless contact problem with elastic material is path independent; that is, the equilibrium state is independent of the load history. From a mechanics point of view, a potential energy (or augmented potential energy with contact penalty function) exists, and all field variables are functions of the current configuration.

The first step is to express the normal contact variational form in terms of displacement variation. By taking the first variation of the normal gap function in Eq. (5.15) and using the variation of the contact consistency condition in Eq. (5.14), the first variation of the normal gap function can be obtained as

$$\bar{g}_n(\mathbf{u}; \bar{\mathbf{u}}) = \bar{\mathbf{u}}^T \mathbf{e}_n, \quad (5.35)$$

where the variation of the natural coordinate at the contact point is canceled by an orthogonal condition. The normal gap function can vary only in a normal direction to the rigid surface, which is physically plausible. By using Eq. (5.35), the normal contact form is expressed in terms of displacement variation as

$$b_N(\mathbf{u}, \bar{\mathbf{u}}) = \omega_n \int_{\Gamma_c} g_n \bar{\mathbf{u}}^T \mathbf{e}_n d\Gamma. \quad (5.36)$$

This contact form originates in the impenetrability condition and the fact that the magnitude of the impenetrability force is proportional to the violation of the impenetrability condition.

Note that $b_N(\mathbf{u}, \bar{\mathbf{u}})$ is linear with respect to $\bar{\mathbf{u}}$ and implicit with respect to \mathbf{u} through g_n and \mathbf{e}_n . Since $b_N(\mathbf{u}, \bar{\mathbf{u}})$ is nonlinear in displacement, the same linearization

procedure is required that was used for the structural energy form in Chaps. 3 and 4. The increment of the normal gap function can be obtained in a similar procedure to Eq. (5.35) as

$$\Delta g_n(\mathbf{u}; \Delta \mathbf{u}) = \mathbf{e}_n^T \Delta \mathbf{u}. \quad (5.37)$$

To obtain the increment of the unit normal vector, it is necessary to compute the increment of natural coordinate ξ_c at the contact point using Eq. (5.14), since the normal vector changes along ξ_c . The increment of Eq. (5.14) solves $\Delta \xi_c$ in terms of $\Delta \mathbf{u}$ as

$$\begin{aligned} & \Delta \left[(\mathbf{x} - \mathbf{x}_c)^T \mathbf{e}_t \right] \\ &= (\Delta \mathbf{u} - \mathbf{t} \Delta \xi_c)^T \mathbf{e}_t + (\mathbf{x} - \mathbf{x}_c)^T \Delta \mathbf{e}_t \\ &= \Delta \mathbf{u}^T \mathbf{e}_t - \|\mathbf{t}\| \Delta \xi_c + (\mathbf{x} - \mathbf{x}_c)^T \mathbf{e}_n \left(\frac{1}{\|\mathbf{t}\|} \mathbf{e}_n^T \mathbf{x}_{c,\xi\xi} \right) \Delta \xi_c = 0. \end{aligned} \quad (5.38)$$

Thus, using the definition in Eq. (5.33), we can calculate the increment of the natural coordinates in terms of increment of displacement, as

$$\Delta \xi_c = \frac{\|\mathbf{t}\|}{c} \mathbf{e}_t^T \Delta \mathbf{u}. \quad (5.39)$$

If \mathbf{e}_3 is the fixed unit vector in the out-of-plane direction, then the increment of the unit normal vector can be obtained from the relation $\mathbf{e}_n = \mathbf{e}_3 \times \mathbf{e}_t$ as

$$\begin{aligned} \Delta \mathbf{e}_n &= \mathbf{e}_3 \times \Delta \mathbf{e}_t \\ &= \mathbf{e}_3 \times \Delta \left[\frac{\mathbf{t}}{\|\mathbf{t}\|} \right] \\ &= \mathbf{e}_3 \times \frac{1}{\|\mathbf{t}\|} [\Delta \mathbf{t} - \mathbf{e}_t (\mathbf{e}_t^T \Delta \mathbf{t})] \\ &= -\frac{1}{\|\mathbf{t}\|} [\mathbf{e}_t (\mathbf{e}_n^T \Delta \mathbf{t})] \\ &= -\frac{\alpha \Delta \xi_c}{\|\mathbf{t}\|} \mathbf{e}_t \\ &= -\frac{\alpha}{c} \mathbf{e}_t (\mathbf{e}_t^T \Delta \mathbf{u}). \end{aligned} \quad (5.40)$$

Thus, from Eqs. (5.37) and (5.40), the linearization of the normal contact form is obtained as

$$b_N^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) = \omega_n \int_{\Gamma_c} \bar{\mathbf{u}}^T \mathbf{e}_n \mathbf{e}_n^T \Delta \mathbf{u} d\Gamma - \omega_n \int_{\Gamma_c} \frac{\alpha g_n}{c} \bar{\mathbf{u}}^T \mathbf{e}_t \mathbf{e}_t^T \Delta \mathbf{u} d\Gamma. \quad (5.41)$$

Note that there is a component in the tangential direction because of the effect of curvature. The first term is the conventional contact tangent term for linear

kinematics. The contribution of the second term is usually small, as the contact violation is reduced. If the contact boundary is linear, the second term disappears as $\alpha = 0$.

In the case of a general nonlinear material with a frictionless contact problem, the principle for virtual work can be written as

$$a(\mathbf{u}, \bar{\mathbf{u}}) + b_N(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (5.42)$$

The above equation is obtained from the first variation of the penalized potential energy function in Eq. (5.31), which is equated to zero to satisfy the Kuhn-Tucker condition. Suppose the current time is t_n and the current iteration counter is $k+1$. Assuming that the external force is independent of displacement, the linearized incremental equation of Eq. (5.42) is obtained as

$$\begin{aligned} & a^*(n\mathbf{u}^k; \Delta\mathbf{u}^{k+1}, \bar{\mathbf{u}}) + b_N^*(n\mathbf{u}^k; \Delta\mathbf{u}^{k+1}, \bar{\mathbf{u}}) \\ &= \ell(\bar{\mathbf{u}}) - a(n\mathbf{u}^k, \bar{\mathbf{u}}) - b_N(n\mathbf{u}^k, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \end{aligned} \quad (5.43)$$

Equation (5.43) is linear in incremental displacement for a given displacement variation. The linearized system of Eq. (5.43) is solved iteratively with respect to incremental displacement until the residual forces on the right side of the equation vanish at each time step.

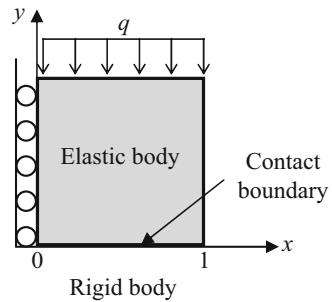
Example 5.5. Frictionless contact of a block A unit square block is under a uniformly distributed load at the top surface and a frictionless contact condition with a rigid body at the bottom surface, as shown in Fig. 5.10. Using the penalty method, calculate the displacement field, penetration, and contact force at the contact interface. Use $EA = 10^5$ N and $q = 1.0$ kN/m and vary the penalty parameter from 10^5 to 10^8 . Assume plane strain with zero Poisson's ratio.

Solution In a two-dimensional problem, the displacement and its variation can be written as $\mathbf{u} = \{u_x, u_y\}^T$ and $\bar{\mathbf{u}} = \{\bar{u}_x, \bar{u}_y\}^T$, respectively. Since the contact surface is flat and parallel to the x -coordinate, the unit normal vector is constant as $\mathbf{e}_n = \{0, 1\}^T$. In this simple problem, the contact boundary can be parameterized by $\xi = x$. Accordingly, the slave contact point and corresponding master point can be written as $\mathbf{x} = \{\xi_c, 0\}^T$ and $\mathbf{x}_c = \{\xi_c, u_y\}^T$. Therefore, the gap function can be defined as

$$g_n = (\mathbf{x} - \mathbf{x}_c)^T \mathbf{e}_n = u_y.$$

Therefore, the contact form in Eq. (5.36) can be written in terms of displacement as

Fig. 5.10 Frictionless contact of an elastic block



$$b_N(\mathbf{u}, \bar{\mathbf{u}}) = \omega_n \int_0^1 u_y \bar{u}_y \Big|_{y=0} dx.$$

The penalized potential energy for a two-dimensional plane strain problem can be written as

$$\Pi + P = \frac{1}{2} \iint_A \mathbf{e}^T \mathbf{D} \mathbf{e} dA - \int_0^1 (-q) u_y \Big|_{y=1} dx + \frac{1}{2} \omega_n \int_0^1 g_n^2 \Big|_{y=1} dx. \quad (5.44)$$

From the assumption of zero Poisson's ratio, the stress-strain matrix \mathbf{D} becomes a diagonal matrix, and all stress-strain relations are decoupled. In addition, since the load is only applied to y -direction, it can be concluded that $\epsilon_{xx} = \gamma_{xy} = 0$. Therefore, the only nonzero displacement component will be u_y . Based on the Rayleigh-Ritz method, the following forms of u_y and its variation are assumed:

$$u_y = a_0 + a_1 y, \quad \bar{u}_y = \bar{a}_0 + \bar{a}_1 y.$$

After substituting these approximations into Eq. (5.44) and taking the variation, we have

$$\bar{\Pi} + \bar{P} = \iint_A E a_1 \bar{a}_1 dA - \int_0^1 (-q)(\bar{a}_0 + \bar{a}_1) dx + \omega_n \int_0^1 a_0 \bar{a}_0 dx = 0.$$

Since \bar{a}_0 and \bar{a}_1 are arbitrary, their coefficients must be zero in order to satisfy the above equation, from which the two coefficients can be determined by

$$a_0 = -\frac{q}{\omega_n}, \quad a_1 = -\frac{q}{EA}.$$

Therefore, the displacement u_y can be determined by

$$u_y = -\frac{q}{\omega_n} - \frac{q}{EA}y, \quad 0 \leq y \leq 1.$$

The first term on the right-hand side is the contact constraint violation due to the penalty method, while the second term represents the constant strain due to the distributed load. The constraint violation will be reduced as the penalty parameter increases. On the other hand, the contact force remains the same as $-\omega_n g_n = -\omega_n u_y|_{y=0} = q$; that is, the product of penetration and penalty parameter remains constant. Note that the contact force is equal and opposite in direction to the distributed load in order to create equilibrium for the block. ■

5.3.5 Frictional Contact Formulation

As mentioned before, frictionless contact is independent of load history. When friction exists at the contact interface, the solution depends on the history of the load applied to the structure. The sequence of the load needs to be considered, and the friction force is determined using not only the current but also the previous location of the contact point. Therefore, it is natural to discuss frictional behavior in the framework of load increment. The current load increment is t_n , and the previous load increment t_{n-1} is converged. For the notational convenience, all variables at load increment t_{n-1} are denoted by a right superscript “0,” and all variables at the current load increment are denoted without any superscript.

The classical Coulomb friction law is commonly used in computational mechanics. However, as mentioned in Sect. 5.2.2, due to the discontinuity in the relationship between slip and friction force, it is difficult to handle in the framework of iterative solution procedures based on the Newton–Raphson method, which assumes that the solution is continuous and smooth. As an alternative, the frictional interface law of Wriggers et al. [9] is employed here. This friction law is a regularized version of Coulomb law, such that the vertical portion of the Coulomb model is changed to an inclined line, as shown in Fig. 5.11. The slope of the regularized line can be related to experimental observation.

The tangent slip form $b_T(\mathbf{u}, \bar{\mathbf{u}})$ in Eq. (5.32) can be expressed in terms of a displacement variation. The first variation of the tangential slip function, presented in Eq. (5.16), becomes

$$\bar{g}_t = \|\mathbf{t}^0\| \bar{\xi}_c = \nu \bar{\mathbf{u}}^T \mathbf{e}_t, \quad (5.45)$$

where a procedure similar to Eq. (5.39) is used. Note that the first variations of $\|\mathbf{t}^0\|$ and ξ_c^0 are zero, since they are the solutions to the previous time increment and fixed at the current time. By using Eq. (5.45), the tangential slip variational form in Eq. (5.32) can be rewritten in terms of the displacement variation as

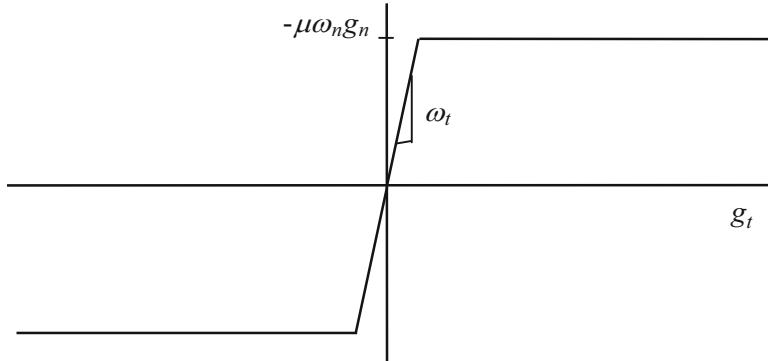


Fig. 5.11 Frictional interface model

$$b_T(\mathbf{u}, \bar{\mathbf{u}}) = \omega_t \int_{\Gamma_c} \nu g_t \bar{\mathbf{u}}^T \mathbf{e}_t d\Gamma. \quad (5.46)$$

The frictional traction force $\omega_t g_t$ works in the tangential direction, is proportional to the tangential slip, and is scaled by curvature through ν . As discussed in Eq. (5.33), the variable $\nu = 1$ when the contact boundary is straight.

The frictional force is bounded above by a compressive normal force multiplied by the friction coefficient in the Coulomb friction law. In the case of a small slip (micro-displacement), however, traction force is proportional to the tangential slip. The penalty parameter ω_t is the proportional constant for this case. An exact stick condition represented by a step function in the classical Coulomb friction law is now regularized by a piecewise linear function, with the penalty parameter ω_t serving as a regularization parameter. As shown in Fig. 5.11, this regularized friction law is reduced to the classical law as $\omega_t \rightarrow \infty$. The regularized stick condition occurs when

$$\omega_t g_t \leq |\mu \omega_n g_n|. \quad (5.47)$$

Otherwise, it becomes a slip condition and $\omega_t g_t = -\mu \omega_n g_n$. In Eq. (5.47), μ is the Coulomb friction coefficient. In the case of a slip condition, the contact variational form has to be modified. Thus, Eq. (5.46) must be divided into two cases as

$$b_T(\mathbf{u}, \bar{\mathbf{u}}) = \begin{cases} \omega_t \int_{\Gamma_c} \nu g_t \bar{\mathbf{u}}^T \mathbf{e}_t d\Gamma & \text{if } |\omega_t g_t| \leq |\mu \omega_n g_n| \\ -\mu \omega_n \text{sgn}(g_t) \int_{\Gamma_c} \nu g_n \bar{\mathbf{u}}^T \mathbf{e}_t d\Gamma & \text{otherwise} \end{cases}. \quad (5.48)$$

Thus, linearization of the tangential slip variational form has to be separated into stick and slip conditions.

5.3.5.1 Linearization of Stick Condition

The first equation in Eq. (5.48) implicitly depends on displacement through v , g_t , and \mathbf{e}_t . The incremental form of g_t can be obtained using the relation in Eq. (5.39) as

$$\begin{aligned}\Delta g_t(\mathbf{u}; \Delta \mathbf{u}) &= \|\mathbf{t}^0\| \Delta \xi_c \\ &= \nu \mathbf{e}_t^T \Delta \mathbf{u}.\end{aligned}\quad (5.49)$$

The incremental form of the unit tangential vector can be derived using a procedure similar to that used in Eq. (5.40) with $\mathbf{e}_t = \mathbf{e}_3 \times \mathbf{e}_n$

$$\begin{aligned}\Delta \mathbf{e}_t &= -\mathbf{e}_3 \times \Delta \mathbf{e}_n \\ &= \frac{\alpha}{c} \mathbf{e}_n (\mathbf{e}_t^T \Delta \mathbf{u}).\end{aligned}\quad (5.50)$$

In addition, the increment of v can be obtained from its definition in Eq. (5.33). After some algebraic calculation, the linearization of Eq. (5.46) leads to the tangential stick bilinear form

$$\begin{aligned}b_T^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) &= \omega_t \int_{\Gamma_c} \nu^2 \bar{\mathbf{u}}^T \mathbf{e}_t \mathbf{e}_t^T \Delta \mathbf{u} d\Gamma \\ &\quad + \omega_t \int_{\Gamma_c} \frac{\alpha \nu g_t}{c} \bar{\mathbf{u}}^T (\mathbf{e}_n \mathbf{e}_t^T + \mathbf{e}_t \mathbf{e}_n^T) \Delta \mathbf{u} d\Gamma \\ &\quad + \omega_t \int_{\Gamma_c} \frac{\nu g_t}{c^2} \left((\gamma \|\mathbf{t}\| - 2\alpha\beta) g_n - \beta \|\mathbf{t}\|^2 \right) \bar{\mathbf{u}}^T \mathbf{e}_t \mathbf{e}_t^T \Delta \mathbf{u} d\Gamma.\end{aligned}\quad (5.51)$$

Again, for the case of a straight contact boundary, only the first terms on the right-hand side of the above equation survives.

The contact bilinear form is the sum of Eqs. (5.41) and (5.51) as

$$b^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) = b_N^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) + b_T^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}).\quad (5.52)$$

In the case of a stick condition, the contact bilinear form in Eq. (5.52) is symmetric with respect to the incremental displacement and variation of displacement. It is noted that the elastic stick contact condition is a conservative system.

5.3.5.2 Linearization of Slip Condition

As the contact point is forced to move along the contact surface, leading to a violation of Eq. (5.47), the slip contact condition is applied and the second equation from Eq. (5.48) is used. In the case of a slip contact condition, the tangential penalty parameter ω_t is related to the impenetrability penalty parameter ω_n according to the following relation:

$$\omega_t = -\mu \omega_n \operatorname{sgn}(g_t). \quad (5.53)$$

The tangential slip form for the slip condition is

$$b_T(\mathbf{u}, \bar{\mathbf{u}}) = \omega_t \int_{\Gamma_c} \nu g_n \bar{\mathbf{u}}^T \mathbf{e}_t d\Gamma. \quad (5.54)$$

The linearization of Eq. (5.54) leads to the tangential slip bilinear form as

$$\begin{aligned} b_T^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) &= \omega_t \int_{\Gamma_c} \nu \bar{\mathbf{u}}^T \mathbf{e}_t \mathbf{e}_n^T \Delta \mathbf{u} d\Gamma \\ &\quad + \omega_t \int_{\Gamma_c} \frac{\alpha \nu g_n}{c} \bar{\mathbf{u}}^T (\mathbf{e}_n \mathbf{e}_t^T + \mathbf{e}_t \mathbf{e}_n^T) \Delta \mathbf{u} d\Gamma \\ &\quad + \omega_t \int_{\Gamma_c} \frac{\nu g_n}{c^2} \left((\gamma \| \mathbf{t} \| - 2\alpha\beta) g_n - \beta \| \mathbf{t} \|^2 \right) \bar{\mathbf{u}}^T \mathbf{e}_t \mathbf{e}_n^T \Delta \mathbf{u} d\Gamma. \end{aligned} \quad (5.55)$$

In the case of a slip condition, the contact bilinear form in Eq. (5.55) is not symmetric with respect to the incremental displacement and variation of the displacement. The system is no longer conservative because frictional slip dissipates energy.

In the case of a general nonlinear material with a frictional contact problem, the principle for virtual work can be written as

$$a(\mathbf{u}, \bar{\mathbf{u}}) + b(\mathbf{u}, \bar{\mathbf{u}}) = \ell(\bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \quad (5.56)$$

The current time is t_n and the current iteration counter is $k+1$. Assuming that the external force is independent of displacement, the linearized incremental equation of Eq. (5.56) is obtained as

$$\begin{aligned} a^*({}^n \mathbf{u}^k; \Delta \mathbf{u}^{k+1}, \bar{\mathbf{u}}) + b^*({}^n \mathbf{u}^k; \Delta \mathbf{u}^{k+1}, \bar{\mathbf{u}}) \\ = \ell(\bar{\mathbf{u}}) - a({}^n \mathbf{u}^k, \bar{\mathbf{u}}) - b({}^n \mathbf{u}^k, \bar{\mathbf{u}}), \quad \forall \bar{\mathbf{u}} \in \mathbb{Z}. \end{aligned} \quad (5.57)$$

Equation (5.57) is linear in incremental displacement for a given displacement variation. This linearized equation is solved iteratively with respect to incremental displacement until the residual forces (the right side of the equation) vanish at each time step.

Example 5.6. Frictional slip of a cantilever beam The cantilever beam in Example 5.4 is now under additional axial load $P = 100$ N at the tip, after the distributed load q is applied. Using the variation of the penalized potential energy, determine the stick or slip condition and calculate the tip displacement. Use friction penalty parameter $\omega_t = 10^6$, axial rigidity $EA = 10^5$ N, and friction coefficient $\mu = 0.5$. Assume the axial displacement in the form of $u(x) = a_0 + a_1 x$.

Solution From Example 5.4, the cantilever beam is in contact with the rigid block with the contact force of $F_c = -\omega_n g_n = 75$ N. From the infinitesimal deformation assumption, the bending behavior of the beam can be decoupled (or sequential) with the axial behavior. Therefore, it is possible to write the penalized potential energy of the axial behavior and take a variation to find an equilibrium. The penalized potential energy becomes

$$\Pi_a = \int_0^L EA(u_{,x})^2 dx - Pu(L) + \frac{1}{2}\omega_t g_t^2 \Big|_{x=L}.$$

The variation of the penalized potential energy becomes

$$\bar{\Pi}_a = \int_0^L EAu_{,x}\bar{u}_{,x} dx - P\bar{u}(L) + \omega_t g_t \bar{g}_t \Big|_{x=L} = 0, \quad \forall \bar{u} \in \mathbb{Z}.$$

The assumed axial displacement must satisfy the essential boundary condition, which is $u(0) = 0$ in this case. Therefore, the first coefficient should be zero, and $u(x) = a_1 x$; only one coefficient needs to be identified. The gradient and its variation of displacement can be written in terms of the unknown coefficient as $u_{,x} = a_1$ and $\bar{u}_{,x} = \bar{a}_1$.

The tangential slip function needs to be expressed in terms of displacement using the definition in Eq. (5.16). In order to simplify the calculation, it can be assumed that the parametric coordinate x has an origin at $x = L$, and it has the same length as the x -coordinate. Based on this setting, it can be derived that $\|\mathbf{x}_{c,\xi}\| = \|\mathbf{t}\| = \|\mathbf{t}^0\| = 1$ and $\xi_c^0 = 0$. In addition, the tangential slip becomes $g_t = \xi_c = u(L) = a_1$.

First, the stick condition is assumed; that is, $\omega_t g_t \leq |\mu \omega_n g_n|$ must be satisfied once the solution is obtained. After substituting the above variables, the variation of the penalized potential energy becomes

$$\bar{a}_1(EAa_1 + \omega_t a_1 - P) = 0, \quad \forall \bar{a}_1 \in \mathbb{R},$$

where \mathbb{R} is the space of real number. In order to satisfy the above equation for all \bar{a}_1 , the terms in the parenthesis must vanish, which can be solved for the unknown coefficient a_1 . Therefore, the axial displacement becomes

$$u(x) = \frac{Px}{EA + \omega_t} = 9.09 \times 10^{-5}x.$$

Using the tip displacement, the stick condition should be checked, as

$$\omega_t g_t = 90.9 > 37.5 = |\mu \omega_n g_n|.$$

Since the assumption of the stick condition is violated, the beam is under the slip condition, where the variation of the penalized potential energy can be written as

$$\bar{\Pi}_a = \int_0^L EAu_{,x}\bar{u}_{,x} dx - P\bar{u}(L) - \mu\omega_n \operatorname{sgn}(g_t)g_n \bar{g}_t|_{x=L} = 0, \quad \forall \bar{u} \in \mathbb{Z}.$$

From the normal contact result in Example 5.4, it can be concluded that $-\mu\omega_n g_n = 37.5 \text{ N}$. Therefore, the above penalized potential energy becomes

$$\bar{a}_1(EAa_1 - P + 37.5) = 0, \quad \forall \bar{a}_1 \in \mathbb{R}$$

which yields $a_1 = 62.5 \times 10^{-5}$ and the tip displacement $u_{\text{tip}} = 0.625 \text{ mm}$, which is consistent with the result in Eq. (5.8). ■

5.4 Finite Element Formulation of Contact Problems

As mentioned before, since the contact formulation is independent of constitutive models, it is enough to discuss finite element formulation of the contact variational form in Eq. (5.30). Then, it can be added to the matrix equation of different materials, for example, elastic material models in Chap. 3 and elastoplastic material models in Chap. 4. Therefore, in the following, only the discretization of the contact variational form will be discussed.

Since the contact problem is solved as a part of finite element analysis, it makes sense to formulate the contact problem in the same context. For that purpose, the contact variational form in Eq. (5.30) is calculated on the boundary of the discretized finite element domain. If the structural domain is discretized by two-dimensional finite elements, then the contact problem is defined on the boundary of two-dimensional finite elements, that is, along a boundary curve. In the case of three dimensions, the contact problem is defined on the boundary surface. In this section, contact conditions in two dimensions are discussed. In order to make the presentation simple, the contact between a flexible body and a rigid body will be discussed first in Sect. 5.4.1, followed by contact between two flexible bodies in Sect. 5.4.2.

5.4.1 Contact Between a Flexible Body and a Rigid Body

The simplest formulation of a contact problem can be obtained when a flexible body is in contact with a rigid body, which is the main topic of this section. In general, it is possible that the rigid body can move to satisfy the equilibrium; but in this text, it is assumed that either the rigid body is fixed or its motion is prescribed. In such a case, it is obvious to choose the flexible body as a slave body and the rigid body as a master body so that the flexible body cannot penetrate the rigid body. In fact, it is sufficient to define the master boundary, not the entire master body. In addition,

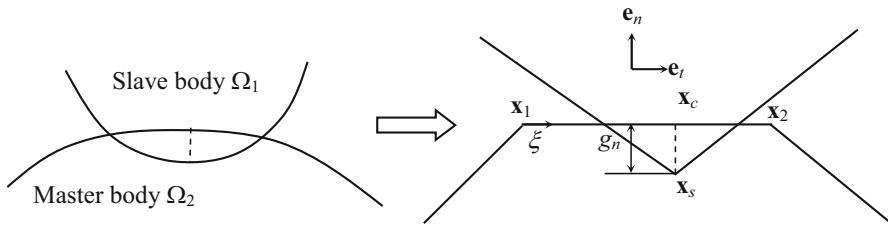


Fig. 5.12 Continuum vs. discrete contact conditions

since the master body is not governed by equilibrium, the contact variational form is only calculated on the slave boundary.

Let Γ_c be a portion of the slave boundary where the slave body penetrates the master body as shown in Fig. 5.12. This boundary is represented by a set of slave nodes that penetrate the master boundary. In the following, a single slave node is considered. Although different ways of defining contact constraints exist in finite elements, in this section it is assumed that the contact constraint is defined using a pair that includes a slave node and a master segment. In addition, only a straight master segment that is defined by two nodes is considered. Therefore, a contact pair can be defined using a slave node and two master nodes, as $\mathbf{X} = \{\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}_2\}^T$. It is possible that one slave node can be associated with different master segments that have a possibility of making contact with the slave node. The two master nodes are ordered in such a way that the master body is located on the right-hand side of the directional line segment from node \mathbf{x}_1 to node \mathbf{x}_2 . The natural coordinate ξ on the master boundary is defined such that it is zero at \mathbf{x}_1 and one at \mathbf{x}_2 .

For a given contact pair \mathbf{X} , the objectives are (1) to find if the contact pair is in contact or in separation and (2) to calculate the contact force and penetration if it is in contact. The first objective is called “contact search.” In a large-scale model, many slave nodes have a possibility of making contact with many master segments. Therefore, the number of contact pairs is huge and a lot of computational time is often consumed in search of contact pairs that are actually in contact. Once these actual contact pairs are identified, the contact force is calculated for these pairs in the second step.

5.4.1.1 Normal Contact

For a given contact pair, the unit normal and tangential vectors can be defined as

$$\mathbf{t} = \mathbf{x}_2 - \mathbf{x}_1, \quad \mathbf{e}_t = \frac{\mathbf{t}}{\|\mathbf{t}\|}, \quad \mathbf{e}_n = \mathbf{e}_3 \times \mathbf{e}_t.$$

Since the master segment is straight and fixed, the above vectors are also fixed. Because of the linear master segment, the contact consistency condition in

Eq. (5.14) and the gap in Eq. (5.15) can explicitly be calculated. First, the gap can be calculated by

$$g_n = (\mathbf{x}_s - \mathbf{x}_l)^T \mathbf{e}_n. \quad (5.58)$$

If $g_n > 0$, then contact does not occur in this pair and no further calculation is required. If $g_n \leq 0$, then one more check is required. That is, the natural coordinate at the contact point must be within $0 \leq \xi_c \leq 1$ so that the contact occurs within this master segment. The natural coordinate at the contact point can be calculated by

$$\xi_c = \frac{1}{\|\mathbf{t}\|} (\mathbf{x}_s - \mathbf{x}_l)^T \mathbf{e}_t.$$

If $\xi_c < 0$ or $\xi_c > 1$, then contact does not occur in this segment and no further calculation is required. If $0 \leq \xi_c \leq 1$, then contact occurs in the segment and needs to calculate the contact force, which acts in the direction to the normal vector and proportional to penetration, as

$$\mathbf{f}_n^c = -\omega_n g_n \mathbf{e}_n.$$

In Newton–Raphson iteration during nonlinear analysis, the tangent stiffness of the above contact force is required. Since \mathbf{e}_n is fixed, no linearization is required. Therefore, only the gap needs to be linearized, which is similar to the variation of the gap in Eq. (5.35), by replacing the displacement variation with the displacement increment. Therefore, the contact stiffness can be obtained by

$$\mathbf{k}_n^c = \omega_n \mathbf{e}_n \mathbf{e}_n^T.$$

In the continuum formulation, $b_N(\mathbf{u}, \bar{\mathbf{u}})$ is expressed as an integral along the boundary Γ_c . However, using the slave–master pair and collocation integration, the boundary integral along the Γ_c is approximated by the summation for those violated slave nodes. Let the contact boundary of body Ω_2 in Fig. 5.12 be represented by piecewise linear master segments, with a slave node on the contact boundary of Ω_1 . Since Γ_c is not known in priori, contact search has to be carried out first to find those violated nodes. Let NC be the number of slave nodes that penetrate the master segment. Then, the discretized contact variational form becomes

$$b_N(\mathbf{u}, \bar{\mathbf{u}}) \approx \sum_{I=1}^{NC} [\bar{\mathbf{u}}^T (\omega_n g_n \mathbf{e}_n)]_I = \sum_{I=1}^{NC} [\bar{\mathbf{u}}^T (-\mathbf{f}_n^c)]_I \equiv \bar{\mathbf{u}}_g^T (-\mathbf{F}_n^c), \quad (5.59)$$

where \mathbf{F}_n^c is the contact force in the global coordinate, which is constructed by adding contact forces at each slave node to the corresponding global degrees of freedom. Since the contact variational form occurs on the left-hand side of Eq. (5.56), it will be moved to the right-hand side as a residual force during a

Newton–Raphson iteration as in Eq. (5.57). Including a negative sign in front of the contact force in Eq. (5.59) is equivalent to adding the contact force to the global residual force.

At this point, it is a good idea to discuss collocation integrals. In general, numerical integration of a function approximates the integral by function values and associated weights at selected integration points. A collocation integral simply chooses the integration points at the node. This choice is a matter of convenience and accuracy. Since most field variables are calculated at nodes in the finite element method, it is convenient to use the nodal values in integration. This is why many finite element programs use collocation integrals for contact analysis. Since the accuracy of numerical integration depends on the number of integration points, a single point integration at a node is less accurate. The weight represents the domain that an integration point covers. For example, if a constant function is integrated over an area with a single integration point, then the weight is the same as the area. However, it is not commonly known that the weight of integration is implicitly included in the function in a collocation integral. For the case of the contact variational form in Eq. (5.59), the weight is included in the gap, g_n . This concept is further explained in Example 5.7.

In a Newton–Raphson iteration during nonlinear analysis, linearization of the contact variational form needs to be calculated, which yields the tangent stiffness matrix. Linearization of $b_N(\mathbf{u}, \bar{\mathbf{u}})$ becomes

$$b_N^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) = \sum_{I=1}^{NC} (\omega_n \bar{\mathbf{u}}^T \mathbf{e}_n \mathbf{e}_n^T \Delta\mathbf{u})_I = \sum_{I=1}^{NC} (\bar{\mathbf{u}}^T \mathbf{K}_n^c \Delta\mathbf{u})_I \equiv \bar{\mathbf{u}}_g^T \mathbf{K}_n^c \Delta\mathbf{u}_g, \quad (5.60)$$

where \mathbf{K}_n^c is the contact stiffness in the global coordinate, which is constructed by adding contact stiffness matrices at each slave node to the corresponding global degrees of freedom. Using a too-large value for the penalty parameter can cause a numerical difficulty because it makes the matrix ill-conditioned.

Example 5.7. Contact force and gap of a block A unit square block is under a uniformly distributed load $q = 1.0 \text{ kN/m}$ on the top surface. The bottom surface is under contact constraint against the rigid floor. When the block is modeled by one and four finite elements, as shown in Fig. 5.13, calculate contact forces and gaps at contact nodes. Assume isotropic material and no friction in the contact interface. Use the penalty parameter $\omega_n = 10^5$.

Solution Since a uniformly distributed load is applied, the finite elements are under constant stress. Therefore, without detailed calculation, the contact forces at the bottom two nodes should be $f_1^c = f_2^c = 500 \text{ N}$. Since the contact force is generated by the gap multiplied by the penalty parameter, the gap at two nodes should be $g_{n1} = g_{n2} = -10^5 \times 500 = -0.005$.

In the case of four elements, the two bottom elements are in contact with the rigid floor. Since each element is under constant stress and distributes the equal contact force to the two nodes, the contact forces at the three nodes become

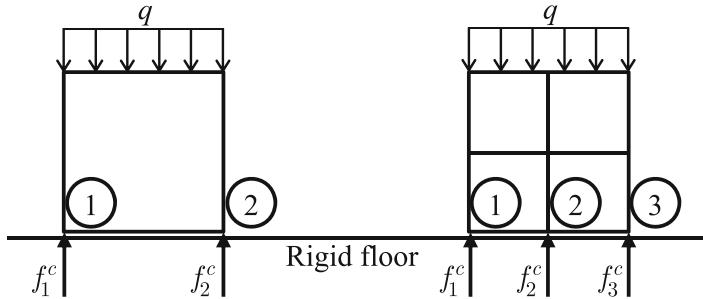


Fig. 5.13 Contact forces of square block

$f_1^c = f_3^c = 250 \text{ N}$ and $f_2^c = 500 \text{ N}$. Again, using the penalty method, the gap at each node can be calculated using the contact force by $g_{n1} = g_{n3} = -0.0025$ and $g_{n2} = -0.005$. As more elements are used along the contact boundary, the gap will become smaller. It is clear that the normal gap changes as the element size changes, which means that the discretized normal gap includes the integration weight implicitly. ■

5.4.1.2 Frictional Slip

Different from the normal contact, the tangential slip under friction requires information from the current as well as the reference configuration. The reference configuration can be the initial state or the previous time increment, but for a large deformation problem, the previous time increment can be more accurate. For the straight master segment, the tangential slip is first defined by

$$g_t = l^0 (\xi_c - \xi_c^0), \quad (5.61)$$

where the right superscript “0” denotes the value evaluated at the previous time, $\xi_c \in [0, 1]$ is the natural coordinate corresponding to the contact point on the master segment, and l^0 is the length of the master segment. Since the master segment is rigid, its length does not change, but the above definition is used in order to be compatible with the case of two flexible bodies in contact.

In the penalty method, the friction force is generated if the tangential slip is not zero, proportional to the tangential penalty parameter ω_t , as

$$\mathbf{f}_t^c = -\omega_t g_t \mathbf{e}_t. \quad (5.62)$$

The above linear relationship is called the stick condition because the tangential slip disappears if the tangential force vanishes. Therefore, the two bodies stick together and behave similar to an elastic material. Also, in the stick condition, the friction force can be understood as a recovery force against tangential deformation. Therefore, the stick condition represents an elastic deformation before slip occurs.

However, the tangential force cannot indefinitely increase in the Coulomb friction model. The magnitude of the tangential force is limited by the friction coefficient multiplied by the normal contact force as in Eq. (5.47). Once the slip is greater than the limit, then the tangential force is limited by

$$\mathbf{f}_t^c = \mu \omega_n \operatorname{sgn}(g_t) g_n \mathbf{e}_t, \quad \text{if } |\omega_t g_t| \geq |\mu \omega_n g_n|, \quad (5.63)$$

where μ is the frictional coefficient. Thus, the tangential contribution is separated into two cases: the stick and slip conditions. The stick condition is applied when the tangential force is small such that only a microscopic relative movement is observed and the frictional force is proportional to the relative deformation. The slip condition is applied when macroscopic movement is occurred with the critical force. In this case, Eq. (5.63) is used to calculate tangential friction force.

For the stick condition, the tangent stiffness of the above frictional force becomes

$$\mathbf{k}_t^c = \omega_t \mathbf{e}_t \mathbf{e}_t^T$$

while the tangent stiffness of the slip condition becomes

$$\mathbf{k}_t^c = \mu \omega_n \operatorname{sgn}(g_t) \mathbf{e}_t \mathbf{e}_n^T.$$

Note that the tangent stiffness matrix is unsymmetric for the slip condition.

Now, the tangential slip form representing the frictional behavior of the contact interface can be discretized by

$$b_T(\mathbf{u}, \bar{\mathbf{u}}) \approx \sum_{I=1}^{NC} [\bar{\mathbf{u}}^T (-\mathbf{f}_t^c)]_I \equiv \bar{\mathbf{u}}_g^T (-\mathbf{F}_t^c), \quad (5.64)$$

where \mathbf{F}_t^c is the frictional force in the global coordinate, which is constructed by adding frictional forces at each slave node to the corresponding global degrees of freedom. And the linearization of the tangential slip form yields

$$b_T^*(\mathbf{u}; \Delta \mathbf{u}, \bar{\mathbf{u}}) \approx \sum_{I=1}^{NC} (\bar{\mathbf{u}}^T \mathbf{k}_t^c \Delta \mathbf{u})_I \equiv \bar{\mathbf{u}}_g^T \mathbf{K}_t^c \Delta \mathbf{u}_g,$$

where \mathbf{K}_t^c is the contact stiffness in the global coordinate, which is constructed by adding contact stiffness matrices at each slave node to the corresponding global degrees of freedom.

5.4.2 Contact Between Two Flexible Bodies

When both the slave and master bodies are flexible, the finite element discretization becomes more complicated as the contact force is applied to both bodies. In this case, the slave nodes are the boundary nodes of the slave finite elements, while the master segments are the edges of master finite elements. In the case of self-contact, the slave finite elements are the same with the master finite elements.

The contact force is directly calculated at the slave node. However, in the case of the master segment, the contact force is applied at the ξ_c location of the segment. Therefore, the contact force is distributed to the two master nodes proportional to the distance from the contact point. For example, when the contact force \mathbf{f}_n^c occurs at ξ_c , this force is distributed to two master nodes by $[\mathbf{f}_{n1}^c, \mathbf{f}_{n2}^c] = [-(1 - \xi_c)\mathbf{f}_n^c, -\xi_c\mathbf{f}_n^c]$. The negative sign is added in the contact force because the direction of contact force is reversed at the master segment. Therefore, in the contact pair, $\hat{\mathbf{x}} = \{\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}_2\}^T$, the contact forces can also be written in the same format as $\hat{\mathbf{f}}_n^c = \{\mathbf{f}_n^c, -(1 - \xi_c)\mathbf{f}_n^c, -\xi_c\mathbf{f}_n^c\}^T$. Considering the contact force as an internal force, the sum of contact forces in a contact pair vanishes. In the following, the superposed “hat” symbol will be used to represent the nodal values in a contact pair, for example, $\hat{\mathbf{u}} = \{\mathbf{u}_s, \mathbf{u}_1, \mathbf{u}_2\}^T$.

In the multi-body contact, the contact variational forms in Eq. (5.32) need to be modified to include the effect of master surface, as

$$\begin{aligned} b_N(\mathbf{u}, \bar{\mathbf{u}}) &= \omega_n \int_{\Gamma_c} g_n \mathbf{e}_n^T (\bar{\mathbf{u}}_s - \bar{\mathbf{u}}_c) d\Gamma, \\ b_T(\mathbf{u}, \bar{\mathbf{u}}) &= \omega_t \int_{\Gamma_c} g_t \left(\nu \mathbf{e}_t^T (\bar{\mathbf{u}}_s - \bar{\mathbf{u}}_c) + \frac{g_n \|\mathbf{t}^0\|}{c} \mathbf{e}_n^T \bar{\mathbf{u}}_{c,\xi} \right) d\Gamma. \end{aligned}$$

Note that the tangential slip form has a normal component; the contact point can move in both normal and tangential directions due to the movement of master surface. For detailed derivations of the above equations, readers are referred to Kim et al. [10].

For notational convenience of derivations, the following sets of vectors (6×1) are defined, which has become quite standard [11]:

$$\begin{aligned} \hat{\mathbf{u}} &= \begin{bmatrix} \mathbf{u}_s \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} \mathbf{e}_n \\ -(1 - \xi_c)\mathbf{e}_n \\ -\xi_c\mathbf{e}_n \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{e}_t \\ -(1 - \xi_c)\mathbf{e}_t \\ -\xi_c\mathbf{e}_t \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{e}_n \\ \mathbf{e}_n \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{e}_t \\ \mathbf{e}_t \end{bmatrix} \\ \mathbf{C}_n &= \mathbf{N} - \frac{g_n}{l} \mathbf{Q}, \quad \mathbf{C}_t = \mathbf{T} + \frac{g_n}{l} \mathbf{P}. \end{aligned} \tag{5.65}$$

Then, the contact variational form $b(\mathbf{u}, \bar{\mathbf{u}})$ can be discretized by

$$\begin{aligned}
b(\mathbf{u}, \bar{\mathbf{u}}) &= b_N(\mathbf{u}, \bar{\mathbf{u}}) + b_T(\mathbf{u}, \bar{\mathbf{u}}) \\
&\approx \sum_{I=1}^{NC} \left[\hat{\bar{\mathbf{u}}}^T (\omega_n g_n \mathbf{N}) \right]_I + \sum_{I=1}^{NC} \left[\hat{\bar{\mathbf{u}}}^T (\omega_t g_t \mathbf{C}_t) \right]_I \\
&\equiv \bar{\mathbf{u}}_g^T (-\mathbf{F}_C),
\end{aligned} \tag{5.66}$$

where \mathbf{F}_C is the contact force in the global coordinate, which is constructed by adding contact forces at slave and master nodes to the corresponding global degrees of freedom.

The linearization of the above contact variational form can be obtained by following a similar procedure as with the previous section. First, linearization of $b_N(\mathbf{u}, \bar{\mathbf{u}})$ becomes

$$b_N^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) = \sum_{I=1}^{NC} \left(\omega_n \hat{\bar{\mathbf{u}}}^T [\mathbf{C}_n \mathbf{C}_n^T] \Delta\hat{\mathbf{u}} \right)_I \equiv \bar{\mathbf{u}}_g^T \mathbf{K}_N \Delta\mathbf{u}_g. \tag{5.67}$$

Linearization of $b_T(\mathbf{u}, \bar{\mathbf{u}})$ should be considered in two different cases.

For the stick condition,

$$\begin{aligned}
b_T^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) &= \sum_{I=1}^{NC} \left(\omega_t \hat{\bar{\mathbf{u}}}^T \left[\mathbf{C}_t \mathbf{C}_t^T + \frac{2g_t}{l} (\text{sym}(\mathbf{C}_n \mathbf{P}^T) - \text{sym}(\mathbf{C}_t \mathbf{Q}^T)) \right] \Delta\hat{\mathbf{u}} \right)_I \\
&\equiv \bar{\mathbf{u}}_g^T \mathbf{K}_T \Delta\mathbf{u}_g.
\end{aligned} \tag{5.68}$$

For the slip condition,

$$\begin{aligned}
b_T^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) &= \sum_{I=1}^{NC} \left(\omega_t \hat{\bar{\mathbf{u}}}^T \left[\mathbf{C}_t \mathbf{N}^T + \frac{2g_N}{l} (\text{sym}(\mathbf{C}_n \mathbf{P}^T) - \text{sym}(\mathbf{C}_t \mathbf{Q}^T)) \right] \Delta\hat{\mathbf{u}} \right)_I, \\
&\equiv \bar{\mathbf{u}}_g^T \mathbf{K}_T \Delta\mathbf{u}_g,
\end{aligned} \tag{5.69}$$

where $\omega_t = \mu \omega_n \text{sgn}(g_t)$ is used for the slip condition in Eq. (5.69) and $\text{sym}(\cdot)$ is the symmetric part of the matrix. Note that the matrix \mathbf{K}_T in the slip condition is not symmetric. Thus, the linearization of the contact variational form is obtained as

$$\begin{aligned}
b^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) &= b_N^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) + b_T^*(\mathbf{u}; \Delta\mathbf{u}, \bar{\mathbf{u}}) \\
&\approx \bar{\mathbf{u}}_g^T \mathbf{K}_N \Delta\mathbf{u}_g + \bar{\mathbf{u}}_g^T \mathbf{K}_T \Delta\mathbf{u}_g \\
&= \bar{\mathbf{u}}_g^T \mathbf{K}_C \Delta\mathbf{u}_g,
\end{aligned} \tag{5.70}$$

where \mathbf{K}_C is the contact tangent stiffness matrix in the global coordinate. After combining with the structural matrix equation and transforming to the physical coordinate, the incremental variational equation and corresponding matrix equation are obtained as

$$\bar{\mathbf{u}}_g^T \mathbf{K} \Delta \mathbf{u}_g + \bar{\mathbf{u}}_g^T \mathbf{K}_C \Delta \mathbf{u}_g = \bar{\mathbf{u}}_g^T \mathbf{F}_{\text{res}} + \bar{\mathbf{u}}_g^T \mathbf{F}_C. \quad (5.71)$$

The incremental discrete variational equation (5.71) must satisfy for all $\bar{\mathbf{u}}_g$ that satisfy the homogeneous essential boundary conditions. One of the common methods in imposing this condition is to delete those rows that correspond to the essential boundary from the above matrix equation. After performing this removal process, we can obtain the reduced form of the incremental matrix equation:

$$(\mathbf{K} + \mathbf{K}_C) \Delta \mathbf{u}_g = \mathbf{F}_{\text{res}} + \mathbf{F}_C. \quad (5.72)$$

As can be seen in the above equation, the contribution from the contact constraints is separated from the contribution from the structural problem. Thus, the contact constraints can be implemented independently of the structural constitutive model. For both the elastic problem in Chap. 3 and elastoplastic problem in Chap. 4, the contact stiffness matrix \mathbf{K}_C and the contact residual force \mathbf{F}_C need to be added to the system matrix equation.

5.4.3 MATLAB Code for Contact Analysis

The MATLAB program **cntelm2d** calculates the contact force and contact tangent stiffness for a contact pair, whose current coordinates are defined in the **ELXY** array. The format of the **ELXY** array is

$$\text{ELXY} = \begin{bmatrix} x_s & x_1 & x_2 \\ y_s & y_1 & y_2 \end{bmatrix}.$$

ELXYP is the same as **ELXY**, except that the array stores the coordinates of the contact pair at the previous time increment. **OMEGAN**, **OMEGAT**, and **CFRI** are, respectively, the two penalty parameters and the coefficient of friction. If **LTAN** is not zero, then **cntelm2d** calculates the contact tangent stiffness matrix **STIFF**. The contact force, **FORCE**, will always be calculated.

The program first checks if contact occurs in the given contact pair. This check is performed in two ways: (1) the gap must be negative and (2) the natural coordinate at the contact point must be between zero and one. Also, the program checks if the contact interface has friction, based on the value of the coefficient of friction. Once these two conditions are satisfied, then the contact force vector, whose dimension is (6×1) , is calculated. If **LTAN** is not zero, then the contact tangent stiffness matrix, whose dimension is (6×6) , is also returned.

PROGRAM cntelm2d

```

function [FORCE, STIFF] = cntelm2d(OMEGAN, OMEGAT, CFRI, ELXY, ELXYP, LTAN)
%*****SEARCH CONTACT POINT AND RETURN STIFFNESS AND RESIDUAL FORCE
% IF CONTACTED FOR NORMAL CONTACT
%*****
%
ZERO = 0.D0; ONE = 1.D0; EPS = 1.E-6; P05 = 0.05; FORCE=[]; STIFF=[];
XT = ELXY(:,3)-ELXY(:,2); XLEN = norm(XT);
if XLEN < EPS, return; end
XTP = ELXYP(:,3)-ELXYP(:,2); XLENP = norm(XTP);
%
% UNIT NORMAL AND TANGENTIAL VECTOR
XT = XT/XLEN;
XTP = XTP/XLENP;
XN = [-XT(2); XT(1)];
%
% NORMAL GAP FUNCTION Gn = (X_s - X_1).N
GAPN = (ELXY(:,1)-ELXY(:,2))'*XN;
%
% CHECK IMPENETRATION CONDITION
if (GAPN >= ZERO) || (GAPN <= -XLEN), return; end
%
% NATURAL COORDINATE AT CONTACT POINT
ALPHA = (ELXY(:,1) - ELXY(:,2))'*XT/XLEN;
ALPHA0 = ((ELXYP(:,1)-ELXYP(:,2))'*XTP)/XLENP;
%
% OUT OF SEGMENT
if (ALPHA > ONE+P05) || (ALPHA < -P05), return; end
%
% CONTACT OCCURS IN THIS SEGMENT
XLAMBN = -OMEGAN*GAPN;
XLAMBT = 0;
LFRIC = 1; if CFRI == 0, LFRIC = 0; end
if LFRIC
    GAPT = (ALPHA - ALPHA0)*XLENP;
    XLAMBT = -OMEGAT*GAPT;
    FRTOL = XLAMBN*CFRI;
    LSLIDE = 0;
    if abs(XLAMBT) > FRTOL
        LSLIDE = 1;
        XLAMBT = -FRTOL*SIGN(ONE,GAPT);
    end
end
%
% DEFINE VECTORS
NN = [XN; -(ONE-ALPHA)*XN; -ALPHA*XN];
TT = [XT; -(ONE-ALPHA)*XT; -ALPHA*XT];
PP = [ZERO; ZERO; -XN; XN];
QQ = [ZERO; ZERO; -XT; XT];

```

```

CN = NN - GAPN*QQ/XLEN;
CT = TT + GAPN*PP/XLEN;
%
% CONTACT FORCE
FORCE = XLAMBN*CN + XLAMBT*CT;
%
% FORM STIFFNESS
if LTAN
STIFF = OMEGAN*(CN*CN');
if LFRIC
TMP1 = -CFRI*OMEGAN*SIGN(ONE,GAPT);
TMP2 = -XLAMBT/XLEN;
if LSLIDE
STIFF = STIFF + TMP1*(CT*CN') + TMP2*(CN*PP' + PP*CN' - CT*QQ' - QQ*CT');
else
STIFF = STIFF + OMEGAT*(CT*CT') + TMP2*(CN*PP' + PP*CN' - CT*QQ' - QQ*CT');
end
end
end
end

```

5.5 Three-Dimensional Contact Analysis

The two-dimensional contact formulation in Sect. 5.3 and its finite element discretization in Sect. 5.4 can be extended to three dimensions. However, three-dimensional contact formulations are quite complicated without providing much insight in physical understandings. In this section, a finite element formulation of three-dimensional contact is introduced without considering continuum variational formulation. In order to simplify the presentation, it is assumed that the master body is a rigid body and the master surface is discretized by four-node quadrilateral elements. Only frictionless contact between a slave node and a master element is considered. An extended formulation of three-dimensional contact formulations can be bound in the work by Laursen and Simo [12] or Kim et al. [13].

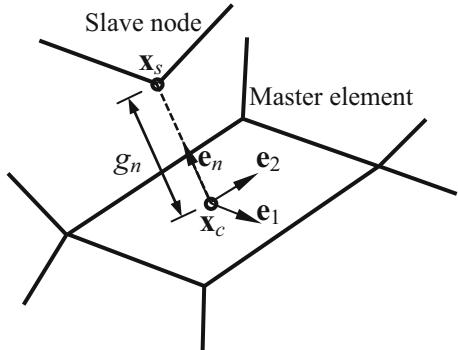
Figure 5.14 shows the contact situation between a flexible slave body and a rigid master body. In the discretized domain, the contact condition between a slave node and a master element is considered. The reference coordinates in finite elements can be used as natural coordinates in contact formulation. Therefore, the master element is represented by the two parameters ξ_1 and ξ_2 such that a point on the element can be expressed as $\mathbf{x}_c(\xi_1, \xi_2)$.

Two tangential vectors in the parametric direction on the master element are defined as

$$\mathbf{t}_\alpha = \mathbf{x}_{c,\alpha}, \quad \alpha = 1, 2, \quad (5.73)$$

where the subscripted comma denotes a partial derivative with respect to the parametric coordinate, i.e., $\mathbf{x}_{c,\alpha} = \partial \mathbf{x}/\partial \xi_\alpha$, $\alpha = 1, 2$. In this section, Greek letters

Fig. 5.14 Contact kinematics and design velocities of two bodies



are used for the index in the direction of the parametric coordinates. Note that \mathbf{t}_1 and \mathbf{t}_2 are not necessarily orthogonal to each other, but are parallel to the contact surface. In the quadrilateral master element, the two tangent vectors can be calculated by differentiating shape functions

$$\mathbf{t}_\alpha = \sum_{I=1}^4 \frac{\partial N_I(\xi_\alpha)}{\partial \xi_\alpha} \mathbf{x}_I, \quad \alpha = 1, 2.$$

The unit outward normal vector on the master surface can be obtained using Eq. (5.73) as

$$\mathbf{e}_n = \frac{\mathbf{t}_1 \times \mathbf{t}_2}{\|\mathbf{t}_1 \times \mathbf{t}_2\|}. \quad (5.74)$$

One of the most important steps in contact analysis process is locating the contact point in an accurate and efficient way. The contact point on the master element corresponding to the slave point can be found from the following consistency condition:

$$\varphi_\alpha = (\mathbf{x}_s - \mathbf{x}_c(\xi_1, \xi_2))^T \mathbf{t}_\alpha(\xi_1, \xi_2) = 0, \quad \alpha = 1, 2, \quad (5.75)$$

which provides the closest projection point \mathbf{x}_c of \mathbf{x}_s , and the corresponding parametric coordinates at the contact point are denoted by (ξ_1^c, ξ_2^c) . For a general master surface, no explicit form of the solution to Eq. (5.75) is available. Finding contact point \mathbf{x}_c efficiently is very important for a large deformation problem. A local Newton–Raphson method can be used to solve nonlinear Eq. (5.75) with a close initial estimate. The contact consistency Eq. (5.75) is two equations with two

unknowns. Using the first-order Taylor series expansion, the following equation for Newton–Raphson iteration can be obtained:

$$\begin{bmatrix} \frac{\partial \varphi_1}{\partial \xi_1} & \frac{\partial \varphi_1}{\partial \xi_2} \\ \frac{\partial \varphi_2}{\partial \xi_1} & \frac{\partial \varphi_2}{\partial \xi_2} \end{bmatrix} \begin{Bmatrix} \Delta \xi_1 \\ \Delta \xi_2 \end{Bmatrix} = \begin{Bmatrix} -\varphi_1 \\ -\varphi_2 \end{Bmatrix}.$$

Once the contact point is found, it is necessary to check if the contact point is within the master element or not. If $\xi_\alpha < -1$ or $\xi_\alpha > 1$, then contact does not occur in this element and no further calculation is required. If $-1 \leq \xi_\alpha \leq 1$, the gap function is defined by the distance between the slave node and the contact point on the master element as

$$g_n = (\mathbf{x}_s - \mathbf{x}_c)^T \mathbf{e}_n \geq 0, \quad (5.76)$$

where the inequality constraint represents the impenetrability condition: the slave point cannot penetrate the master surface. If the gap at the contact point is greater than zero, contact does not occur and no further calculation is required. The violated region of constraint Eq. (5.76) is penalized by applying the contact force, which acts in the direction of the normal vector and proportional to penetration, as

$$\mathbf{f}_n^c = -\omega_n g_n \mathbf{e}_n.$$

In a Newton–Raphson iteration during nonlinear analysis, the tangent stiffness of the above contact force is required. Since \mathbf{e}_n is fixed, no linearization is required. Therefore, only the gap needs to be linearized, which is similar to the variation of the gap in Eq. (5.35) by replacing the displacement variation with the displacement increment. Therefore, the contact stiffness can be obtained by

$$\mathbf{k}_n^c = \omega_n \mathbf{e}_n \mathbf{e}_n^T.$$

The assembly process is identical to two-dimensional contact.

The MATLAB program **cntelm3d** calculates the contact force and contact tangent stiffness for a contact pair, whose current coordinates are defined in the **ELXY** array. Since the master segment is discretized by four-node quadrilateral elements, the format of the **ELXY** array is

$$\text{ELXY} = \begin{bmatrix} x_s & x_1 & x_2 & x_3 & x_4 \\ y_s & y_1 & y_2 & y_3 & y_4 \\ z_s & z_1 & z_2 & z_3 & z_4 \end{bmatrix}.$$

OMEGAN and **Ltan** are the same with **cntelm2d**.

The master element is parameterized by two natural coordinates, **XI** and **ETA**, and the contact point XC is found by determining these two natural coordinates at the contact point. Different from a two-dimensional contact problem, these natural coordinates cannot be determined explicitly; that is, the contact consistency condition must be solved iteratively using the Newton–Raphson-type method. Since the convergence of the Newton–Raphson method strongly depends on the initial estimate, cntelm3d projects the slave node to the master element and approximately estimates the natural coordinates by projecting the projected slave node to the two tangent vectors.

```
% INITIAL CONTACT POINT ESTIMATE.
[T1, T2, XS] = CUTL(0, 0, ELXY);
XN = cross(T1, T2); XN = XN/norm(XN);
XI = (XS'*T1) / (2*norm(T1)^2);
ETA = (XS'*T2) / (2*norm(T2)^2);
GN = XN'*XS;
```

If the estimated **XI** and **ETA** are out of their ranges with a safety margin $[-2, 2]$, then it is clear that the current contact pair is not in contact and no further calculation is performed. Also, if the estimated gap is positive with a safety margin, it is also concluded that no contact occurs.

Once the initial estimate is within the thresholds, then Newton–Raphson iteration is performed to find the accurate contact point. Since the master element is linear quadrilateral, the iteration should converge within two or three iterations. Once the accurate contact point is determined and if the gap is negative, then the contact force is calculated proportional to the amount of penetration. Also, if **Ltan** is not zero, then the contact tangent stiffness matrix, whose dimension is (6×6) , is also returned.

```
PROGRAM cntelm3d
function [FORCE, STIFF] = cntelm3d(OMEGAN, ELXY, LTAN)
%*****
% CALCULATE CONTACT FORCE AND STIFFNESS FOR NORMAL CONTACT FOR 3D
%*****
%
EPS=1.E-6; TL1=2; TL2=0.1; TL3=1.01; FORCE=[]; STIFF=[];
%
% INITIAL CONTACT POINT ESTIMATE.
[T1, T2, XS] = CUTL(0, 0, ELXY);
XN = cross(T1, T2); XN=XN/norm(XN);
XI = (XS'*T1) / (2*norm(T1)^2);
ETA = (XS'*T2) / (2*norm(T2)^2);
GN = XN'*XS;
XX=(ELXY(:,2)-ELXY(:,3)+ELXY(:,4)-ELXY(:,5))/4;
%
% INITIAL SCREENING OF OUT OF BOUNDS
if((XI<-TL1) || (XI>TL1) || (ETA<-TL1) || (ETA>TL1) || GN>TL2), return; end
%
```

```
% FIND EXACT CONTACT POINT THROUGH NEWTON-RAPHSON METHOD
for ICOUNT=1:20
    [T1, T2, XS] = CUTL(ETA, XI, ELXY);
    A=[-T1'^2*T1, XS'^2*XX-T2'^2*T1; XS'^2*XX-T2'^2*T1, -T2'^2*T2];
    B=[-XS'^2*T1; -XS'^2*T2];
    DXI=A\B;
    XI=XI+DXI(1); ETA=ETA+DXI(2);
    if(norm(DXI)<EPS), break; end
end
%
% CHECK THE RANGE OF NATURAL COORD .
if((XI<-TL3) || (XI>TL3) || (ETA<-TL3) || (ETA>TL3)), return; end
%
% NORMAL GAP FUNCTION AND CONTACT FORCE
XN = cross(T1, T2); XN=XN/norm(XN);
GN = XN'*XS;
if GN>0, return; end
FORCE = -OMEGAN*GN*XN;
%
% FORM STIFFNESS (NONFRICTION)
if LTAN, STIFF = OMEGAN*(XN*XN'); end
end
function [T1, T2, XS] = CUTL(ETA, XI, ELXY)
%*****
% COMPUTE COORD. OF CENTEROID AND TWO TANGENT VECTORS
%*****
XNODE=[0 -1 1 1 -1; 0 -1 -1 1 1];
T1=zeros(3,1); T2=zeros(3,1); XC=zeros(3,1); XS=zeros(3,1);
for J = 1:3
    T1(J) = sum(XNODE(1,2:5).*(1+ETA*XNODE(2,2:5)).*ELXY(J,2:5)./4);
    T2(J) = sum(XNODE(2,2:5).*(1+XI*XNODE(1,2:5)).*ELXY(J,2:5)./4);
    XC(J) = sum((1+XI*XNODE(1,2:5)).*(1+ETA*XNODE(2,2:5)).*ELXY(J,2:5)./4);
    XS(J) = ELXY(J,1) - XC(J);
end

```

5.6 Contact Analysis Procedure and Modeling Issues

The contact formulations in the previous sections are relatively straightforward compared to nonlinear constitutive models in the previous chapters. However, in practice, users often experience difficulty of solving nonlinear problems due to contact. The lack of convergence and significant amount of calculation error can be caused by poorly modeled contact conditions. Therefore, it is important to understand the modeling characteristics of contact problems, which is the objective of this section.

5.6.1 *Contact Analysis Procedure*

In general, contact analysis requires three steps: (1) defining contact pairs and types, (2) searching for the contact point, and (3) calculating contact force and tangent stiffness.

5.6.1.1 **Definition of Contact Pairs and Types**

Since the user does not know the location of the contact boundary, it is necessary to define contact pairs that are already in contact or have a possibility of contact. This is especially important for a large deformation problem where the structural boundary can change its shape significantly during the analysis. Many commercial programs provide a tool to generate all contact pairs automatically or with minimum user actions. In addition to contact pairs, it is also necessary to define the properties of contact interface, including (a) weld contact, (b) rough contact, (3) stick contact, and (4) slip contact.

In weld contact, the slave node is bonded to the master element and there is no relative motion in the interface. There will be no contact search, and all contact pairs are assumed already in contact, which makes this formulation fastest. Conceptually, this is equivalent to the rigid link element or multipoint constraint; the only difference is that the force in the interface is decomposed into normal and tangential components. However, the interface is still under infinitesimal elastic deformation, as it is a part of an elastic body.

Rough contact is similar to weld contact, except that the contact interface may not be initially in contact or the initial contact point can be separated. But once it is in contact, it behaves similar to a weld contact. Therefore, its behavior is similar to the case where the contact interface is rough such that there is no relative motion in the interface, independent of the magnitude of normal contact force.

Stick contact is similar to rough contact in the sense that the contact interface can be closed or separated. The difference from rough contact is that the interface can have a relative motion similar to elastic deformation. When a tangential force acts on the contact interface, rough contact behaves like a rigid link, while stick contact shows a small elastic deformation in the interface. The user needs to specify the tangential stiffness. It is noted that the tangential stiffness matrix is symmetric.

Slip contact is the most general contact formation, where the contact point can be closed or separated. In addition, the contact interface can have a relative motion governed by the Coulomb friction model. In practice, the stick condition is first applied before the slip condition is used in many contact algorithms. Different from stick contact, the tangent stiffness of slip contact is unsymmetric.

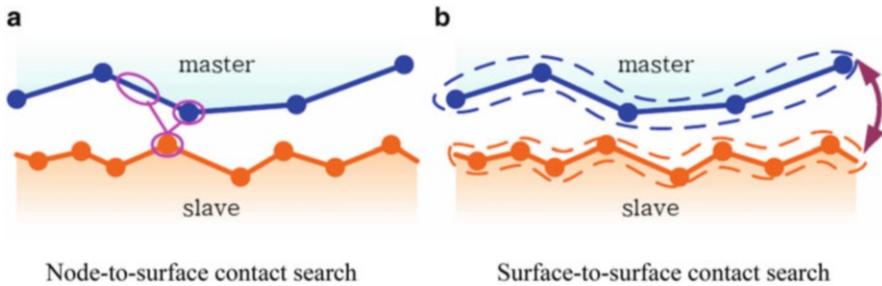


Fig. 5.15 Contact search methods. (a) Node-to-surface contact search. (b) Surface-to-surface contact search

5.6.1.2 Contact Search

The easiest way of performing a contact search is for the user to specify the master element with which that a slave node will contact. This is only possible when deformation is small and no relative motion exists in the contact interface. Slave and master nodes are often located at the same position and connected by a compression-only spring (node-to-node contact). This type of contact pairing works for very limited cases where the user makes slave elements and master elements such that both elements coincide at the contact interface. Also, the contact surface must be simple enough so that the user knows the exact contact region in advance.

In general cases, however, the user does not know about the contact pairs that are actually in contact. Instead, the user specifies all possible candidates. During the contact analysis, the program searches for all contact pairs and determines those pairs that are actually in contact, that is, the violated pairs of the impenetrability condition. Since contact pairs include all possible pairs, the number of pairs is significantly large. For example, if 1,000 slave nodes have a possibility of contacting 1,000 master elements, then theoretically it is necessary to check one million contact pairs. Considering this is required during a single iteration of nonlinear analysis, the program will repeat this search numerous times in order to finish iterative nonlinear analysis. Therefore, it is important to effectively search for contact pairs. Sometimes, it is useful to store the currently contacting master element information for a given slave node, such that in the following iteration, the contact search is performed for only neighboring master elements of the previous one.

In general, contact search is categorized by a node-to-surface and surface-to-surface search (Fig. 5.15). The former is to search if a slave node penetrates the master surface, which is often used when the master surface is rigid. The latter is to search for an impenetrability condition between a slave surface and a master surface. This is useful when the two flexible bodies are under a large slip such that the distinction of slave and master is unclear. Although the latter represents the impenetrability condition more accurately, it takes more computational time due to bidirectional contact (Fig. 5.16).

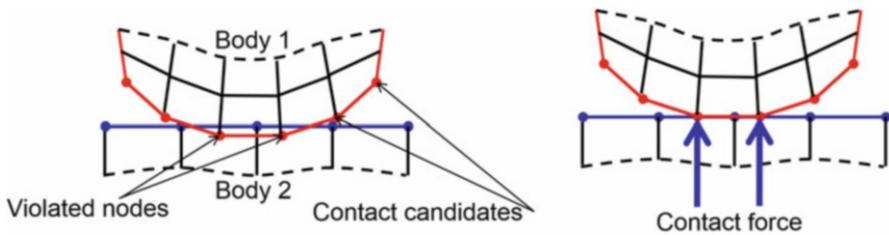


Fig. 5.16 Contact search and contact force

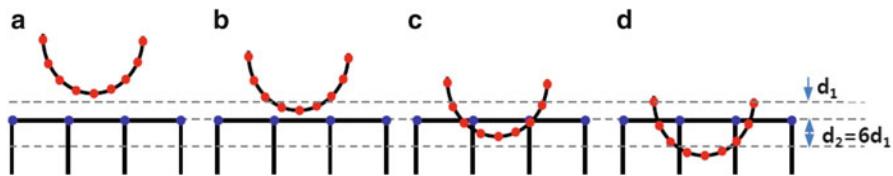


Fig. 5.17 Contact tolerance and detecting contact

Contact tolerance: Since searching for all possible contact pairs is very expensive, commercial programs often use the concept of contact tolerance, which is the minimum distance to search for contact. The default value can be 1 % of the contact element length. This can be used for detecting bodies about to make contact as well as excluding bodies that are on opposite sides. The contact tolerance can be used for compensating for geometric tolerance in the case of weld contact. If two contact surfaces are within the contact tolerance, they are considered in contact and the contact force is calculated. In the case of rough and general contact, contact pair is established when two surfaces are within the contact tolerance.

For example, the two separate bodies in Fig. 5.17 are going to be in contact. Contact tolerance is set in two ways: d_1 for separation and d_2 for penetration. In the cases of (b) and (c), the initial separation or penetration is within the tolerance, the contact pair is established, and the convergence analysis is performed by generating appropriate contact force. In the case of (c), since penetration is relatively large, the load increment is bisected to reduce the amount of penetration. However, in the cases of (a) and (d), the initial separation or penetration is larger than the contact tolerance, the search algorithm fails to detect contact, and, as a result, the two surfaces will be penetrated without contact.

Therefore, an appropriate load increment, as in Fig. 5.18a, should be used in order to make initial contact detection. If the load increment is too large, as in Fig. 5.18c, then the contact search algorithm fails to detect contact because the movement is larger than the contact tolerance. In this case, the contact condition is not established, and, as a result, a too-large penetration occurs in the next load increment, which may completely miss contact detection. In the case of Fig. 5.18b, the contact surface is within the contact tolerance; even if a large penetration may occur, the contact pair will be generated, and the impenetrability condition will be satisfied through bisecting the load increment.

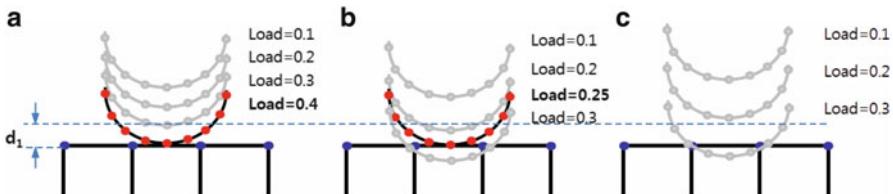


Fig. 5.18 The effect of load increment in contact detection

5.6.1.3 Contact Force and Tangent Stiffness

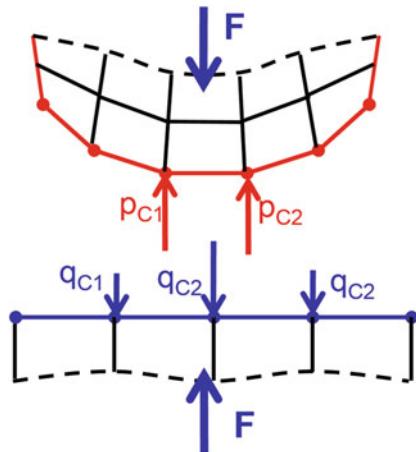
Once the contact pairs are actually in contact (or violated the impenetrability condition), either the penalty method or the Lagrange multiplier method can be applied to satisfy contact constraint. The penalty method is simple and intuitive but allows a small amount of constraint violation. That is, the impenetrability condition will be slightly violated. The amount of violation can be controlled by the penalty parameter. A large penalty parameter allows only a small amount of violation, but a too-large penalty parameter can cause numerical instability because it makes the stiffness matrix ill-conditioned.

Contact stiffness: In practice, the penalty parameter is better selected based on material stiffness, element size, and element height normal to the contact interface. Therefore, it is often called the contact stiffness. If two contacting bodies have different material stiffness, it is calculated based on the softer material. A large value of contact stiffness can reduce penetration, but can also cause a problem in convergence. Therefore, a proper value of contact stiffness must be determined based on allowable penetration, which requires experience. Normally many programs suggest the contact stiffness based on the elastic modulus of contacting bodies and allow users to change it by multiplying a scale factor with a default of one. The user can start with a small initial scale factor and gradually increase it until a reasonable penetration can be achieved.

Tangential stiffness: If the contact stiffness is for the normal contact, tangential stiffness is for the frictional force in the contact interface. Since the frictional force is generated through normal contact force, it depends on the contact stiffness, and its behavior is more complicated because of friction. In the penalty formulation, an elastic stick condition applies before slip occurs under a tangential load. If the tangential load is removed, then the body returns to its original state. The tangential stiffness controls this stick condition. If the tangential stiffness is too large, then the contact interface shows slip without stick. If too small, then the stick condition will be overextended.

Contact force: When two bodies are in contact, the contact force in the interface can be considered as either an internal or external force, depending on how the system is defined. If a free-body diagram is constructed of each body separately,

Fig. 5.19 Contact force on slave nodes and master elements



then the contact force is the externally applied force on the boundary. From this viewpoint, the contact problem is called boundary nonlinearity, because both the boundary and force are unknown. However, if the free-body diagram includes both contacting bodies, then the contact force can be viewed as an internal force. If the entire system is in equilibrium, then all internal forces must vanish. Therefore, the contact force on the slave nodes must be equal and opposite in direction to the contact force on the master elements. This can also be viewed from Newton's third law: equal and opposite forces act on interface. Figure 5.19 shows two contacting bodies in equilibrium. Because the individual bodies as well as both bodies together are in equilibrium, the following relation should be satisfied:

$$\mathbf{F} = \sum_{i=1}^{N_p} p_{ci} = \sum_{i=1}^{N_q} q_{ci}. \quad (5.77)$$

It is noted that in Eq. (5.77), individual p_{ci} and q_{ci} are different in magnitudes because of discretization. The force distribution can be different. However, the resultants should be the same, as the two bodies are in equilibrium.

5.6.2 Contact Modeling Issues

In this section, several modeling issues in contact analysis are summarized. The contents that are covered in this section are by no means complete. However, users should be familiar to these issues in order to solve convergence problems as well as accuracy of analysis results.

Fig. 5.20 Definition of slave and master

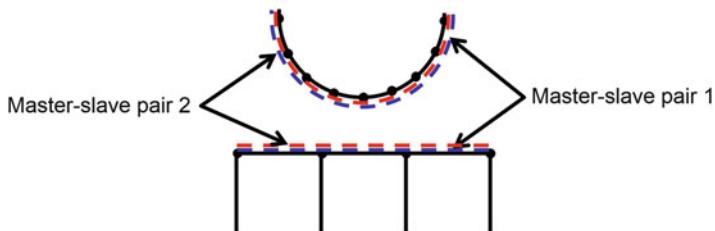
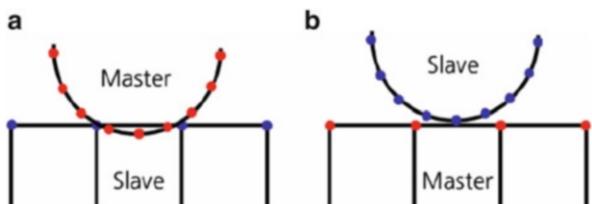


Fig. 5.21 Alternating definition of slave–master pairs in order to prevent penetration from either body

5.6.2.1 Definition of Slave and Master

When two bodies are in contact, the slave–master concept distinguishes body 1 from body 2. Although there is no theoretical reason to distinguish body 1 from body 2, the distinction is often made for numerical convenience. One body is called a slave body, while the other is called a master body. Then, the contact condition is imposed such that the slave body cannot penetrate into the master body. This means that hypothetically the master body can penetrate into the slave body, which is not physically possible but numerically possible because it is not checked. There is not much difference in a fine mesh, but the results can be quite different in a coarse mesh, as shown in Fig. 5.20. When a curved boundary with a fine mesh is selected as a master body, a straight slave boundary with a coarse mesh shows a significant amount of penetration, even if none of slave nodes penetrate into the master body. Therefore, it is important to select the slave and master body in order to minimize this type of numerical error. In general, in order to minimize penetration, a flat and stiff body is selected as a master body, while a concave and soft body is selected as a slave body. Also, it is suggested that a body with a fine mesh be a slave and a body with a coarse mesh be a master. In the case of flexible–rigid body contact, the rigid body is selected as a master body and the flexible one as a slave body.

No matter how the slave and master are selected, it is possible that a master node can penetrate into the slave element. In order to prevent penetration from either body, it is necessary to define the slave–master pair twice by changing their role, as shown in Fig. 5.21. Some surface-to-surface algorithms use this technology to prevent penetration from either body.

5.6.2.2 Flexible Contact vs. Flexible–Rigid Contact

Since all bodies are flexible in the viewpoint of mechanics, it seems natural to model all contacting bodies as flexible and apply contact conditions between flexible bodies. However, since modeling is an abstraction of physical phenomena, it is possible to consider one body as a rigid body, even if in reality it is flexible. Therefore, in such a case, a flexible–rigid body contact condition can be applied. The question is why we want to use flexible–rigid body contact and when we can apply that condition.

The flexible–flexible contact can be applied when two bodies have a similar stiffness and both can deform. For example, metal-on-metal contact can be modeled as flexible–flexible contact. However, when the stiffness of two bodies are significantly different, such as contact between rubber and metal, the behavior of metal can be approximated as a rigid body, because the deformation of metal can be negligible compared to that of rubber. However, this can also depend on physical behavior of the system. For example, if a rubber ball impacts on a thin metal plate, then the plate needs to be modeled as a flexible body because the deformation of the plate can be large.

There are obvious advantages of using flexible–rigid body contact over two flexible–body contact. When two bodies have a large difference in stiffness, the stiffness matrix becomes ill-conditioned and the matrix equation loses many significant digits. Therefore, accurate calculation becomes difficult. In addition, as shown in previous section, the numerical implementation of flexible–rigid body contact formulation is much easier than multi-body contact formulation.

5.6.2.3 Sensitivity of Mesh Discretization

At the continuum level, it is assumed that the contact boundary varies smoothly and the boundary is differentiated two or three times in deriving contact force and tangent stiffness. In the numerical model, however, the contact boundary is approximated by piecewise continuous curves (or straight lines), and only C^0 continuity is guaranteed across the element boundary. Therefore, the slope of the contact boundary is not continuous. Unfortunately, the contact force is very sensitive to the boundary discretization and strongly depends on this slope: contact force acts in the normal direction of the contact boundary. Therefore, if the actual contact point is near the boundary of two elements with a large slope change, it is possible that the Newton–Raphson iteration may have difficulty in convergence.

Another important aspect related to mesh is the distribution of contact stress/contact pressure. As shown in Fig. 5.22, if a uniform pressure is applied on top of a slave body, it is natural to think that the contact pressure on the bottom surface will also be uniform. However, due to the effect of a large master surface at the bottom, the contact pressure is high on the edge of the contacting region. Therefore, the

Fig. 5.22 Contact stress distribution under uniform pressure load

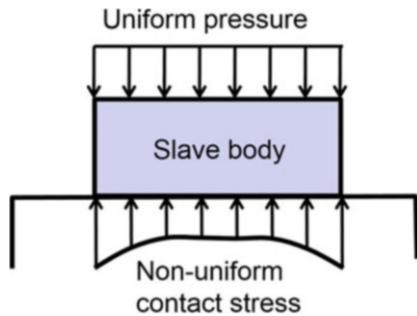
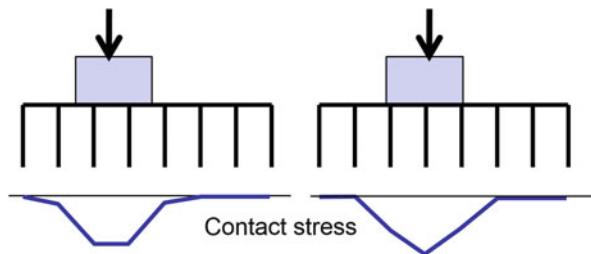


Fig. 5.23 Variation of contact stress distribution as a function of block location



contact stress/pressure is not uniform. Theoretically, the contact stress on the edge can be twice the inside contact stress.

Another important observation on contact stress distribution is that it is sensitive to mesh discretization. As shown in Fig. 5.23, the contact stress distribution is different for different locations of the block. Therefore, it is dangerous to determine the maximum contact stress using a single coarse mesh. It is always recommended to perform mesh sensitivity study to show convergence of contact stress.

5.6.2.4 Rigid-Body Motion

Rigid-body motion in contact is one of the most commonly confused concepts to users. This is also a good example of contact boundary conditions that are different from the displacement boundary conditions. Figure 5.24 shows a cylindrical slave body between two rigid masters. It is assumed that the slave body slightly penetrates into the lower master body, while it has a slight gap with the upper master body. Since the contact force is generated proportional to penetration, the upward contact force will be applied at the lower part of the cylinder, which will move the cylinder upward, as in Fig. 5.24a. Next, the body now penetrates the upper master body because of the previous upward motion. Then, the contact force is now applied from the upper master body and it is not in contact with the lower master

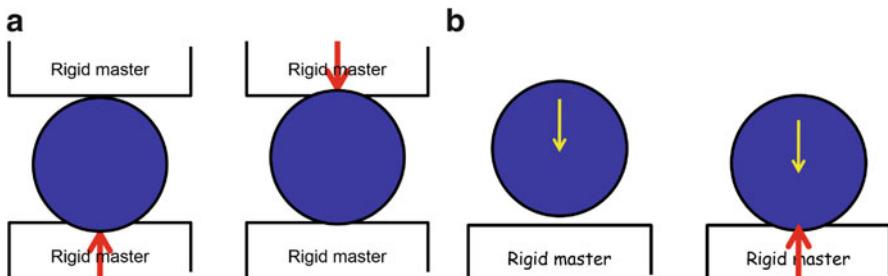


Fig. 5.24 The effect of rigid-body motion in contact

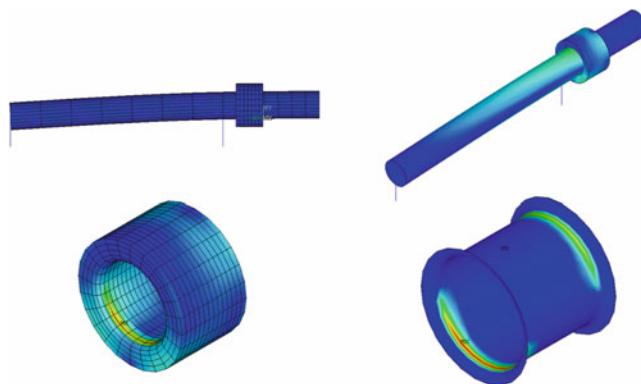


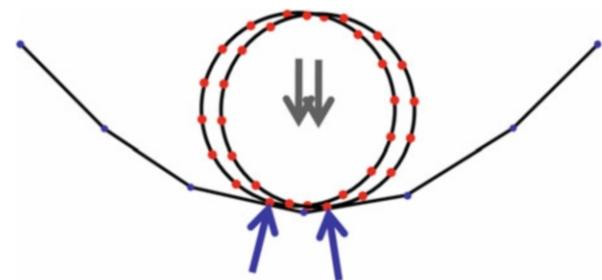
Fig. 5.25 Contact stress at bushing due to shaft bending

body, which will cause a downward contact force. Under this situation, the slave body can either oscillate between the two master surfaces (Fig. 5.24a) or fly out if it is in contact with a single master body (Fig. 5.24b). In fact, without contact, the cylinder is not well constrained. Even if in real physics a body can be stable between two contacting bodies, in numerical analysis, it is better to constrain the flexible body without contact, so that the rigid-body motion can be removed. When a body has rigid-body motion, an initial gap can cause a singular matrix (infinite/very large displacements). The same is true when there is an initial overlap. In order to remove rigid-body motion, it is possible to add a small, artificial bar element so that the body is well constrained while minimally affecting analysis results, as shown in Fig. 5.25, where the shaft is constrained by two bar elements.

5.6.2.5 Convergence Difficulty

Common difficulties in contact analysis are (a) the contact condition does not work, i.e., penetration occurs, and (b) the Newton–Raphson iteration does not converge. The former is related to the contact definition or a too-large load increment.

Fig. 5.26 Discontinuity of contact force by nonsmooth contact boundary



Therefore, this type of problem can be solved relatively easily. On the other hand, the lack of convergence is the most common difficulty in nonlinear analysis, and it is not trivial to find the cause because they can be caused by different reasons.

As the convergence of Newton–Raphson method depends on the initial estimate, it is possible that the method can improve the convergence by starting with the initial estimate that is close to the solution. In the increment force method in Chap. 2, the solution is a function of load increment. A small increment means that the solution from a previous increment is close to the solution in the current load increment. Therefore, using a small load increment is the most common remedy when convergence cannot be obtained. Many commercial programs have the capability to automatically control the load increment. When a given load increment does not converge, then the current increment is reduced by half or a quarter and convergence iteration is retried. This bisection process is repeated until the convergence can be achieved or the program stops when the maximum allowed bisections are consumed or the minimal size of load increment is not converged.

If the Newton–Raphson iteration failed to converge with the smallest load increment, the problem resides in fundamental issues. The basic assumption in Newton–Raphson method is that the nonlinear function is smooth with respect to input parameters. In the context of contact analysis, this can be interpreted as the contact force varies smoothly throughout deformation. Unfortunately, this is a strong assumption in finite element analysis because of discretization. As shown in Fig. 5.26, the slope of finite elements is discontinuous across the element boundary, especially when the contact boundary is curved. As illustrated in the figure, this discontinuity can make the contact force oscillate between two master elements and discontinuously change the direction of contact force. In order to minimize such a situation, it is necessary to use more elements to represent the curve boundary. As a rule of thumb, it is recommended to generate about 10 contact elements along the 90° corner fillet or use higher-order elements.

A nonsmooth contact boundary can also affect the accuracy of contact analysis. As an example, Fig. 5.27 shows contact between a shaft and a hole. In Fig. 5.27a, both the shaft and hole are discretized by 15 linear elements along the circumference. When the mesh locations of both parts are different, the inaccuracy of representing circular geometry significantly affects contact results. Some nodes

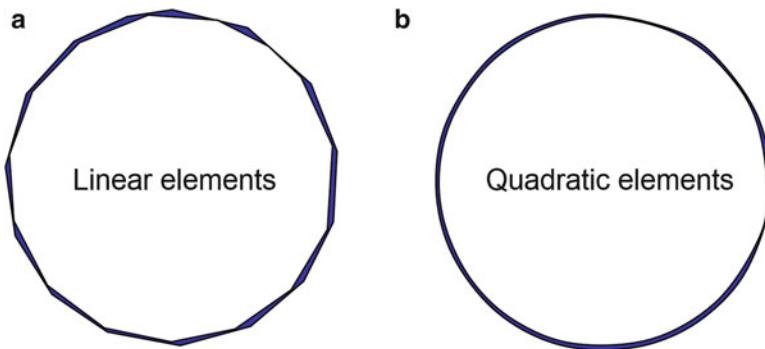


Fig. 5.27 Discretization of circular shaft and hole using (a) linear and (b) quadratic elements

are out of contact, while others are under excessive contact force due to over-penetration. Therefore, the contact stress contour does not show a smooth variation of contact stress. Rather, a localized random and discrete contact stress distribution may be observed. On the other hand, if higher-order elements are used as in Fig. 5.27b, the two contact boundaries become much more conforming and smooth contact stress distribution can be obtained.

5.7 Exercises

- P5.1 For the beam contact problem in Sect. 5.2.1, determine the contact force and tip deflection using the Lagrange multiplier method. Choose the gap g as a Lagrange multiplier.
- P5.2 For the beam contact problem in Sect. 5.2.1, determine the contact force and tip deflection using the Lagrange multiplier method. Model the beam using a two-node Euler beam element. Compare the results with the results in Sect. 5.2.1, and explain the reason for different results.
- P5.3 For the frictional contact problem in Sect. 5.2.2, determine the frictional force and slip displacement using the Lagrange multiplier method. Choose the slip u_{tip} as a Lagrange multiplier.
- P5.4 During a Newton–Raphson iteration, a rectangular plane element is in contact with a rigid surface as shown in the figure. Due to the penalty method, the penetration of $g = -1 \times 10^{-4}$ m is observed with penalty parameter $\omega_n = 10^6$. In the two-dimensional problem, the element has eight degrees of freedom $\{u_{1x}, u_{1y}, u_{2x}, u_{2y}, u_{3x}, u_{3y}, u_{4x}, u_{4y}\}^T$. Calculate the contact force and contact stiffness matrix in terms of 8×1 vector and 8×8 matrix, respectively.

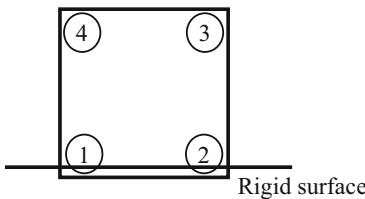


Fig. P5.4 Contact of a rectangular block

P5.5 A sphere of radius $r = 8$ mm is pressed against a rigid flat plane. Using a commercial program, determine the contact radius, a , for a given load $F = (30 \times 2\pi)$ N. Assume a linear elastic material with Young's modulus $E = 1,000$ N/mm 2 and Poisson's ratio $\nu = 0.3$. Use an axisymmetric model. Compare the finite element result with the analytical contact radius of $a = 1.010$ mm.

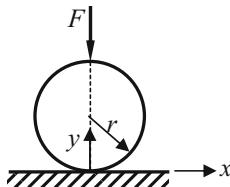


Fig. P5.5 Contact of a sphere

P5.6 A long rubber cylinder with radius $r = 200$ mm is pressed between two rigid plates using a maximum imposed displacement of $\delta_{\max} = 200$ mm. Determine the force-deflection response. Use Mooney-Rivlin material with $A_{10} = 0.293$ MPa and $A_{01} = 0.177$ MPa. Assume a plane strain condition and symmetry. Compare the results with the target results of $F = 250$ N at $\delta = 100$ mm and $F = 1,400$ N at $\delta = 200$ mm.

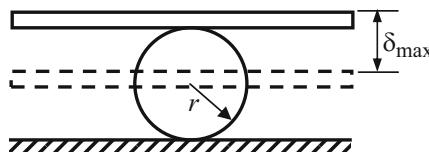


Fig. P5.6 Rubber cylinder contact problem

P5.7 Two long cylinders of radii $R_1 = 10$ mm and $R_2 = 13$ mm, in frictionless contact with their axes parallel to each other, are pressed together with a force per unit length, $F = 3,200$ N/mm. Determine the semi-contact length b and the approach distance d . Both materials are linear elastic with $E_1 = 30,000$ N/mm 2 and $\nu_1 = 0.25$ for Cylinder 1 and $E_2 = 29,120$ N/mm 2 and $\nu_2 = 0.3$ for Cylinder 2. Assume a plane stress condition with a unit

thickness and symmetry. Compare the results with the target results of $d = -0.4181$ mm and $b = 1.20$ mm.

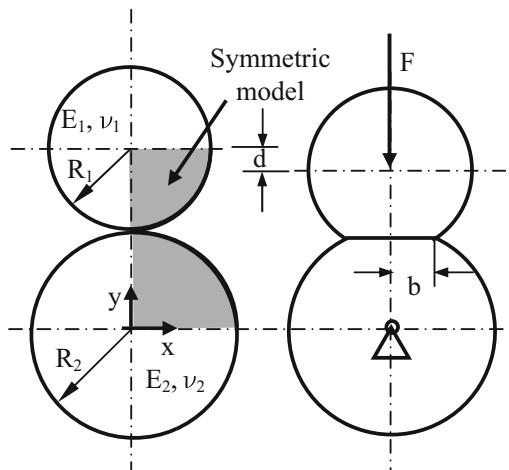


Fig. P5.7 Hertzian contact problem

P5.8 Deep drawing is a manufacturing process that can create a complex shape out of a simply shaped plate (blank). The deep-drawing configuration is shown in the figure, which is composed of a blank, punch, die, and blank holder. The thickness of the initial blank is 0.78 mm. The die is fixed throughout the entire process, while the punch moves down by 30 mm to shape the blank. The holder controls the slip of the blank by applying friction force. The fillet radii of both punch and die are 5 mm. After the maximum downstroke of the punch, both the punch and holder are removed. Then, the blank will experience elastic springback. The objective of this project is to simulate the final geometry of the blank after springback.

Model the process using an axisymmetric problem. You may use CAX4R elements. The whole simulation is divided by three steps. (1) The blank holder is pushed (displacement control) to provide about 100 kN of holding force. (2) While the blank holder is fixed at the location of step (1), the punch is moved down by 30 mm. (3) Punch, die, and blank holder are removed so that the blank is elastically deformed by springback. It is possible to change processes.

The following results need to be submitted: (1) deformed shape plots of five different steps, (2) graph of radial position vs. radial strain, and (3) graph of radial position vs. thickness change, (4) graph of punch displacement vs. punch force, and (5) comparison of deformed shapes at the maximum stroke and after springback.

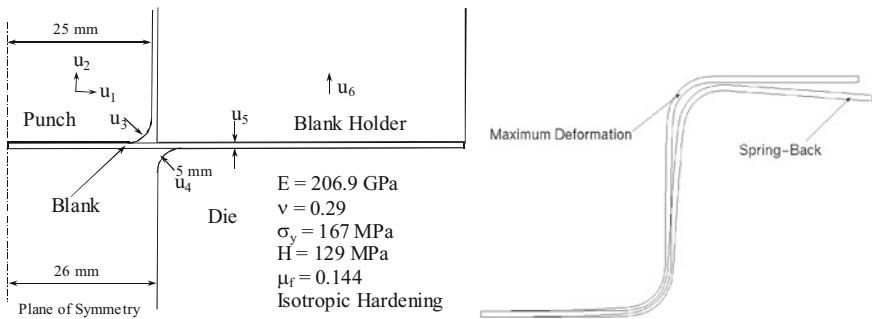


Fig. P5.8 Deep-drawing problem

References

1. Duvaut G, Lions JL. Inequalities in mechanics and physics. Berlin: Springer-Verlag; 1976.
2. Kikuchi N, Oden JT. Contact problems in elasticity: a study of variational inequalities and finite element method. Philadelphia: SIAM; 1988.
3. Ciarlet PG. The finite element method for elliptic problems. New York: North-Holland; 1978.
4. Klarbring A. A mathematical programming approach to three-dimensional contact problems with friction. Comput Methods Appl Mech Eng. 1986;58:175–200.
5. Kwak BM. Complementarity-problem formulation of 3-dimensional frictional contact. J Appl Mech Trans ASME. 1991;58(1):134–40.
6. Luenberger DG. Linear and nonlinear programming. Boston: Addison-Wesley; 1984.
7. Barthold FJ, Bischoff D. Generalization of Newton type methods to contact problems with friction. J Mec Theor Appl. 1988;7 Suppl 1:97–110.
8. Arora JS. Introduction to optimum design. 2nd ed. San Diego: Elsevier; 2004.
9. Wriggers P, Van TV, Stein E. Finite element formulation of large deformation impact-contact problems with friction. Comput Struct. 1990;37:319–31.
10. Kim NH, Park YH, Choi KK. An optimization of a hyperelastic structure with multibody contact using continuum-based shape design sensitivity analysis. Struct Multidiscip Optim. 2001;21(3):196–208.
11. DeSalvo GJ, Swanson JA. ANSYS engineering analysis system, user's manual, vols. I and II. Houston: Swanson Analysis Systems Inc.; 1989.
12. Laursen TA, Simo JC. A continuum-based finite element formulation for the implicit solution of multibody large deformation frictional contact problems. Int J Numer Methods Eng. 1993;36:2451–3485.
13. Kim NH, Yi KY, Choi KK. A material derivative approach in design sensitivity analysis of 3-D contact problems. Int J Solids Struct. 2002;39(8):2087–108.

Index

A

Assembly, 57
Associative plasticity, 291

B

Back stress, 282, 334
Backward Euler method, 291
Balance of momentum, 37
Basis vectors, 4
Baushinger effect, 278
Broyden, Fletcher, Goldfarb, and Shanno (BFGS) method, 107
Bisection, 116
Boundary condition, 38, 54
 essential, 54
 natural, 54
Boundary valued problem, 38, 54
Bulk modulus, 192

C

Cauchy–Green tensor, 145, 147, 176, 191, 327, 331, 343
 left, 147, 327, 331, 343
 right, 145, 327, 343
Cauchy’s Lemma, 20
Consistency condition, 372, 380
 contact, 380, 409
Constitutive relation, 31
Constrained optimization, 384
 contact, 384
Contact force, 372, 410, 417
 normal, 410
Contact form, 387

 normal, 387
 tangential, 387
Contact pair, 413
Contact problem, 367
Contact search, 414
Contact stiffness, 410, 416
Contact tolerance, 415
Contraction, 8
Convergence, 94, 421
Convex set, 382
Coulomb friction, 375, 393
Critical displacement, 180
Critical load, 179, 181, 183
 actual load factor, 183
 load factor, 181
 one-point, 181
 two-point, 181
Cross product. *See* Vector, product

D

Deformation field, 27
Deformation gradient, 144, 330
 relative, 330
Deviator, 274
Directional derivative, 385
Displacement field, 27
Displacement gradient, 144
Dissipation function, 327, 328
Dissipation inequality, 328
Distortion energy theory, 268
Divergence, 11
Divergence theorem, 12
Dual vector, 10
Dyadic product, 5

E

- Effective plastic strain, 282
- Eigenvalue, 23, 182
- Eigenvector, 23
- Elastic domain, 282, 326
- Elasticity matrix, 34
- Elasticity tensor, 32
- Elastic limit, 32
- Elastic modulus, 243
- Elastic predictor, 291
- Elastoplasticity, 241, 273, 308, 325, 360
 - finite deformation, 360
 - finite rotation, 308
 - infinitesimal, 273
 - multiplicative plasticity, 325
- Euclidean norm, 157

F

- Failure envelope, 267
- Finite element, 50, 51, 62
 - shape function, 62
- Flow potential, 283
- Form, 44
 - energy bilinear, 44
 - load linear, 44
- Frame indifference, 21
- Fréchet differentiable, 43
- Free energy, 327, 332
- Friction, 374

G

- Gap, 370, 390
- Gap function, 410
- Gauss integration, 65
- Gauss' theorem, 47
- Generalized Hooke's law, 31, 32
- Generalized solution, 40
- Gradient, 11
- Green's identity, 14

H

- Hooke's law, 15
 - generalized, 15
- Hydrostatic pressure, 192
- Hyper-elastic material, 184

I

- Impenetrability, 372
- Impenetrability condition, 379, 380

- Incremental force method, 109
- Initial stiffness, 170, 298
- Inner product, 4
- Integration-by-parts, 13
- Interpolation function, 53
- Invariant, 185
- Isoparametric mapping, 62
- Isotropic hardening, 282

J

- Jacobian, 94
- Jacobian matrix, 116

K

- Kinematically admissible displacement, 40
- Kinematic hardening, 282, 283
- Kronecker delta symbol, 4, 164
- Kuhn–Tucker condition, 284, 329

L

- Lagrange multiplier, 284, 368, 372, 376
- Lagrangian strain, 167
- Lame's constants, 33, 163, 281
- Laplace operator, 11
- Lie derivative, 327
- Load step, 110
- Lower and upper (LU) decomposition, 101

M

- Master, 371
- Master element, 408
- Material description, 168
- Matrix, 5, 23, 34
 - determinant, 23
 - elasticity, 34
- Modified Newton–Raphson method, 101–103
- Mooney–Rivlin material, 186–187

N

- Natural coordinate, 379
 - contact problem, 379
- Necking, 32
- Neo–Hookean material, 186
- Newton–Raphson method, 93, 168
- Nonlinear elastic problem, 162
- Nonlinearity, 162, 241, 367
 - boundary, 367
 - force, 90–91

- geometric, 85–87, 164
kinematic, 89–90
material, 87–89, 241
Nonlinear solution procedure, 91
Norm, 5, 8
Normal gap, 380
- O**
Objective rate, 360
Operator, linear, 81
- P**
Penalty, 368, 372, 377
Penalty method, 386
Penalty parameter, 373, 386
Penetration, 372
Permutation, 10, 156
Plane strain, 34
Plane stress, 34
Plastic consistency
 parameter, 283
Plastic corrector, 291
Plastic modulus, 246, 284, 334
Poisson’s ratio, 33
Polar decomposition, 150
Potential energy, 166, 384, 386
Principal stress direction, 22, 24
Principal stretch, 332
Principle of minimum potential
 energy, 39
Principle of virtual work, 46
Projection, 4, 290, 380
Proportional limit, 32
- R**
Reference element, 65
Residual, 94, 116
Residual load, 170, 299
Return mapping, 292, 333, 360
Reynolds transport theorem, 13
Rigid-body motion, 421
Rigid body rotation, 315
Rotation tensor, 150
- S**
Secant method, 104
Secant stiffness matrix, 107
Shape function, 53
Shear modulus, 33
- Slave, 371
Slave-master, 368
Slave node, 399
Slip, 375, 393
Slip condition, 394
Sobolev space, 40
Solution, 44, 52
 generalized, 44
 trial, 52
Spatial description, 174
Spatial velocity gradient, 327
Spin tensor, 315
Stick, 375
Stick condition, 395
Stiffness matrix, 57, 64, 296
 consistent, 296
 solid, 64
Strain, 7, 26, 28–30, 145, 147, 167,
 170, 174, 268, 332, 334
 deviatoric, 30, 268
 effective plastic, 334
 elastic principal stretch, 332
 engineering, 174
 engineering shear, 28
 Eulerian, 147
 infinitesimal, 145, 172
 Lagrangian, 145, 167, 170
 normal, 28
 shear, 28
 symmetric, 29
 tensorial shear, 28
 volumetric, 30
Strain energy, 39, 163, 281
 elastic, 281
Strain energy density, 268
 distortion, 268
Strain hardening, 32
Stress, 17, 18, 20–22, 24, 31, 32, 159,
 160, 174, 268, 291, 314, 316, 326
 Cauchy, 159, 174, 314
 deviatoric, 21, 268
 first Piola–Kirchhoff, 159, 318
 invariant, 24
 Kirchhoff, 160, 326
 mean, 21
 normal, 20
 principal, 22
 second Piola–Kirchhoff, 159
 shear, 20
 symmetry, 18
 tensor, 17
 trial, 291
 ultimate, 32

- Stress (*cont.*)
 - uniaxial, 31
 - yield, 32
 Stress rate, 315
 - Jaumann, 315
 Stress vector, 15
 Stretch tensor, 150
 Strong form, 38
 Structural energy form, 168, 185,
 - 298, 337
 - elastic, 168
 - elastoplasticity, 298
 - finite deformation, 337
 - nonlinear, 175
 St. Venant–Kirchhoff material, 163
 Surface traction, 15
- T**
 Tangential slip, 379, 380
 Tangential traction force, 387
 Tangent modulus, 243
 Tangent operator, 297, 336, 337
 - consistent, 297, 337
 - material, 336
 - spatial, 336
 Tangent stiffness matrix, 94
 Tensor, 5–7, 9, 10, 17, 32
 - Cartesian, 5
 - elasticity, 32
 - identity, 5
 - orthogonal, 9
 - skew, 6, 10
 - spin, 7
 - stress, 17
 symmetric, 6
 Tensor product, 269
 Time step, 110
 Total Lagrangian formulation, 168
 Trace, 8, 268
 Transpose, 3
 Trial function, 50
- U**
 Updated Lagrangian formulation, 174
- V**
 Variational equation, 43, 167
 Variational inequality, 383
 Vector, 3, 10
 - dual, 10
 - product, 10
 Virtual displacement, 42
 Virtual work, 391
 - contact, 391
- W**
 Weak form, 44, 115, 166, 249, 298, 383
 Work, 39
- Y**
 Yield criterion, 282
 - von Mises, 282
 Yield function, 282, 327
 Yield surface, 282
 Young's modulus, 33, 83