SPECTRAL/*HP* FINITE ELEMENT MODELS FOR FLUIDS AND STRUCTURES

A Dissertation

by

GREGORY STEVEN PAYETTE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2012

Major Subject: Mechanical Engineering

SPECTRAL/*HP* FINITE ELEMENT MODELS FOR FLUIDS AND STRUCTURES

A Dissertation

by

GREGORY STEVEN PAYETTE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	J. N. Reddy
Committee Members,	H. A. Hogan
	M. T. Schobeiri
	J. M. Roesset
Head of Department,	Jerald A. Caton

May 2012

Major Subject: Mechanical Engineering

ABSTRACT

Spectral/*hp* Finite Element Models
for Fluids and Structures. (May 2012)
Gregory Steven Payette, B.S., University of Idaho;
M.S., Texas A&M University
Chair of Advisory Committee: Dr. J. N. Reddy

We consider the application of high-order spectral/*hp* finite element technology to the numerical solution of boundary-value problems arising in the fields of fluid and solid mechanics. For many problems in these areas, high-order finite element procedures offer many theoretical and practical computational advantages over the low-order finite element technologies that have come to dominate much of the academic research and commercial software of the last several decades. Most notably, we may avoid various forms of locking which, without suitable stabilization, often plague low-order least-squares finite element models of incompressible viscous fluids as well as weak-form Galerkin finite element models of elastic and inelastic structures.

The research documented in this dissertation includes applications of spectral/*hp* finite element technology to an analysis of the roles played by the linearization and minimization operators in least-squares finite element models of nonlinear boundary-value problems, a novel least-squares finite element model of the incompressible Navier-Stokes equations with improved local mass conservation, weak-form Galerkin finite element models of viscoelastic beams and a high-order seven parameter continuum shell element for the numerical simulation of the fully geometrically nonlinear mechanical response of isotropic, laminated composite and functionally graded elastic shell structures. In addition, we also present a simple and efficient sparse global finite element coefficient matrix assembly operator that may be readily parallelized for use

on shared memory systems. We demonstrate, through the numerical simulation of carefully chosen benchmark problems, that the finite element formulations proposed in this study are efficient, reliable and insensitive to all forms of numerical locking and element geometric distortions.

To Cami, Nathan, William and James; and to my parents

ACKNOWLEDGMENTS

I wish to express my sincere thanks and heartfelt gratitude to my advisor Dr. J. N. Reddy for his unwavering support during the course of my graduate studies and research work at Texas A&M University. I will ever be indebted to him for his dedication to teaching and also for the genuine encouragement and confidence that he has always placed in me. I am especially gratefully for the flexibility he afforded me to pursue many of my own interests, in taking courses and also in conducting academic research.

I would also like to thank the members of my dissertation committee: Dr. Harry Hogan, Dr. José Roesset and Dr. Taher Schobeiri. I thoroughly enjoyed the semesters that I spent working with Dr. Hogan as a teaching assistant. I had the pleasure of taking my first graduate level course on the finite element method at Texas A&M University from Dr. Roesset. It was Dr. Schobeiri who provided me with an introduction to tensor analysis using curvilinear coordinates, a subject which plays a crucial role in both the theory and the finite element analysis of shells.

I am particularly grateful for the financial support that was provided me during my Ph.D. studies by the Oscar S. Wyatt Endowed Chair, the Air Force Office of Scientific Research through the MURI Grant FA9550-09-1-0686 and the Excellence in Engineering Graduate Fellowship awarded jointly by Sandia National Laboratories and the Dwight Look College of Engineering at Texas A&M University.

The numerical simulations presented in Chapters II, IV and VI were conducted using the computing resources of the Texas A&M Supercomputing Facility; their support is gratefully acknowledged. Special thanks are also due to Dr. Roman Arciniega, Dr. Kalyana Nakshatrala and Mr. Venkat Vallala; each for their friendship and willingness to engage in fruitful discussions on a variety of topics, including (but not

limited to): physics, mathematics, philosophy, history, politics and life.

I wish to thank my wife Cami for her love and support during my time as a student at Texas A&M University. The realization of my educational aspirations has been, to a large extent, her sacrifice and for that I am truly grateful. I must also thank our three boys: Nathan, William and James for their love and patience.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

A. Background

In the numerical simulation of a wide range of physical phenomena (mathematically described in terms of boundary or initial boundary-value problems), the finite element method has emerged as one of the most powerful tools for obtaining accurate, efficient and stable approximate solutions. Since the publication of the groundbreaking work of Turner et al. [1], the scientific literature and, *much more importantly*, practical engineering software based on the finite element method have grown at a remarkable pace, spanning many fields of engineering and applied science. At present, the finite element method is widely recognized as the premier computational procedure for the numerical simulation of solid mechanics problems. Outside the realm of the mechanics of solids, however, the method has yet to receive such a level of acceptance and prominence. This is especially noteworthy in computational fluid dynamics (CFD), a field that is presently dominated by low-order finite difference and finite volume technologies.

The genesis of most finite element models is the weak-form Galerkin formulation. It is now well-known that the success of finite element procedures, based on the Galerkin formulation, in obtaining favorable numerical solutions of boundary-value problems is closely tied to the degree to which the weak formulation coincides with an unconstrained minimization problem [2]. More generally, whenever any weak formulation (based on the Galerkin, Petrov-Galerkin, weighted residual, or least-squares methods, among others) is equivalent to the problem of minimizing an unconstrained

---

The journal model is *IEEE Transactions on Automatic Control.*

convex quadratic functional, the finite element model inherits the following highly desirable mathematical properties:

1. The numerical solution becomes an orthogonal projection of the exact solution onto the trial space of a given conforming finite element discretization. As a result, the numerical solution represents the "best approximation" of the exact solution in the trial space (as measured by a well defined *energy* norm).

2. No highly restrictive compatibility requirements (such as the discrete inf-sup condition) ever arise that must be additionally satisfied by the discrete conforming function spaces of the various dependent variables.

3. The resulting linear algebraic system of global finite element equations are always symmetric and positive-definite (a property that may be exploited by both direct as well as iterative solvers).

This *ideal* setting for finite element approximation, stemming from the unconstrained minimization of a convex quadratic functional, is sometimes termed a *variational setting*.

In retrospect, it is now clear that the finite element method emerged in perhaps the most favorable of settings; i.e., the analysis of linear elastic structural components. The method initially arose as a direct extension of the classical Ritz method [3], wherein the numerical solution is sought via a direct and discrete minimization of the total potential energy functional. The combination of the method's successful application to problems in linear elasticity along with its versatility in handling irregular domains and complex boundary conditions led researchers to extend the finite element method, in the context of the weak-form Galerkin procedure, to boundary-value problems whose weak formulations cannot be construed as global minimizers.

For many such problems it was soon discovered that many of the most attractive features of the finite element method exhibited in the solution of solid mechanics problems, were no longer present.

In recent years, there has been a large body of work attempting to recover *some* of the attractive features of the ideal *variational setting* for problems whose Galerkin based weak formulations are either estranged or completely divorced from any notion of unconstrained functional minimization. Many of the advocated procedures may be viewed as stabilized Galerkin formulations and include methods such as the SUPG [4, 5], penalty [6, 7] and Galerkin least-squares [8], among others. Unfortunately, the success of these methods is often intertwined with ad-hoc parameters that require mesh and/or solution dependent fine-tuning. Furthermore, it is worth noting that although the various stabilized Galerkin formulations can often sidestep the discrete inf-sup condition, they cannot generally inherit the best approximation property nor produce symmetric positive-definite coefficient matrices for the case when the governing equations contain non-self-adjoint operators.

In addition to the stabilized Galerkin formulations, there has also been renewed interest over the past two decades in developing finite element models for problems outside the realm of solid mechanics that recover most, *if not all*, of the attractive features of the ideal *variational setting*. One such formulation is based on the least-squares method and allows for a finite element model to be developed for any boundary-value problem in a setting of unconstrained functional minimization (see for example Refs. [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]). The least-squares method is based on the notion of *residual* minimization, wherein a least-squares functional is constructed from the sum of the squares of the norms of the partial differential equation residuals (where the norms of standard Sobolev spaces are typically employed). Such functionals are purely mathematical in nature and do not have the meaning of

energy of a system. The weak form is obtained via a direct minimization of the least-squares functional. The finite element model is then obtained in the usual way, and inherits the desirable properties discussed previously for the ideal *variational setting*.

B. Motivation for the present study

In previous work concerned with developing effective finite element models for structures and fluids, predominantly low-order polynomial finite element procedures have been adopted, primarily through the use of the weak-form Galerkin formulation. As discussed previously, the Galerkin procedure is typically sufficient to achieve a favorable setting for the numerical simulation of deformable solids. For viscous fluids on the other hand, we find the least-squares method to be better suited for attaining a reliable computational environment for finite element approximations. Throughout this work, we further advocate the use of high-order polynomial approximations to improve the discrete setting for various formulations for fluids and structures.

To motivate the need for polynomial refinement (or p-refinement), we recall that although unconstrained minimization principles offer a highly attractive setting for finite element approximation, adequate solution convergence properties under h-refinement alone cannot always be realized. We recall that for weak forms resulting from the unconstrained minimization of a quadratic functional, error estimates of the following type can often be established for a given conforming finite element approximation [19]

$$||\mathbf{u} - \mathbf{u}_{hp}||_{\Omega,s} \leq Ch^{p+1-s}, \quad s = 0, 1 \tag{1.1}$$

In the above expression $\mathbf{u}$ is the analytical solution, $\mathbf{u}_{hp}$ is the finite element solution and C is a constant. The quantity $||\cdot||_{\Omega,s}$ is the norm associated with the Sobolev space $\mathbf{H}^s(\Omega)$ and Ω is the domain on which the problem is posed (see Chapter II

for details). The quantity h is a measure of the average element size in $\bar{\Omega}$ and the symbol p is the polynomial order of the finite element approximation within a given element. Under *ideal* conditions the constant C will depend on $\mathbf{u}$, Ω and the material properties of the given boundary-value problem; and is therefore, independent of h.

Unfortunately, h-refinement alone does not always constitute an effective means of improving the finite element solution. For example, in the finite element approximation of the mechanical response of structural components (such as beams, plates and shells) the constant C becomes adversely large in the limit as the thickness tends to zero. Furthermore, equal low-order interpolation of the dependent variables inevitably leads to various forms of numerical locking that cannot be directly overcome without the use of severe mesh refinement. To overcome such deficiencies, most researchers employ *stabilized* low-order finite element technology using either: (a) a displacement-based formulation with selective reduced integration or (b) a mixed variational formulation based on the Hu-Washizu principle (e.g., the assumed strain and enhanced strain procedures). It is worth noting that low-order stabilization procedures often necessitate additional ad-hoc fixes such as hour-glass control.

Another important example where h-refinement yields a non-optimal computational procedure arises in least-squares finite element models of the Navier-Stokes equations governing flows of incompressible fluids. Out of practicality, the majority of such finite element models are constructed from least-squares functionals whose energy norms are not $\mathbf{H}^s(\Omega)$-norm equivalent (i.e., are non-$\mathbf{H}^s(\Omega)$-coercive). For these finite element models, the constant C either depends on the mesh parameter h and/or little may be inferred directly from Eq. (1.1). To improve the performance of low-order least-squares finite element formulations, ad-hoc reduced integration and/or collocation procedures have often been adopted.

The present study is motivated by the observation that many of the deficiencies

encountered in finite element models constructed from unconstrained minimization principles may be largely circumvented or avoided entirely whenever a sufficiently adequate polynomial order p is employed in constructing the finite element approximation $\mathbf{u}_{hp}$ within each element. In particular, whenever an appropriate level of p-refinement is utilized, efficient finite element procedures are obtained which do not require any of the sophisticated ad-hoc tricks that are so often required to improve the numerical solutions associated with low-order finite element formulations. As a result, we are free in the numerical implementation to employ full integration and allow the high-order finite element function spaces to naturally avoid any inconsistencies found in low-order approximations that otherwise result in locking.

C. Scope of the research

The research began at Texas A&M University in the Fall of 2007 and is largely concerned with developing efficient finite element models for fluids and structures based on high-order spectral/hp finite element technology. The research encompasses an analysis of the least-squares method as applied in the finite element solution of nonlinear boundary-value problems [20], a novel least-squares formulation of the steady and non-stationary incompressible Navier-Stokes equations with enhanced local mass conservation and weak-form Galerkin finite element models of viscoelastic beams based on the Euler-Bernoulli, Timoshenko and third-order Reddy beam theories [21]. In addition, we also present a general shell element for the numerical simulation of the finite deformation of isotropic, laminated composite and functionally graded elastic shell structures. Our aim throughout this research has been to apply novel mathematical models and numerical solution strategies to a variety of problem sets in continuum mechanics, wherein the additional benefits obtained from employing high-

order spectral/*hp* finite element technology are substantial.

The dissertation is organized as follows. In Chapter II we present an overview of the steps involved in developing and arriving at finite element models using high-order spectral/*hp* finite element technology. We also document highly practical strategies, developed during the course of the present research, for implementing high-order finite element procedures in parallel computing environments using the OpenMP paradigm. Of significant importance is a discussion on a simple and efficient shared memory based sparse global coefficient matrix assembly operator (an algorithm which has been implemented numerically in C++ and successfully utilized on practical finite element problems containing as many as half a million degrees of freedom).

Chapters III and IV are concerned with least-squares finite element models of nonlinear boundary-value problems with specific applications to viscous incompressible fluid flows. In Chapter III we provide a critical examination of the consequences associated with exchanging the order of application of the minimization and linearization operators in least-squares finite element formulations of nonlinear boundary-values problems. In our analysis, we consider the abstract setting for an L_2-norm least-squares formulation of an abstract nonlinear boundary-value problem. We further provide a thorough discussion of possible forms taken by the *linearized* least-squares weak formulation, when linearization is either performed before or after minimization of the least-squares functional in the context of both the Picard and Newton linearization procedures. We show both mathematically and also by way of numerical experiments that although the least-squares principle suggests that minimization ought to be performed prior to linearization, such an approach is often impractical and not necessary. In Chapter IV we present a novel least-squares finite element formulation for both the steady and non-stationary incompressible Navier-Stokes equations based on the standard velocity-pressure-vorticity first-order system, but with

enhanced element-level mass conservation. The proposed formulation comes with little additional computational cost (as compared to the standard velocity-pressure-vorticity least-squares formulation) and does not compromise the unconstrained minimization setting that is so attractive in least-squares finite element models. We showcase the performance of the proposed least-squares formulation (in improving local mass conservation) through the numerical simulation of a variety of important steady-state and non-stationary fluid flow problems.

In Chapters V and VI we consider applications of spectral/hp finite element technology to problems in solid mechanics, namely, viscoelastic beams and elastic shells. In Chapter V, we present efficient finite element models for initially straight viscoelastic beam structures subjected to loading conditions that induce large displacements, moderate rotations and small strains. The finite element models are constructed using the kinematic assumptions of the Euler-Bernoulli, Timoshenko and third-order Reddy beam theories. The viscoelastic constitutive equations are efficiently discretized in time using the trapezoidal rule in conjunction with a two-point recurrence formula. The resulting finite element models are shown to be void of both membrane and shear locking. In Chapter VI we propose a general high-order continuum shell finite element for use in the analysis of the fully geometrically nonlinear mechanical response of thin and thick isotropic, laminated composite and functionally graded elastic shell structures. The shell formulation is based on a 7-parameter expansion of the displacement field; thereby allowing for the use of fully three-dimensional constitutive equations while avoiding the need for a rotation tensor in the kinematical description. The shell element is shown, through the numerical simulation of carefully chosen benchmark problems, to be insensitive to all forms of numerical locking and severe geometric distortions. Finally, in Chapter VII we provide concluding remarks and offer suggestions for future research directions.

CHAPTER II

NUMERICAL IMPLEMENTATION OF HIGH-ORDER SPECTRAL/*HP* FINITE ELEMENT PROCEDURES

In this chapter, we present a general overview of fundamental steps involved in developing and arriving at finite element models of boundary-value problems using high-order spectral/*hp* finite element technology. We also document highly practical strategies, many of which were developed during the course of this study, for implementing high-order finite element procedures for moderately large sparse finite element systems on shared-memory based parallel computing architectures.

The chapter is organized as follows. We begin by providing an overview of some of the basic notation and standard terminology that is employed throughout this dissertation. We then review the standard one-dimensional C^0 spectral nodal basis functions that we utilize to develop high-order finite element interpolation functions for multi-dimensional spectral/*hp* finite elements. Since high-order finite element procedures necessitate high-order quadrature rules, we also review basic formulas needed to determine the points and weights of the Gauss-Legendre quadrature rule (for the general case where an arbitrary number of quadrature points are desired).

We also discuss in this chapter efficient algorithms for implementing high-order finite element technology in parallel on shared-memory systems. Most notably, we present a global finite element assembly operator that may be readily parallelized using the OpenMP paradigm. The set of algorithms constituting the global assembly operator were developed during the course of the present research and have been successfully implemented using the C++ programming language. The assembly operator efficiently constructs a sparse representation of the global finite element coefficient matrix using a *compressed row* (or *compressed column*) storage format. As a result,

the operator is applicable to finite element equations consisting of well over 100,000 degrees of freedom and may be used in conjunction with any number of modern sparse equation solver libraries (e.g., UMFPACK, PARDISO, MUMPS, etc.). We improve system memory requirements in the numerical implementation of high-order spectral/*hp* finite element technology by adopting element-level static condensation [22], wherein the interior degrees of freedom of each element are implicitly eliminated prior to invoking the global assembly operator. Finally, we showcase the performance of the high-order finite element procedures discussed throughout this chapter through the numerical simulation of an example problem possessing roughly half a million total degrees of freedom.

A. The abstract finite element problem

1. Notation

Before beginning our discussion on high-order spectral/*hp* finite element procedures and their efficient numerical implementation, we find it prudent to introduce some standard notation that will be used throughout this dissertation. We assume that Ω is an open bounded subset of $\mathbb{R}^{nd}$, where nd denotes the number of spatial dimensions. The boundary of Ω is denoted by $\Gamma = \partial\Omega = \bar{\Omega} - \Omega$, where $\bar{\Omega}$ represents the closure of Ω. A typical point belonging to $\bar{\Omega}$ is denoted as $\mathbf{x}$. We employ the customary designations for the Sobolev spaces $H^s(\Omega)$ and $H^s(\Gamma)$ where $s \geqslant 0$ [23]. The corresponding norms are given as $||\cdot||_{\Omega,s}$ and $||\cdot||_{\Gamma,s}$. Likewise the inner products associated with these spaces are denoted as $(\,\cdot\,,\,\cdot\,)_{\Omega,s}$ and $(\,\cdot\,,\,\cdot\,)_{\Gamma,s}$ respectively. The product spaces $\mathbf{H}^s(\Omega) = [H^s(\Omega)]^{nd}$ are constructed in the usual way.

Throughout this study we favor the so-called "Gibbs notation" for tensor analysis as opposed to the "Ricci notation" which is popular in the continuum mechanics

community. As a result, the tensor product of vectors $\mathbf{u}$ and $\mathbf{v}$ is given as $\mathbf{uv}$ as opposed to $\mathbf{u} \otimes \mathbf{v}$. Likewise, the gradient of vector $\mathbf{u}$ is represented with respect to an orthogonal Cartesian coordinate system as $\nabla\mathbf{u} = (\partial u_j/\partial x_i)\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j$ rather than $\nabla\mathbf{u} = (\partial u_i/\partial x_j)\hat{\mathbf{e}}_i \otimes \hat{\mathbf{e}}_j$. The former expression follows naturally whenever ∇ is defined as a vector differential operator of the form $\nabla \equiv \hat{\mathbf{e}}_i\partial/\partial x_i$.

2. Weak formulations

The classical form of a typical boundary or initial boundary-value problem is not well suited for numerical approximation via the finite element method. Instead, a given boundary or initial boundary-value problem must be first recast into the form of a generalized variational boundary-value problem, also know as a weak formulation, prior to numerical discretization using the finite element method. In the present work we construct weak formulations of various boundary and initial value problems based upon the classical weak-form Galerkin formulation as well as through the use of least-squares variational principles. Weak formulations typically involve integral statements over Ω and Γ that are in a *generalized sense* equivalent to the original set of partial differential equations and natural boundary conditions. In general, a weak formulation (based on either the weak-form Galerkin or least-squares models) of a general boundary-value problem may be stated as follows: find $\mathbf{u} \in \mathcal{V}$ such that

$$\mathcal{B}(\mathbf{w}, \mathbf{u}) = \mathcal{F}(\mathbf{w}) \quad \forall\ \mathbf{w} \in \mathcal{W} \tag{2.1}$$

where $\mathcal{B}(\mathbf{w}, \mathbf{u})$ is a bilinear form, $\mathcal{F}(\mathbf{w})$ is a linear form and $\mathcal{V}$ and $\mathcal{W}$ are function spaces (e.g., appropriate subsets of the Sobolev space $\mathbf{H}^1(\Omega)$). The quantity $\mathbf{u}$ represents the set of dependent variables (associated with the variational boundary-value problem) and $\mathbf{w}$ represents the corresponding weighting or test function. Unlike classical solutions that are defined unambiguously point-wise, weak solutions exist with

respect to test functions and are therefore understood in the context of distributions. As a result, the weak solution (and its derivatives) is typically defined unambiguously in Ω up to a set of measure zero. We note that Eq. (2.1) is not limited to the analysis of linear problems only, but is also applicable to nonlinear generalized boundary-value problems that have been linearized in the context of an appropriate iterative solution procedure (e.g., a fixed point iteration scheme such as the methods of Picard or Newton).

3. High-order spectral/*hp* finite element models

We now proceed to describe the high-order spectral/*hp* finite element technology that is employed throughout the present research. To this end we note that the finite element model associated with Eq. (2.1) is obtained by restricting the solution space to a finite dimensional sub-space $\mathcal{V}^{hp}$ of the infinite dimensional function space $\mathcal{V}$ and the weighting function to a finite dimensional sub-space $\mathcal{W}^{hp} \subset \mathcal{W}$. As a result, in the discrete case we seek to find $\mathbf{u}_{hp} \in \mathcal{V}^{hp}$ such that

$$\mathcal{B}(\mathbf{w}_{hp}, \mathbf{u}_{hp}) = \mathcal{F}(\mathbf{w}_{hp}) \quad \forall\ \mathbf{w}_{hp} \in \mathcal{W}^{hp} \tag{2.2}$$

We assume that the domain $\bar{\Omega} \subset \mathbb{R}^{nd}$ is discretized into a set of NE non-overlapping sub-domains $\bar{\Omega}^e$, called finite elements, such that $\bar{\Omega} \approx \bar{\Omega}^{hp} = \bigcup_{e=1}^{\mathrm{NE}} \bar{\Omega}^e$. The geometry of each element is characterized using the standard isoparametric bijective mapping from the master element $\hat{\Omega}^e$ to the physical element $\bar{\Omega}^e$. In the present study we restrict the classes of elements considered to lines in $\mathbb{R}^1$, four sided quadrilaterals in $\mathbb{R}^2$ and six faced bricks in $\mathbb{R}^3$ (although numerical results are presented for $nd = 1$ and 2 only). As a result we can simply define the geometry of the master element as $\hat{\Omega}^e = [-1, +1]^{nd}$. The natural coordinates associated with $\hat{\Omega}^e$ (when $nd = 3$) are defined as $\boldsymbol{\xi} = (\xi^1, \xi^2, \xi^3) = (\xi, \eta, \zeta)$ (and may be truncated appropriately whenever

$nd < 3$). We note in passing that the continuum shell element presented in Chapter VI is obtained by mapping the master element $\hat{\Omega}^e = [-1, +1]^2$ onto a two-dimensional manifold $\bar{\Omega}^e$ constituting the approximate mid-surface of the eth element. As a result, the finite element approximation of the shell mid-plane will generally consist of a curved two-dimensional surface imbedded in three-dimensional space.

In this work we employ a family of finite elements constructed using high polynomial order interpolation functions. The quantity h in the definition of the sub-spaces $\mathcal{V}^{hp}$ and $\mathcal{W}^{hp}$ represents the average size of all the elements in a given finite element discretization. Likewise, the symbol p denotes the polynomial degree (or p-level) of the finite element interpolation functions associated with each element in the model. As a result, the discrete solution may be refined by either increasing the number of elements (i.e., reducing h) in $\bar{\Omega}^{hp}$ (h-refinement), increasing the polynomial order of the approximate solution within each element $\bar{\Omega}^e$ (p-refinement) or through an appropriate and systematic combination of both h-refinement and p-refinement.

Within a typical finite element $\bar{\Omega}^e$, the set of dependent variables $\mathbf{u}$ is approximated using the following general interpolation formula

$$\mathbf{u}(\mathbf{x}) \approx \mathbf{u}_{hp}(\mathbf{x}) = \sum_{i=1}^{n} \boldsymbol{\Delta}_i^e \psi_i(\boldsymbol{\xi}) \qquad \text{in } \hat{\Omega}^e \tag{2.3}$$

where $\psi_i(\boldsymbol{\xi})$ are the nd-dimensional Lagrange interpolation functions, $\boldsymbol{\Delta}_i^e$ is an array containing the value of $\mathbf{u}_{hp}(\mathbf{x})$ at the location of the ith node in $\bar{\Omega}^e$ and $n = (p+1)^{nd}$ is the number of nodes in $\bar{\Omega}^e$. The above definition is unambiguous due to the employment of the standard isoparametric mapping $\hat{\Omega}^e \rightleftarrows \bar{\Omega}^e$ (used in the characterization of the geometry of each element). Note that $\{\boldsymbol{\Delta}_i^e\}_{i=1}^n$ constitutes a set of n arrays for the eth element, where the size of each array is equal to the total number of variables comprising $\mathbf{u}$. In the current research, all interpolants appearing in Eq. (2.3) are of polynomial order p, and are hence non-hierarchal.

There are a variety of ways in which high-order *nd*-dimensional interpolation functions may be formulated. For our analysis we construct these polynomial functions from tensor products of the one-dimensional C^0 spectral nodal interpolation functions

$$\varphi_j(\xi) = \frac{(\xi-1)(\xi+1)L'_p(\xi)}{p(p+1)L_p(\xi_j)(\xi-\xi_j)} \qquad \text{in } [-1,+1] \tag{2.4}$$

where $L_p(\xi)$ is the Legendre polynomial of order p and $L'_p(\xi)$ represents the derivative of $L_p(\xi)$ with respect to ξ. The quantities ξ_j represent the locations of the nodes associated with the one-dimensional interpolants (with respect to the natural coordinate ξ). The one-dimensional nodal points are defined as the roots of the following expression

$$(\xi-1)(\xi+1)L'_p(\xi) = 0 \qquad \text{in } [-1,+1] \tag{2.5}$$

The nodal points $\{\xi_j\}_{j=1}^{p+1}$ found in solving Eq. (2.5) are known as the Gauss-Lobatto-Legendre (GLL) points. Whenever $p \leq 2$, the GLL points are equally spaced within the standard interval $[-1,+1]$. When $p > 2$ the GLL points are distributed unequally with discernable bias given to the end points of the interval. The bias associated with the spacing of the GLL points increases with p. In Figure 1 we plot the high-order interpolation functions $\{\varphi_j\}_{j=1}^{p+1}$ generated for the case where $p = 6$. In this figure we show the interpolation functions associated with both an equal as well as a GLL spacing of the nodal points in the standard bi-unit interval. The interpolation functions constructed using equal nodal spacing clear exhibit oscillations (often termed the *Runge effect*) near the end points of the standard interval. These oscillations become more pronounced as the p-level is increased. The spectral interpolation functions, on the other hand are free of the *Runge effect*. Finite element coefficient matrices constructed using spectral interpolation functions are as a result better conditioned than matrices formulated using elements with equally spaced nodes.

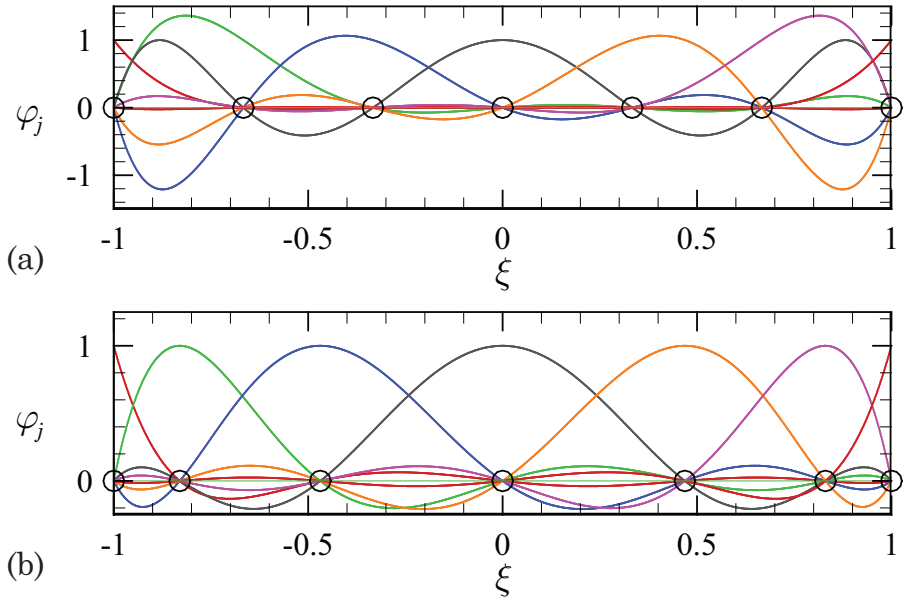


Fig. 1. High polynomial order one-dimensional C^0 Lagrange interpolation functions. Cases shown are for $p = 6$ with: (a) equal spacing of the element nodes and (b) unequal nodal spacing associated with GLL points.

It is worthwhile to note that the spectral nodal basis functions $\{\varphi_j\}_{j=1}^{p+1}$ may be viewed as standard Lagrange interpolation functions, with the locations of the unequally spaced nodal points given in terms of the roots of Eq. (2.5). As a result, it is possible to write the spectral interpolants of order p using the following classical formula for Lagrange polynomials

$$\varphi_j(\xi) = \prod_{i=1, i\neq j}^{p+1} \frac{\xi - \xi_i}{\xi_j - \xi_i} \tag{2.6}$$

Although less elegant than Eq. (2.4), the above expression is better suited for numerical implementation in a general purpose finite element program. Furthermore, the above equation may also be easily utilized to produce a simple formula for calculating derivatives of the one-dimensional spectral interpolation functions.

In order to generate the spectral-interpolation functions, it is necessary to be able to evaluate high-order Legendre polynomials of arbitrary orders. For completeness we recall that the lowest order Legendre polynomials are of the form $L_0(\xi) = 1$ and $L_1(\xi) = \xi$. All subsequent Legendre polynomials may be determined through the use of the following well-known three-point recurrence formula

$$L_{p+1}(\xi) = \left[(2p+1)\xi L_p(\xi) - pL_{p-1}(\xi)\right]/(p+1) \tag{2.7}$$

We also have the following useful expression for calculating the first derivative of the Legendre polynomials

$$\frac{(\xi-1)(\xi+1)}{p} L'_p(\xi) = \xi L_p(\xi) - L_{p-1}(\xi) \tag{2.8}$$

The multi-dimensional high-order interpolation functions $\psi_i(\boldsymbol{\xi})$ may be constructed by taking simple tensor products of the one-dimensional spectral interpolants. For example, in two-dimensions, the high-order interpolation functions may be defined as

$$\psi_i(\xi,\eta) = \varphi_j(\xi)\varphi_k(\eta) \qquad \text{in } \hat{\Omega}^e = [-1,+1]^2 \tag{2.9}$$

where $i = j + (k-1)(p+1)$ and $j,k = 1,\ldots,p+1$. A variety of high-order two-dimensional master elements are depicted in Figure 2. In this study we restrict our analysis to problems that may be solved using either one or two-dimensional master elements. For the sake of completeness, however, we note in passing that in three-dimensions, the high-order interpolants can be expressed as

$$\psi_i(\xi,\eta,\zeta) = \varphi_j(\xi)\varphi_k(\eta)\varphi_l(\zeta) \qquad \text{in } \hat{\Omega}^e = [-1,+1]^3 \tag{2.10}$$

where $i = j + [k-1+(l-1)(p+1)](p+1)$ and $j,k,l = 1,\ldots,p+1$.

Finite elements whose interpolation functions are constructed in terms of tensor

products of $\varphi_j(\xi)$ are commonly referred to as spectral elements in the literature [22]. Such elements are merely standard high-order Lagrange type finite elements, where the locations of the unequally spaced nodes in $\hat{\Omega}^e$ are taken as tensor products of the roots of Eq. (2.5).

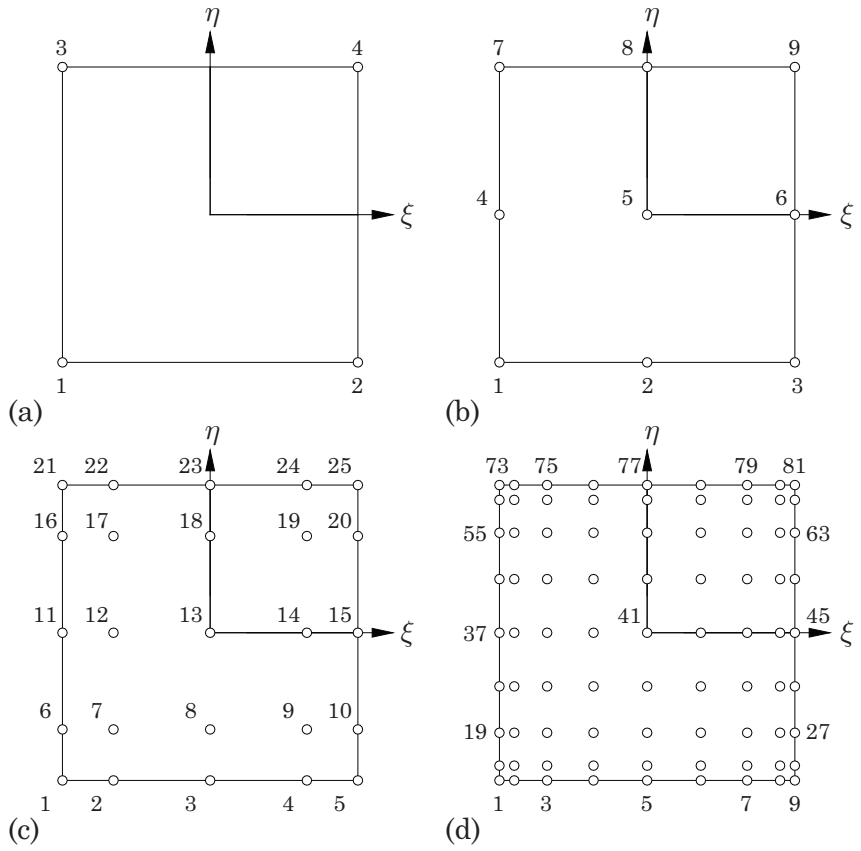


Fig. 2. Examples of various high polynomial order spectral/hp quadrilateral master elements $\hat{\Omega}^e$: (a) a 4 noded element, $p = 1$ (b) a 9 noded element, $p = 2$ (c) a 25 noded element, $p = 4$ and (d) an 81 noded element, $p = 8$.

The finite element method naturally leads to a set of linear algebraic equations for each element associated with a given finite element discretization. Substitution of Eq. (2.3) as well as an appropriate discrete test function $\mathbf{w}_{hp}$ into Eq. (2.2) yields

the following set of equations for the eth element of the finite element model

$$[K^e]\{\Delta^e\} = \{F^e\} \tag{2.11}$$

In the above expression $[K^e]$ is the element coefficient matrix, $\{\Delta^e\}$ is a vector containing the essential variables associated with each node of the element and $\{F^e\}$ is the element force vector. The element coefficient matrix and force vector are obtained respectively by restricting evaluation of the bilinear form $\mathcal{B}(\mathbf{w}_{hp}, \mathbf{u}_{hp})$ and linear form $\mathcal{F}(\mathbf{w}_{hp})$ to the domain $\bar{\Omega}^e$.

In this work we utilize the standard Gauss-Legendre quadrature rules in the numerical integration of all terms appearing in the element coefficient matrix and force vector. Unless explicitly stated otherwise, we employ *full integration* of all integrals and *do not* resort to selective under-integration of *any* terms in the coefficient matrix or force vector. Numerical results are typically obtained using a quadrature rule of at least NGP $= p + 1$, where NGP represents the number of quadrature points in the direction of a given natural coordinate associated with $\hat{\Omega}^e$. Since high-order methods require the use of high-order quadrature rules, we note that the Gauss-Legendre quadrature points are obtained as the roots of the Legendre polynomial of order NGP. The Gauss-Legendre quadrature weights may be obtained from the following expression

$$w_i = \frac{2}{(1-\xi_i^2)L'_{\mathrm{NGP}}(\xi_i)^2} \tag{2.12}$$

where $\{\xi_i\}_{i=1}^{\mathrm{NGP}}$ are the quadrature points (which are distinct from and should not be confused with the GLL points). The Gauss-Legendre quadrature points and weights as well as the GLL points may be accurately determined within a user pre-defined numerical tolerance through the use of a symbolic algebra package such as Maple.

The set of equations for a given finite element discretization is obtained by com-

bining the equations associated with each element into the following global system of linear algebraic equations

$$[K]\{\Delta\} = \{F\} \tag{2.13}$$

where

$$[K] = \mathop{\mathbf{A}}_{e=1}^{\mathrm{NE}} [K^e], \qquad \{F\} = \mathop{\mathbf{A}}_{e=1}^{\mathrm{NE}} \{F^e\} \tag{2.14}$$

In the above expressions, $\mathbf{A}$ is a symbolic representation of the global finite element assembly operator. Efficient, shared-memory based parallel algorithms for the global assembly operator $\mathbf{A}$ will be discussed in the subsequent section for the case where the global coefficient matrix $[K]$ is sparse (i.e., populated primarily by zeros).

For additional details on the computer implementation of the finite element method, including descriptions of the bijective isoparametric mapping $\hat{\Omega}^e \rightleftarrows \bar{\Omega}^e$ and the global assembly operator $\mathbf{A}$ (for full and banded matrices), we refer to the books of Reddy [24] and Bathe [25]. For further details on construction of the spectral interpolation functions, we refer to the book by Karniadakis and Sherwin [22].

B. Shared-memory based parallel implementation of high-order finite element procedures

Having established the general high-order finite element technology that will be used throughout this work, we turn our attention to efficient numerical implementation strategies that may be adopted in a general finite element framework. In particular we will focus our discussion on numerical implementation techniques that may be readily incorporated in a parallel computing environment based on the OpenMP paradigm. OpenMP is an Application Programming Interface (API) that supports multithreading on computer architectures that admit shared-memory multiprocessing. This form of parallelization may be employed on a standard desktop (possessing

multiple cores) or on a single node of a supercomputer. Unlike the more general Message Passing Interface (MPI), the use of OpenMP is restricted to programs involving tasks that may be accomplished by a set of processors which all have access to the same pool of shared memory.

The purpose of the current discussion is to present simple strategies, developed mostly during the course of this research, for adapting serial finite element code for efficient parallel execution on shared-memory systems. Paramount to this process is the ability to assemble the global sparse coefficient matrix in a manner that is fast, memory efficient and in a form that is appropriate for linkage with modern sparse solver libraries. We will illustrate what we feel are the key concepts in the context of a one-dimensional steady-state heat transfer problem. Although deceptively simple, the fundamental ideas for parallelization introduced through this problem may be readily utilized in the analysis of a much larger class of problems posed in multiple dimensions and solved using high-order finite element technology. The scope of our discussion will be limited to presenting key concepts, and we will therefore refrain from reviewing the various OpenMP pre-compiler directives that are specific to a given programming language. We close this section by commenting on general element-level operations, such as static condensation, that may be readily adopted to enhance the performance of high-order finite element procedures.

1. A one-dimensional example problem

a. Problem description

In this example (adapted from Reddy [24]), we consider the one-dimensional steady-state transfer of heat through a wall composed of three separate constituents. The governing equation for the temperature field $T(x)$ (based on Fourier's law of heat

conduction) may be expressed as

$$-\frac{d}{dx}\left(k\frac{dT}{dx}\right) = 0 \qquad \text{in } \Omega = (0, L) \tag{2.15}$$

where k is the thermal conductivity and $L = 8.5$ cm is the length of the domain. The boundary conditions for the problem are defined as

$$T(0) = T_0, \qquad -k\frac{dT}{dx}\bigg|_{x=L} = \beta[T(L) - T_\infty] \tag{2.16}$$

where β is the convective heat transfer coefficient and T_∞ is the far-field temperature of the air on the right hand side of the domain. An illustration of the simple heat transfer problem is provided in Figure 3 (a). This figure also provides numerical values used for the problem including the lengths h_i and thermal conductivities k_i of each composite layer.

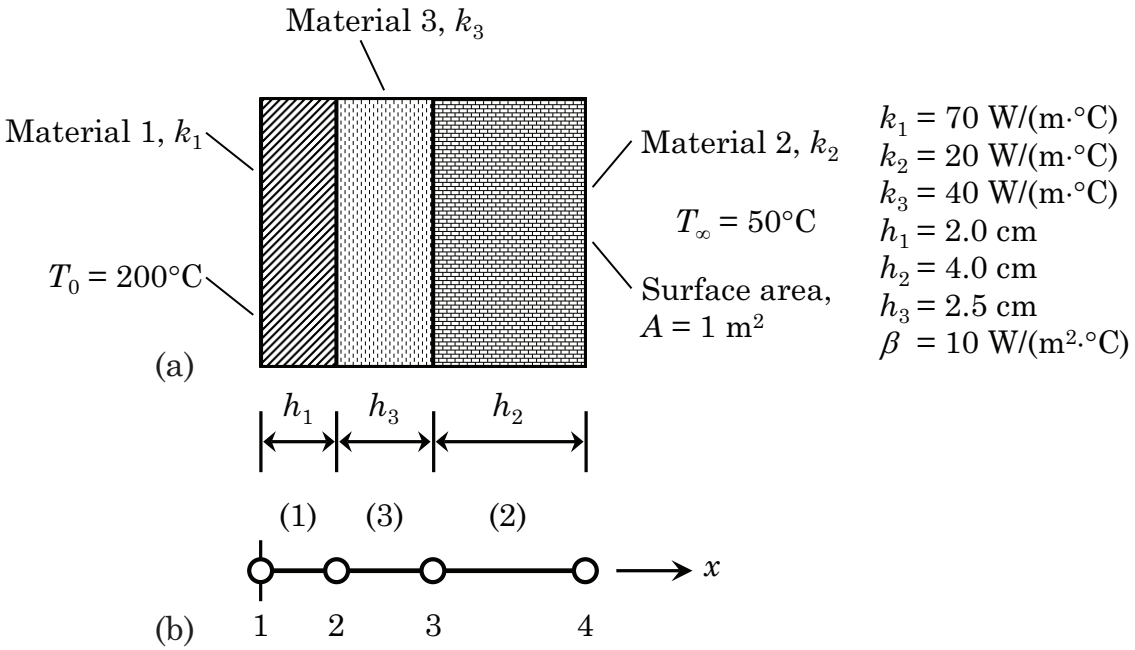


Fig. 3. A one-dimensional heat transfer problem: (a) problem description and (b) finite element discretization of $\bar{\Omega}$ using three linear finite elements.

b. Variational form of the problem

The variational form of the above boundary-value problem, based on the weak-form Galerkin formulation, may be stated as follows: find $T \in \mathcal{V}$ such that

$$\mathcal{B}(w,T) = \mathcal{F}(w) \qquad \forall\ w \in \mathcal{W} \tag{2.17}$$

where the bilinear form $\mathcal{B}(w,T)$ and linear functional $\mathcal{F}(w)$ are given as

$$\mathcal{B}(w,T) = \int_\Omega k\frac{dw}{dx}\frac{dT}{dx}dx + \beta w(L)T(L) \tag{2.18a}$$

$$\mathcal{F}(w) = \beta T_\infty w(L) \tag{2.18b}$$

The function spaces $\mathcal{V}$ and $\mathcal{W}$ associated with the weak-form Galerkin formulation of the problem are of the form

$$\mathcal{V} := \left\{T :\ T \in H^1(\Omega),\ T(0) = T_0\right\} \tag{2.19a}$$

$$\mathcal{W} := \left\{w :\ w \in H^1(\Omega),\ w(0) = 0\right\} \tag{2.19b}$$

c. Discrete element-level finite element equations

We represent the computational domain using a finite element mesh consisting of three linear elements (i.e., $p = 1$) as shown in Figure 3 (b). The nodes are numbered consecutively from left to right. For reasons which will become apparent later, we choose to number the elements in a less structured fashion. The element connectivity array denoted by ECON for the problem is of the form

$$\mathsf{ECON} = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 4 & 3 \end{bmatrix}^{\mathrm{T}} \tag{2.20}$$

The finite element coefficient matrices and force vectors may be determined using the following formulas

$$[K^e] = \frac{k_e}{h_e}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \beta \begin{bmatrix} 0 & 0 \\ 0 & \alpha_e \end{bmatrix}, \qquad \{F^e\} = \beta T_\infty \begin{Bmatrix} 0 \\ \alpha_e \end{Bmatrix} \tag{2.21}$$

where $\alpha_1 = \alpha_3 = 0$ and $\alpha_2 = 1$. As a result, the finite element matrices and vectors for each element may be expressed as

$$[K^1] = \begin{bmatrix} 3{,}500 & -3{,}500 \\ -3{,}500 & 3{,}500 \end{bmatrix}, \qquad \{F^1\} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.22a}$$

$$[K^2] = \begin{bmatrix} 500 & -500 \\ -500 & 510 \end{bmatrix}, \qquad \{F^2\} = \begin{Bmatrix} 0 \\ 500 \end{Bmatrix} \tag{2.22b}$$

$$[K^3] = \begin{bmatrix} 1{,}600 & -1{,}600 \\ -1{,}600 & 1{,}600 \end{bmatrix}, \qquad \{F^3\} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.22c}$$

d. Element-level specification of essential boundary conditions

At this point it is customary to construct the finite element equations for the system using the global finite element assembly operator **A**. Following global assembly, it is conventional to then modify the system of equations to account for the essential boundary condition associated with node 1. For large sparse systems of finite element equations, however, we find that such an approach is not attractive as it requires searches and sorts that can greatly reduce performance. An alternative procedure that is computationally efficient is to apply the essential boundary conditions at the element level, prior to global assembly. Such an approach yields the following modified

coefficient matrix and force vector for element 1

$$[K^1] = \begin{bmatrix} 3{,}500 & 0 \\ 0 & 3{,}500 \end{bmatrix}, \qquad \{F^1\} = \begin{Bmatrix} 700{,}000 \\ 700{,}000 \end{Bmatrix} \tag{2.23}$$

Note that the essential boundary condition (for local node 1 of element 1) has been applied in a manner that both preserves symmetry and conditioning of the element coefficient matrix. We refer to Reddy [24] for details on maintaining symmetry when applying essential boundary conditions.

Application of the essential boundary conditions at the element level may be facilitated in a general finite element program, through the creation of the following one-dimensional arrays during the pre-processing stage of the finite element simulation

$$\mathsf{BC\_p} = (1, 2, 2, 2), \qquad \mathsf{BC\_n} = (1), \qquad \mathsf{BC\_v} = (200) \tag{2.24}$$

$\mathsf{BC\_p}$ may be viewed as an array of integers used in accessing the components of arrays $\mathsf{BC\_n}$ and $\mathsf{BC\_v}$. In general, the size of $\mathsf{BC\_p}$ is $\mathrm{NE}+1$. Likewise, the integer array $\mathsf{BC\_n}$ and double array $\mathsf{BC\_v}$ are each of length $\mathsf{BC\_p}(\mathrm{NE}+1)-1$ (which is the total number of element-level essential boundary conditions). The arrays $\mathsf{BC\_n}$ and $\mathsf{BC\_v}$ contain the local node numbers and numerical values of the essential boundary conditions. By local node numbers, we mean the node numbering associated with the master element $\hat{\Omega}^e$ (i.e., $i = 1, \ldots, (p+1)^{nd}$), as opposed to the global node numbering associated with the physical element $\bar{\Omega}^e$. The $\mathsf{BC\_p}$, $\mathsf{BC\_n}$ and $\mathsf{BC\_v}$ arrays are used as follows: provided that $\mathsf{BC\_p}(e+1) - \mathsf{BC\_p}(e) > 0$, the local node numbers associated with the element-level boundary conditions for element e are stored in $\mathsf{BC\_n}(\mathsf{BC\_p}(e), \ldots, \mathsf{BC\_p}(e+1)-1)$. Likewise, the corresponding numerical values are stored in $\mathsf{BC\_v}(\mathsf{BC\_p}(e), \ldots, \mathsf{BC\_p}(e+1)-1)$.

In the present example problem the $\mathsf{BC\_p}$, $\mathsf{BC\_n}$ and $\mathsf{BC\_v}$ arrays are somewhat

trivial as there is only one element-level essential boundary condition. Had we replaced the convection boundary condition on the right hand side of the domain with the strong boundary condition $T(L) = T_\infty$, then the element-level essential boundary condition arrays would have been of the form

$$\mathsf{BC\_p} = (1, 2, 3, 3), \qquad \mathsf{BC\_n} = (1, 2), \qquad \mathsf{BC\_v} = (200, 50) \tag{2.25}$$

In general, the BC_p, BC_n and BC_v arrays may be constructed for any finite element discretization based solely on NE, p, ECON and the global essential boundary condition data. Application of essential boundary conditions at the element level for multi-dimensional problems involving multiple degrees of freedom per node is achieved in a manner that is largely analogous to the procedures outlined in the present one-dimensional case study. The major difference encountered in higher dimensions is the need to sometimes apply the same boundary condition multiple times (since a given boundary node will often be shared by neighboring elements).

e. Sparse construction of global coefficient matrix

Prior to global assembly, the element-level equations for a particular finite element are completely independent of the equations associated with any other element. As a result, the element-level operations of constructing and applying boundary conditions to $[K^e]$ and $\{F^e\}$, may be readily performed in a parallel computing environment. Parallel construction of the global finite element system from the element-level equations in a manner that is both fast and memory efficient is a far less trivial task. The purpose of this section, therefore, is to present strategies developed during the course of this research for efficient construction of the global sparse system of equations in a manner that can be readily accomplished in parallel. To motivate our discussion, we present in BOX 1 an overview of the primary steps involved in our parallel shared-

BOX 1. Processing stage of a low-order finite element simulation.

1. Loop over all finite elements: $e = 1, \mathrm{NE}$ (parallel)
   - Build element coefficient matrix $[K^e]$ and force vector $\{F^e\}$
   - Apply essential boundary conditions to $[K^e]$ and $\{F^e\}$
   - Add components of $[K^e]$ into the global sparse coefficient matrix $[K]$
   - Add components of $\{F^e\}$ into the global force vector $\{F\}$
2. Sort global sparse coefficient matrix $[K]$ into *compressed row* (or *compressed column*) form (parallel)
   - Sort column (or row) indices of each row (or column) of $[K]$ in non-decreasing order
   - Sum repeated entries of $[K]$ to enforce compatibility of primary variables
   - Remove "numerical" zeros from sparse matrix $[K]$
3. Solve global system of equations using an appropriate linear solver library (parallel)

memory based implementation of the general *processing* stage of a given finite element simulation. These procedures will be expanded upon for efficient use with high-order finite elements in Section B-2. The steps outlined in BOX 1 are applicable to any

finite element program regardless of whether the model problem is linear, nonlinear, quasi-static or transient. Since we have previously addressed building $[K^e]$ and $\{F^e\}$, we will focus the remainder of our discussion on parallel construction of the global system of sparse finite element equations.

In the current example, the full system of equations may be obtained by combining Eqs. (2.23), (2.22b) and (2.22c) into the following set of linear algebraic equations

$$\begin{bmatrix} 3{,}500^{\dagger} & 0^{\dagger} & 0 & 0 \\ 0^{\dagger} & 5{,}100^{\dagger *} & -1{,}600^{*} & 0 \\ 0 & -1{,}600^{*} & 2{,}100^{\diamondsuit *} & -500^{\diamondsuit} \\ 0 & 0 & -500^{\diamondsuit} & 510^{\diamondsuit} \end{bmatrix} \begin{Bmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \end{Bmatrix} = \begin{Bmatrix} 700{,}000^{\dagger} \\ 700{,}000^{\dagger *} \\ 0^{\diamondsuit *} \\ 500^{\diamondsuit} \end{Bmatrix} \tag{2.26}$$

Since all boundary conditions have been applied at the element level, the above system constitutes the final set of finite element equations for our simple example problem. The symbols $\dagger$, $\diamondsuit$ and $*$, corresponding with $e = 1$, 2 and 3 respectively, are included to illustrate which finite elements contribute to which coefficients of the global set of equations. Invoking the linear solver yields

$$\Delta_1 = 200.00^{\circ}\mathrm{C}, \quad \Delta_2 = 199.58^{\circ}\mathrm{C}, \quad \Delta_3 = 198.67^{\circ}\mathrm{C}, \quad \Delta_4 = 195.76^{\circ}\mathrm{C} \tag{2.27}$$

For very large problems it is impractical to construct the coefficient matrix $[K]$ in the form given in Eq. (2.26). Throughout this work we employ a *compressed row* (or *compressed column*) representation of $[K]$. This sparse storage format is closely related to *storage by indices*, whose data structure consists of ne (number of equations), nnz (number of non-zero entries in $[K]$) and the following arrays

$$\mathsf{k\_i} = (1, 2, 2, 3, 3, 3, 4, 4) \tag{2.28a}$$

$$\mathsf{k\_j} = (1, 2, 3, 2, 3, 4, 3, 4) \tag{2.28b}$$

$$\mathsf{k\_v} = (3500, 5100, -1600, -1600, 2100, -500, -500, 510) \tag{2.28c}$$

The integer arrays $\mathsf{k\_i}$ and $\mathsf{k\_j}$ contain the row and column addresses of the non-zero entries in $[K]$ respectively. The double precision real array $\mathsf{k\_v}$ contains the values of $[K]$ as accessed by $\mathsf{k\_i}$ and $\mathsf{k\_j}$. All three arrays are of size nnz. Note that the arrays are sorted by row and then column in ascending order. As a result, it is possible to abandon $\mathsf{k\_i}$ in favor of the following integer array

$$\mathsf{k\_p} = (1, 2, 4, 7, 9) \tag{2.29}$$

which is of size $\mathsf{ne} + 1$. As a result, the column indices of entries in row i are stored in $\mathsf{k\_j}(\mathsf{k\_p}(i), \ldots, \mathsf{k\_p}(i+1) - 1)$. The corresponding numerical values are stored in $\mathsf{k\_v}(\mathsf{k\_p}(i), \ldots, \mathsf{k\_p}(i+1) - 1)$. The data structure associated with the *compressed row* form of $[K]$ therefore consists of ne, $\mathsf{k\_p}$, $\mathsf{k\_j}$ and $\mathsf{k\_v}$. The *compressed row* (or the very similar *compressed column*) representation of $[K]$ is the sparse form of the global coefficient matrix typically required by modern linear solver libraries (e.g., UMFPACK, PARDISO, MUMPS, etc.).

Efficient construction of the *compressed row* form of $[K]$ for an arbitrary finite element discretization is a non-trivial task. We now proceed to describe a set of simple procedures for constructing $[K]$ that may be readily accomplished in parallel using the OpenMP paradigm. To simplify our discussion, we will present the key ideas in terms of the *storage by indices* data structure. First, it is important to note that nnz is not generally known prior to global assembly. However, if $\mathtt{size}([K^e]) = \mathsf{nke} \times \mathsf{nke}$ (where nke is the number of equations for a given element), then nnz may be bounded from above as $\mathsf{nnz} \leq \mathrm{NE} \times \mathsf{nke}^2$. We therefore initialize $\mathsf{k\_i}$, $\mathsf{k\_j}$ and $\mathsf{k\_v}$ to be of size

$\mathsf{nnzmax} = \mathrm{NE} \times \mathsf{nke}^2$ which for the current example problem yields

$$\mathsf{k\_i} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0) \tag{2.30a}$$

$$\mathsf{k\_j} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0) \tag{2.30b}$$

$$\mathsf{k\_v} = (0,0,0,0,0,0,0,0,0,0,0,0,0,0) \tag{2.30c}$$

Our next objective is to populate the entries in the above arrays with the components K_{ij}^e of the element-level coefficient matrices. To avoid race conditions in the parallel implementation of the algorithm, we will initially assign each component K_{ij}^e to a unique location in the sparse coefficient matrix. Such an assignment may be accomplished through the introduction of a unique integer $k \in [1, \ldots, \mathrm{NE} \times \mathsf{nke}^2] \subset \mathbb{N}$ associated with each component K_{ij}^e that may be determined from the following formula

$$k = \mathsf{pnt}((e-1)\mathsf{nke} + i) + j - 1 \tag{2.31}$$

where pnt is a one-dimensional array of size $\mathrm{NE} \times \mathsf{nke}$ constructed during the preprocessing stage of the analysis. In the current example problem we define pnt as

$$\mathsf{pnt} = (1^\dagger, 3^\dagger, 7^\diamondsuit, 11^\diamondsuit, 5^*, 9^*) \tag{2.32}$$

We therefore have the following formulas for constructing $\mathsf{k\_i}$, $\mathsf{k\_j}$ and $\mathsf{k\_v}$

$$\mathsf{k\_i}(k) = \mathsf{ECON}(e,i), \qquad \mathsf{k\_j}(k) = \mathsf{ECON}(e,j) \qquad \mathsf{k\_v}(k) = K_{ij}^e \tag{2.33}$$

which as applied to the current problem yields

$$\mathsf{k\_i} = (1^\dagger, 1^\dagger, 2^\dagger, \underline{2}^\dagger, \underline{2}^*, 2^*, \underline{3}^\diamondsuit, 3^\diamondsuit, 3^*, \underline{3}^*, 4^\diamondsuit, 4^\diamondsuit) \tag{2.34a}$$

$$\mathsf{k\_j} = (1^\dagger, 2^\dagger, 1^\dagger, \underline{2}^\dagger, \underline{2}^*, 3^*, \underline{3}^\diamondsuit, 4^\diamondsuit, 2^*, \underline{3}^*, 3^\diamondsuit, 4^\diamondsuit) \tag{2.34b}$$

$$\mathsf{k\_v} = (3500^{\dagger}, 0^{\dagger}, 0^{\dagger}, \underline{3500}^{\dagger}, \underline{1600}^{*}, -1600^{*}, \underline{500}^{\diamondsuit}, -500^{\diamondsuit}, -1600^{*}, \underline{1600}^{*}, \\ -500^{\diamondsuit}, 510^{\diamondsuit}) \tag{2.34c}$$

The symbols †, $\diamondsuit$ and $*$ are again included in the above expressions to more readily identify to the reader, which finite elements are associated with which coefficients of pnt, k_i, k_j and k_v for the present example problem. It should be apparent that the operations described in Eq. (2.33) may be readily accomplished in parallel. Furthermore, it should also be noted that the elements of the pnt array have been specifically defined in the pre-processing stage of the analysis such that the entries appearing in k_i are naturally sorted in non-decreasing order. The meaning of the pnt array should be clear: it is used to contiguously place the ith row of $[K^e]$ into k_i, k_j and k_v, starting at location $\mathsf{pnt}((e-1)\mathsf{nke}+i)$.

To obtain the sparse coefficient matrix in the form of the *storage by indices* data structure, we sort into non-decreasing order the columns associated with each given row, enforce compatibility of the primary variables and then remove any "numerically" zero entries. Sorting the columns of a given row may be readily facilitated via a robust sorting algorithm such as `quicksort`. Once sorted, compatibility of the primary variables may be achieved by summing the coefficient matrix values associated with any duplicate sets of indices (e.g., see the underlined terms appearing above in k_i, k_j and k_v). Finally, any entries in k_v that are considered "numerically" zero (i.e., whose magnitudes are less than a prescribed tolerance TOL) may then be removed from the coefficient matrix. Carrying out each of these operations results in

$$\mathsf{k\_i} = (1, 2, 2, 3, 3, 3, 4, 4, 0, 0, 0, 0) \tag{2.35a}$$

$$\mathsf{k\_j} = (1, 2, 3, 2, 3, 4, 3, 4, 0, 0, 0, 0) \tag{2.35b}$$

$$\mathsf{k\_v} = (3500, 5100, -1600, -1600, 2100, -500, -500, 510, 0, 0, 0, 0) \tag{2.35c}$$

It is important to note that the sorting and compatibility enforcement operations may be performed over each row via parallel processing. The final operation of removing zeros, on the other hand is inherently serial, yet requires relatively little computational expense (i.e., $\mathcal{O}(\mathsf{nnzmax})$). As a result, the overwhelming majority of computations needed to: (a) build the element-level equations and (b) construct the global finite element coefficient matrix, may be performed in a parallel computing environment. It should be readily apparent that once truncated to size $\mathsf{nnz} = 8$, the k_i, k_j and k_v arrays correspond identically with those given in Eq. (2.28).

It should be noted that up until now, we have devoted our attention to the construction of the sparse form of $[K]$ and have said nothing regarding the global assembly of $\{F\}$. It turns out that even when the global system of finite element equations is large, the global force vector $\{F\}$ may be adequately stored using a simple one-dimensional array. As a result, global construction of $\{F\}$ is completely straightforward in the serial case and by comparison requires only a modicum of additional programming logic to achieve an efficient parallel implementation.

The algorithms described above for construction of the sparse form of $[K]$ using the *storage by indices* data structure are efficient and easy to parallelize on shared memory systems using the OpenMP paradigm. With relatively modest modifications, the procedures may be adapted to directly construct the *compressed row* form of the sparse coefficient matrix without ever explicitly forming k_i. Furthermore, the strategies can also be further generalized to capitalize on any symmetry in the global system of equations. The main critique of the overall algorithm is that it in general requires a somewhat greater amount of memory to construct the sparse form of the global coefficient matrix than is actually needed to store the sparse form of $[K]$. In practice, however, we find that this need for extra memory does not constitute a computationally onerous requirement and is hence of little practical concern.

2. Additional element-level operations

The procedures outlined in the previous section have been presented in the context of the finite element analysis of a simple one-dimensional boundary-value problem using standard low-order finite element technology. These procedures may be readily adapted for use in the finite element analysis of multi-dimensional boundary-value problems using both low and high-order finite element discretizations. Efficient numerical implementation of high-order finite element technology, however, requires the deployment of a few additional procedures that are not necessarily required in low-order finite element formulations. The purpose of the current discussion, therefore, is to review what we feel are the most crucial element-level operations that may be used to substantially improve the competitiveness of high-order finite element formulations.

In general, the element-level equations associated with the eth finite element in a typical finite element discretization are given in terms of Eq. (2.11). During numerical construction, however, it is typical to partition the element-level equations for a given element into the following equivalent form

$$\begin{bmatrix} [K^{11}] & \cdots & [K^{1\mathsf{n}}] \\ \vdots & \ddots & \vdots \\ [K^{\mathsf{n}1}] & \cdots & [K^{\mathsf{n}\mathsf{n}}] \end{bmatrix} \begin{Bmatrix} \{\Delta^{(1)}\} \\ \vdots \\ \{\Delta^{(\mathsf{n})}\} \end{Bmatrix} = \begin{Bmatrix} \{F^{(1)}\} \\ \vdots \\ \{F^{(\mathsf{n})}\} \end{Bmatrix} \tag{2.36}$$

where n is the number of dependent variables constituting $\mathbf{u}$. We note that the element-level equations have been partitioned with respect to the $\{\Delta^{(j)}\}$ arrays, where each array represents a column vector containing the values of the jth component of $\mathbf{u}_{hp}$ as evaluated at the element nodes. The components of each $\{\Delta^{(j)}\}$ array are related to $\mathbf{\Delta}_i^e$ (defined in Eq. (2.3)) by the formula $\mathbf{\Delta}_i^e = \{\Delta_i^{(1)} \cdots \Delta_i^{(n)}\}^{\mathrm{T}}$.

For general nd-dimensional finite element problems, it is typically possible to

express the components of each element sub-coefficient matrix $[K^{\alpha\beta}]$ as

$$\begin{aligned} K_{ij}^{\alpha\beta} &= \int_{\bar{\Omega}^e} \sum_{l=0}^{nd} \sum_{m=0}^{nd} \mathcal{C}_{lm}^{\alpha\beta}(\mathbf{x}, \mathbf{u}_{hp}(\mathbf{x})) \mathcal{S}_{ij}^{lm}(\mathbf{x}) d\bar{\Omega}^e \\ &= \int_{\hat{\Omega}^e} \sum_{l=0}^{nd} \sum_{m=0}^{nd} \mathcal{C}_{lm}^{\alpha\beta}(\mathbf{x}(\boldsymbol{\xi}), \mathbf{u}_{hp}(\mathbf{x}(\boldsymbol{\xi}))) \mathcal{S}_{ij}^{lm}(\mathbf{x}(\boldsymbol{\xi})) J(\boldsymbol{\xi}) d\hat{\Omega}^e \end{aligned} \tag{2.37}$$

where $J(\boldsymbol{\xi})$ is the Jacobian of the isoparametric coordinate transformation for the element (i.e., the determinant of the Jacobian matrix) and $i, j = 1, \ldots, (p+1)^{nd}$. A similar expression may also be produced for determining the coefficients F_i^{α} of the force vector. The quantities $\mathcal{S}_{ij}^{lm}$ represent products of the interpolation functions (and their spatial derivatives) of the form

$$\mathcal{S}_{ij}^{00} = \psi_i \psi_j, \qquad \mathcal{S}_{ij}^{0m} = \psi_i \frac{\partial \psi_j}{\partial x_m}, \qquad \mathcal{S}_{ij}^{l0} = \frac{\partial \psi_i}{\partial x_l} \psi_j, \qquad \mathcal{S}_{ij}^{lm} = \frac{\partial \psi_i}{\partial x_l} \frac{\partial \psi_j}{\partial x_m} \tag{2.38}$$

where $l, m = 1, \ldots, nd$ and x_m are the components of $\mathbf{x}$ as expressed with respect to some fixed Cartesian coordinate system (i.e., $\mathbf{x} = x_m \hat{\mathbf{e}}_m$). The spatial derivatives of the interpolation functions are of course evaluated in terms of the natural coordinates $\boldsymbol{\xi}$ using the components of the inverse Jacobian matrix associated with the isoparametric mapping from $\hat{\Omega}^e$ to $\bar{\Omega}^e$. Note that $\mathcal{S}_{ij}^{lm}$ possesses the following symmetry $\mathcal{S}_{ij}^{lm} = \mathcal{S}_{ji}^{ml}$. The coefficients $\mathcal{C}_{lm}^{\alpha\beta}$ may be constant, spatially varying and/or dependent on the components of the dependent variable $\mathbf{u}_{hp}(\mathbf{x})$. It is our observation that the coefficients $\mathcal{C}_{lm}^{\alpha\beta}$ can, in general, become quite involved; this is especially true in shell finite element formulations as well as for least-squares based finite element models. We recall that, throughout this work, Gauss-Legendre quadrature rules are employed exclusively in evaluation of the element-level coefficient matrices and force vectors. To achieve an attractive level of performance in the numerical integration of the element-level partitioned coefficient matrices, we find it *imperative* to:

1. Decompose each partitioned coefficient matrix into the form given in Eq. (2.37).

2. At a given quadrature point explicitly evaluate the components of $\mathcal{C}_{lm}^{\alpha\beta}$, prior to looping over i and j in the numerical evaluation of $K_{ij}^{\alpha\beta}$.

Similar observations can also be made regarding construction of F_i^α.

A major disadvantage of high-order finite elements is that the connectivity between the degrees of freedom of a given element and also between neighboring elements increases with p. To emphasize the implications of this increased connectivity, we consider the following scenario: suppose region $\bar{\Omega}$ is discretized using two distinct finite element meshes; in the first case a standard low-order finite element discretization is employed and in the second we utilized a high-order finite element mesh, where $p > 1$. Assuming that both meshes have the same total number of nodes, the global coefficient matrix associated with the latter discretization will always be *more dense* than the coefficient matrix for the former. This is a direct consequence of the large element-level coefficient matrices that are naturally generated when high-order finite elements are employed. High-order discretizations, however, typically require far fewer total degrees of freedoms (as compared with low-order discretizations) to obtain reliable numerical solutions. Even with this advantage, however, a high-order discretization will inevitably require more computer memory resources to store the global coefficient matrix.

The onerous memory requirements associated with high-order finite element models may be reduced through the use of element-level static condensation [22, 24]. As we will demonstrate, static condensation reduces global memory requirements and allows for significant parallelization in the global solution procedure. In an effort to present the key ideas, we rearrange the element-level equations for the eth element

into the following form

$$\begin{bmatrix} [K^{\mathcal{BB}}] & [K^{\mathcal{BI}}] \\ [K^{\mathcal{IB}}] & [K^{\mathcal{II}}] \end{bmatrix} \begin{Bmatrix} \{\Delta^{(\mathcal{B})}\} \\ \{\Delta^{(\mathcal{I})}\} \end{Bmatrix} = \begin{Bmatrix} \{F^{(\mathcal{B})}\} \\ \{F^{(\mathcal{I})}\} \end{Bmatrix} \tag{2.39}$$

In the above expression we have partitioned the element-level system of equations with respect to the element boundary degrees of freedom $\{\Delta^{(\mathcal{B})}\}$ and the element interior degrees of freedom $\{\Delta^{(\mathcal{I})}\}$. Since the interior degrees of freedom for element e do not contribute to the element-level equations of any other element, it is possible to implicitly remove them from Eq. (2.39). This process yields the following *condensed* set of equations for the element boundary degrees of freedom

$$[\bar{K}^e]\{\Delta^{(\mathcal{B})}\} = \{\bar{F}^e\} \tag{2.40}$$

where the effective element coefficient matrix $[\bar{K}^e]$ and force vector $\{\bar{F}^e\}$ are of the form

$$[\bar{K}^e] = [K^{\mathcal{BB}}] - [K^{\mathcal{BI}}][K^{\mathcal{II}}]^{-1}[K^{\mathcal{IB}}] \tag{2.41a}$$

$$\{\bar{F}^e\} = \{F^{(\mathcal{B})}\} - [K^{\mathcal{BI}}][K^{\mathcal{II}}]^{-1}\{F^{(\mathcal{I})}\} \tag{2.41b}$$

It is important to note that $[K^{\mathcal{II}}]^{-1}$ need not be evaluated explicitly. Instead, the operations $[K^{\mathcal{II}}]^{-1}[K^{\mathcal{IB}}]$ and $[K^{\mathcal{II}}]^{-1}\{F^{(\mathcal{I})}\}$ may be performed via Gaussian elimination with partial pivoting using the standard LAPACK subroutine `dgesv`.

It is now possible to formulate the global system of finite element equations in terms of the element boundary degrees of freedom only. This system necessitates only a fraction of the memory required to compute the full system of finite element equations (formulated in terms of all degrees of freedom). Once the element boundary degrees of freedom have been determined by the global solver, the interior degrees of freedom may be obtained by solving the following set of equations for each finite

BOX 2. Processing stage of a high-order finite element simulation.

1. Loop over all finite elements: $e = 1, \mathrm{NE}$ (parallel)
   - Build element coefficient matrix $[K^e]$ and force vector $\{F^e\}$
   - Apply essential boundary conditions to $[K^e]$ and $\{F^e\}$
   - Perform static condensation to construct $[\bar{K}^e]$ and $\{\bar{F}^e\}$
   - Add components of $[\bar{K}^e]$ into the global sparse coefficient matrix $[K]$
   - Add components of $\{\bar{F}^e\}$ into the global force vector $\{F\}$
2. Sort global sparse coefficient matrix $[K]$ into *compressed row* (or *compressed column*) form (parallel)
   - Sort column (or row) indices of each row (or column) of $[K]$ in non-decreasing order
   - Sum repeated entries of $[K]$ to enforce compatibility of primary variables
   - Remove "numerical" zeros from sparse matrix $[K]$
3. Solve global system of equations for all element boundary degrees of freedom using an appropriate linear solver library (parallel)
4. Loop over all finite elements: $e = 1, \mathrm{NE}$ (parallel)
   - Solve Eq. (2.42) for interior degrees of freedom $\{\Delta^{(\mathcal{I})}\}$

element

$$[K^{\mathcal{I}\mathcal{I}}]\{\Delta^{(\mathcal{I})}\} = \{F^{(\mathcal{I})}\} - [K^{\mathcal{I}\mathcal{B}}]\{\Delta^{(\mathcal{B})}\} \tag{2.42}$$

The finite element solution procedure for high-order finite element discretizations, given in BOX 2, is obtained by augmenting the steps presented in BOX 1 to also include element-level static condensation.

At this point we find it prudent to note that whenever element-level static condensation is adopted, nnzmax (used to allocate memory for constructing the sparse global coefficient matrix) may be determined using the following formula

$$\mathsf{nnzmax} = \mathrm{NE}\{\mathsf{n}[(p+1)^{nd} - (p-1)^{nd}]\}^2 \tag{2.43}$$

where we recall that NE is the number of elements and n is the number of degrees of freedom per node. We can likewise show that for an equivalent low-order mesh (possessing the same total number of nodes and the p-level taken as 1) the quantity nnzmax is of the form

$$\mathsf{nnzmax} = \mathrm{NE}(2^{nd}\mathsf{n})^2 p^{nd} \tag{2.44}$$

where NE and p are the number of elements and p-level of the original *high-order* discretization (as opposed to the equivalent low-order mesh). When $nd = 2$, it turns out that $\mathsf{nnzmax} = 16 \times \mathrm{NE}(p\,\mathsf{n})^2$ for both low and high-order finite element discretizations (provided of course that element-level static condensation is adopted). As a result, system memory requirements associated with constructing the global sparse coefficient matrix are equivalent in both cases. This highly desirable result is especially attractive in the high-order shell finite element formulation discussed in Chapter VI.

A typical two-dimensional high-order spectral/*hp* finite element mesh is shown in Figure 4. In this figure we show both the full finite element mesh (where all nodes are depicted) and also the statically condensed mesh (where only the element boundary

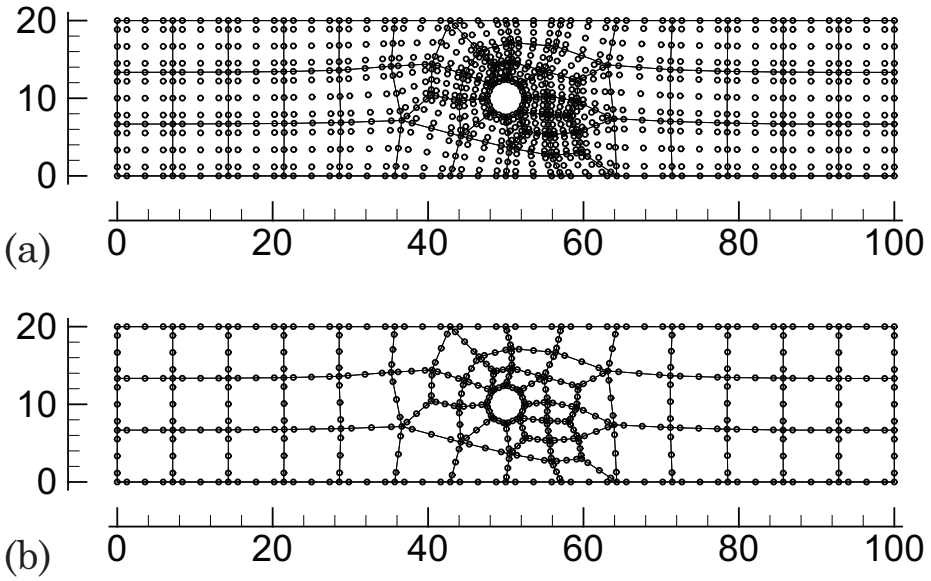


Fig. 4. A high-order spectral/*hp* finite element discretization of a two-dimensional region. Case shown is for $p = 4$: (a) finite element mesh showing the elements and nodes and (b) finite element mesh showing the elements and element boundary nodes.

nodes are shown). It is worth noting that the computer implementation of element-level static condensation in high-order finite element models requires either: (a) the user to carefully number the global nodes such that the element-boundary nodes are numbered first or (b) the computer program to automatically re-number the global nodes associated with the statically condensed mesh. In this work we have adopted the latter approach, as it is far less restrictive on the program user. Either way, it is important to note that it is still necessary for the computer program to generate a data structure for the element connectivity array associated with the statically condensed finite element mesh. For additional details on static condensation as applied to high-order finite element models, we refer to the book by Karniadakis and Sherwin [22] and the journal paper by Couzy and Deville [26].

C. Numerical example: a verification benchmark

We now wish to numerically demonstrate the performance of the shared-memory based parallelization strategies advocated in the previous section. As an example problem, we consider the steady low Reynolds number two-dimensional flow of a viscous incompressible fluid past a circular cylinder. The computational domain $\bar{\Omega}$ on which the problem is posed is defined as the set difference between the closed rectangular region $[-25, 25] \times [-15, 15]$ and an open unit-diameter circular cylinder centered about the origin. The fluid along the top, bottom and left hand sides of the domain is traveling with a unit horizontal velocity. A no slip condition is taken along the circular cylinder and an appropriate outflow boundary condition is utilized along the right hand side of the domain (see Chapters III and IV for details). The Reynolds number for the flow is taken to be 40.

For the finite element discretization, we use 1,920 quadrilateral elements, as shown in Figure 5, and employ an eighth-order polynomial expansion within each element; this amounts to 123,904 total nodal points in the finite element mesh. The finite element formulation of the problem is obtained through the use of least-squares based finite element technology as applied to the first-order vorticity form of the Navier-Stokes equations. The finite element discretization, therefore, contains a total of 495,616 degrees of freedom; and as a result, is suitable for showcasing the performance of the algorithms used in our numerical implementation. The discretization, however, is *far more* dense than is actually required to obtain a reliable numerical solution.

For the numerical implementation, we utilize the C++ programming language and IBM's AIX v11.1 compiler. Efficient parallelization is achieved by combin-

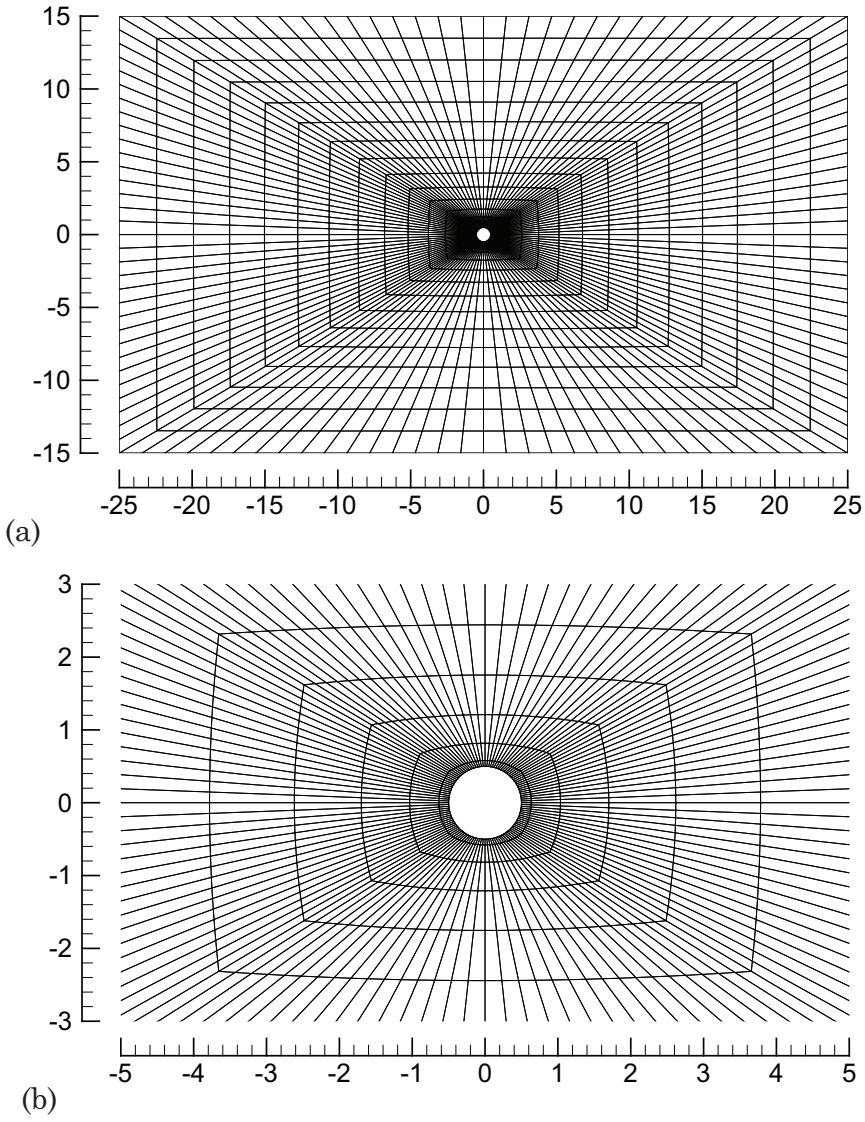


Fig. 5. Finite element mesh used in the solution of steady fluid flow past a circular cylinder: (a) full view of the finite element mesh and (b) close up view of the finite element mesh in the vicinity of the cylinder.

ing the algorithms described earlier in this chapter with appropriate placements of pre-compiler directives prior to parallelizable `for` loops using the C/C++ specific OpenMP syntax `#pragma omp parallel for`. The current test problem has been solved using the computational resources available at the Texas A&M Supercomputing Facility at Texas A&M University. The simulations were run on the Hydra supercomputer, an IBM Cluster-1600, that is made up of IBM's 1.9 GHz RISC Power5+ processors. Each node is a symmetric multi-processor (SMP) system with 16 processors and 25 GB of usable shared memory.

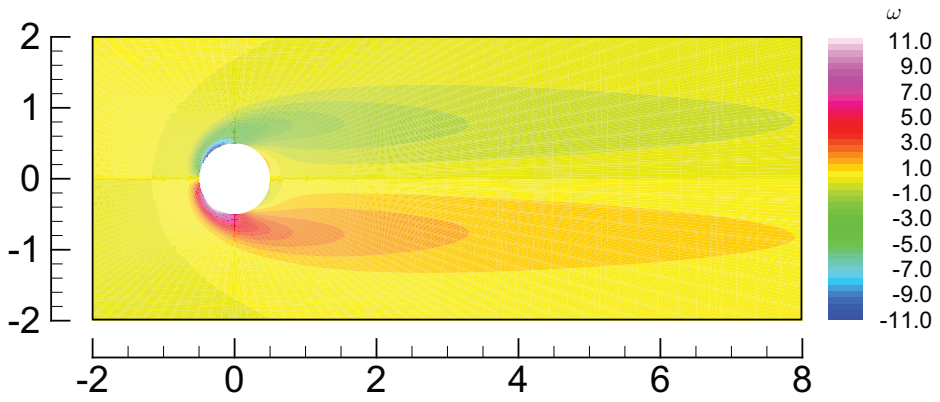


Fig. 6. Vorticity field ω for the steady flow of a viscous incompressible fluid past a circular cylinder at $\mathrm{Re} = 40$.

The non-dimensionalized vorticity field ω in the vicinity of the cylinder for the test problem, as obtained in parallel using 16 processors, is shown in Figure 6. In Figure 7 we provide a comparison of the theoretical performance with the actually observed speedups for steps 1, 2 and 4 of BOX 2 (i.e., the general *processing* stage of the finite element simulation). Near ideal performance is achieved for steps 1 and 4; however, the speedup observed for step 2 is clearly not optimal. Since the number of operations associated with this step is far less than the number of computations

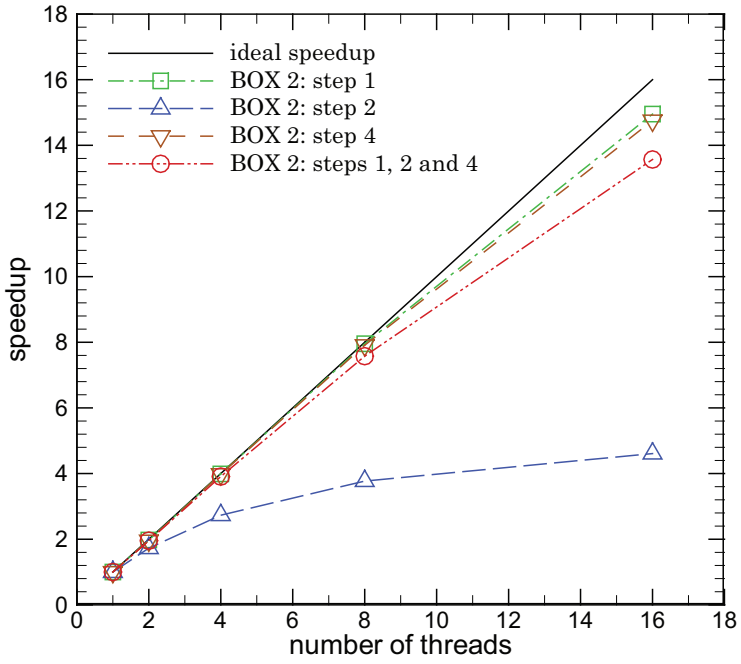


Fig. 7. Parallel performance observed in the finite element solution of the low Reynolds number flow of a viscous incompressible fluid past a circular cylinder.

needed to carry out steps 1 and 4 (see Table I for a comparison of wall clock times), the cumulative parallel performance is actually quite competitive (e.g., 94.7% of ideal performance is achieved with 8 threads and 84.8% of ideal performance is obtained using 16 threads). We note in passing that the results presented in Figure 7 and Table I have been averaged over the 6 nonlinear solution iterations required to satisfy a nonlinear convergence criteria of 10^{-6}.

For the current example, step 3 of BOX 2 was performed using the external UMFPACK library [27, 28, 29, 30] (a set of routines for solving sparse unsymmetric linear systems directly using the multifrontal method). Due to element-level static condensation, only 119,296 equations needed to actually be solved at the global level (24.07% of the original system of 495,616 equations). This translated into tremendous

Table I. Elapsed wall clock time for various steps of the finite element *processing* stage of a given nonlinear iteration for steady flow past a cylinder (np is the number of processors or threads).

	Elapsed wall clock time (in seconds)				
Processing procedure(s)	np = 1	np = 2	np = 4	np = 8	np = 16
BOX 2: step 1	52.31	26.47	13.11	6.58	3.50
BOX 2: step 2	4.14	2.40	1.52	1.10	0.90
BOX 2: step 4	40.00	20.43	10.05	5.06	2.71
BOX 2: steps 1, 2 and 4	96.45	49.30	24.68	12.74	7.11

memory savings in both the construction and direct factorization of $[K]$; in this example less than 2.5 GB of RAM was actually required during the entire solution process. Of course, even less memory would have been necessary had an appropriate iterative solver, such as the preconditioned conjugate gradient method, been employed. The present case study clearly demonstrates that the high-order finite element procedures and algorithms discussed in this chapter may be readily utilized to efficiently solve non-trivial finite element problems on shared-memory systems.

CHAPTER III

LEAST-SQUARES FINITE ELEMENT FORMULATIONS FOR NONLINEAR BOUNDARY-VALUE PROBLEMS: AN ANALYSIS OF THE MINIMIZATION AND LINEARIZATION OPERATIONS*

In this chapter we consider application of spectral/*hp* finite element procedures to the solution of nonlinear systems of partial differential equations using least-squares variational principles. The chapter is motivated in part by the considerable attention the least-squares method has received in recent years, particularly as applied in the numerical solution of the Navier-Stokes equations governing flows of viscous incompressible fluids. Although we will discuss these equations in particular, our ultimate objective is to provide a more general discussion of least-squares variational principles as applied to nonlinear boundary-value problems. More specifically, we will discuss the specific roles played by the *minimization* and *linearization* operators in nonlinear least-squares finite element models and demonstrate in what manner the numerical solution is affected by exchanging the application order of these operations.

The chapter is organized as follows. We begin by providing a brief overview of the least-squares method, with an emphasis on L_2-norm based least-squares formulations that are *practical* for numerical implementation. We then consider the abstract setting for an L_2-norm based least-squares formulation of an abstract first-order nonlinear boundary-value problem. The least-squares weak formulation is developed for this abstract system via direct minimization of the least-squares functional through the aid of the Gâteaux derivative. We provide a thorough discussion of possible forms taken

*The numerical results reported in this chapter appear in the article "On the roles of minimization and linearization in least-squares finite element models of nonlinear boundary-value problems" by G. S. Payette and J. N. Reddy, *J. Comp. Phys.*, vol. 230, pp. 3589–3613, 2011. Copyright (2011) Elsevier Science.

by the *linearized* weak formulation, when linearization is either performed before or after minimization of the least-squares functional in the context of both the Picard and Newton linearization schemes. We show that although the underlying least-squares principle suggests that minimization ought to be performed prior to linearization, such an approach is often impractical and not necessary. Finally, we underscore the differences between the various linearization schemes adopted in the abstract formulation, by numerically solving several nonlinear two-dimensional verification benchmark boundary-value problems using least-squares finite element models. As a first example we solve a nonlinear form of the Poisson equation. We also present three numerical solutions of the incompressible Navier-Stokes equations, including steady flow past a circular cylinder, flow over a backward facing step and lid-driven cavity flow. For each benchmark, we provide a detailed assessment of the performance of each least-squares finite element formulation.

A. An overview of the least-squares method

It is well known that the success of weak-form Galerkin finite element models in obtaining favorable numerical solutions to partial differential equations is intimately connected with the notion of global minimization of unconstrained quadratic functionals [2]. When the Galerkin based weak formulation of a set of partial differential equations can be obtained equivalently through the minimization of a quadratic functional, the finite element solution becomes an orthogonal projection of the exact solution onto the trial space associated with the finite element discretization. The resulting numerical solution represents the best possible approximation of the exact solution in the trial space as measured with respect to the energy norm of the functional. When the energy norm $\|\mathbf{u}\|_E$ can be shown to be equivalent to a more standard

norm associated with an appropriate Hilbert space (e.g., the $\mathbf{H}^1(\Omega)$ norm), optimal convergence rates of the finite element solution can be established. Such a setting, often referred to as a *variational setting*, is ideal for finite element approximation and is exemplified by the case of linear elasticity [3].

Unfortunately, finite element models based on the weak-form Galerkin procedure often depart from the ideal variational setting; this is especially the case for many problems arising outside the realm of solid mechanics. For example, application of the Galerkin method to the Stokes equations results in a constrained variational problem, whose discrete solution must satisfy restrictive compatibility conditions [31]. The weak-form Galerkin finite element model of the Navier-Stokes equations on the other hand is completely divorced from any minimization principles and further inherits the discrete inf-sup condition of the Stokes problem [32]. Numerical solutions are far from optimal as characterized by the need for severe mesh refinement in order to suppress spurious oscillations of the solution. A considerable amount of research in recent years has been devoted to modifications of the weak-form Galerkin approach in the hope of obtaining a more favorable setting for the numerical solution. Stabilized finite element formulations such as the penalty [6, 7], SUPG [4, 5] and Galerkin least-squares [8] have been proposed and extensively researched. These schemes have yet to gain wide acceptance, due in part to the associated temporal and mesh dependent ad-hoc parameters that must be fine tuned in each formulation.

Finite element models based on least-squares variational principles often offer an appealing alternative to the more popular weak-form Galerkin approach. This is especially relevant in the analysis of partial differential equations containing non-self-adjoint operators, as is found in the Navier-Stokes equations. Although not as popular as weak-form Galerkin formulations, least-squares models of partial differential equations have been an active field of research since at least the early 1970's [33].

In 1976, Eason [34] compiled an extensive review containing well over 200 references to least-squares methods as applied to the solution of partial differential equations. Since the publication of this review article, least-squares finite element models have continued to receive substantial attention and discussion in the literature.

Least-squares variational formulations allow us to define an unconstrained convex least-squares functional $\mathcal{J}(\mathbf{u})$ in terms of the sum of the squares of the norms of the partial differential equation residuals [35]; where standard inner product based Sobolev norms are typically employed (e.g., norms associated with $L_2(\Omega)$ or $H^k(\Omega)$, where $k \in \mathbb{N}$). If the governing equations (augmented by the appropriate boundary conditions) are well posed, it can be readily shown that the exact solution coincides with the minimizer of the least-squares functional. As a result, in the least-squares method the weak formulation is obtained via direct minimization of $\mathcal{J}(\mathbf{u})$. The concept of minimization of the partial differential equation residuals is, therefore, at the heart of the least-squares formulation. For the case of linear partial differential equations, it is always possible to associate with the least-squares functional a well defined energy norm $\|\mathbf{u}\|_E$. If it can be shown that the energy norm induced by the least-squares functional is equivalent to an appropriate standard norm, such as the $\mathbf{H}^1(\Omega)$ norm, optimal convergence rates can be established for the least-squares finite element model. Under such conditions the least-squares finite element formulation constitutes an ideal *variational setting*, regardless of whether or not such a setting is achieved by the associated weak-form Galerkin finite element formulation [2].

To maintain practicality in the numerical implementation, it becomes computationally advantageous to construct the least-squares functional in terms of the sum of the squares of the $L_2(\Omega)$ norms of the first-order form of the partial differential equation residuals. Regrettably, it is not always possible to establish *a priori* norm equivalence (or $H^1(\Omega)$-coercivity) of the resulting least-squares formulation. As iden-

tified by Bochev [36] using Agmon, Douglis, Nirenberg (ADN) elliptic theory [37], it is typically possible to construct a least-squares functional that is $H^1(\Omega)$-coercive [2, 38, 39, 40]. Unfortunately, the optimal choice of norms can: (a) depend on the nature of the boundary conditions of a given problem and (b) result in an unattractive computational implementation. It is important to note that departure from the ideal variational setting (i.e., using a least-squares functional that is non-$H^1(\Omega)$-coercive) does not typically result in disastrous consequences for least-squares finite element models. Even when a given formulation is non-$H^1(\Omega)$-coercive, the least-squares finite element model always: (a) possesses the best approximation property with respect to a well-defined norm (i.e., the energy norm $\|\mathbf{u}\|_E$) and (b) avoids restrictive compatibility requirements on the finite element function spaces (i.e., the discrete inf-sup condition never arises). That the least-squares method is always based on a minimization principle ensures a robust setting that is often lacking in Galerkin based weak formulations. It is well known, however, that non-$H^1(\Omega)$-coercive low-order finite element implementations are often prone to locking whenever full numerical integration techniques are employed in evaluating the coefficient matrices. In the context of the Navier-Stokes equations, it has been shown that such issues may be largely avoided through the use of collocation or selective reduced integration strategies [9, 10, 11, 12, 13, 14]. On the other hand, the combination of high-order finite element technology with least-squares variational principles has also shown great promise in recent years. In particular, building off of the earlier work of Jiang and Sonnad [41] and Bell and Surana [42, 43], Proot and Gerritsma [44, 45, 46, 47] and Pontaza and Reddy [15, 16, 17, 18] demonstrated numerically that *hp*-least-squares finite element models are capable of yielding highly accurate results even when the least-squares functional cannot be shown to be $H^1(\Omega)$-coercive *a priori*.

Least-squares finite element models offer several additional attractive features as

compared with weak-form Galerkin formulations. In the case of linear analysis, the least-squares formulation always admits a symmetric positive-definite (SPD) coefficient matrix, regardless of whether or not such symmetry is manifest in the governing partial differential equations. As a result, extremely robust direct as well as iterative solution algorithms (such as the preconditioned conjugate gradient method) can be employed in the solution process [48, 49] and only half of the global coefficient matrix need be stored in memory. This is not the case when the weak-form Galerkin scheme is applied to non-self-adjoint systems of equations [50]. As mentioned previously, the least-squares formulation does not suffer from the restrictive inf-sup condition. This is highly desirable in the numerical discretization of fluid mechanics problems, as it allows the velocity and pressure to be approximated using the same bases of interpolation [35]. Finally, least-squares formulations are also free from the need for numerical dissipation through the use of upwind techniques. As a result, ad-hoc stabilization is not needed in the analysis of convection dominated problems [35].

Least-squares formulations are certainly not without their own deficiencies. Most problems in physics possess at the very minimum second order spatial differential operators. Since no weakening of these operators is typically possible through the employment of Green's identities (as can be readily accomplished in weak-form Galerkin formulations), least-squares models typically require higher regularity of the approximate solution within each element. Higher regularity requirements negatively affect the condition number of the coefficient matrix and also the continuity requirement of the solution across element boundaries. High regularity requirements may be avoided by constructing the least-squares finite element model in terms of an equivalent lower-order system by the introduction of additional independent auxiliary variables [35]. The resulting *mixed* formulation permits the use of standard Lagrange interpolation functions and also improves the conditioning of the global coefficient matrix [48].

However, such benefits are gained at the expense of an increase in size of the global system of equations. It can be argued that such a formulation is at least somewhat useful, however, as the auxiliary variables often represent important physical quantities of interest (e.g., the heat flux, vorticity, stress, etc.). Other drawbacks to least-squares formulations, in the context of fluid mechanics, include lack of local mass conservation and poor coupling between the velocity and pressure in transient problems. Least-squares formulations seeking to address these issues have been adopted by Chang and Nelson [51], Pontaza [52], Prabhakar and Reddy [53, 54, 55] and Prabhakar et al. [56]. Additionally, in Chapter IV we propose a least-squares finite element model of the incompressible Navier-Stokes equations with improved local mass conservation.

B. The minimization and linearization procedures

As stated previously, the fundamental principle for the least-squares method, as applied to a given boundary-value problem, is that the function minimizing the least-squares functional coincides with the exact solution. The necessary condition for minimization naturally requires the first variation of the least-squares functional to be identically zero; carrying out this procedure produces the weak form of the governing equations. In the least-squares literature it is common to refer to the resulting weak formulation as the Euler (or Euler-Lagrange) equation of the least-squares variational boundary-value problem (see for example Bochev [2]). The Euler equation resulting from invoking the necessary condition forms the basis of the least-squares finite element model. As noted previously, the least-squares variational principle associated with linear systems of partial differential equations always produces a symmetric bilinear form and as a result, a symmetric system of finite element equations. This is a highly attractive property of the least-squares method. When the governing equa-

tions are nonlinear, however, symmetry of the nonlinear Euler equation is not always guaranteed.

Paramount to the solution of a system of nonlinear algebraic equations (as arise in the finite element approximation of a nonlinear boundary-value problem) is the need for linearization. In the context of an iterative solution procedure, the role of linearization is to facilitate the solution to the original nonlinear equations through the successive solution of an appropriate linearized form of the equations [50]. In this work we consider the two most common iterative solution procedures, namely the methods of Picard and Newton. The Picard scheme enjoys a large radius of convergence accompanied by a slow convergence rate. Conversely, Newton's method offers a quadratic rate of convergence when the assumed solution is near the true solution point. However, this method possesses a much smaller radius of convergence than what is generally exhibited when the Picard scheme is employed.

In practice, it is possible to adopt one of two approaches when constructing least-squares finite element models of nonlinear boundary-value problems. In the first approach, linearization of the nonlinear partial differential equations is performed prior to minimization of the least-squares functional. A major motivation for this approach is the desire to maintain symmetry and positive-definiteness in the resulting finite element coefficient matrix. As pointed out by Jiang, such a framework also yields "minimization problems of quadratic functionals which have been well studied" [35]. Whenever this approach is adopted, we say that the least-squares finite element model is constructed via *linearization before minimization*. In general, however, the discrete minimizer resulting from this approach will be associated with the linearized governing equations (as opposed to the actual set of PDEs). In the second approach, we instead construct the least-squares functional from the original set of nonlinear partial differential equations [42, 43] and then require the first variation of this *true*

least-squares functional to be identically zero. The motivation for this approach has to do with the least-squares principle itself. Since the least-squares method is independent of both the discretization procedure and the iterative nonlinear solution scheme, it is easy to see that this second approach is mathematically consistent with the underlying least-squares variational principle. When this procedure is chosen, we say that the finite element model is formulated through *linearization after minimization*. The drawback to this approach, however, is that the resulting linearized system will in general be non-symmetric and non-positive definite.

It is worthwhile to note that in the context of finite element models based on weak-form Galerkin formulations, there is no distinction between linearization of the governing partial differential equations before or after creation of the weak form. In essence, both approaches are equivalent, owing to the fact that the Galerkin procedure constitutes a linear operation which acts on the nonlinear set of governing equations. For the case of least-squares finite element formulations on the other hand, *linearization before minimization* is clearly not equivalent to *linearization after minimization* [20]. Throughout the remainder of this chapter, we highlight the differences between these two approaches and discuss in what context the numerical solution is affected by interchanging the application order of the minimization and linearization operations.

C. Abstract least-squares formulations of nonlinear boundary-value problems

In this section we present the steps involved in developing and arriving at weak formulations, based on the least-squares method, for nonlinear boundary-value problems that can be readily utilized to construct least-squares finite element models. To insure a general treatment of the subject, we present the fundamental concepts and procedures in the context of an abstract boundary-value problem. The least-squares

functional for this general problem is defined in terms of the sum of the squares of the L_2 norms of the abstract equation residuals. The Euler equation associated with the problem is then developed through an appropriate minimization of the least-squares functional with respect to the solution variable(s). We provide a thorough discussion of possible forms taken by the linearized Euler equation, when linearization is performed before or after minimization of the functional in the context of both the Picard and Newton linearization schemes. We also present a simple mathematical analysis of Newton's method as applied to least-squares formulations (both before and after minimization).

1. The abstract nonlinear boundary-value problem

We recall from notation introduced in Chapter II that Ω and Γ represent, respectively, the domain and boundary upon which a typical boundary-value problem may be posed. In addition, we follow the customary procedure of partitioning the boundary Γ into Dirichlet Γ^{D} and Neumann Γ^{N} parts, such that $\Gamma = \Gamma^{\mathrm{D}} \bigcup \Gamma^{\mathrm{N}}$ and $\Gamma^{\mathrm{D}} \bigcap \Gamma^{\mathrm{N}} = \varnothing$. We consider the following abstract boundary-value problem

$$\mathcal{L}(\mathbf{u}) = \mathbf{f} \qquad \text{in } \Omega \tag{3.1a}$$

$$\mathbf{u} = \mathbf{u}^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{D}} \tag{3.1b}$$

$$g(\mathbf{u}) = \mathbf{h} \qquad \text{on } \Gamma^{\mathrm{N}} \tag{3.1c}$$

where $\mathcal{L}$ is a nonlinear first-order spatial partial differential operator, $\mathbf{u}$ is the dependent variable, $\mathbf{f}$ is the forcing function and $\mathbf{u}^{\mathrm{p}}$ is the prescribed essential boundary condition. The flux or Neumann boundary condition for the problem is expressed in terms of the operator g and the prescribed function $\mathbf{h}$. The boundary conditions are of course understood in the sense of traces [23, 57]. We assume that the function g is

linear in $\mathbf{u}$ and that the problem is well-posed.

2. The L_2 least-squares functional and associated minimization principle

At the center of the least-squares method and the resulting least-squares finite element model is the least-squares functional. In keeping with our desire to maintain an appropriate level of practicality in the numerical implementation, we construct the least-squares functional in terms of the sum of the squares of the L_2 norms of the abstract equation residuals

$$\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h}) = \frac{1}{2}\Big(\|\mathcal{L}(\mathbf{u})-\mathbf{f}\|^2_{\Omega,0} + \|g(\mathbf{u})-\mathbf{h}\|^2_{\Gamma^{\mathrm{N}},0}\Big) \tag{3.2}$$

It is important to note that the least-squares functional has been defined such that the Neumann boundary condition is enforced *weakly* as a consequence of the minimization procedure. Hence, in the numerical implementation there will be no need to constrain the finite element function spaces to satisfy the natural boundary conditions.

The abstract minimization principle associated with the least-squares method may be stated as follows: find $\mathbf{u} \in \mathcal{V}$ such that

$$\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h}) \leqslant \mathcal{J}(\tilde{\mathbf{u}};\mathbf{f},\mathbf{h}) \quad \text{for all } \tilde{\mathbf{u}} \in \mathcal{V} \tag{3.3}$$

The function space $\mathcal{V}$ associated with the least-squares problem is defined as

$$\mathcal{V} = \big\{\mathbf{u} : \mathbf{u} \in \mathbf{H}^1(\Omega),\ \mathbf{u} = \mathbf{u}^{\mathrm{p}} \text{ on } \Gamma^{\mathrm{D}}\big\} \tag{3.4}$$

The necessary condition for minimization requires that the first variation of $\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h})$ be identically zero. The minimization procedure may be readily facilitated through the use of the *Gâteaux derivative* [23]. We recall that the Gâteaux derivative (or

Gâteaux variation) of a general functional $\Pi(\mathbf{u})$ in the direction of $\delta\mathbf{u}$ is defined as

$$\delta\Pi(\mathbf{u},\delta\mathbf{u}) = \frac{d}{d\varepsilon}\Pi(\mathbf{u}+\varepsilon\delta\mathbf{u})\Big|_{\varepsilon=0} = D\Pi(\mathbf{u})[\delta\mathbf{u}] = \nabla\Pi(\mathbf{u})\cdot\delta\mathbf{u} \tag{3.5}$$

where the symbolic derivative (or gradient) operator ∇ acts with respect to the dependent variable $\mathbf{u}$. With the above formula in mind, the first variation of $\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h})$, denote by $\mathcal{G}(\mathbf{u},\delta\mathbf{u})$, can be expressed as

$$\begin{aligned}\mathcal{G}(\mathbf{u},\delta\mathbf{u}) &= \delta\mathcal{J}(\mathbf{u},\delta\mathbf{u};\mathbf{f},\mathbf{h}) = \frac{d}{d\varepsilon}\mathcal{J}(\mathbf{u}+\varepsilon\delta\mathbf{u};\mathbf{f},\mathbf{h})\Big|_{\varepsilon=0}\\ &= (\delta\mathcal{L}(\mathbf{u},\delta\mathbf{u}),\mathcal{L}(\mathbf{u})-\mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}),g(\mathbf{u})-\mathbf{h})_{\Gamma^{\mathrm{N}},0}\\ &= (\nabla\mathcal{L}(\mathbf{u})\cdot\delta\mathbf{u},\mathcal{L}(\mathbf{u})-\mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}),g(\mathbf{u})-\mathbf{h})_{\Gamma^{\mathrm{N}},0} = 0\end{aligned} \tag{3.6}$$

where $\delta\mathbf{u}\in\mathcal{W}$ is an admissible variation of $\mathbf{u}$. The linear vector space of kinematically admissible variations $\mathcal{W}$ is of the form

$$\mathcal{W} = \left\{\delta\mathbf{u} : \delta\mathbf{u}\in\mathbf{H}^1(\Omega),\ \delta\mathbf{u}=\mathbf{0}\text{ on }\Gamma^{\mathrm{D}}\right\} \tag{3.7}$$

The Euler equation associated with the abstract least-squares problem is to find $\mathbf{u}\in\mathcal{V}$ such that Eq. (3.6) holds for all $\delta\mathbf{u}\in\mathcal{W}$.

When the governing equations for the physical system are linear, Eq. (3.6) can be expressed conveniently as

$$\mathcal{B}(\delta\mathbf{u},\mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \tag{3.8}$$

where the bilinear form $\mathcal{B}(\delta\mathbf{u},\mathbf{u})$ and linear form $\mathcal{F}(\delta\mathbf{u})$ are given as

$$\mathcal{B}(\delta\mathbf{u},\mathbf{u}) = (\mathcal{L}(\delta\mathbf{u}),\mathcal{L}(\mathbf{u}))_{\Omega,0} + (g(\delta\mathbf{u}),g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \tag{3.9a}$$

$$\mathcal{F}(\delta\mathbf{u}) = (\mathcal{L}(\delta\mathbf{u}),\mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}),\mathbf{h})_{\Gamma^{\mathrm{N}},0} \tag{3.9b}$$

We can associate with the least-squares based weak formulation a well defined energy

norm $\|\mathbf{u}\|_E$ of the form

$$\|\mathbf{u}\|_E = \sqrt{\mathcal{J}(\mathbf{u};\mathbf{0},\mathbf{0})} \tag{3.10}$$

If it can be shown that $\|\mathbf{u}\|_E$ is equivalent to the norm of a standard Sobolev space $\mathcal{X}$ (e.g., $\mathcal{X} = \mathbf{H}^1(\Omega)$) in the sense that $c_1\|\mathbf{u}\|_{\mathcal{X}}^2 \leq \|\mathbf{u}\|_E^2 \leq c_2\|\mathbf{u}\|_{\mathcal{X}}^2$ for all $\mathbf{u} \in \mathcal{V}$ where c_1 and c_2 are positive constants, then we say that $\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h})$ is $\mathcal{X}$-norm equivalent or $\mathcal{X}$-coercive. Under such conditions the least-squares method constitutes an ideal variational setting, and in particular optimal convergence rates under h-refinement may be established for a conforming finite element discretization [19]. Unfortunately, L_2-norm based least-squares functionals will not generally be $\mathcal{X}$-norm equivalent. However, as demonstrated by Pontaza [58], such functionals can still recover an optimal variational setting for least-squares finite element models whenever an appropriate level of p-refinement is employed (where $p \geq 4$ is typically sufficient).

Clearly, in the linear case the bilinear form $\mathcal{B}(\delta\mathbf{u},\mathbf{u})$ is symmetric irrespective of the particular form of $\mathcal{L}$. When the differential operator $\mathcal{L}$ is nonlinear, however, this is no longer the case. In the following sections we discuss procedures for linearizing the least-squares based weak formulation, both before and after functional minimization.

3. Linearization before minimization

Eq. (3.6) constitutes the proper setting for the least-squares variational formulation. However, in general, this approach yields a non-symmetric coefficient matrix for nonlinear problems. As a result, many authors choose to create a linearized version of $\mathcal{L}$ prior to construction and minimization of the least-squares functional. Two approaches are commonly advocated, namely the Picard method of successive substitution and the Newton scheme. In both approaches we replace the nonlinear operator

$\mathcal{L}$ with the linearized operator $\mathcal{L}_{\text{Lin}}$ defined as

$$\mathcal{L}(\mathbf{u}) \cong \mathcal{L}_{\text{Lin}}(\mathbf{u};\mathbf{u}_0) = \tilde{\mathcal{L}}(\mathbf{u};\mathbf{u}_0) + \hat{\mathcal{L}}(\mathbf{u}_0) \tag{3.11}$$

In the above expression $\mathbf{u}_0$ represents a characteristic state about which the solution is linearized. In the discrete numerical implementation $\mathbf{u}_0$ is taken either as a guess or as a known quantity from the immediate previous iteration. The linearized operator $\mathcal{L}_{\text{Lin}}$ is decomposed into the sum of $\tilde{\mathcal{L}}$, which is linear in $\mathbf{u}$, and $\hat{\mathcal{L}}$ which depends on $\mathbf{u}_0$ only.

In the Picard scheme $\mathcal{L}_{\text{Lin}}(\mathbf{u};\mathbf{u}_0)$ is replaced with $\mathcal{L}_{\text{Pic}}(\mathbf{u};\mathbf{u}_0)$, denoting a Picard linearization of $\mathcal{L}$. An important artifact of the Picard scheme is that $\mathcal{L}_{\text{Pic}}$ does not necessarily represent a unique linearization of $\mathcal{L}$, since a nonlinear operator can be linearized in more than one way. In Newton's method, the operator $\mathcal{L}_{\text{Lin}}(\mathbf{u};\mathbf{u}_0)$ is replaced with the Newton operator $\mathcal{L}_{\text{New}}(\mathbf{u};\mathbf{u}_0)$ defined as

$$\mathcal{L}_{\text{New}}(\mathbf{u};\mathbf{u}_0) = \mathcal{L}(\mathbf{u}_0) + \nabla\mathcal{L}(\mathbf{u}_0)\cdot(\mathbf{u}-\mathbf{u}_0) \tag{3.12}$$

where the gradient operator ∇ now acts with respect to $\mathbf{u}_0$. The abstract least-squares functional $\mathcal{J}(\mathbf{u};\mathbf{f},\mathbf{h})$ given in Eq. (3.2) is therefore replaced by the following approximation

$$\mathcal{J}_{\text{Lin}}(\mathbf{u};\mathbf{u}_0,\tilde{\mathbf{f}},\mathbf{h}) = \frac{1}{2}\Big(\|\tilde{\mathcal{L}}(\mathbf{u};\mathbf{u}_0)-\tilde{\mathbf{f}}\|^2_{\Omega,0} + \|g(\mathbf{u})-\mathbf{h}\|^2_{\Gamma^{\text{N}},0}\Big) \tag{3.13}$$

where $\tilde{\mathbf{f}} = \mathbf{f} - \hat{\mathcal{L}}(\mathbf{u}_0)$. The above expression has been defined in terms of the operators $\tilde{\mathcal{L}}$ and $\hat{\mathcal{L}}$ appearing in Eq. (3.11), and is thus applicable for use in the context of both the Picard and Newton methods. The Euler equation corresponding with minimization of $\mathcal{J}_{\text{Lin}}$ can be expressed as

$$\mathcal{G}_{\text{Lin}}(\mathbf{u},\delta\mathbf{u};\mathbf{u}_0) = (\tilde{\mathcal{L}}(\delta\mathbf{u};\mathbf{u}_0),\tilde{\mathcal{L}}(\mathbf{u};\mathbf{u}_0)-\tilde{\mathbf{f}})_{\Omega,0} + (g(\delta\mathbf{u}),g(\mathbf{u})-\mathbf{h})_{\Gamma^{\text{N}},0} \tag{3.14}$$

The above expression can be written equivalently as

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \tag{3.15}$$

where the bilinear form $\mathcal{B}(\delta\mathbf{u}, \mathbf{u})$ and linear form $\mathcal{F}(\delta\mathbf{u})$ are given as

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = (\tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0), \tilde{\mathcal{L}}(\mathbf{u}; \mathbf{u}_0))_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \tag{3.16a}$$

$$\mathcal{F}(\delta\mathbf{u}) = (\tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0), \mathbf{f} - \hat{\mathcal{L}}(\mathbf{u}_0))_{\Omega,0} + (g(\delta\mathbf{u}), \mathbf{h})_{\Gamma^{\mathrm{N}},0} \tag{3.16b}$$

The above forms apply to both the Picard and Newton linearization schemes. However, for the Newton scheme, the bilinear and linear forms can be reduced to

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = (\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}, \nabla\mathcal{L}(\mathbf{u}_0) \cdot \mathbf{u})_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \tag{3.17a}$$

$$\mathcal{F}(\delta\mathbf{u}) = (\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}, \mathbf{f} - \mathcal{L}(\mathbf{u}_0) + \nabla\mathcal{L}(\mathbf{u}_0) \cdot \mathbf{u}_0)_{\Omega,0} + (g(\delta\mathbf{u}), \mathbf{h})_{\Gamma^{\mathrm{N}},0} \tag{3.17b}$$

Clearly the bilinear form is symmetric and positive-definite regardless of which linearization scheme is employed. As a result fast and robust solution procedures may be employed in the actual finite element implementation (such as sparse forms of either the Cholesky decomposition or the preconditioned conjugate gradient method). We also note that when linearization is performed prior to minimization, we are able to associate with the linearized least-squares functional the following energy norm

$$\|\mathbf{u}\|_E = \sqrt{\mathcal{J}_{\mathrm{Lin}}(\mathbf{u}; \mathbf{u}_0, \mathbf{0}, \mathbf{0})} \tag{3.18}$$

4. Linearization after minimization

We now consider the case where construction and minimization of the least-squares functional is performed without first linearizing the governing partial differential equations. As a result, we work directly in terms of the *true* least-squares functional $\mathcal{J}(\mathbf{u}; \mathbf{f}, \mathbf{h})$ and the nonlinear Euler equation $\mathcal{G}(\mathbf{u}, \delta\mathbf{u})$ resulting from appropriate func-

tional minimization (see Eqs. (3.2) and (3.6)). An interesting implication of working in this setting is that we can no longer define an energy norm associated with the least-squares functional. We note that the Euler equation given by Eq. (3.6) is nonlinear in $\mathbf{u}$. This expression can be solved through the use of an appropriate iterative solution scheme, where linearization is fundamental to the iterative procedure. In the case of the Picard method, the Euler equation can be linearized as

$$\mathcal{G}_{\text{Pic}}(\mathbf{u}, \delta\mathbf{u}) \equiv (\mathcal{P}(\delta\mathbf{u}; \mathbf{u}_0), \mathcal{L}_{\text{Pic}}(\mathbf{u}; \mathbf{u}_0) - \mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}) - \mathbf{h})_{\Gamma^{\text{N}},0} \tag{3.19}$$

where $\mathcal{P}(\delta\mathbf{u}; \mathbf{u}_0)$ represents a Picard linearization of $\nabla\mathcal{L}(\mathbf{u}) \cdot \delta\mathbf{u}$. It is imperative to note that the Gâteaux variation and Picard linearization *do not commute* when applied to $\mathcal{L}(\mathbf{u})$. As a result, linearization prior to minimization is clearly not equivalent to linearization after minimization when the Picard scheme is employed. We note that it is always possible to decompose $\mathcal{P}(\delta\mathbf{u}; \mathbf{u}_0)$ into the following sum

$$\mathcal{P}(\delta\mathbf{u}; \mathbf{u}_0) = \tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0) + \mathcal{L}_{\text{Add}}(\delta\mathbf{u}; \mathbf{u}_0) \tag{3.20}$$

The quantity $\mathcal{L}_{\text{Add}}(\delta\mathbf{u}; \mathbf{u}_0)$ represents an additional term present in the Picard linearization of $\nabla\mathcal{L}(\mathbf{u}) \cdot \delta\mathbf{u}$ that is not accounted for when linearization is performed prior to minimization. The Picard linearization of the nonlinear Euler equation can therefore be expressed as

$$\begin{aligned} \mathcal{G}_{\text{Pic}}(\mathbf{u}, \delta\mathbf{u}) = (\tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0), \mathcal{L}_{\text{Pic}}(\mathbf{u}; \mathbf{u}_0) - \mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}) - \mathbf{h})_{\Gamma^{\text{N}},0} \\ + \underline{(\mathcal{L}_{\text{Add}}(\delta\mathbf{u}; \mathbf{u}_0), \mathcal{L}_{\text{Pic}}(\mathbf{u}; \mathbf{u}_0) - \mathbf{f})_{\Omega,0}} \end{aligned} \tag{3.21}$$

The underlined term in the above expression is not present when linearization is performed prior to minimization. The linearized Euler equation can also be written as

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) + \tilde{\mathcal{B}}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \tag{3.22}$$

where $\mathcal{B}(\delta\mathbf{u}, \mathbf{u})$ is a bilinear form, $\tilde{\mathcal{B}}(\delta\mathbf{u}, \mathbf{u})$ is a mixed form (i.e., it contains both bilinear and linear forms) and $\mathcal{F}(\delta\mathbf{u})$ is a linear form. These quantities are defined as

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = (\tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0), \tilde{\mathcal{L}}(\mathbf{u}; \mathbf{u}_0))_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \tag{3.23a}$$

$$\tilde{\mathcal{B}}(\delta\mathbf{u}, \mathbf{u}) = \underline{(\mathcal{L}_{\text{ Add}}(\delta\mathbf{u}; \mathbf{u}_0), \mathcal{L}_{\mathrm{Pic}}(\mathbf{u}; \mathbf{u}_0) - \mathbf{f})_{\Omega,0}} \tag{3.23b}$$

$$\mathcal{F}(\delta\mathbf{u}) = -(\tilde{\mathcal{L}}(\delta\mathbf{u}; \mathbf{u}_0), \hat{\mathcal{L}}(\mathbf{u}_0) - \mathbf{f})_{\Omega,0} + (g(\delta\mathbf{u}), \mathbf{h})_{\Gamma^{\mathrm{N}},0} \tag{3.23c}$$

We have employed the decomposition of $\mathcal{L}_{\mathrm{Pic}}$ defined in Eq. (3.11) in arriving at the above expressions for $\mathcal{B}(\delta\mathbf{u}, \mathbf{u})$ and $\mathcal{F}(\delta\mathbf{u})$.

It should be clear that the Picard linearization of the nonlinear Euler equation differs from Eq. (3.15) only on account of $\tilde{\mathcal{B}}(\delta\mathbf{u}, \mathbf{u})$. It readily follows that this mixed form is identically zero, for all $\delta\mathbf{u} \in \mathcal{W}$, whenever the true solution for $\mathbf{u}$ is inserted into Eq. (3.22) (assuming of course that $\mathbf{u}_0 \to \mathbf{u}$). In the finite element implementation, however, we replace $\mathbf{u}$ and $\delta\mathbf{u}$ with their discrete counterparts $\mathbf{u}_{hp} \in \mathcal{V}^{hp}$ and $\delta\mathbf{u}_{hp} \in \mathcal{W}^{hp}$, and as a result $\tilde{\mathcal{B}}(\delta\mathbf{u}_{hp}, \mathbf{u}_{hp}) \neq 0$. Consequently, the finite element solution obtained using Eq. (3.15) will not exactly coincide with the numerical results procured via Eq. (3.22). Clearly, application of Picard's linearization scheme prior to minimization induces error in the numerical solution (as compared to the case where minimization is performed first). We do expect, however, that under proper mesh refinement, the error induced by neglecting the discrete mixed term will diminish. A major shortcoming of applying the Picard method after minimization is that we can no longer guarantee symmetry of the resulting bilinear form.

Eq. (3.6) can also be linearized using Newton's method. In this case $\mathcal{G}(\mathbf{u}, \delta\mathbf{u})$ is replaced with

$$\mathcal{G}_{\mathrm{New}}(\Delta\mathbf{u}, \delta\mathbf{u}; \mathbf{u}_0) \equiv \mathcal{G}(\mathbf{u}_0, \delta\mathbf{u}) + D\mathcal{G}(\mathbf{u}_0, \delta\mathbf{u})[\Delta\mathbf{u}] \tag{3.24}$$

where $\Delta\mathbf{u} = \mathbf{u} - \mathbf{u}_0$ and $D\mathcal{G}(\mathbf{u}_0, \delta\mathbf{u})[\Delta\mathbf{u}]$ is the tangent operator defined as

$$\begin{aligned} D\mathcal{G}(\mathbf{u}_0, \delta\mathbf{u})[\Delta\mathbf{u}] &= \nabla\mathcal{G}(\mathbf{u}_0, \delta\mathbf{u}) \cdot \Delta\mathbf{u} \\ &= (\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}, \nabla\mathcal{L}(\mathbf{u}_0) \cdot \Delta\mathbf{u})_{\Omega,0} + (g(\delta\mathbf{u}), g(\Delta\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \\ &\quad + (\nabla(\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}) \cdot \Delta\mathbf{u}, \mathcal{L}(\mathbf{u}_0) - \mathbf{f})_{\Omega,0} \end{aligned} \tag{3.25}$$

Since $\delta\mathbf{u}$ remains constant during the increment $\Delta\mathbf{u}$, it follows that

$$\nabla(\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}) \cdot \Delta\mathbf{u} = (\nabla\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}) \cdot \Delta\mathbf{u} = (\nabla\nabla\mathcal{L}(\mathbf{u}_0) \cdot \Delta\mathbf{u}) \cdot \delta\mathbf{u} \tag{3.26}$$

The above implies that the tangent operator is symmetric. As a result, the expression $\mathcal{G}_{\mathrm{New}}(\Delta\mathbf{u}, \delta\mathbf{u}; \mathbf{u}_0) = 0$ can be written in the usual manner

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \tag{3.27}$$

where the bilinear form $\mathcal{B}(\delta\mathbf{u}, \mathbf{u})$ and linear form $\mathcal{F}(\delta\mathbf{u})$ are given as

$$\begin{aligned} \mathcal{B}(\delta\mathbf{u}, \mathbf{u}) &= (\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}, \nabla\mathcal{L}(\mathbf{u}_0) \cdot \mathbf{u})_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \\ &\quad + \underline{((\nabla\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}) \cdot \mathbf{u}, \mathcal{L}(\mathbf{u}_0) - \mathbf{f})_{\Omega,0}} \end{aligned} \tag{3.28a}$$

$$\begin{aligned} \mathcal{F}(\delta\mathbf{u}) &= (\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}, \mathbf{f} - \mathcal{L}(\mathbf{u}_0) + \nabla\mathcal{L}(\mathbf{u}_0) \cdot \mathbf{u}_0)_{\Omega,0} + (g(\delta\mathbf{u}), \mathbf{h})_{\Gamma^{\mathrm{N}},0} \\ &\quad + \underline{((\nabla\nabla\mathcal{L}(\mathbf{u}_0) \cdot \delta\mathbf{u}) \cdot \mathbf{u}_0, \mathcal{L}(\mathbf{u}_0) - \mathbf{f})_{\Omega,0}} \end{aligned} \tag{3.28b}$$

It is interesting to note that symmetry in the bilinear form is guaranteed when linearization is performed subsequent to minimization, if the linearization is employed in the framework of an iterative Newton solution procedure. As a result, the coefficient matrix of the finite element model will always be symmetric, even when the first variation of $\mathcal{J}(\mathbf{u}; \mathbf{f}, \mathbf{h})$ is not.

When the underlined terms in the above expression are neglected, the scheme is equivalent to performing linearization prior to minimization. Justification for such

omission can be seen by rewriting the linearized equations above as

$$\tilde{\mathcal{B}}(\delta\mathbf{u},\mathbf{u}) + \hat{\mathcal{B}}(\delta\mathbf{u},\mathbf{u}) = \tilde{\mathcal{F}}(\delta\mathbf{u}) \tag{3.29}$$

where $\tilde{\mathcal{B}}(\delta\mathbf{u},\mathbf{u})$ is a bilinear form, $\hat{\mathcal{B}}(\delta\mathbf{u},\mathbf{u})$ is a mixed form and $\tilde{\mathcal{F}}(\delta\mathbf{u})$ is a linear form. We define these quantities as

$$\tilde{\mathcal{B}}(\delta\mathbf{u},\mathbf{u}) = (\nabla\mathcal{L}(\mathbf{u}_0)\cdot\delta\mathbf{u}, \nabla\mathcal{L}(\mathbf{u}_0)\cdot\mathbf{u})_{\Omega,0} + (g(\delta\mathbf{u}), g(\mathbf{u}))_{\Gamma^{\mathrm{N}},0} \tag{3.30a}$$

$$\hat{\mathcal{B}}(\delta\mathbf{u},\mathbf{u}) = ((\nabla\nabla\mathcal{L}(\mathbf{u}_0)\cdot\delta\mathbf{u})\cdot(\mathbf{u}-\mathbf{u}_0), \mathcal{L}(\mathbf{u}_0)-\mathbf{f})_{\Omega,0} \tag{3.30b}$$

$$\tilde{\mathcal{F}}(\delta\mathbf{u}) = (\nabla\mathcal{L}(\mathbf{u}_0)\cdot\delta\mathbf{u}, \mathbf{f}-\mathcal{L}(\mathbf{u}_0)+\nabla\mathcal{L}(\mathbf{u}_0)\cdot\mathbf{u}_0)_{\Omega,0} + (g(\delta\mathbf{u}),\mathbf{h})_{\Gamma^{\mathrm{N}},0} \tag{3.30c}$$

We can clearly see that as $\mathbf{u}_0 \to \mathbf{u}$, the mixed form $\hat{\mathcal{B}}(\delta\mathbf{u},\mathbf{u})$ goes to zero. The same is also true in the discrete case (i.e., $\hat{\mathcal{B}}(\delta\mathbf{u}_{hp},\mathbf{u}_{hp}) \to 0$ as $\mathbf{u}_0 \to \mathbf{u}_{hp}$). We therefore conclude that application of Newton's method prior to minimization of the least-squares functional is equivalent to modifying the search direction of the Newton scheme as applied after minimization. As a result, we expect both schemes to yield the same solution, provided that the initial guess is such that convergence is possible.

5. A simple analysis of Newton's method

In this section we seek to gain a deeper understanding of the iterative solution process associated with application of Newton's method in least-squares formulations (both before and after minimization of the least-squares functional). To simplify the mathematical analysis, we consider application of the least-squares principle to the problem of finding the simple root of a nonlinear differentiable function $f(x)$ defined on the interval $I = [a,b]$. To this end we define the *true* least-squares functional as

$$\mathcal{J}(x) = \frac{1}{2}f(x)^2 \tag{3.31}$$

In the least-squares formulation of the problem we seek to find $\alpha \in I$ such that $\mathcal{J}(\alpha) \leqslant \mathcal{J}(x)$ for all $x \in I$. Invoking the minimization principle yields the nonlinear Euler equation which may be expressed as

$$\mathcal{G}(x, \delta x) = f'(x) f(x) \delta x \tag{3.32}$$

where $f'(x)$ is the derivative of f with respect to x. Linearization of the Euler equation using Newton's method yields

$$\mathcal{B}(\delta x, x) = \mathcal{F}(\delta x) \tag{3.33}$$

where the bilinear form $\mathcal{B}(\delta x, x)$ and linear form $\mathcal{F}(\delta x)$ are given as

$$\mathcal{B}(\delta x, x) = [f'(x_0)^2 + \underline{f''(x_0) f(x_0)}] x \delta x \tag{3.34a}$$

$$\mathcal{F}(\delta x) = [f'(x_0)(-f(x_0) + f'(x_0) x_0) + \underline{f''(x_0) f(x_0) x_0}] \delta x \tag{3.34b}$$

The underlined terms are absent in the event that Newton's method is applied prior to minimization of $\mathcal{J}(x)$. Solving the above expression for x yields

$$x = x_0 - \frac{f'(x_0) f(x_0)}{f'(x_0)^2 + \gamma \underline{f''(x_0) f(x_0)}} \tag{3.35}$$

where $\gamma = 1.0$.

It is interesting to note that application of Newton's method to the *true* least-squares problem is equivalent to applying Newton's method in the solution of $h(x) = g(x) f(x) = 0$ where $g(x) = f'(x)$. It can be easily shown, however, that when the least-squares functional is defined in terms of a Newton linearization of $f(x)$, the least-squares problem is completely equivalent to simply finding the solution of $f(x) = 0$ using Newton's method. Clearly, the convergence properties associated with each Newton linearization scheme will be distinct. It has been shown by Gerlach

[59] that it is possible to *define* $g(x)$ such that application of Newton's method to $h(x) = g(x)f(x) = 0$ will possess superior convergence properties than a direct application of Newton's procedure to $f(x) = 0$. In such formulations, $g(x)$ is typically defined such that $h''(x)$ (or even higher derivatives of $h(x)$) goes to zero in the vicinity of α. For example if $g(x)$ is *defined* as $g(x) = 1/\sqrt{|f'(x)|}$, Newton's procedure produces an algorithm with a cubic convergence rate (also known as Halley's method [59]). Such a procedure may be obtained by artificially setting γ to $-1/2$ in Eq. (3.35). Since Eq. (3.35) is neither Newton's procedure nor Halley's method (as applied in the solution of $f(x) = 0$) it is only reasonable to be cautious when applying this iterative scheme to the solution of nonlinear equations. Clearly, linearization after minimization will tend to increase the nonlinearity of the resulting Euler equation. It therefore seems plausible to assume that such an increase may tend to negatively affect the radius of convergence of the iterative solution procedure.

Although a rigorous mathematical analysis on the convergence behavior of the two Newton formulations is beyond the scope of this dissertation, we offer the following by way of simple analysis: We invoke Taylor's theorem, which may be applied in the exact evaluation of a differentiable function $h(x)$ at its root α in the vicinity of a characteristic state x_n as

$$h(\alpha) = 0 = h(x_n) + h'(x_n)(\alpha - x_n) + \frac{h''(\xi)}{2}(\alpha - x_n)^2 \tag{3.36}$$

where ξ is between α and x_n. Combing the above expression with Newton's scheme (i.e., $x_{n+1} = x_n - h(x_n)/h'(x_n)$) and taking the absolute value yields

$$|\alpha - x_{n+1}| = \frac{|h''(\xi)|}{2|h'(x_n)|}(\alpha - x_n)^2 \tag{3.37}$$

The above expression is typically utilized to prove the quadratic convergence rate of Newton's method as applied in the solution of $h(x) = 0$. If we apply the above

expression in the evaluation of $h(x) = f(x)$ and $h(x) = f'(x)f(x)$, we obtain

$$|\alpha - \bar{x}_{n+1}| = \bar{C}(\bar{\xi}, x_n)(\alpha - x_n)^2 \tag{3.38a}$$

$$|\alpha - \tilde{x}_{n+1}| = \tilde{C}(\tilde{\xi}, x_n)(\alpha - x_n)^2 \tag{3.38b}$$

where $\bar{C}(\bar{\xi}, x_n)$ and $\tilde{C}(\tilde{\xi}, x_n)$ are of the form

$$\bar{C}(\bar{\xi}, x_n) = \frac{|f''(\bar{\xi})|}{2|f'(x_n)|} \tag{3.39a}$$

$$\tilde{C}(\tilde{\xi}, x_n) = \frac{|3f'(\tilde{\xi})f''(\tilde{\xi}) + f'''(\tilde{\xi})f(\tilde{\xi})|}{2|f'(x_n)^2 + f''(x_n)f(x_n)|} \tag{3.39b}$$

It is important to note that the above equations have been defined such that the characteristic state x_n is the same in both expressions. The updated quantities $\bar{x}_{n+1}$ and $\tilde{x}_{n+1}$ will of course be distinct. In general it is difficult to assess how $\bar{C}(\bar{\xi}, x_n)$ compares with $\tilde{C}(\tilde{\xi}, x_n)$. We consider two limit cases. In the case that ($f''(\bar{\xi}) \to 0$ and $f''(\tilde{\xi}) \to 0$), $\bar{C}(\bar{\xi}, x_n)$ also tends to zero while the value of $\tilde{C}(\tilde{\xi}, x_n)$ depends on $|f'''(\tilde{\xi})f(\tilde{\xi})|/(2|f'(x_n)^2 + f''(x_n)f(x_n)|)$. Also, in the limit as $|\alpha - x_n| \to 0$, it is obvious that $\tilde{C} \to 3\bar{C}$. In both of these limit cases, linearization prior to minimization appears to produce slightly superior convergence properties than does linearization after minimization of $\mathcal{J}(x)$ (although both schemes have quadratic convergence rates). However, general convergence behavior cannot be obtained through a simple extrapolation of these limit cases.

The above discussion has been restricted to the problem of finding the simple root of a nonlinear differentiable function of a single variable. Clearly, blind extrapolation of the characteristics observed in this simple problem to more general classes of problems involving the least-squares method is tenuous. However, the qualities observed in this simple problem are consistent with our numerical findings in the least-squares finite element analysis of nonlinear boundary-value problems. As

demonstrated numerically in Section E, application of Newton's method (after linearization) typically results in a solution procedure possessing: (a) a small radius of convergence (as compared with Newton's method applied prior to minimization) and (b) a rate of convergence similar to Newton's procedure (as applied before minimization), assuming of course that solution convergence is possible.

6. General remarks on abstract least-squares problem

At this point, it is worthwhile to summarize some of the more pertinent qualities associated with the abstract least-squares formulation. In particular, the following observations and conclusions can be drawn:

1. The roles of functional minimization and nonlinear operator linearization have distinct and separate purposes. In general, these operations *do not* commute.

2. Minimization of the least-squares functional is the fundamental variational principle upon which the least-squares finite element model is predicated. The principle is independent of the iterative nonlinear solution procedure and the discretization scheme, and as a result should be applied to the *true* least-squares functional constructed from the governing equations associated with a given physical system. If the model problem is nonlinear, the least-squares functional should be constructed and minimized without first introducing any linearizing assumptions. In the continuous setting, the minimizer of the least-squares functional coincides with the true solution of the governing partial differential equations.

3. Linearization is merely a means of facilitating the solution of the nonlinear Euler equation, and may be employed in conjunction with an appropriate iterative solution scheme (e.g., the methods of Picard or Newton).

4. In general, when the nonlinear operator is linearized prior to construction of the least-squares functional, the *discrete* minimizer is associated with the linearized least-squares functional as opposed to the proper or *true* least-squares functional (constructed in terms of the nonlinear operator). Hence, minimization of the *linearized* least-squares functional is *not equivalent* to minimization of the *true* least-squares functional.

5. It is often computationally advantageous to linearize the partial differential equations prior to construction and minimization of the least-squares functional. The coefficient matrix components are simpler to formulate and faster to compute when linearization is applied prior to minimization. The Picard scheme produces the simplest form for the coefficient matrix and its use is especially convenient when complicated constitutive models are involved. The Newton scheme, on the other hand yields a slightly more complicated expression for the coefficient matrix. However, both methods as applied prior to minimization guarantee symmetry and positive-definiteness of the resulting global coefficient matrix. Due in part to the formulative and computational simplicity, performing linearization prior to minimization is the preferred approach of most researchers.

6. In the discrete setting, application of the Picard method prior to minimization introduces error into the resulting Euler equation (as compared to performing linearization after invoking the minimization principle). This error is proportional to how well the least-squares variational boundary-value problem is satisfied by the finite element approximation of the weak solution. We expect that the magnitude and effect of this error to diminish when proper mesh refinement is employed. This expectation is indeed realized in the numerical results presented later in this chapter.

7. The Euler equation resulting from minimization of the *true* least-squares functional is in general non-symmetric in its bilinear form. A Picard linearization of this expression will also, in general, yield a non-symmetric bilinear form. Applying the Newton scheme in the solution of the nonlinear Euler equation always yields a symmetric bilinear form, regardless of the form taken by $\mathcal{L}$. The resulting bilinear form, however, will not necessarily be positive-definite.

8. The Newton scheme can be interpreted geometrically as a numerical solution procedure for a nonlinear set of equations that relies on the concept of a search direction (given by the tangent operator). The abstract formulation reveals that linearization using Newton's method prior to minimization introduces an error in the search direction of the scheme, as compared with application of Newton's approach after minimization of the least-squares functional. We expect, however, that both Newton schemes will yield identical discrete numerical results, assuming of course that the characteristic states are initially chosen such that solution convergence is possible. Hence, exchanging the order of application of the minimization and Newton linearization operations has no effect on the *converged* numerical results (i.e., the discrete minimizers coincide). As we will demonstrate numerically, however, the order of application of these operators significantly affects the radii and rates of convergence of a given numerical implementation. Bell and Surana [42, 43] correctly recognized that the expression in the tangent operator, given in Eq. (3.26), for the Newton case (after minimization) may be neglected without affecting the final state of the converged solution. Such omission is equivalent to linearizing the governing equations using Newton's method prior to minimization.

D. Numerical implementation of the least-squares method

The procedures outlined in Chapter II may be used to produce the least-squares finite element model of a given nonlinear first-order system of partial differential equations. This is accomplished by replacing the function spaces $\mathcal{V}$ and $\mathcal{W}$ with the finite dimensional sub-spaces $\mathcal{V}^{hp} \subset \mathcal{V}$ and $\mathcal{W}^{hp} \subset \mathcal{W}$ associated with a given high-order spectral/hp finite element discretization. This naturally leads to a set of nonlinear algebraic equations that can be expressed as

$$[K(\{\Delta\}^{(k-1)})]\{\Delta\}^{(k)} = \{F\}^{(k)} \tag{3.40}$$

where the index k denotes the iteration number of the nonlinear iterative solution procedure. The coefficient matrix $[K(\{\Delta\}^{(k-1)})]$ and force vector $\{F\}^{(k)}$ are constructed from the bilinear and linear forms respectively of the given least-squares based weak formulation. We adopt the Gauss-Legendre quadrature rules (see Chapter II) in the numerical evaluation of all integrals appearing in the finite element equations. For nonlinear least-squares based finite element implementations, the order of the quadrature rule plays a critical role on the reliability of the resulting numerical solution. For a fine discretization, it is our experience that a quadrature rule of NGP $= p + 1$ is typically sufficient to produce dependable numerical results. For coarse meshes, however, it is typically expedient to employ higher order quadrature formulas (e.g., $p + 1 <$ NGP $\leq p + 5$) to insure that the integrity of the numerical solution is not polluted by errors associated with the numerical integration scheme.

It is important to note that direct application of nonlinear least-squares finite element models with linearization performed after minimization, often yield iterative solution schemes that diverge. To restore convergence for these cases, we employ a relaxation scheme in the numerical solution. To this end we call $\{\Delta\}^{(*)}$ the solution

calculated at the kth iteration, and let $\{\Delta\}^{(k-1)}$ denote the solution known from the previous iteration (i.e., iteration $k-1$). We define the modified current solution $\{\Delta\}^{(k)}$ as

$$\{\Delta\}^{(k)} = \omega(\{\Delta\}^{(*)} - \{\Delta\}^{(k-1)}) + \{\Delta\}^{(k-1)} \tag{3.41}$$

where $\omega \in \mathbb{R}^+$ is the relaxation parameter. We note that when $\omega = 1.0$, $\{\Delta\}^{(k)} = \{\Delta\}^{(*)}$. We postulate the following simple expression for the relaxation parameter

$$\omega(\varepsilon) = 1 - (1 - \omega_0)(2 - \varepsilon)\varepsilon \tag{3.42}$$

where ω_0 is a constant specified by the user and ε represents the error in the iterative solution

$$\varepsilon = \frac{\|\{\Delta\}^{(*)} - \{\Delta\}^{(k-1)}\|}{\|\{\Delta\}^{(*)}\|} \tag{3.43}$$

In the above expression $\|\cdot\|$ is the Euclidean norm. Eq. (3.42) has been specifically designed such that $\omega \to 1.0$ as $\varepsilon \to 0$.

It is important to be able to estimate how well the numerical solution approximates the exact solution for a particular problem. When the exact solution is known, the following expression represents an appropriate error measure of the numerical solution

$$\mathcal{E}_u = \|u - u_{hp}\|_{\Omega,0} / \|u\|_{\Omega,0} \tag{3.44}$$

where u and u_{hp} represent respectively, the exact and finite element solution for a given dependent variable. When the exact solution is unavailable we utilize an *a posteriori* evaluation of the least-squares functional $\mathcal{J}(\mathbf{u}_{hp}; \mathbf{f}, \mathbf{h})$ as an estimate for the error in the finite element solution. We note that this approach represents a global error estimate of the sum of all solution variables. Alternatively, $\mathcal{J}(\mathbf{u}_{hp}; \mathbf{f}, \mathbf{h})$ may also be evaluated for each element separately as a means of identifying regions where mesh refinement may be necessary.

E. Numerical examples: verification benchmarks

In what follows, we apply the preceding abstract formulations to the numerical solution of several nonlinear boundary-value problems using least-squares finite element models. It is our intent to illustrate through these numerical experiments, various artifacts associated with how the numerical solution is affected by exchanging the order of application of the minimization and linearization operators. To this end we first consider the numerical solution of a two-dimensional nonlinear Poisson equation. We then present numerical solutions of the stationary incompressible form of the Navier-Stokes equations including flow past a circular cylinder, flow over a backward facing step and lid-driven cavity flow.

1. A nonlinear Poisson equation

We consider the solution of a nonlinear Poisson equation, governing the diffusion of heat or chemical species. The problem may be formally stated as follows: find $u(\mathbf{x})$ such that

$$-\nabla \cdot (k \nabla u) = f \qquad \text{in } \Omega \tag{3.45a}$$

$$u = u^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{D}} \tag{3.45b}$$

$$-\hat{\mathbf{n}} \cdot k \nabla u = q^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{N}} \tag{3.45c}$$

where $\hat{\mathbf{n}}$ is the outward unit normal and k is a nonlinear function of u given by the following formula

$$k = k_0 + k_u u > 0 \tag{3.46}$$

where k_0 and k_u are constants. We assume the boundary conditions and data are given such that the problem is well-posed. Direct application of the least-squares method to this boundary-value problem is certainly possible. However, this will require a high

degree of regularity in the finite element solution such as $u \in H^2(\Omega)$. To maintain practicality in the numerical implementation, we recast the above problem into an equivalent first-order form. The problem can be stated as follows: find u and $\mathbf{q}$ such that

$$\nabla \cdot \mathbf{q} - f = 0 \qquad \text{in } \Omega \tag{3.47a}$$

$$\mathbf{q} + k\nabla u = \mathbf{0} \qquad \text{in } \Omega \tag{3.47b}$$

$$\nabla \times \mathbf{q} = \mathbf{0} \qquad \text{in } \Omega \tag{3.47c}$$

$$u = u^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{D}} \tag{3.47d}$$

$$\hat{\mathbf{n}} \cdot \mathbf{q} = q^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{N}} \tag{3.47e}$$

where $\mathbf{q}$ is a flux quantity (e.g., in the context of the heat equation, $\mathbf{q}$ is the heat flux). The first-order system in Eq. (3.47) involves physical variables, although one may select an alternative set of first-order equations that may not be physical.

We define the *true* least-squares functional for the nonlinear Poisson equation as

$$\mathcal{J}(u, \mathbf{q}; f) = \frac{1}{2} \int_{\Omega} \{(\nabla \cdot \mathbf{q} - f)^2 + [\mathbf{q} + (k_0 + k_u u)\nabla u]^2 + (\nabla \times \mathbf{q})^2\} d\Omega \tag{3.48}$$

The Neumann boundary condition may also be included, if needed, in the definition of the least-squares functional. It is also possible to construct a linearized form of the above expression by performing linearization of the nonlinear parts of the Poisson equation prior to construction of the least-squares functional. In the Picard approach we replace $u\nabla u$ with $u_0 \nabla u$. Likewise, in Newton's scheme, $u\nabla u_0 + u_0 \nabla u - u_0 \nabla u_0$ is used in substitution of the nonlinear term. The linearized Euler equation (or weak formulation) resulting from invoking the minimization principle for the nonlinear Poisson equation may be stated as follows: find $\mathbf{u} \in \mathcal{V}$ such that

$$\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \quad \text{for all } \delta\mathbf{u} \in \mathcal{W} \tag{3.49}$$

where $\mathbf{u} = (u, \mathbf{q})$ and $\delta\mathbf{u} = (\delta u, \delta\mathbf{q})$ are ordered pairs, introduced to simplify the discussion. Likewise, $\mathcal{V}$ and $\mathcal{W}$ are appropriate function spaces (see for example Eqs. (3.4) and (3.7)). In the case of Picard linearization, the bilinear form $\mathcal{B}(\delta\mathbf{u}, \mathbf{u})$ and linear form $\mathcal{F}(\delta\mathbf{u})$ are given as

$$\begin{aligned}\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = \int_\Omega \Big\{ &[\delta\mathbf{q} + (k_0 + k_u u_0)\nabla\delta u] \cdot [\mathbf{q} + (k_0 + k_u u_0)\nabla u] \\ &+ (\nabla \cdot \delta\mathbf{q})(\nabla \cdot \mathbf{q}) + (\nabla \times \delta\mathbf{q}) \cdot (\nabla \times \mathbf{q}) \\ &+ \underline{k_u \delta u \nabla u_0 \cdot [\mathbf{q} + (k_0 + k_u u_0)\nabla u]} \Big\} d\Omega \end{aligned} \tag{3.50a}$$

$$\mathcal{F}(\delta\mathbf{u}) = \int_\Omega f(\nabla \cdot \delta\mathbf{q}) d\Omega \tag{3.50b}$$

Likewise, for the case of Newton linearization the bilinear and linear forms can be expressed as

$$\begin{aligned}\mathcal{B}(\delta\mathbf{u}, \mathbf{u}) = \int_\Omega \Big\{ &[\delta\mathbf{q} + (k_0 + k_u u_0)\nabla\delta u + k_u \delta u \nabla u_0] \cdot [\mathbf{q} + (k_0 + k_u u_0)\nabla u \\ &+ k_u u \nabla u_0] + (\nabla \cdot \delta\mathbf{q})(\nabla \cdot \mathbf{q}) + (\nabla \times \delta\mathbf{q}) \cdot (\nabla \times \mathbf{q}) \\ &+ \underline{k_u (\delta u \nabla u + u \nabla \delta u) \cdot [\mathbf{q}_0 + (k_0 + k_u u_0)\nabla u_0]} \Big\} d\Omega \end{aligned} \tag{3.51a}$$

$$\begin{aligned}\mathcal{F}(\delta\mathbf{u}) = \int_\Omega \Big\{ &f(\nabla \cdot \delta\mathbf{q}) + (k_u u_0 \nabla u_0) \cdot [\delta\mathbf{q} + (k_0 + k_u u_0)\nabla\delta u + k_u \delta u \nabla u_0] \\ &+ \underline{k_u (\delta u \nabla u_0 + u_0 \nabla \delta u) \cdot [\mathbf{q}_0 + (k_0 + k_u u_0)\nabla u_0]} \Big\} d\Omega \end{aligned} \tag{3.51b}$$

The underlined terms above are present when linearization is performed after minimization. The above expressions are consistent with our findings for the abstract problem. In particular, the converged solutions for both Newton schemes should coincide, while the Picard linearization (before minimization) introduces error in the discrete setting that is proportional to how well the governing equations are satisfied by the numerical solution.

a. A manufactured solution

To demonstrate convergence properties of the various least-squares formulations for the nonlinear Poisson equation, we seek to compare numerical results with an appropriate analytic solution. Since obtaining exact solutions of nonlinear equations is often a formidable task, we resort to the method of manufactured solutions. In this approach we postulate a solution to the partial differential equation, and then find a forcing function f that makes the solution exact. Ideally we would like to come up with a solution that is infinitely differentiable, to insure that the numerical solution cannot be trivially satisfied by the finite element basis functions.

We consider the domain for the problem to be given as $\bar{\Omega} = [0, \sqrt{7\pi/2}] \times [0, 1]$. The closed form analytic solution chosen for the problem is of the form

$$u(x, y) = [e^x \cos(x^2) + \pi^2 x] \sin(\pi y) \tag{3.52}$$

The equation parameters for k are taken as $k_0 = 1$ and $k_u = 100$. We note that the proposed solution does not satisfy the Poisson equation when $f = 0$. To make the solution exact we specify f such that Eq. (3.45a) is satisfied. This choice for f is sometimes termed the consistent forcing function. The expression is somewhat complicated but can be easily determined using a symbolic algebra software package such as Maple or Mathematica.

In the numerical implementation of the problem, the boundary conditions are applied by specifying the exact solution for u along the whole boundary (i.e., $\Gamma = \Gamma^{\mathrm{D}}$ and $\Gamma^{\mathrm{N}} = \varnothing$). The following values are utilized as initial guesses at the beginning of the iterative nonlinear solution procedure: $u = 15$ and $q_x = q_y = 0$ in Ω. The finite element model consists of a uniform 6×2 mesh of rectangular elements. The mesh is refined by systematically increasing the p-level of the finite element approximation

functions. Nonlinear convergence is declared once $\varepsilon < 10^{-6}$. An example mesh and accompanying finite element solution of the Poisson equation is given in Figure 8.

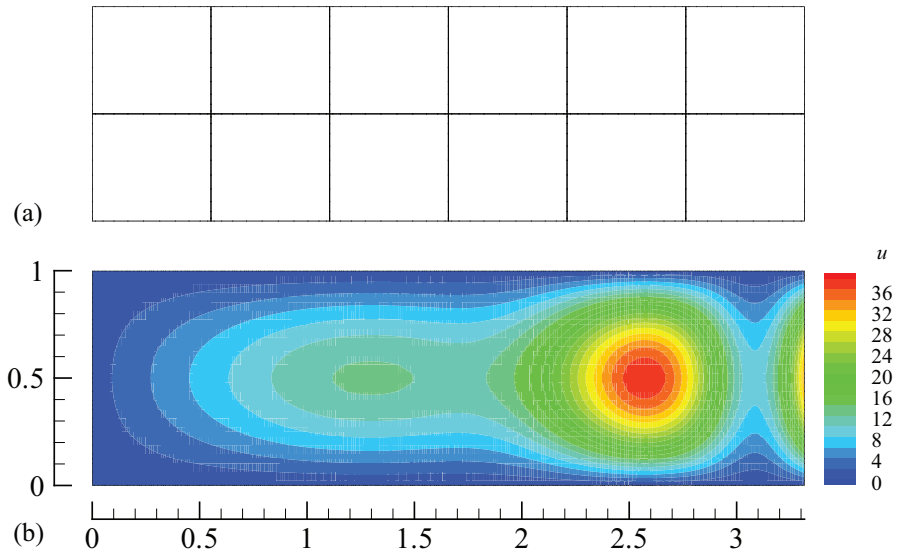


Fig. 8. Nonlinear Poisson equation: (a) 6×2 finite element mesh and (b) finite element solution for u at $p = 11$.

The solution error $\mathcal{E}_u$ associated with u_{hp} is evaluated using the pseudo-metric given in Eq. (3.44). As expected, exponential decay of $\mathcal{E}_u$ is observed under p-refinement as can be seen in Figure 9; this is true of all four least-squares based formulations. It is worth noting that all formulations with the exception of the Picard method (with linearization performed prior to minimization) yield equivalent values for $\mathcal{E}_u$ for a given p-level. We also, see that this Picard scheme also yields the largest value of $\mathcal{E}_u$. Clearly, the error inherent in the Picard scheme (before minimization) does not prevent solution convergence under mesh refinement. It does, however, prevent this scheme from being as competitive with the other least-squares finite element formulations. In addition, we see that application of Newton's method (before minimization) yields identical numerical results as compared with the two schemes

where linearization is performed subsequent to minimization. These conclusions are consistent with our findings from the abstract least-squares problem.

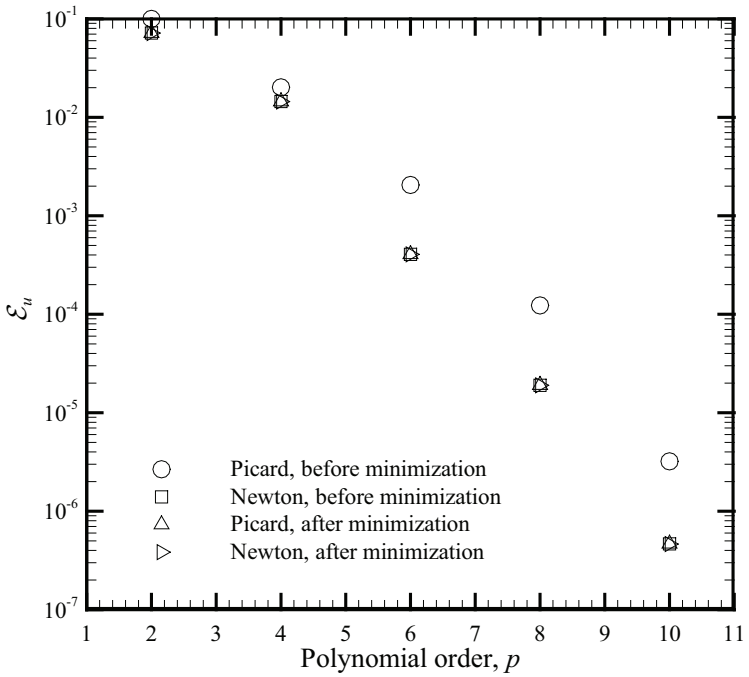


Fig. 9. Convergence of u_{hp} as measured by $\mathcal{E}_u$ for various least-squares formulations of the nonlinear Poisson equation (uniform 6×2 mesh with p-refinement).

In Figure 10 we also plot the solution for u along the horizontal centerline (i.e., $y = 0.5$) of the domain for $p = 2$ and $p = 6$. We compare the results obtained from the finite element solutions (i.e., the Newton (before minimization) and Picard (before minimization)) with the exact solution. We clearly see that for $p = 2$, the Newton (before minimization) scheme produces more accurate results for u than the Picard (before minimization) formulation. As the mesh is refined, however, the Picard method also produces acceptable results. On the finer mesh, where $p = 6$, we see that the results of both schemes are nearly identical. We have chosen to not to include results for the Newton (after minimization) formulation in this plot as the numerical

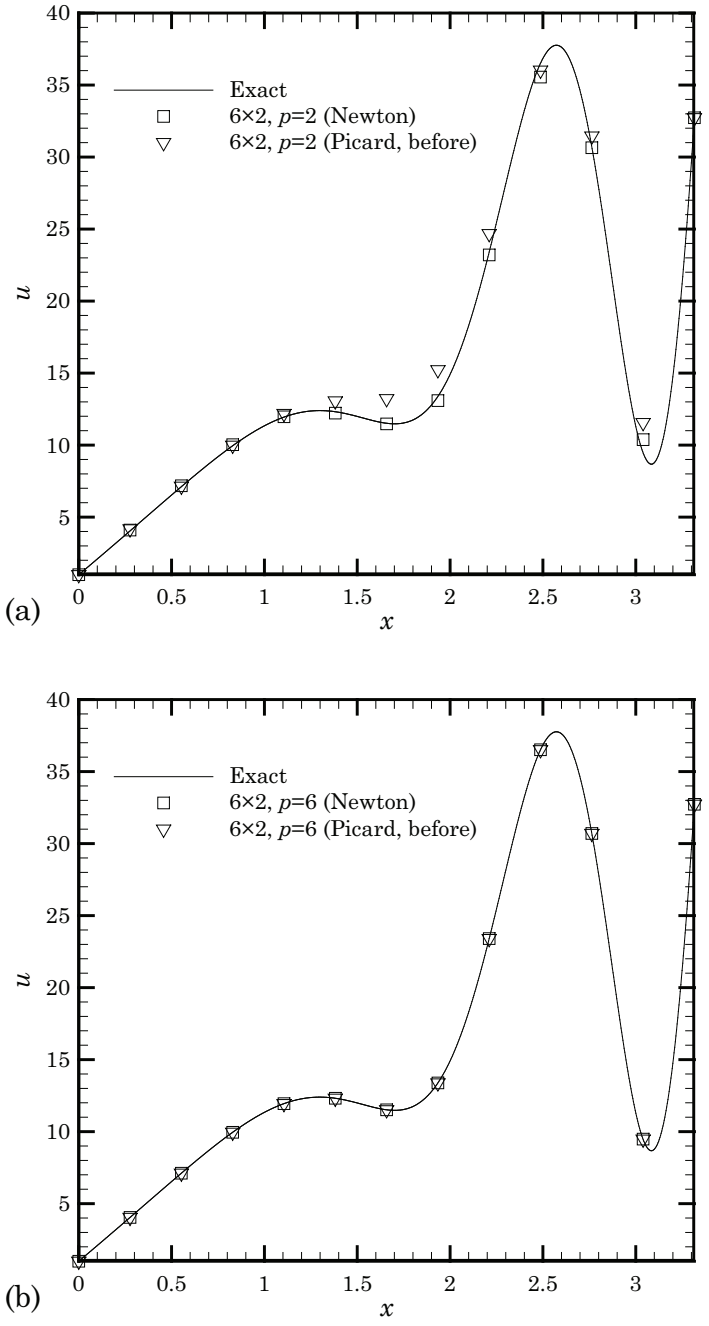


Fig. 10. Analytic and least-squares finite element solutions for u of the nonlinear Poisson equation along the horizontal mid-line of the domain: (a) $p = 2$ and (b) $p = 6$.

values are equivalent to those obtained by the Newton (before minimization) scheme.

Table II. Manufactured solution of a nonlinear Poisson equation: Number of iterations required to satisfy the nonlinear solution convergence criterion for various least-squares finite element implementations (termination criteria $\varepsilon = 10^{-6}$).

		Number of nonlinear iterations			
Least-squares formulation	ω_0	$p = 2$	$p = 4$	$p = 6$	$p = 8$
Picard (before)	0.50	207	28	44	112
Newton (before)	1.00	9	6	6	6
Newton (before)	0.50	8	7	7	7
Picard (after)	0.50	56	37	33	41
Newton (after)	0.50	7	7	7	7
Newton (mixed)	0.75	6	6	6	7

Finally, we also evaluate the nonlinear iterative solution convergence behavior of each least-squares finite element model. In all simulations, the nonlinear equations have been solved without the use of load stepping. The numbers of nonlinear iterations required to satisfy the convergence criterion at various p-levels are listed in Table II. Relaxation was employed in many of the finite element simulations. For both Picard formulations, the application of relaxation resulted in significant reductions in the required number of iterations. However, even with relaxation, these schemes always required a large number of iterations for convergence. In the case of the Newton linearization (performed after minimization), complete solution divergence was observed whenever relaxation was not employed. On the other hand, the Newton formulation (with linearization applied prior to minimization) performed well with or without relaxation. We also considered a mixed Newton formulation, where 5 Newton (before minimization) iterations were employed prior to the use of Newton (after minimization) iterations. This mixed formulation performed slightly better than the Newton (after minimization) formulation in several cases. Overall we

find that all Newton schemes (when relaxation is employed) offer similar convergence properties. Clearly the Newton scheme before minimization possesses a much larger radius of convergence as compared with the Newton scheme after minimization.

2. The incompressible Navier-Stokes equations

We next turn our attention to the incompressible form of the stationary Navier-Stokes equations, which constitutes a very popular application for least-squares variational principles. The classical problem for the non-dimensional form of the incompressible Navier-Stokes equations can be stated as follows: find the velocity vector $\mathbf{v}(\mathbf{x})$ and pressure $p(\mathbf{x})$ such that

$$\mathbf{v} \cdot \nabla \mathbf{v} = -\nabla p + \frac{1}{\mathrm{Re}} \nabla \cdot (\nabla \mathbf{v} + \nabla \mathbf{v}^{\mathrm{T}}) + \mathbf{b} \qquad \text{in } \Omega \tag{3.53a}$$

$$\nabla \cdot \mathbf{v} = 0 \qquad \text{in } \Omega \tag{3.53b}$$

$$\mathbf{v} = \mathbf{v}^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{D}} \tag{3.53c}$$

$$\hat{\mathbf{n}} \cdot \sigma = \mathbf{t}^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{N}} \tag{3.53d}$$

where Re is the Reynolds number, $\mathbf{b}$ is the body force, σ is the Cauchy stress tensor and $\hat{\mathbf{n}}$ is the outward unit normal. The Cauchy stress is given in terms of the following constitutive equation

$$\sigma = -p\mathbf{I} + \frac{1}{\mathrm{Re}} (\nabla \mathbf{v} + \nabla \mathbf{v}^{\mathrm{T}}) \tag{3.54}$$

Note that in this dissertation we have employed a typical abuse of notation by allowing p to represent both the p-level of the finite element solution as well as the pressure field in the Navier-Stokes equations.

There are many first-order formulations of the Navier-Stokes equations that have been presented in the literature [2, 35] that can be used to construct finite element models of least-squares type. One of the most popular schemes is the

velocity-pressure-vorticity $(\mathbf{v}, p, \omega)$ formulation. In this formulation the vorticity vector $\omega = \nabla \times \mathbf{v}$ is introduced, along with the vector identity

$$\nabla \times (\nabla \times \mathbf{v}) = -\nabla^2 \mathbf{v} + \nabla(\nabla \cdot \mathbf{v}) \tag{3.55}$$

As a result, we are able to restate the original problem in terms of the following equivalent first-order system: find the velocity $\mathbf{v}(\mathbf{x})$, pressure $p(\mathbf{x})$ and vorticity $\omega(\mathbf{x})$ such that

$$\mathbf{v} \cdot \nabla \mathbf{v} + \nabla p + \frac{1}{\mathrm{Re}} \nabla \times \omega = \mathbf{b} \qquad \text{in } \Omega \tag{3.56a}$$

$$\omega - \nabla \times \mathbf{v} = \mathbf{0} \qquad \text{in } \Omega \tag{3.56b}$$

$$\nabla \cdot \mathbf{v} = 0 \qquad \text{in } \Omega \tag{3.56c}$$

$$\mathbf{v} = \mathbf{v}^{\mathrm{p}} \qquad \text{on } \Gamma_{\mathbf{v}} \tag{3.56d}$$

$$\omega = \omega^{\mathrm{p}} \qquad \text{on } \Gamma_{\omega} \tag{3.56e}$$

$$\hat{\mathbf{n}} \cdot \tilde{\sigma} = \tilde{\mathbf{t}}^{\mathrm{p}} \qquad \text{on } \Gamma^{\mathrm{N}} \tag{3.56f}$$

where Γ^{D} has been partitioned such that $\Gamma^{\mathrm{D}} = \Gamma_{\mathbf{v}} \bigcup \Gamma_{\omega}$ and $\Gamma_{\mathbf{v}} \bigcap \Gamma_{\omega} = \varnothing$. We note that the incompressibility constraint has been imposed in the construction of the momentum equation. The pseudo-traction boundary condition $\tilde{\mathbf{t}}^{\mathrm{p}}$ is given in terms of the pseudo-stress tensor $\tilde{\sigma}$ defined as

$$\tilde{\sigma} = -p\mathbf{I} + \frac{1}{\mathrm{Re}} \nabla \mathbf{v} \tag{3.57}$$

For three-dimensional analysis it is helpful to augment the above equations by the compatibility condition $\nabla \cdot \omega = 0$.

We associate with the stationary first-order form of the incompressible Navier-

Stokes equations the following *true* least-squares functional

$$\mathcal{J}(\mathbf{v},p,\omega;\mathbf{b},\tilde{\mathbf{t}}^{\mathrm{p}}) = \frac{1}{2}\Bigg(\|\mathbf{v}\cdot\nabla\mathbf{v}+\nabla p+\frac{1}{\mathrm{Re}}\nabla\times\omega-\mathbf{b}\|^2_{\Omega,0}+\|\nabla\cdot\mathbf{v}\|^2_{\Omega,0} + \|\omega-\nabla\times\mathbf{v}\|^2_{\Omega,0}\Bigg) \tag{3.58}$$

The outflow boundary condition may also be directly accounted for in the definition of the least-squares functional (see Chapter IV for details). Linearized versions of the least-squares functional may be obtained by replacing the nonlinear convective term with following Picard or Newton approximations

$$\mathbf{v}\cdot\nabla\mathbf{v}|_{\mathrm{Pic}} = \mathbf{v}_0\cdot\nabla\mathbf{v} \tag{3.59a}$$

$$\mathbf{v}\cdot\nabla\mathbf{v}|_{\mathrm{New}} = \mathbf{v}\cdot\nabla\mathbf{v}_0+\mathbf{v}_0\cdot\nabla\mathbf{v}-\mathbf{v}_0\cdot\nabla\mathbf{v}_0 \tag{3.59b}$$

The linearized least-squares based weak formulation resulting from invoking the minimization principle may be stated as follows: find $\mathbf{u}\in\mathcal{V}$ such that for all $\delta\mathbf{u}\in\mathcal{W}$ the following expression holds

$$\mathcal{B}(\delta\mathbf{u},\mathbf{u}) = \mathcal{F}(\delta\mathbf{u}) \tag{3.60}$$

where $\mathbf{u}=(\mathbf{v},p,\omega)$, $\delta\mathbf{u}=(\delta\mathbf{v},\delta p,\delta\omega)$ and $\mathcal{V}$ and $\mathcal{W}$ are appropriate function spaces. When the Picard linearization scheme is employed, the bilinear form $\mathcal{B}(\delta\mathbf{u},\mathbf{u})$ and linear functional $\mathcal{F}(\delta\mathbf{u})$ are given as

$$\begin{aligned}\mathcal{B}(\delta\mathbf{u},\mathbf{u}) = \int_\Omega \Bigg[&\Bigg(\mathbf{v}_0\cdot\nabla\delta\mathbf{v}+\nabla\delta p+\frac{1}{\mathrm{Re}}\nabla\times\delta\omega\Bigg)\cdot\Bigg(\mathbf{v}_0\cdot\nabla\mathbf{v}+\nabla p \\ &+\frac{1}{\mathrm{Re}}\nabla\times\omega\Bigg)+(\nabla\cdot\delta\mathbf{v})(\nabla\cdot\mathbf{v})+(\delta\omega-\nabla\times\delta\mathbf{v})\cdot(\omega-\nabla\times\mathbf{v}) \\ &+\underline{(\delta\mathbf{v}\cdot\nabla\mathbf{v}_0)\cdot(\mathbf{v}\cdot\nabla\mathbf{v}_0)}\Bigg]d\Omega\end{aligned} \tag{3.61a}$$

$$\begin{aligned}\mathcal{F}(\delta\mathbf{u}) = \int_\Omega \Bigg[&\Bigg(\mathbf{v}_0\cdot\nabla\delta\mathbf{v}+\nabla\delta p+\frac{1}{\mathrm{Re}}\nabla\times\delta\omega\Bigg)\cdot\mathbf{b}-\underline{(\delta\mathbf{v}\cdot\nabla\mathbf{v}_0)\cdot} \\ &\underline{\Bigg(\nabla p_0+\frac{1}{\mathrm{Re}}\nabla\times\omega_0-\mathbf{b}\Bigg)}\Bigg]d\Omega\end{aligned} \tag{3.61b}$$

When Newton's method is applied, the bilinear form and linear form are

$$\mathcal{B}(\delta\mathbf{u},\mathbf{u}) = \int_\Omega \Bigg[\left(\delta\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\delta\mathbf{v} + \nabla\delta p + \frac{1}{\mathrm{Re}}\nabla\times\delta\omega\right)\cdot \left(\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\mathbf{v} + \nabla p + \frac{1}{\mathrm{Re}}\nabla\times\omega\right) + (\nabla\cdot\delta\mathbf{v})(\nabla\cdot\mathbf{v}) + (\delta\omega - \nabla\times\delta\mathbf{v})\cdot(\omega - \nabla\times\mathbf{v}) + \underline{(\delta\mathbf{v}\cdot\nabla\mathbf{v} + \mathbf{v}\cdot\nabla\delta\mathbf{v})\cdot \left(\mathbf{v}_0\cdot\nabla\mathbf{v}_0 + \nabla p_0 + \frac{1}{\mathrm{Re}}\nabla\times\omega_0 - \mathbf{b}\right)}\Bigg] d\Omega \tag{3.62a}$$

$$\mathcal{F}(\delta\mathbf{u}) = \int_\Omega \Bigg[\left(\delta\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\delta\mathbf{v} + \nabla\delta p + \frac{1}{\mathrm{Re}}\nabla\times\tilde{\omega}\right)\cdot(\mathbf{b} + \mathbf{v}_0\cdot\nabla\mathbf{v}_0) + \underline{(\delta\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\delta\mathbf{v})\cdot\left(\mathbf{v}_0\cdot\nabla\mathbf{v}_0 + \nabla p_0 + \frac{1}{\mathrm{Re}}\nabla\times\omega_0 - \mathbf{b}\right)}\Bigg] d\Omega \tag{3.62b}$$

The terms underlined above appear when minimization is performed prior to linearization.

a. Low Reynolds number flow past a circular cylinder

In this example we consider flow past a circular cylinder, a problem that has been studied extensively by way of experiment [60, 61, 62] and is a standard benchmark for numerical computation [63, 64, 65]. It is well-known from both experimentation [62] and numerical modeling [66, 67] that for moderately low Reynolds numbers ($5 < \mathrm{Re} < 46.1$) the flow is spatially stationary and characterized by two symmetric regions of circulation directly downwind of the cylinder. The size of the standing vortices in the wake region is proportional to the Reynolds number.

Ideally we would like to model the flow in a manner such that end effects (due to truncation of the problem to a geometrically finite computational domain) do not corrupt the integrity of the numerical solution. To this end we take $\bar{\Omega}$ to be the set difference between the rectangular region $[-25, 25] \times [-15, 15]$ and an open circular region with unit diameter centered about the origin. The computational domain

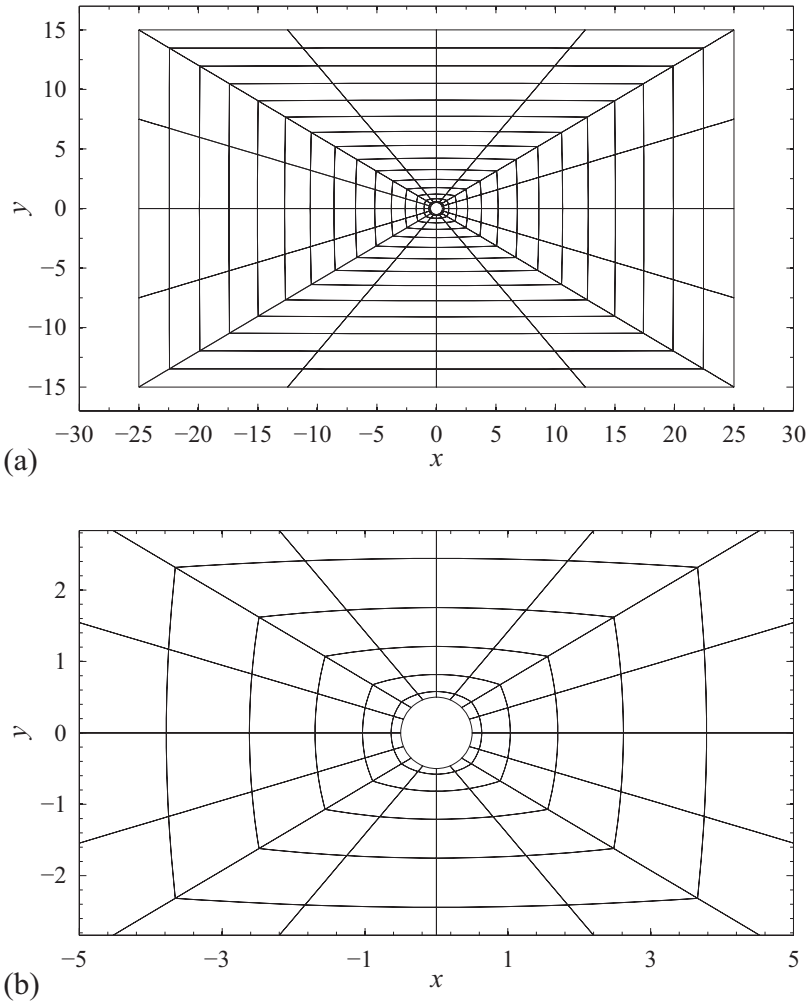


Fig. 11. Finite element discretization of the computational domain $\bar{\Omega}$ for the analysis of steady flow past a circular cylinder: (a) view of the complete mesh and (b) close up view of mesh near the cylinder.

$\bar{\Omega}^{hp} \cong \bar{\Omega}$ consists of 240 non-uniform finite elements, with 15 element layers in the radial direction and 16 along the circumference of the cylinder as shown in Figure 11. The smallest elements are placed in the vicinity of the cylinder to ensure adequate numerical resolution in the anticipated wake region. The mesh geometry is characterized using an isoparametric formulation which, when combined with high-order finite element technology, allows for a highly accurate approximation of the cylinder surface. As in the Poisson benchmark problem, we refine the mesh by systematically increasing the p-level of the finite element approximation functions. We consider the cases where $p = 2$, 4, 6 and 8; which amounts to 3,968, 15,616, 34,944 and 61,952 total degrees of freedom for each corresponding finite element discretization. The boundary conditions coincide with those used in the parallel performance benchmark problem given in Chapter II Section C where the Reynolds number is taken to be 40. The outflow boundary condition is enforced weakly through the least-squares functional (see Chapter IV) with $\mathbf{t}^{\mathrm{p}}$ taken as zero along the right hand side of $\bar{\Omega}$. Nonlinear convergence is declared for a given numerical simulation once the relative error in the solution is less than 10^{-6}.

Since this problem does not admit an analytic solution, we obtain a reasonable *a posteriori* estimate for the error via a numerical evaluation of $\mathcal{J}$ during the post-processing stage of the analysis. Exponential decay of the least-squares functional, shown in Figure 12, is clearly visible as the polynomial order of the numerical solution is increased. As expected, each least-squares formulation produces identical converged results (for a given p-level) with the exception of the Picard scheme (applied prior to minimization). The value of $\mathcal{J}$ for this scheme is only slightly greater than the values determined by the other three formulations.

Figure 13 shows the numerically determined pressure coefficients along the surface of the cylinder. When the mesh is coarse ($p = 2$), linearization before minimiza-

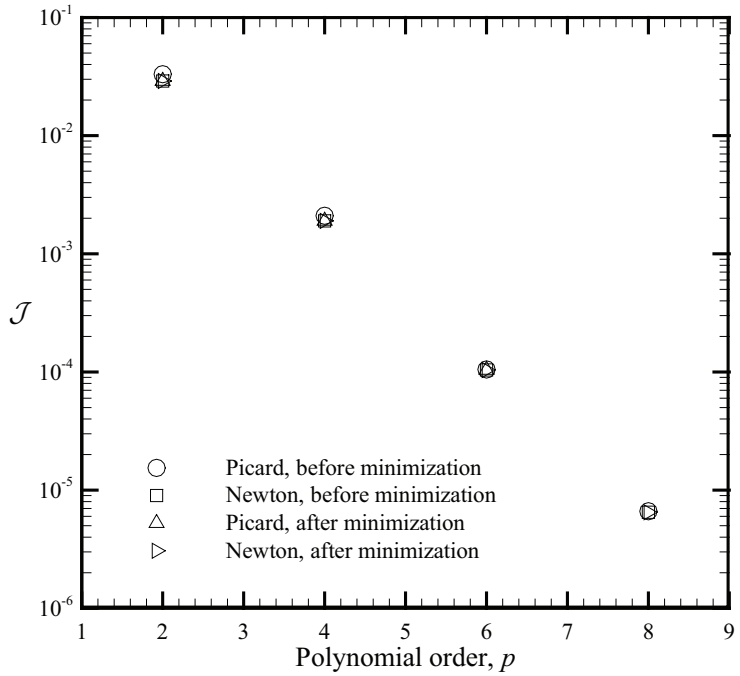


Fig. 12. Convergence of the least-squares finite element solutions under p-refinement as measured in terms of the least-squares functional $\mathcal{J}$ for steady flow past a circular cylinder at Re = 40.

tion using Picard's method yields substantially different results than does Newton's method (where by Newton's method we mean either Newton scheme, as they both yield identical results). Neither solution at this p-level, however, constitutes an appropriate converged solution for the problem. We see that as the p-level is increased to 4, the Picard (before minimization) and Newton schemes begin to coincide. Finally, at $p = 6$ we observe virtually no difference between the results of either scheme. The computed values at this p-level and higher were found to be in excellent agreement with the empirical work conducted by Grove et al. [61]. In Figure 14 we also show the pressure field and velocity component v_y in the vicinity of the cylinder for the Newton solution at $p = 8$. Streamlines are also shown highlighting the size of the circulation

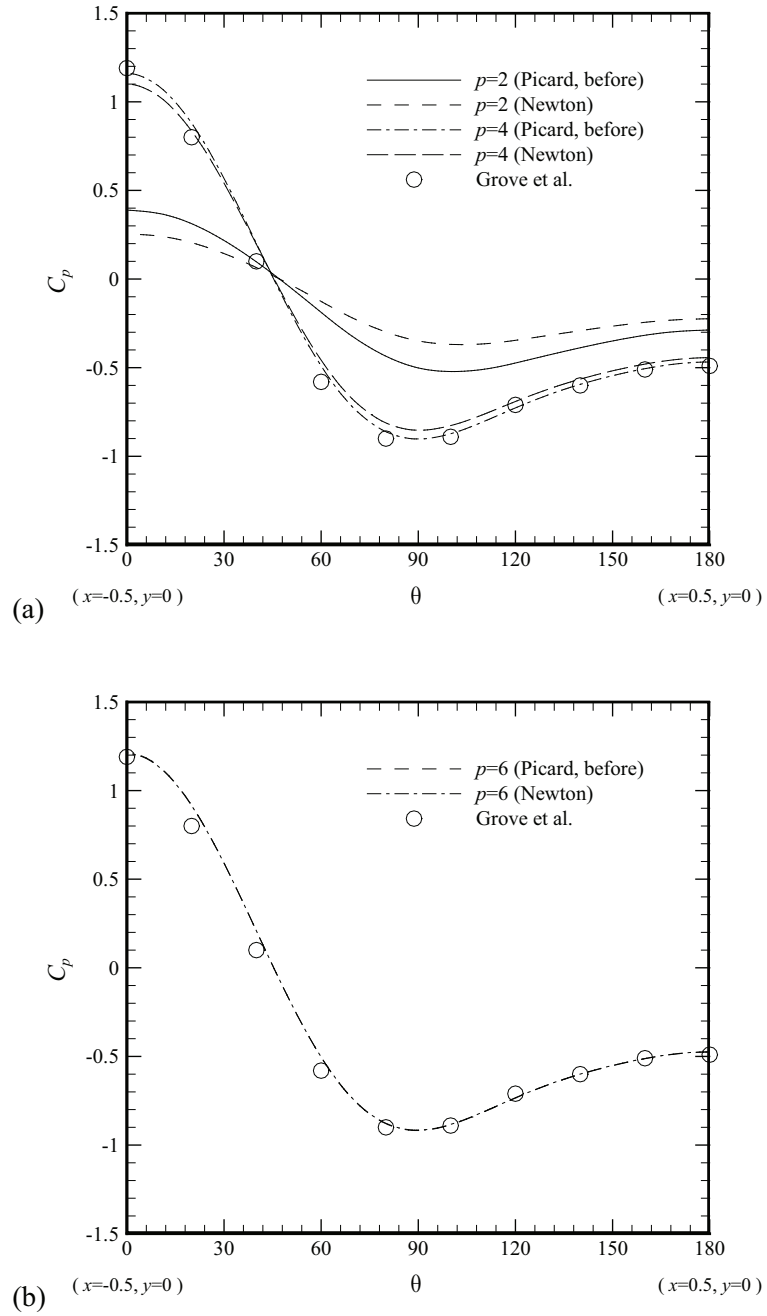


Fig. 13. A comparison of the numerically computed pressure coefficient C_p along the surface of the cylinder at Re = 40 with the experimental data obtained by Grove et al. [61]: (a) non-converged numerical solutions and (b) fully-converged numerical solutions.

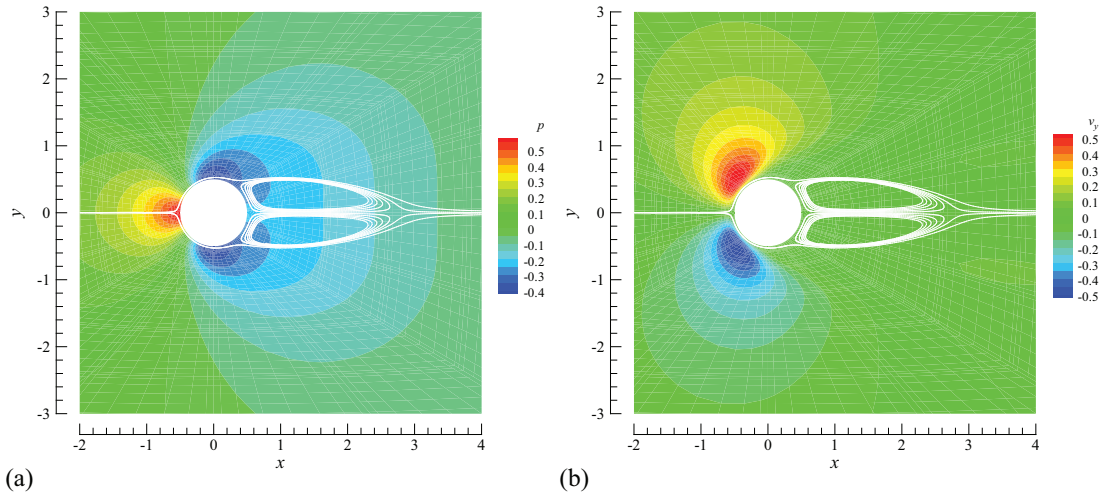


Fig. 14. Steady flow past a circular cylinder at Re = 40: (a) pressure field and streamlines and (b) velocity component v_y and streamlines.

regions. Our numerical calculations predict the wake region to extend 4.50 cylinder radii downstream of the cylinder, which is in excellent agreement with the numerical results reported by Kawaguti and Jain [63].

Table III. Steady flow past a circular cylinder: Number of iterations required to satisfy the nonlinear solution convergence criterion for various least-squares finite element implementations (termination criteria $\varepsilon = 10^{-6}$).

		Number of nonlinear iterations			
Least-squares formulation	ω_0	$p = 2$	$p = 4$	$p = 6$	$p = 8$
Picard (before)	1.0	11	15	15	15
Newton (before)	1.0	10	9	7	6
Picard (after)	0.5	39	161	179	–
Newton (mixed)	1.0	7	7	7	7

A word on the nonlinear iterative convergence behavior of each finite element scheme is also in order. In each formulation we solved the equations without the employment of load steps. A summary of the total number of iterations needed to achieve

the desired termination criteria is summarized in Table III. When linearization was performed prior to minimization, solution convergence was possible without the need for relaxation. For the case of the Picard linearization after minimization, however, convergence was extremely slow and could not be achieved without relaxation. Due to such poor rates of convergence, a solution at $p = 8$ was not attempted. Using the Newton linearization scheme after minimization produced divergent results, with or without solution relaxation. We therefore considered a mixed Newton scheme where 3 Newton (before minimization) iterations were used prior to subsequent Newton (after minimization) iterations. This method produced very good results in terms of minimizing the total number of iterations required for convergence. For low p-levels this approach slightly outperformed the Newton (before minimization) scheme. Overall, the Newton (before minimization and mixed) schemes exhibited much better convergence rates than the Picard formulations.

b. Steady flow over a backward facing step

In this example we consider the flow of a viscous incompressible fluid over a backward facing step. This problem was studied by way of experiment and also numerical simulation by Armaly et al. [68]. Laminar, transition and turbulent flows were empirically assessed for $70 < \mathrm{Re} < 8{,}000$, and numerically simulated for steady-state cases up to a Reynolds number of 1,250. In our numerical study, we evaluate the stationary solution of the two-dimensional problem at $\mathrm{Re} = 800$, using the simplified step configuration proposed in the benchmark solution of Gartling [69].

The computational domain for the problem is given as $\bar{\Omega} = [0, 30] \times [-0.5, 0.5]$ as shown in Figure 15. The fluid enters the domain on the left hand side of $\bar{\Omega}$ on $0 \leq y \leq 0.5$. The velocity vector at the inlet is assumed to be horizontal with the x-component given by the parabolic expression $\bar{v}_x = 24y(0.5 - y)$. The components

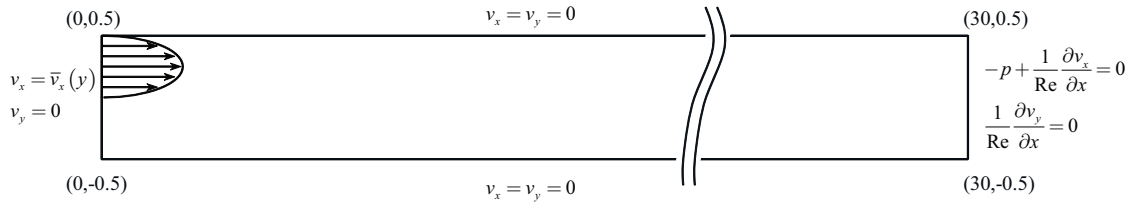


Fig. 15. Geometry and boundary conditions for steady flow of an incompressible viscous fluid over a backward facing step at Re = 800.

of the velocity are taken to be zero along all solid surfaces in accordance with the non-slip condition. The outflow boundary condition is enforced weakly by taking $\mathbf{t}^{\mathrm{p}} = \mathbf{0}$ in the least-squares functional.

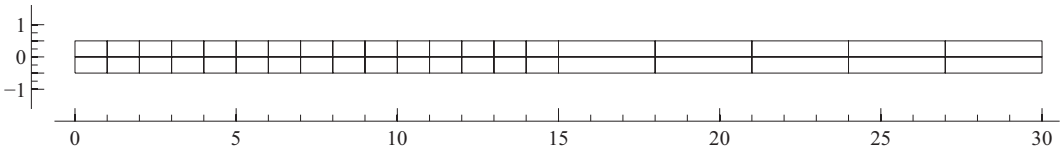


Fig. 16. Finite element mesh for analysis of stationary incompressible viscous flow over a backward facing step.

We discretize the computational domain into a set of 40 rectangular finite elements, with 20 elements along the channel length and 2 along the channel height as shown in Figure 16. The majority of the elements are positioned within 15 units of the channel inlet to insure proper resolution of all variables within the flow separation regions anticipated downstream of the step. We once again refine the discrete solution by systematically increasing the number of nodes in each finite element. We arrive at the numerical solution at Re = 800, by solving a series of problems at intermediate Reynolds numbers. We begin by solving the problem at Re = 100 followed by Re = 200 and so on until we reach Re = 800. For each intermediate problem, we

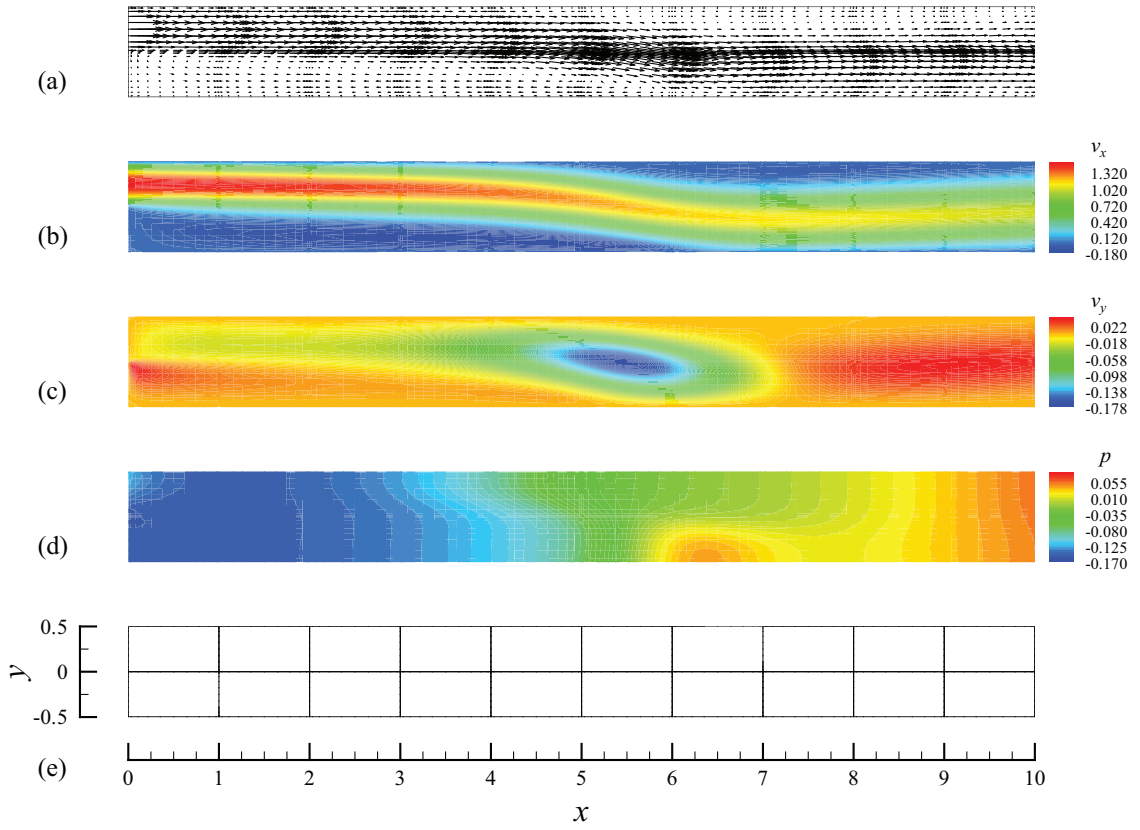


Fig. 17. Steady flow of an incompressible viscous fluid over a backward facing step at Re = 800: (a) velocity vector field at finite element nodes, (b) velocity component v_x, (c) velocity component v_y, (d) pressure field and (e) finite element mesh directly behind step.

utilize the converged solution from the previous problem in the series as the initial guess. As an initial guess for the problem where Re = 100, we assume all variables to be zero. Nonlinear convergence is declared for each problem when the Euclidean norm of the difference between the nonlinear solution increments is less than 10^{-4}.

In Figure 17 we show the least-squares finite element solution of the problem as determined using a polynomial of order 10 within each element. The velocity vectors are depicted along with contour plots of the velocity components and pressure field.

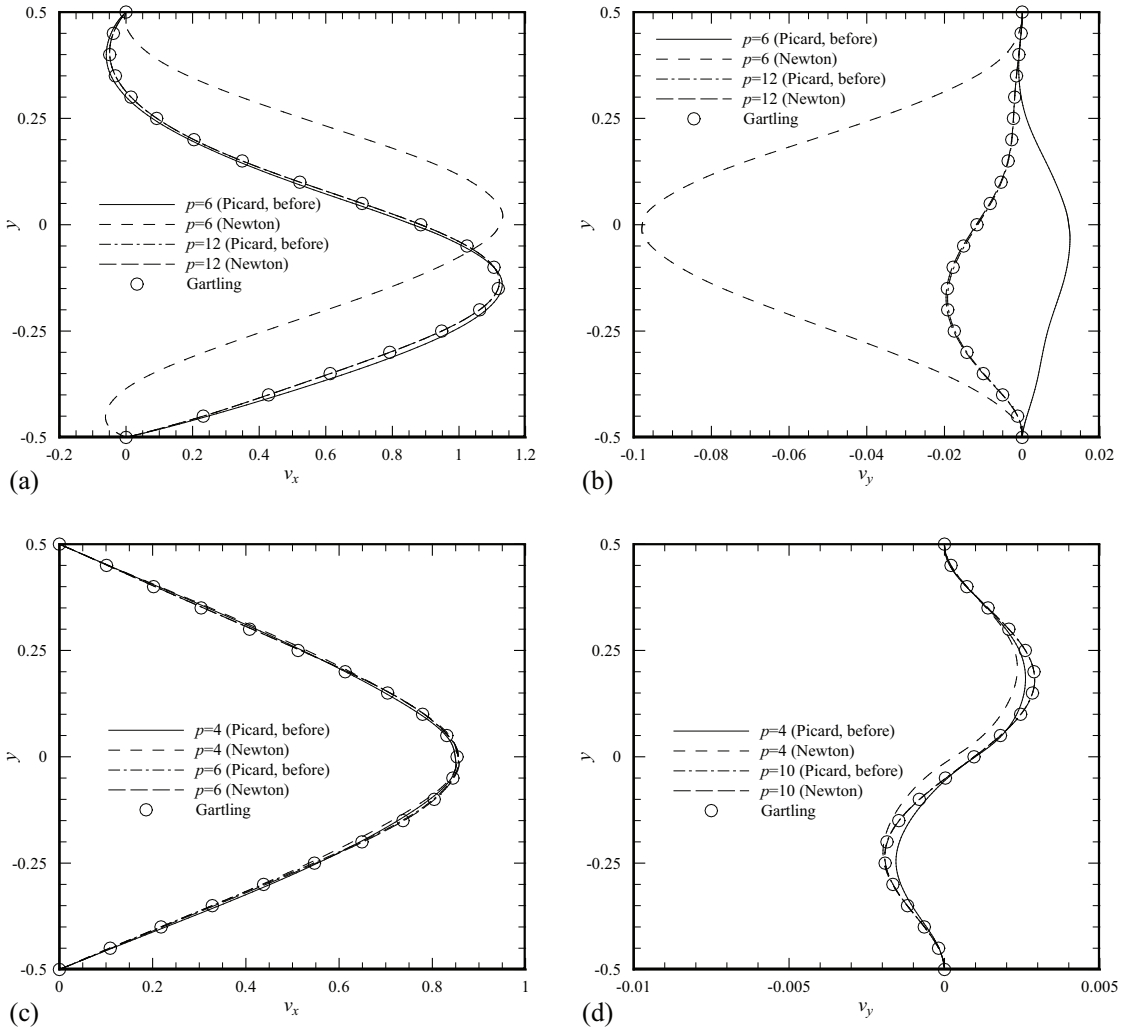


Fig. 18. Comparison of numerically computed velocity components for the steady flow of a viscous fluid over a backward facing step at Re = 800 with the published results of Gartling [69, 7]: (a) horizontal velocity profile at $x = 7$, (b) vertical velocity profile at $x = 7$, (c) horizontal velocity profile at $x = 15$ and (d) vertical velocity profile at at $x = 15$.

The flow is characterized by a large recirculation zone directly behind the step on the low side of the channel that extends roughly 6.1 units beyond the step. A second region of flow separation and recirculation is also present on the top side of the channel that develops around 4.9 units downstream of the step and extends to approximately $x = 10.5$.

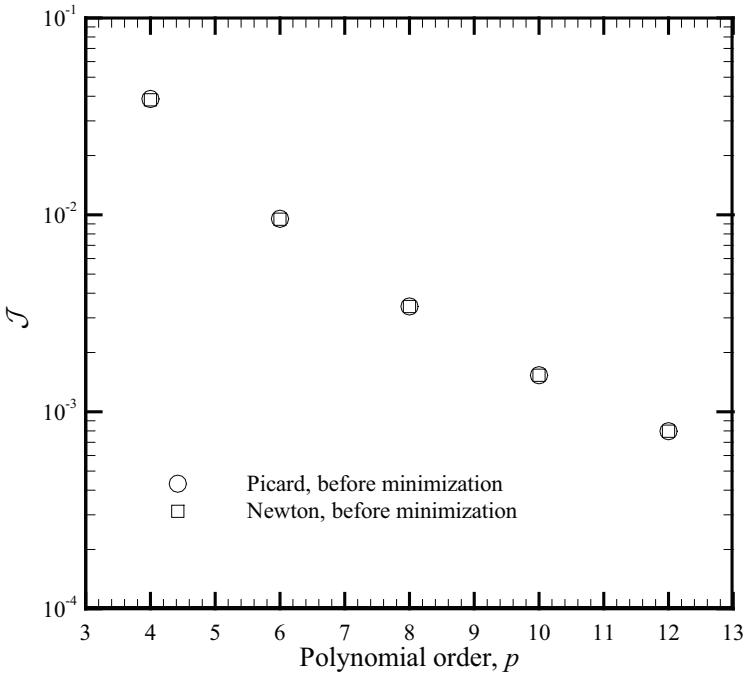


Fig. 19. Convergence of the least-squares finite element solutions under p-refinement as measured in terms of the least-squares functional $\mathcal{J}$ for steady flow past a backward facing step at $\mathrm{Re} = 800$.

In Figure 18 we compare our numerical solutions for the components of the velocity vector along $x = 7$ and $x = 15$ with the results reported by Gartling [69], where a weak-form Galerkin finite element model was employed. The converged results for the Picard (linearization before minimization) and Newton schemes are in excellent agreement with the published data. As expected, the Picard and Newton

formulations yield dissimilar results when the finite element mesh is too coarse to allow for convergence. It is interesting to note, however, that the Picard scheme offers a somewhat better approximation of the velocity components on the coarse mesh than does Newton's method at $x = 7$ and $x = 15$. The reason for this phenomenon is unclear; however, as anticipated from the abstract problem, both schemes converge to the same solution under proper mesh refinement.

Table IV. Steady flow over backward facing step: Number of iterations required to satisfy the nonlinear solution convergence criterion for various least-squares finite element implementations, where $p = 6$ (termination criteria $\varepsilon = 10^{-4}$).

	Number of nonlinear iterations			
Renolds number	Picard (before)	Newton (before)	Picard (after)	Newton (mixed)
100	13	6	81	6
200	20	4	313	4
300	28	5	616	5
400	39	6	774	6
500	53	7	641	8
600	63	8	501	7
700	69	11	405	8
800	74	15	380	10

In Figure 19 we plot the value of the least-squares functional for the Picard (before minimization) and Newton (before minimization) schemes as a function of the p-level. Although $\mathcal{J}$ is always greater for the Picard scheme as compared with Newton's method, the actual numerical values are nearly identical. A summary of the total number of iterations required to reach the desired termination criteria at each Reynolds number is summarized in Table IV for $p = 6$. The Picard and Newton finite element solutions (where linearization was performed prior to minimization) were obtained without the use of relaxation. For the Picard (after minimization) formulation, a relaxation parameter of $\omega_0 = 0.5$ was utilized. Even with the aid of

relaxation, however, the scheme still suffered from a severely poor rate of convergence. We were unable to obtain a converged solution for the Newton (after minimization) formulation, with or without relaxation. As in the previous example, we once again introduced a mixed Newton formulation in an attempt to recover a convergent solution. In the mixed approach, we utilized the Newton (before minimization) formulation in the iterative solution scheme until ε was less than 0.05, at which point we switched to the Newton (after minimization) formulation. Solution relaxation was found to be unnecessary in the mixed approach. On average the mixed formulation performed comparably to the Newton (before minimization) scheme and in some cases superior at this p-level. However, at higher p-levels we find little difference between the convergence behaviors of these Newton formulations (especially at Re = 700, 800). As expected, the Newton schemes require far fewer iterations than their Picard counterparts.

c. Lid-driven cavity flow

As a final verification benchmark, we consider the classical two-dimensional lid-driven cavity flow problem. The computational domain is taken as the unit square given as $\bar{\Omega} = [0,1]\times[0,1]$ and the boundary conditions are specified in terms of the components of the velocity vector and the pressure at a single point. On the bottom and left and right sides of the cavity the components of the velocity are taken to be zero. Along the top surface a horizontal velocity profile is specified using the following expression

$$v_x(x) = \begin{cases} \tanh(50x) & 0 \leqslant x \leqslant 0.5 \\ -\tanh[50(x-1)] & 0.5 < x \leqslant 1.0 \end{cases} \tag{3.63}$$

The above boundary condition essentially prescribes v_x as unity along the majority of the top surface of the cavity with a smooth and abrupt transition to $v_x = 0$ at the

corners. The boundary condition is applied in this way to avoid *singularities* in the solution in the vicinity of the upper corners [55]. High-order methods are sensitive to such singularities, and the above boundary condition insures, in this sense, a well posed problem. The pressure is taken to be zero at the single point $(x, y) = (0.5, 0)$. In our analysis, we consider the steady-state solution of the problem at a Reynolds number of 3,200. We compare our numerical solutions with the tabulated finite difference results reported by Ghia et al. [70], who used $v_x = 1$ at all points of the lid except at $x = 0$ and $x = 1$, where they used $v_x = 0$.

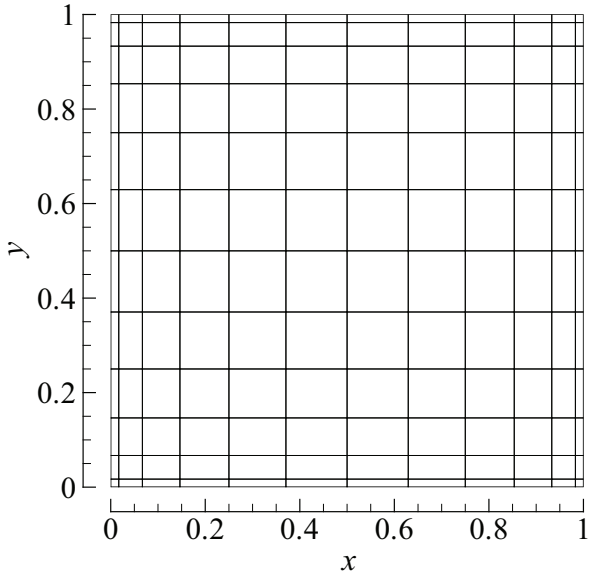


Fig. 20. Finite element mesh for the lid-driven cavity flow problem.

The domain is discretized into a non-uniform set of 144 rectangular finite elements as shown in Figure 20. The mesh is graded such that smaller elements are placed near the boundaries to insure proper resolution of the numerical solution in the regions of the boundary layers and anticipated vortices. As in previous examples, the mesh is refined by increasing the p-level of the solution within each finite element.

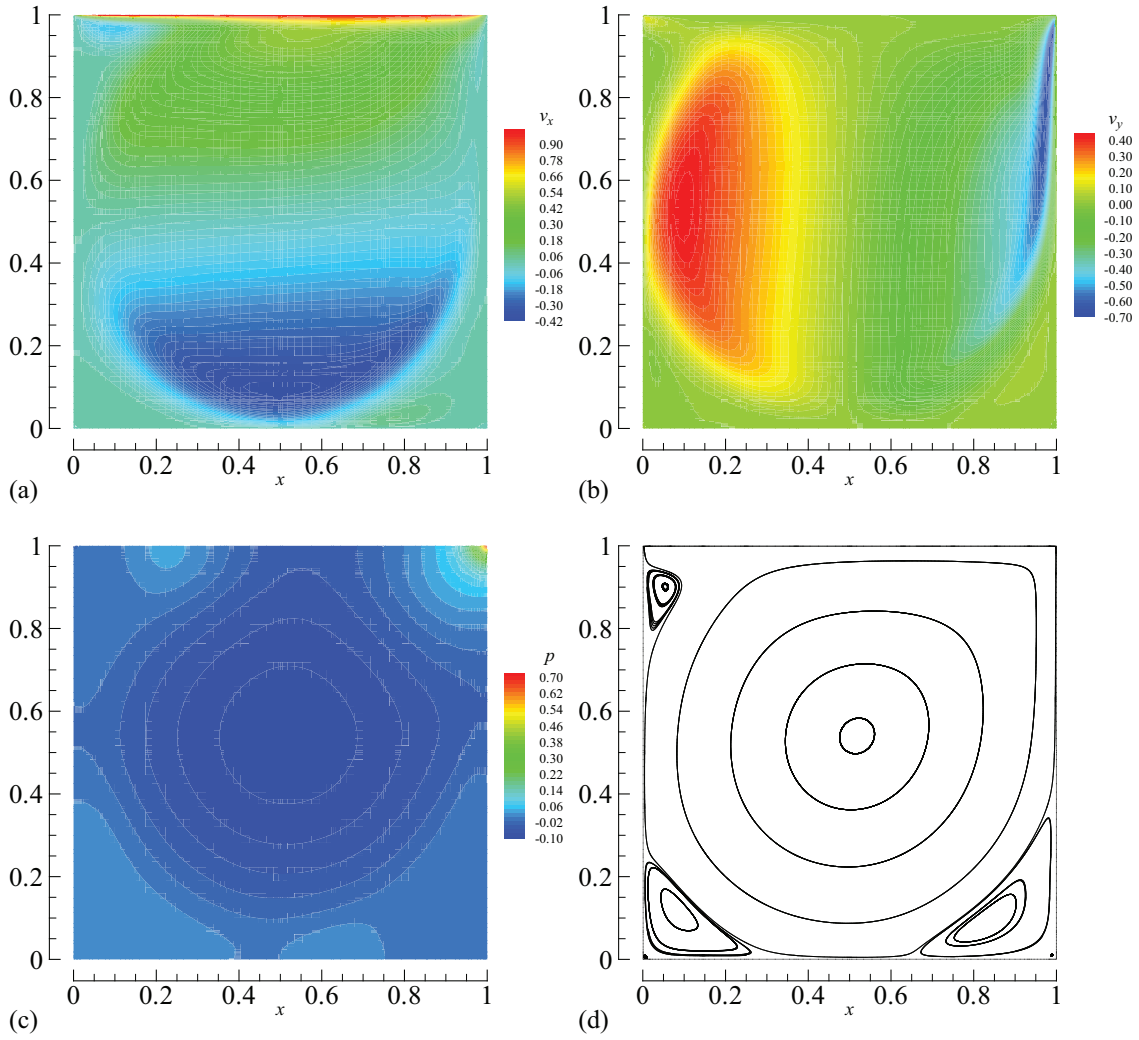


Fig. 21. Two-dimensional lid-driven cavity flow at Re = 3,200: (a) velocity component v_x, (b) velocity component v_y, (c) pressure field and (d) streamline patterns in cavity highlighting standing vortices.

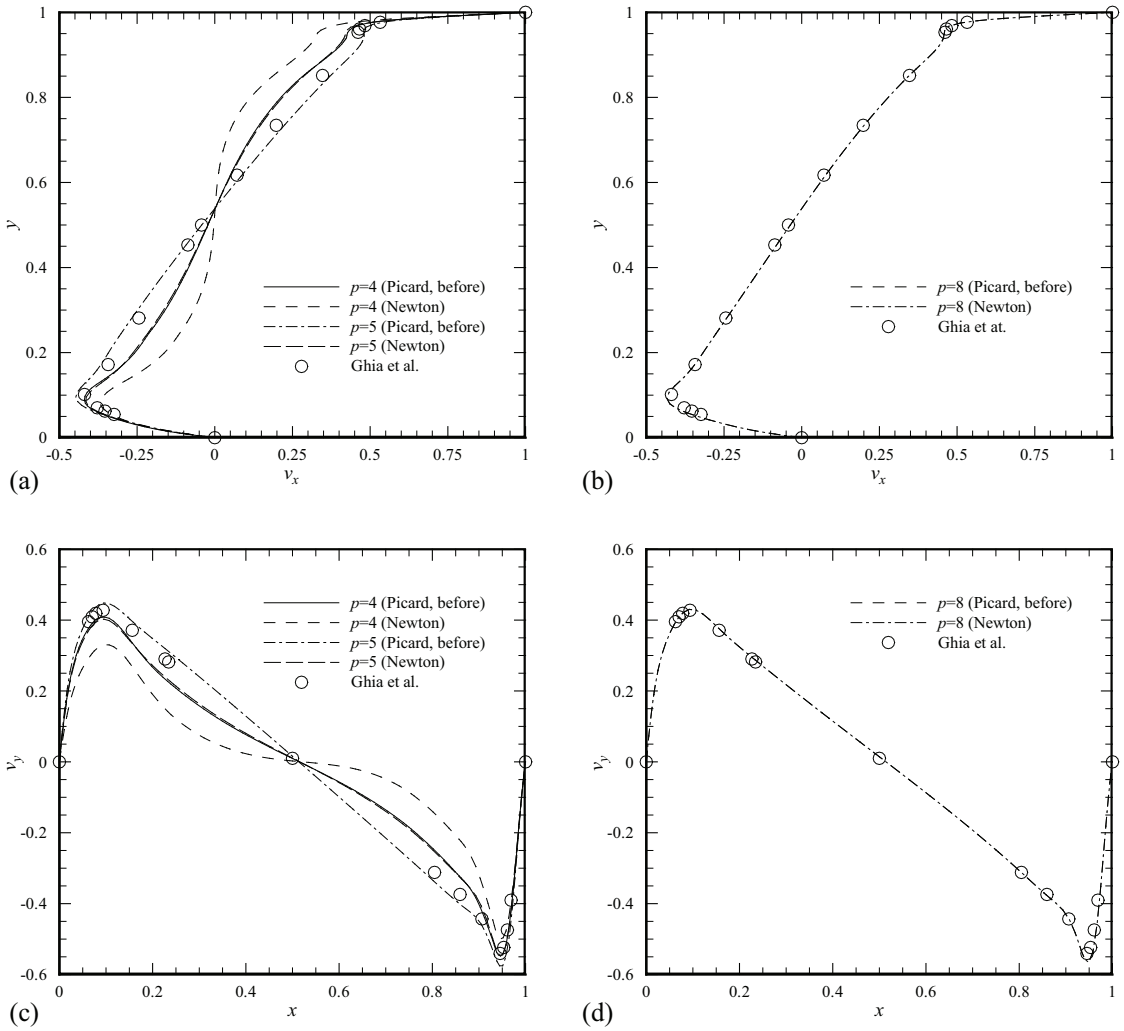


Fig. 22. Comparison of numerically computed velocity components along vertical and horizontal mid-lines of lid-driven cavity with published results of Ghia et al. [70] at Re = 3,200: (a) non-converged numerical solutions for horizontal velocity v_x profiles along vertical centerline, (b) converged numerical solutions for horizontal velocity v_x profiles along vertical centerline, (c) non-converged numerical solutions for vertical velocity v_y profiles along horizontal centerline and (d) converged numerical solutions for vertical velocity v_y profiles along horizontal centerline.

We utilize polynomials of orders 4, 5, 6, 7, 8 and 9 in our analysis which correspond with 9,604, 14,884, 21,316, 28,900, 37,636 and 47,524 total degrees of freedom. As in the previous example, the desired solution at Re = 3,200 is obtained by solving a series of problems at intermediate Reynolds numbers. In this case we solve a series of seven problems beginning with the first problem posed with Re = 457.14 and culminating with the final desired solution at Re = 3,200. Nonlinear convergence is considered to be achieved in each problem once the Euclidean norm of the difference between the nonlinear iterative solution increments is less than 10^{-4}. Due to poor convergence properties, a numerical solution using the Picard scheme (as applied after minimization) was not attempted.

The velocity components, pressure field and streamlines are shown in Figure 21 for the numerical solution obtained using Newton's method (before minimization) at $p = 9$. The flow is characterized by a large region of rotation that is just off-set from the geometric center of the cavity. Secondary vortices are also present in the regions near the bottom (left and right) and top left corners of the domain. The streamline patterns match well with the published results of Ghia et al. [70].

In Figure 22 we compare our least-squares finite element solutions along the vertical and horizontal mid-planes of the cavity with the tabulated results of Ghia et al. [70]. We once again find that when the mesh is too coarse to yield a convergent solution, the numerical results differ for the Picard (linearization before minimization) and Newton schemes. However, as expected both schemes yield identical results when the mesh is properly refined. As in the previous example, we are surprised to find that the Picard scheme offers a slightly better approximation of the velocity components on the coarse mesh along the mid-lines of the cavity.

The value of the least-squares functional as a function of the p-level exhibits characteristics similar to those discussed in previous examples. In particular, the

value of $\mathcal{J}$ in both Newton schemes is always slightly less than (although nearly identical to) the value as determined using the Picard (before minimization) method. We were once again able to obtain convergent solutions using the Picard and Newton schemes (with linearization performed before minimization) without the need for solution relaxation. Solution convergence could not be achieved for the Newton (after minimization) scheme with or without relaxation. As a result, we again utilized a mixed Newton scheme, where Newton (before minimization) iterations where performed until the relative error ε was less than 0.01 at which point we switched to the Newton (after minimization) scheme. The total number of iterations required for solution convergence at a p-level of 6 is summarized for each scheme in Table V.

Table V. Two dimensional cavity driven flow: Number of iterations required to satisfy the nonlinear solution convergence criterion for various least-squares finite element implementations, where $p = 6$ (termination criteria $\varepsilon = 10^{-4}$).

	Number of nonlinear iterations		
Reynolds number	Picard (before)	Newton (before)	Newton (mixed)
457.14	12	7	7
914.29	12	5	5
1371.43	12	4	4
1828.57	14	4	4
2285.71	17	4	4
2742.86	20	4	4
3200.00	23	3	3

CHAPTER IV

A LEAST-SQUARES FINITE ELEMENT FORMULATION FOR VISCOUS INCOMPRESSIBLE FLUID FLOWS WITH ENHANCED ELEMENT-LEVEL MASS CONSERVATION

Tremendous progress has been achieved over the last few decades in the field of computational fluid dynamics. The advent of the digital computer and in particular parallel processing has made it possible to numerically simulate complex flow patterns that just a few years ago could only have been investigated using experimental procedures and dimensional analysis. Much of the success and breakthroughs in the numerical simulation of the Navier-Stokes equations for incompressible fluids have come in the context of low-order finite difference and finite volume technologies. Although the finite element method has become the dominate method of choice in the numerical analysis of solids, it has yet to receive such widespread acceptance when applied to fluid flow problems. It is well known, however, that finite element procedures offer many advantages over finite difference and finite volume methods. In particular, the finite element method can naturally deal with complex regions, complicated boundary conditions and possesses a rich mathematical foundation. As a result, there has been a renewed interest in recent years in developing efficient and accurate finite element models of the incompressible Navier-Stokes equations.

The majority of finite element models for fluids are based on the weak-form Galerkin procedure. It is well-known, however, that application of this method can lead to a non-optimal setting for a given finite element discretization [2, 50]. As discussed in Chapter III, application of the weak-form Galerkin method to the incompressible Navier-Stokes equations expressed in terms of the velocities and pressure results in a finite element model that must satisfy the restrictive discrete inf-sup or

LBB condition [31]; this effectively precludes the use of equal interpolation of the velocity and pressure fields in the numerical implementation. Even when the LBB condition is satisfied, the finite element solution may still be plagued by spurious oscillations in convection dominated problems. Stabilized weak-form Galerkin finite element formulations such as the SUPG [4, 5], penalty [6, 7] and Galerkin least-squares [8] have received considerable attention over the last few decades and have greatly improved the discrete setting for the finite element solution. Unfortunately, the success of these methods is often intertwined with ad-hoc parameters that must be fine tuned for a given flow problem.

Least-squares finite element models for the numerical simulation of viscous incompressible fluid flows have received substantial attention in the academic literature in recent years and offer an appealing alternative to the more popular weak-form Galerkin procedure. The least-squares formulation allows for the construction of finite element models for fluids that possess many of the attractive qualities associated with the well-known Ritz method [3]; for example global minimization, best approximation with respect to a well-defined norm and symmetric positive-definiteness of the resulting finite element coefficient matrix [2]. We refer to Chapter III for a detailed discussion on both the least-squares method and advantages it holds over the traditional weak-form Galerkin procedure.

It is well-known, however, that least-squares finite element models of both the steady and non-stationary form of the incompressible Navier-Stokes can be plagued by poor *local* (or element-level) mass conservation [71]. This is especially true whenever low-order elements are employed in a given finite element discretization. The discrete violation of the requirement that the velocity be a solenoidal vector is often attributed to the fact that, in least-squares formulations, local satisfaction of the governing PDEs is sacrificed in favor of global minimization of the governing equation residuals [51]. In

transient flow problems, lack of velocity-pressure coupling [52] has also been identified as a source for poor local mass conservation. We must emphasize that the violation *is not* merely numerical noise and, depending on the nature of the domain and boundary conditions, can actually be quite substantial [51].

Several techniques have been proposed to improve local mass conservation in least-squares finite element models. For example Deang and Gunzburger [71] advocated weighting the continuity equation residual in the definition of the least-squares functional, where the chosen weight may be either uniform across the whole problem domain or distinct for each element [72]. Chang and Nelson [51], on the other hand, combined the least-squares method with Lagrange multipliers to exactly enforce element-level mass conservation. In this approach the continuity equation is treated as an additional constraint for each element that is enforced in the discrete setting through a set of NE Lagrange multipliers. Although successful [51, 73], this approach comes at the expense of increasing the system size of the finite element equations and compromising the unconstrained minimization setting that is so attractive for least-squares finite element models [74]. Recently Heys et al. [75] demonstrated improved mass balance using a least-squares functional based on a novel first-order reformulation of the incompressible Navier-Stokes equations. It is worth noting that increasing the p-level also tends to improves mass conservation [17].

For non-stationary flows, lack of strong velocity-pressure coupling can also compromise local mass conservation and further lead to total instability in space-time decoupled finite element simulations [17]. Pontaza showed that the employment of a regularized form of the continuity equation in least-squares formulations can greatly enhance velocity-pressure coupling and as a direct consequence local mass conservation [52]. Similar approaches have also been advocated in the iterative penalty formulations of Prabhakar and Reddy [53, 54, 55] and Prabhakar et al. [56]. For a

more mathematical analysis of such optimization methods as applied to the Stokes problem, we refer to the work of Bochev and Gunzburger [76].

The purpose of this chapter is to present a novel least-squares finite element model for both the steady and non-stationary incompressible Navier-Stokes equations based on the standard velocity-pressure-vorticity first-order system, but with enhanced element-level mass conservation. The proposed formulation may be viewed as a direct extension of the work of Chang and Nelson [51] and also Pontaza [52]. For the steady flow case, we recast the constrained minimization problem of Chang and Nelson [51] into an unconstrained minimization problem through the use of the penalty method. This approach is quite natural, as the traditional least-squares method is itself in a sense a multi-equation penalty formulation (where penalization is applied *to all* of the partial differential equation residuals). For non-stationary flows, a penalty formulation is proposed that enhances local mass conservation while improving velocity-pressure coupling and overall numerical stability.

A. The non-stationary incompressible Navier-Stokes equations

We consider the non-stationary incompressible flow of a viscous fluid as described by the Navier-Stokes equations. The problem may be stated in non-dimensional form as follows: find the velocity $\mathbf{v}(\mathbf{x},t)$ and pressure $p(\mathbf{x},t)$ such that

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v} + \nabla p - \frac{1}{\mathrm{Re}}\nabla\cdot(\nabla\mathbf{v} + \nabla\mathbf{v}^{\mathrm{T}}) = \mathbf{b} \qquad \text{in } \Omega\times(0,\tau] \tag{4.1a}$$

$$\nabla\cdot\mathbf{v} = 0 \qquad \text{in } \Omega\times(0,\tau] \tag{4.1b}$$

$$\mathbf{v}(\mathbf{x},0) = \bar{\mathbf{v}}(\mathbf{x}) \qquad \text{in } \Omega \tag{4.1c}$$

$$\mathbf{v} = \mathbf{v}^{\mathrm{p}}(\mathbf{x},t) \qquad \text{on } \Gamma^{\mathrm{D}}\times(0,\tau] \tag{4.1d}$$

$$\hat{\mathbf{n}}\cdot\sigma = \mathbf{t}^{\mathrm{p}}(\mathbf{x},t) \qquad \text{on } \Gamma^{\mathrm{N}}\times(0,\tau] \tag{4.1e}$$

where $\tau \in \mathbb{R}^+$ is the time parameter, Re is the Reynolds number, $\mathbf{b}$ is the body force, σ is the stress tensor (see Eq. (3.54)) and $\hat{\mathbf{n}}$ is the outward unit normal vector to the boundary. In addition, $\bar{\mathbf{v}}(\mathbf{x})$ is the initial velocity profile in Ω, $\mathbf{v}^{\mathrm{p}}(\mathbf{x}, t)$ is the prescribed velocity on Γ^{D} and $\mathbf{t}^{\mathrm{p}}(\mathbf{x}, t)$ is the traction specified on Γ^{N}. We assume that $\nabla \cdot \bar{\mathbf{v}} = 0$ in Ω and that the problem is well posed. Whenever $\Gamma^{\mathrm{N}} = \varnothing$ we further prescribe the pressure at a single point in $\bar{\Omega}$.

1. The velocity-pressure-vorticity first-order system

As discussed in Chapter III, the Navier-Stokes equations as expressed in terms of the primitive variables $\mathbf{v}(\mathbf{x}, t)$ and $p(\mathbf{x}, t)$ are poorly suited for direct implementation in a least-squares finite element formulation. To allow for the use of practical C^0 basis functions in the numerical implementation we introduce the vorticity vector $\omega = \nabla \times \mathbf{v}$, which allows us to recast the Navier-Stokes equations in terms of the following equivalent first-order system problem statement: find the velocity $\mathbf{v}(\mathbf{x}, t)$, pressure $p(\mathbf{x}, t)$ and vorticity $\omega(\mathbf{x}, t)$ such that

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} + \nabla p + \frac{1}{\mathrm{Re}} \nabla \times \omega = \mathbf{b} \qquad \text{in } \Omega \times (0, \tau] \tag{4.2a}$$

$$\nabla \cdot \mathbf{v} = 0 \qquad \text{in } \Omega \times (0, \tau] \tag{4.2b}$$

$$\omega - \nabla \times \mathbf{v} = \mathbf{0} \qquad \text{in } \Omega \times (0, \tau] \tag{4.2c}$$

$$\mathbf{v}(\mathbf{x}, 0) = \bar{\mathbf{v}}(\mathbf{x}) \qquad \text{in } \Omega \tag{4.2d}$$

$$\mathbf{v} = \mathbf{v}^{\mathrm{p}}(\mathbf{x}, t) \qquad \text{on } \Gamma_{\mathbf{v}} \times (0, \tau] \tag{4.2e}$$

$$\omega = \omega^{\mathrm{p}}(\mathbf{x}, t) \qquad \text{on } \Gamma_{\omega} \times (0, \tau] \tag{4.2f}$$

$$\hat{\mathbf{n}} \cdot \tilde{\sigma} = \tilde{\mathbf{t}}^{\mathrm{p}}(\mathbf{x}, t) \qquad \text{on } \Gamma^{\mathrm{N}} \times (0, \tau] \tag{4.2g}$$

In the above expressions $\omega^{\mathrm{p}}(\mathbf{x}, t)$ is the prescribed vorticity on Γ_{ω}, $\tilde{\sigma} = -p\mathbf{I} + 1/\mathrm{Re} \nabla \mathbf{v}$ is the pseudo stress tensor and $\tilde{\mathbf{t}}^{\mathrm{p}}(\mathbf{x}, t)$ is the pseudo traction vector specified on Γ^{N}.

The Dirichlet part of the boundary has been partitioned such that $\Gamma^{\mathrm{D}} = \Gamma_{\mathbf{v}} \bigcup \Gamma_{\omega}$ and $\Gamma_{\mathbf{v}} \bigcap \Gamma_{\omega} = \varnothing$. The expression $\nabla(\nabla \cdot \mathbf{v})$ has been eliminated from the momentum equation on account of the solenoidal nature of the velocity field; as a result the outflow condition given in Eq. (4.2g) is preferred over Eq. (4.1e) [77]. For three-dimensional problems it is helpful to augment the first-order system with the seemingly redundant compatibility condition $\nabla \cdot \omega = 0$ in $\Omega \times (0, \tau]$ [35].

2. Temporal discretization

In this work we employ a space-time decoupled finite element approximation of the dependent variables. At each time step we approximate the time derivative of the velocity field using the backwards difference formula of order n (or BDFn)

$$\frac{\partial \mathbf{v}^{s+1}}{\partial t} \cong \frac{1}{\Delta t_{s+1}} \Big(\gamma_0^n \mathbf{v}^{s+1} - \sum\nolimits_{q=0}^{n-1} \beta_q^n \mathbf{v}^{s-q} \Big) \tag{4.3}$$

where $\Delta t_{s+1} = t_{s+1} - t_s$ is the time increment and γ_0^n and β_q^n are the temporal integration parameters. It is well-known that the backward difference formulas are particularly useful in the numerical solutions of stiff partial differential equations and differential-algebraic equations (DAEs). The backward difference formulas are especially valuable in achieving numerical stability and typically provide sufficient numerical dissipation of spurious high-frequency modes [78]. In this chapter we adopt the BDF1 and BDF2 formulas. Since the BDF2 time integrator is non-self-starting, we employ the BDF1 formula in the first few time steps.

3. The standard L_2-norm based least-squares formulation

The standard least-squares functional associated with the first-order vorticity form of the Navier-Stokes equations is constructed in terms of the sum of the squares of the L_2 norms of the partial differential equation residuals. In the space-time decoupled

formulation, we define the least-squares functional associated with the current time step $t = t_{s+1}$ as

$$\mathcal{J}_{\Delta t}(\mathbf{v}, p, \omega; \tilde{\mathbf{b}}, \tilde{\mathbf{t}}^{\mathrm{p}}) = \frac{1}{2}\Bigg(\alpha\|\lambda_0^n\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\mathbf{v} + \nabla p + \frac{1}{\mathrm{Re}}\nabla\times\omega - \tilde{\mathbf{b}}\|^2_{\Omega,0} + \|\nabla\cdot\mathbf{v}\|^2_{\Omega,0} + \|\omega - \nabla\times\mathbf{v}\|^2_{\Omega,0} + \|\mathbf{n}\cdot\tilde{\sigma} - \tilde{\mathbf{t}}^{\mathrm{p}}\|^2_{\Gamma^{\mathrm{N}},0}\Bigg) \tag{4.4}$$

where the quantities λ_0^n and $\tilde{\mathbf{b}}$ are defined as

$$\lambda_0^n = \gamma_0^n/\Delta t, \qquad \tilde{\mathbf{b}} = \mathbf{b} + \mathbf{v}_0\cdot\nabla\mathbf{v}_0 + \frac{1}{\Delta t}\sum\nolimits_{q=0}^{n-1}\beta_q^n\mathbf{v}^{s-q} \tag{4.5}$$

All quantities appearing in the definition of $\mathcal{J}_{\Delta t}$ are evaluated at the current time step $t = t_{s+1}$ unless explicitly noted otherwise. Newton's method has been employed in linearizing the momentum equation prior to minimization [20]. The weighting parameter α is taken as $\alpha = (\Delta t)^2$ to insure the discrete minimization problem is not extraneously dominated by the momentum equation residual in the limit as $\Delta t \to 0$.

The least-squares based weak formulation resulting from minimization of $\mathcal{J}_{\Delta t}$ may be stated as follows: find $\mathbf{u} \in \mathcal{V}$ such that

$$\mathcal{B}_{\Delta t}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}_{\Delta t}(\delta\mathbf{u}) \quad \text{for all } \delta\mathbf{u} \in \mathcal{W} \tag{4.6}$$

where $\mathbf{u} = (\mathbf{v}, p, \omega)$, $\delta\mathbf{u} = (\delta\mathbf{v}, \delta p, \delta\omega)$ and $\mathcal{V}$ and $\mathcal{W}$ are appropriate function spaces (see for example Eqs. (3.4) and (3.7)). The bilinear form $\mathcal{B}_{\Delta t}(\delta\mathbf{u}, \mathbf{u})$ and linear functional $\mathcal{F}_{\Delta t}(\delta\mathbf{u})$ are given as

$$\begin{aligned}\mathcal{B}_{\Delta t}(\delta\mathbf{u}, \mathbf{u}) = \int_\Omega \Bigg[\alpha\Bigg(\lambda_0^n\delta\mathbf{v} + \delta\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\delta\mathbf{v} + \nabla\delta p + \frac{1}{\mathrm{Re}}\nabla\times\delta\omega\Bigg)\cdot \\ \Bigg(\lambda_0^n\mathbf{v} + \mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\mathbf{v} + \nabla p + \frac{1}{\mathrm{Re}}\nabla\times\omega\Bigg) \\ + (\nabla\cdot\delta\mathbf{v})(\nabla\cdot\mathbf{v}) + (\delta\omega - \nabla\times\delta\mathbf{v})\cdot(\omega - \nabla\times\mathbf{v})\Bigg] d\Omega \\ + \int_{\Gamma^{\mathrm{N}}}\Bigg(-\delta p\hat{\mathbf{n}} + \frac{1}{\mathrm{Re}}\hat{\mathbf{n}}\cdot\nabla\delta\mathbf{v}\Bigg)\cdot\Bigg(-p\hat{\mathbf{n}} + \frac{1}{\mathrm{Re}}\hat{\mathbf{n}}\cdot\nabla\mathbf{v}\Bigg) d\Gamma^{\mathrm{N}}\end{aligned} \tag{4.7a}$$

$$\mathcal{F}_{\Delta t}(\delta\mathbf{u}) = \int_{\Omega} \alpha\left(\lambda_0^n \delta\mathbf{v} + \delta\mathbf{v}\cdot\nabla\mathbf{v}_0 + \mathbf{v}_0\cdot\nabla\delta\mathbf{v} + \nabla\delta p + \frac{1}{\mathrm{Re}}\nabla\times\delta\omega\right)\cdot\tilde{\mathbf{b}}d\Omega \quad (4.7b)$$
$$+ \int_{\Gamma^{\mathrm{N}}}\left(-\delta p\hat{\mathbf{n}} + \frac{1}{\mathrm{Re}}\hat{\mathbf{n}}\cdot\nabla\delta\mathbf{v}\right)\cdot\tilde{\mathbf{t}}^{\mathrm{p}}d\Gamma^{\mathrm{N}}$$

The least-squares finite element model associated with the above standard $(\mathbf{v}, p, \omega)$-space-time decoupled least-squares functional $\mathcal{J}_{\Delta t}$ often suffers from poor *local* mass conservation and can lead to an ill-behaved response (most notability in the pressure) as we march in time; this is especially true when Δt is small and α is taken as unity.

4. A modified L_2-norm based least-squares formulation with improved element-level mass conservation

The purpose of this section is to present a modified least-squares formulation that both enhances local mass conservation and improves velocity-pressure coupling. To this end we first recall that in traditional Galerkin based weak formulations, the pressure may be clearly identified as a Lagrange multiplier whose role is to enforce the divergence free constraint on the velocity field. In least-squares formulations, however, the pressure no longer possesses this well-defined role. In an effort to improve the function of the pressure in enforcing the continuity equation, Pontaza [52] proposed a penalized least-squares finite element model based on the following *regularized* form of the divergence free condition for the velocity

$$\nabla\cdot\mathbf{v} = -\epsilon\Delta p \qquad \text{in } \Omega\times(0,\tau] \quad (4.8)$$

where ϵ is a small parameter, $\Delta p = p_{k+1} - p_k$ and the index $k \in \mathbb{N}$ pertains to the iterative penalization of the divergence free constraint. The incompressibility constraint is recovered in either the limit as $\epsilon \to 0$ or $k \to \infty$ (assuming of course that the sequence $\{\Delta p\}_{k=0}^{\infty}$ is Cauchy). In practice the regularization may be adopted in conjunction with the iterative Newton solution procedure. Pontaza [52] demonstrated

numerically that using Eq. (4.8) in place of Eq. (4.2b) in the construction of $\mathcal{J}_{\Delta t}$ results in a significant improvement in the evolution of p for non-stationary flows.

With the regularized continuity equation in mind, we propose a novel *unconstrained* least-squares formulation that both enhances element-level mass conservation for steady flows and improves the temporal evolution of the pressure for non-stationary flows. The basic idea is to add directly to $\mathcal{J}_{\Delta t}$ a penalized sum of the squares of the appropriately *normalized* element-level integrals of Eq. (4.8). To make the concept clear, we consider the integral of Eq. (4.8) over an *arbitrary*, possibly time dependent, region $\mathcal{P}(t)$

$$\hat{Q}^{\mathcal{P}}(t) = \oint_{\partial\mathcal{P}(t)} \hat{\mathbf{n}} \cdot \mathbf{v} d\Gamma^{\mathcal{P}(t)} + \int_{\mathcal{P}(t)} \epsilon \Delta p d\mathcal{P}(t) \tag{4.9}$$

When the second term on the right hand side is neglected, the quantity $\hat{Q}^{\mathcal{P}}(t)$ may be clearly identified in the discrete setting as the volumetric flow rate imbalance associated with region $\mathcal{P}(t)$. Replacing $\mathcal{P}(t)$ with Ω^e in the above equation allows us to obtain the following expression for the eth element of the finite element discretization

$$\hat{Q}^{e}(t) = \oint_{\Gamma^e} \hat{\mathbf{n}} \cdot \mathbf{v} d\Gamma^{e} + \int_{\Omega^e} \epsilon \Delta p d\Omega^{e} \tag{4.10}$$

We find it is useful to normalize the above expression as $Q^e(t) = \hat{Q}^e(t)/\mu(\Omega^e)$ where $\mu(\Omega^e)$ denotes the Lebesgue measure or nd-dimensional volume of Ω^e. As a result, $Q^e(t)$ represents (when $\epsilon = 0$) the volumetric flow rate imbalance per nd-dimensional volume of Ω^e.

We are now in a position to define the following *modified* space-time decoupled L_2-norm least-squares functional for the first-order vorticity form of the incompressible Navier-Stokes equations

$$\mathcal{J}^{\star}_{\Delta t}(\mathbf{v}, p, \omega; \tilde{\mathbf{b}}, \tilde{\mathbf{t}}^{\mathrm{p}}) = \mathcal{J}_{\Delta t}(\mathbf{v}, p, \omega; \tilde{\mathbf{b}}, \tilde{\mathbf{t}}^{\mathrm{p}}) + \frac{\gamma}{2} \sum_{e=1}^{\mathrm{NE}} (Q^e)^2 \tag{4.11}$$

where γ is a global weight or penalty parameter. The *modified* least-squares based weak formulation resulting from minimization of $\mathcal{J}^{\star}_{\Delta t}$ may be stated as follows: find $\mathbf{u} \in \mathcal{V}$ such that

$$\mathcal{B}^{\star}_{\Delta t}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{F}^{\star}_{\Delta t}(\delta\mathbf{u}) \quad \text{for all } \delta\mathbf{u} \in \mathcal{W} \tag{4.12}$$

where the bilinear form $\mathcal{B}^{\star}_{\Delta t}(\delta\mathbf{u}, \mathbf{u})$ and linear form $\mathcal{F}^{\star}_{\Delta t}(\delta\mathbf{u})$ may be expressed as

$$\mathcal{B}^{\star}_{\Delta t}(\delta\mathbf{u}, \mathbf{u}) = \mathcal{B}_{\Delta t}(\delta\mathbf{u}, \mathbf{u}) + \gamma \sum_{e=1}^{\mathrm{NE}} \left(\oint_{\Gamma^e} \hat{\mathbf{n}} \cdot \delta\mathbf{v} d\Gamma^e + \int_{\Omega^e} \epsilon \delta p d\Omega^e \right) \times \tag{4.13a}$$
$$\left(\oint_{\Gamma^e} \hat{\mathbf{n}} \cdot \mathbf{v} d\Gamma^e + \int_{\Omega^e} \epsilon p d\Omega^e \right) / \mu(\Omega^e)^2$$

$$\mathcal{F}^{\star}_{\Delta t}(\delta\mathbf{u}) = \mathcal{F}_{\Delta t}(\delta\mathbf{u}) + \gamma \sum_{e=1}^{\mathrm{NE}} \left(\oint_{\Gamma^e} \hat{\mathbf{n}} \cdot \delta\mathbf{v} d\Gamma^e + \int_{\Omega^e} \epsilon \delta p d\Omega^e \right) \times \tag{4.13b}$$
$$\int_{\Omega^e} \epsilon p_0 d\Omega^e / \mu(\Omega^e)^2$$

Unlike $\mathcal{J}_{\Delta t}$, the modified least-squares functional $\mathcal{J}^{\star}_{\Delta t}$ clearly includes both element-level mass conservation as well as velocity-pressure coupling. Working in terms of $\mathcal{J}^{\star}_{\Delta t}$ leads to an unconstrained minimization problem that may be viewed as an attractive alternative to the Lagrange multiplier based least-squares model of the Stokes equations proposed by Chang and Nelson [51]. For stationary flows, we find that it is sufficient to take $\epsilon = 0$.

B. Numerical examples: verification benchmarks

In this section we present numerical results obtained using the proposed least-squares formulation. The problems have been selected to assess the capabilities of the formulation to: (a) generally improve element-level mass conservation and (b) enhance velocity-pressure coupling and overall numerical stability in non-stationary flows.

1. Stationary flow

In what follows, we test the performance of the proposed least-squares formulation to improve mass conservation for problems involving steady fluid flows. We utilize the stationary least-squares functionals $\mathcal{J}$ and $\mathcal{J}^\star$, obtained by setting $\alpha = 1$ and $\gamma_0^n = \beta_q^n = 0$ (where $q = 0, \ldots, n-1$) in the definitions of $\mathcal{J}_{\Delta t}$ and $\mathcal{J}_{\Delta t}^\star$ respectively.

a. Kovasznay flow

In this first example, we numerically examine a well-known incompressible fluid flow problem possessing an analytic solution. The solution is due to Kovasznay [79] and is posed on a two dimensional square region defined as $\bar{\Omega} = [-0.5, 1.5] \times [-0.5, 1.5]$. The proposed solution is of the form

$$v_x = 1 - e^{\lambda x}\cos(2\pi y), \qquad v_x = \frac{\lambda}{2\pi}e^{\lambda x}\sin(2\pi y), \qquad p = p_{\text{ref}} - \frac{1}{2}e^{2\lambda x} \tag{4.14}$$

where the parameter λ is given as $\lambda = \text{Re}/2 - [(\text{Re}/2)^2 + (2\pi)^2]^{1/2}$ and p_{ref} is a reference pressure (which in the current study is taken to be zero).

We discretize the domain $\bar{\Omega}$ into 8 non-uniform rectangular finite elements as depicted in Figure 23 (a). Figure 23 (b) shows the numerically computed horizontal velocity component v_x. The boundary conditions for the problem are applied by specifying the exact solution for the velocity vector $\mathbf{v}$ along the entire boundary through an employment of Eq. (4.14). We specify no boundary conditions for the vorticity and only prescribe the pressure at the single point $\mathbf{x} = (-0.5, 0)$. In this study the mesh is refined by systematically increasing the p-level of the finite element approximation within each element. Nonlinear convergence for a given numerical simulation is declared once the relative Euclidean norm of the solution residuals, $\|\Delta^k - \Delta^{k-1}\|/\|\Delta^k\|$, is less than 10^{-6}. All reported numerical results have been

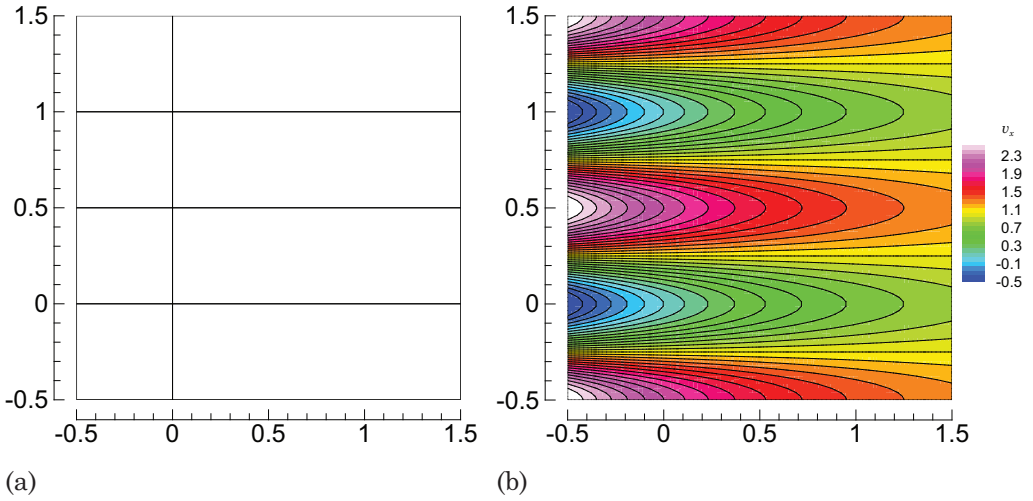


Fig. 23. Kovasznay flow: (a) spectral/*hp* finite element discretization of domain $\bar{\Omega}$ and (b) numerical solution of horizontal velocity field v_x for Re = 40.

obtained for a Reynolds number of 40.

In Table VI we report the decay of the $L_2(\Omega)$-norm error measures for the velocity, pressure and vorticity fields under p-refinement, where the penalty parameter γ is varied from 0 to 100. We also show the decay of the *unmodified* least-squares functional $\mathcal{J}$. We observe exponential decay in the error measures for all variables as the p-level is increased. This observation is true for all values of γ considered. Figure 24 shows the evolution of the error measures under p-refinement for the case where $\gamma = 100$. Clearly, the inclusion of element-level mass conservation in the definition of $\mathcal{J}^\star$ does not pollute the integrity of the finite element solution.

Figure 25 shows the decay of the normalized volumetric flow rate imbalance Q^e for element 1 under p-refinement for $\gamma = 0, 1, 10$ and 100, where $\bar{\Omega}^1 = [-0.5, 0] \times [-0.5, 0]$. The normalized volumetric flow rate imbalance Q^e associated with each element in $\bar{\Omega}^{hp}$ for p-levels 3 and 7 is also provided in Figure 26. Although p-refinement clearly improves local mass conservation, significant additional enhancement may be obtained

Table VI. Kovasznay flow: Decay of the unmodified least-squares functional $\mathcal{J}$ and convergence of the numerically computed velocity components, pressure and vorticity in the $L_2(\Omega)$-norm under p-refinement for various values of the penalty parameter γ.

γ	p-level	$\mathcal{J}^{1/2}$	$\|v_x - v_x^{hp}\|_{\Omega,0}$	$\|v_y - v_y^{hp}\|_{\Omega,0}$	$\|p - p_{hp}\|_{\Omega,0}$	$\|\omega - \omega_{hp}\|_{\Omega,0}$
0	3	1.5061 E-01	1.7124 E-02	6.2962 E-03	7.3897 E-03	1.9290 E-01
	5	4.5738 E-03	2.6574 E-04	1.1853 E-04	1.4078 E-04	3.9751 E-03
	7	6.7688 E-05	2.1663 E-06	1.1941 E-06	6.9539 E-07	4.4836 E-05
	9	5.8185 E-07	1.2070 E-08	8.0882 E-09	5.4114 E-09	3.2600 E-07
1	3	1.5061 E-01	1.7089 E-02	6.2757 E-03	7.4704 E-03	1.9274 E-01
	5	4.5738 E-03	2.6434 E-04	1.1868 E-04	1.3214 E-04	3.9711 E-03
	7	6.7688 E-05	2.1659 E-06	1.1941 E-06	6.9467 E-07	4.4836 E-05
	9	5.8185 E-07	1.2067 E-08	8.0881 E-09	5.3552 E-09	3.2600 E-07
10	3	1.5061 E-01	1.7080 E-02	6.2675 E-03	7.4982 E-03	1.9267 E-01
	5	4.5739 E-03	2.6425 E-04	1.1873 E-04	1.2996 E-04	3.9703 E-03
	7	6.7688 E-05	2.1658 E-06	1.1941 E-06	6.9442 E-07	4.4836 E-05
	9	5.8185 E-07	1.2067 E-08	8.0881 E-09	5.3407 E-09	3.2600 E-07
100	3	1.5061 E-01	1.7079 E-02	6.2662 E-03	7.5023 E-03	1.9266 E-01
	5	4.5739 E-03	2.6424 E-04	1.1874 E-04	1.2968 E-04	3.9702 E-03
	7	6.7688 E-05	2.1657 E-06	1.1941 E-06	6.9438 E-07	4.4836 E-05
	9	5.8185 E-07	1.2067 E-08	8.0881 E-09	5.3381 E-09	3.2600 E-07

through constructing the least-squares finite element model using the modified least-squares functional $\mathcal{J}^{\star}$ as opposed to the standard least-squares functional $\mathcal{J}$.

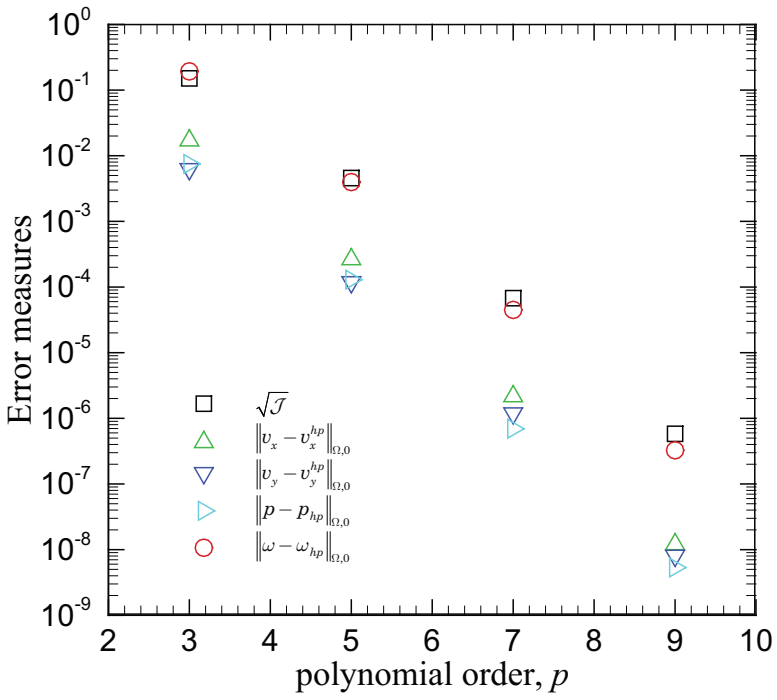


Fig. 24. Convergence of numerically computed velocity, pressure and vorticity under p-refinement to the analytic solution of Kovasznay for $\gamma = 100$. The decay of the square root of the unmodified least-squares functional $\mathcal{J}$ is also shown.

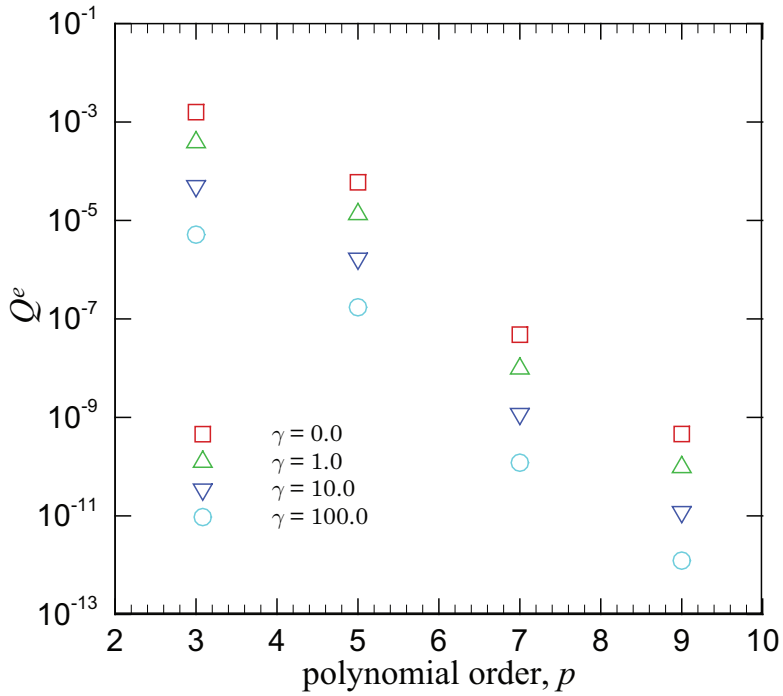


Fig. 25. Decay of normalized volumetric flow rate imbalance Q^e under p-refinement for various values of γ for Kovasznay flow. Results are for element 1, where $\bar{\Omega}^1 = [-0.5, 0] \times [-0.5, 0]$.

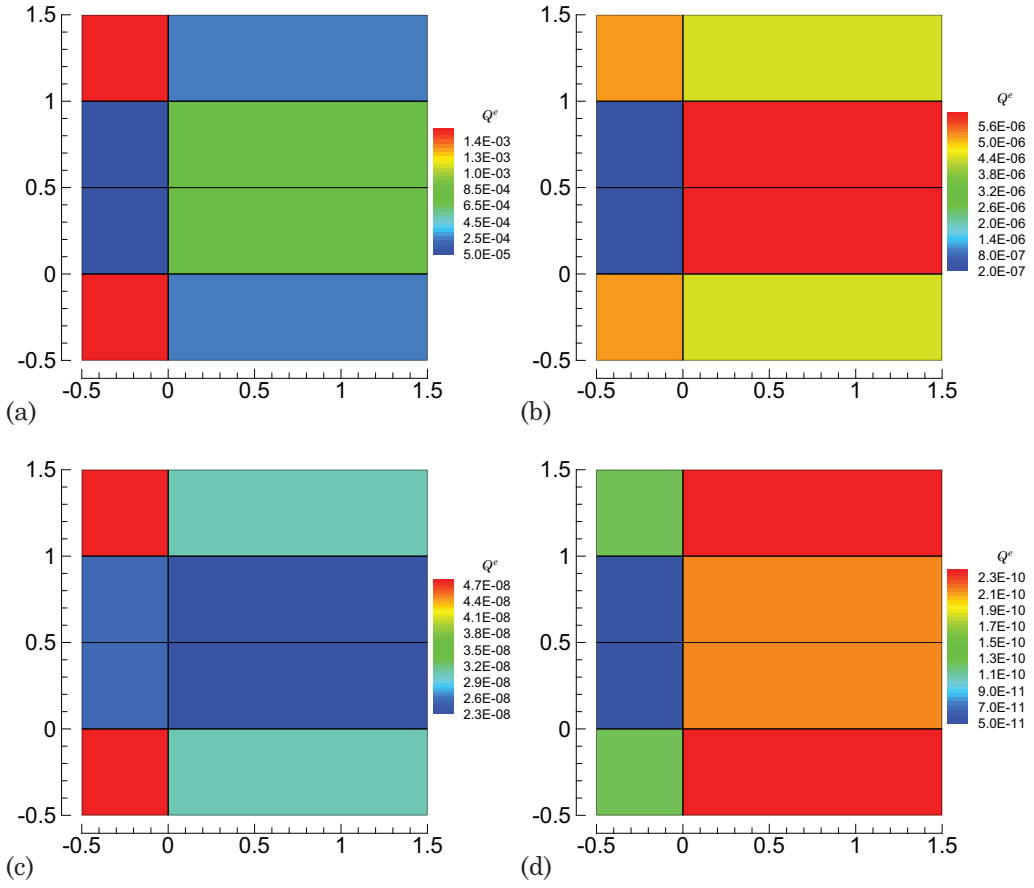


Fig. 26. Normalized volumetric flow rate imbalance Q^e for each finite element in $\bar{\Omega}^{hp}$ for Kovasznay flow. Results are for various polynomial orders and values of γ: (a) p-level = 3 and $\gamma = 0$, (b) p-level = 3 and $\gamma = 100$, (c) p-level = 7 and $\gamma = 0$ and (d) p-level = 7 and $\gamma = 100$.

b. Flow in a 1×2 rectangular cavity at Re = 1,500

In this next example, we test the modified least-squares finite element formulation on a problem posed at an elevated Reynolds number. To this end we consider a two-dimensional lid-driven cavity flow problem as posed on a rectangular domain with an aspect ratio of 2, where $\bar{\Omega} = [0, 1] \times [0, 2]$. As in the previous example, the boundary conditions are specified in terms of the velocity vector and the pressure at a single point only. Along the bottom and left and right sides of the cavity, the velocity is taken to be zero in accordance with the no slip condition. It is common practice in the literature to prescribe a unit value for the horizontal velocity component v_x along the entire top surface of the cavity. In the context of high-order finite elements, however, such a boundary condition produces undesirable singularities in the vicinity of the upper cavity corners. In an effort to avoid an ill-posed *discrete* problem, the horizontal velocity profile is instead prescribed in terms of the following expression

$$v_x^{\mathrm{p}}(x) = \begin{cases} \tanh(50x) & 0 \leqslant x \leqslant 0.5 \\ -\tanh[50(x-1)] & 0.5 < x \leqslant 1.0 \end{cases} \tag{4.15}$$

which allows for a smooth transition from 1.0 to zero in the neighborhoods of the corners as can be seen in Figure 27. The pressure is taken to be zero at $\mathbf{x} = (0.5, 0)$. The Reynolds number for the problem is taken to be 1,500.

The finite element mesh consists of a 12×24 discretization of 288 elements as shown in Figure 28 (a). The mesh is graded so as to adequately resolve the anticipated boundary layers and regions of circulation near the cavity walls. Nonlinear convergence of the iterative solution procedure is declared once the relative norm of the solution residuals, $\|\Delta^k - \Delta^{k-1}\|/\|\Delta^k\|$, is less than 10^{-6}. We employ a continuation approach, wherein we arrive at the solution at Re = 1,500 by solving a series of problems posed at intermediate Reynolds numbers. We begin by solving the problem

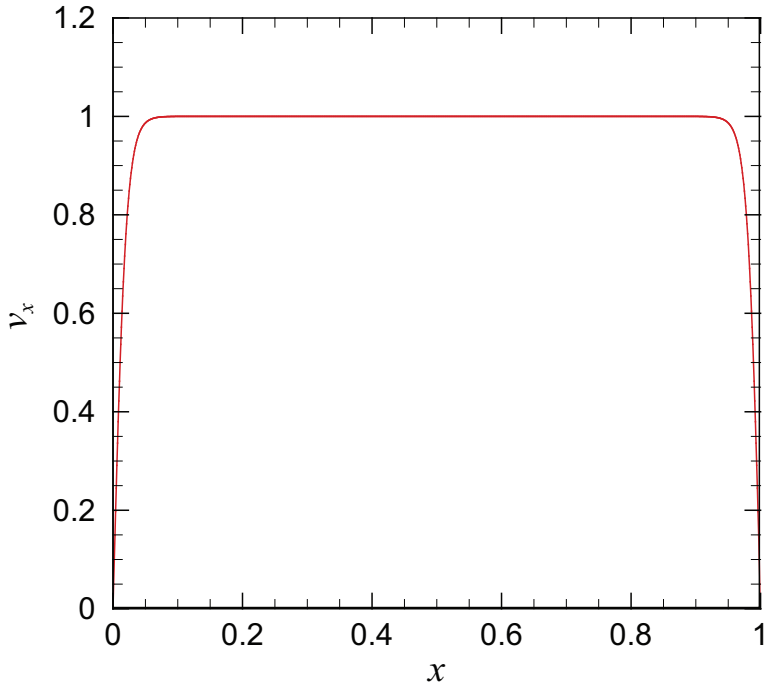


Fig. 27. Specified horizontal velocity profile v_x along the top surface of the 2-D lid–driven cavity flow problem with aspect ratio of 2.

at Re = 300 followed by Re = 600 and so on until we reach Re = 1,500. For each problem, the converged solution taken from the immediate previous problem in the series is used as the initial guess. The problem is solved using the modified least-squares finite element formulation taking γ as 0, 0.1, 1.0 and 10.0.

Figure 28 (b) shows the vorticity field and streamlines for the problem. The flow is characterized by two large regions of circulation, with smaller vortex regions also present in the vicinity of the bottom as well as the upper left hand corners of the domain. In Figure 28 (c) we show the horizontal velocity component v_x along the vertical centerline of the domain as determined using a p-level of 9 and $\gamma = 10.0$. The streamlines and horizontal velocity component v_x along the vertical centerline are visually in excellent agreement with the results reported by Gupta and Kalita [80].

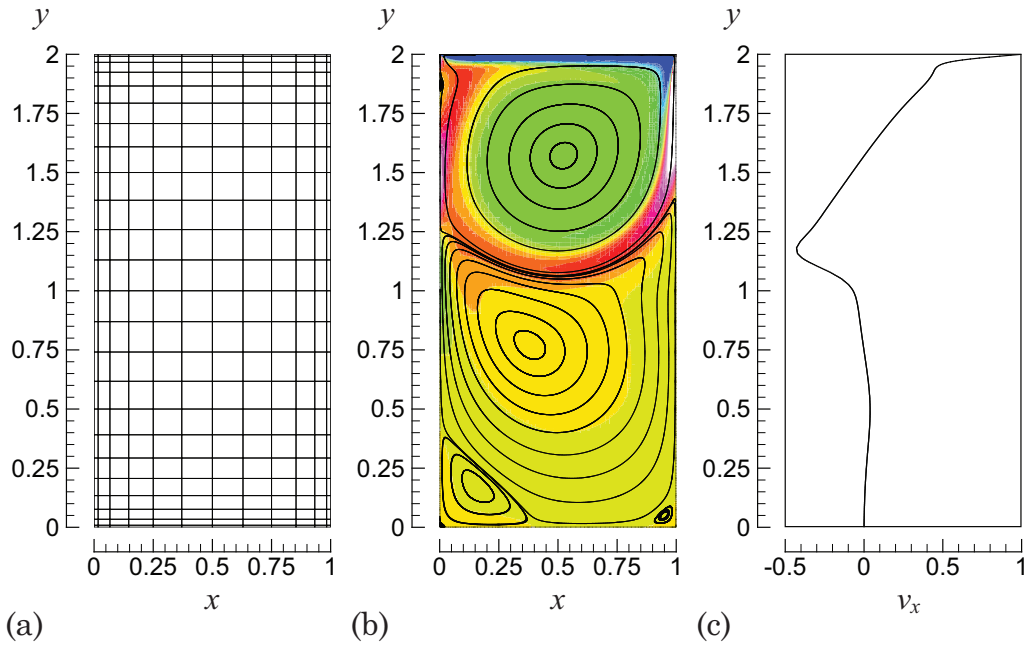


Fig. 28. 2-D lid-driven cavity flow with aspect ratio of 2 at Re = 1,500. Numerical results obtained for p-level of 9: (a) finite element mesh, (b) vorticity field and streamlines and (c) horizontal velocity component v_x along vertical cavity centerline.

In Figure 29 we show the decay of the normalized volumetric flow rate imbalance Q^e for element 107, where the geometric centroid of the element is located at $\mathbf{x} = (0.3103, 0.8053)$. In this figure, both the p-level and the penalty parameter γ are varied. We also provide in Figure 30 an illustration of the normalized volumetric flow rate imbalance Q^e for all elements in the discretization at p-levels 5 and 7 for various values for the penalty parameter γ. Clearly, both the polynomial order as well as the value of γ are significant factors in improving element-level mass conservation for this problem. It is interesting to note that substantial improvement in element-level mass conservation is obtained even for the case where $\gamma = 1.0$.

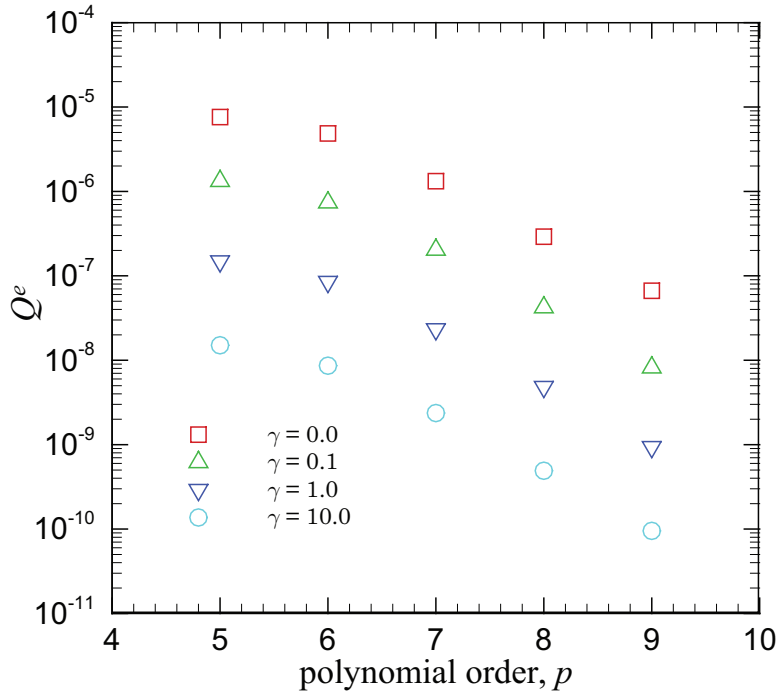


Fig. 29. Decay of normalized volumetric flow rate imbalance Q^e under p-refinement for various values of γ for 2-D lid-driven flow in a rectangular cavity. Results shown are for element 107 with geometric centroid located at $\mathbf{x} = (0.3103, 0.8053)$.

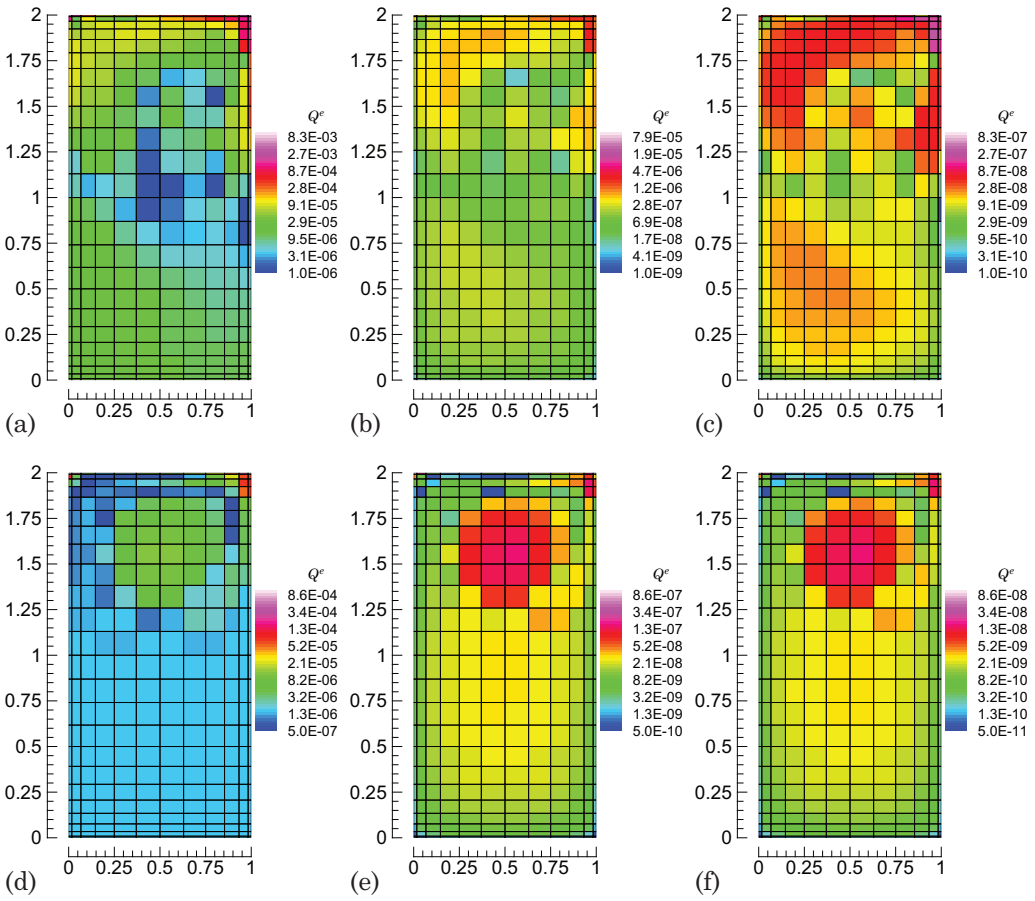


Fig. 30. Normalized volumetric flow rate imbalance Q^e for each finite element in $\bar{\Omega}^{hp}$ for 2-D lid-driven flow in a rectangular cavity. Results are for various polynomial orders and values of γ: (a) p-level = 5 and $\gamma = 0$, (b) p-level = 5 and $\gamma = 1.0$, (c) p-level = 5 and $\gamma = 10.0$, (d) p-level = 7 and $\gamma = 0$, (e) p-level = 7 and $\gamma = 1.0$ and (f) p-level = 7 and $\gamma = 10.0$.

c. Flow past a large cylinder in a narrow channel at Re = 40

As a final steady flow example, we consider a problem constituting a much more rigorous test for mass conservation, namely flow past a circular cylinder in a narrow channel. In this problem (which is similar to the test problem used by Chang and Nelson [51]) we take for the domain $\bar{\Omega}$, the set difference between the closed rectangular region $[-5, 10] \times [-1, 1]$ and an open circular region with unit diameter centered about the origin. Along the inflow part of the boundary (i.e., the left hand side of the domain) we prescribe a parabolic horizontal velocity profile $v_x^{\mathrm{p}}(y) = \frac{3}{2}(1 - y^2)$, which is consistent with Poiseuille flow. The pseudo-traction is taken to be zero on the right hand side of the domain and a no-slip condition is utilized on all other parts of the boundary. The Reynolds number, based on the diameter of the cylinder and the average horizontal velocity at the inlet is taken to be 40.

The finite element mesh utilized in our numerical investigation is shown in Figure 31 (a). The problem is solved by varying the p-level incrementally from 2 to 7. At each p-level we further investigate the influence of the penalty parameter on improving mass conservation by solving the problem for $\gamma = 0$, 1, 10 and 100. We adopt the same nonlinear convergence criteria for the iterative solution procedure that was used in the two previous stationary benchmark problems. The horizontal velocity profile in the domain is shown in Figure 31 (b). The element-level normalized volumetric flow rate imbalance Q^e for each element in the discretization are shown in Figure 31 (c)–(f) for p-levels 2 and 3, where γ is taken as either 0 or 100. In Figure 32 we show the horizontal velocity profile along the gap between the cylinder and the top of the channel at $x = 0$. The so-called "exact" solution in this figure is the finite element solution obtained using a p-level of 7 and $\gamma = 0$. Figure 33 shows the observed volumetric flow rate imbalance for element 115 of the finite element model, where

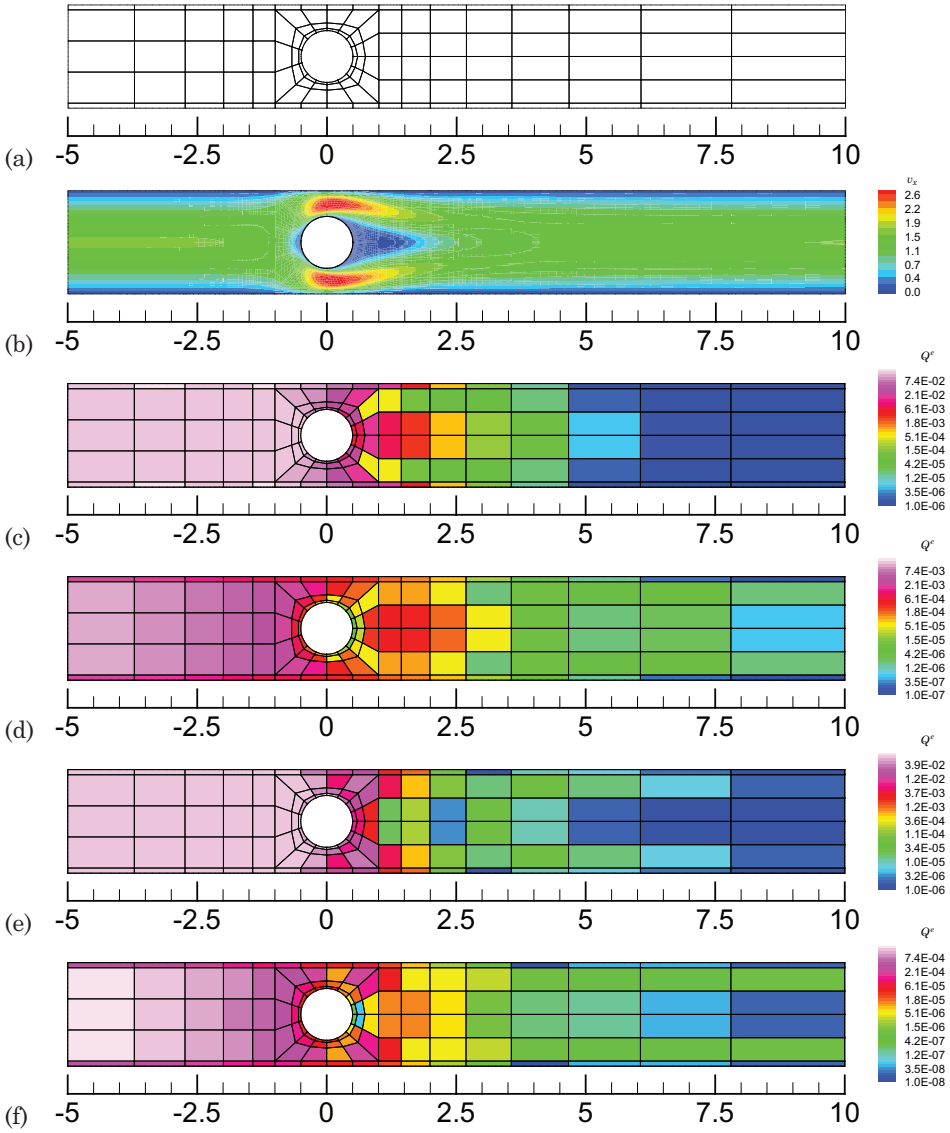


Fig. 31. Finite element mesh, horizontal velocity component v_x and normalized volumetric flow rate imbalance Q^e for each finite element in $\bar{\Omega}^{hp}$ for steady flow past a large circular cylinder in a narrow channel at Re = 40: (a) finite element mesh, (b) horizontal velocity component v_x, (c) Q^e for p-level = 2 and $\gamma = 0$, (d) Q^e for p-level = 2 and $\gamma = 100$, (e) Q^e for p-level = 3 and $\gamma = 0$ and (f) Q^e for p-level = 3 and $\gamma = 100$.

both the p-level and penalty parameter γ are varied. The normalized volumetric flow rate past the crown of the cylinder for the various finite element discretizations is summarized in Table VII. Clearly, mass conservation is improved by constructing the finite element model in terms of the modified least-squares functional $\mathcal{J}^{\star}$.

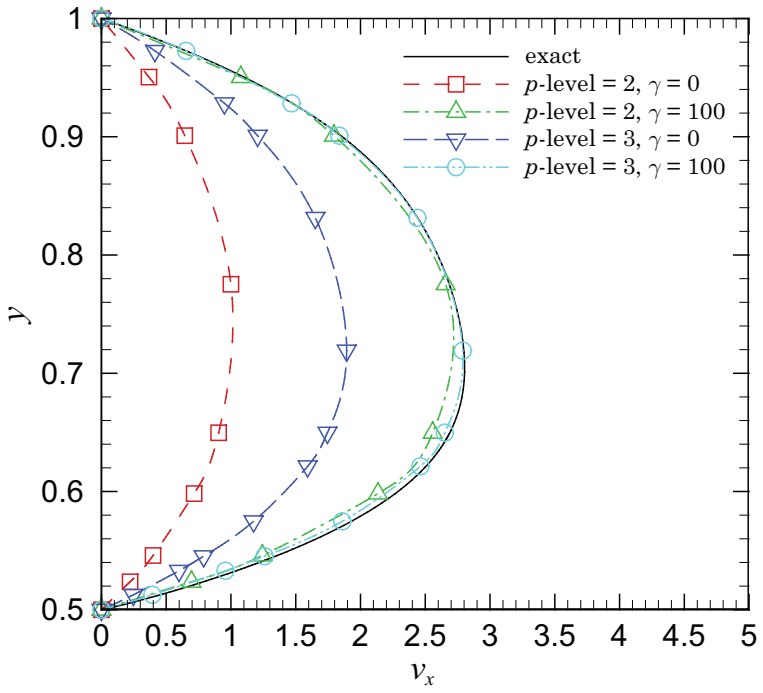


Fig. 32. Horizontal velocity $v_x(0, y)$ profiles along the gap between the top of the circular cylinder and the channel wall at $x = 0$ for flow past a large cylinder in a narrow channel.

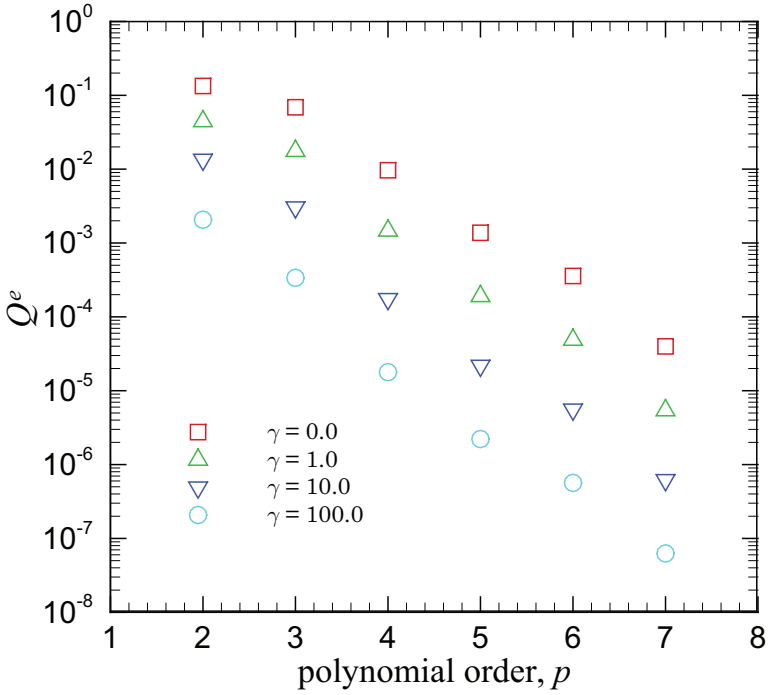


Fig. 33. Decay of normalized volumetric flow rate imbalance Q^e under p-refinement for various values of γ for flow past a large cylinder in a narrow channel. Results shown are for element 115 with geometric centroid located at $\mathbf{x} = (-0.8365, 0)$.

Table VII. Normalized volumetric flow rate past the crown $(x = 0, y)$ of the large circular cylinder.

	Normalized volumetric flow rate			
p-level	$\gamma = 0$	$\gamma = 1$	$\gamma = 10$	$\gamma = 100$
2	0.34870	0.53887	0.81899	0.97124
3	0.65973	0.81925	0.95983	0.99538
4	0.95250	0.98485	0.99767	0.99975
5	0.99335	0.99951	0.99971	0.99997
6	0.99828	0.99951	0.99993	0.99999

2. Transient flow

In this section we assess the performance of the proposed least-squares formulation to improve mass conservation, velocity-pressure coupling and overall numerical stability in the numerical simulation of non-stationary fluid flows. Unless otherwise stated, we take $\alpha = (\Delta t)^2$ in the definition of $\mathcal{J}^{\star}_{\Delta t}$ for each numerical simulation.

a. Flow past a circular cylinder at Re = 100

As an inaugural non-stationary example we consider the standard flow past a circular cylinder problem, where the Reynolds number is taken as 100. For the computational domain Ω, we take the set difference between the open square region $(-15.5, 25.5) \times (-20.5, 20.5)$ and a closed unit-diameter circle that is centered about the origin. The spatial discretizations $\bar{\Omega}^{hp}$ that are employed in the finite element simulations are shown in Figure 34. The top mesh contains 2,004 quadratic elements (i.e., the p-level is 2). Likewise, the bottom mesh contains 501 elements, where the p-level is taken as 4. Each discretization contains 8,216 nodes and 32,864 total degrees of freedom.

All flow fields are initially taken to be zero. The horizontal velocity component v_x is then gradually increased in time along the left, top and bottom sides of $\bar{\Omega}^{hp}$ in accordance with the formula $v_x^{\mathrm{p}}(t) = v_\infty \tanh(t)$; the free-stream velocity v_∞ is taken to be 1.0. A no-slip boundary condition is used along the circular cylinder and the outflow boundary condition (along the right hand side of the domain) is enforced weakly by taking the pseudo-traction $\tilde{\mathbf{t}}^{\mathrm{p}}$ as zero in the definition of $\mathcal{J}^{\star}_{\Delta t}$. We employ the BDF2 time integrator with a uniform time step size of $\Delta t = 0.1$ sec. Since the BDF2 integration formula is non-self-starting, we utilize the BDF1 formula for the first 10 time steps. A total of 3,000 time steps are employed in each transient finite element simulation. A nonlinear convergence criteria of $\varepsilon = 10^{-6}$, defined in terms

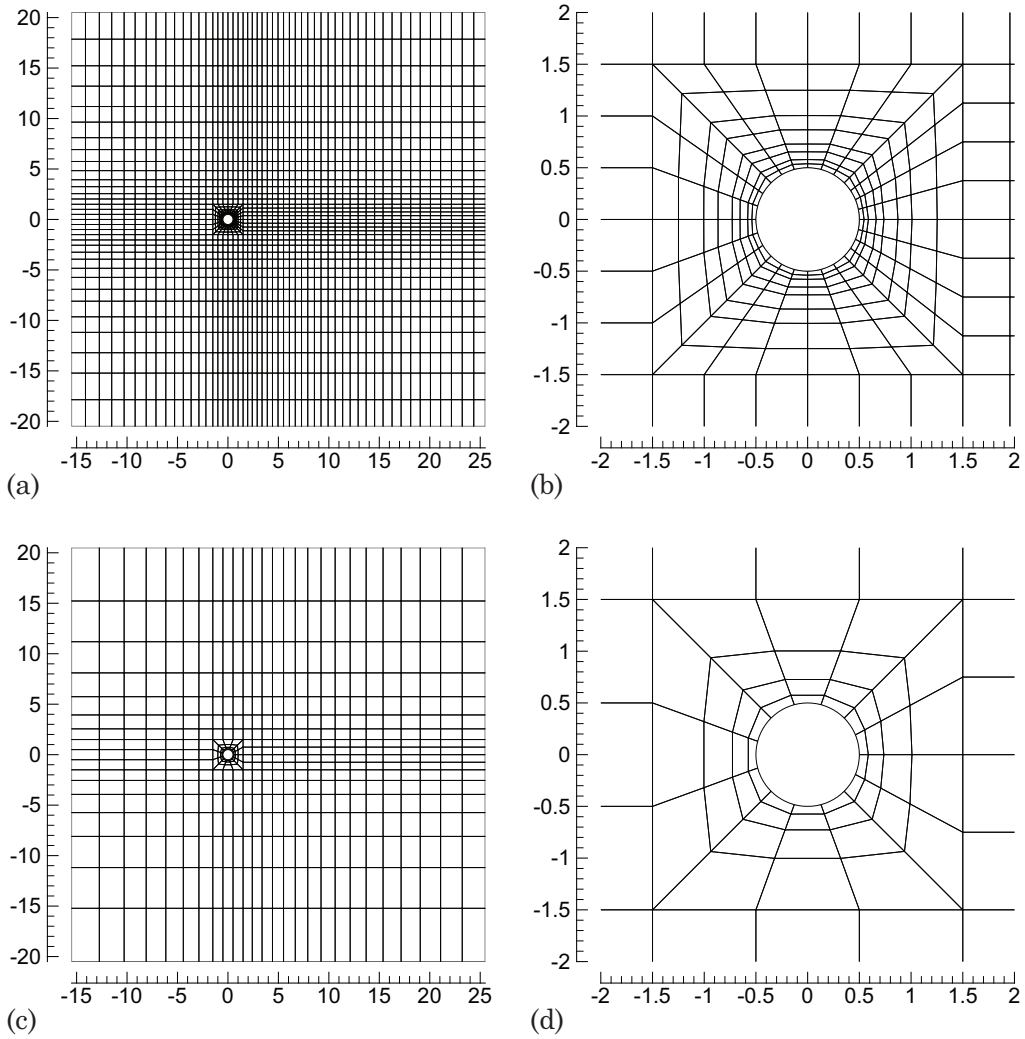


Fig. 34. Spectral/*hp* finite element discretizations used to numerically simulate the unsteady viscous flow of an incompressible fluid past a circular cylinder: (a) computational domain $\bar{\Omega}^{hp}$ for a p-level of 2, (b) close-up view of $\bar{\Omega}^{hp}$ in the vicinity of the cylinder for a p-level of 2, (c) computational domain $\bar{\Omega}^{hp}$ for a p-level of 4 and (d) close-up view of $\bar{\Omega}^{hp}$ in the vicinity of the cylinder for a p-level of 4.

of the relative Euclidean norm of the residuals in the nodal velocities between two successive iterations, is adopted at each time step. This typically requires only 2 or 3 nonlinear iterations. The linearized algebraic equations are constructed and solved using the sparse finite element equation solution procedures outlined in Chapter II; the UMFPACK direct solver library [27, 28, 29, 30] is utilized in the numerical solution of the global sparse set of finite element equations. We solve for the temporal evolution of the fluid using the modified least-squares functional $\mathcal{J}^{\star}_{\Delta t}$ for the cases where γ is either 0 or 100. Although all results reported below have been obtained by taking ϵ as zero, we note in passing that we have also obtained reliable solutions using $\epsilon = 0.005$ and 0.01. The additional velocity-pressure coupling associated with a non-zero value for ϵ, however, typically demands a greater number of nonlinear iterations to meet the nonlinear convergence criterion of $\varepsilon = 10^{-6}$.

In Figure 35 we show the time history of the velocity components, vorticity and pressure at the spatial point $(x, y) = (1, 0)$ as computed using the finite element mesh shown in Figure 34 (c) with γ taken as 100. The dimensional pressure field shown in Figure 35 (d) has been obtained by scaling the non-dimensional pressure field by a factor of 100. The *virtually* stationary flow pattern that forms during the early stages of the simulation becomes noticeably unstable between 150 and 175 sec. The instability eventually results in a well-defined periodic swirling of vortices that are shed in the wake region immediately downwind of the cylinder. This oscillatory behavior is commonly referred to as the von Kármán vortex street. We measure the non-dimensional period to be $\mathcal{T} = 6.035$; this translates into a non-dimensional shedding frequency (or Strouhal number) of St = 0.1657. This is in very close agreement with St = 0.1653 reported by Pontaza and Reddy [17] using a space-time coupled spectral/*hp* least-squares finite element simulation.

Instantaneous contours of the velocity components v_x and v_y along with the

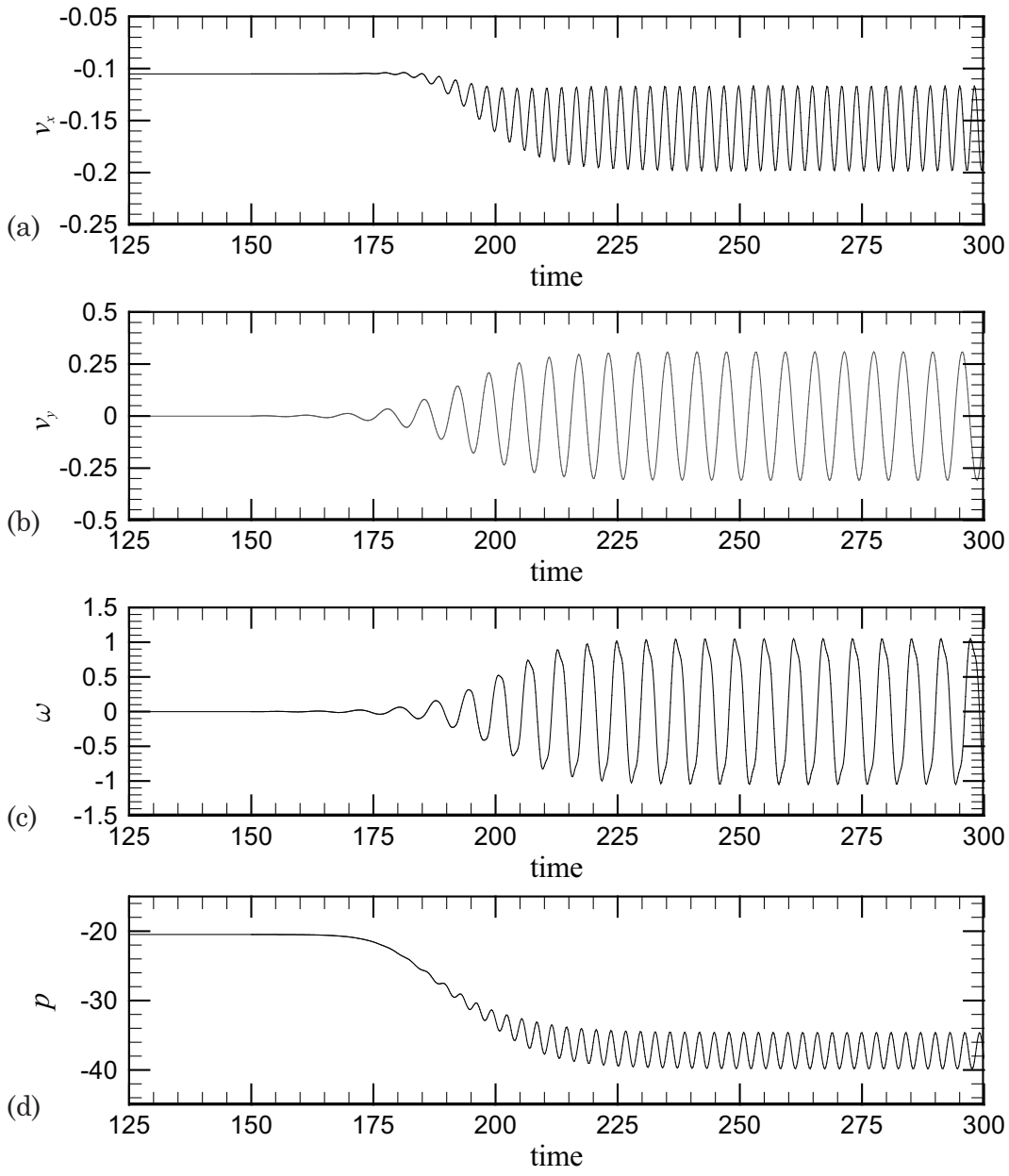


Fig. 35. Time history of the flow fields behind the circular cylinder at $(x, y) = (1, 0)$ as determined using a p-level of 4 in the spatial discretization: (a) horizontal velocity component v_x, (b) vertical velocity component v_y, (c) vorticity ω and (d) pressure field p.

pressure field p in the wake region are depicted in Figure 36 at $t = 280$ sec. We also provide in Figure 37 snapshots of the vorticity field ω during the course of a single shedding cycle. Both figures have been generated using the numerical results obtained using a p-level of 4 and $\gamma = 100$. We see that within a given period $\mathcal{T}$ two eddies are shed from the cylinder into the wake region, one originating from the top and the other from the bottom of the cylinder. The former eddy spins clockwise while the latter rotates in the counterclockwise direction. The outflow boundary condition, imposed weakly through the least-squares functional $\mathcal{J}^{\star}_{\Delta t}$, clearly allows the fluid to leave the computational domain in a physically reasonable manner.

In an effort to showcase the performance of the modified least-squares formulation in improving local mass conservation, we present in Figure 38 the normalized volumetric flow rate imbalance Q^e for the finite elements in a neighborhood of the wake region behind the circular cylinder. The reported results are for the numerical solution obtained at $t = 260$ sec. using the spatial discretization shown in Figure 34 (c). We see that element-level mass conservation is clearly improved by taking γ as 100 as opposed to 0 in the modified least-squares formulation. The improvement is particularly noticeable for the smaller elements in $\bar{\Omega}^{hp}$ that are closest to the cylinder. To assess general mass conservation for the fluid flowing past the circular cylinder, we post-compute the absolute value of the volumetric flow rate

$$Q(t) = \left| \oint_{\Gamma^s} \hat{\mathbf{n}} \cdot \mathbf{v}_{hp}(t) d\Gamma^s \right| \tag{4.16}$$

across the closed surface $\Gamma^s = \partial\Omega^s$, where $\Omega^s = (-1.5, 1.5)^2$. We numerically evaluate Q at each time step using the Gauss-Legendre quadrature rule. In Figure 39 we trace the time history of Q for both spatial discretizations where γ is again taken as either 0 or 100. For both spatial discretizations we observe significant improvement in mass conservation across Γ^s when γ is taken as 100.

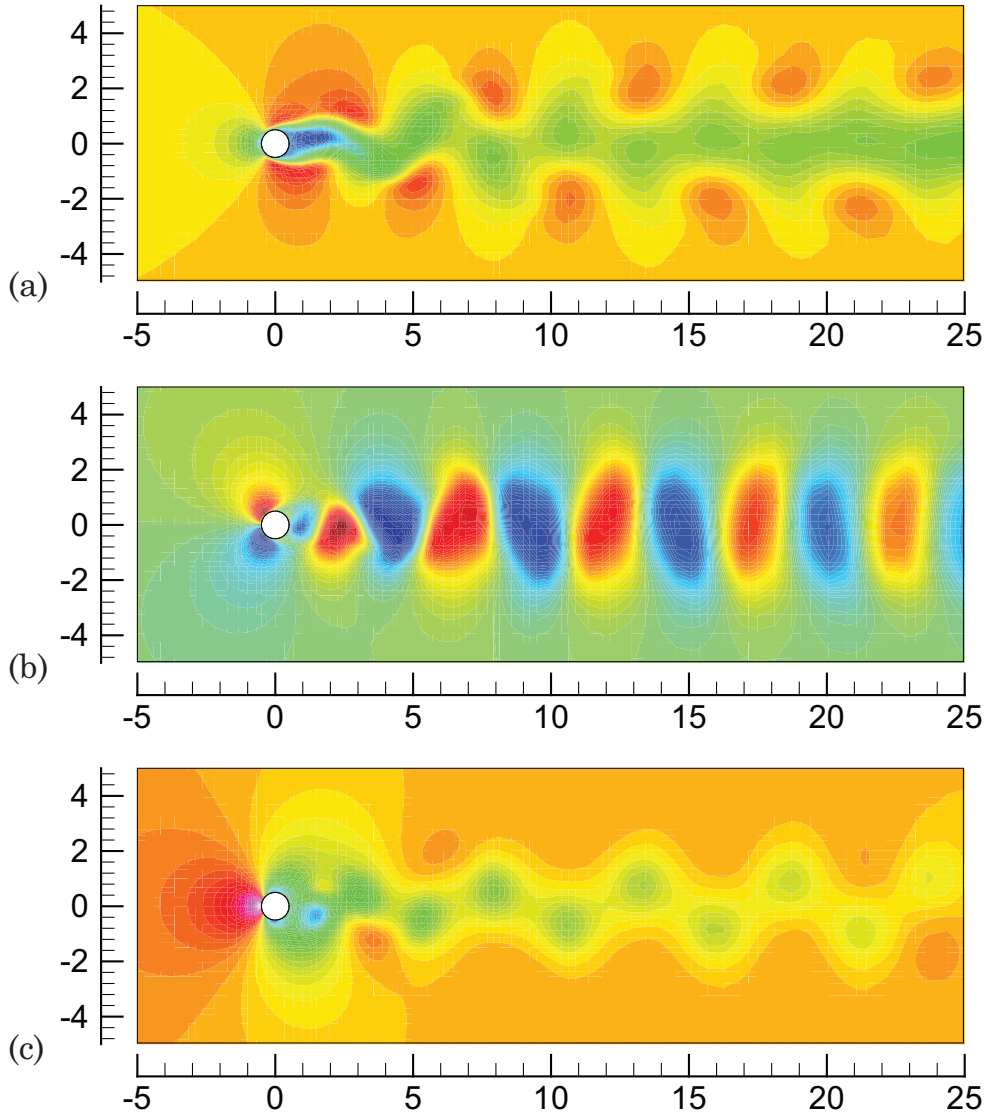


Fig. 36. Instantaneous contours for flow past a circular cylinder at $t = 280$ sec., where the finite element mesh associated with a p-level of 4 has been utilized: (a) horizontal velocity component v_x, (b) vertical velocity component v_y and (c) pressure field p.

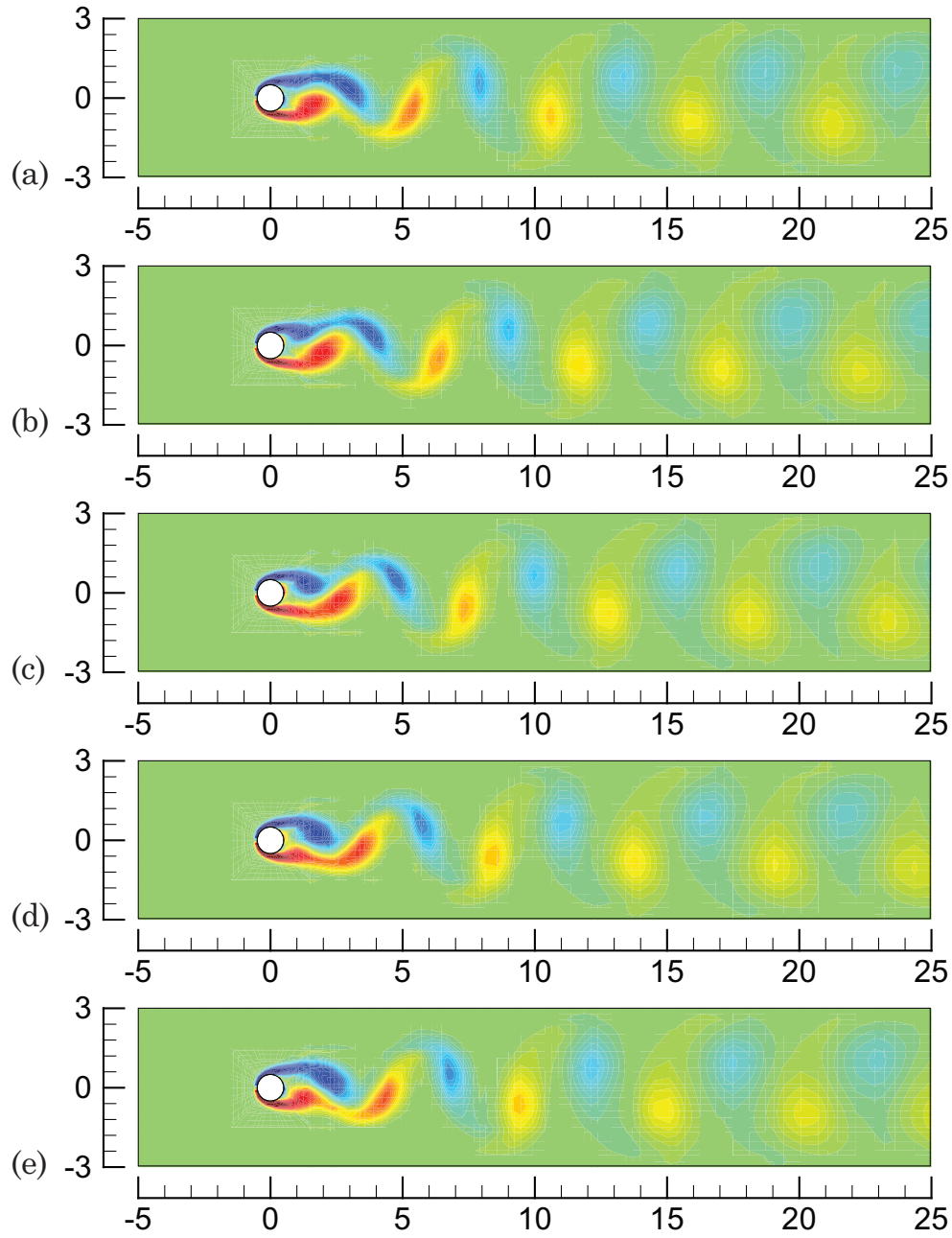


Fig. 37. Time history of vorticity contours behind the circular cylinder at five successive discrete instances in time. The finite element mesh associated with a p-level of 4 has been employed: (a) $t = 280.0$ sec., (b) $t = 281.2$ sec., (c) $t = 282.4$ sec., (d) $t = 283.6$ sec. and (e) $t = 284.8$ sec.

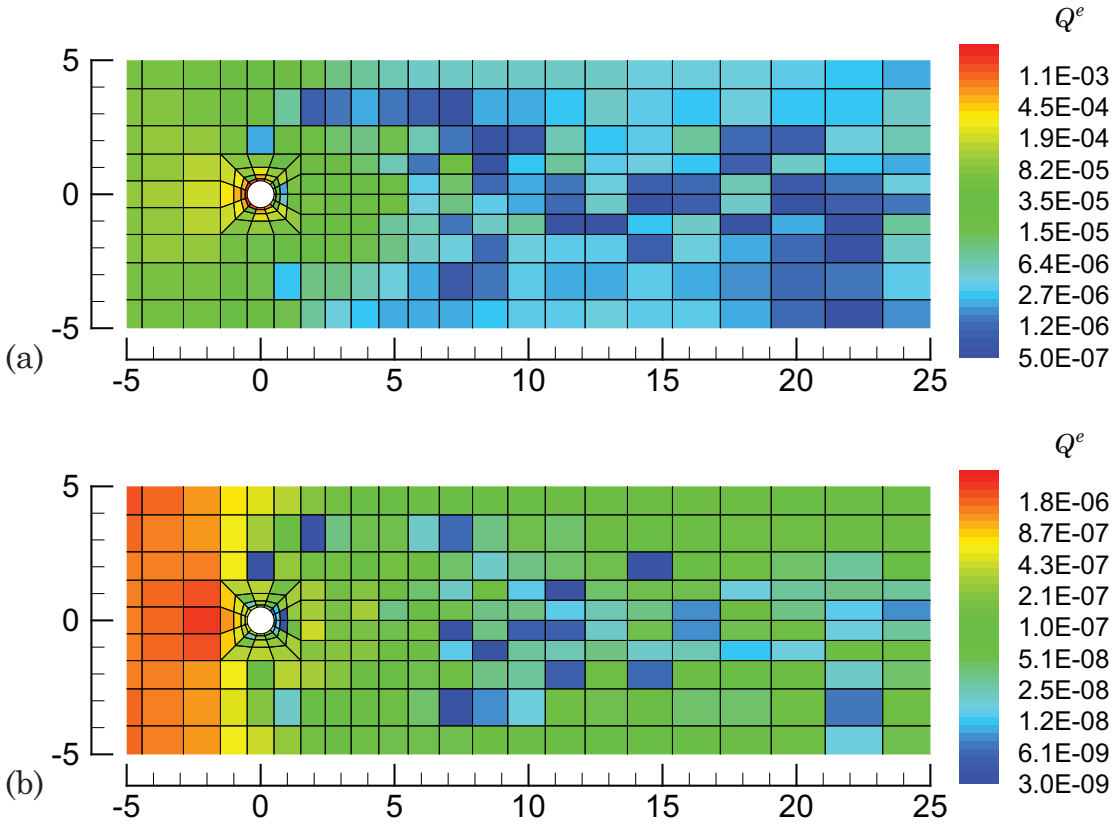


Fig. 38. Normalized volumetric flow rate imbalance Q^e for the finite elements in the vicinity of the wake region behind the circular cylinder. The results shown are a snapshot taken at $t = 260$ sec. using a p-level of 4 in the spatial discretization: (a) $\gamma = 0$ and (b) $\gamma = 100$.

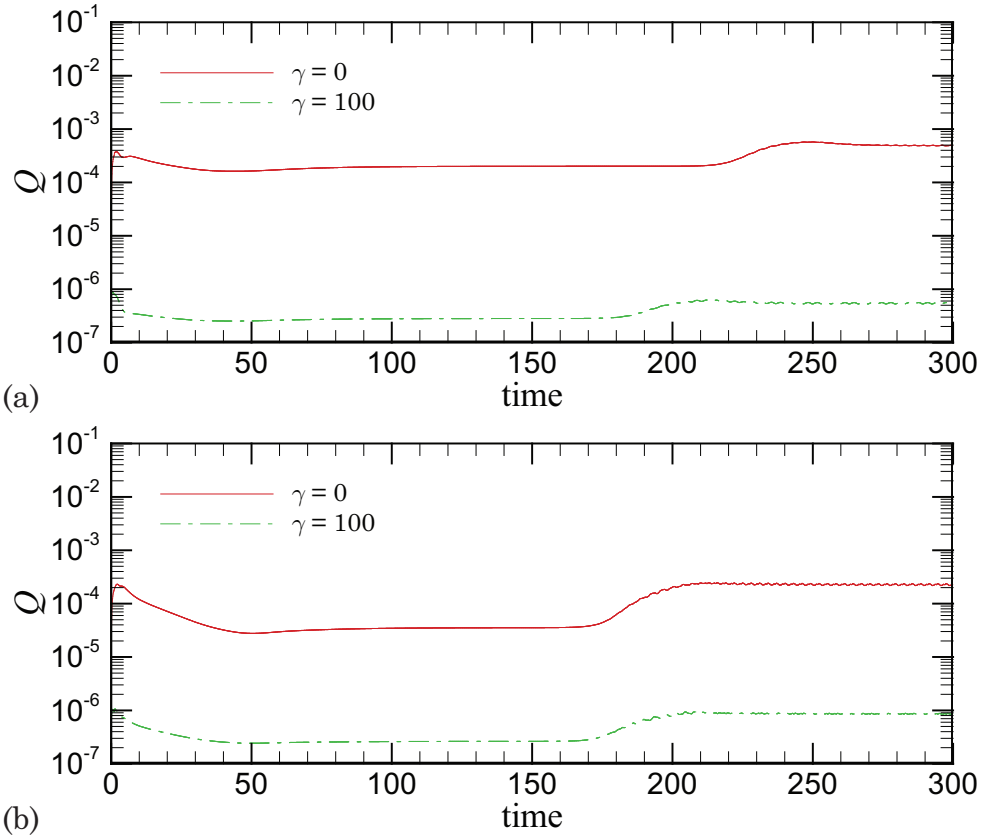


Fig. 39. Time history of the volumetric flow rate Q past the closed surface Γ^s, where $\Gamma^s = \partial\Omega^s$ is the boundary associated with the region $\Omega^s = (-1.5, 1.5)^2$. The reported results are obtained using the spectral/hp spatial discretizations shown in Figure 34 where: (a) the p-level is 2 and (b) the p-level is 4.

b. Flow past a large cylinder in a narrow channel at Re = 100

In this next example we revisit the obstructed channel flow problem introduced previously as a steady flow benchmark. To obtain a non-stationary problem, we raise the Reynolds number from 40 to 100 and vertically translate the circular cylinder 0.01 spatial units upward. The channel is again taken to be 15 units in length and 2 in height, with the center of the cylinder placed 5 units from the inlet side of the domain. The finite element mesh is nearly identical to the one employed in the stationary flow problem (see Figure 31 (a)). A close up view of the mesh (in the vicinity of the cylinder) used in the current study is shown in Figure 40.

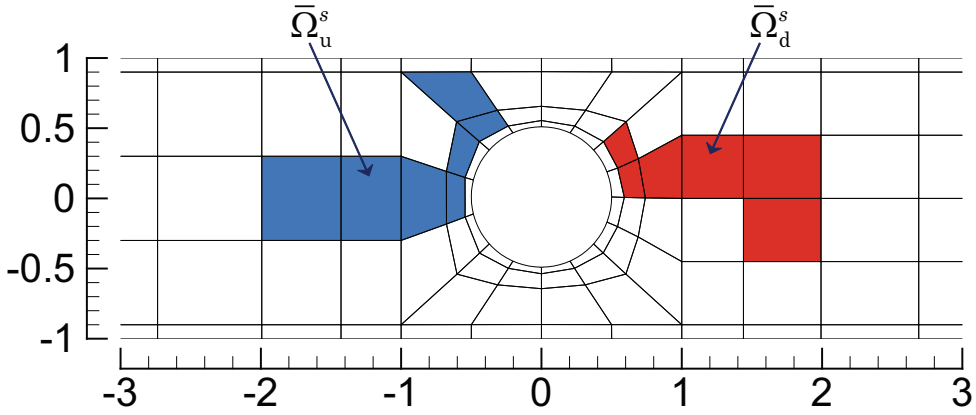


Fig. 40. Close-up view of the finite element mesh used to simulate unsteady flow through a channel with a circular obstruction. The shaded regions $\bar{\Omega}^s_u$ and $\bar{\Omega}^s_d$ are control volumes used in the post-processing stage to assess the severity of mass conservation violation for a given finite element simulation.

All flow fields are taken initially to be zero. For the inflow boundary condition, taken along the left hand side of the computational domain, we specify a time dependent parabolic horizontal velocity profile $v_x^p(y,t) = \frac{3}{2}(1-y^2)\tanh(t)$. The prescribed outflow and no slip boundary conditions are the same as those described for

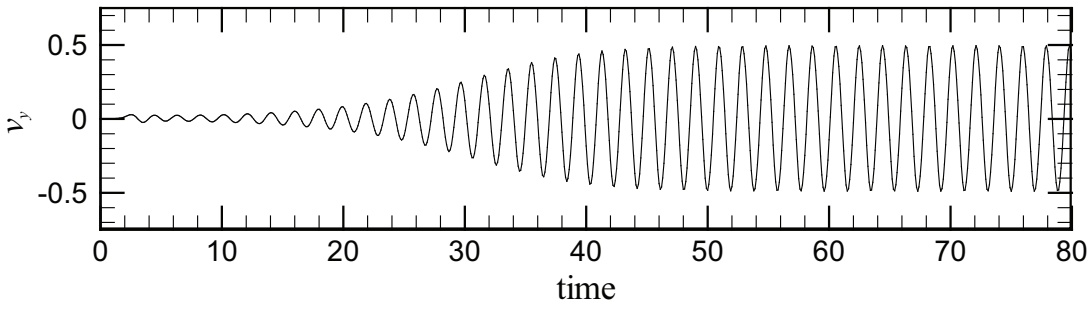


Fig. 41. Time history of the vertical velocity component v_y downstream from the circular cylinder at $(x, y) = (2, 0)$ as determined using a p-level of 6 in the spatial discretization. The time step is $\Delta t = 0.02$, $\gamma = 100$ and $\epsilon = 0.005$.

the steady-state version of the problem. As in the previous non-stationary example, we again employ the BDF2 time integrator (where the BDF1 integrator is utilized for the first 10 time steps). We solve the problem over a total time interval of 100 sec. The nonlinear convergence criteria, defined in terms of the relative Euclidean norm of the residuals in the nodal velocities between two successive iterations, is taken as $\varepsilon = 10^{-4}$ at each time step. The UMFPACK direct solver library is again utilized in the solution of the sparse global system of finite element equations. The problem is solved using the modified least-squares functional $\mathcal{J}^{\star}_{\Delta t}$ for all possible combinations of the following parameters: $\gamma = 0$ and 100, $\epsilon = 0$ and 0.005, p-level = 4 and 6 and $\Delta t = 0.05$. To verify that the numerical solutions are indeed sufficiently resolved in time, we also solved the problem using a time increment of $\Delta t = 0.02$ for the case where the p-level is 6, $\gamma = 100$ and $\epsilon = 0.005$.

The time history of the vertical velocity component v_y at the spatial point $(x, y) = (2, 0)$ is shown in Figure 41. The non-symmetric domain allows the instability in the flow to propagate quickly such that a well-defined periodic response is reached at around 50 sec. into the simulation. From Figure 41 we measure the

non-dimensional period for a typical vortex shedding cycle to be $\mathcal{T} = 1.93$ which corresponds with a non-dimensional shedding frequency of $\mathrm{St} = 0.5181$. The presence of the channel clearly results in a much shorter shedding cycle than what was observed in the previously presented *external* flow past a cylinder problem.

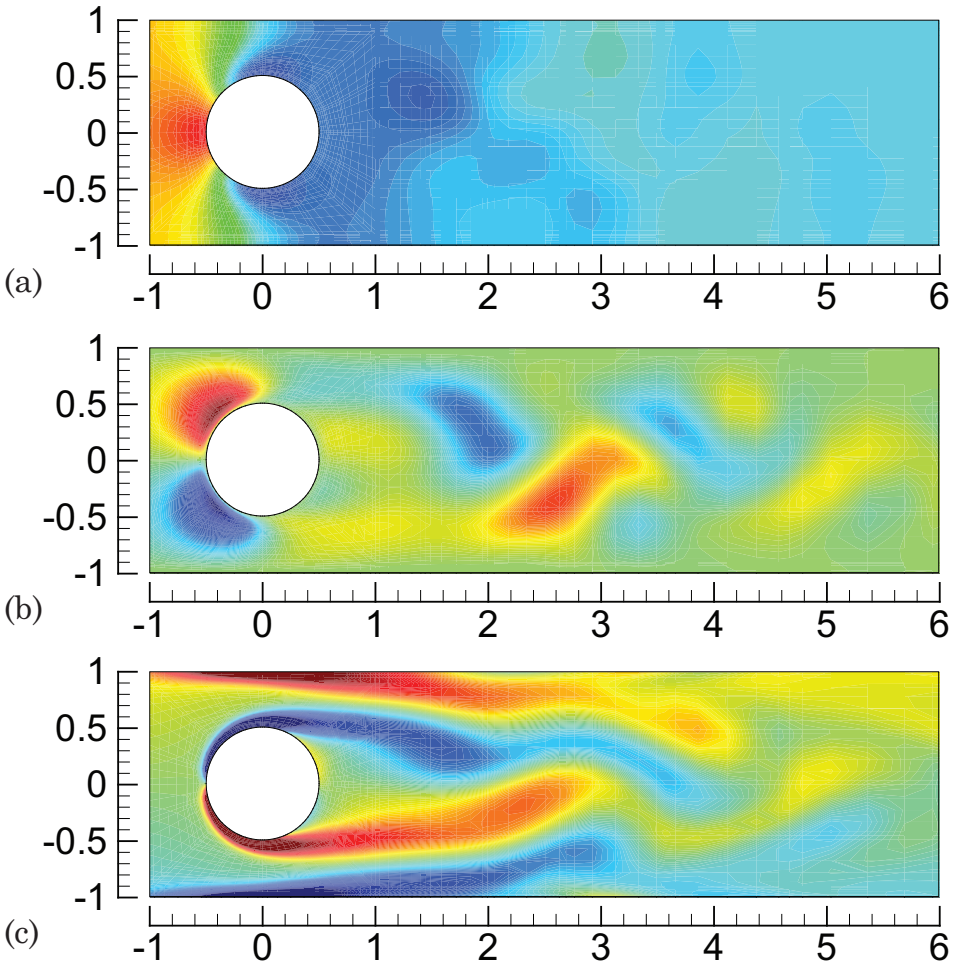


Fig. 42. Instantaneous contours for flow in a channel past a circular cylinder at $t = 75.5$ sec., where the finite element mesh associated with a p-level of 6 has been utilized: (a) pressure field p, (b) vertical velocity component v_y and (c) vorticity ω. The results shown are for the case where $\Delta t = 0.05$, $\gamma = 100$ and $\epsilon = 0.005$.

Contours of the instantaneous pressure p, velocity component v_y and vorticity

ω in the wake region are shown in Figure 42 at $t = 75.5$ sec. for a p-level of 6. The swirling of vortices that develops in the wake region clearly becomes suppressed (due to the channel walls) as the fluid travels further downstream past the cylinder. In Figure 43 we present a snapshot of the normalized volumetric flow rate imbalance Q^e for all elements in the computational discretization $\bar{\Omega}^{hp}$ at $t = 75.5$ sec., using a p-level of 4. We take $\epsilon = 0$ in the post-processing of Q^e for each element. In Figure 44 we trace the time histories of the volumetric flow rate Q (obtained using Eq. (4.16)) through the closed boundaries of the upstream and downstream control regions $\bar{\Omega}^s_{\mathrm{u}}$ and $\bar{\Omega}^s_{\mathrm{d}}$ shown in Figure 40, again using a spatial discretization with a p-level of 4. Noticeable improvement in mass conservation is observed for each element $\bar{\Omega}^e$ in $\bar{\Omega}^{hp}$ and also for the control regions for the case where $\gamma = 100$.

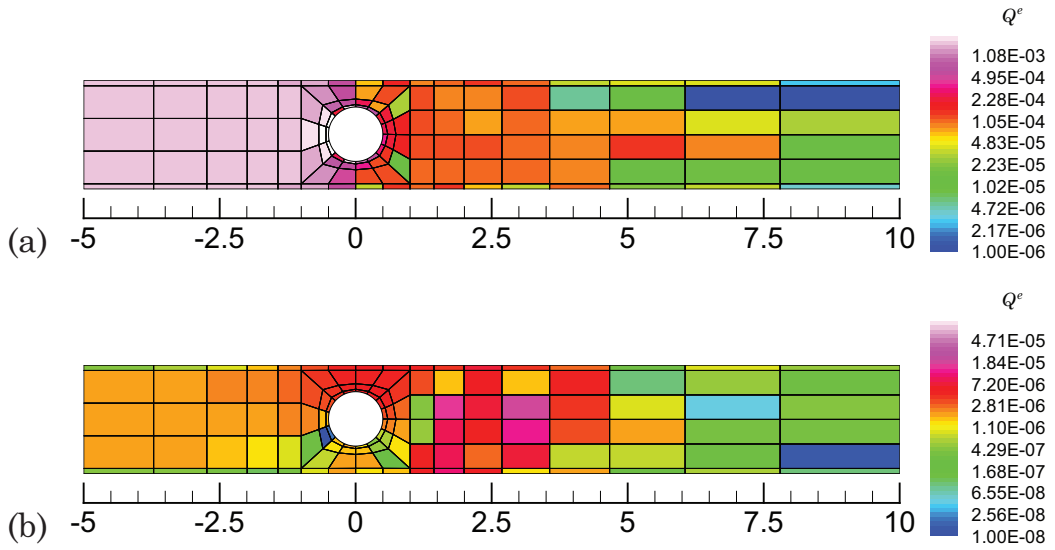


Fig. 43. Normalized volumetric flow rate imbalance Q^e for each finite element in $\bar{\Omega}^{hp}$ for flow in a channel past a circular cylinder at $t = 75.5$ sec. The results shown have been obtained using a time increment of $\Delta t = 0.05$ sec.: (a) the p-level is 4 and $\gamma = 0$ and (b) the p-level is 4, $\gamma = 100$ and $\epsilon = 0.005$.

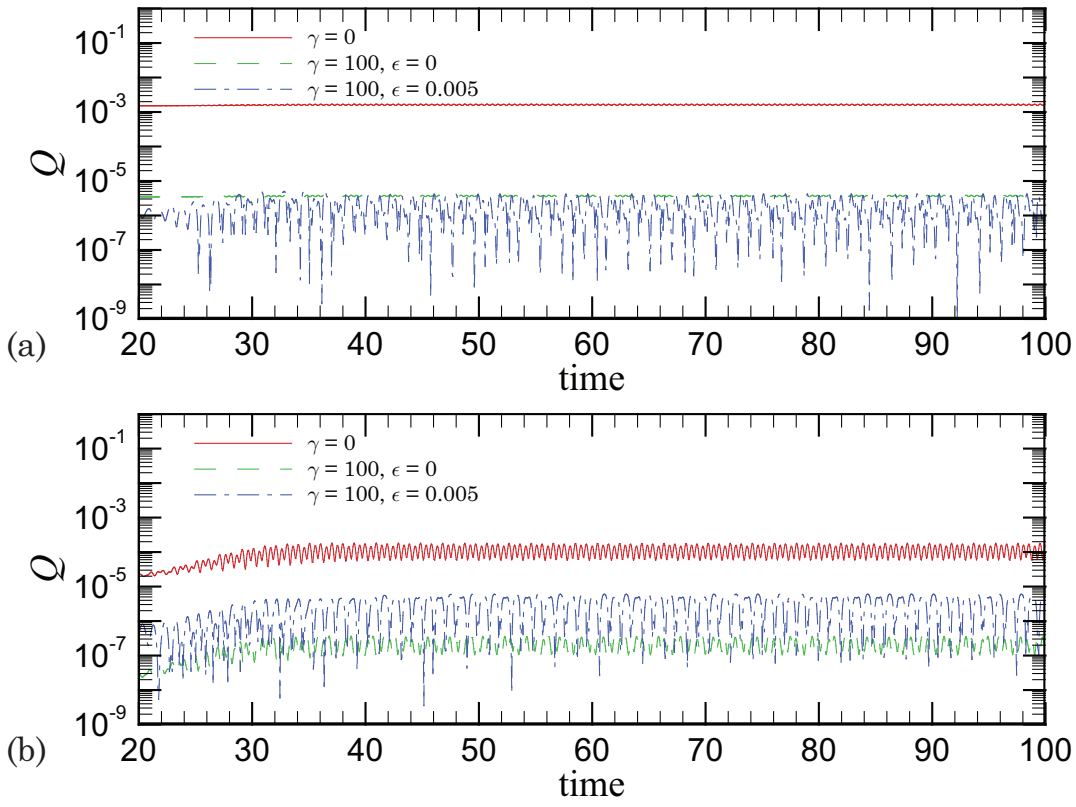


Fig. 44. Time histories of the volumetric flow rate Q past: (a) the closed surface $\Gamma^s_{\rm u}$ and (b) the closed surface $\Gamma^s_{\rm d}$. The surfaces $\Gamma^s_{\rm u} = \partial\Omega^s_{\rm u}$ and $\Gamma^s_{\rm d} = \partial\Omega^s_{\rm d}$ are the boundaries of the upstream and downstream control volumes shown in Figure 40. The results are for the case where the p-level is 4 and $\Delta t = 0.05$.

Time histories of the volumetric flow rate Q past the crown of the cylinder are plotted in Figure 45 for different values of γ and ϵ at p-levels 4 and 6. The results shown have been normalized by the long term prescribed inlet volumetric flow rate. Similar results were observed at the domain exit and also at $x = 1.0$. In the upper two plots (where the p-level is 4 and 6 respectively) we observe excellent mass conservation when $\gamma = 100$ for both values chosen for ϵ.

In an effort to demonstrate that the proposed formulation improves velocity-

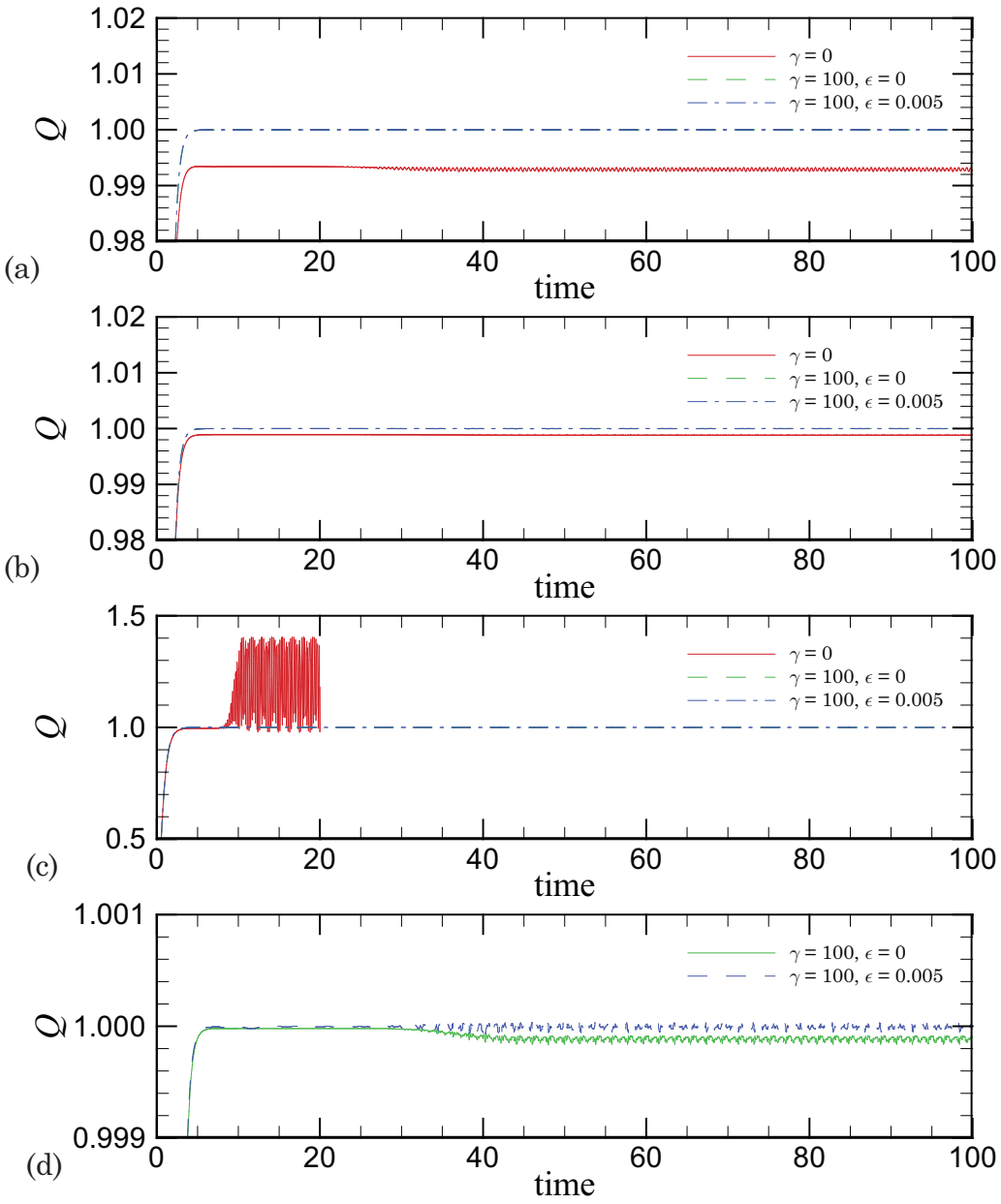


Fig. 45. Time histories of the normalized volumetric flow rate Q past the crown $(x = 0, y)$ of the large circular cylinder. All results have been obtained using a time increment of $\Delta t = 0.05$ sec.: (a) the p-level is 4 and $\alpha = (\Delta t)^2$, (b) the p-level is 6 and $\alpha = (\Delta t)^2$ (c) the p-level is 6 and $\alpha = 1.0$ and (d) the p-level is 6 and $\alpha = 1.0$.

pressure coupling and overall numerical stability in the simulation of transient flows, we show in the lower two plots of Figure 45 the normalized volumetric flow rates past the crown of the cylinder using a non-scaled version of $\mathcal{J}^{\star}_{\Delta t}$ (i.e., taking $\alpha = 1$) for a p-level of 6. For the standard least-squares formulation, obtained by setting $\gamma = 0$, we observe spurious oscillations in all fields, which eventually leads to total instability of the finite element solution procedure; this simulation was, therefore, manually terminated at $t = 20$ sec. The cases where $\gamma = 100$ yield reliable results for all fields; furthermore, excellent mass conservation is observed despite the fact that we have employed no scaling of the momentum equation residual in the definition of the least-squares functional. When $\gamma = 100$ and $\epsilon = 0.005$ we observe "exact" mass conservation up to 3 decimal places for all time. For the current example problem it is clear that the modified least-squares formulation has the ability to: (a) improve mass conservation and (b) enhance numerical stability in the simulation of non-stationary flows.

CHAPTER V

VISCOELASTIC BEAMS*

In this chapter we develop and numerically implement high-order finite element models for the quasi-static and fully transient mechanical response of initially straight viscoelastic beams subjected to loads that induce large displacements, moderate rotations and small strains. Beams are among the most commonly employed structural members and are encountered in virtually all systems of structural design. The kinematic assumptions upon which theories for beams are based are, to a large extent, independent of the actual constitutive makeup of a given beam structure. Closure of most analytic or numerical models for beams, however, is most often achieved through the additional assumption of elastic material response. The usefulness of the models resulting from this additional conjecture cannot be overstated; this is especially true in the analysis of metallic and ceramic based structural components. There are many engineering materials, however, that cannot be adequately modeled using the classical elasticity assumption. One such category, which constitutes the focus of this chapter, is the set of viscoelastic solid materials (we will restrict our attention to *linear* viscoelastic solids). Prominent examples that often fall into this category of materials include metals at elevated temperatures, polymers, rubbers and concrete. These materials are often highly favored for use in structural components, due to their natural ability to dampen out structural vibrations. Robust, efficient finite element technology for the analysis of viscoelastic beams, is therefore of particular importance.

The theoretical foundations of viscoelasticity are well established. We refer to

---

*Part of the numerical results reported in this chapter appear in the article "Nonlinear quasi-static finite element formulations for viscoelastic Euler-Bernoulli and Timoshenko beams" by G. S. Payette and J. N. Reddy, *Comm. Numer. Meth. Eng.*, vol. 26, pp. 1736–1755, 2010.

the standard texts of Flügge [81], Christensen [82], Findley [83] and Reddy [84] for an overview on the theory of viscoelastic material behavior, as well as the classical analytical solution techniques that may be used to solve simple viscoelastic boundary-value problems. For example, in Flügge [81], the Laplace transform procedure is employed to obtain exact expressions for the transverse deflection of viscoelastic beams. Another important analytical solution method, discussed by Christensen and Findley et al. [82, 83], is the correspondence principle which under certain loading conditions allows linear elasticity solutions to be converted into viscoelasticity solutions through the use of integral transformations. Analytical solutions based on the Laplace transform method or correspondence principle, however, are typically limited to very simple geometric configurations, boundary conditions and material models.

Numerical methods provide a powerful framework for obtaining approximate solutions to viscoelasticity problems. The finite element method, in particular, has been employed with great success in the analysis of viscoelastic bodies by many researchers. Of paramount importance, in the formulation of a numerical procedure for solving viscoelastic boundary-value problems, is the ability to efficiently integrate the viscoelastic constitutive equations in time. Keeping this in mind, Taylor et al. [85] employed the finite element method in conjunction with a two-point recurrence relation to solve viscoelasticity problems such that solution data from only the immediate previous time step (as opposed to the entire deformation history) is needed in determining a body's configuration at the current time step. Similarly, Oden and Armstrong [86] presented a finite element framework for thermoviscoelasticity and conducted numerical experiments involving thick-walled cylinders subjected to time-dependent boundary conditions. In their work, they extended the applicability of recurrence-based temporal integration formulas to also include nonlinear boundary-value problems. Additional general finite element formulations for viscoelastic con-

tinua can be found in Refs. [87, 88, 89, 90].

Although three-dimensional finite element formulations are applicable to continua in general, it is often computationally advantageous to specialize these models to structural elements such as beams, plates and shells. When appropriately employed, finite element formulations for structures can offer the prospect of highly accurate numerical solutions, often at a mere fraction of the computational expense needed to conduct a fully three-dimensional simulation. A variety of beam theory based finite element models have been presented in the literature for the analysis of viscoelastic structures. The majority of these formulations employ some form of either the Euler-Bernoulli or Timoshenko beam theories and are mostly restricted to small strain analysis. The formulations differ in how the convolution form of the viscoelastic constitutive equations are temporally discretized. A popular approach adopted by many researchers is to employ the Laplace transform method directly in the construction of the finite element equations [91, 92, 93]. In this approach, quantities associated with the time domain, including the convolution integral, are transformed into variables associated with the s coordinate of the Laplace space. A successful numerical simulation therefore requires an efficient and accurate inversion of the solution in s space back to the time domain. Many of the key ideas are presented in work of Aköz and Kadioglu [92], wherein a Timoshenko beam element is developed using mixed variational principles. In their work, the finite element model requires numerical inversion from the Laplace-Carson domain back to the time domain. Temel et al. [93] utilized the Durbin's inverse Laplace transform method in an analysis of cylindrical helical rods (based on the Timoshenko beam hypotheses).

Additional numerical formulations for viscoelastic beams have been constructed using the Fourier transform method [94], the anelastic displacement (ADN) procedure [95, 96], the Golla-Hughes-McTavish (GHM) method [97, 98, 99, 100] and the

trapezoidal rule [101]. It can be shown that when the relaxation moduli are given in the form of Prony series, the convolution form of the linear viscoelastic constitutive equations may be equivalently expressed as a set of ordinary differential equations (in terms of a collection of internal strain variables). Numerical discretization procedures exploiting this ODE form of the viscoelastic constitutive equations have been successfully adopted in the works of Johnson et al. [102] and Austin and Inman [103]. It is worth noting that finite element models for sandwich beams (based on the Euler-Bernoulli and Timoshenko beam hypotheses) have also been developed by Galucio et al. [104] using fractional derivative viscoelastic constitutive models.

The viscoelastic beam finite element formulations described above are restricted to a class of problems involving infinitesimal strains and small deflections. As a direct consequence, these models lack the ability to account for various geometrically nonlinear effects that can become significant whenever the externally applied loads are sufficiently large. The objective of the present chapter, therefore, is to develop a family of efficient locking-free nonlinear finite element models based on the Euler-Bernoulli (EBT), Timoshenko (TBT) and Reddy third-order (RBT) beam theories that can be readily applied to the analysis of quasi-static and fully transient viscoelastic beam structures.

The chapter is organized as follows. We first review the kinematic assumptions that form the basis for each of the three beam theories considered in the present study. An effective strain tensor (a simplification of the Green-Lagrange strain) is then introduced along with the assumed linear viscoelastic constitutive model. The finite element formulation for each beam theory is then derived from the principle of virtual displacements, or equivalently through the use of the weak-form Galerkin procedure. In the fully discretized finite element models, the convolution integrals (emanating from the viscoelastic constitutive equations) are temporally approximated using the

trapezoidal rule in conjunction with a two-point recurrence formula. We conclude the chapter by presenting numerical results for quasi-static and fully transient verification benchmark problems. We shown that all forms of locking may be avoided through the use of either: (a) low-order finite elements with selective employment of full and reduced numerical integration strategies or (b) fully integrated finite elements constructed from high-order polynomial interpolation functions of both Lagrange and Hermite type.

A. Kinematics of deformation

There are a variety of beam theories that have been successfully employed in the mechanical analysis of structural elements [105]. Such theories are typically formulated in terms of truncated Taylor series expansions of the components of the displacement field; where the expansions are taken with respect to the thickness coordinate. Before presenting the resulting simplified displacement fields for the beam theories considered in this work, we first introduce some notation that is somewhat unique to the current chapter. We let $\mathcal{B} \subset \mathbb{R}^3$, an open and bounded set, denote the material or reference configuration occupied by the beam at $t = 0$. The material configuration may be expressed as $\mathcal{B} = \Omega \times A$, where $\Omega = (0, L)$ and L is the initial length of the beam. In addition the quantity A represents the undeformed cross-sectional area of the beam. A typical material point belonging to $\mathcal{B}$ is denoted as $\mathbf{X} = (X, Y, Z)$. Likewise the spatial or current configuration of the beam at time t is denoted by $\mathcal{B}_t$ and an associated point is given as $\mathbf{x} = (x, y, z)$. The motion of the beam is a one parameter family of configurations (where the time t is the parameter) that may be expressed in terms of the standard bijective mapping $\boldsymbol{\chi} : \mathcal{B} \times \mathbb{R} \to \mathcal{B}_t$. As a result, the location of point $\mathbf{X}$ at time t is given as $\mathbf{x} = \boldsymbol{\chi}(\mathbf{X}, t)$. The displacement may be

expressed in the usual manner as $\mathbf{u}(\mathbf{X},t) = \boldsymbol{\chi}(\mathbf{X},t) - \mathbf{X}$.

1. The displacement fields for the EBT, TBT and RBT

The most simple and commonly used beam theory is the Euler-Bernoulli beam theory (EBT), which is based on the displacement field

$$u(X,Z,t) = u_0(X,t) - Z\frac{\partial w_0}{\partial X} \tag{5.1a}$$

$$w(X,Z,t) = w_0(X,t) \tag{5.1b}$$

where the X coordinate is taken along the beam length, the Z coordinate along the thickness direction of the beam, u_0 is the axial displacement of a point on the mid-plane $(X,0,0)$ of the beam and w_0 represents the transverse deflection of the mid-plane. The Euler-Bernoulli displacement field implies that straight lines orthogonal to the mid-surface before deformation remain so after deformation. The major deficiency associated with the EBT is failure to account for deformations associated with shearing.

A slightly more complicated theory is the Timoshenko beam theory (TBT) and is based on the displacement field

$$u(X,Z,t) = u_0(X,t) + Z\phi_x(X,t) \tag{5.2a}$$

$$w(X,Z,t) = w_0(X,t) \tag{5.2b}$$

When the deformation is small the parameter $\phi_x(X,t)$ may be interpreted as the rotation of the transverse normal about the Y axis. The Timoshenko beam theory relaxes the normality assumption of the Euler Bernoulli theory and admits a constant state of shear strain across a given cross section. Since the actual shear strain for a beam is at least quadratic, the TBT necessitates the use of shear correction coefficients

in order to accurately predict the transverse displacements of thick beams.

The final beam theory considered in this chapter is the third-order Reddy beam theory (RBT). In the RBT, the displacement field (for a beam with a rectangular cross section) takes the following form

$$u(X,Y,Z,t) = u_0(X,t) + Z\phi_x(X,t) - Z^3 c_1\Big(\phi_x(X,t) + \frac{\partial w_0}{\partial X}\Big) \tag{5.3a}$$

$$w(X,Y,Z,t) = w_0(X,t) \tag{5.3b}$$

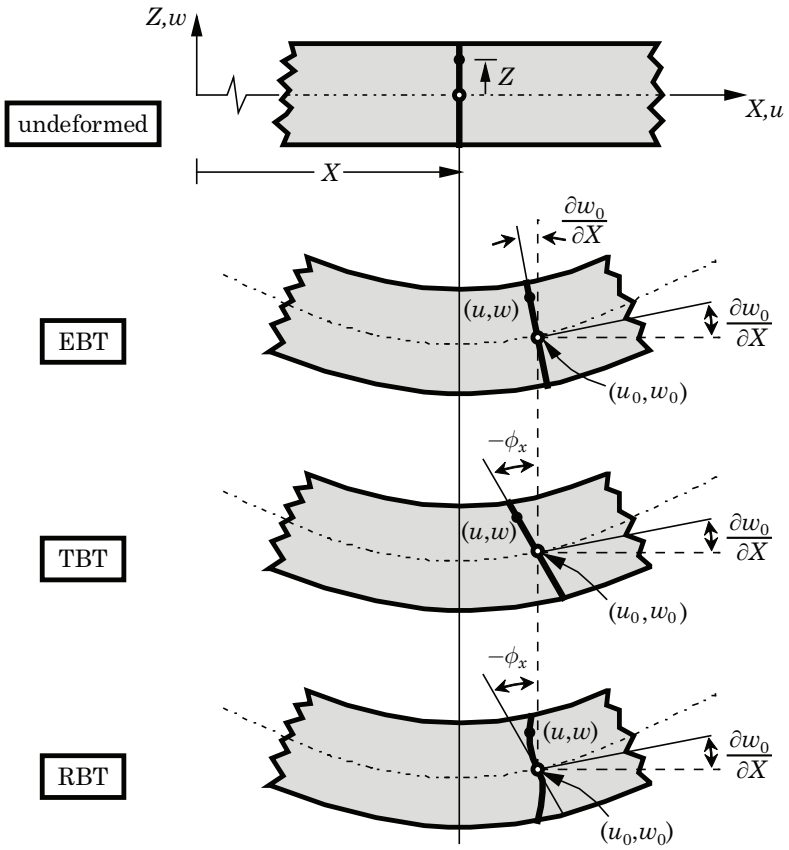


Fig. 46. Deformation of a beam structure according to the Euler-Bernoulli, Timoshenko and third-order Reddy beam theories (adapted from Reddy [106]).

In the above expression $c_1 = 4/(3h^2)$, where h is the height of the beam and b is the

beam width. The displacement field of the Reddy beam theory suggests that a straight line perpendicular to the undeformed mid-plane becomes a cubic curve following deformation. As a result, the Reddy beam theory provides a more realistic prediction (as compared with the EBT and TBT) of the shear strain along the cross-section of a beam and as a result circumvents the need for shear correction factors. Figure 46 shows the kinematics of deformation of a transverse normal for a beam structure as predicted by each beam theory. It is important to note that the displacement field of Reddy's third-order beam theory contains the other two lower-order beam theories as special cases. The TBT is recovered by setting $c_1 = 0$ and the EBT is obtained by replacing ϕ_x with $-\partial w_0/\partial X$. Since the lower-order theories are in this sense *contained* within the RBT, we will restrict the scope of our remaining discussion to developing a finite element model for viscoelastic beams based on the third-order Reddy beam theory only. Numerical results, however, will be presented using all three beam theories. For details specific to the Euler-Bernoulli and Timoshenko beam theories, we refer to the article by Payette and Reddy [21].

2. The effective strain tensor for the simplified theory

In the mechanical analysis of deformable solids, it is imperative to employ stress and strain measures that are consistent with the deformations realized (see [50, 107]). When the deformations of the body are large, there are a variety of strain measures that may be employed. In our formulation we employ a total Lagrangian description of the deformation (more precisely, we employ a *simplified* description that includes only some of the nonlinearities present in the Lagrangian formulation). In the Lagrangian description, the Green-Lagrange strain tensor $\mathbf{E}$ constitutes an appropriate measure of the strain at a point in the body. For the present analysis the non-zero Cartesian

components of $\mathbf{E}$ may be expressed as

$$E_{XX} = \frac{\partial u}{\partial X} + \frac{1}{2}\left[\left(\underline{\frac{\partial u}{\partial X}}\right)^2 + \left(\frac{\partial w}{\partial X}\right)^2\right] \tag{5.4a}$$

$$E_{XZ} = \frac{1}{2}\left(\frac{\partial u}{\partial Z} + \frac{\partial w}{\partial X} + \underline{\frac{\partial u}{\partial X}\frac{\partial u}{\partial Z}}\right) \tag{5.4b}$$

$$E_{ZZ} = \frac{1}{2}\left(\underline{\frac{\partial u}{\partial Z}}\right)^2 \tag{5.4c}$$

In the present formulation we wish to develop a finite element framework that is applicable under loading conditions that produce large transverse displacements, moderate rotations (10-15°) and small strains [106]. Under such conditions it is possible to neglect the underlined terms in the above definition of the Green-Lagrange strain tensor. Consequently, we employ a reduced form of the Green-Lagrange strain tensor, denoted by ε, whose non-zero components may be expressed as

$$\varepsilon_{xx} = \frac{\partial u_0}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0}{\partial x}\right)^2 + z\frac{\partial \phi_x}{\partial x} - z^3 c_1\left(\frac{\partial \phi_x}{\partial x} + \frac{\partial^2 w_0}{\partial x^2}\right) \tag{5.5a}$$

$$\gamma_{xz} = 2\varepsilon_{xz} = \left(1 - c_2 z^2\right)\left(\phi_x + \frac{\partial w_0}{\partial x}\right) \tag{5.5b}$$

where $c_2 = 3c_1$. The strain components associated with the linearized strain tensor ε are commonly called the von Kármán strain components. This *simplified* strain tensor will be used in both the viscoelastic constitutive equations and the virtual work statement. For a comparison of numerical results obtained using the above simplified theory with the full nonlinear theory for elastic structures, we refer to the work of Başar et al. [108]. It is important to note that the material coordinates appearing in the definition of the reduced strain components and throughout the remainder of this chapter are denoted as (x, y, z) as a reminder that the present formulation is applicable to small strains and moderate rotations, and is therefore a linearization of the more general finite deformation theory.

3. Linear viscoelastic constitutive equations

For linear viscoelastic materials, the constitutive equations relating the components of the second Piola-Kirchhoff stress tensor $\mathbf{S}$ to the Green-Lagrange strain $\mathbf{E}$ may be expressed in terms of the following set of integral equations

$$\mathbf{S}(t) = \mathbb{C}(0) : \mathbf{E}(t) + \int_0^t \dot{\mathbb{C}}(t-s) : \mathbf{E}(s) ds \tag{5.6}$$

where $\dot{\mathbb{C}}(t-s) \equiv d\mathbb{C}(t-s)/d(t-s)$ and $\mathbb{C}(t)$ is the fourth-order viscoelasticity relaxation tensor. Note that throughout this chapter, a dot appearing above a given variable always denotes differentiation with respect to the inclosed arguments (e.g., $\dot{f}(t) = df(t)/dt$ and $\dot{f}(t-s) = df(t-s)/d(t-s)$). Replacing $\mathbf{E}$ with ε yields

$$\sigma_{xx}(\mathbf{x}, t) = E(0)\varepsilon_{xx}(\mathbf{x}, t) + \int_0^t \dot{E}(t-s)\varepsilon_{xx}(\mathbf{x}, s) ds \tag{5.7a}$$

$$\sigma_{xz}(\mathbf{x}, t) = G(0)\gamma_{xz}(\mathbf{x}, t) + \int_0^t \dot{G}(t-s)\gamma_{xz}(\mathbf{x}, s) ds \tag{5.7b}$$

where σ_{xx} and σ_{xz} are the nonzero components of second Piola-Kirchhoff stress tensor used in the present *simplified* formulation. The quantities $E(t)$ and $G(t)$ are the relaxation moduli. The specific forms of $E(t)$ and $G(t)$ will depend upon the material model employed. For the present analysis we assume that these relaxation functions can be expanded as Prony series of order NPS as

$$E(t) = E_0 + \sum_{l=1}^{\text{NPS}} \bar{E}_l(t), \qquad G(t) = G_0 + \sum_{l=1}^{\text{NPS}} \bar{G}_l(t) \tag{5.8}$$

where $\bar{E}_l(t)$ and $\bar{G}_l(t)$ have been defined as (following the generalized Maxwell model)

$$\bar{E}_l(t) = E_l e^{-t/\tau_l^E}, \qquad \bar{G}_l(t) = G_l e^{-t/\tau_l^G} \tag{5.9}$$

The Prony series representation of the viscoelastic relaxation moduli will prove critical in the implementation of efficient temporal numerical integration algorithms of the

viscoelastic constitutive equations.

B. The weak-form Galerkin finite element model

1. The Galerkin based weak formulation

The weak-form Galerkin finite element model of the third-order Reddy beam theory may be developed by applying the principle of virtual work to a typical beam as viewed in the reference configuration. The dynamic form of the virtual work statement may therefore be expressed as

$$\begin{aligned}\mathcal{G}(\delta\mathbf{u},\mathbf{u}) &= -\delta\mathcal{K}(\delta\mathbf{u},\mathbf{u}) + \delta\mathcal{W}_{\mathrm{I}}(\delta\mathbf{u},\mathbf{u}) + \delta\mathcal{W}_{\mathrm{E}}(\delta\mathbf{u},\mathbf{u}) \\ &= \int_{\mathcal{B}} \big(\delta\mathbf{u}\cdot\rho_0\ddot{\mathbf{u}} + \delta\mathbf{E}:\mathbf{S} - \delta\mathbf{u}\cdot\rho_0\mathbf{b}\big)dV - \int_{\Gamma_\sigma}\delta\mathbf{u}\cdot\mathbf{t}_0 dS \\ &\cong \int_0^L\int_A \big(\delta\mathbf{u}\cdot\rho_0\ddot{\mathbf{u}} + \delta\varepsilon:\sigma - \delta\mathbf{u}\cdot\rho_0\mathbf{b}\big)dAdx - \int_{\Gamma_\sigma}\delta\mathbf{u}\cdot\mathbf{t}_0 dS \equiv 0\end{aligned} \tag{5.10}$$

where $\delta\mathcal{K}$ is the virtual kinetic energy, $\delta\mathcal{W}_{\mathrm{I}}$ is the internal virtual work and $\delta\mathcal{W}_{\mathrm{E}}$ is the external virtual work. The additional quantities ρ_0, $\mathbf{b}$ and $\mathbf{t}_0$ are the density, body force and traction vector, respectively. The above expression constitutes the weak form of the classical Euler-Lagrange equations of motion of a continuous body.

Since the generalized displacements (u_0, w_0, ϕ_x) depend only on x and t, it is possible to pre-integrate the virtual work statement over A. As a result, the computational domain for the problem reduces to the material line $\bar{\Omega} = [0, L]$ taken along $(x, y = 0, z = 0)$. The finite element discretization is therefore obtained by partitioning this material line into a set of NE finite elements, as described in Chapter II, where the domain of the eth element may be expressed as $\bar{\Omega}^e = [x_a^e, x_b^e]$. The resulting variational problem associated with the weak formulation of the Reddy beam equations may therefore be expressed as follows: find $(u_0, w_0, \phi_x) \in \mathcal{V} = \mathcal{Q}\times\mathcal{X}\times\mathcal{Y}$ such that for all $(\delta u_0, \delta w_0, \delta\phi_x) \in \mathcal{W} = \tilde{\mathcal{Q}}\times\tilde{\mathcal{X}}\times\tilde{\mathcal{Y}}$ the following expressions hold

within each element:

$$0 = \int_{x_a}^{x_b} \left(I_0 \delta u_0 \ddot{u}_0 + \frac{\partial \delta u_0}{\partial x} N_{xx} - \delta u_0 f \right) dx - \delta u_0(x_a) Q_1 - \delta u_0(x_b) Q_5 \tag{5.11a}$$

$$\begin{aligned} 0 = \int_{x_a}^{x_b} \Bigg[& I_0 \delta w_0 \ddot{w}_0 + \frac{\partial \delta w_0}{\partial x} \left(c_1^2 I_6 \frac{\partial \ddot{w}_0}{\partial x} - J_4 \ddot{\phi}_x \right) + \frac{\partial \delta w_0}{\partial x} \Bigg(\frac{\partial w_0}{\partial x} N_{xx} + Q_x \\ & - c_2 R_x \Bigg) - \frac{\partial^2 \delta w_0}{\partial x^2} c_1 P_{xx} - \delta w_0 q \Bigg] dx - Q_2 \delta w_0(x_a) - Q_6 \delta w_0(x_b) \\ & - Q_3 \left(-\frac{\partial \delta w_0}{\partial x} \right) \Bigg|_{x=x_a} - Q_7 \left(-\frac{\partial \delta w_0}{\partial x} \right) \Bigg|_{x=x_b} \end{aligned} \tag{5.11b}$$

$$\begin{aligned} 0 = \int_{x_a}^{x_b} \Bigg[& \delta \phi_x \left(-J_4 \frac{\partial \ddot{w}_0}{\partial x} + K_2 \ddot{\phi}_x \right) + \delta \phi_x (Q_x - c_2 R_x) \\ & + \frac{\partial \delta \phi_x}{\partial x} (M_{xx} - c_1 P_{xx}) \Bigg] dx - Q_4 \delta \phi_x(x_a) - Q_8 \delta \phi_x(x_b) \end{aligned} \tag{5.11c}$$

The function spaces comprising the product spaces $\mathcal{V}$ and $\mathcal{W}$ are defined as

$$\mathcal{Q} := \left\{ u_0 : \quad u_0 \in H^1(\Omega) \times C^2(\mathcal{I}), \quad u_0 = u_0^{\mathrm{p}} \text{ on } \Gamma_u \right\} \tag{5.12a}$$

$$\mathcal{X} := \left\{ w_0 : \quad w_0 \in H^2(\Omega) \times C^2(\mathcal{I}), \quad w_0 = w_0^{\mathrm{p}} \text{ on } \Gamma_w, \quad -\partial_x w_0 = \varphi_0^{\mathrm{p}} \text{ on } \Gamma_\varphi \right\} \tag{5.12b}$$

$$\mathcal{Y} := \left\{ \phi_x : \quad \phi_x \in H^1(\Omega) \times C^2(\mathcal{I}), \quad \phi_x = \phi_x^{\mathrm{p}} \text{ on } \Gamma_\phi \right\} \tag{5.12c}$$

$$\tilde{\mathcal{Q}} := \left\{ \delta u_0 : \ \delta u_0 \in H^1(\Omega) \times C(\mathcal{I}), \quad \delta u_0 = 0 \quad \text{on } \Gamma_u \right\} \tag{5.12d}$$

$$\tilde{\mathcal{X}} := \left\{ \delta w_0 : \delta w_0 \in H^2(\Omega) \times C(\mathcal{I}), \ \delta w_0 = 0 \quad \text{on } \Gamma_w, -\partial_x \delta w_0 = 0 \quad \text{on } \Gamma_\varphi \right\} \tag{5.12e}$$

$$\tilde{\mathcal{Y}} := \left\{ \delta \phi_x : \delta \phi_x \in H^1(\Omega) \times C(\mathcal{I}), \ \delta \phi_x = 0 \quad \text{on } \Gamma_\phi \right\} \tag{5.12f}$$

where $H^m(\Omega)$ is the Sobolev space of order m, $\mathcal{I} = [0, \tau]$ is the time interval (where $\tau > 0$) and $\partial_x(\,) \equiv \partial(\,)/\partial x$. The quantities Γ_u, Γ_w, Γ_φ and Γ_ϕ each represent a set of points along $\bar{\Omega}$ where u_0, w_0, $-\partial_x w_0$ and ϕ_x are specified respectively. For the sake of brevity we have omitted the superscript e from quantities appearing in Eq. (5.11) and throughout the remainder of this work (e.g., x_a and x_b). The quantities f and q appearing above are the distributed axial and transverse loads respectively. We have

also introduced the following constants

$$I_i = \rho_0 D_i = \rho_0 \int_A z^i dA, \qquad J_4 = c_1(I_4 - c_1 I_6), \qquad K_2 = I_2 - 2c_1 I_4 + c_1^2 I_6 \quad (5.13)$$

The internal stress resultants N_{xx}, M_{xx}, P_{xx}, Q_x and R_x are defined as

$$\begin{Bmatrix} N_{xx} \\ M_{xx} \\ P_{xx} \end{Bmatrix} = \int_A \begin{Bmatrix} 1 \\ z \\ z^3 \end{Bmatrix} \sigma_{xx} dA, \qquad \begin{Bmatrix} Q_x \\ R_x \end{Bmatrix} = \int_A \begin{Bmatrix} 1 \\ z^2 \end{Bmatrix} \sigma_{xz} dA \quad (5.14)$$

and can be expressed in terms of the generalized displacements (u_0, w_0, ϕ_x) through the use of the viscoelastic constitutive equations. The quantities N_{xx}, M_{xx} and Q_x are the internal axial force, bending moment and shear force. In addition, P_{xx} and R_x are higher order stress resultants that arise in the third-order beam theory due to the cubic expansion of the axial displacement field. The quantities Q_j (where j=1,. . . ,8) are the externally applied generalized nodal forces.

2. The semi-discrete finite element equations

In this section we develop the semi-discrete finite element equations associated with the third-order Reddy beam theory. Within a typical finite element the generalized displacements (u_0, w_0, ϕ_x) may be adequately approximated using the following interpolation formulas

$$u_0(x,t) \cong \sum_{j=1}^{n} \Delta_j^{(1)}(t)\psi_j^{(1)}(x) \quad (5.15a)$$

$$w_0(x,t) \cong \sum_{j=1}^{2n} \Delta_j^{(2)}(t)\psi_j^{(2)}(x) \quad (5.15b)$$

$$\phi_x(x,t) \cong \sum_{j=1}^{n} \Delta_j^{(3)}(t)\psi_j^{(1)}(x) \quad (5.15c)$$

where a space-time decoupled formulation has been adopted and n represents the number of nodes per element. Since the weak formulation requires w_0, $\partial w_0/\partial x$ and $\partial^2 w_0/\partial x^2$ to all belong to $L_2(\Omega) \times C^2(\mathcal{I})$, the discrete setting naturally dictates that $\psi_j^{(2)}$ be at the very minimum $C^1(\bar{\Omega})$ functions. As a result, $\psi_j^{(1)}$ are standard $(n-1)$th-order Lagrange interpolation functions (see Chapter II), while $\psi_j^{(2)}$ are $(2n-1)$th-order Hermite interpolation functions. Inserting the above approximations into Eq. (5.11) results in the semi-discrete finite element equations for the RBT which may be expressed at the current time t as

$$[M^e]\{\ddot{\Delta}^e\} + [K^e]\{\Delta^e\} + \int_0^t \{\Lambda^e(t,s)\}ds = \{F^e\} \tag{5.16}$$

The element-level equations may be partitioned into the following equivalent set of expressions

$$[M^{\alpha\beta}]\{\ddot{\Delta}^{(\beta)}\} + [K^{\alpha\beta}]\{\Delta^{(\beta)}\} + \int_0^t \{\Lambda^{(\alpha)}(t,s)\}ds = \{F^{(\alpha)}\} \tag{5.17}$$

where α and β range from 1 to 3 and Einstein's summation convention is implied over β. The components of the partitioned coefficient matrices and vectors may be expressed as

$$M_{ij}^{11} = \int_{x_a}^{x_b} I_0 \psi_i^{(1)} \psi_j^{(1)} dx \tag{5.18a}$$

$$M_{ij}^{12} = M_{ji}^{21} = 0 \tag{5.18b}$$

$$M_{ij}^{13} = M_{ji}^{31} = 0 \tag{5.18c}$$

$$M_{ij}^{22} = \int_{x_a}^{x_b} \left(I_0 \psi_i^{(2)} \psi_j^{(2)} + c_1^2 I_6 \frac{d\psi_i^{(2)}}{dx} \frac{d\psi_j^{(2)}}{dx} \right) dx \tag{5.18d}$$

$$M_{ij}^{23} = M_{ji}^{32} = -\int_{x_a}^{x_b} J_4 \frac{d\psi_i^{(2)}}{dx} \psi_j^{(1)} dx \tag{5.18e}$$

$$M_{ij}^{33} = \int_{x_a}^{x_b} K_2 \psi_i^{(1)} \psi_j^{(1)} dx \tag{5.18f}$$

$$K_{ij}^{11} = \int_{x_a}^{x_b} E(0) D_0 \frac{d\psi_i^{(1)}}{dx} \frac{d\psi_j^{(1)}}{dx} dx \tag{5.19a}$$

$$K_{ij}^{12} = \frac{1}{2} K_{ji}^{21} = \frac{1}{2} \int_{x_a}^{x_b} \left(E(0) D_0 \frac{\partial w_0(x,t)}{\partial x} \right) \frac{d\psi_i^{(1)}}{dx} \frac{d\psi_j^{(2)}}{dx} dx \tag{5.19b}$$

$$K_{ij}^{13} = K_{ji}^{31} = 0 \tag{5.19c}$$

$$K_{ij}^{22} = \int_{x_a}^{x_b} \left[\frac{1}{2} E(0) D_0 \left(\frac{\partial w_0(x,t)}{\partial x} \right)^2 \frac{d\psi_i^{(2)}}{dx} \frac{d\psi_j^{(2)}}{dx} + G(0) \hat{A}_s \frac{d\psi_i^{(2)}}{dx} \frac{d\psi_j^{(2)}}{dx} \right. \tag{5.19d}$$
$$\left. + E(0) c_1^2 D_6 \frac{d^2\psi_i^{(2)}}{dx^2} \frac{d^2\psi_j^{(2)}}{dx^2} \right] dx$$

$$K_{ij}^{23} = K_{ji}^{32} = \int_{x_a}^{x_b} \left(G(0) \hat{A}_s \frac{d\psi_i^{(2)}}{dx} \psi_j^{(1)} - E(0) L_4 \frac{d^2\psi_i^{(2)}}{dx^2} \frac{d\psi_j^{(1)}}{dx} \right) dx \tag{5.19e}$$

$$K_{ij}^{33} = \int_{x_a}^{x_b} \left(E(0) M_2 \frac{d\psi_i^{(1)}}{dx} \frac{d\psi_j^{(1)}}{dx} + G(0) \hat{A}_s \psi_i^{(1)} \psi_j^{(1)} \right) dx \tag{5.19f}$$

$$\Lambda_i^1(t,s) = \int_{x_a}^{x_b} \dot{E}(t-s) D_0 \frac{d\psi_i^{(1)}}{dx} \left[\frac{\partial u_0(x,s)}{\partial x} + \frac{1}{2} \left(\frac{\partial w_0(x,s)}{\partial x} \right)^2 \right] dx \tag{5.20a}$$

$$\Lambda_i^2(t,s) = \int_{x_a}^{x_b} \left\{ \dot{E}(t-s) D_0 \frac{\partial w_0(x,t)}{\partial x} \frac{d\psi_i^{(2)}}{dx} \left[\frac{\partial u_0(x,s)}{\partial x} + \frac{1}{2} \left(\frac{\partial w_0(x,s)}{\partial x} \right)^2 \right] \right. \tag{5.20b}$$
$$+ \dot{E}(t-s) \frac{d^2\psi_i^{(2)}}{dx^2} \left(c_1^2 D_6 \frac{\partial^2 w_0(x,s)}{\partial x^2} - L_4 \frac{\partial \phi_x(x,s)}{\partial x} \right)$$
$$\left. + \dot{G}(t-s) \hat{A}_s \frac{d\psi_i^{(2)}}{dx} \left(\frac{\partial w_0(x,s)}{\partial x} + \phi_x(x,s) \right) \right\} dx$$

$$\Lambda_i^3(t,s) = \int_{x_a}^{x_b} \left[\dot{E}(t-s) \frac{d\psi_i^{(1)}}{dx} \left(M_2 \frac{\partial \phi_x(x,s)}{\partial x} - L_4 \frac{\partial^2 w_0(x,s)}{\partial x^2} \right) \right. \tag{5.20c}$$
$$\left. + \dot{G}(t-s) \hat{A}_s \psi_i^{(1)} \left(\frac{\partial w_0(x,s)}{\partial x} + \phi_x(x,s) \right) \right] dx$$

$$F_i^1 = \int_{x_a}^{x_b} \psi_i^{(1)} f dx + \psi_i^{(1)}(x_a) Q_1 + \psi_i^{(1)}(x_b) Q_5 \tag{5.21a}$$

$$F_i^2 = \int_{x_a}^{x_b} \psi_i^{(2)} q dx + Q_2 \psi_i^{(2)}(x_a) + Q_6 \psi_i^{(2)}(x_b) + Q_3 \left(-\frac{d\psi_i^{(2)}}{dx} \right) \Big|_{x=x_a} + Q_7 \left(-\frac{d\psi_i^{(2)}}{dx} \right) \Big|_{x=x_b} \tag{5.21b}$$

$$F_i^3 = Q_4 \psi_i^{(1)}(x_a) + Q_8 \psi_i^{(1)}(x_b) \tag{5.21c}$$

In the above equations we have made extensive use of the following constants

$$\hat{A}_s = D_0 - 2D_2 c_2 + D_4 c_2^2 \tag{5.22a}$$

$$L_4 = c_1(D_4 - D_6 c_1) \tag{5.22b}$$

$$M_2 = D_2 - 2D_4 c_1 + D_6 c_1^2 \tag{5.22c}$$

3. The fully-discrete finite element equations

In this section we develop the fully discretized finite element equations for the Reddy beam theory. We begin by partitioning the time interval $\mathcal{I} = [0, \tau]$ into a set of N non-overlapping subintervals such that $\mathcal{I} = \bigcup_{k=1}^{N} \mathcal{I}_k$, where $\mathcal{I}_k = [t_k, t_{k+1}]$ and $t_k < t_{k+1}$ for all $\mathcal{I}_k \subset \mathcal{I}$. The solution may then be obtained incrementally by solving an initial value problem within each subinterval $\mathcal{I}_k$, where we assume that the solution is known at $t = t_k$. Within each subregion it is therefore necessary to introduce approximations for both the temporal derivatives of the generalized displacements (resulting from the inertia terms) as well as the convolution integrals (resulting from the viscoelastic constitutive model of the material). The temporal derivatives of the generalized displacements may be adequately approximated through the use of the Newmark scheme [109] or one of its variants [110]. Since temporal integration of the inertia terms is relatively straightforward, we restrict the current discussion to discretization of the quasi-static form of the semi-discrete finite element equations only.

In the present work, we approximate the convolution integrals present in the semi-discrete form of the finite element equations using the trapezoidal rule within each time subinterval. It is important to note, however, that a *naive* application of the trapezoidal rule (or any other approximation scheme for that matter) in the numerical integration of the convolution terms will result in a computationally unattractive solution procedure requiring storage of the generalized displacements for the entire deformation history. When N is large, the computational time expended at a given time step can become dominated by the task of evaluating the convolution integrals. Of course the storage required will also negatively affect the amount of memory needed in a given simulation. Since the viscoelastic relaxation moduli are expressed in terms of Prony series, it is possible to develop an efficient recurrence based temporal integration algorithm that requires only the storage of the generalized displacements and a set of internal variables evaluated at the Gauss points, both from the immediate previous time step only.

We assume, without loss of generality, that the quasi-static semi-discrete finite element equations have been successfully integrated temporally up until $t = t_k$. Our goal, therefore, is to numerically integrate the finite element equations over the subinterval $\mathcal{I}_k$ to obtain the solution for the generalized displacements at $t = t_{k+1}$. Before proceeding we must emphasize that all subsequent discussions regarding efficient recurrence based temporal integration strategies rely on the following multiplicative decompositions of the Prony series terms appearing in the definition of the relaxation moduli [111]

$$\dot{\bar{E}}_l(t_{k+1} - s) = e^{-\Delta t_{k+1}/\tau_l^E} \dot{\bar{E}}_l(t_k - s), \qquad \dot{\bar{G}}_l(t_{k+1} - s) = e^{-\Delta t_{k+1}/\tau_l^G} \dot{\bar{G}}_l(t_k - s) \quad (5.23)$$

where $\Delta t_{k+1} = t_{k+1} - t_k$ is the time step associated with subinterval $\mathcal{I}_k$. With the above formulas in mind, we note that the components of $\Lambda_i^\alpha(t_{k+1}, s)$ may be

conveniently expressed as

$$\Lambda_i^\alpha(t_{k+1}, s) = \sum_{j=1}^{n_\alpha} {}^j\bar{\Lambda}_i^\alpha(t_{k+1}, s) \tag{5.24}$$

where $n_1 = 1$, $n_2 = 3$ and $n_3 = 2$. The components ${}^j\bar{\Lambda}_i^\alpha(t_{k+1}, s)$ can be decomposed multiplicatively using the following general formula

$${}^j\bar{\Lambda}_i^\alpha(t_{k+1}, s) = \sum_{m=1}^{\text{NGP}} \sum_{l=1}^{\text{NPS}} {}^j_m\chi_i^\alpha(t_{k+1})^j_l\beta^\alpha(\Delta t_{k+1})^j_{lm}\kappa^\alpha(t_k, s)W_m \tag{5.25}$$

In the above expression we have employed the Gauss-Legendre quadrature rule in evaluation of all spatial integrals (resulting in summation over m). The quantity W_m represents the mth quadrature weight associated with the Gauss-Legendre quadrature rule. Summation over l is due to the Prony series representation of the relaxation moduli. The multiplicative decomposition of each ${}^j\bar{\Lambda}_i^\alpha(t_{k+1}, s)$ is essential for the recurrence based integration strategy. The components of ${}^j_m\chi_i^\alpha(t_{k+1})$ are used to store the discrete finite element test functions as well as any nonlinear quantities associated with the first variation of the simplified Green-Lagrange strain tensor. In the present formulation the components of ${}^j_m\chi_i^\alpha(t_{k+1})$ are defined as

$${}^1_m\chi_i^1(t_{k+1}) = \frac{d\psi_i^{(1)}(x_m)}{dx} \tag{5.26a}$$

$${}^1_m\chi_i^2(t_{k+1}) = \frac{\partial w_0(x_m, t_{k+1})}{\partial x}\frac{d\psi_i^{(2)}(x_m)}{dx} \tag{5.26b}$$

$${}^2_m\chi_i^2(t_{k+1}) = \frac{d^2\psi_i^{(2)}(x_m)}{dx^2} \tag{5.26c}$$

$${}^3_m\chi_i^2(t_{k+1}) = \frac{d\psi_i^{(2)}(x_m)}{dx} \tag{5.26d}$$

$${}^1_m\chi_i^3(t_{k+1}) = {}^1_m\chi_i^1(t_{k+1}) \tag{5.26e}$$

$${}^2_m\chi_i^3(t_{k+1}) = \psi_i^{(1)}(x_m) \tag{5.26f}$$

In the above expression, x_m represents the value of x as evaluated at the mth quadra-

ture point of a given finite element. The isoparametric mapping $\hat{\Omega}^e \rightleftarrows \bar{\Omega}^e$ used to characterize the geometry of each element allows for simple evaluation of such expressions. The components of ${}^{j}_{l}\beta^{\alpha}(\Delta t_{k+1})$ are defined as

$$\begin{aligned} &{}^{1}_{l}\beta^{1}(\Delta t_{k+1}) = {}^{1}_{l}\beta^{2}(\Delta t_{k+1}) = {}^{2}_{l}\beta^{2}(\Delta t_{k+1}) = {}^{1}_{l}\beta^{3}(\Delta t_{k+1}) = e^{-\Delta t_{k+1}/\tau_l^E} \\ &{}^{3}_{l}\beta^{2}(\Delta t_{k+1}) = {}^{2}_{l}\beta^{3}(\Delta t_{k+1}) = e^{-\Delta t_{k+1}/\tau_l^G} \end{aligned} \tag{5.27}$$

Likewise, the components of ${}^{j}_{lm}\kappa^{\alpha}(t_k, s)$ may be determined using the following formulas

$${}^{1}_{lm}\kappa^{1}(t_k, s) = \dot{\bar{E}}_l(t_k - s) D_0 \left[\frac{\partial u_0(x_m, s)}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0(x_m, s)}{\partial x} \right)^2 \right] \tag{5.28a}$$

$${}^{1}_{lm}\kappa^{2}(t_k, s) = {}^{1}_{lm}\kappa^{1}(t_k, s) \tag{5.28b}$$

$${}^{2}_{lm}\kappa^{2}(t_k, s) = \dot{\bar{E}}_l(t_k - s)\left(c_1^2 D_6 \frac{\partial^2 w_0(x_m, s)}{\partial x^2} - L_4 \frac{\partial \phi_x(x_m, s)}{\partial x} \right) \tag{5.28c}$$

$${}^{3}_{lm}\kappa^{2}(t_k, s) = \dot{\bar{G}}_l(t_k - s)\hat{A}_s\left(\frac{\partial w_0(x_m, s)}{\partial x} + \phi_x(x_m, s) \right) \tag{5.28d}$$

$${}^{1}_{lm}\kappa^{3}(t_k, s) = \dot{\bar{E}}_l(t_k - s)\left(M_2 \frac{\partial \phi_x(x_m, s)}{\partial x} - L_4 \frac{\partial^2 w_0(x_m, s)}{\partial x^2} \right) \tag{5.28e}$$

$${}^{2}_{lm}\kappa^{3}(t_k, s) = {}^{3}_{lm}\kappa^{2}(t_k, s) \tag{5.28f}$$

It is important to note that the components of ${}^{j}_{lm}\kappa^{\alpha}(t_k, s)$ have been defined such that the following multiplicative recurrence formulas hold

$${}^{j}_{lm}\kappa^{\alpha}(t_{k+1}, s) = {}^{j}_{l}\beta^{\alpha}(\Delta t_{k+1}) {}^{j}_{lm}\kappa^{\alpha}(t_k, s) \tag{5.29}$$

The above expressions are admissible on account of the assumption that the relaxation parameters are expressed in terms of Prony series.

We assume that at $t = t_k$ the components of the following expression are known

$$\int_0^{t_k} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_k, s) ds = \sum_{m=1}^{\text{NGP}} \sum_{l=1}^{\text{NPS}} {}^{j}_{m}\chi^{\alpha}_{i}(t_k) {}^{j}_{lm}X^{\alpha}(t_k) W_m \tag{5.30}$$

where ${}^{j}_{lm}X^{\alpha}(t_k)$ is a set of history variables (stored at the quadrature points of each element) that are of the form

$$ {}^{j}_{lm}X^{\alpha}(t_k) = \int_0^{t_k} {}^{j}_{lm}\kappa^{\alpha}(t_k, s)ds \tag{5.31} $$

We note that ${}^{j}_{lm}X^{\alpha}(0) = 0$. At $t = t_k$ the above history variables are known and there is no need to explicitly evaluate the expression appearing on the right hand side of Eq. (5.31). At the subsequent time step $t = t_{k+1}$ Eq. (5.30) may be written as

$$ \begin{aligned} \int_0^{t_{k+1}} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds &= \int_0^{t_k} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds + \int_{t_k}^{t_{k+1}} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds \\ &= \sum_{m=1}^{\text{NGP}} \sum_{l=1}^{\text{NPS}} {}^{j}_{m}\chi^{\alpha}_{i}(t_{k+1})^{j}_{l}\beta^{\alpha}(\Delta t_{k+1})^{j}_{lm}X^{\alpha}(t_k)W_m \\ &\quad + \int_{t_k}^{t_{k+1}} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds \end{aligned} \tag{5.32} $$

It is important to note that we have expressed the first integral on the right hand side of the above equation in terms of ${}^{j}_{lm}X^{\alpha}(t_k)$ (which is known from the previous time step). To integrate the remaining expression over the subinterval $\mathcal{I}_k$ we employ the trapezoidal rule which may be expressed as

$$ \begin{aligned} \int_{t_k}^{t_{k+1}} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds &\cong \frac{\Delta t_{k+1}}{2}\left[{}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, t_k) + {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, t_{k+1})\right] \\ &= \frac{\Delta t_{k+1}}{2} \sum_{m=1}^{\text{NGP}} \sum_{l=1}^{\text{NPS}} {}^{j}_{m}\chi^{\alpha}_{i}(t_{k+1})^{j}_{l}\beta^{\alpha}(\Delta t_{k+1})\left[{}^{j}_{lm}\kappa^{\alpha}(t_k, t_k)\right. \\ &\quad \left. + {}^{j}_{lm}\kappa^{\alpha}(t_k, t_{k+1})\right]W_m \end{aligned} \tag{5.33} $$

As a result, Eq. (5.32) can be written in the following simplified form

$$ \int_0^{t_{k+1}} {}^{j}\bar{\Lambda}^{\alpha}_{i}(t_{k+1}, s)ds = \sum_{m=1}^{\text{NGP}} \sum_{l=1}^{\text{NPS}} {}^{j}_{m}\chi^{\alpha}_{i}(t_{k+1})^{j}_{lm}X^{\alpha}(t_{k+1})W_m \tag{5.34} $$

where

$$
\begin{aligned}
{}^{j}_{lm}X^{\alpha}(t_{k+1}) = \frac{\Delta t_{k+1}}{2}{}^{j}_{l}\beta^{\alpha}(\Delta t_{k+1})\left[{}^{j}_{lm}\kappa^{\alpha}(t_k,t_k) + {}^{j}_{lm}\kappa^{\alpha}(t_k,t_{k+1})\right] \\
+ {}^{j}_{l}\beta^{\alpha}(\Delta t_{k+1}){}^{j}_{lm}X^{\alpha}\left(t_k\right)
\end{aligned}
\tag{5.35}
$$

As a result, in Eq. (5.34) we have developed a general expression for integrating the viscoelastic terms up to any discrete instance in time. The expression relies on a recurrence relationship defined in terms of the set of history variables ${}^{j}_{lm}X^{\alpha}(t_{k+1})$. These variables must be stored in memory at the immediate previous time step and may be updated to the subsequent time step in accordance with the procedure outlined in Eq. (5.35). The history variables may be expressed explicitly as

$$
\begin{aligned}
{}^{1}_{lm}X^{1}(t_{k+1}) = \frac{\Delta t_{k+1}}{2}D_0\Bigg\{\dot{\bar{E}}_l(\Delta t_{k+1})\left[\frac{\partial u_0(x_m,t_k)}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0(x_m,t_k)}{\partial x}\right)^2\right] \\
+ \dot{\bar{E}}_l(0)\left[\frac{\partial u_0(x_m,t_{k+1})}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0(x_m,t_{k+1})}{\partial x}\right)^2\right]\Bigg\} \\
+ e^{-\Delta t_{k+1}/\tau_l^E}{}^{1}_{lm}X^{1}(t_k)
\end{aligned}
\tag{5.36a}
$$

$$
{}^{1}_{lm}X^{2}(t_{k+1}) = {}^{1}_{lm}X^{1}(t_{k+1}) \tag{5.36b}
$$

$$
\begin{aligned}
{}^{2}_{lm}X^{2}(t_{k+1}) = \frac{\Delta t_{k+1}}{2}\Bigg[\dot{\bar{E}}_l(\Delta t_{k+1})\left(c_1^2 D_6\frac{\partial^2 w_0(x_m,t_k)}{\partial x^2} - L_4\frac{\partial \phi_x(x_m,t_k)}{\partial x}\right) \\
+ \dot{\bar{E}}_l(0)\left(c_1^2 D_6\frac{\partial^2 w_0(x_m,t_{k+1})}{\partial x^2} - L_4\frac{\partial \phi_x(x_m,t_{k+1})}{\partial x}\right)\Bigg] \\
+ e^{-\Delta t_{k+1}/\tau_l^E}{}^{2}_{lm}X^{2}(t_k)
\end{aligned}
\tag{5.36c}
$$

$$
\begin{aligned}
{}^{3}_{lm}X^{2}(t_{k+1}) = \frac{\Delta t_{k+1}}{2}\hat{A}_s\Bigg[\dot{\bar{G}}_l(\Delta t_{k+1})\left(\frac{\partial w_0(x_m,t_k)}{\partial x} + \phi_x(x_m,t_k)\right) \\
+ \dot{\bar{G}}_l(0)\left(\frac{\partial w_0(x_m,t_{k+1})}{\partial x} + \phi_x(x_m,t_{k+1})\right)\Bigg] + e^{-\Delta t_{k+1}/\tau_l^G}{}^{3}_{lm}X^{2}(t_k)
\end{aligned}
\tag{5.36d}
$$

$$
\begin{aligned}
{}^{1}_{lm}X^{3}(t_{k+1}) = \frac{\Delta t_{k+1}}{2}\Bigg[\dot{\bar{E}}_l(\Delta t_{k+1})\left(M_2\frac{\partial \phi_x(x_m,t_k)}{\partial x} - L_4\frac{\partial^2 w_0(x_m,t_k)}{\partial x^2}\right) \\
+ \dot{\bar{E}}_l(0)\left(M_2\frac{\partial \phi_x(x_m,t_{k+1})}{\partial x} - L_4\frac{\partial^2 w_0(x_m,t_{k+1})}{\partial x^2}\right)\Bigg] \\
+ e^{-\Delta t_{k+1}/\tau_l^E}{}^{1}_{lm}X^{3}(t_k)
\end{aligned}
\tag{5.36e}
$$

$$ {}^{2}_{lm}X^{3}(t_{k+1}) = {}^{3}_{lm}X^{2}(t_{k+1}) \tag{5.36f} $$

It is now possible to express the fully discretized finite element equations at the current time step as

$$ [\bar{K}]_{k+1}\{\Delta\}_{k+1} = \{F\}_{k+1} - \{\tilde{Q}\}_{k+1} \tag{5.37} $$

where

$$ \bar{K}^{11}_{ij} = \int_{x_a}^{x_b} \left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) D_0 \frac{d\psi_i^{(1)}}{dx}\frac{d\psi_j^{(1)}}{dx} dx \tag{5.38a} $$

$$ \bar{K}^{12}_{ij} = \frac{1}{2}\bar{K}^{21}_{ji} = \frac{1}{2}\int_{x_a}^{x_b} \left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) D_0 \frac{\partial w_0(x,t_{k+1})}{\partial x}\frac{d\psi_i^{(1)}}{dx}\frac{d\psi_j^{(2)}}{dx} dx \tag{5.38b} $$

$$ \bar{K}^{13}_{ij} = \bar{K}^{31}_{ji} = 0 \tag{5.38c} $$

$$ \begin{aligned} \bar{K}^{22}_{ij} = \int_{x_a}^{x_b} \Bigg[& \frac{1}{2}\left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) D_0 \left(\frac{\partial w_0(x,t_{k+1})}{\partial x} \right)^2 \frac{d\psi_i^{(2)}}{dx}\frac{d\psi_j^{(2)}}{dx} \\ & + \left(E\,(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) c_1^2 D_6 \frac{d^2\psi_i^{(2)}}{dx^2}\frac{d^2\psi_j^{(2)}}{dx^2} \\ & + \left(G\,(0) + \frac{\Delta t_{k+1}}{2}\dot{G}(0) \right) \hat{A}_s \frac{d\psi_i^{(2)}}{dx}\frac{d\psi_j^{(2)}}{dx} \Bigg] dx \end{aligned} \tag{5.38d} $$

$$ \begin{aligned} \bar{K}^{23}_{ij} = \bar{K}^{32}_{ji} = \int_{x_a}^{x_b} \Bigg[& \left(G(0) + \frac{\Delta t_{k+1}}{2}\dot{G}(0) \right) \hat{A}_s \frac{d\psi_i^{(2)}}{dx}\psi_j^{(1)} \\ & - \left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) L_4 \frac{d^2\psi_i^{(2)}}{dx^2}\frac{d\psi_j^{(1)}}{dx} \Bigg] dx \end{aligned} \tag{5.38e} $$

$$ \begin{aligned} \bar{K}^{33}_{ij} = \int_{x_a}^{x_b} \Bigg[& \left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0) \right) M_2 \frac{d\psi_i^{(1)}}{dx}\frac{d\psi_j^{(1)}}{dx} \\ & + \left(G(0) + \frac{\Delta t_{k+1}}{2}\dot{G}(0) \right) \hat{A}_s \psi_i^{(1)}\psi_j^{(1)} \Bigg] dx \end{aligned} \tag{5.38f} $$

and

$$ \tilde{Q}_i^{\alpha} = \sum_{j=1}^{n_\alpha} {}^{j}\bar{Q}_i^{\alpha} \tag{5.39} $$

The components of ${}^{j}\bar{Q}_i^{\alpha}(t_{k+1})$ are of the form

$$
\begin{aligned}
{}^{1}\bar{Q}_i^{1} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{E}(\Delta t_{k+1})D_0\frac{d\psi_i^{(1)}}{dx}\left[\frac{\partial u_0(x,t_k)}{\partial x}+\frac{1}{2}\left(\frac{\partial w_0(x,t_k)}{\partial x}\right)^2\right]dx \qquad (5.40a)\\
&\quad+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^E}\frac{d\psi_i^{(1)}(x_m)}{dx}{}^{1}_{lm}X^1(t_k)W_m\\
{}^{1}\bar{Q}_i^{2} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{E}(\Delta t_{k+1})D_0\frac{\partial w_0(x,t_{k+1})}{\partial x}\frac{d\psi_i^{(2)}}{dx}\left[\frac{\partial u_0(x,t_k)}{\partial x}\right. \qquad (5.40b)\\
&\quad\left.+\frac{1}{2}\left(\frac{\partial w_0(x,t_k)}{\partial x}\right)^2\right]dx+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^E}\frac{\partial w_0(x_m,t_{k+1})}{\partial x}\times\\
&\quad\frac{d\psi_i^{(2)}(x_m)}{dx}{}^{1}_{lm}X^2(t_k)W_m\\
{}^{2}\bar{Q}_i^{2} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{E}(\Delta t_{k+1})\frac{d^2\psi_i^{(2)}}{dx^2}\left(c_1^2D_6\frac{\partial^2 w_0(x,t_k)}{\partial x^2}-L_4\frac{\partial\phi_x(x,t_k)}{\partial x}\right)dx \qquad (5.40c)\\
&\quad+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^E}\frac{d^2\psi_i^{(2)}(x_m)}{dx^2}{}^{2}_{lm}X^2(t_k)W_m\\
{}^{3}\bar{Q}_i^{2} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{G}(\Delta t_{k+1})\hat{A}_s\frac{d\psi_i^{(2)}}{dx}\left(\frac{\partial w_0(x,t_k)}{\partial x}+\phi_x(x,t_k)\right)dx \qquad (5.40d)\\
&\quad+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^G}\frac{d\psi_i^{(2)}(x_m)}{dx}{}^{3}_{lm}X^2(t_k)W_m\\
{}^{1}\bar{Q}_i^{3} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{E}(\Delta t_{k+1})\frac{d\psi_i^{(1)}}{dx}\left(M_2\frac{\partial\phi_x(x,t_k)}{\partial x}-L_4\frac{\partial^2 w_0(x,t_k)}{\partial x^2}\right)dx \qquad (5.40e)\\
&\quad+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^E}\frac{d\psi_i^{(1)}(x_m)}{dx}{}^{1}_{lm}X^3(t_k)W_m\\
{}^{2}\bar{Q}_i^{3} &= \frac{\Delta t_{k+1}}{2}\int_{x_a}^{x_b}\dot{G}(\Delta t_{k+1})\hat{A}_s\psi_i^{(1)}\left(\frac{\partial w_0(x,t_k)}{\partial x}+\phi_x(x,t_k)\right)dx \qquad (5.40f)\\
&\quad+\sum_{m=1}^{\text{NGP}}\sum_{l=1}^{\text{NPS}}e^{-\Delta t_{k+1}/\tau_l^G}\psi_i^{(1)}(x_m){}^{2}_{lm}X^3(t_k)W_m
\end{aligned}
$$

4. A Newton based iterative solution procedure

The fully discretized finite element equations are nonlinear due to the use of the von Kármán strain components in the definition of the effective strain tensor ε. In our

work, we adopt the Newton procedure in the iterative solution of the nonlinear finite element equations. The resulting linearized finite element equations are of the form

$$[T^e]^{(r)}_{k+1}\{\delta\Delta^e\}^{(r+1)}_{k+1} = -([\bar{K}^e]^{(r)}_{k+1}\{\Delta^e\}^{(r)}_{k+1} - \{F^e\}^{(r)}_{k+1} + \{\tilde{Q}^e\}^{(r)}_{k+1}) \tag{5.41}$$

where $\{\delta\Delta^e\}^{(r+1)}_{k+1}$ represents the incremental solution at the $(r+1)$th nonlinear iteration. The total global solution at the $(r+1)$th iteration is obtained as

$$\{\Delta\}^{(r+1)}_{k+1} = \{\delta\Delta\}^{(r+1)}_{k+1} + \{\Delta\}^{(r)}_{k+1} \tag{5.42}$$

The element tangent stiffness matrix $[T^e]^{(r)}_{k+1}$ appearing in the Newton linearization of the finite element equations may be expressed (using Einstein's summation convention over n) as

$$T^e_{ij} = \bar{K}^e_{ij} + \frac{\partial \bar{K}^e_{in}}{\partial \Delta^e_j}\Delta^e_n + \frac{\partial \tilde{Q}^e_i}{\partial \Delta^e_j} \tag{5.43}$$

All quantities comprising the tangent stiffness matrix are formulated using the solution from the rth iteration. The partial derivatives are taken with respect to the solution at the current time step. The components of the tangent stiffness matrix may be determined using the following general formulas

$$T^{11}_{ij} = \bar{K}^{11}_{ij} \tag{5.44a}$$

$$T^{12}_{ij} = T^{21}_{ji} = 2\bar{K}^{12}_{ij} \tag{5.44b}$$

$$T^{13}_{ij} = \bar{T}^{31}_{ji} = 0 \tag{5.44c}$$

$$\begin{aligned} T^{22}_{ij} = \int_{x_a}^{x_b} D_0 \Bigg\{ \left(E(0) + \frac{\Delta t_{k+1}}{2}\dot{E}(0)\right)\left[\frac{\partial u_0(x,t_{k+1})}{\partial x} + \left(\frac{\partial w_0(x,t_{k+1})}{\partial x}\right)^2\right] \\ + \frac{\Delta t_{k+1}}{2}\dot{E}(\Delta t_{k+1})\left[\frac{\partial u_0(x,t_k)}{\partial x} + \frac{1}{2}\left(\frac{\partial w_0(x,t_k)}{\partial x}\right)^2\right]\Bigg\}\frac{d\psi^{(2)}_i}{dx}\frac{d\psi^{(2)}_j}{dx}dx \\ + \sum_{m=1}^{\mathrm{NGP}}\sum_{l=1}^{\mathrm{NPS}} e^{-\Delta t_{k+1}/\tau^E_l}\frac{d\psi^{(2)}_i(x_m)}{dx}\frac{d\psi^{(2)}_j(x_m)}{dx}{}_{1}{}_{lm}X^2(t_k)W_m + \bar{K}^{22}_{ij} \end{aligned} \tag{5.44d}$$

$$T_{ij}^{23} = T_{ji}^{32} = \bar{K}_{ij}^{23} \tag{5.44e}$$

$$T_{ij}^{33} = \bar{K}_{ij}^{33} \tag{5.44f}$$

Clearly the tangent stiffness coefficient matrix is symmetric.

5. Numerical locking and high-order finite element interpolation functions

It is well-known that low-order finite elements for beams are prone to locking [50, 105, 3] when quadrature rules are employed that result in exact integration of the element coefficient matrices and force vectors. To circumvent the locking phenomena, we consider two philosophically dissimilar numerical procedures. In the first approach, we employ the lowest order element admissible in the formulation (i.e., a two-node element). Selective full and one point Gauss-Legendre quadrature rules are applied; where reduced integration techniques are employed on all nonlinear expressions associated with the finite element model. This element is denoted as an RBT-2-R element (meaning a two-node reduced integration RBT element). It is worth noting that this element requires a splitting of the history variables into subsets associated with the full and reduced integration points. In the second approach, we construct the Reddy beam finite elements using high polynomial order expansions of the dependent variables, by systematically increasing the number of nodes per finite element. In this approach, the same quadrature formulas may be adopted in the evaluation of all expressions appearing in the coefficient matrices and force vectors of the finite element model. The resulting elements are denoted in this work as RBT-n elements, where n represents the number of nodes per element. In Section C we also present numerical results obtained using Euler-Bernoulli and Timoshenko beam elements denoted as EBT-n and TBT-n respectively.

The high-order Lagrange interpolation functions $\{\psi_i^{(1)}\}_{i=1}^n$ are constructed, with

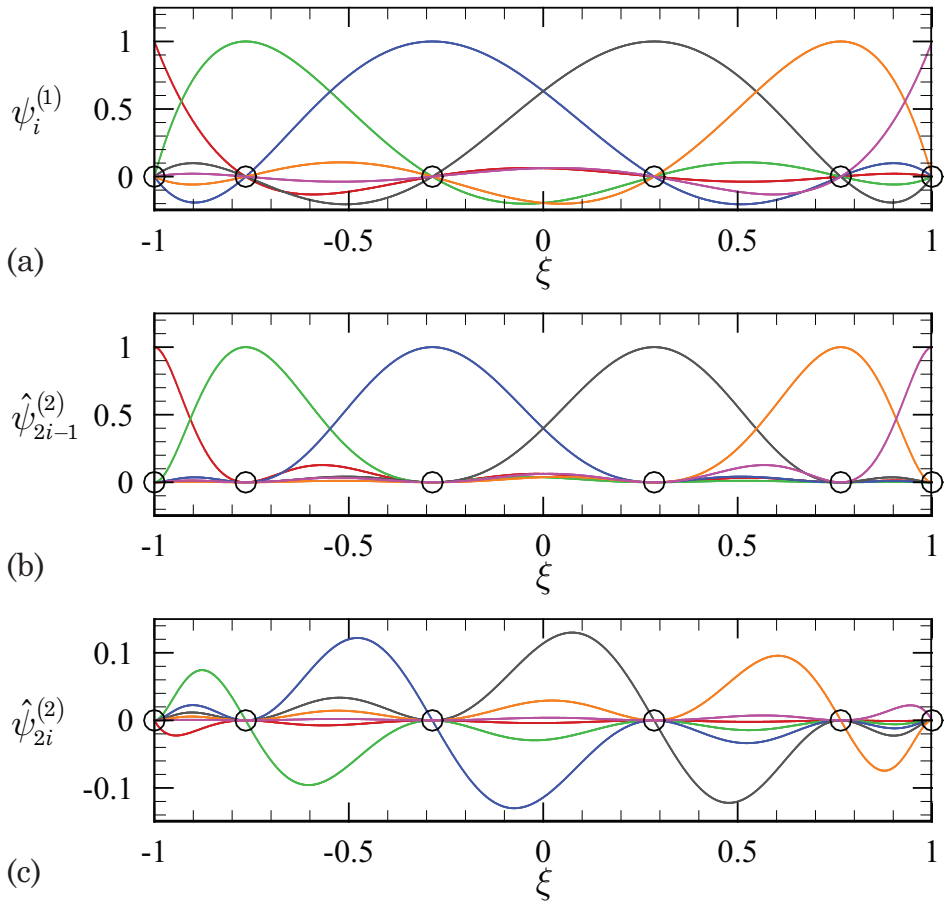


Fig. 47. Interpolation functions for a high-order RBT finite element where $n = 6$ and $i = 1, \ldots, n$: (a) Lagrange interpolation functions $\psi_i^{(1)}$, (b) Hermite interpolation functions $\hat{\psi}_{2i-1}^{(2)}$ and (c) Hermite interpolation functions $\hat{\psi}_{2i}^{(2)}$.

respect to the natural coordinate ξ, using the C^0 spectral nodal interpolation formula given in Eq. (2.4). The high-order Hermite interpolation functions $\{\hat{\psi}_i^{(2)}\}_{i=1}^{2n}$, on the other hand, may be developed for the *master element* $\hat{\Omega}^e$ through the use of the following expression

$$\hat{\psi}_i^{(2)}(\xi) = \sum_{j=1}^{2n} c_i^{j-1} \xi^{j-1} \tag{5.45}$$

The coefficients c_i^{j-1} appearing in the above equation may be determined by imposing the following compatibility conditions on the interpolation functions

$$\hat{\psi}^{(2)}_{2i-1}(\xi_j) = -\frac{d\hat{\psi}^{(2)}_{2i}}{d\xi}\bigg|_{\xi=\xi_j} = \delta_{ij}, \qquad \frac{d\hat{\psi}^{(2)}_{2i-1}}{d\xi}\bigg|_{\xi=\xi_j} = \hat{\psi}^{(2)}_{2i}(\xi_j) = 0 \tag{5.46}$$

where i and j both range from 1 to n. The Hermite interpolation functions $\{\psi_i^{(2)}\}_{i=1}^{2n}$ associated with the *physical element* $\bar{\Omega}^e$ may be determined as

$$\psi^{(2)}_{2i-1}(\xi) = \hat{\psi}^{(2)}_{2i-1}(\xi), \qquad \psi^{(2)}_{2i}(\xi) = J^e\hat{\psi}^{(2)}_{2i}(\xi) \tag{5.47}$$

where $J^e = dx/d\xi$ is the Jacobian of the element coordinate transformation $\hat{\Omega}^e \rightleftarrows \bar{\Omega}^e$. With the above formulas in mind, we are now in a position to be able to generate the interpolation functions for a beam element possessing any number of nodes per element. The standard lowest order two-node element may be obtained as a special case. The interpolation functions associated with a six-node RBT finite element are shown in Figure 47.

C. Numerical examples: verification benchmarks

In this section, numerical results are presented and tabulated for the mechanical response of viscoelastic beam structures obtained using the proposed finite element formulation for the Euler-Bernoulli, Timoshenko and third-order Reddy beam theories. The results have been obtained using the Newton solution procedure described previously. Nonlinear convergence is declared at the current step once the Euclidean norm of the normalized difference between the nonlinear iterative solution increments (i.e., $\|\{\Delta\}_{k+1}^{(r+1)} - \{\Delta\}_{k+1}^{(r)}\|/\|\{\Delta\}_{k+1}^{(r+1)}\|$), is less than 10^{-6}.

The material model utilized in the quasi-static numerical studies is based upon the experimental results tabulated by Lai and Bakker [112] for a glassy amorphous

polymer material (PMMA). The Prony series parameters for the viscoelastic relaxation modulus given in Table VIII were calculated by Payette and Reddy [21] from the published compliance parameters [112]. As in the work of Chen [91] and Payette and Reddy [21], we assume that Poisson's ratio is time invariant. As a result, the shear relaxation modulus is given as

$$G(t) = \frac{E(t)}{2(1+\nu)} \tag{5.48}$$

where Poisson's ratio is taken to be $\nu = 0.40$ [113].

Table VIII. Viscoelastic relaxation parameters for a PMMA.

E_0	205.7818 ksi		
E_1	43.1773 ksi	τ_1^E	9.1955×10^{-1} sec.
E_2	9.2291 ksi	τ_2^E	9.8120×10^{0} sec.
E_3	22.9546 ksi	τ_3^E	9.5268×10^{1} sec.
E_4	26.2647 ksi	τ_4^E	9.4318×10^{2} sec.
E_5	34.6298 ksi	τ_5^E	9.2066×10^{3} sec.
E_6	40.3221 ksi	τ_6^E	8.9974×10^{4} sec.
E_7	47.5275 ksi	τ_7^E	8.6852×10^{5} sec.
E_8	46.8108 ksi	τ_8^E	8.5143×10^{6} sec.
E_9	58.6945 ksi	τ_9^E	7.7396×10^{7} sec.

1. Quasi-static mechanical response

a. Deflection of a thin beam under uniform loading

In this first example we consider a viscoelastic beam of length $L = 100$ in. and cross section 1 in. $\times$ 1 in. At $t = 0$ sec. the beam is subjected to a uniform vertically distributed load $q_0 = 0.25$ lb_f/in that is maintained for 1,800 sec. Due to symmetry about $x = L/2$, it is only necessary to computationally model half of the physical domain; as a result, we take $\bar{\Omega}^{hp} = [0, L/2]$. To assess the performance of various

finite element discretizations in circumventing the locking phenomena, we consider the following three sets of boundary conditions (for the RBT):

1. Hinged at both ends

$$w_0(0,t) = u_0(L/2,t) = \frac{\partial w_0}{\partial x}(L/2,t) = \phi_x(L/2,t) = 0 \tag{5.49}$$

2. Pinned at both ends

$$u_0(0,t) = w_0(0,t) = u_0(L/2,t) = \frac{\partial w_0}{\partial x}(L/2,t) = \phi_x(L/2,t) = 0 \tag{5.50}$$

3. Clamped at both ends

$$\begin{aligned} u_0(0,t) = w_0(0,t) = \frac{\partial w_0}{\partial x}(0,t) = \phi_x(0,t) = 0 \\ u_0(L/2,t) = \frac{\partial w_0}{\partial x}(L/2,t) = \phi_x(L/2,t) = 0 \end{aligned} \tag{5.51}$$

Similar boundary conditions may also be adopted for the Euler-Bernoulli and Timoshenko beam theories.

In the numerical implementation we discretize the computational domain $\bar{\Omega}^{hp}$ using 10 $\mathbb{Q}$BT-2 elements (11 nodes), 5 $\mathbb{Q}$BT-3 elements (11 nodes), 3 $\mathbb{Q}$BT-4 elements (10 nodes) and 2 $\mathbb{Q}$BT-6 elements (11 nodes), where $\mathbb{Q}$ = E, T or R. An equal time increment $\Delta t = 1.0$ sec. has been employed for all time steps. Five load steps were used in each simulation at $t = 0$ to insure nonlinear convergence of the instantaneous elastic response. At each subsequent time step the finite element equations were solved iteratively using the Newton procedure, which typically required only 2 or 3 nonlinear iterations. In Tables IX and X and we summarize the numerical results for the maximum vertical deflection of the viscoelastic beam for the three different sets of boundary conditions listed above. In Figures 48 and 49 we also plot the maximum vertical deflections for the three cases as obtained using the RBT-6 discretization.

Table IX. Quasi-static EBT and TBT finite element solutions for the maximum vertical deflection $w_0(L/2,t)$ of a viscoelastic beam under uniform load q_0 with three different boundary conditions.

Time, t	EBT-2-R	TBT-2	TBT-3	TBT-4	TBT-6
Hinged-hinged					
0	7.2961	0.8629	7.0098	7.2939	7.2980
200	8.6194	1.0194	8.1966	8.6151	8.6217
400	8.7617	1.0363	8.3221	8.7571	8.7641
600	8.8486	1.0465	8.3986	8.8439	8.8510
800	8.9183	1.0548	8.4598	8.9134	8.9207
1,000	8.9775	1.0618	8.5118	8.9725	8.9799
1,200	9.0287	1.0678	8.5567	9.0236	9.0311
1,400	9.0733	1.0731	8.5958	9.0681	9.0758
1,600	9.1126	1.0778	8.6301	9.1073	9.1150
1,800	9.1474	1.0819	8.6605	9.1420	9.1498
Pinned-pinned					
0	1.2481	0.7258	1.2452	1.2453	1.2452
200	1.3278	0.8210	1.3244	1.3243	1.3242
400	1.3358	0.8307	1.3324	1.3323	1.3322
600	1.3407	0.8366	1.3372	1.3371	1.3370
800	1.3446	0.8413	1.3411	1.3410	1.3409
1,000	1.3478	0.8452	1.3443	1.3442	1.3441
1,200	1.3507	0.8486	1.3471	1.3470	1.3469
1,400	1.3531	0.8516	1.3496	1.3495	1.3494
1,600	1.3553	0.8542	1.3517	1.3516	1.3515
1,800	1.3572	0.8565	1.3536	1.3535	1.3534
Clamped-clamped					
0	0.9110	0.1727	0.8832	0.9102	0.9109
200	1.0000	0.2038	0.9707	0.9988	0.9997
400	1.0089	0.2071	0.9795	1.0077	1.0086
600	1.0144	0.2092	0.9848	1.0130	1.0140
800	1.0187	0.2108	0.9891	1.0173	1.0183
1,000	1.0223	0.2122	0.9927	1.0210	1.0220
1,200	1.0255	0.2134	0.9957	1.0241	1.0251
1,400	1.0282	0.2144	0.9984	1.0268	1.0278
1,600	1.0306	0.2154	1.0008	1.0292	1.0302
1,800	1.0327	0.2162	1.0029	1.0313	1.0323

Table X. Quasi-static RBT finite element solutions for the maximum vertical deflection $w_0(L/2,t)$ of a viscoelastic beam under uniform load q_0 with three different boundary conditions.

Time, t	RBT-2	RBT-2-R	RBT-3	RBT-4	RBT-6
Hinged-hinged					
0	5.4740	7.2840	7.2277	7.2946	7.2980
200	6.1234	8.6052	8.5170	8.6169	8.6217
400	6.1895	8.7473	8.6552	8.7592	8.7641
600	6.2295	8.8340	8.7396	8.8460	8.8510
800	6.2615	8.9035	8.8071	8.9156	8.9207
1,000	6.2886	8.9627	8.8646	8.9748	8.9799
1,200	6.3119	9.0138	8.9143	9.0259	9.0311
1,400	6.3322	9.0584	8.9575	9.0705	9.0758
1,600	6.3499	9.0975	8.9956	9.1098	9.1150
1,800	6.3656	9.1322	9.0293	9.1445	9.1498
Pinned-pinned					
0	1.2442	1.2493	1.2452	1.2452	1.2452
200	1.3233	1.3291	1.3242	1.3242	1.3242
400	1.3313	1.3371	1.3322	1.3322	1.3322
600	1.3361	1.3420	1.3370	1.3370	1.3370
800	1.3399	1.3459	1.3409	1.3409	1.3409
1,000	1.3432	1.3492	1.3441	1.3441	1.3441
1,200	1.3460	1.3520	1.3470	1.3470	1.3469
1,400	1.3484	1.3545	1.3494	1.3494	1.3494
1,600	1.3506	1.3566	1.3515	1.3515	1.3515
1,800	1.3525	1.3585	1.3534	1.3534	1.3534
Clamped-clamped					
0	0.9037	0.9098	0.9106	0.9108	0.9109
200	0.9918	0.9992	0.9993	0.9995	0.9997
400	1.0007	1.0082	1.0083	1.0084	1.0086
600	1.0060	1.0136	1.0136	1.0138	1.0140
800	1.0103	1.0180	1.0180	1.0182	1.0183
1,000	1.0139	1.0216	1.0216	1.0218	1.0220
1,200	1.0170	1.0248	1.0247	1.0249	1.0251
1,400	1.0197	1.0275	1.0275	1.0277	1.0278
1,600	1.0221	1.0299	1.0298	1.0300	1.0302
1,800	1.0242	1.0321	1.0319	1.0322	1.0323

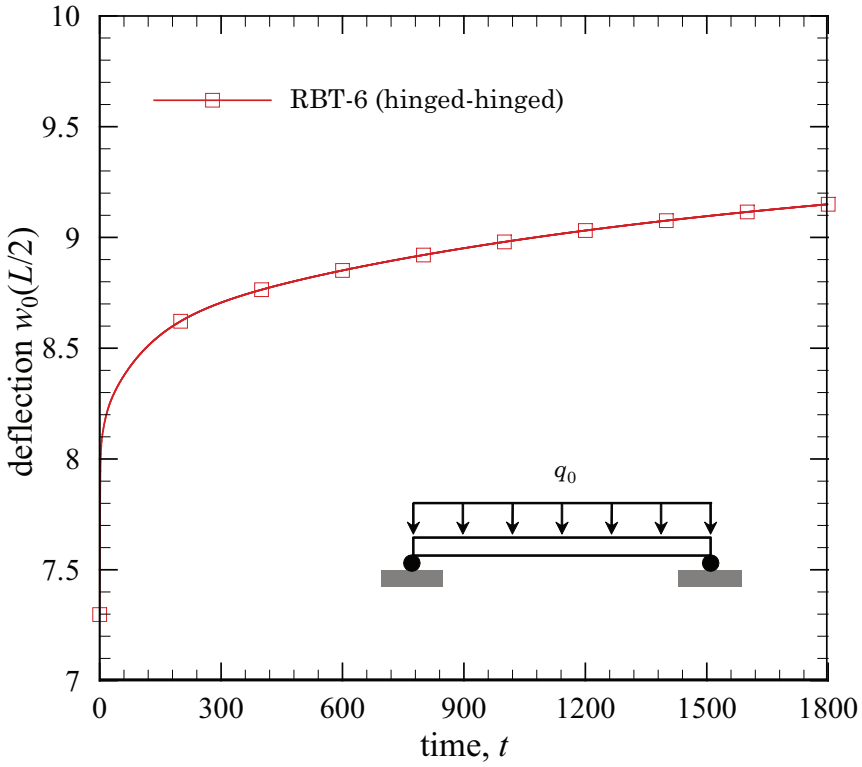


Fig. 48. Maximum vertical deflection $w_0(L/2, t)$ of a hinged-hinged viscoelastic beam subjected to a uniform vertically distributed load q_0, where two RBT-6 elements have been used in the finite element discretization.

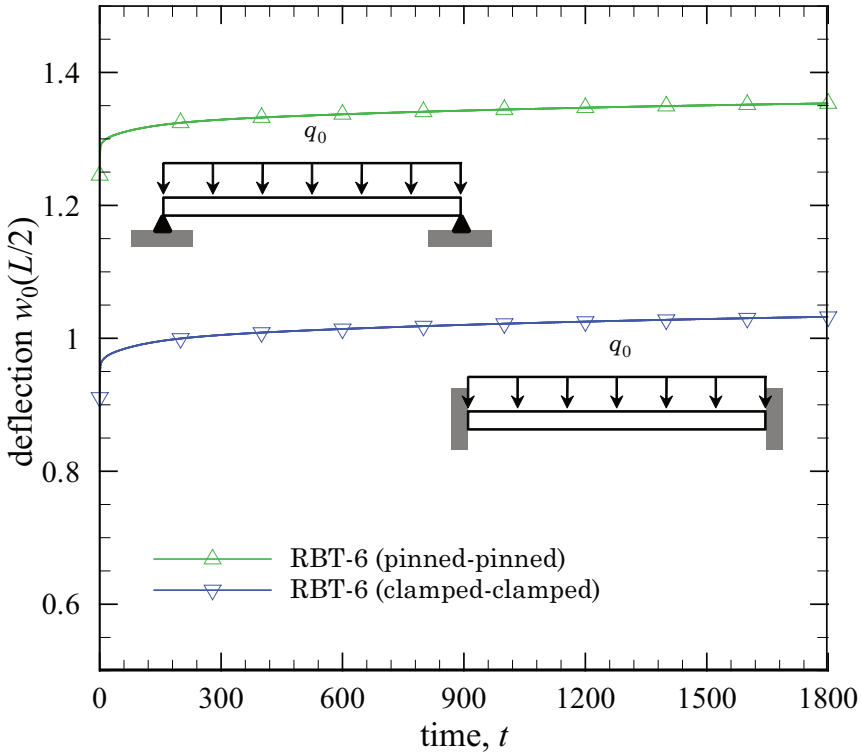


Fig. 49. Maximum vertical deflection $w_0(L/2, t)$ of both pinned-pinned and clamped–clamped viscoelastic beams subjected to a uniform vertically distributed load q_0, where two RBT-6 elements have been used in the finite element discretization.

It is evident from the numerical results that the TBT element suffers from numerical locking whenever low-order polynomial expansions are employed. This is also true of the RBT-2 element when used to solve the hinged-hinged problem. In general, however, we find that the RBT-n elements are less prone to locking than are the TBT-n elements. In fact, the RBT-3 element is almost completely locking free. On the other hand, the Timoshenko beam equations require the use of Lagrange interpolation functions only, and are hence simple to construct. The solutions resulting from the TBT-6 and RBT-6 discretizations are spatially fully converged and can actually be obtained using a coarse grid consisting of only a single element. As a result, the overall computational cost associated with these single element discretizations is actually quite low. It is interesting to note that the low-order EBT-2-R and RBT-2-R also provide reliable numerical results for the transverse deflection.

Table XI. Analytical and finite element solutions for the maximum quasi-static vertical deflection $w_0(L/2,t)$ of a hinged-hinged beam under uniform transverse loading q_0.

	Maximum vertical deflection, $w_0(L/2,t)$					
Time, t	Exact	$\Delta t = 0.1$	$\Delta t = 1.0$	$\Delta t = 2.0$	$\Delta t = 5.0$	$\Delta t = 10.0$
0	7.2980	7.2980	7.2980	7.2980	7.2980	7.2980
200	8.5429	8.5437	8.6217	8.8493	10.2278	14.7260
400	8.6827	8.6835	8.7641	8.9993	10.4291	15.1493
600	8.7680	8.7689	8.8510	9.0910	10.5524	15.4107
800	8.8364	8.8372	8.9207	9.1645	10.6513	15.6214
1,000	8.8945	8.8954	8.9799	9.2270	10.7356	15.8022
1,200	8.9448	8.9456	9.0311	9.2811	10.8087	15.9597
1,400	8.9886	8.9895	9.0758	9.3282	10.8726	16.0983
1,600	9.0271	9.0280	9.1150	9.3697	10.9288	16.2210
1,800	9.0612	9.0621	9.1498	9.4064	10.9787	16.3306

For the hinged-hinged beam configuration, the vertical deflection coincides with the exact solution of the geometrically linear theory. In Table XI we compare numer-

ical results obtained using two RBT-6 beam elements with the exact solution for the Timoshenko beam theory given by Flügge [81] as

$$w_0(L/2,t) = \frac{5q_0L^4}{384D_2}\left[1 + \frac{8(1+\nu)}{5\kappa}\left(\frac{h}{L}\right)^2\right]D(t) \tag{5.52}$$

where $D(t)$ is the creep compliance and κ is the shear correction factor. The error in the numerical solution due to temporal discretization based on the trapezoidal rule tends to over-predict the deflection of the beam as is evident in Table XI (where numerical solutions obtained using various time increment sizes are compared).

b. Deflection of a thick beam under uniform loading

In this next example we consider a thick viscoelastic beam (i.e., $L/h < 20$) to demonstrate the ability of the RBT finite element formulation to accurately account for deformations associated with shearing. We modify the thin beam problem given in the previous example by taking $L = 10$ in., $q = 25.0$ lb_f/in and $\Delta t = 1.0$ sec. All other geometric and material parameters are the same as in the previous example. In Table XII numerical results are presented for the transverse deflection of pinned-pinned and clamped-clamped beams. The same number of elements (per element type) are employed as in the previous example. In Table XII we also compare results from the Reddy beam theory with finite element solutions obtained using a low-order reduced integration finite element model based on the Euler-Bernoulli beam theory (which does not account for deformations associated with shearing). The numerical results obtained for the RBT element compare well with numerical solutions obtained using TBT finite elements [21].

Table XII. Comparison of the quasi-static finite element solutions for the maximum vertical deflection $w_0(L/2,t)$ of thick pinned-pinned and clamped-clamped viscoelastic beams under uniform transverse loading q_0.

	Maximum vertical deflection, $w_0(L/2,t)$					
	pinned-pinned			clamped-clamped		
Time t	EBT-2-R	RBT-2-R	RBT-6	EBT-2-R	RBT-2-R	RBT-6
0	0.07184	0.07362	0.07367	0.01459	0.01645	0.01653
200	0.08437	0.08643	0.08649	0.01724	0.01943	0.01952
400	0.08571	0.08779	0.08785	0.01752	0.01975	0.01985
600	0.08652	0.08862	0.08869	0.01769	0.01995	0.02004
800	0.08717	0.08929	0.08935	0.01783	0.02010	0.02020
1,000	0.08773	0.08985	0.08992	0.01795	0.02024	0.02033
1,200	0.08821	0.09034	0.09041	0.01805	0.02035	0.02045
1,400	0.08862	0.09076	0.09083	0.01814	0.02045	0.02055
1,600	0.08899	0.09114	0.09121	0.01822	0.02054	0.02064
1,800	0.08931	0.09147	0.09154	0.01829	0.02062	0.02072

c. Deflection of a thin beam under time-dependent loading

For this example we employ the geometric parameters, material properties and hinged-hinged boundary conditions utilized in the first numerical example. We replace the stationary uniformly distributed load with the following quasi-static transverse load

$$q(t) = q_0 \left\{ H(t) - \frac{1}{\tau(\beta-\alpha)} \left[(t-\alpha\tau)H(t-\alpha\tau) - (t-\beta\tau)H(t-\beta\tau) \right] \right\} \qquad (5.53)$$

where $q_0 = 0.25$ lb$_\mathrm{f}$/in, $\tau = 200$ sec. and $H(t)$ is the Heaviside function. The parameters α and β (where $0 \leq \alpha \leq \beta \leq 1$) are constants whose values may be appropriately adjusted. The load function above is constant for $0 < t < \alpha\tau$ and decays linearly from $t = \alpha\tau$ to $t = \beta\tau$, after which the load is maintained at zero. We utilize the above loading function to numerically demonstrate that the finite element model correctly predicts that the viscoelastic beam will eventually recover its original configuration upon removal of all externally applied mechanical loads. The numerical solution for

the problem, as obtained using two RBT-6 elements, is presented in Figure 50 for various values of α and β. It is clear that in all cases, the beam does tend to recover its original configuration following removal of the externally applied load.

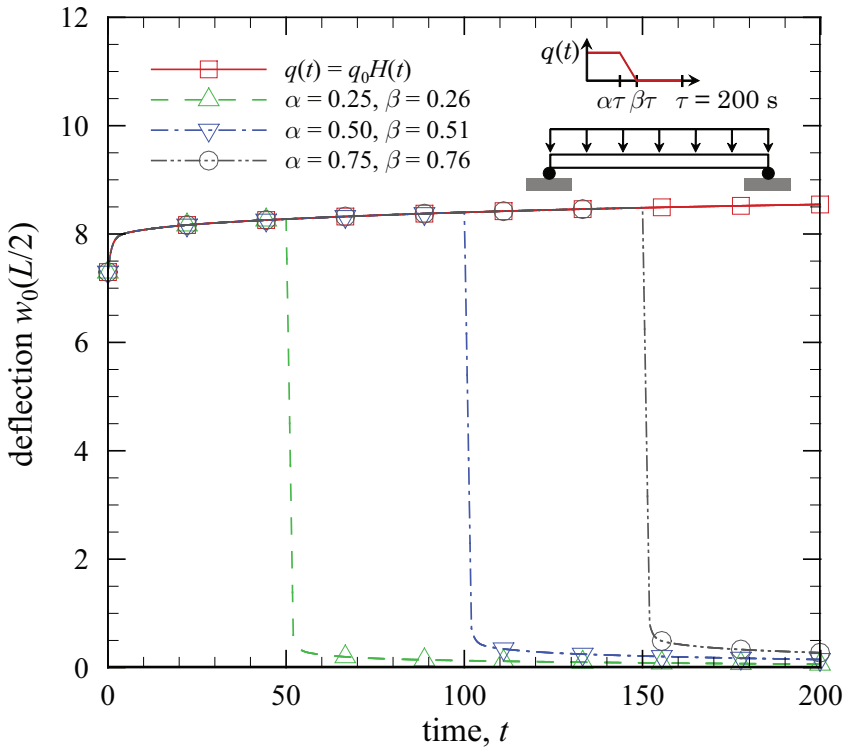


Fig. 50. Maximum vertical deflection $w_0(L/2, t)$ of a hinged-hinged viscoelastic beam subjected to a time-dependent transversely distributed load $q(t)$.

2. Fully-transient mechanical response

a. Forced vibrational response of hinged beams

As a final example, we consider the fully transient response of viscoelastic beams under mechanical loading as modeled using the third-order Reddy beam theory. For this example we employ a simple three parameter solid model utilized previously by Chen [91]. In the standard three parameter solid model, the relaxation modulus may

be expressed as

$$E(t) = \frac{k_1 k_2}{k_1 + k_2}\left(1 - e^{-t/\tau_1^E}\right) + k_1 e^{-t/\tau_1^E} \tag{5.54}$$

where in the present example $k_1 = 9.8 \times 10^7$ N/m^2 and $k_2 = 2.45 \times 10^7$ N/m^2. The relaxation time is of the form $\tau_1^E = \eta/(k_1 + k_2)$ and the material density is taken to be $\rho_0 = 500$ kg/m^3. A constant Poisson ratio of $\nu = 0.3$ is assumed.

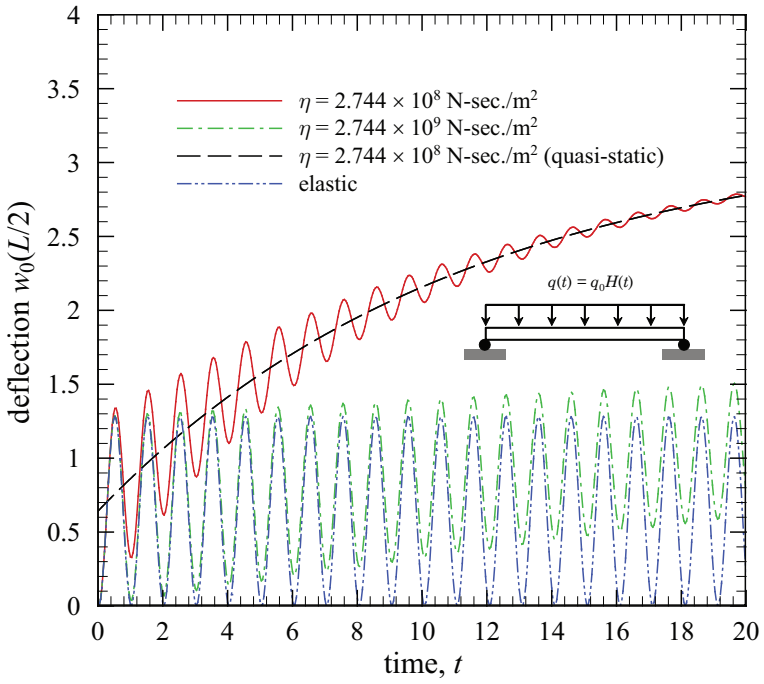


Fig. 51. A comparison of the time-dependent vertical response $w_0(L/2, t)$ (with units of mm) of hinged-hinged beams due to a suddenly applied transversely distributed load $q(t)$. Results are for both viscoelastic as well as elastic beams.

We consider a beam with hinged boundary conditions at both ends. The beam length, width and thickness are taken to be $L = 10$ m, $b = 2$ m and $h = 0.5$ m respectively. We consider two loading scenarios. In loading scenario (1) a uniformly distributed transverse load is specified along the entire length of the beam as $q(t) = q_0 H(t)$ N/m, where $q_0 = 10$. Likewise, for loading scenario (2) a periodic concentrated

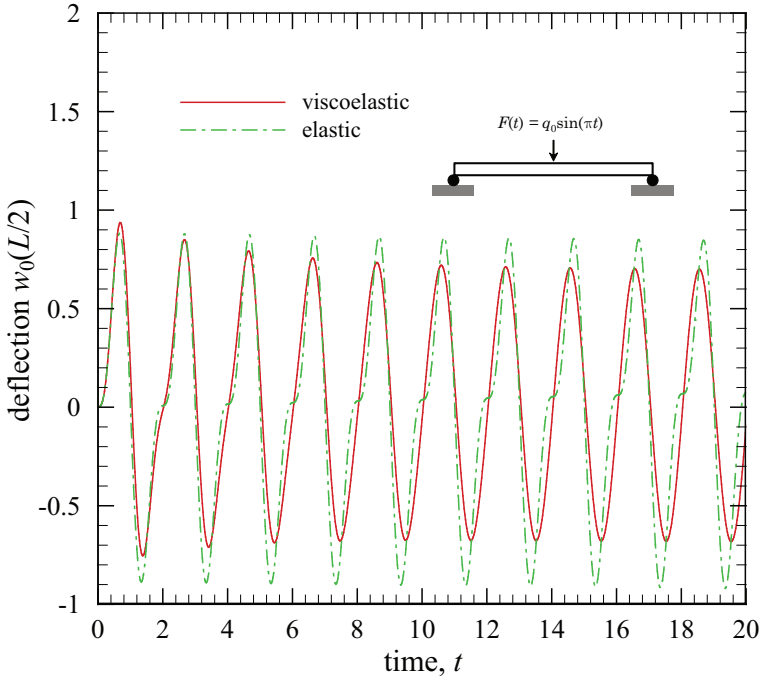


Fig. 52. A comparison of the time-dependent vertical response $w_0(L/2, t)$ (with units of mm) of hinged-hinged viscoelastic and elastic beams due to a periodic concentrated load $F(t)$ (where $\eta = 2.744 \times 10^8$ N-sec./m^2).

force is applied at the center of the beam as $F(t) = q_0 \sin(\pi t)$ N, where $q_0 = 50$. In the finite element discretization of both problems, we employ two RBT-6 elements of equal size. As in the previous examples, symmetry is once again exploited in construction of the finite element meshes. We utilize the Newmark-β procedure [109] for performing temporal discretization of the inertia terms appearing in the fully transient beam finite element equations. The Newmark integration parameters are chosen in accordance with the constant-average acceleration method [24]. Both transient problems are solved over a total time interval of 20 sec. For loading scenario (1) we employ 500 time steps, while 1,000 time steps are utilized for loading scenario (2).

Numerical results for loading scenarios (1) and (2) are presented in Figures 51

and 52 respectively. In each figure we present both viscoelastic and elastic solutions (where the Young's modulus is obtained in the elastic case by taking $E_{\text{elastic}} = E(0)$). As expected, the viscoelastic effects tend to add damping to what would otherwise be purely elastic response. In Figure 51 we present fully transient viscoelastic results using two different values for η. This problem is also solved using the quasi-static viscoelasticity solution procedure. It is evident that the transient viscoelastic solution approaches the steady-state quasi-static viscoelastic response once a sufficiently long enough period of time has transpired. For loading scenario (2) the viscoelastic material properties clearly reduce the overall amplitude of the forced vibrational response. An overall smoothing of the beam response is also observed for this problem.

CHAPTER VI

A NONLINEAR SHELL FINITE ELEMENT FORMULATION FOR ISOTROPIC, LAMINATED COMPOSITE AND FUNCTIONALLY GRADED SHELL STRUCTURES

In this chapter we present a weak-form Galerkin finite element formulation for the general analysis of elastic shear-deformable shell structures using an improved first-order shell theory with seven independent parameters. As shells constitute some of the most prevalent and significant of structural components employed in engineering design, efficient and accurate procedures for their numerical simulation are of great importance. Robust algorithms are particularly crucial, as shells are widely recognized as the prima donnas of structures [114] due to the fact that small changes in geometry and loading can culminate in large changes in the mechanical response.

In the finite element analysis of shells there are primarily four categories of element types that may be adopted: (a) facet-shell elements, (b) 3-D elasticity elements or layerwise theory elements, (c) continuum shell elements (or degenerated shell elements) and (d) 2-D shell theory elements. The facet-elements were developed during the emergence of the finite element method and consist of planar elasticity elements with additional plate-like bending analysis capabilities. These simple elements are still available in many of the commercial finite element software. The 3-D elasticity elements are of course the most general elements; however, their use becomes prohibitively expensive whenever thin and/or multi-layered composite shell structures are to be analyzed. For the case of composite shells, finite elements based on a layerwise theory offer a less expensive computational procedure as compared with 3-D elasticity elements since aspect ratio requirements are dictated by the mid-surface mesh only (see for example Reddy [115, 116] and Robbins and Reddy [117, 118]).

The majority of recent advances in the finite element analysis of shells have come in the areas of continuum shell elements and 2-D shell theory elements. Continuum shell elements were initially introduced by Ahmad et al. [119] and are constructed by mapping a two-dimensional master element onto a surface in $\mathbb{R}^3$ constituting the mid-plane of the element. An isoparametric approach is typically adopted in characterizing the *approximate* mid-plane of the shell element as well as approximating the displacement field. The approximate three-dimensional geometry of the shell element is usually recovered in the continuum approach by prescribing a unit normal at each node that is interpolated using the standard basis functions of the element. The formulation is completed by imposing appropriate kinematic assumptions on the displacement field. Although no shell theory is explicitly invoked, the resulting formulation may be identified as a shell element with qualities consistent with a first-order shear deformation shell model.

In contrast to the so-called continuum elements, shell elements based on a shell theory, are formulated using an exact analytical description of the undeformed mid-surface of the shell. The shell mid-surface is therefore represented using a 2-D *chart* $\boldsymbol{\phi}(\omega^1, \omega^2)$; i.e., a smooth injective mapping from $\bar{\boldsymbol{\omega}} \subset \mathbb{R}^2$, the closure of the open bounded region $\boldsymbol{\omega}$, onto $\bar{\Omega} \subset \mathbb{R}^3$. The three-dimensional shell geometry is then obtained by the 3-D chart: $\boldsymbol{\Phi}(\omega^1, \omega^2, \omega^3) = \boldsymbol{\phi}(\omega^1, \omega^2) + \omega^3 \mathbf{a}_3(\omega^1, \omega^2)$, where $\mathbf{a}_3$ is the analytic unit normal to the mid-surface, $\omega^3 \in [-h/2, h/2]$ and h is the shell thickness. In shell theory the fundamental kinematic assumptions, stress and strain measures, constitutive model, virtual work statement and governing equations are all expressed with respect to the general curvilinear coordinates $(\omega^1, \omega^2, \omega^3)$ used to characterize the three-dimensional shell geometry. Furthermore, the finite element mesh is constructed on $\bar{\boldsymbol{\omega}}$ as opposed to $\bar{\Omega}$. Examples of finite element models for shells constructed directly from shell theories can be found in Refs. [120, 121, 122, 123] and

more recently in the work of Arciniega and Reddy [124, 125]. To the casual observer, the shell theory and continuum based shell finite elements appear quite different from each other. As discussed by Büechter and Ramm, however, the formulations are actually quite similar and when based on the same mechanical assumptions "differ only in the kind of discretization" [126].

The underlying kinematic assumptions adopted in the vast majority of continuum and shell theory based shell finite element models require either 5 or 6 independent variables in the characterization of the displacement field. In 5-parameter models, thickness changes are neglected and as a result, the plane-stress condition must be invoked [127]. In addition a rotation tensor is typically introduced in finite rotation implementations to exactly enforce the inextensibility condition [120, 108]. The rotation tensor may be parametrized by means of rotational degrees of freedom; however, depending on the adopted parametrization, singularities and other rank-deficiencies can arise (see for example Betsch et al. [128]). 6-parameter formulations, on the other hand, may be employed in conjunction with fully three-dimensional constitutive equations; however, such implementations are unfortunately hindered by an erroneous state of constant normal strain through the thickness, a phenomena known as Poisson locking [129]. It is crucial to note that this form of locking is an artifact of the mathematical model and not the discrete finite element implementation.

In recent years there has been significant attention devoted to shell finite element formulations that may be employed with unmodified fully three-dimensional constitutive equations. Motivation for these models stems from the desire to circumvent many of the problems associated with the incorporation of the plane-stress assumption. Such formulations account for thickness stretching and provide reasonable representations of all components of the through-thickness stress states of thin and thick shell structures. These models are usually called 7-parameter formula-

tions, as they involve seven independent parameters in the kinematical description. In a 7-parameter model, the transverse displacement is expanded up to a quadratic term, which essentially mitigates Poisson locking when three-dimensional constitutive equations are adopted. Some of the notable works on 7-parameter shell formulations include Sansour [130] and Bischoff and Ramm [129, 131].

It is well-known that low-order finite element implementations for shells suffer from various forms of locking whenever a purely displacement-based formulation is adopted. The locking phenomena occurs on account of inconsistencies that arise in the discrete finite element representation of the membrane and transverse shear energies. In recent years, the issue of locking has been most prominently addressed through the use of low-order finite element technology using Hu-Washizu type mixed variational principles. Among the successful low-order implementations are the assumed strain (see Dvorkin and Bathe [132] and Hinton and Huang [133]) and enhanced strain (see Simo and Rifai [134]) formulations. High-order finite element implementations have also been advocated in recent years as a means of eliminating the locking phenomena completely. Mostly notably, whenever a sufficient degree of p-refinement is employed, highly reliable locking free numerical solutions may be obtained in a purely displacement-based setting. Among the relevant works are the least-squares finite element formulations of Pontaza and Reddy [135, 136] and Moleiro et al. [137, 138] for the linear analysis of plates and the tensor-based (i.e., shell theory based) weak-form Galerkin finite element models of Arciniega and Reddy [124, 125] for the finite deformation analysis of isotropic, laminated composite and functionally graded shells.

In this chapter we present an improved first-order shear deformation continuum shell finite element formulation for use in the analysis of the fully geometrically nonlinear mechanical response of thin and thick isotropic, composite and functionally graded elastic shell structures. We adopt a 7-parameter formulation which naturally

circumvents the need for a rotation tensor in the kinematical description and allows us to use fully three-dimensional constitutive equations in the numerical implementation. Many of the advances in recent years in the area of locking-free shell finite element formulations have been in the context of low-order elements and mixed variational principles. In the present work, however, we utilize high-order spectral/*hp* type quadrilateral finite element technology in a purely displacement-based finite element setting which naturally allows us to obtain: (a) highly accurate approximations of arbitrary shell geometries and (b) reliable numerical results that are completely locking-free. In the computer implementation, the Schur complement method is adopted at the element level to statically condense out all degrees of freedom interior to each element in a given finite element discretization. This procedure vastly improves computer memory requirements in the numerical implementation of the resulting shell element and allows for significant parallelization of the global solver. The use of spectral/*hp* finite element technology provides an efficient mechanism for reducing errors associated with the isoparametric approximation of arbitrary shell geometries. This constitutes an important departure from the tensor based shell finite element formulation proposed previously in the work of Arciniega and Reddy [124, 125], where a chart was employed to insure exact parametrization of the shell mid-surface.

The shell finite element framework presented in this chapter is applicable to the fully geometrically nonlinear analysis of elastic shell structures based on the St. Venant Kirchhoff material model. The formulation requires as input the prescription of the three-dimensional coordinates of the shell mid-surface as well as two sets of directors (one set normal and the other tangent to the mid-surface) at each node in the shell finite element model. Each of these quantities is approximated discretely using the standard spectral/*hp* finite element interpolation functions within a given shell element. The prescribed tangent vector is particularly useful, as it allows for

the simple construction of the local bases associated with the principle orthotropic material directions of each lamina in a given composite. This allows us to freely adopt skewed and/or arbitrarily curved quadrilateral shell elements in actual finite element simulations. We show, through the numerical simulation of carefully chosen non-trivial benchmark problems, that the proposed shell element is insensitive to all forms of numerical locking and severe geometric distortions.

A. Parametrization of the reference configuration of the shell

A shell structure is by definition a solid body with one geometric dimension being significantly smaller than the other two. In the classical theory of shells, this concept is made mathematically precise through the definition of a mid-surface Ω (where Ω, an open bounded set, is the reference or undeformed mid-surface of the shell) imbedded in physical space $\mathbb{R}^3$ and a thickness parameter h. The mid-surface is characterized using either a single or a set of two-dimensional charts from $\mathbb{R}^2$ into $\mathbb{R}^3$ (e.g., in the single chart case $\boldsymbol{\phi} : \boldsymbol{\omega} \to \Omega \subset \mathbb{R}^3$, where $\boldsymbol{\omega} \subset \mathbb{R}^2$).

In this work we immediately dispense with this exact parametrization of Ω and instead introduce an appropriate finite element based approximation of the mid-surface. To this end, we assume that the closure of Ω (denoted by $\bar{\Omega}$) has been approximated by a conforming set of high-order spectral/hp quadrilateral finite elements. We denote the resulting finite element approximation of $\bar{\Omega}$ as $\bar{\Omega}^{hp}$. This isoparametric characterization of the mid-surface leads to the following standard finite element type approximation

$$\underline{\mathbf{X}} = \boldsymbol{\phi}^e(\xi^1, \xi^2) = \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2) \underline{\mathbf{X}}^k \qquad \text{in } \hat{\Omega}^e \tag{6.1}$$

within a given element, where $\underline{\mathbf{X}}$ represents a point on the approximate mid-surface and ψ_k are the two-dimensional spectral/hp basis functions. In the above expression

$\underline{\mathbf{X}}^k$ are the locations in $\mathbb{R}^3$ of the mid-surface nodes of the eth element (note that all finite element nodes reside on $\bar{\Omega}^{hp}$). The element nodal coordinates $\underline{\mathbf{X}}^k$ (as well as all other subsequent nodal quantities) are given with respect to a fixed orthonormal Cartesian coordinate system with basis vectors: $\{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \hat{\mathbf{E}}_3\}$; as a result $\underline{\mathbf{X}}^k = \underline{X}_i^k \hat{\mathbf{E}}_i$ (where Einstein's summation convention is implied over i). The master element $\hat{\Omega}^e$ used in the isoparametric characterization of the approximate element mid-surface $\bar{\Omega}^e$ (i.e., $\boldsymbol{\phi}^e : \hat{\Omega}^e \to \bar{\Omega}^e \subset \bar{\Omega}^{hp}$) is taken as the standard bi-unit square $\hat{\Omega}^e = [-1, +1]^2$. It should be clear that p-refinement offers us an attractive mechanism for reducing errors in the computational model associated with approximating $\bar{\Omega}$ by $\bar{\Omega}^{hp}$.

At each point of the mid-surface $\bar{\Omega}^e$ of a given element we define the vectors

$$\mathbf{a}_\alpha = \frac{\partial \underline{\mathbf{X}}}{\partial \xi^\alpha} \equiv \underline{\mathbf{X}}_{,\alpha} \tag{6.2}$$

which are linearly independent and thus form a local basis of the tangent plane. We follow the customary convention and allow Greek indices to range over 1 and 2 and Latin indices over 1, 2 and 3. The unit normal vector may be defined as

$$\mathbf{a}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{||\mathbf{a}_1 \times \mathbf{a}_2||} \tag{6.3}$$

We see that for each $(\xi^1, \xi^2) \in \hat{\Omega}^e$, the vectors $\mathbf{a}_i$ define a basis for $\mathbb{R}^3$. In the current work, we will be largely unconcerned with $\mathbf{a}_3$ and instead utilize a finite element approximation of the unit normal defined within a given element as

$$\hat{\mathbf{n}} = \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2) \hat{\mathbf{n}}^k \tag{6.4}$$

The present formulation, therefore, requires as input the mid-surface locations $\underline{\mathbf{X}}$ and the unit normals $\hat{\mathbf{n}}$, both evaluated at the finite element nodes.

We are now in a position to characterize the three-dimensional geometry of the undeformed configuration of a typical shell element $\bar{\mathcal{B}}_0^e$ and as a consequence $\bar{\mathcal{B}}_0^{hp}$ (i.e.,

the finite element approximation of the three-dimensional undeformed shell configuration $\bar{\mathcal{B}}_0$). Assuming a constant thickness h (not to be confused with the mesh parameter) we define the position vector in the shell element as

$$\mathbf{X} = \mathbf{\Phi}^e(\xi^1, \xi^2, \xi^3) = \boldsymbol{\phi}^e(\xi^1, \xi^2) + \xi^3 \frac{h}{2}\hat{\mathbf{n}} = \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2)\left(\underline{\mathbf{X}}^k + \xi^3 \frac{h}{2}\hat{\mathbf{n}}^k\right) \tag{6.5}$$

where $\xi^3 \in [-1, +1]$. The process of parametrizing $\bar{\mathcal{B}}_0^e$ is summarized in Figure 53.

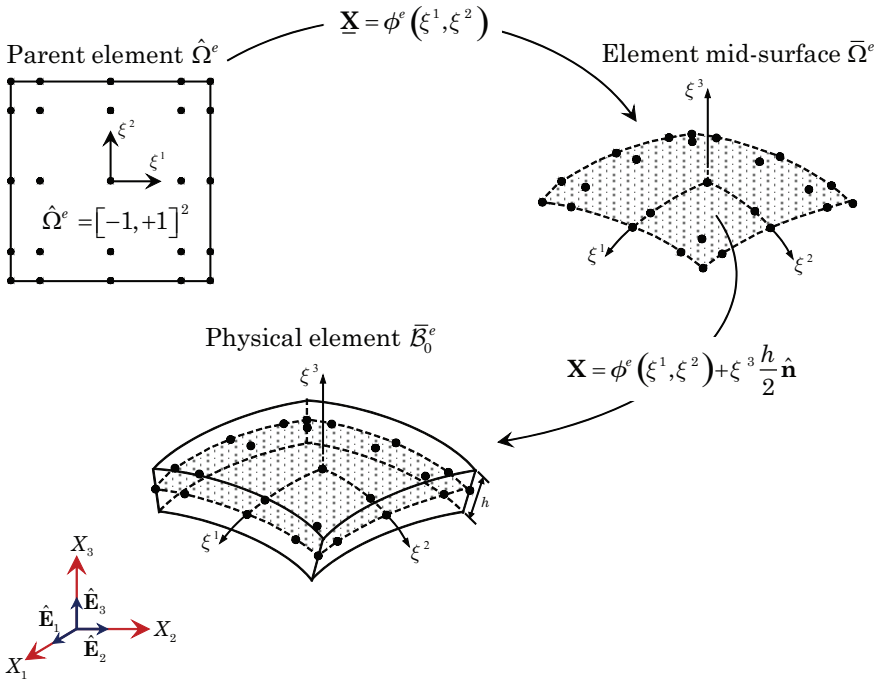


Fig. 53. The process of approximating the three-dimensional geometry of a shell element in the reference configuration based on a isoparametric map from the parent element to the finite element approximation of the mid-surface followed by an additional map to account for the shell thickness.

At each point of the shell element $\bar{\mathcal{B}}_0^e$ (not necessary on the mid-surface $\bar{\Omega}^e$) we define a set of covariant basis vectors

$$\mathbf{g}_i = \frac{\partial \mathbf{X}}{\partial \xi^i} \equiv \mathbf{X}_{,i} \tag{6.6}$$

Using Eq. (6.5) allows us to express the shell basis vectors as

$$\mathbf{g}_\alpha = \mathbf{a}_\alpha + \xi^3 \frac{h}{2}\hat{\mathbf{n}}_{,\alpha}, \qquad \mathbf{g}_3 = \frac{h}{2}\hat{\mathbf{n}} \tag{6.7}$$

In Figure 54 we provide an illustration of the vectors $\mathbf{a}_\alpha$ and $\mathbf{g}_\alpha$ at points A and B respectively, in a typical shell element $\bar{\mathcal{B}}_0^e$. Note that A resides on the mid-surface $\bar{\Omega}^e$, while B lies directly above A in the direction of the unit normal $\hat{\mathbf{n}}$.

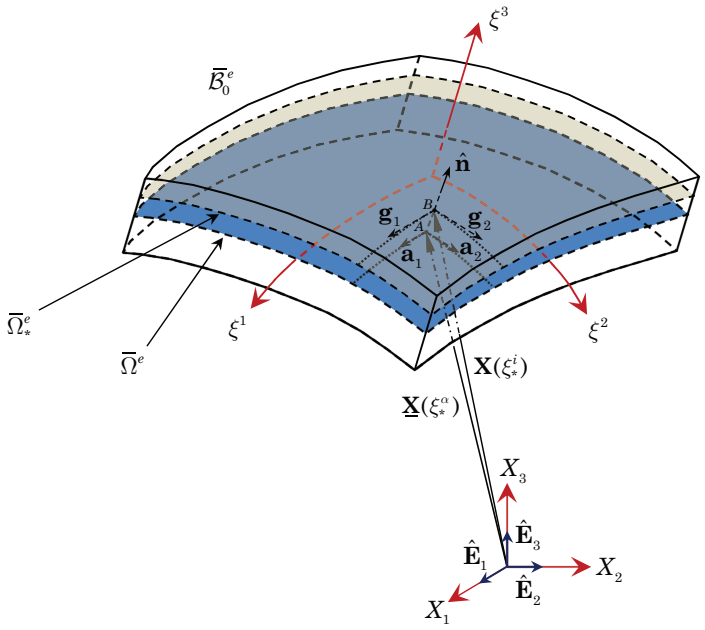


Fig. 54. Geometry of a typical shell finite element $\bar{\mathcal{B}}_0^e$ in the reference configuration. The basis vectors $\mathbf{a}_\alpha$ and $\mathbf{g}_\alpha$ as well as the finite element representation of the unit normal $\hat{\mathbf{n}}$ are also shown.

The covariant basis vectors $\mathbf{g}_i$ allow us to write a differential line element in $\bar{\mathcal{B}}_0^e$ in terms of the curvilinear coordinates (ξ^1, ξ^2, ξ^3) as

$$d\mathbf{X} = d\mathbf{X}_1 + d\mathbf{X}_2 + d\mathbf{X}_3 = \mathbf{g}_1 d\xi^1 + \mathbf{g}_2 d\xi^2 + \mathbf{g}_3 d\xi^3 \tag{6.8}$$

which can be expressed in matrix form as

$$\begin{Bmatrix} dX_1 \\ dX_2 \\ dX_3 \end{Bmatrix} = \{d\xi\}^{\mathrm{T}}[J] = \begin{Bmatrix} d\xi^1 \\ d\xi^2 \\ d\xi^3 \end{Bmatrix}^{\mathrm{T}} \begin{bmatrix} \frac{\partial X_1}{\partial \xi^1} & \frac{\partial X_2}{\partial \xi^1} & \frac{\partial X_3}{\partial \xi^1} \\ \frac{\partial X_1}{\partial \xi^2} & \frac{\partial X_2}{\partial \xi^2} & \frac{\partial X_3}{\partial \xi^2} \\ \frac{\partial X_1}{\partial \xi^3} & \frac{\partial X_2}{\partial \xi^3} & \frac{\partial X_3}{\partial \xi^3} \end{bmatrix} \tag{6.9}$$

The quantity $[J]$ is the Jacobian matrix, which is always invertible. The inverse of the Jacobian matrix is for our purposes denoted as $[J^\star]$. Likewise, the determinant of $[J]$ is simply referred to as J. It is easy to show that a differential volume element in $\bar{\mathcal{B}}_0^e$ is given as

$$d\bar{\mathcal{B}}_0^e = d\mathbf{X}_1 \cdot (d\mathbf{X}_2 \times d\mathbf{X}_3) = J d\xi^1 d\xi^2 d\xi^3 \tag{6.10}$$

We associate with the covariant basis, a dual or contravariant set of basis vectors $\mathbf{g}^i$ defined by the relation

$$\mathbf{g}^i \cdot \mathbf{g}_j = \delta^i_j \tag{6.11}$$

where δ^i_j is the Kronecker delta. The contravariant basis vectors may also be determined from the following formulas

$$\mathbf{g}^1 = \frac{\mathbf{g}_2 \times \mathbf{g}_3}{J}, \qquad \mathbf{g}^2 = \frac{\mathbf{g}_3 \times \mathbf{g}_1}{J}, \qquad \mathbf{g}^3 = \frac{\mathbf{g}_1 \times \mathbf{g}_2}{J} \tag{6.12}$$

The covariant and contravariant basis vectors may be alternatively defined in terms of the components of Jacobian matrix and its inverse

$$\mathbf{g}_i = J_{ij}\hat{\mathbf{E}}_j, \qquad \mathbf{g}^i = J^\star_{ji}\hat{\mathbf{E}}_j \tag{6.13}$$

For completeness, we provide the following formulas for evaluating the components of $[J]$

$$J_{11} = \frac{\partial X_1}{\partial \xi^1} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^1} \left(\underline{X}_1^k + \xi^3 \frac{h}{2} \hat{n}_1^k \right) \tag{6.14a}$$

$$J_{12} = \frac{\partial X_2}{\partial \xi^1} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^1} \left(\underline{X}_2^k + \xi^3 \frac{h}{2} \hat{n}_2^k \right) \tag{6.14b}$$

$$J_{13} = \frac{\partial X_3}{\partial \xi^1} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^1} \left(\underline{X}_3^k + \xi^3 \frac{h}{2} \hat{n}_3^k \right) \tag{6.14c}$$

$$J_{21} = \frac{\partial X_1}{\partial \xi^2} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^2} \left(\underline{X}_1^k + \xi^3 \frac{h}{2} \hat{n}_1^k \right) \tag{6.14d}$$

$$J_{22} = \frac{\partial X_2}{\partial \xi^2} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^2} \left(\underline{X}_2^k + \xi^3 \frac{h}{2} \hat{n}_2^k \right) \tag{6.14e}$$

$$J_{23} = \frac{\partial X_3}{\partial \xi^2} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^2} \left(\underline{X}_3^k + \xi^3 \frac{h}{2} \hat{n}_3^k \right) \tag{6.14f}$$

$$J_{31} = \frac{\partial X_1}{\partial \xi^3} = \frac{h}{2} \sum_{k=1}^{n} \psi_k \hat{n}_1^k \tag{6.14g}$$

$$J_{32} = \frac{\partial X_2}{\partial \xi^3} = \frac{h}{2} \sum_{k=1}^{n} \psi_k \hat{n}_2^k \tag{6.14h}$$

$$J_{33} = \frac{\partial X_3}{\partial \xi^3} = \frac{h}{2} \sum_{k=1}^{n} \psi_k \hat{n}_3^k \tag{6.14i}$$

Likewise, the components of $[J^\star]$ may be determined as

$$J_{11}^\star = (J_{22}J_{33} - J_{23}J_{32})/J \tag{6.15a}$$

$$J_{12}^\star = (J_{13}J_{32} - J_{12}J_{33})/J \tag{6.15b}$$

$$J_{13}^\star = (J_{12}J_{23} - J_{13}J_{22})/J \tag{6.15c}$$

$$J_{21}^\star = (J_{23}J_{31} - J_{21}J_{33})/J \tag{6.15d}$$

$$J_{22}^\star = (J_{11}J_{33} - J_{13}J_{31})/J \tag{6.15e}$$

$$J_{23}^\star = (J_{13}J_{21} - J_{11}J_{23})/J \tag{6.15f}$$

$$J_{31}^\star = (J_{21}J_{32} - J_{22}J_{31})/J \tag{6.15g}$$

$$J_{32}^\star = (J_{12}J_{31} - J_{11}J_{32})/J \tag{6.15h}$$

$$J_{33}^\star = (J_{11}J_{22} - J_{12}J_{21})/J \tag{6.15i}$$

where J is of the form

$$J = J_{11}(J_{22}J_{33} - J_{23}J_{32}) - J_{12}(J_{21}J_{33} - J_{23}J_{31}) + J_{13}(J_{21}J_{32} - J_{22}J_{31}) \tag{6.16}$$

B. The displacement field and strain measures

We now consider the motion $\boldsymbol{\chi}(\mathbf{X}, t)$ of the shell from the reference finite element configuration $\bar{\mathcal{B}}_0^{hp}$ to the current or spatial finite element configuration $\bar{\mathcal{B}}_t^{hp}$. To this end we recall that the displacement of a material point from the reference configuration to the current configuration may be expressed in the usual manner as

$$\mathbf{u}(\mathbf{X}, t) = \boldsymbol{\chi}(\mathbf{X}, t) - \mathbf{X} = \mathbf{x}(\mathbf{X}, t) - \mathbf{X} \tag{6.17}$$

We next assume that at any point within a typical shell element $\bar{\mathcal{B}}_0^e$, the displacement vector may be approximated by a Taylor series expansion with respect to the thickness curvilinear coordinate ξ^3

$$\mathbf{u}(\mathbf{X}(\xi^i), t) \cong \mathbf{u}^{(0)}(\xi^\alpha, t) + \xi^3\mathbf{u}^{(1)}(\xi^\alpha, t) + \frac{(\xi^3)^2}{2}\mathbf{u}^{(2)}(\xi^\alpha, t) + \cdots \tag{6.18}$$

where $\mathbf{u}^{(j)}(\xi^\alpha, t) = \partial^j \mathbf{u}(\xi^i, t)/\partial(\xi^3)^j|_{\xi^3=0}$.

We wish to truncate the Taylor series approximation for $\mathbf{u}$ such that the resulting shell model is asymptotically consistent with three-dimensional solid mechanics [127]; thereby allowing for the use of fully three-dimensional constitutive equations in the mathematical model and subsequent numerical implementation. We therefore restrict the displacement field to the following seven-parameter expansion

$$\mathbf{u}(\xi^i) = \underline{\mathbf{u}}(\xi^\alpha) + \xi^3\frac{h}{2}\boldsymbol{\varphi}(\xi^\alpha) + (\xi^3)^2\frac{h}{2}\boldsymbol{\psi}(\xi^\alpha) \tag{6.19}$$

where each $\mathbf{u}^{(j)}(\xi^\alpha, t)$ ($j = 0, 1$ and 2) has been renamed, and for $j = 1$ and 2, scaled by some factor of h. For the sake of brevity, we omit the time parameter t from the

above expressions and the subsequent discussion. The generalized displacements $\underline{\mathbf{u}}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$ may be expressed as

$$\underline{\mathbf{u}}(\xi^\alpha) = \underline{u}_i(\xi^\alpha)\hat{\mathbf{E}}_i, \qquad \boldsymbol{\varphi}(\xi^\alpha) = \varphi_i(\xi^\alpha)\hat{\mathbf{E}}_i, \qquad \boldsymbol{\psi}(\xi^\alpha) = \Psi(\xi^\alpha)\hat{\mathbf{n}}(\xi^\alpha) \tag{6.20}$$

The quantity $\underline{\mathbf{u}}$ represents the mid-plane displacement and $\boldsymbol{\varphi}$ is the so-called difference vector (which gives the change in the mid-surface director). The seventh parameter Ψ is included to circumvent spurious stresses in the thickness direction, caused in the six-parameter formulation by an artificial constant normal strain (a phenomena referred to as Poisson locking [129]).

The position occupied by a material point belonging to $\bar{\mathcal{B}}_0^e$ at the current time t may be evaluated by substituting the assumed displacement field into Eq. (6.17) which upon rearrangement yields

$$\mathbf{x} = \mathbf{X} + \mathbf{u} = \underline{\mathbf{x}} + \xi^3\frac{h}{2}\hat{\bar{\mathbf{n}}} + (\xi^3)^2\frac{h}{2}\Psi\hat{\mathbf{n}} \tag{6.21}$$

where $\underline{\mathbf{x}} = \underline{\mathbf{X}} + \underline{\mathbf{u}}$ (a point on the deformed mid-surface) and $\hat{\bar{\mathbf{n}}} = \hat{\mathbf{n}} + \boldsymbol{\varphi}$ (a *pseudo*-director associated with the deformed mid-surface). It is important to note that unlike $\hat{\mathbf{n}}$; the director $\hat{\bar{\mathbf{n}}}$ is in general neither a unit vector nor is it normal to the deformed mid-surface.

We define the finite element approximation of the displacement field given by Eq. (6.19) as

$$\mathbf{u}(\xi^i) = \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2)\left(\underline{\mathbf{u}}^k + \xi^3\frac{h}{2}\boldsymbol{\varphi}^k + (\xi^3)^2\frac{h}{2}\Psi^k\hat{\mathbf{n}}(\xi^\alpha)\right) \tag{6.22}$$

where $\hat{\mathbf{n}}(\xi^\alpha)$ is given by Eq. (6.4). Note that we interpolate Ψ and $\hat{\mathbf{n}}$ separately in the finite element approximation of $\boldsymbol{\psi}$ (as opposed to interpolating the product $\Psi\hat{\mathbf{n}}$ as a single entity). The derivative of the displacement field with respect to the curvilinear

coordinates of the element may be expressed as

$$\mathbf{u}_{,\alpha} = \sum_{k=1}^{n} \frac{\partial \psi_k}{\partial \xi^\alpha} \left[\underline{\mathbf{u}}^k + \xi^3 \frac{h}{2} \boldsymbol{\varphi}^k + (\xi^3)^2 \frac{h}{2} \left(\Psi^k \hat{\mathbf{n}}(\xi^\beta) + \hat{\mathbf{n}}^k \Psi(\xi^\beta) \right) \right] \tag{6.23a}$$

$$\mathbf{u}_{,3} = h \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2) \left(\frac{1}{2} \boldsymbol{\varphi}^k + \xi^3 \Psi^k \hat{\mathbf{n}}(\xi^\beta) \right) \tag{6.23b}$$

We recall from continuum mechanics that the deformation gradient $\mathbf{F}$ may be defined as

$$\mathbf{F} = (\nabla_0 \mathbf{x})^{\mathrm{T}} = \mathbf{x}_{,i} \mathbf{g}^i = \bar{\mathbf{g}}_i \mathbf{g}^i \tag{6.24}$$

where $\bar{\mathbf{g}}_i = \mathbf{g}_i + \mathbf{u}_{,i}$ are the covariant basis vectors associated with the deformed finite element configuration of the three-dimensional shell $\bar{\mathcal{B}}_t^{hp}$. The nabla symbol ∇_0 is the material gradient operator. It should be clear that $\mathbf{F}$ is a two-point tensor relating differential line segments in the material configuration to their associated differential line segments in the deformed configuration.

We next define the Green-Lagrange strain tensor $\mathbf{E}$ as

$$\begin{aligned} \mathbf{E} &= \frac{1}{2} \Big(\mathbf{F}^{\mathrm{T}} \cdot \mathbf{F} - \mathbf{I} \Big) \\ &= \frac{1}{2} \Big(\mathbf{u}_{,i} \cdot \mathbf{g}_j + \mathbf{g}_i \cdot \mathbf{u}_{,j} + \mathbf{u}_{,i} \cdot \mathbf{u}_{,j} \Big) \mathbf{g}^i \mathbf{g}^j \end{aligned} \tag{6.25}$$

which relates the difference in the squares of differential spatial and reference configuration line segments as

$$(ds)^2 - (dS)^2 = d\mathbf{x} \cdot d\mathbf{x} - d\mathbf{X} \cdot d\mathbf{X} = 2 d\mathbf{X} \cdot \mathbf{E} \cdot d\mathbf{X} \tag{6.26}$$

The covariant components of the Green-Lagrange strain tensor (i.e., the coefficients of the second order tensor contravariant bases $\mathbf{g}^i \mathbf{g}^j$ appearing in Eq. (6.25)) may be expanded in terms of the thickness coordinate ξ^3 as

$$E_{ij}(\xi^m) = \varepsilon_{ij}^{(0)} + \xi^3 \varepsilon_{ij}^{(1)} + \underline{(\xi^3)^2 \varepsilon_{ij}^{(2)} + (\xi^3)^3 \varepsilon_{ij}^{(3)} + (\xi^3)^4 \varepsilon_{ij}^{(4)}} \tag{6.27}$$

where $\varepsilon_{ij}^{(n)} = \varepsilon_{ij}^{(n)}(\xi^\alpha)$. In the present formulation we neglect all covariant components of $\mathbf{E}$ that are of higher order than linear in ξ^3. The retained covariant components may be determined as

$$\varepsilon_{\alpha\beta}^{(0)} = \frac{1}{2}\left(\underline{\mathbf{u}}_{,\alpha}\cdot\mathbf{a}_\beta + \mathbf{a}_\alpha\cdot\underline{\mathbf{u}}_{,\beta} + \underline{\mathbf{u}}_{,\alpha}\cdot\underline{\mathbf{u}}_{,\beta}\right) \tag{6.28a}$$

$$\varepsilon_{\alpha\beta}^{(1)} = \frac{h}{4}\left[\underline{\mathbf{u}}_{,\alpha}\cdot\left(\hat{\mathbf{n}}_{,\beta}+\boldsymbol{\varphi}_{,\beta}\right) + \left(\hat{\mathbf{n}}_{,\alpha}+\boldsymbol{\varphi}_{,\alpha}\right)\cdot\underline{\mathbf{u}}_{,\beta} + \boldsymbol{\varphi}_{,\alpha}\cdot\mathbf{a}_\beta + \mathbf{a}_\alpha\cdot\boldsymbol{\varphi}_{,\beta}\right] \tag{6.28b}$$

$$\varepsilon_{\alpha 3}^{(0)} = \frac{h}{4}\left[\underline{\mathbf{u}}_{,\alpha}\cdot\left(\hat{\mathbf{n}}+\boldsymbol{\varphi}\right) + \mathbf{a}_\alpha\cdot\boldsymbol{\varphi}\right] \tag{6.28c}$$

$$\varepsilon_{\alpha 3}^{(1)} = \frac{h}{2}\left\{\frac{h}{4}\left[\boldsymbol{\varphi}_{,\alpha}\cdot\hat{\mathbf{n}} + \left(\hat{\mathbf{n}}_{,\alpha}+\boldsymbol{\varphi}_{,\alpha}\right)\cdot\boldsymbol{\varphi}\right] + \left(\mathbf{a}_\alpha + \underline{\mathbf{u}}_{,\alpha}\right)\cdot\boldsymbol{\psi}\right\} \tag{6.28d}$$

$$\varepsilon_{33}^{(0)} = \frac{h^2}{8}\left(2\hat{\mathbf{n}}+\boldsymbol{\varphi}\right)\cdot\boldsymbol{\varphi} \tag{6.28e}$$

$$\varepsilon_{33}^{(1)} = \frac{h^2}{2}\left(\hat{\mathbf{n}}+\boldsymbol{\varphi}\right)\cdot\boldsymbol{\psi} \tag{6.28f}$$

We see from the above expressions that in the six-parameter formulation (obtained by taking $\boldsymbol{\psi} = \mathbf{0}$) the strain component $\varepsilon_{33}^{(1)}$ is identically zero.

C. Constitutive equations

The underlying kinematic assumptions of the adopted shell finite element formulation can be applied in the context of a multitude of material models (e.g., Cauchy elastic, hyperelastic, viscoelastic, elasto-plastic, etc.). In this work we assume that the material response remains in the elastic regime. Furthermore, we assume that the second Piola Kirchhoff stress tensor $\mathbf{S}$ is related to the Green-Lagrange strain tensor $\mathbf{E}$ by the following relation

$$\mathbf{S} = \mathbb{C} : \mathbf{E} \tag{6.29}$$

where $\mathbb{C} = \mathbb{C}^{ijkl}\mathbf{g}_i\mathbf{g}_j\mathbf{g}_k\mathbf{g}_l$ is the fourth-order elasticity tensor. We require the elasticity tensor to be independent of the shell deformation. However, we do allow $\mathbb{C}$ to be non-

homogeneous (i.e., a function of $\mathbf{X}$). This frame-indifferent hyperelastic constitutive model is often called linear elastic (not to be confused with the theory of linear elasticity). In the numerical implementation, we rely on the following component representation of the set of constitutive equations

$$S^{ij} = \mathbb{C}^{ijkl} E_{kl} \tag{6.30}$$

The adopted material model may also be expressed in matrix form as

$$\begin{Bmatrix} S^{11} \\ S^{22} \\ S^{33} \\ S^{23} \\ S^{13} \\ S^{12} \end{Bmatrix} = \begin{bmatrix} \mathbb{C}^{1111} & \mathbb{C}^{1122} & \mathbb{C}^{1133} & \mathbb{C}^{1123} & \mathbb{C}^{1113} & \mathbb{C}^{1112} \\ \mathbb{C}^{1122} & \mathbb{C}^{2222} & \mathbb{C}^{2233} & \mathbb{C}^{2223} & \mathbb{C}^{2213} & \mathbb{C}^{2212} \\ \mathbb{C}^{1133} & \mathbb{C}^{2233} & \mathbb{C}^{3333} & \mathbb{C}^{3323} & \mathbb{C}^{3313} & \mathbb{C}^{3312} \\ \mathbb{C}^{1123} & \mathbb{C}^{2223} & \mathbb{C}^{3323} & \mathbb{C}^{2323} & \mathbb{C}^{2313} & \mathbb{C}^{2312} \\ \mathbb{C}^{1113} & \mathbb{C}^{2213} & \mathbb{C}^{3313} & \mathbb{C}^{2313} & \mathbb{C}^{1313} & \mathbb{C}^{1312} \\ \mathbb{C}^{1112} & \mathbb{C}^{2212} & \mathbb{C}^{3312} & \mathbb{C}^{2312} & \mathbb{C}^{1312} & \mathbb{C}^{1212} \end{bmatrix} \begin{Bmatrix} E_{11} \\ E_{22} \\ E_{33} \\ 2E_{23} \\ 2E_{13} \\ 2E_{12} \end{Bmatrix} \tag{6.31}$$

where the coefficient matrix $[\mathbb{C}^{ijkl}]$ appearing in the above expression is the matrix form of the contravariant components of the elasticity tensor $\mathbb{C}$. It should be evident that there are in general 21 unique contravariant components of $\mathbb{C}$.

For completeness, we recall that the second Piola Kirchhoff stress tensor $\mathbf{S}$ is defined as

$$\mathbf{S} = J_{\mathbf{F}} \mathbf{F}^{-1} \cdot \sigma \cdot \mathbf{F}^{-\mathrm{T}} \tag{6.32}$$

where σ is the *true* or Cauchy stress tensor and $J_{\mathbf{F}} = \det\mathbf{F}$. The symmetry of $\mathbf{S}$ follows from the symmetry of σ.

1. Isotropic shells: homogeneous and functionally graded

We now specialize the assumed constitutive model for use in the context of isotropic shells. We consider the homogeneous case and also the scenario where the material

is functionally graded through the thickness of the shell. Homogeneous shells are abundant and can be found in piping, pressure vessels, ship hulls, large roofs and the bodies of automobiles. Functionally graded shells on the other hand have been advocated for use in high temperature environments with applications in reactor vessels, turbines and other machine parts [116]. These materials are typically composed of metals and ceramics to maximize the strength and toughness properties of the former and the thermal and corrosion resistance attributes of the latter.

For isotropic materials, the fourth-order elasticity tensor may be expressed as

$$\mathbb{C} = \lambda \mathbf{I}\mathbf{I} + 2\mu \mathbb{I} \tag{6.33}$$

The Lamé parameters λ and μ are related to the Young's modulus E and Poisson's ratio ν by the following expressions

$$\lambda = \frac{\nu E}{(1+\nu)(1-2\nu)} \tag{6.34a}$$

$$\mu = \frac{E}{2(1+\nu)} \tag{6.34b}$$

The quantities $\mathbf{I} = \delta_{ij}\hat{\mathbf{E}}_i\hat{\mathbf{E}}_j$ and $\mathbb{I} = \frac{1}{2}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\hat{\mathbf{E}}_i\hat{\mathbf{E}}_j\hat{\mathbf{E}}_k\hat{\mathbf{E}}_l$ are the second and fourth-order identity tensors respectively. These tensors may also be expressed with respect to the covariant basis vectors $\mathbf{g}_i$ as

$$\mathbf{I} = \mathbf{G} = g^{ij}\mathbf{g}_i\mathbf{g}_j \tag{6.35a}$$

$$\mathbb{I} = \frac{1}{2}(g^{ik}g^{jl} + g^{il}g^{jk})\mathbf{g}_i\mathbf{g}_j\mathbf{g}_k\mathbf{g}_l \tag{6.35b}$$

where $g^{ij} = \mathbf{g}^i \cdot \mathbf{g}^j$ are the contravariant components of the *Riemannian metric* tensor $\mathbf{G}$ in the reference configuration. We can therefore express the contravariant components of $\mathbb{C}$ as

$$\mathbb{C}^{ijkl} = \lambda g^{ij}g^{kl} + \mu(g^{ik}g^{jl} + g^{il}g^{jk}) \tag{6.36}$$

Although $\mathbb{C}$ depends on only the Lamé parameters, the 21 contravariant components associated with the matrix $[\mathbb{C}^{ijkl}]$ are in general distinct from one another.

For the homogeneous case, the Young's modulus and Poisson's ratio are constant throughout the shell structure. For functionally graded structures, we assume that the shell is composed of two isotropic constituents. In such cases, we allow the Young's modulus to vary with respect to the shell thickness coordinate ξ^3 as prescribed by the following smooth function

$$E(\xi^3) = (E^+ - E^-) f^+(\xi^3) + E^- \tag{6.37}$$

where

$$f^+(\xi^3) = \left(\frac{\xi^3 + 1}{2} \right)^n \tag{6.38}$$

The quantities E^- and E^+ constitute the Young's moduli at the bottom ($\xi^3 = -1$) and top ($\xi^3 = +1$) surfaces of the shell respectively. Eq. (6.37) constitutes a power-law variation of E through the shell thickness (where the non-negative constant n is the power-law parameter). Note that E^- and E^+ are recovered throughout the thickness in the limits where $n \to \infty$ and $n \to 0$ respectively. As in the homogeneous case, functionally graded shells may also be described using Eq. (6.36) if the Lamé parameters are taken as functions of ξ^3.

2. Laminated composite shells

In this work we are also concerned with the numerical simulation of laminated composite shell structures. A composite laminae is a thin sheet (plate or shell like) of material, typically composed of two distinct constituents, which together possess desirable mechanical properties that cannot be exhibited by the individual materials acting in bulk alone. A laminated composite shell is a collection of stacked laminae

(where the stacking sequence is typically prescribed in a manner which maximizes the desired stiffness of the composite). In our analysis, we treat each laminae as an orthotropic layer of material. We further assume that for a given structure, perfect bonding exists between each layer and that the continuum hypothesis holds.

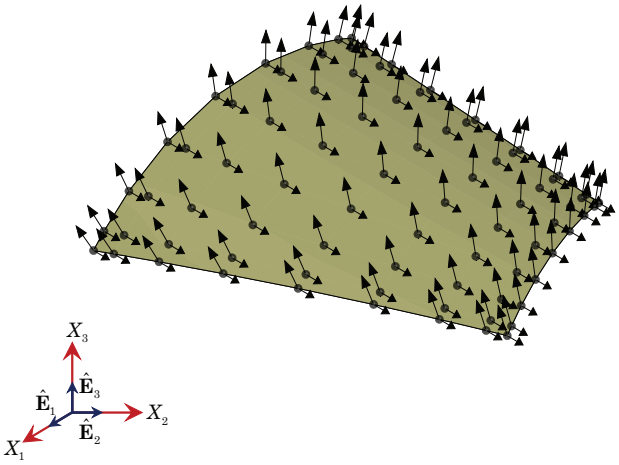


Fig. 55. The mid-surface $\bar{\Omega}^e$ of a typical high-order spectral/hp shell finite element (case shown is for a p-level of 8). The unit normals $\hat{\mathbf{n}}^k$ and tangents $\hat{\mathbf{t}}^k$ are also shown at the element nodes.

To simplify the discussion, we initially restrict our attention to a shell composed of a single orthotopic layer (i.e., one laminae). Next, we define at each node in $\bar{\Omega}^{hp}$ a unit vector $\hat{\mathbf{t}}$ that is tangent to the finite element approximation of the mid-plane. The discrete tangents are utilized to define a continuous tangent vector field in $\bar{\Omega}^{hp}$. Within each element, the tangent field is represented using the following standard interpolation formula

$$\hat{\mathbf{t}} = \sum_{k=1}^{n} \psi_k(\xi^1, \xi^2)\hat{\mathbf{t}}^k \tag{6.39}$$

The tangent vector $\hat{\mathbf{t}}$ is prescribed in a manner that allows us to easily construct a local orthogonal Cartesian basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ associated with the principle directions

of the orthotropic laminae. In Figure 55 we show the geometry, nodes, unit normals $\hat{\mathbf{n}}^k$ and unit tangents $\hat{\mathbf{t}}^k$ for a typical high-order spectral/*hp* shell finite element. Note that the direction of $\hat{\mathbf{t}}$ need not coincide with the direction of either $\hat{\mathbf{a}}_1$ or $\hat{\mathbf{a}}_2$.

We next express the elasticity tensor $\mathbb{C}$ for the shell with respect to the local basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ (which we will soon define) as

$$\mathbb{C} = \bar{\mathbb{C}}_{ijkl}\hat{\mathbf{e}}_i\hat{\mathbf{e}}_j\hat{\mathbf{e}}_k\hat{\mathbf{e}}_l \tag{6.40}$$

Assuming an orthotropic material model allows us to express the coefficients $\bar{\mathbb{C}}_{ijkl}$ in matrix form as

$$[\bar{\mathbb{C}}_{ijkl}] = \begin{bmatrix} \bar{\mathbb{C}}_{1111} & \bar{\mathbb{C}}_{1122} & \bar{\mathbb{C}}_{1133} & 0 & 0 & 0 \\ \bar{\mathbb{C}}_{1122} & \bar{\mathbb{C}}_{2222} & \bar{\mathbb{C}}_{2233} & 0 & 0 & 0 \\ \bar{\mathbb{C}}_{1133} & \bar{\mathbb{C}}_{2233} & \bar{\mathbb{C}}_{3333} & 0 & 0 & 0 \\ 0 & 0 & 0 & \bar{\mathbb{C}}_{2323} & 0 & 0 \\ 0 & 0 & 0 & 0 & \bar{\mathbb{C}}_{1313} & 0 \\ 0 & 0 & 0 & 0 & 0 & \bar{\mathbb{C}}_{1212} \end{bmatrix} \tag{6.41}$$

The components of the coefficient matrix $[\bar{\mathbb{C}}_{ijkl}]$ may be determined in terms of the Engineering parameters: E_1, E_2, E_3, ν_{12}, ν_{13}, ν_{23}, G_{12}, G_{13} and G_{23} as

$$[\bar{\mathbb{C}}_{ijkl}] = \begin{bmatrix} \frac{1}{E_1} & -\frac{\nu_{12}}{E_1} & -\frac{\nu_{13}}{E_1} & 0 & 0 & 0 \\ -\frac{\nu_{12}}{E_1} & \frac{1}{E_2} & -\frac{\nu_{23}}{E_2} & 0 & 0 & 0 \\ -\frac{\nu_{13}}{E_1} & -\frac{\nu_{23}}{E_2} & \frac{1}{E_3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{G_{23}} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{G_{13}} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{G_{12}} \end{bmatrix}^{-1} \tag{6.42}$$

Note that 9 independent material parameters are required to define the orthotropic form of the elasticity tensor $\mathbb{C}$.

We next address construction of the local basis $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ for a typical shell element. Without loss of generality we take $\hat{\mathbf{e}}_3 = \hat{\mathbf{n}}$. Next we assume that $\hat{\mathbf{e}}_1$ may be obtained locally in terms of a proper finite rotation of the tangent $\hat{\mathbf{t}}$ about the unit normal $\hat{\mathbf{n}}$, where the angle of rotation is θ. In this work we always define $\hat{\mathbf{t}}$ such that it is sufficient to take θ as constant in $\bar{\Omega}^e$ and throughout $\bar{\Omega}^{hp}$. Given the preceding assumptions, we can show from geometry that the local basis for the principle directions of the material is given as

$$\hat{\mathbf{e}}_1 = \hat{\mathbf{t}} \cos\theta + \hat{\mathbf{n}} \times \hat{\mathbf{t}} \sin\theta \tag{6.43a}$$

$$\hat{\mathbf{e}}_2 = -\hat{\mathbf{t}} \sin\theta + \hat{\mathbf{n}} \times \hat{\mathbf{t}} \cos\theta \tag{6.43b}$$

$$\hat{\mathbf{e}}_3 = \hat{\mathbf{n}} \tag{6.43c}$$

It should be clear that when $\theta = 0$, the in-plane principle basis vectors of the material reduce to: $\hat{\mathbf{e}}_1 = \hat{\mathbf{t}}$ and $\hat{\mathbf{e}}_2 = \hat{\mathbf{n}} \times \hat{\mathbf{t}}$. It is important to note that the in-plane *material* basis vectors $\hat{\mathbf{e}}_\alpha$ are constructed independent from the in-plane *natural* basis vectors $\hat{\mathbf{a}}_\alpha$. As a result, we may freely employ unstructured skewed and/or curved quadrilateral finite elements in the numerical discretization of complex shell structures. Key to the success of the present formulation is an appropriate prescription of the discrete tangent vector $\hat{\mathbf{t}}$ and angle of rotation θ. In Figure 56 we show the unit normal $\hat{\mathbf{n}}$, unit tangent $\hat{\mathbf{t}}$, rotation angle θ and local material basis vectors $\hat{\mathbf{e}}_i$ at a point on the mid-surface of a typical shell element.

In the numerical implementation we require the contravariant components $\mathbb{C}^{ijkl}$ of the elasticity tensor $\mathbf{\mathbb{C}}$. These may be obtained by contracting Eq. (6.40) with $\mathbf{g}^i \mathbf{g}^j \mathbf{g}^k \mathbf{g}^l$ which yields

$$\mathbb{C}^{ijkl} = T_{im} T_{jn} T_{kp} T_{lq} \bar{\mathbb{C}}_{mnpq} \tag{6.44}$$

where the components of T_{ij} are defined as

$$T_{ij} = \mathbf{g}^i \cdot \hat{\mathbf{e}}_j = J^{\star}_{ki} \bar{E}_{jk} \tag{6.45}$$

and $\bar{E}_{jk} = \hat{\mathbf{e}}_j \cdot \hat{\mathbf{E}}_k$. We see that evaluation of $\mathbb{C}^{ijkl}$ requires 5 matrix multiplications. In the actual numerical implementation, however, we have generated the C++ code for evaluating the 21 independent coefficients in $\mathbb{C}^{ijkl}$ using the symbolic algebra software Maple. The expressions are quite involved and are hence not provided in the text of this dissertation.

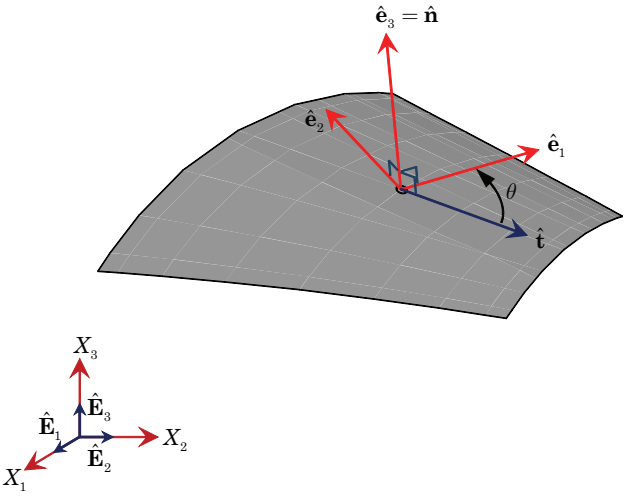


Fig. 56. The unit normal $\hat{\mathbf{n}}$, unit tangent $\hat{\mathbf{t}}$, rotation angle θ and local basis vectors $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \hat{\mathbf{e}}_3\}$ at a point on the mid-surface $\bar{\Omega}^e$ of a typical shell finite element.

The above discussion has been limited to the analysis of composite shells composed of a single orthotropic layer. For multi-layered composites, we define a unit tangent vector $\hat{\mathbf{t}}$ along with a set of orientation angles $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{\text{NL}})$ associated with each ply (where NL is the total number of layers). We number each layer in the laminated composite in order from the bottom ply to the top laminae. Within a given layer (say layer q) we obtain the local material basis vectors $\{\hat{\mathbf{e}}_1^q, \hat{\mathbf{e}}_2^q, \hat{\mathbf{e}}_3^q\}$ using θ_q

in place of θ in Eq. (6.43). Once the local basis vectors are known, we may determine the components $\mathbb{C}^{ijkl}$ throughout the qth ply using Eq. (6.44).

D. Weak formulation and discrete numerical implementation

The finite element model is developed using the standard weak-form Galerkin procedure, which is equivalent to the principle of virtual displacements. We restrict our formulation to static or quasi-static analysis, and therefore omit the inertial terms. The principle of virtual work may be stated as follows: find $\Phi \in \mathcal{V}$ such that for all $\delta\Phi \in \mathcal{W}$ the following weak statement holds

$$\mathcal{G}(\delta\Phi, \Phi) = \delta\mathcal{W}_{\mathrm{I}}(\delta\Phi, \Phi) + \delta\mathcal{W}_{\mathrm{E}}(\delta\Phi, \Phi) \equiv 0 \tag{6.46}$$

The quantities $\delta\mathcal{W}_{\mathrm{I}}$ and $\delta\mathcal{W}_{\mathrm{E}}$ are the internal and external virtual work, respectively. These quantities may be defined with respect to the undeformed configuration as

$$\delta\mathcal{W}_{\mathrm{I}} = \int_{\mathcal{B}_0} \delta\mathbf{E} : \mathbf{S} d\mathcal{B}_0 \tag{6.47a}$$

$$\delta\mathcal{W}_{\mathrm{E}} = -\int_{\mathcal{B}_0} \delta\mathbf{u} \cdot \rho_0 \mathbf{b}_0 d\mathcal{B}_0 - \int_{\Gamma_\sigma} \delta\mathbf{u} \cdot \mathbf{t}_0 ds \tag{6.47b}$$

where ρ_0 is the density, $\mathbf{b}_0$ is the body force and $\mathbf{t}_0$ is the traction vector (which are all expressed with respect to the reference configuration). The function space of admissible configurations $\mathcal{V}$ and linear vector space of admissible variations $\mathcal{W}$ are defined for the continuous problem as

$$\mathcal{V} = \left\{ \ \Phi = (\underline{\mathbf{u}}, \boldsymbol{\varphi}, \Psi) : \qquad \Phi \in \mathbf{H}^1(\Omega) \times \mathbf{H}^1(\Omega) \times H^1(\Omega), \ \ \Phi = \Phi^{\mathrm{p}} \text{ on } \Gamma^{\mathrm{D}} \right\} \tag{6.48a}$$

$$\mathcal{W} = \left\{ \delta\Phi = (\delta\underline{\mathbf{u}}, \delta\boldsymbol{\varphi}, \delta\Psi) : \delta\Phi \in \mathbf{H}^1(\Omega) \times \mathbf{H}^1(\Omega) \times H^1(\Omega), \ \delta\Phi = \mathbf{0} \quad \text{on } \Gamma^{\mathrm{D}} \right\} \tag{6.48b}$$

where Γ^{D} is the part of the boundary on which Φ is specified.

In the numerical implementation, we restrict Φ and $\delta\Phi$ to their appropriate high-

order spectral/hp finite element sub-spaces: $\Phi_{hp} \in \mathcal{V}^{hp}$ and $\delta\Phi_{hp} \in \mathcal{W}^{hp}$. This results in the following discrete variation problem: find $\Phi_{hp} \in \mathcal{V}^{hp}$ such that

$$\mathcal{G}(\delta\Phi_{hp}, \Phi_{hp}) = \delta\mathcal{W}_{\mathrm{I}}(\delta\Phi_{hp}, \Phi_{hp}) + \delta\mathcal{W}_{\mathrm{E}}(\delta\Phi_{hp}, \Phi_{hp}) \equiv 0 \quad \forall\ \delta\Phi_{hp} \in \mathcal{W}^{hp} \tag{6.49}$$

We refer the reader to Chapter II for details on the high-order spectral/hp basis functions. Evaluation of the internal virtual work statement for the eth element of the discrete problem yields

$$\begin{aligned}
\delta\mathcal{W}_I^e &= \int_{\mathcal{B}_0^e} (\delta\varepsilon^{(0)} + \xi^3\delta\varepsilon^{(1)}) : \mathbb{C} : (\varepsilon^{(0)} + \xi^3\varepsilon^{(1)}) d\mathcal{B}_0^e \\
&= \int_{\hat{\Omega}^e}\int_{-1}^{+1} \big(\delta\varepsilon_{ij}^{(0)} + \xi^3\delta\varepsilon_{ij}^{(1)}\big)\mathbb{C}^{ijkl}\big(\varepsilon_{kl}^{(0)} + \xi^3\varepsilon_{kl}^{(1)}\big) J d\xi^3 d\hat{\Omega}^e \\
&= \int_{\hat{\Omega}^e} \big[\mathbb{A}^{ijkl}\delta\varepsilon_{ij}^{(0)}\varepsilon_{kl}^{(0)} + \mathbb{B}^{ijkl}\big(\delta\varepsilon_{ij}^{(0)}\varepsilon_{kl}^{(1)} + \delta\varepsilon_{ij}^{(1)}\varepsilon_{kl}^{(0)}\big) + \mathbb{D}^{ijkl}\delta\varepsilon_{ij}^{(1)}\varepsilon_{kl}^{(1)}\big] d\hat{\Omega}^e
\end{aligned} \tag{6.50}$$

where $\int_{\hat{\Omega}^e}(\,\cdot\,)d\hat{\Omega}^e = \int_{-1}^{+1}\int_{-1}^{+1}(\,\cdot\,)d\xi^1 d\xi^2$. The quantities $\mathbb{A}^{ijkl}$, $\mathbb{B}^{ijkl}$ and $\mathbb{D}^{ijkl}$ are the contravariant components of the effective *extensional*, *extensional-bending coupling* and *bending* fourth-order stiffnesses respectively. The components may be determined as

$$\{\mathbb{A}^{ijkl}, \mathbb{B}^{ijkl}, \mathbb{D}^{ijkl}\} = \int_{-1}^{+1} \{1, \xi^3, (\xi^3)^2\}\mathbb{C}^{ijkl} J d\xi^3 \tag{6.51}$$

It is crucial to note that the stiffness components have been systematically defined such that they include the Jacobian determinant J. In the computer implementation, we perform the above integration numerically using the Gauss-Legendre quadrature rule (with 50 quadrature points taken along the thickness direction of each laminae). Therefore, no thin-shell approximating assumptions are imposed on either J or $\mathbb{C}^{ijkl}$ in the finite element model.

The external virtual work consists of body forces and tractions. For each element, we decompose the boundary of the shell into top $\Gamma_{\sigma,+}^e$, bottom $\Gamma_{\sigma,-}^e$ and lateral $\Gamma_{\sigma,S}^e$ surfaces. As a result, the external virtual work for a typical shell element may be

expressed as

$$\begin{aligned}\delta\mathcal{W}^e_{\mathrm{E}} &= -\int_{\mathcal{B}^e_0}\delta\mathbf{u}\cdot\rho_0\mathbf{b}d\mathcal{B}^e_0 - \int_{\Gamma^e_\sigma}\delta\mathbf{u}\cdot\mathbf{t}_0 ds \\ &= -\int_{\mathcal{B}^e_0}\delta\mathbf{u}\cdot\rho_0\mathbf{b}d\mathcal{B}^e_0 - \int_{\Gamma^e_{\sigma,+}}\delta\mathbf{u}\cdot\mathbf{t}^+_0 ds^+ - \int_{\Gamma^e_{\sigma,-}}\delta\mathbf{u}\cdot\mathbf{t}^-_0 ds^- \\ &\quad - \int_{\Gamma^e_{\sigma,S}}\delta\mathbf{u}\cdot\mathbf{t}^S_0 ds^S\end{aligned} \tag{6.52}$$

The traction boundary conditions along the top and bottom of the shell element may be expressed as

$$\int_{\Gamma^e_{\sigma,+}}\delta\mathbf{u}\cdot\mathbf{t}^+_0 ds^+ = \int_{\hat{\Omega}^e}\sum_{k=1}^{n}\psi_k(\xi^1,\xi^2)\left(\delta\underline{\mathbf{u}}^k + \frac{h}{2}\delta\boldsymbol{\varphi}^k + \frac{h}{2}\delta\Psi^k\hat{\mathbf{n}}\right)\cdot\mathbf{t}^+_0 J^+ d\hat{\Omega}^e \tag{6.53a}$$

$$\int_{\Gamma^e_{\sigma,-}}\delta\mathbf{u}\cdot\mathbf{t}^-_0 ds^- = \int_{\hat{\Omega}^e}\sum_{k=1}^{n}\psi_k(\xi^1,\xi^2)\left(\delta\underline{\mathbf{u}}^k - \frac{h}{2}\delta\boldsymbol{\varphi}^k + \frac{h}{2}\delta\Psi^k\hat{\mathbf{n}}\right)\cdot\mathbf{t}^-_0 J^- d\hat{\Omega}^e \tag{6.53b}$$

where the following quantities have been used:

$$J^+ = ||\mathbf{g}^+_1\times\mathbf{g}^+_2||, \qquad J^- = ||\mathbf{g}^-_1\times\mathbf{g}^-_2|| \tag{6.54a}$$

$$\mathbf{g}^+_\alpha = \mathbf{g}_\alpha(\xi^1,\xi^2,+1), \qquad \mathbf{g}^-_\alpha = \mathbf{g}_\alpha(\xi^1,\xi^2,-1) \tag{6.54b}$$

1. The nonlinear solution procedure

In general, the discrete virtual work statement constitutes a highly nonlinear set of algebraic equations. As in previous chapters, we employ Newton's method in the solution of the resulting equations. To facilitate a numerical solution for problems involving very large deformations, we further imbed the iterative Newton procedure within an incremental load stepping algorithm. For post-buckling analysis we employ a cylindrical arc-length solution procedure.

a. The basic iterative Newton procedure

The basic Newton method proceeds as follows: given a characteristic solution state Φ_{hp}^n (not necessarily satisfying the virtual work statement) we seek to find $\Delta\Phi_{hp}^{n+1}$ satisfying the following linearized expression

$$\mathcal{G}(\delta\Phi_{hp}, \Phi_{hp}^n) + D\mathcal{G}(\delta\Phi_{hp}, \Phi_{hp}^n)[\Delta\Phi_{hp}^{n+1}] = 0 \tag{6.55}$$

where $\Delta\Phi_{hp}^{n+1} = \Phi_{hp}^{n+1} - \Phi_{hp}^n$. To simplify the present discussion, we introduce the following notation for the discrete quantities: $\hat{\Phi}^n = \Phi_{hp}^n$, $\delta\hat{\Phi} = \delta\Phi_{hp}$ and $\Delta\hat{\Phi}^{n+1} = \Delta\Phi_{hp}^{n+1}$. Next, assuming that the applied loads are deformation independent, the discrete tangent operator $D\mathcal{G}(\delta\Phi_{hp}, \Phi_{hp}^n)[\Delta\Phi_{hp}^{n+1}]$ may be evaluated within a typical element as

$$D\mathcal{G}^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}] = D\mathcal{G}_G^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}] + D\mathcal{G}_M^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}] \tag{6.56}$$

The geometric tangent operator $D\mathcal{G}_G^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}]$ and material tangent operator $D\mathcal{G}_M^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}]$ are determined within the element as

$$\begin{aligned} D\mathcal{G}_G^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}] &= \int_{\mathcal{B}_0^e} \Big(D\delta\varepsilon_{ij}^{(0)}[\Delta\hat{\Phi}^{n+1}] + \xi^3 D\delta\varepsilon_{ij}^{(1)}[\Delta\hat{\Phi}^{n+1}] \Big) S^{ij} d\mathcal{B}_0^e \qquad (6.57a) \\ &= \int_{\hat{\Omega}^e} \Big(D\delta\varepsilon_{ij}^{(0)}[\Delta\hat{\Phi}^{n+1}] N^{ij} + D\delta\varepsilon_{ij}^{(1)}[\Delta\hat{\Phi}^{n+1}] M^{ij} \Big) d\hat{\Omega}^e \\ D\mathcal{G}_M^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}] &= \int_{\mathcal{B}_0^e} \Big(\delta\varepsilon_{ij}^{(0)} + \xi^3 \delta\varepsilon_{ij}^{(1)} \Big) \mathbb{C}^{ijkl} \Big(D\varepsilon_{kl}^{(0)}[\Delta\hat{\Phi}^{n+1}] \qquad (6.57b) \\ &\qquad + \xi^3 D\varepsilon_{kl}^{(1)}[\Delta\hat{\Phi}^{n+1}] \Big) d\mathcal{B}_0^e \\ &= \int_{\hat{\Omega}^e} \Big[\Big(\mathbb{A}^{ijkl} \delta\varepsilon_{ij}^{(0)} + \mathbb{B}^{ijkl} \delta\varepsilon_{ij}^{(1)} \Big) D\varepsilon_{kl}^{(0)}[\Delta\hat{\Phi}^{n+1}] \\ &\qquad + \Big(\mathbb{B}^{ijkl} \delta\varepsilon_{ij}^{(0)} + \mathbb{D}^{ijkl} \delta\varepsilon_{ij}^{(1)} \Big) D\varepsilon_{kl}^{(1)}[\Delta\hat{\Phi}^{n+1}] \Big] d\hat{\Omega}^e \end{aligned}$$

The contravariant components of the internal stress resultants N^{ij} and M^{ij} appearing in the discrete tangent operator may be evaluated as

$$N^{ij} = \int_{-1}^{+1} S^{ij} J d\xi^3 = \left(\mathbb{A}^{ijkl} \varepsilon_{kl}^{(0)} + \mathbb{B}^{ijkl} \varepsilon_{kl}^{(1)} \right) \Big|_{\Phi_{hp} = \hat{\Phi}^n} \tag{6.58a}$$

$$M^{ij} = \int_{-1}^{+1} \xi^3 S^{ij} J d\xi^3 = \left(\mathbb{B}^{ijkl} \varepsilon_{kl}^{(0)} + \mathbb{D}^{ijkl} \varepsilon_{kl}^{(1)} \right) \Big|_{\Phi_{hp} = \hat{\Phi}^n} \tag{6.58b}$$

Upon substitution of the discrete finite element solution variables and trial functions into the linearized virtual work statement, we arrive at a system of highly complex equations for the eth element in the finite element model of the form

$$[K^e]^{(n)} \{\delta \Delta^e\}^{(n+1)} = \{F^e\}^{(n)} \tag{6.59}$$

where $[K^e]^{(n)}$ is the element tangent coefficient matrix, $\{F^e\}^{(n)}$ is the element force vector and $\{\delta \Delta^e\}^{(n+1)} = \{\Delta^e\}^{(n+1)} - \{\Delta^e\}^{(n)}$ is the incremental solution. Due to the incredible complexity of the above system of equations (there are 22,050 unique terms in the discrete tangent operator $D\mathcal{G}^e(\delta\hat{\Phi}, \hat{\Phi}^n)[\Delta\hat{\Phi}^{n+1}]$), the symbolic algebra software Maple has been utilized in the construction of $[K^e]^{(n)}$ and $\{F^e\}^{(n)}$. As discussed in Chapter II, we partition Eq. (6.59) into the following equivalent form

$$\begin{bmatrix} [K^{11}]^{(n)} & \cdots & [K^{17}]^{(n)} \\ \vdots & \ddots & \vdots \\ [K^{71}]^{(n)} & \cdots & [K^{77}]^{(n)} \end{bmatrix} \begin{Bmatrix} \{\delta \Delta^{(1)}\}^{(n+1)} \\ \vdots \\ \{\delta \Delta^{(7)}\}^{(n+1)} \end{Bmatrix} = \begin{Bmatrix} \{F^{(1)}\}^{(n)} \\ \vdots \\ \{F^{(7)}\}^{(n)} \end{Bmatrix} \tag{6.60}$$

The components of each element sub-coefficient matrix and force vector may be expressed as

$$K_{ij}^{\alpha\beta(n)} = \int_{\hat{\Omega}^e} \sum_{l=0}^{2} \sum_{m=0}^{2} \mathcal{C}_{lm}^{\alpha\beta}(\underline{\mathbf{X}}(\xi^1, \xi^2), \hat{\mathbf{n}}(\xi^1, \xi^2), \Phi_{hp}^n(\xi^1, \xi^2)) \mathcal{S}_{ij}^{lm}(\xi^1, \xi^2) d\hat{\Omega}^e \tag{6.61a}$$

$$F_i^{\alpha(n)} = \int_{\hat{\Omega}^e} \sum_{l=0}^{2} \mathcal{F}_l^{\alpha}(\underline{\mathbf{X}}(\xi^1, \xi^2), \hat{\mathbf{n}}(\xi^1, \xi^2), \Phi_{hp}^n(\xi^1, \xi^2)) \mathcal{T}_i^l(\xi^1, \xi^2) d\hat{\Omega}^e \tag{6.61b}$$

where $i, j = 1, \ldots, (p+1)^2$ and $\alpha, \beta = 1, \ldots, 7$. The functions $\mathcal{S}_{ij}^{lm}$ and $\mathcal{T}_i^l$ are of the form

$$\mathcal{S}_{ij}^{00} = \psi_i \psi_j, \qquad \mathcal{S}_{ij}^{0m} = \psi_i \frac{\partial \psi_j}{\partial \xi^m}, \qquad \mathcal{S}_{ij}^{l0} = \frac{\partial \psi_i}{\partial \xi^l} \psi_j, \qquad \mathcal{S}_{ij}^{lm} = \frac{\partial \psi_i}{\partial \xi^l} \frac{\partial \psi_j}{\partial \xi^m} \tag{6.62a}$$

$$\mathcal{T}_i^0 = \psi_i, \qquad \mathcal{T}_i^1 = \frac{\partial \psi_i}{\partial \xi^1}, \qquad \mathcal{T}_i^2 = \frac{\partial \psi_i}{\partial \xi^2} \tag{6.62b}$$

where l and m each range from 1 to 2. The coefficients $\mathcal{C}_{lm}^{\alpha\beta}$ and $\mathcal{F}_l^\alpha$ (which are independent of i and j) are quite involved; in the numerical implementation we have obtained these quantities symbolically using Maple and have then translated the resulting expressions into C++ code. At this point it is worth noting that interpolating Ψ and $\hat{\mathbf{n}}$ separately in the finite element approximation of $\boldsymbol{\psi}$ (refer to Eq. (6.22)) is crucial in insuring that $\mathcal{C}_{lm}^{\alpha\beta}$ and $\mathcal{F}_l^\alpha$ are indeed independent of the i and j indices in $K_{ij}^{\alpha\beta(n)}$ and $F_i^{\alpha(n)}$.

The components of the element coefficient matrix and force vector are obtained numerically using the Gauss-Legendre quadrature rule, where $p+1$ quadrature points are taken in each coordinate direction of $\hat{\Omega}^e$. At a given integration point $(\xi_I^1, \xi_J^2) \in \hat{\Omega}^e$ we evaluate numerically, based on Eq. (6.51), the components of $\mathbb{A}^{ijkl}$, $\mathbb{B}^{ijkl}$ and $\mathbb{D}^{ijkl}$. Once the effective stiffnesses are known we determine $\mathcal{C}_{lm}^{\alpha\beta}(\xi_I^1, \xi_J^2)$ and $\mathcal{F}_l^\alpha(\xi_I^1, \xi_J^2)$ and then apply the summation procedure of the Gauss-Legendre quadrature rule to the components of $K_{ij}^{\alpha\beta(n)}$ and $F_i^{\alpha(n)}$. Repeating this process at each quadrature point in $\hat{\Omega}^e$ insures an efficient numerical implementation.

Upon application of the global finite element assembly operator $\mathbf{A}$, discussed in detail in Chapter II, we arrive at the following global system of linearized algebraic equations

$$[K]^{(n)} \{\delta\Delta\}^{(n+1)} = \{F\}^{(n)} \tag{6.63}$$

which may be constructed and solved recurrently for $\{\delta\Delta\}^{(n+1)}$ until a pre-defined

nonlinear convergence criterion has been satisfied. As part of the global assembly process, we employ element-level static condensation (see Chapter II). It is worth mentioning that in terms of system memory requirements (associated with constructing and storing the sparse form of $[K]^{(n)}$) our present high-order spectral/*hp* finite element formulation for shells is very much comparable with standard low-order shell finite element implementations. Furthermore, unlike many low-order discretizations, the present shell finite element formulation is completely displacement-based. As demonstrated in Section E, highly accurate numerical results may be obtained using the proposed shell element without the need for solution stabilization (e.g., reduced integration, assumed strain and/or mixed interpolation).

b. The incremental/iterative Newton and cylindrical arc-length procedures

We now discuss the incremental/iterative Newton procedure as well as the cylindrical arc-length method. These nonlinear solution strategies are necessary when solving problems involving very large deformations and/or rotations. In both solution schemes, we assume that the external loads are applied in increments. Next, we express the discrete weak formulation, given by Eq. (6.49), at the current load step $t + \Delta t$ as

$$^{t+\Delta t}\{R\} = {}^{t+\Delta t}\{F^{\text{int}}\} - {}^{t+\Delta t}\lambda\{F^{\text{ext}}\} \equiv 0 \tag{6.64}$$

where $^{t+\Delta t}\{F^{\text{int}}\}$ is a column vector obtained from the internal virtual work and $\{F^{\text{ext}}\}$ is a constant vector (independent of the load step) constructed from the externally applied virtual work. The quantity $^{t+\Delta t}\lambda$ is the load factor associated with the current load step. Linearizing the above expression using Newton's method yields

$$^{t+\Delta t}[K]^{(n)}\{\delta\Delta\}^{(n+1)} = -{}^{t+\Delta t}\{R\}^{(n)} + \delta\lambda^{(n+1)}\{F^{\text{ext}}\} \tag{6.65}$$

where $\{\delta\Delta\}^{(n+1)}$ and $\delta\lambda^{(n+1)}$ are defined as

$$\{\delta\Delta\}^{(n+1)} = {}^{t+\Delta t}\{\Delta\}^{(n+1)} - {}^{t+\Delta t}\{\Delta\}^{(n)} \tag{6.66a}$$

$$\delta\lambda^{(n+1)} = {}^{t+\Delta t}\lambda^{(n+1)} - {}^{t+\Delta t}\lambda^{(n)} \tag{6.66b}$$

In the incremental/iterative Newton solution procedure, ${}^{t+\Delta t}\lambda$ is prescribed by the user, and hence $\delta\lambda^{(n+1)} = 0$. In this case we solve for a sequence of shell configurations $\{{}^{k\Delta t}\Phi_{hp}\}_{k=1}^{N}$ associated with the prescribed load parameters $\{{}^{k\Delta t}\lambda\}_{k=1}^{N}$. In solving for configuration ${}^{t+\Delta t}\Phi_{hp}$, the coefficient matrix ${}^{t+\Delta t}[K]^{(0)}$ and residual ${}^{t+\Delta t}\{R\}^{(0)}$ are constructed using the converged solution ${}^{t}\Phi_{hp}$ from load step t.

For the vast majority of nonlinear problems, the incremental/iterative Newton procedure is adequate. However, in the numerical simulation of the post-buckling of shell structures, such a naive strategy may fail to trace the equilibrium path through the limit points. For these problems we employ an arc-length procedure, wherein a constraint equation is proposed to control the load factor associated with a given load step. For general details on the historical development of the arc-length method we refer to the work of Riks [139, 140] and Crisfield [141] (detailed explanations of the method may also be found in the texts of Bathe [25] and Reddy [50]).

In the arc-length solution procedure, we introduce the following additive decomposition of the incremental solution

$$\{\delta\Delta\}^{(n+1)} = \{\delta\bar{\Delta}\}^{(n+1)} + \delta\lambda^{(n+1)}\{\delta\tilde{\Delta}\}^{(n+1)} \tag{6.67}$$

Using the above expression along with Eq. (6.65) allows us to obtain the following two sets of linearized equations for $\{\delta\bar{\Delta}\}^{(n+1)}$ and $\{\delta\tilde{\Delta}\}^{(n+1)}$

$${}^{t+\Delta t}[K]^{(n)}\{\delta\bar{\Delta}\}^{(n+1)} = -{}^{t+\Delta t}\{R\}^{n} \tag{6.68a}$$

$${}^{t+\Delta t}[K]^{(n)}\{\delta\tilde{\Delta}\}^{(n+1)} = \{F^{\text{ext}}\} \tag{6.68b}$$

Once the above equations have been solved (and assuming of course that $\delta\lambda^{(n+1)}$ is known), we may obtain $\{\delta\Delta\}^{(n+1)}$ using Eq. (6.67). Next we define the solution increments ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)}$ and ${}^{t+\Delta t}\hat{\lambda}^{(n+1)}$, between configurations ${}^{t+\Delta t}\Phi_{hp}^{(n+1)}$ and ${}^{t}\Phi_{hp}$, as

$$\begin{aligned} {}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)} &= {}^{t+\Delta t}\{\Delta\}^{(n+1)} - {}^{t}\{\Delta\} && \text{(6.69a)} \\ &= {}^{t+\Delta t}\{\hat{\Delta}\}^{(n)} + \{\delta\bar{\Delta}\}^{(n+1)} + \delta\lambda^{(n+1)}\{\delta\tilde{\Delta}\}^{(n+1)} \\ {}^{t+\Delta t}\hat{\lambda}^{(n+1)} &= {}^{t+\Delta t}\lambda^{(n+1)} - {}^{t}\lambda && \text{(6.69b)} \\ &= {}^{t+\Delta t}\hat{\lambda}^{(n)} + \delta\lambda^{(n+1)} \end{aligned}$$

With the above formulas in mind, we are able to define the standard spherical arc-length constraint equation for $\delta\lambda^{(n+1)}$ as

$$\begin{aligned} {}^{t+\Delta t}\mathcal{K}^{(n+1)} &= ||{}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)}||^2 + \beta({}^{t+\Delta t}\hat{\lambda}^{(n+1)})^2||\{F^{\text{ext}}\}||^2 - ({}^{t+\Delta t}\Delta L)^2 \\ &= a_1(\delta\lambda^{(n+1)})^2 + a_2\delta\lambda^{(n+1)} + a_3 = 0 \end{aligned} \tag{6.70}$$

where ${}^{t+\Delta t}\Delta L$ is the so-called arc-length, β is a scaling parameter and $||\cdot||$ denotes the Euclidean norm. The constraint ${}^{t+\Delta t}\mathcal{K}^{(n+1)}$ is a quadratic equation in $\delta\lambda^{(n+1)}$ with coefficients: a_1, a_2 and a_3 given as

$$a_1 = ||\{\delta\tilde{\Delta}\}^{(n+1)}||^2 + \beta||\{F^{\text{ext}}\}||^2 \tag{6.71a}$$

$$a_2 = 2\big[\big({}^{t+\Delta t}\{\hat{\Delta}\}^{(n)} + \{\delta\bar{\Delta}\}^{(n+1)}\big)^{\mathrm{T}}\{\delta\tilde{\Delta}\}^{(n+1)} + \beta^{t+\Delta t}\hat{\lambda}^{(n)}||\{F^{\text{ext}}\}||^2\big] \tag{6.71b}$$

$$a_3 = ||{}^{t+\Delta t}\{\hat{\Delta}\}^{(n)} + \{\delta\bar{\Delta}\}^{(n+1)}||^2 + \beta({}^{t+\Delta t}\hat{\lambda}^{(n)})^2||\{F^{\text{ext}}\}||^2 - ({}^{t+\Delta t}\Delta L)^2 \tag{6.71c}$$

The two possible solutions for the constraint equation may be expressed as

$$\delta\lambda_1^{(n+1)} = \frac{-a_2 + \sqrt{a_2^2 - 4a_1a_3}}{2a_1}, \qquad \delta\lambda_2^{(n+1)} = \frac{-a_2 - \sqrt{a_2^2 - 4a_1a_3}}{2a_1} \tag{6.72}$$

We select $\delta\lambda_i^{(n+1)}$ such that the inner product of ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)}$ with ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n)}$ is

positive. This insures that we do not march backwards along the previously computed solution path. In the event that both $\delta\lambda_1^{(n+1)}$ and $\delta\lambda_2^{(n+1)}$ yield positive inner products of ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)}$ with ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n)}$, we select $\{\delta\Delta\}^{(n+1)}$ such that ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n+1)}$ is closest to ${}^{t+\Delta t}\{\hat{\Delta}\}^{(n)}$ in the Euclidean metric. For the first iteration of a given load step, we select ${}^{t+\Delta t}\{\hat{\Delta}\}^{(1)}$ such that the inner product of ${}^{t+\Delta t}\{\hat{\Delta}\}^{(1)}$ with ${}^{t}\{\hat{\Delta}\}$ is positive [142] (where ${}^{t}\{\hat{\Delta}\}$ is the converged incremental solution from load step t). In the numerical implementation we take $\beta = 0$, which results in the well-known cylindrical arc-length procedure (we refer to Crisfield [141] for a discussion on the importance of β).

To initialize the arc-length solution method (at the first load step: Δt and initial iteration: $n = 0$) we take ${}^{\Delta t}\lambda^{(0)} = 0$, *prescribe* an appropriate value for $\delta\lambda^{(1)}$ (typically we define $\{F^{\text{ext}}\}$ such that it is sufficient to take $\delta\lambda^{(1)} = 1$) and then solve Eq. (6.68b) for $\{\delta\tilde{\Delta}\}^{(1)}$. We then take $\{\delta\Delta\}^{(1)} = \delta\lambda^{(1)}\{\delta\tilde{\Delta}\}^{(1)}$ and define the arc-length ${}^{\Delta t}\Delta L$ for the subsequent nonlinear iterations as

$$ {}^{\Delta t}\Delta L = \delta\lambda^{(1)}||\delta\tilde{\Delta}^{(1)}|| \tag{6.73} $$

To improve the efficiency of the arc-length method, we adjust the arc-length ${}^{t+\Delta t}\Delta L$ depending on how many iterations were required to achieve nonlinear solution convergence at the immediate previous load step t. We adopt the following formula from the literature [142]

$$ {}^{t+\Delta t}\Delta L = {}^{t}\Delta L \sqrt{{}^{t+\Delta t}I/{}^{t}I} \tag{6.74} $$

where ${}^{t}I$ is the *actual* number of iterations required for convergence at the immediate previous load step and ${}^{t+\Delta t}I$ is the *desired* number of iterations required to satisfy the convergence criterion at the current load step. We typically take $4 \leq {}^{t+\Delta t}I \leq 6$, which naturally reduces the arc-length in the vicinity of limit points and increases the arc-length whenever nonlinear convergence is quickly achieved.

E. Numerical examples: verification benchmarks

In this section we present numerical results for various standard shell benchmark problems. The problems include various plates and cylinders, a hemisphere with an 18° hole and a pinched hyperboloidal shell. The problems have been selected to showcase the capabilities of the proposed shell finite element formulation in solving some of the more challenging finite deformation problems for elastic shells found in the literature. We are particularly interested in comparing solutions obtained using the present shell finite element formulation with the numerical results reported by Arciniega and Reddy [124], wherein a tensor-based shell finite element model was adopted. In this previous research, a given shell geometry was prescribed exactly at the quadrature points while high-order Lagrange type basis functions (with *equal* spacing of the element nodes) were utilized for the numerical solution.

We construct the finite element approximation of the undeformed mid-surface geometry for each example problem by mapping the nodal positions of a conforming finite element discretization of $\bar{\boldsymbol{\omega}} \subset \mathbb{R}^2$ (a closed and bounded region) onto the nodal locations associated with $\bar{\Omega}^{hp} \subset \mathbb{R}^3$. The coordinates of $\mathbb{R}^2$ are denoted as (ω^1, ω^2) and unless otherwise stated we take $\bar{\boldsymbol{\omega}} = [0,1] \times [0,1]$. The discrete mapping used to characterize the nodal coordinates of $\bar{\Omega}^{hp}$ is also employed to prescribe the nodal values for $\hat{\mathbf{n}}$ and $\hat{\mathbf{t}}$. A convergence criterion of 10^{-6} is adopted in all numerical examples.

1. A cantilevered plate strip under an end load

As a first example problem, we consider the mechanical response of a cantilevered plate strip subjected to a distributed end shear load q as shown in Figure 57, where $L = 10$, $b = 1$ and $h = 0.1$. We consider an isotropic plate and also a multi-layered

composite laminate with material properties given as

$$\text{Isotropic:} \qquad E = 1.2 \times 10^6, \quad \nu = 0.0 \tag{6.75a}$$

$$\text{Orthotropic:} \begin{cases} E_1 = 1.0 \times 10^6, & E_2 = E_3 = 0.3 \times 10^6 \\ G_{23} = 0.12 \times 10^6, & G_{13} = G_{12} = 0.15 \times 10^6 \\ \nu_{23} = 0.25, & \nu_{13} = \nu_{12} = 0.25 \end{cases} \tag{6.75b}$$

The isotropic problem has been considered by many authors (see for example Refs. [143, 144, 145, 146, 147, 148, 149]), while a composite version of the problem has been proposed recently by Arciniega [150].

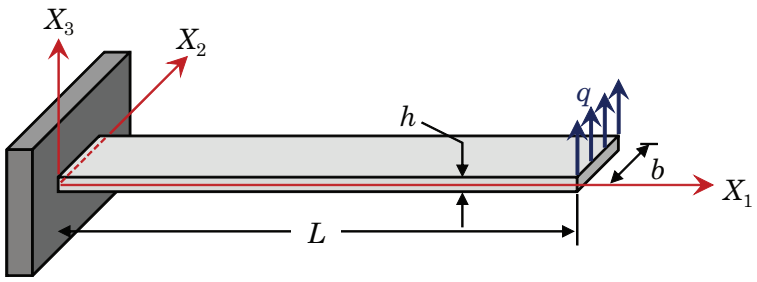


Fig. 57. A cantilevered plate strip subjected at its end to a vertically applied shear force.

We employ a regular finite element mesh consisting of 4 elements, with the p-level taken as 4. The unit normal and unit tangent vectors are prescribed as: $\hat{\mathbf{n}} = \hat{\mathbf{E}}_3$ and $\hat{\mathbf{t}} = \hat{\mathbf{E}}_1$. In Figure 58 we show the computed axial and vertical deflections of the plate tip for the isotropic case. The calculated deflections are in excellent agreement with the numerical results reported by Sze et al. [149]. In Figure 59 we trace the transverse tip deflections vs. the applied load q for four different lamination schemes. We see that the stacking sequence $(90°/0°/90°)$ yields the most flexible response while the $(0°/90°/0°)$ laminate exhibits the greatest stiffness. As expected, the non-symmetric

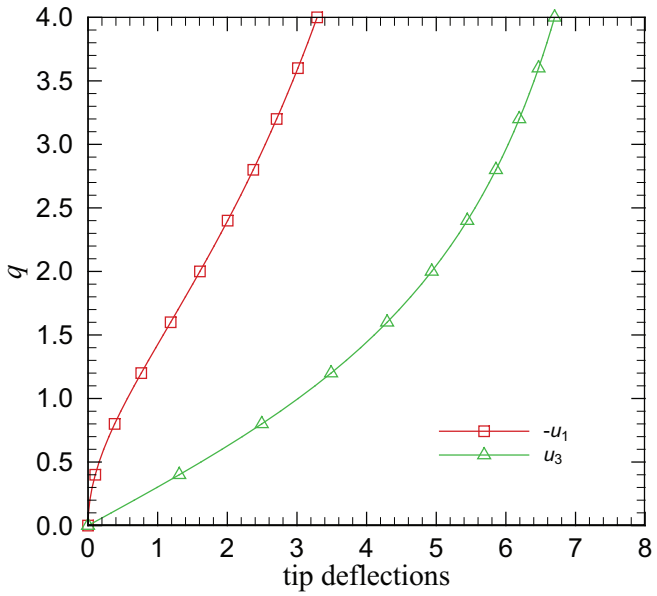


Fig. 58. Tip deflections vs. shear load q for an isotropic cantilevered plate strip under end loading.

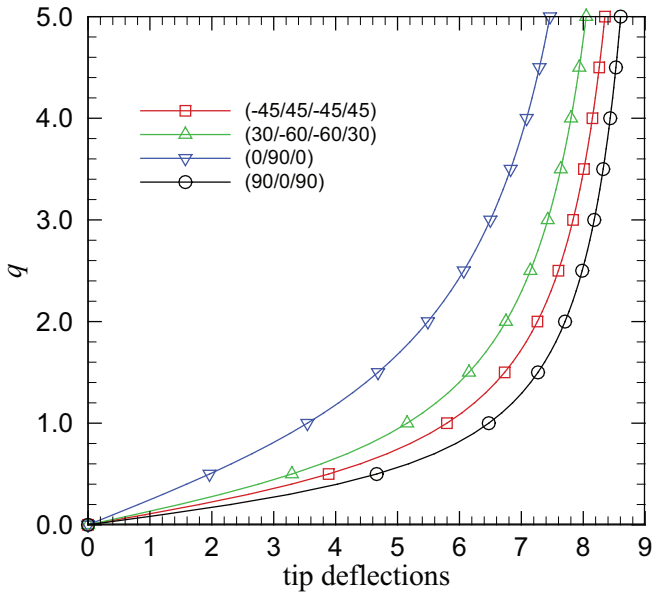


Fig. 59. Vertical tip deflections u_3 vs. shear load q for laminated composite cantilevered plate strips under end loading.

stacking sequence (-45°/45°/-45°/45°) also leads to lateral deflection of the plate in the direction of the X_2 coordinate. The composite plate results compare nicely with the results reported by Arciniega [150]. For completeness, we show in Figure 60 the undeformed and various deformed mid-surface configurations of the isotropic plate strip.

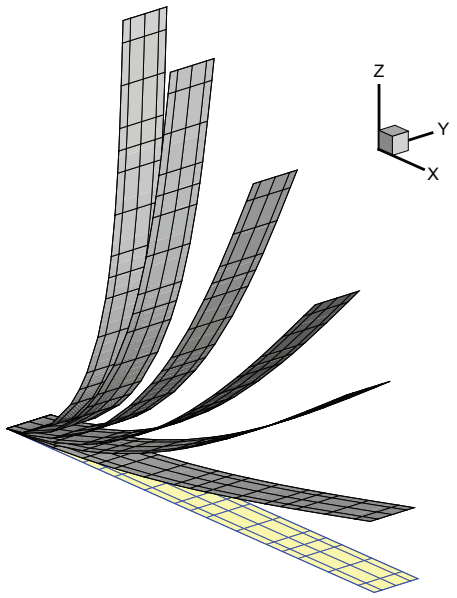


Fig. 60. Undeformed and various deformed mid-surface configurations of an isotropic cantilevered plate strip subjected at its end to a vertical shear force ($q = 0.4, 1.2, 2, 4, 10$ and 20).

2. Post-buckling of a plate strip

In this next example we wish to determine the post-buckling behavior of an isotropic plate strip subjected to an end compressive load q as shown in Figure 61. The material properties for the problem are those employed by Massin and Al Mikdad [151], given as

$$E = 2.0 \times 10^{11}, \quad \nu = 0.3 \tag{6.76}$$

In addition, the geometric parameters are prescribed as $L = 0.5$, $b = 0.075$ and $h = 0.0045$. The analytical solution, first obtained by Leonhard Euler, may be found in the well-known text on the linearized theory of elasticity by Timoshenko [152].

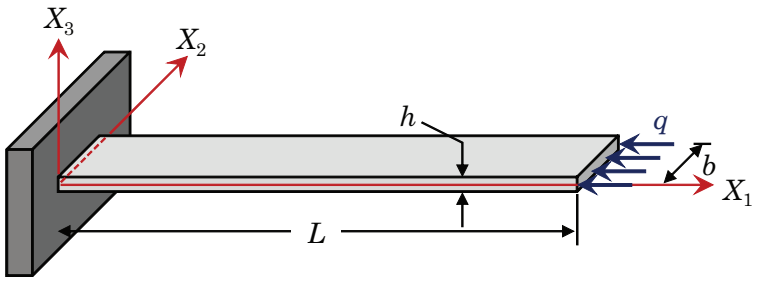


Fig. 61. A cantilevered plate strip subjected at its end to a compressive axial force.

To instigate post-buckling behavior of the plate beyond the limit point, we introduce a perturbation technique, wherein the load is prescribed slightly out-of-plane at an angle of 1/1000 radians (see Massin and Al Mikdad [151] for a similar approach). In Figure 62 we trace the axial and transverse deflections of the plate tip vs. the externally applied load P, where P is the net resultant force associated with the distributed load q. We also show in this figure the Euler-Bernoulli beam theory based critical buckling load $P_{\text{cr}} = EI(\pi/2L)^2 = 1124.21$, where $I = bh^3/12$ is the second moment of area about the X_2 axis. We see that post-buckling occurs in the numerical simulation in the immediate vicinity of the critical load P_{cr}. We find that our computed tip deflections are in excellent agreement with the numerical results reported by Arciniega and Reddy [124]. In Figure 63 we further show the undeformed and various post-buckled mid-surface configurations of the plate strip. Although the cylindrical arc-length method may be employed for this problem, the reported results have been obtained using the incremental/iterative Newton procedure. This is admissible since the applied load is non-decreasing when traversing the limit point.

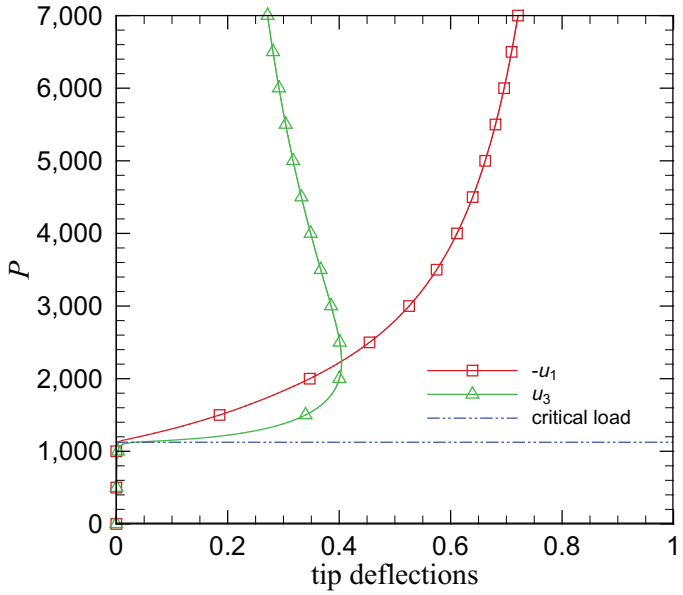


Fig. 62. Tip deflections vs. compressive load P for the cantilevered plate strip (a mesh of 4 elements with the p-level taken as 8 has been employed).

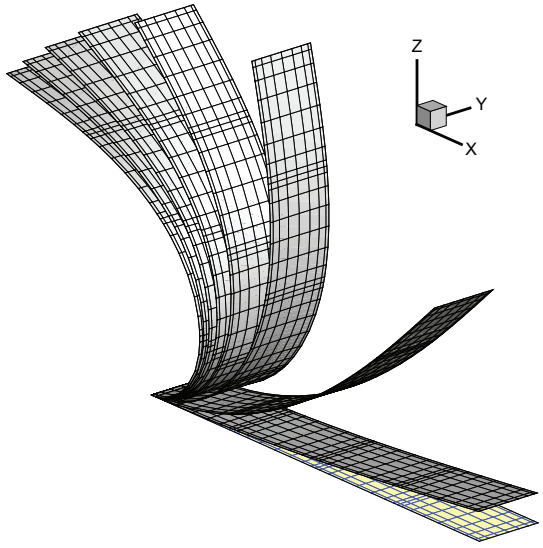


Fig. 63. Undeformed and various post-buckled deformed mid-surface configurations of the axially loaded cantilevered plate strip (P = 1,125, 1,250, 2,000, 3,000, 4,000, 5,000, 6,000, 7,000).

3. A slit annular plate under an end shear force

We now examine a plate problem whose geometry cannot be exactly characterized in terms of the isoparametric map given in Eq. (6.1). The problem consists of a slit cantilevered annular plate as shown in Figure 64 that is subjected to a line shear load q at its free end. We take $R_i = 6$, $R_o = 10$ and $h = 0.03$. We consider an isotropic plate and also a multi-layered composite laminate, where the material properties are taken as

$$\text{Isotropic:} \qquad E = 21.0 \times 10^6, \quad \nu = 0.0 \tag{6.77a}$$

$$\text{Orthotropic:} \begin{cases} E_1 = 20.0 \times 10^6, & E_2 = E_3 = 6.0 \times 10^6 \\ G_{23} = 2.4 \times 10^6, & G_{13} = G_{12} = 3.0 \times 10^6 \\ \nu_{23} = 0.25, & \nu_{13} = \nu_{12} = 0.3 \end{cases} \tag{6.77b}$$

Numerical solutions for the isotropic case may be found in Refs. [126, 153, 154, 155, 156, 148, 149] among others, while a laminated composite version of the problem has been solved by Arciniega and Reddy [124].

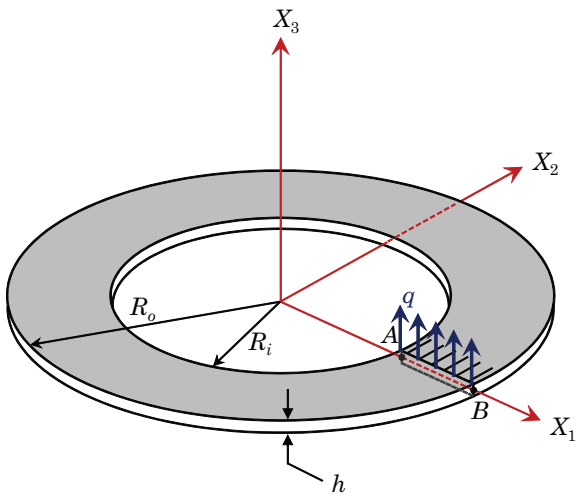


Fig. 64. A cantilevered slit annular plate subjected at its end to a vertical shear force.

We employ a finite element mesh consisting of 4 elements with the p-level taken as 8. The nodal coordinates of the mid-surface $\bar{\Omega}^{hp}$ are obtained using the following formula

$$\underline{\mathbf{X}} = [R_i + (R_o - R_i)\omega^1][\cos(2\pi\omega^2)\hat{\mathbf{E}}_1 + \sin(2\pi\omega^2)\hat{\mathbf{E}}_2] \tag{6.78}$$

The unit normal vector is given as $\hat{\mathbf{n}} = \hat{\mathbf{E}}_3$ and the unit tangent vector (used for the laminated composite problem) is defined at the nodes as

$$\hat{\mathbf{t}} = \cos(2\pi\omega^2)\hat{\mathbf{E}}_1 + \sin(2\pi\omega^2)\hat{\mathbf{E}}_2 \tag{6.79}$$

Each numerical simulation is conducted using the incremental/iterative Newton procedure with 80 load steps.

The transverse tip deflections vs. the net applied force $P = (R_o - R_i)q$ at points A and B are shown for the isotropic case in Figure 65. The computed deflections agree very well with the tabulated displacement values reported by Sze et al. [149]. In Figure 66 we trace the tip deflections at point B vs. the applied load P for four distinct lamination schemes. Our computed results are found to be in excellent agreement with the displacements reported by Arciniega and Reddy [124] for each set of stacking sequences. In Figure 67 we show the undeformed and various deformed mid-surface configurations of the isotropic plate and the (-45°/45°/-45°/45°) laminated composite structure. Clearly, both structures undergo very large deformations which are qualitatively quite similar.

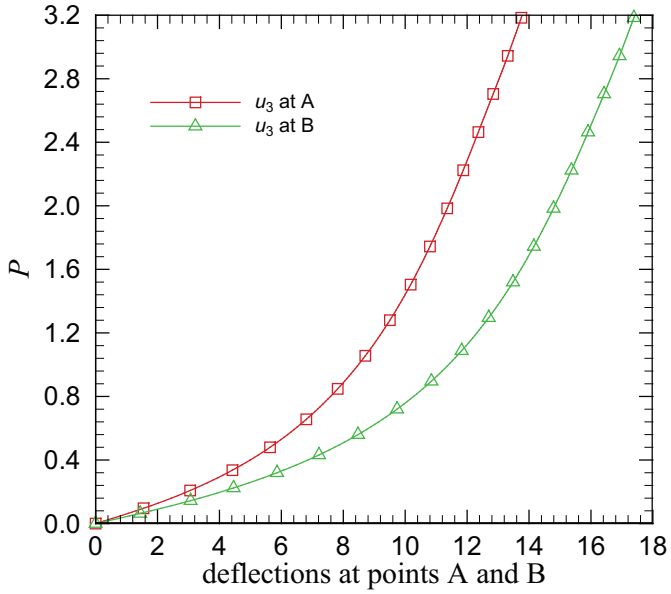


Fig. 65. Tip deflections at points A and B vs. shear force P for the isotropic slit annular plate.

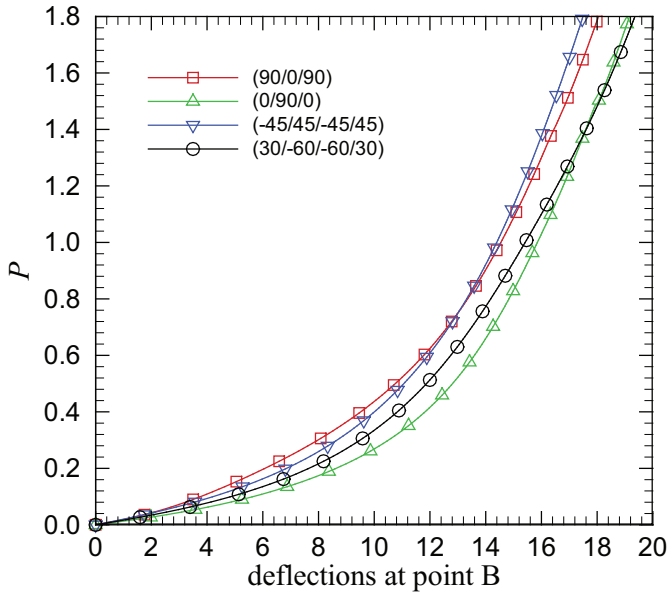
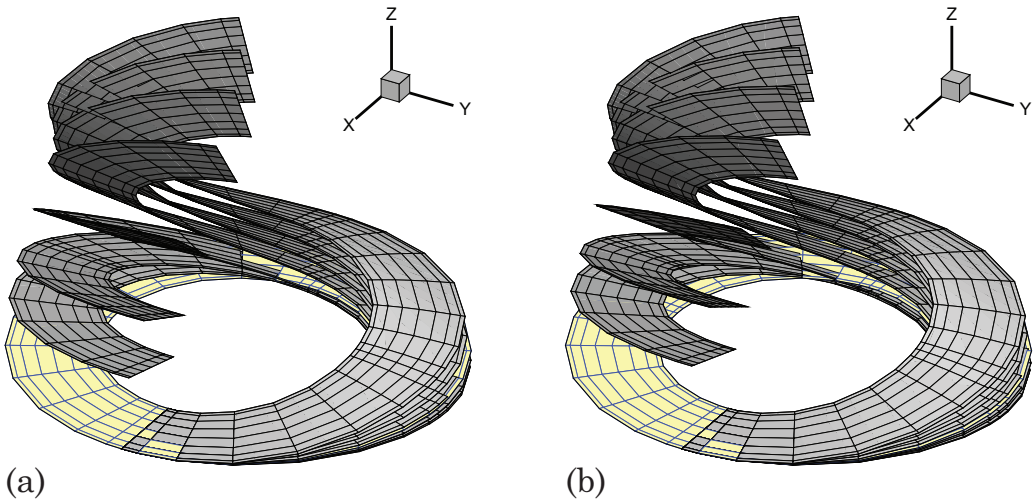


Fig. 66. Vertical tip deflections u_3 at point B vs. shear force P for various laminated composite slit annular plates.



(a) (b)

Fig. 67. Undeformed and various deformed mid-surface configurations of two annular plates: (a) an isotropic plate, where $P = 0.16$, 0.32, 0.64, 1.28, 1.92, 2.56 and 3.20 and (b) a laminated composite plate with (-45°/45°/-45°/45°) stacking sequence, where $P = 0.09$, 0.18, 0.36, 0.72, 1.08, 1.44 and 1.80.

4. A cylindrical panel subjected to a point load

We next examine the mechanical response of various thin cylindrical rooflike panels, each subjected to a point force P as shown in Figure 68. Variants of this problem are found throughout the literature (see for example Refs. [143, 146, 153, 154, 147, 124] among others) and are especially popular on account of the snap-through behavior. In the present example we take $\alpha = 0.1$ rad., $a = 508$ mm and $R = 2{,}540$ mm (where R is the radius of the undeformed mid-surface). We perform a parametric study by considering the following three cases for the shell thickness: $h = 25.4, 12.7$ and 6.35 mm. We investigate isotropic, laminated composite and functionally graded shell configurations with material properties given as

$$\text{Isotropic:} \qquad E = 3{,}102.75 \text{ N/mm}, \quad \nu = 0.3 \tag{6.80a}$$

$$\text{Orthotropic:} \quad \begin{cases} E_1 = 3{,}300 \text{ N/mm}, & E_2 = E_3 = 1{,}100 \text{ N/mm} \\ G_{23} = 440 \text{ N/mm}, & G_{13} = G_{12} = 660 \text{ N/mm} \\ \nu_{23} = 0.25, & \nu_{13} = \nu_{12} = 0.25 \end{cases} \tag{6.80b}$$

$$\text{Functionally graded:} \quad \begin{cases} E^- = 70 \text{ GPa}, & E^+ = 151 \text{ GPa} \\ \nu^- = 0.3, & \nu^+ = 0.3 \end{cases} \tag{6.80c}$$

For the laminated composite shell problems, we consider the following lamination schemes: (90°/0°/90°), (0°/90°/0°), (-45°/45°/-45°/45°) and (30°/-60°/-60°/30°).

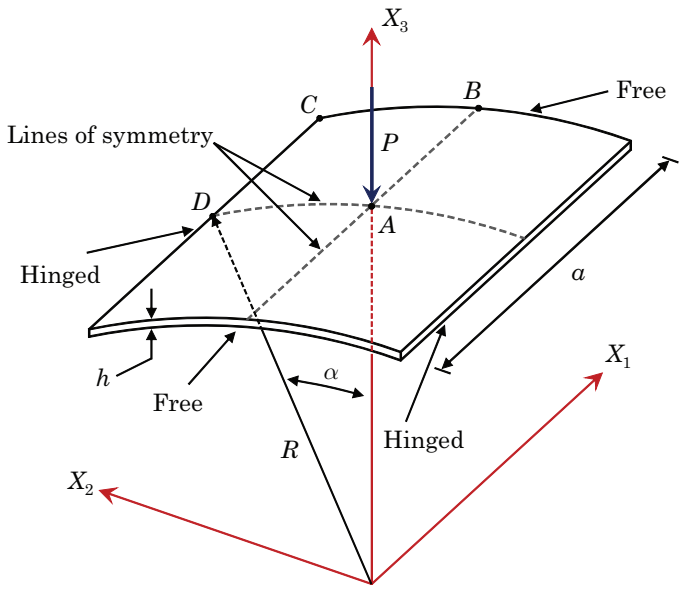


Fig. 68. A shallow cylindrical panel subjected at its center to a vertical point load.

The finite element nodal values for the mid-surface coordinates and the unit normal vector are obtained using the following formulas

$$\underline{\mathbf{X}} = \frac{a}{2}\omega^1 \hat{\mathbf{E}}_1 + R[\sin(\alpha\omega^2)\hat{\mathbf{E}}_2 + \cos(\alpha\omega^2)\hat{\mathbf{E}}_3] \tag{6.81a}$$

$$\hat{\mathbf{n}} = \sin(\alpha\omega^2)\hat{\mathbf{E}}_2 + \cos(\alpha\omega^2)\hat{\mathbf{E}}_3 \tag{6.81b}$$

where the full physical domain may be parametrized by taking $\bar{\boldsymbol{\omega}} = [-1, 1]^2$. The unit tangent vector is prescribed as $\hat{\mathbf{t}} = \hat{\mathbf{E}}_1$. With the exception of the angled-ply laminates (-45°/45°/-45°/45°) and (30°/-60°/-60°/30°), all numerical simulations are conducted using one quarter of the physical domain by taking $\bar{\boldsymbol{\omega}} = [0, 1]^2$ and invoking appropriate symmetry boundary conditions. We employ a uniform 2×2 mesh for the quarter model and a 4×4 discretization for the full domain using a p-level of 4. Along the hinged edges, we take the nodal translations and X_1 component of the difference vector as zero.

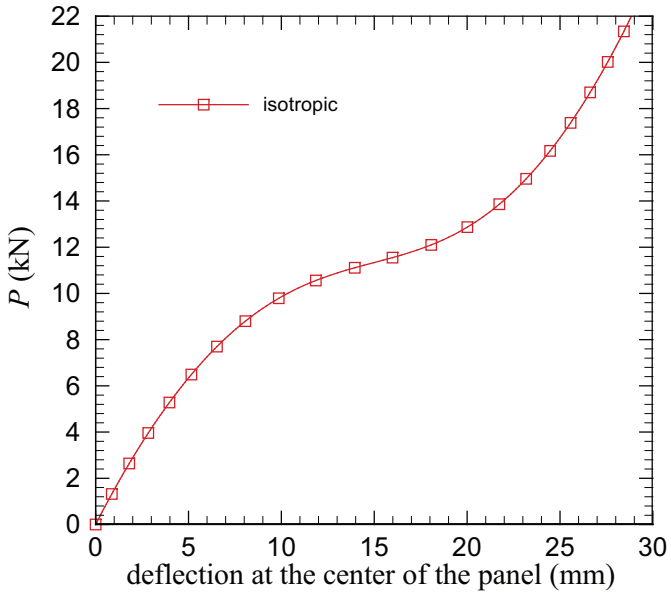


Fig. 69. Vertical deflection of a shallow isotropic cylindrical panel under point loading (case shown is for $h = 25.4$ mm).

In Figure 69 we show the deflection of the isotropic shell at point A vs. the applied load P for the case where $h = 25.4$ mm. The results, which agree strongly with those reported by Arciniega [150], have been obtained using the incremental/iterative Newton procedure. In Figures 70 through 73 we trace the center deflections

vs. P for the isotropic and laminated composite panels for the cases where $h = 12.7$ and 6.35 mm. Each numerical simulation has been conducted using the cylindrical arc-length method. The results are in excellent agreement with the tabulated values given by Sze et al. [149]; and for the angled-ply laminates, the solutions presented by Arciniega and Reddy [124]. It is evident that decreasing the shell thickness greatly increases the complexity of the equilibrium path associated with the arc-length based numerical solution. For example, we observe from Figures 72 and 73 that laminates (0°/90°/0°), (-45°/45°/-45°/45°) and (30°/-60°/-60°/30°) exhibit highly involved equilibrium paths when $h = 6.35$ mm.

Numerical results for metal-ceramic functionally graded panels, for the cases where $h = 12.7$ and 6.35 mm, are shown in Figures 74 and 75. The metal (aluminum) is taken as the bottom material and the ceramic (zirconia) as the top constituent, with the elastic properties given in Eq. (6.80c). As in the isotropic and laminated composite cases, the complexity of the equilibrium paths of the functionally graded panels increases as the shell thickness h is reduced. We adopt the cylindrical arc-length procedure and vary the power-law parameter n to obtain the numerical solutions. The results shown in Figures 74 and 75 are visually in unison with the deflection curves provided by Arciniega and Reddy [124].

An artifact of the snap-through phenomena is the mathematical existence of multiple solution configurations for certain loading scenarios. For example, the 6.35 mm thick laminate panel with stacking sequence (0°/90°/0°) possesses 5 equilibrium configurations for the case where $P = 0$ kN. These configurations (including the undeformed configuration) are shown in Figure 76, from left to right and top to bottom, in the order in which they occur in traveling along the equilibrium path (shown in Figure 72). Further mathematical analysis (which is beyond the scope of this dissertation) is required to assess the stability of each solution configuration.

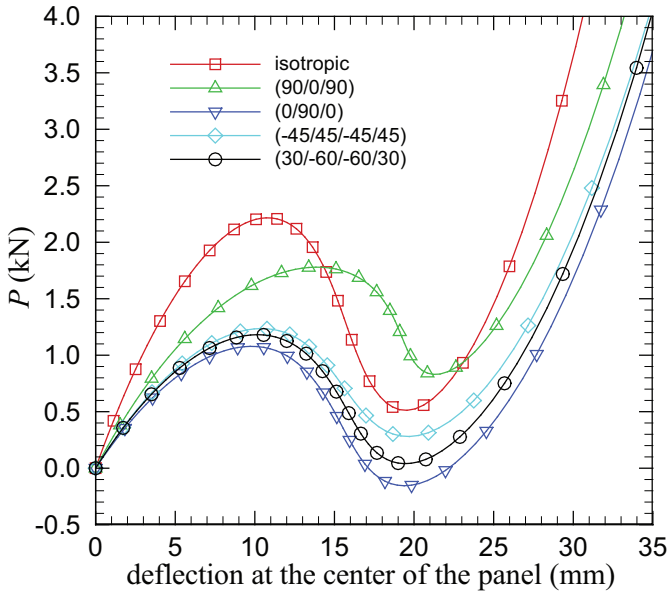


Fig. 70. Vertical deflection of an isotropic and laminated composite shallow cylindrical panels under point loading (cases shown are for $h = 12.7$ mm).

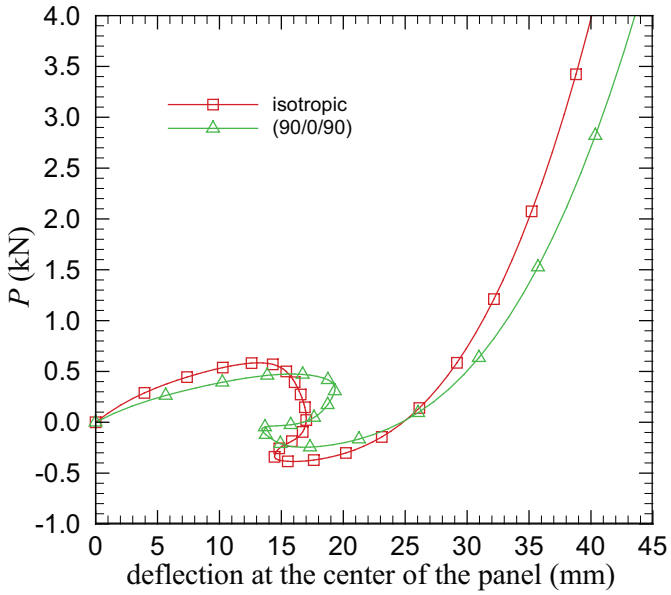


Fig. 71. Vertical deflection of an isotropic and a laminated composite shallow cylindrical panel under point loading (cases shown are for $h = 6.35$ mm).

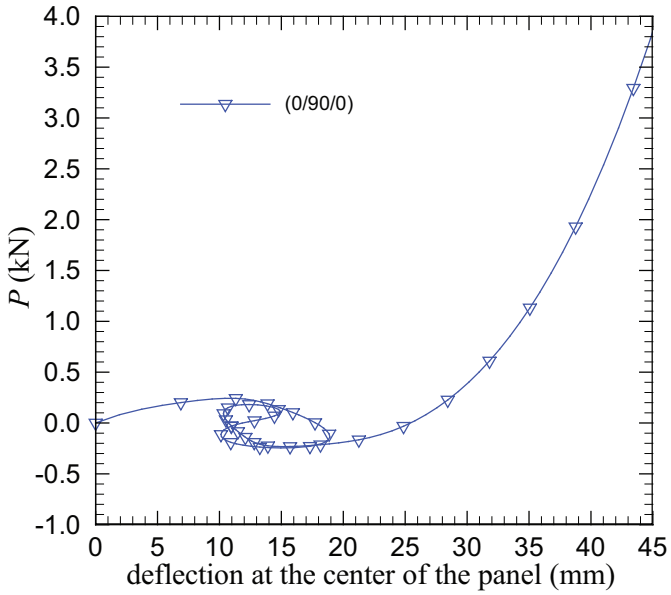


Fig. 72. Vertical deflection of a laminated composite shallow cylindrical panel under point loading (case shown is for $h = 6.35$ mm).

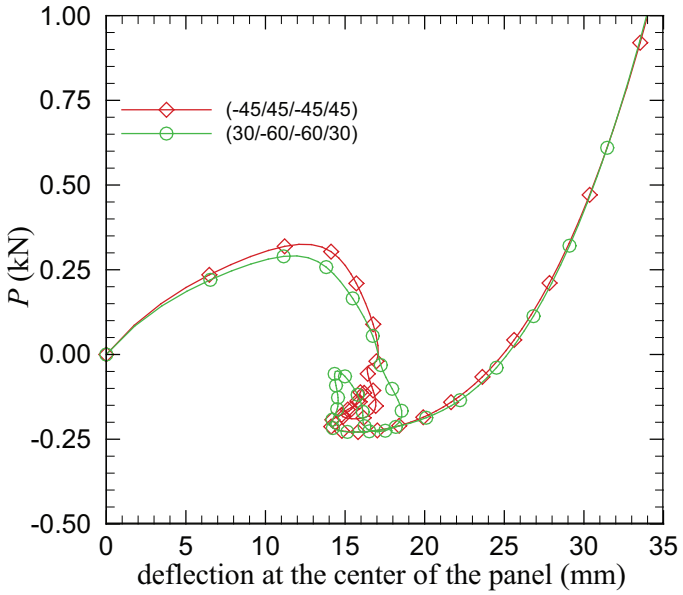


Fig. 73. Vertical deflection of laminated composite shallow cylindrical panels under point loading (cases shown are for $h = 6.35$ mm).

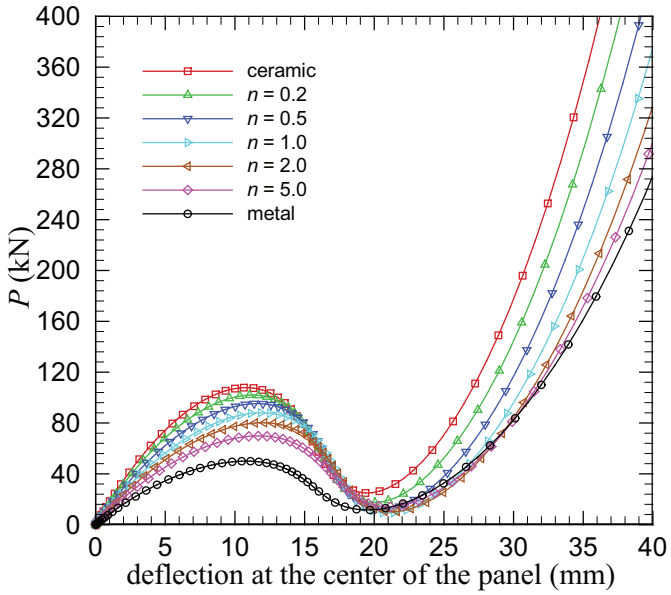


Fig. 74. Vertical deflection of functionally graded metal-ceramic shallow cylindrical panels under point loading (cases shown are for $h = 12.7$ mm).

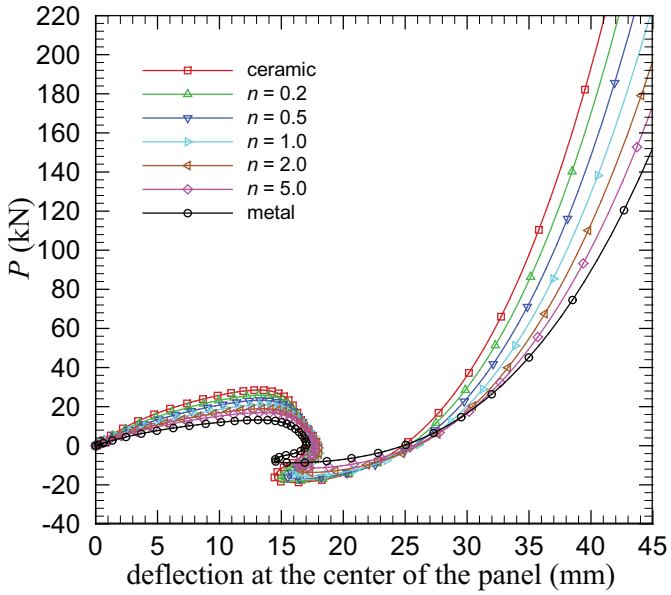


Fig. 75. Vertical deflection of functionally graded metal-ceramic shallow cylindrical panels under point loading (cases shown are for $h = 6.35$ mm).

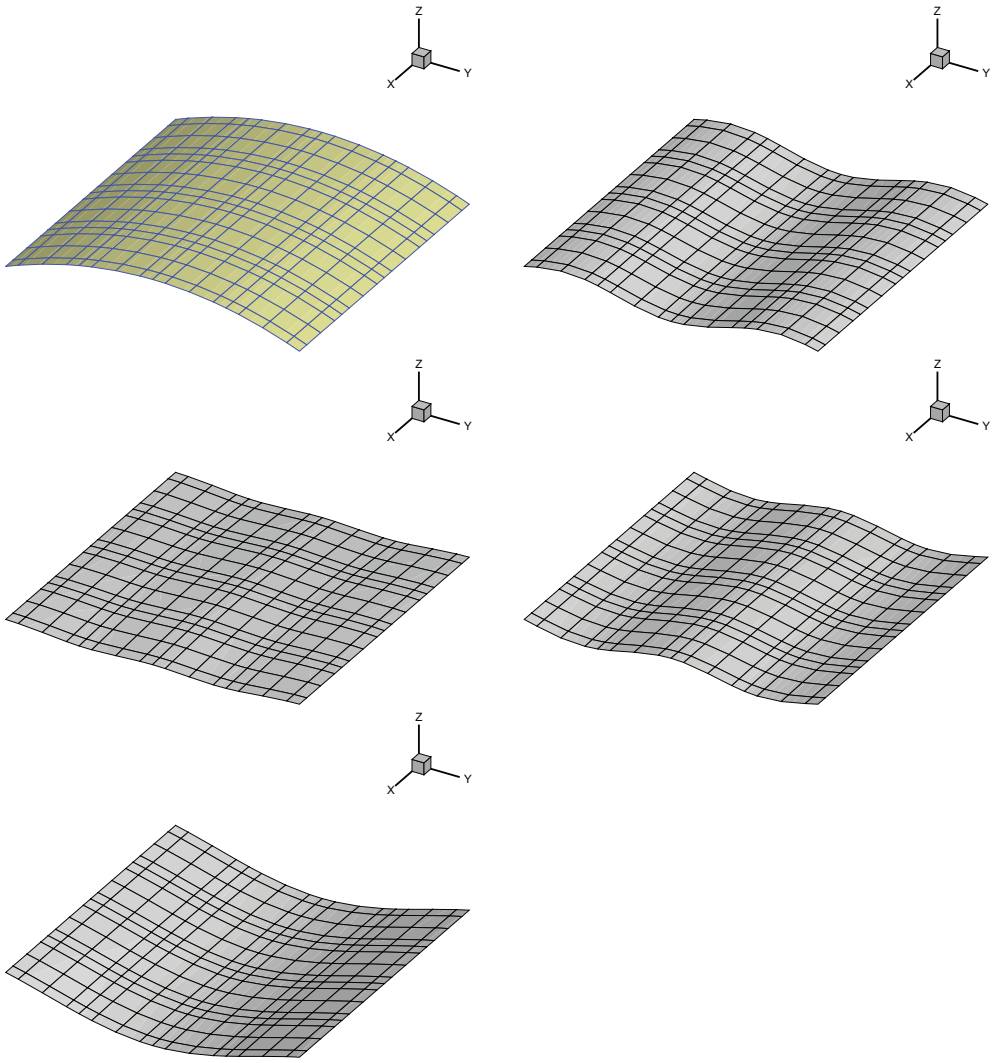


Fig. 76. Undeformed and various deformed mid-surface configurations of the $(0°/90°/0°)$ stacking sequence laminated composite shallow cylindrical panel (cases shown are for $h = 6.35$ mm and $P = 0$ kN). The vertical component of each mid-surface configuration has been magnified by a factor of 4.

5. Pull-out of an open-ended cylindrical shell

In this example we consider the mechanical deformation of an open-ended cylinder, as shown in Figure 77, subjected to two pull-out point forces P. Unlike the previous example, in this problem we apply the loads such that the shell undergoes very large displacements and rotations. As a result, this problem constitutes a severe test of shell finite element formulations and has been addressed in Refs. [153, 155, 156, 148, 149, 124] among others. The isotropic material properties are taken as

$$E = 10.5 \times 10^6, \quad \nu = 0.3125 \tag{6.82}$$

The geometric parameters are taken as: $L = 10.35$, $h = 0.094$ and $R = 4.953$ (where we have taken R as the radius of the *undeformed* mid-surface as opposed to the radius of the *inner* surface of the shell).

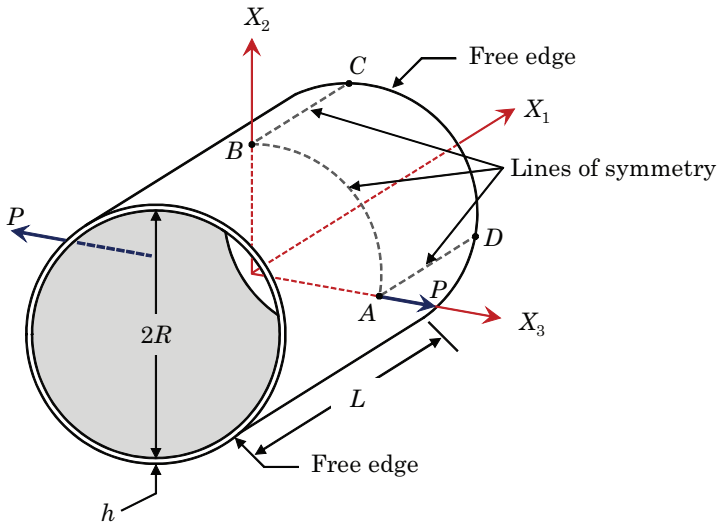


Fig. 77. An open-ended cylindrical shell subjected to two point loads.

Symmetry in the geometry, material properties and loading allow us to construct the numerical model using only an octant of the actual open-ended cylinder. For the

numerical model we employ a regular 2×2 mesh (with the p-level taken as 8) of the shell octant containing points A, B, C and D. The incremental/iterative Newton procedure is adopted using a total of 80 load steps.

The radial deflections vs. the net applied pulling force P are shown in Figure 78 for points A, B and C. The computed deflections are in excellent agreement with results of Sze et al. [149] and also Arciniega and Reddy [124]. The mechanical response of the shell is interesting in that the deformation is initially bending dominated; however, membrane forces clearly play an increasingly significant role as the load is intensified, resulting in a pronounced overall stiffening of the structure. In Figure 79 we show the undeformed and various deformed mid-surface configurations for the open-ended cylindrical shell pull-out problem. The overall deflections and rotations are clearly quite large, especially for the final shell configuration (i.e., the case where $P = 40{,}000$).

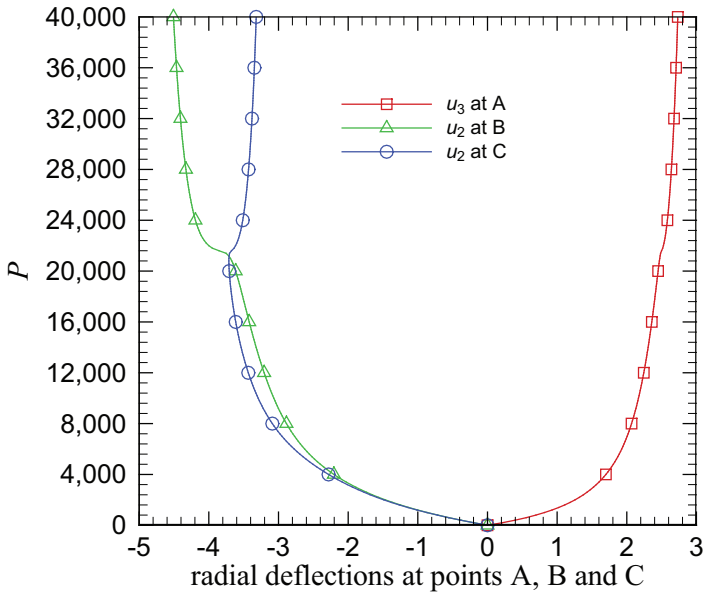


Fig. 78. Radial deflections at points A, B and C vs. pull-out force P for the open-ended cylindrical shell.

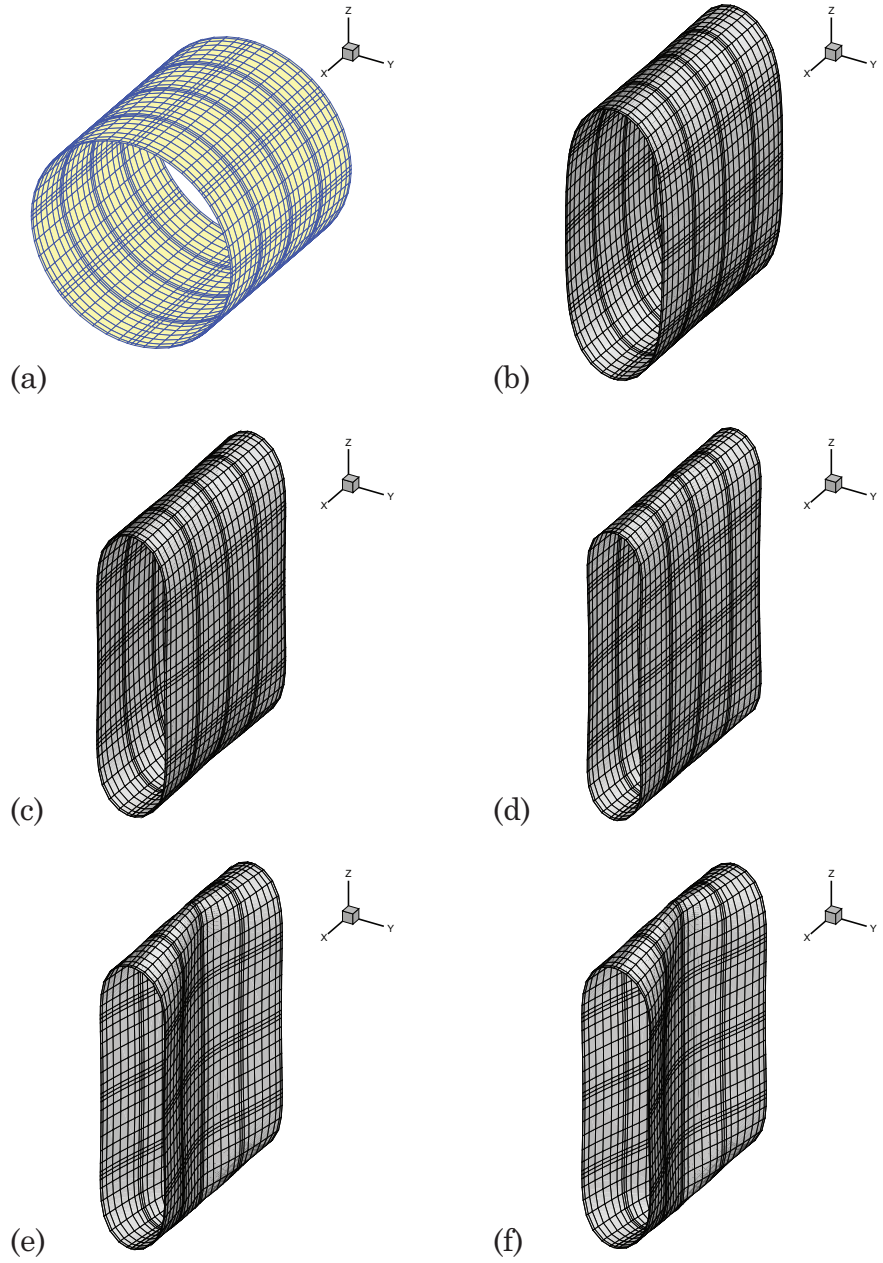


Fig. 79. Undeformed and various deformed mid-surface configurations of the open-ended cylindrical shell: (a) undeformed configuration, (b) P = 5,000, (c) P = 10,000, (d) P = 20,000, (e) P = 30,000 and (f) P = 40,000.

6. A pinched half-cylindrical shell

In this next example we consider a half-cylindrical shell subjected to a single point force P as shown in Figure 80. Numerical solutions for this problem may be found in Refs. [153, 155, 147, 148, 149, 124] among others. We employ the following material properties for the isotropic and laminated composite versions of the problem

$$\text{Isotropic:} \qquad E = 2.0685 \times 10^7, \quad \nu = 0.3 \tag{6.83a}$$

$$\text{Orthotropic:} \begin{cases} E_1 = 2{,}068.5, & E_2 = E_3 = 517.125 \\ G_{23} = 198.8942, & G_{13} = G_{12} = 795.6 \\ \nu_{23} = 0.3, & \nu_{13} = \nu_{12} = 0.3 \end{cases} \tag{6.83b}$$

The geometric parameters are taken as $L = 304.8$, $R = 101.6$ (where R is the radius of the mid-surface) and $h = 3.0$.

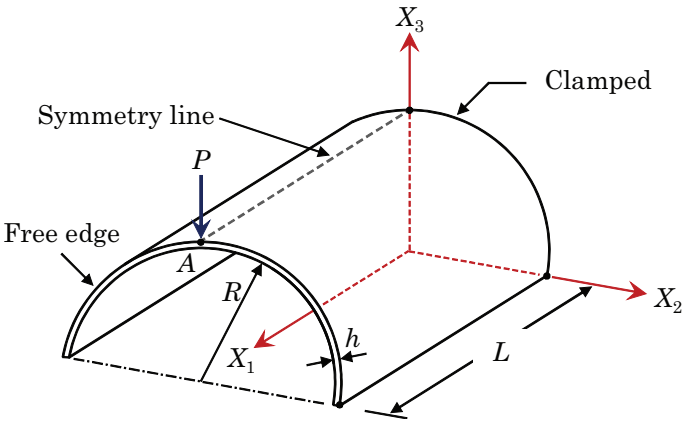


Fig. 80. A half-cylindrical shell subjected to a single point load.

As in the previous example, we exploit symmetry of the problem by performing the finite element simulations using half of the physical domain of the shell (see the

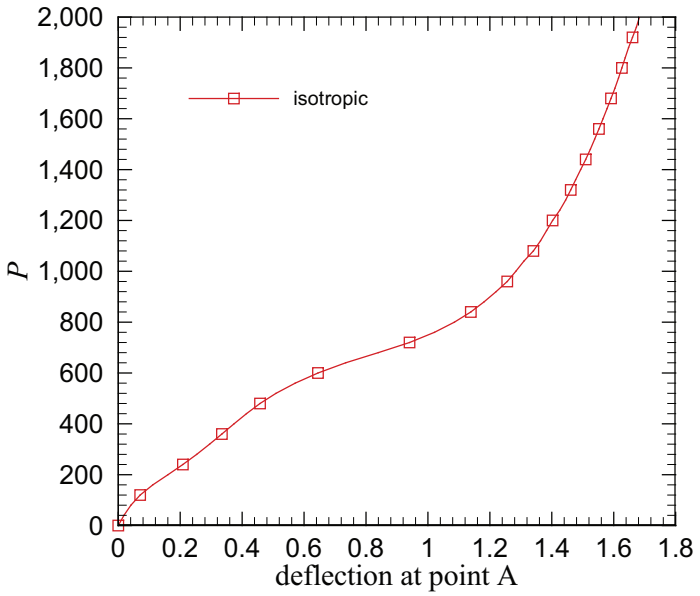


Fig. 81. Vertical deflection at point A of an isotropic half-cylindrical shell under point loading.

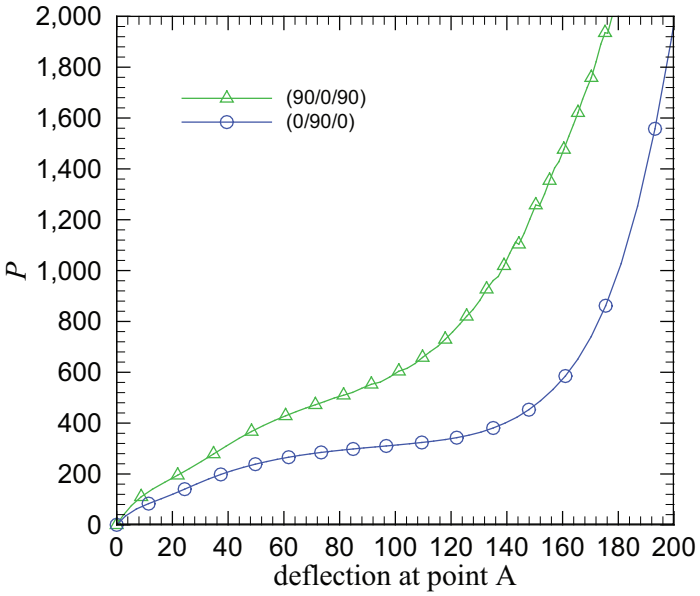


Fig. 82. Vertical deflection at point A of two laminated composite half-cylindrical shells under point loading.

line of symmetry shown in Figure 80); a regular 4×4 mesh is adopted for each simulation with the p-level taken as 8. For the support boundary conditions along the bottom longitudinal edges, we take the vertical deflection and X_3 component of the difference vector as zero. For the laminated composite simulations, with stacking sequences given as: $(90°/0°/90°)$ and $(0°/90°/0°)$, we prescribe the unit tangent vector as $\hat{\mathbf{t}} = \hat{\mathbf{E}}_1$.

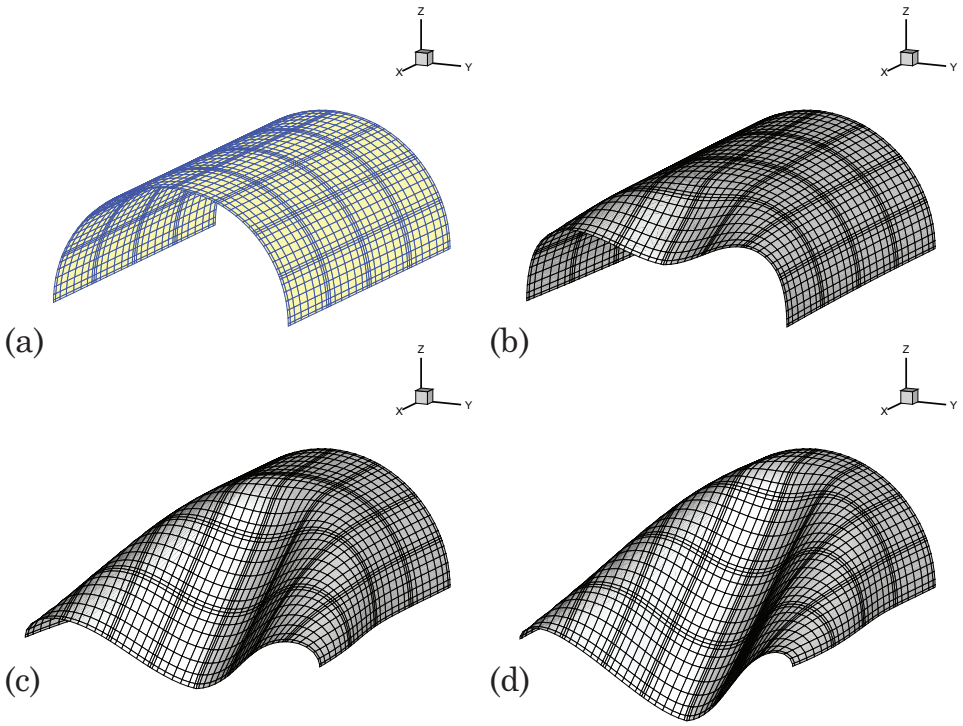


Fig. 83. Undeformed and various deformed mid-surface configurations of the isotropic pinched half-cylindrical shell: (a) undeformed configuration, (b) $P = 600$, (c) $P = 1{,}200$ and (d) $P = 2{,}000$.

In Figures 81 and 82 we trace the vertical displacements at point A of the isotropic and laminated composite cylinders. The cylindrical arc-length procedure has been used in each numerical simulation to smoothly traverse the limit points. We find that the computed displacements agree well with the results reported by Arciniega and

Reddy [124]. Finally, in Figure 83 we show the undeformed and various deformed mid-surface configurations of the pinched isotropic half-cylinder.

7. A pinched hemisphere with an 18° hole

We now consider a pinched isotropic hemisphere with an 18° circular cutout. This problem is widely recognized as one of the most severe shell benchmark problems involving finite deformations and has been addressed by many researchers (see for example Refs. [144, 126, 157, 155, 156, 147, 148, 149] among others). The computational domain (i.e., one quarter of the hemisphere) is shown in Figure 84. The external loads for the problem consist of four alternating radial point forces P, prescribed along the equator at 90° intervals. The mid-surface radius and shell thickness are taken as $R = 10.0$ and $h = 0.04$ respectively; furthermore, the material properties are prescribed as

$$E = 6.825 \times 10^7, \quad \nu = 0.3 \tag{6.84}$$

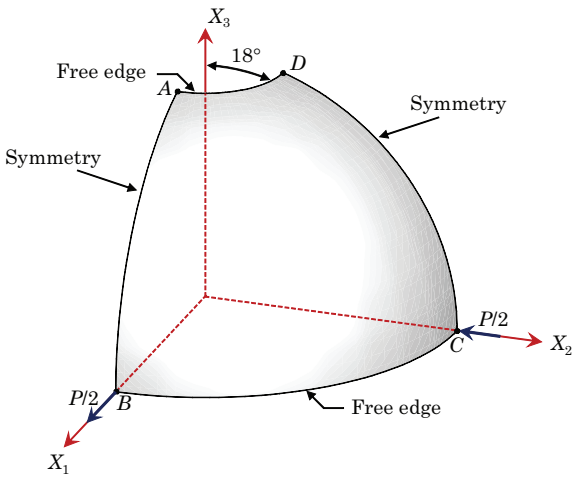


Fig. 84. A pinched hemisphere with an 18° hole (the computational domain shown above is one quarter of the physical domain of the shell).

For the finite element mesh we employ a regular 8×8 discretization with the p-level taken as 4. The finite element nodes on the mid-surface $\bar{\Omega}^{hp}$ are obtained using the following formula

$$\begin{aligned} \underline{\mathbf{X}} = R \Big\{ & \sin[\alpha + (\pi/2 - \alpha)\omega^1][\cos(\pi\omega^2/2)\hat{\mathbf{E}}_1 + \sin(\pi\omega^2/2)\hat{\mathbf{E}}_2] \\ & + \cos[\alpha + (\pi/2 - \alpha)\omega^1]\hat{\mathbf{E}}_3 \Big\} \end{aligned} \tag{6.85}$$

where $\alpha = 18^\circ = \pi/10$ rad. The incremental/iterative Newton method is used in the solution procedure with 80 load steps and $P_{\max}$ is taken as 400. In addition to the symmetry boundary conditions, we also require the X_3 component of the displacement of the node located at point B to be zero.

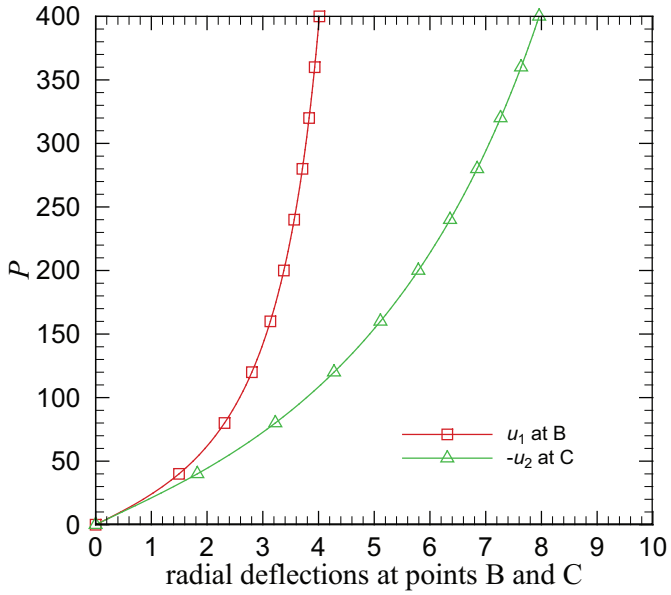


Fig. 85. Radial deflections at points B and C of the pinched hemisphere.

Figure 85 shows the radial deflections at points B and C vs. the applied pinching force P. Our reported deflections compare quite well with the numerical results tabulated by Sze et al. [149]. In Figure 86 we show the undeformed and three deformed

mid-surface configurations of the pinched hemisphere.

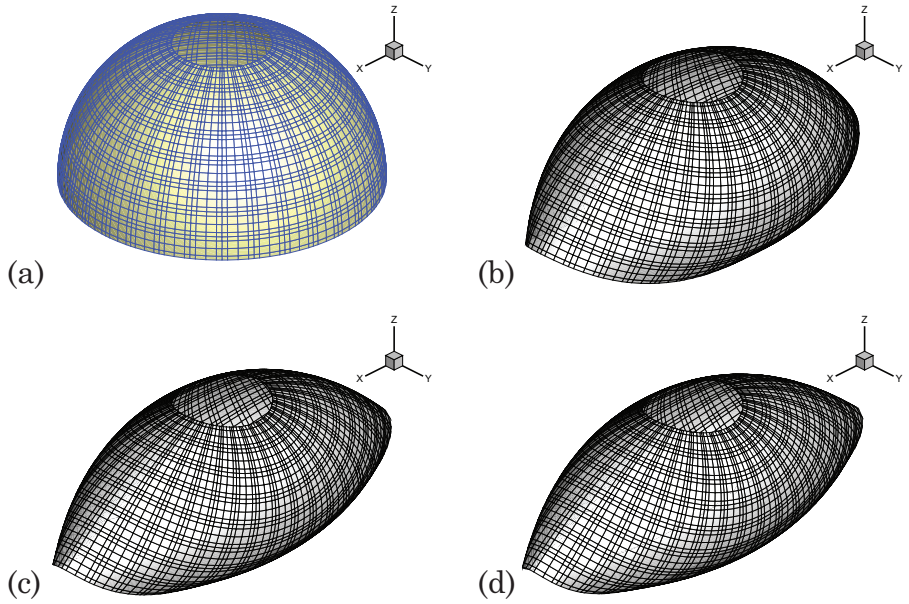


Fig. 86. Undeformed and various deformed mid-surface configurations of the pinched hemispherical shell: (a) undeformed configuration, (b) $P = 150$, (c) $P = 300$ and (d) $P = 400$.

8. A pinched composite hyperboloidal shell

As a final numerical example we consider the finite deformation of a laminated composite hyperboloidal shell that is loaded by four alternating radial point forces P. This challenging benchmark, originally proposed by Başar et al. [108], was designed to test the capabilities of shell elements in handling geometrically complex shell structures undergoing very large displacements and rotations. The problem has been considered by Wagner and Gruttmann [158], Balah and Ghamedy [159] and more recently by Arciniega and Reddy [124]. The computational domain (i.e., one octant of the actual hyperboloid) is shown in Figure 87. The orthotropic material properties for each

lamina are taken as

$$\begin{aligned} E_1 &= 40.0 \times 10^6, \quad & E_2 &= E_3 = 1.0 \times 10^6 \\ G_{23} &= 0.6 \times 10^6, \quad & G_{13} &= G_{12} = 0.6 \times 10^6 \\ \nu_{23} &= 0.25, & \nu_{13} &= \nu_{12} = 0.25 \end{aligned} \tag{6.86}$$

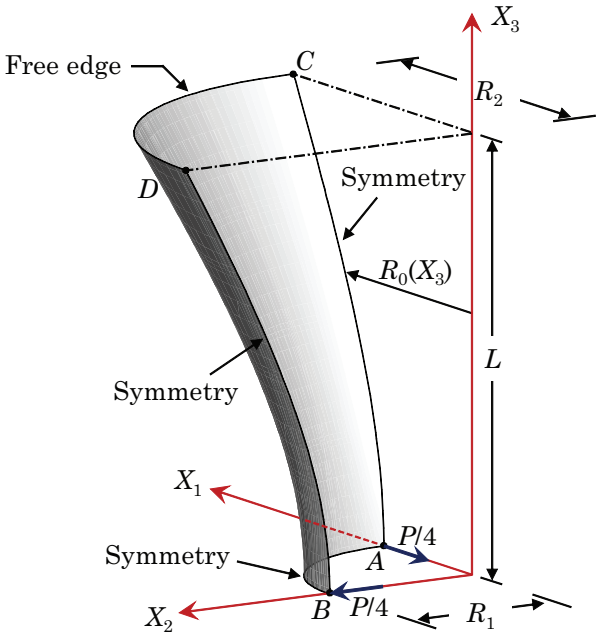


Fig. 87. A pinched laminated composite hyperboloidal shell (the computational domain shown above is one octant of the physical domain of the shell).

We employ three finite element discretizations of the computational domain (see Figure 88) including: a structured 4×4 mesh, an unstructured 4×4 mesh and a structured 5×5 mesh; where in all cases the p-level is taken as 8. The unstructured mesh is utilized to showcase the ability of the proposed shell element to accurately solve nontrivial laminated composite shell problems using skewed elements. Each mesh is generated by mapping the nodal coordinates of an appropriate conforming

discretization of $\bar{\boldsymbol{\omega}} = [0,1]^2$ onto the finite element approximation of the mid-surface $\bar{\Omega}^{hp}$ of the composite hyperboloid using the following formula

$$\underline{\mathbf{X}} = R_0(\omega^2)\left[\cos(\pi\omega^1/2)\hat{\mathbf{E}}_1 + \sin(\pi\omega^1/2)\hat{\mathbf{E}}_2\right] + L\omega^2\hat{\mathbf{E}}_3 \tag{6.87}$$

where $R_0(\omega^2) = R_1\sqrt{1+(L\omega^2/C)^2}$. The geometric parameters are taken as $R_1 = 7.5$, $C = 20/\sqrt{3}$, $L = 20.0$ and $h = 0.04$. The unit normal and tangent vectors are defined at the finite element nodes using the following expressions

$$\hat{\mathbf{n}} = \frac{\partial\underline{\mathbf{X}}/\partial\omega^1 \times \partial\underline{\mathbf{X}}/\partial\omega^2}{||\partial\underline{\mathbf{X}}/\partial\omega^1 \times \partial\underline{\mathbf{X}}/\partial\omega^2||} \tag{6.88a}$$

$$\hat{\mathbf{t}} = -\sin(\pi\omega^1/2)\hat{\mathbf{E}}_1 + \cos(\pi\omega^1/2)\hat{\mathbf{E}}_2 \tag{6.88b}$$

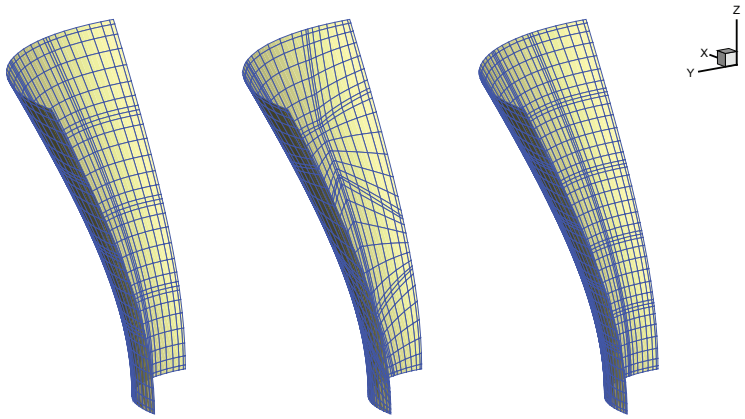


Fig. 88. Finite element discretizations of the composite hyperboloid, where the p-level is 8: (a) a 4×4 structured discretization, (b) a 4×4 unstructured discretization and (c) a 5×5 structured discretization.

In Figures 89 and 90 we show various displacement components vs. the applied load P at points A, B, C and D of the hyperboloidal shell for the composite lamination

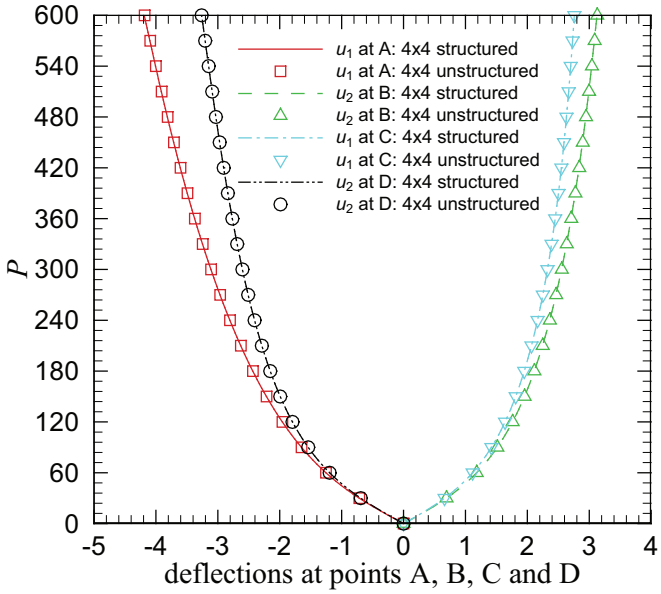


Fig. 89. Deflections at points A, B, C and D of the pinched (0°/90°/0°) stacking sequence laminated composite hyperboloidal shell.

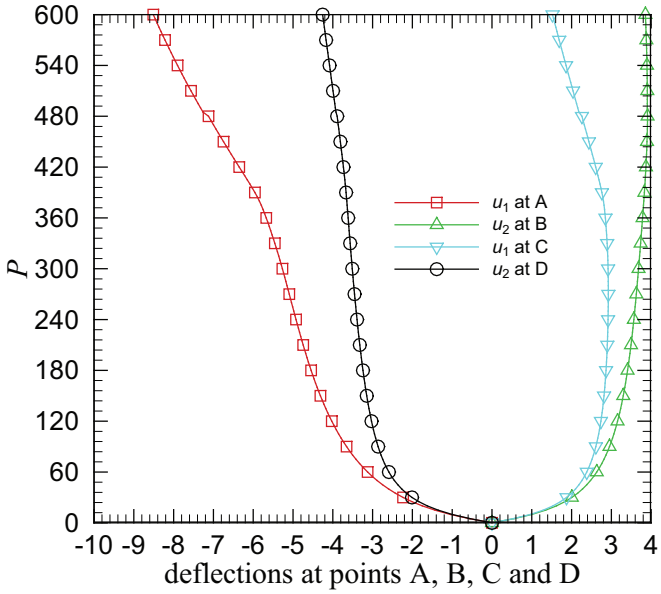


Fig. 90. Deflections at points A, B, C and D of the pinched (90°/0°/90°) stacking sequence laminated composite hyperboloidal shell.

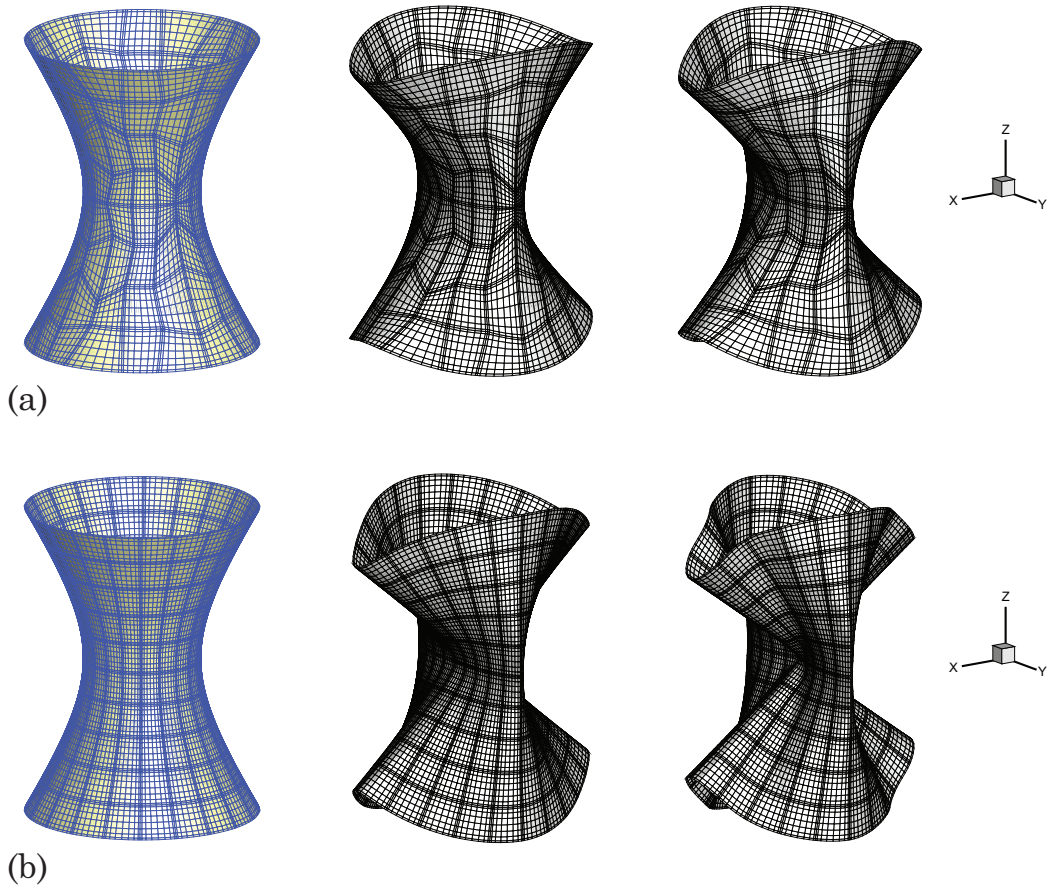


Fig. 91. Undeformed and various deformed mid-surface configurations of two pinched laminated composite hyperboloidal shells: (a) (0°/90°/0°): undeformed configuration, $P = 250$ and $P = 500$ (from left to right) and (b) (90°/0°/90°): undeformed configuration, $P = 250$ and $P = 500$ (from left to right).

schemes: (0°/90°/0°) and (90°/0°/90°). The computed displacements for the stacking sequence (0°/90°/0°), obtained using both structured and unstructured meshes, are in excellent agreement with the results of Başar et al. [108] and Arciniega and Reddy [124]. The displacements calculated for the laminate (90°/0°/90°) (obtained using the regular 5×5 discretization shown in Figure 88 (c)), however, are greater than the values reported by Başar et al. [108] but are also somewhat less than the results

obtained by Arciniega and Reddy [124]. In Figure 91 we show various mid-surface configurations of the hyperboloid for each composite laminate. All numerical results have been obtained via the incremental/iterative Newton procedure using 120 load steps.

CHAPTER VII

CONCLUSIONS

A. Summary and concluding remarks

In this dissertation we have presented finite element formulations for fluid and solid mechanics problems using high-order spectral/*hp* finite element technology. Our aim throughout this work *has not* been to indiscriminately champion high-order finite element procedures in the numerical simulation of all phenomena associated with these disciplines. On the contrary, our primary objective has been to adopt novel mathematical models and innovative discretization procedures in the numerical simulation of fluids and solids, wherein the additional benefits of employing high-order spectral/*hp* finite element technology are pronounced. We find that for many such problem sets (especially those whose weak formulations may be identified as global minimizers), high-order finite element procedures offer the prospect of highly accurate numerical solutions that are completely devoid of all forms of locking. As a result ad-hoc tricks (e.g., reduced integration and/or mixed interpolation) required to stabilize low-order finite element formulations are unnecessary.

An overview of the steps involved in developing and arriving at efficient finite element models using spectral/*hp* finite element technology were presented in Chapter II. The presentation was quite general and therefore applicable to finite element problems posed in 1, 2 or 3 spatial dimensions. A notable contribution contained in this chapter was a sparse global finite element assembly operator that admits parallelization on shared-memory computer systems using the OpenMP paradigm. To improve system memory requirements, we implemented an element-level static condensation technique, wherein the interior degrees of freedom of each element were

implicitly eliminated prior to assembly of the global system of finite element equations. Robustness and efficiency of the proposed global assembly operator and static condensation technique were assessed through the numerical simulation of a finite element problem possessing nearly half a million total degrees of freedom.

Chapter III was one of two chapters devoted to finite element models formulated using the least-squares method. In this chapter we examined the roles of minimization and linearization in least-squares finite element models of nonlinear boundary-value problems. As the least-squares method is independent of: (a) the adopted discretization procedure and (b) the chosen solution scheme, the underlying least-squares principle demands that minimization of the least-squares functional be performed prior to linearization (where linearization is introduced in the context of an appropriate fixed point iterative solution procedure). With this in mind, we discussed practical consequences associated with exchanging the order of application of the minimization and linearization operations in least-squares finite element models of nonlinear boundary-value problems. In the analysis we relied on an examination of the abstract mathematical setting of the least-squares method, a simple analysis of Newton's procedure as applied to least-squares problems and on qualities observed in numerical experiments. Overall we find that although the least-squares principle suggests that minimization ought to be performed prior to linearization, such an approach is often impractical and not necessary.

In Chapter IV we presented a novel least-squares finite element formulation of the steady-state and non-stationary incompressible Navier-Stokes equations with enhanced local mass conservation. The proposed formulation was a modification of the standard L_2-norm least-squares formulation of the Navier-Stokes equations based on the equivalent velocity-pressure-vorticity first-order system. In the new formulation, we modified the standard least-squares functional to also include an appropriately

penalized sum of the squares of the element-level integrals of a regularized form of the continuity equation. As a consequence, the resulting finite element model directly inherited terms in the bilinear form and linear functional (which could be adjusted based on the penalty parameter) that tend to improve element-level mass conservation. A notable quality of the formulation was that improved mass conservation could be attained without introducing additional variables or compromising the unconstrained minimization setting for the numerical solution. Numerical simulations confirmed that the proposed formulation could significantly improve mass conservation for both steady and non-stationary flows. For transient flows, the formulation was further shown to enhance velocity-pressure coupling and overall numerical stability (most notably for cases where the momentum equation residual, appearing in the least-squares functional, was not weighted by the square of the time step).

Chapters V and VI were devoted to solid mechanics problems. In Chapter V we presented finite element models of viscoelastic beam structures based on the kinematic hypotheses of the Euler-Bernoulli, Timoshenko and third-order Reddy beam theories. The formulations (valid for beams undergoing moderately large transverse displacements and rotations) were obtained by replacing the Green-Lagrange strain with the von Kármán strain in both the constitutive equations and the virtual work statement. The linear viscoelastic constitutive equations, taken in convolution form, were temporally discretized using a two-point recurrence formula. High polynomial order Hermite basis functions were introduced in the interpolation of the transverse displacements for the Euler-Bernoulli and third-order Reddy beam theories. These high-order, globally C^1 continuous interpolants, were prescribed using the standard GLL points (see Chapter II) as the nodal locations for the master element. Carefully chosen quasi-static and fully transient benchmark example problems were solved to showcase the insensitivity of each beam element to both membrane and shear locking.

Finally in Chapter VI we proposed a high-order spectral/*hp* continuum shell finite element for the numerical simulation of the *fully* finite deformation mechanical response of isotropic, laminated composite and functionally graded elastic shell structures. The shell element was based on a modified first-order shell theory using a 7-parameter expansion of the displacement field. The seventh parameter was included to allow for the use of fully three-dimensional constitutive equations in the numerical implementation. The finite element coefficient matrices and force vectors were evaluated numerically using appropriate high-order Gauss-Legendre quadrature rules at the appropriate quadrature points of the element mid-surface. The virtual work statement was further integrated numerically through the shell thickness at each quadrature point of the mid-surface; hence no thin-shell approximations were imposed in the numerical implementation. For laminated composite shells, we introduced a user prescribed vector field (defined at the nodes) tangent to the shell mid-surface. This discrete tangent vector allowed for simple construction of the local bases associated with the principle orthotropic material directions of each lamina. As a result, we were free to employ skewed and/or arbitrarily curved elements in actual finite element simulations. We demonstrated, through the numerical simulation of carefully chosen benchmark problems, that the proposed shell element was insensitive to all forms of numerical locking and severe geometric distortions.

B. Topics of ongoing and future research

In this dissertation we have considered applications of high-order spectral/*hp* finite element technology to problems posed in the fields of fluid mechanics and solid mechanics. Our discussion has thus far been limited to a study of these topics in isolation from each another. Of particular interest going forward is the combination of our de-

veloped finite element technology to problems involving the fully coupled interaction of fluids and solids (i.e., fluid-structure interaction). In what follows we briefly describe an example problem that is solved using a least-squares finite element model of the incompressible Navier-Stokes equations that can handle moving fluid boundaries (a preliminary step towards implementation of a general purpose fluid-structure interaction code).

We consider the flow of a viscous incompressible fluid inside a square cavity, where the cavity under consideration is a bi-unit square centered at the origin. A 0.28 units diameter solid circular cylinder is positioned at the origin of the cavity at $t = 0$. Immediately following $t = 0$, the cylinder begins to translate with an instantaneous unit velocity in the x direction. We impose no-slip type boundary conditions along all solid surfaces, including the cylinder Γ_{cyl} and cavity walls Γ_{walls}, where $\Gamma = \Gamma_{\text{cyl}} \cup \Gamma_{\text{walls}}$. This amounts to specifying $\mathbf{v} = \mathbf{0}$ on Γ_{walls} and $v_x = 1$ on Γ_{cyl}. We prescribe the pressure to be zero at the single node located at $(x, y) = (-1, 0)$. The Reynolds number for the flow is taken as $\text{Re} = 100$ by specifying $\rho = 1$, $\mu = 1/100$ and a characteristic unit length. The initial boundary-value problem is posed on the time interval $t = (0, 0.7]$.

The computational domain is discretized into 480 non-uniform finite elements, where we place 40 element layers along the circumference of the cylinder and 12 in the radial direction. A depiction of the finite element mesh at $t = 0$ and $t = 0.70$ is shown in Figure 92. Mesh refinement is employed near the cylinder to ensure acceptable numerical resolution of all variables in the wake region downstream of the cylinder. We solve the problem using a p-level of 4 in each finite element, which amounts to 31,360 total degrees of freedom in the numerical model. We employ Newton's method to linearize the finite element equations and adopt a time step size of $\Delta t = 0.005$. The α-family of time approximation is utilized in the temporal discretization (with α

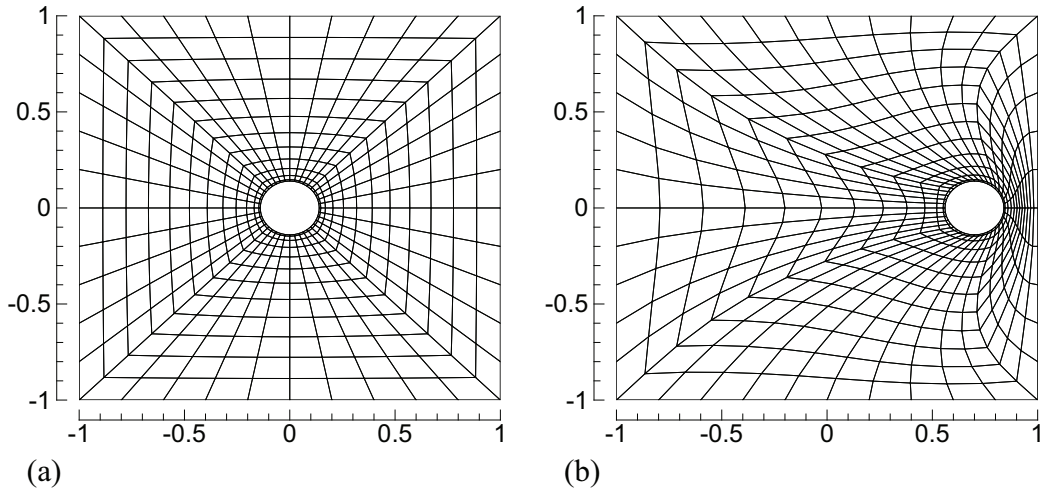


Fig. 92. Finite element discretization of computational domain $\bar{\Omega}$ for the analysis of transient incompressible flow inside a square cavity induced by the motion of a circular cylinder: (a) fluid mesh at $t = 0$ and (b) fluid mesh at $t = 0.70$.

taken as 0.5); the first-order backward difference scheme is employed in the first few iterations.

A word on the adopted fluid mesh motion scheme is in order. The evolution of the deforming fluid mesh is determined at each time step using a standard pseudo-elasticity formulation (see Belytschko et al. [107]) that is implemented in conjunction with the arbitrary Lagrangian Eulerian (ALE) formulation. In this approach, we solve a linear elasticity boundary-value problem with Dirichlet boundary conditions at each time step on the fluid domain. The position of the cylinder at the current time step is used directly as a boundary condition in the mesh motion scheme to determine the new locations and velocities of the nodes of the fluid mesh. A weak-form Galerkin finite element model of the pseudo-elasticity equations is employed. To prevent excessive distortion of elements in the model we specify the Young's modulus for

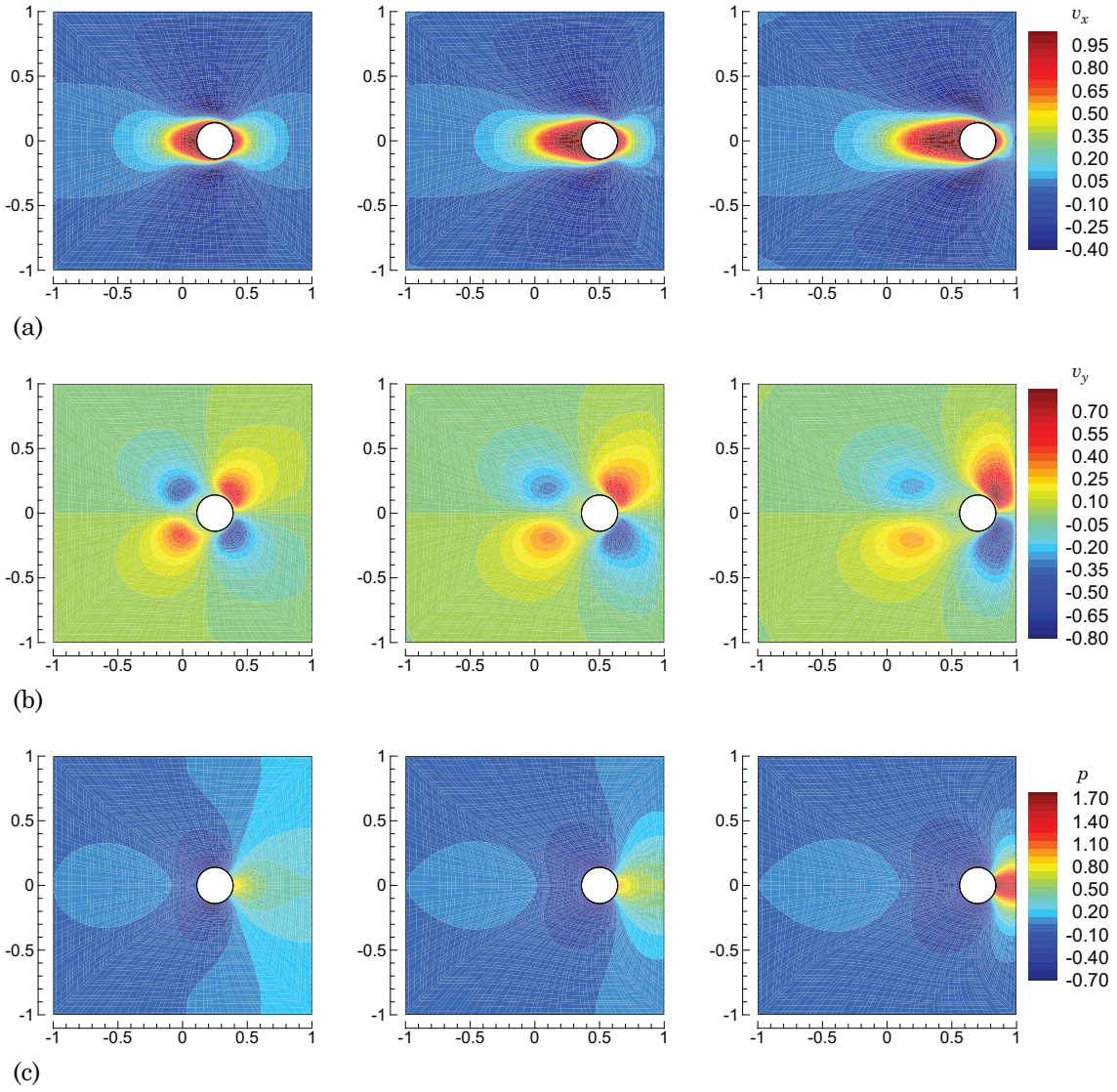


Fig. 93. Transient flow of an incompressible viscous fluid inside a square cavity induced by a moving cylinder at $t = 0.25, 0.50$ and 0.70 (from left to right respectively): (a) velocity component v_x (b) velocity component v_y and (c) non-dimensional pressure field p.

the eth finite element as $E_e = E_0 \mu(\Omega^e)^{-0.5}$ where $\mu(\Omega^e)$ is the area of the element; the strictly-positive quantity E_0 is arbitrary. In Figure 93 we present snapshots of the numerical results for the velocity components and pressure at $t = 0.25$, 0.50 and 0.70. We see that at this Reynolds number the flow field is symmetric about the y-axis. Our numerical results agree well with the high-order weak-form spectral element solution presented by Bodard et al. [160]. Based on these preliminary results, the prospect of extending our work to the numerical simulation of fluid-structure interaction problems using high-order spectral/hp finite element procedures appears promising.

In addition to the fluid-structure interaction computational technology, the shell element developed herein may also be extended to the analysis of laminated composite shell structures experiencing: (a) inelastic response, (b) fracture and (c) damage. These topics are of particular interest to the structures community.

REFERENCES

[1] M. J. Turner, R. W. Clough, H. C. Martin, and L. J. Topp, "Stiffness and deflection analysis of complex structures," *Journal of the Aeronautical Sciences*, vol. 23, no. 9, pp. 805–854, 1956.

[2] P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods*, Springer, New York, 2009.

[3] J. N. Reddy, *Energy Principles and Variational Methods in Applied Mechanics*, John Wiley and Sons, Ltd, New York, 2nd edition, 2002.

[4] T. J. R. Hughes and A. N. Brooks, "Multi-dimensional upwind scheme with no crosswind diffusion," *American Society of Mechanical Engineers, Applied Mechanics Division, AMD*, vol. 34, pp. 19–35, 1979.

[5] A. N. Brooks and T. J. R. Hughes, "Streamline upwind/Petrov-Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier-Stokes equations," *Comput. Methods Appl. Mech. Engrg.*, vol. 32, no. 1–3, pp. 199–259, 1982.

[6] J. N. Reddy, "Penalty-finite-element analysis of 3-D Navier-Stokes equations," *Comput. Methods Appl. Mech. Engrg.*, vol. 35, pp. 87–97, 1982.

[7] J. N. Reddy and D. K. Gartling, *The Finite Element Method in Heat Transfer and Fluid Dynamics*, CRC Press, Boca Raton, FL, 3rd edition, 2010.

[8] T. J. R. Hughes, L. P. Franca, and G. M. Hulbert, "A new finite element formulation for computational fluid dynamics: VIII. The Galerkin/least-squares method for advective-diffusive equations," *Comput. Methods Appl. Mech. Engrg.*, vol. 73, no. 2, pp. 173–189, 1989.

[9] B. N. Jiang and C. L. Chang, "Least-squares finite elements for the Stokes problem," *Comput. Methods Appl. Mech. Engrg.*, vol. 78, no. 3, pp. 297–311, 1990.

[10] B. N. Jiang and L. A. Povinelli, "Least-squares finite element method for fluid dynamics," *Comput. Methods Appl. Mech. Engrg.*, vol. 81, no. 1, pp. 13–37, 1990.

[11] B. N. Jiang and G. F. Carey, "Least-squares finite element methods for compressible Euler equations," *Int. J. Numer. Meth. Fluids*, vol. 10, pp. 557–568, 1990.

[12] B. N. Jiang, "A least-squares finite element method for incompressible Navier-Stokes problems," *Int. J. Numer. Meth. Fluids*, vol. 14, no. 7, pp. 843–859, 1992.

[13] B. N. Jiang, T. L. Lin, and L. A. Povinelli, "Large-scale computation of incompressible viscous flow by least-squares finite element method," *Comput. Methods Appl. Mech. Engrg.*, vol. 114, no. 3–4, pp. 213–231, 1994.

[14] B. N. Jiang, "On the least-squares method," *Comput. Methods Appl. Mech. Engrg.*, vol. 152, no. 1–2, pp. 239–257, 1998.

[15] J. P. Pontaza and J. N. Reddy, "Spectral/*hp* least-squares finite element formulation for the Navier-Stokes equations," *J. Comput. Phys.*, vol. 190, no. 2, pp. 523–549, 2003.

[16] J. P. Pontaza, X. Diao, J. N. Reddy, and K. S. Surana, "Least-squares finite element models of two-dimensional compressible flows," *Finite Elem. Anal. Des.*, vol. 40, no. 5–6, pp. 629–644, 2004.

[17] J. P. Pontaza and J. N. Reddy, "Space-time coupled spectral/*hp* least-squares finite element formulation for the incompressible Navier-Stokes equations," *J. Comput. Phys.*, vol. 197, no. 2, pp. 418–459, 2004.

[18] J. P. Pontaza and J. N. Reddy, "Least-squares finite element formulations for viscous incompressible and compressible fluid flows," *Comput. Methods Appl. Mech. Engrg.*, vol. 195, no. 19–22, pp. 2454–2494, 2006.

[19] P. B. Bochev and M. D. Gunzburger, "Finite element methods of least-squares type," *SIAM Rev.*, vol. 40, no. 4, pp. 789–837, 1998.

[20] G. S. Payette and J. N. Reddy, "On the roles of minimization and linearization in least-squares finite element models of nonlinear boundary-value problems," *J. Comput. Phys.*, vol. 230, no. 9, pp. 3589–3613, 2011.

[21] G. S. Payette and J. N. Reddy, "Nonlinear quasi-static finite element formulations for viscoelastic Euler-Bernoulli and Timoshenko beams," *Comm. Numer. Meth. Eng.*, vol. 26, no. 12, pp. 1736–1755, 2010.

[22] G. E. Karniadakis and S. J. Sherwin, *Spectral/hp Element Methods for CFD*, Oxford University Press, Oxford, 1999.

[23] J. N. Reddy, *Applied Functional Analysis And Variational Methods in Engineering*, McGraw-Hill, New York, 1986.

[24] J. N. Reddy, *An Introduction to the Finite Element Method*, McGraw-Hill, New York, 3rd edition, 2006.

[25] K.-J. Bathe, *Finite Element Procedures*, Prentice Hall, Upper Saddle River, New Jersey, 1996.

[26] W. Couzy and M. O. Deville, "A fast Schur complement method for the spectral element discretization of the incompressible Navier-Stokes equations," *J. Comput. Phys.*, vol. 116, no. 1, pp. 135–142, 1995.

[27] T. A. Davis and I. S. Duff, "An unsymmetric-pattern multifrontal method for sparse LU factorization," *SIAM J. Matrix Anal. Appl.*, vol. 18, pp. 140–158, 1997.

[28] T. A. Davis and I. S. Duff, "A combined unifrontal/multifrontal method for unsymmetric sparse matrices," *ACM Trans. Math. Softw.*, vol. 25, pp. 1–20, 1999.

[29] T. A. Davis, "A column pre-ordering strategy for the unsymmetric-pattern multifrontal method," *ACM Trans. Math. Softw.*, vol. 30, pp. 165–195, 2004.

[30] T. A. Davis, "Algorithm 832: UMFPACK V4.3–an unsymmetric-pattern multifrontal method," *ACM Trans. Math. Softw.*, vol. 30, pp. 196–199, 2004.

[31] F. Brezzi and J. Douglas, "Stabilized mixed methods for the Stokes problem," *Numer. Math.*, vol. 53, pp. 225–235, 1988.

[32] P. Raviart V. Girault, *Finite Element Methods for Navier-Stokes Equations*, Springer, Berlin, 1986.

[33] J. H. Bramble and A. H. Schatz, "Least squares methods for 2mth order elliptic boundary-value problems," *Math. Comput.*, vol. 25, no. 113, pp. 1–32, 1971.

[34] E. D. Eason, "A review of least-squares methods for solving partial differential equations," *Int. J. Numer. Meth. Engng*, vol. 10, no. 5, pp. 1021–1046, 1976.

[35] B. N. Jiang, *The Least-Squares Finite Element Method*, Springer-Verlag, New York, 1998.

[36] P. B. Bochev, "Analysis of least-squares finite element methods for the Navier-Stokes equations," *SIAM J. Numer. Anal.*, vol. 34, no. 5, pp. 1817–1844, 1997.

[37] S. Agmon, A. Douglis, and L. Nirenberg, "Estimates near the boundary for solutions of elliptic partial differential equations satisfying general boundary conditions II," *Commun. Pure Appl. Math.*, vol. 17, no. 1, pp. 35–92, 1964.

[38] P. Bochev, Z. Cai, T. A. Manteuffel, and S. F. McCormick, "Analysis of velocity-flux first-order system least-squares principles for the Navier-Stokes equations: Part I," *SIAM J. Numer. Anal.*, vol. 35, no. 3, pp. 990–1009, 1998.

[39] P. Bochev, T. A. Manteuffel, and S. F. McCormick, "Analysis of velocity-flux least-squares principles for the Navier-Stokes equations: Part II," *SIAM J. Numer. Anal.*, vol. 36, no. 4, pp. 1125–1144, 1999.

[40] J. H. Bramble and J. E. Pasciak, "Least-squares methods for Stokes equations based on a discrete minus one inner product," *J. Comput. Appl. Math.*, vol. 74, no. 1–2, pp. 155–173, 1996.

[41] B. N. Jiang and V. Sonnad, "Least-squares solution of incompressible Navier-Stokes equations with the p-version of finite elements," *Comput. Mech.*, vol. 15, pp. 129–136, 1994.

[42] B. C. Bell and K. S. Surana, "A space-time coupled p-version least-squares finite element formulation for unsteady fluid dynamics problems," *Int. J. Numer. Meth. Engng*, vol. 37, no. 20, pp. 3545–3569, 1994.

[43] B. C. Bell and K. S. Surana, "A space-time coupled p-version least squares finite element formulation for unsteady two-dimensional Navier-Stokes equations," *Int. J. Numer. Meth. Engng*, vol. 39, no. 15, pp. 2593–2618, 1996.

[44] M. M. J. Proot and M. I. Gerritsma, "A least-squares spectral element formulation for the Stokes problem," *J. Sci. Comput.*, vol. 17, no. 1–4, pp. 285–296, 2002.

[45] M. M. J. Proot and M. I. Gerritsma, "Least-squares spectral elements applied to the Stokes problem," *J. Comput. Phys.*, vol. 181, no. 2, pp. 454–477, 2002.

[46] M. I. Gerritsma and M. M. J. Proot, "Analysis of a discontinuous least squares spectral element method," *J. Sci. Comput.*, vol. 17, no. 1–4, pp. 297–306, 2002.

[47] B. De Maerschalck, M. I. Gerritsma, and M. M. J. Proot, "Space-time least-squares spectral elements for convection-dominated unsteady flows," *AIAA Journal*, vol. 44, no. 3, pp. 558–565, 2006.

[48] G. F. Carey and B. N. Jiang, "Least-squares finite element method and preconditioned conjugate gradient solution," *Int. J. Numer. Meth. Engng*, vol. 24, no. 7, pp. 1283–1296, 1987.

[49] G. F. Carey and B. N. Jiang, "Nonlinear preconditioned conjugate gradient and least-squares finite elements," *Comput. Methods Appl. Mech. Engrg.*, vol. 62, no. 2, pp. 145–154, 1987.

[50] J. N. Reddy, *An Introduction to Nonlinear Finite Element Analysis*, Oxford University Press, Oxford, 2004.

[51] C. L. Chang and J. J. Nelson, "Least-squares finite element method for the Stokes problem with zero residual of mass conservation," *SIAM J. Numer. Anal.*, vol. 34, no. 2, pp. 480–489, 1997.

[52] J. P. Pontaza, "A least-squares finite element formulation for unsteady incompressible flows with improved velocity-pressure coupling," *J. Comput. Phys.*,

vol. 217, no. 2, pp. 563–588, 2006.

[53] V. Prabhakar and J. N. Reddy, "A stress-based least-squares finite-element model for incompressible Navier-Stokes equations," *Int. J. Numer. Meth. Fluids*, vol. 54, no. 11, pp. 1369–1385, 2007.

[54] V. Prabhakar and J. N. Reddy, "Spectral/*hp* penalty least-squares finite element formulation for the steady incompressible Navier-Stokes equations," *J. Comput. Phys.*, vol. 215, no. 1, pp. 274–297, 2006.

[55] V. Prabhakar and J. N. Reddy, "Spectral/*hp* penalty least-squares finite element formulation for unsteady incompressible flows," *Int. J. Numer. Meth. Fluids*, vol. 58, no. 3, pp. 287–306, 2008.

[56] V. Prabhakar, J. P. Pontaza, and J. N. Reddy, "A collocation penalty least-squares finite element formulation for incompressible flows," *Comput. Methods Appl. Mech. Engrg.*, vol. 197, no. 6–8, pp. 449–463, 2008.

[57] B. D. Reddy, *Introductory Functional Analysis: With Applications to Boundary Value Problems And Finite Elements*, Springer-Verlag, New York, 1997.

[58] J. P. Pontaza, "Least-squares variational principles and the finite element method: theory, formulations, and models for solid and fluid mechanics," *Finite Elem. Anal. Des.*, vol. 41, no. 7–8, pp. 703–728, 2005.

[59] J. Gerlach, "Accelerated convergence in Newton's method," *SIAM Rev.*, vol. 36, pp. 272–276, June 1994.

[60] D. J. Tritton, "Experiments on the flow past a circular cylinder at low Reynolds numbers," *J. Fluid Mech.*, vol. 6, pp. 547–567, 1959.

[61] A. S. Grove, F. H. Shair, E. E. Petersen, and A. Acrivos, "An experimental investigation of the steady separated flow past a circular cylinder," *J. Fluid Mech.*, vol. 19, pp. 60–80, 1964.

[62] M. Provansal, C. Mathis, and L. Boyer, "Bénard-von Kármán instability: transient and forced regimes," *J. Fluid Mech.*, vol. 182, pp. 1–22, 1987.

[63] M. Kawaguti and P. Jain, "Numerical study of a viscous fluid past a circular cylinder," *J. Phys. Soc. Japan*, vol. 21, pp. 2055–2062, 1966.

[64] S. C. R. Dennis and G. Z. Chang, "Numerical solutions for steady flow past a circular cylinder at Reynolds numbers up to 100," *J. Fluid Mech.*, vol. 42, pp. 471–489, 1970.

[65] B. Fornberg, "A numerical study of steady viscous flow past a circular cylinder," *J. Fluid Mech.*, vol. 98, no. 04, pp. 819–855, 1980.

[66] C. P. Jackson, "A finite-element study of the onset of vortex shedding in flow past variously shaped bodies," *J. Fluid Mech.*, vol. 182, pp. 23–45, 1987.

[67] J. Dušek, P. Le Gal, and P. Fraunié, "A numerical and theoretical study of the first Hopf bifurcation in a cylinder wake," *J. Fluid Mech.*, vol. 264, pp. 59–80, 1994.

[68] B. F. Armaly, F. Durst, J. C. F. Pereira, and B. Schonung, "Experimental and theoretical investigation of backward-facing step flow," *J. Fluid Mech.*, vol. 127, pp. 473–96, 1983.

[69] D. K. Gartling, "A test problem for outflow boundary conditions–flow over a backward-facing step," *Int. J. Numer. Meth. Fluids*, vol. 11, pp. 953–967, 1990.

[70] U. Ghia, K. N. Ghia, and C. T. Shin, "High-Re solution for incompressible flow using the Navier-Stokes equations and the multigrid method," *J. Comput. Phys.*, vol. 48, pp. 387–411, 1982.

[71] J. M. Deang and M. D. Gunzburger, "Issues related to least-squares finite element methods for the Stokes equations," *SIAM J. Sci. Comput.*, vol. 20, no. 3, pp. 878–906, 1998.

[72] P. Bolton and R. W. Thatcher, "On mass conservation in least-squares methods," *J. Comput. Phys.*, vol. 203, no. 1, pp. 287–304, 2005.

[73] X. Ye, "A least-squares finite-element method for the Stokes equations with improved mass balances," *Comput. Math. Appl.*, vol. 38, no. 3–4, pp. 229–237, 1999.

[74] M. M. J. Proot and M. I. Gerritsma, "Mass- and momentum conservation of the least-squares spectral element method for the Stokes problem," *J. Sci. Comput.*, vol. 27, pp. 389–401, 2006.

[75] J. J. Heys, E. Lee, T. A. Manteuffel, and S. F. McCormick, "An alternative least-squares formulation of the Navier-Stokes equations with improved mass conservation," *J. Comput. Phys.*, vol. 226, no. 1, pp. 994–1006, 2007.

[76] P. B. Bochev and M. D. Gunzburger, "Least-squares finite-element methods for optimization and control problems for the Stokes equations," *Comput. Math. Appl.*, vol. 48, no. 7–8, pp. 1035–1057, 2004.

[77] R. L. Sani and P. M. Gresho, "Résumé and remarks on the open boundary condition minisymposium," *Int. J. Numer. Methods Fluids*, vol. 18, no. 10, pp. 983–1008, 1994.

[78] R. W. Johnson, *The Handbook of Fluid Dynamics*, CRC Press, Boca Raton, FL, 1998.

[79] L. I. G. Kovasznay, "Laminar flow behind a two-dimensional grid," *Proc. Camb. Phil. Soc.*, vol. 44, pp. 58–62, 1948.

[80] M. M. Gupta and J. C. Kalita, "A new paradigm for solving Navier-Stokes equations: streamfunction-velocity formulation," *J. Comput. Phys.*, vol. 207, no. 1, pp. 52–68, 2005.

[81] W. Flügge, *Viscoelasticity*, Springer, Berlin, Heidelberg, 2nd edition, 1975.

[82] R. M. Christensen, *Theory of Viscoelasticity*, Academic Press, New York, 2nd edition, 1982.

[83] W. N. Findley, J. S. Lai, and K. Onaran, *Creep and relaxation of nonlinear viscoelastic materials*, North-Holland Pub. Co., New York, 1976.

[84] J. N. Reddy, *An Introduction to Continuum Mechanics with Applications*, Cambridge University Press, New York, 2008.

[85] R. L. Taylor, K. S. Pister, and G. L. Goudreau, "Thermomechanical analysis of viscoelastic solids," *Int. J. Numer. Meth. Engng*, vol. 2, no. 1, pp. 45–59, 1970.

[86] J. T. Oden and W. H. Armstrong, "Analysis of nonlinear, dynamic coupled thermoviscoelasticity problems by the finite element method," *Comput. Struct.*, vol. 1, no. 4, pp. 603–621, 1971.

[87] G. A. Holzapfel and G. Reiter, "Fully coupled thermomechanical behaviour of viscoelastic solids treated with finite elements," *Int. J. Eng. Sci.*, vol. 33, no. 7, pp. 1037–1058, 1995.

[88] S. Yi, S. F. Ling, M. Ying, H. H. Hilton, and J. R. Vinson, "Finite element formulation for anisotropic coupled piezoelectro-hygro-thermo-viscoelasto-dynamic problems," *Int. J. Numer. Meth. Engng*, vol. 45, no. 11, pp. 1531–1546, 1999.

[89] S. Hartmann, "Computation in finite-strain viscoelasticity: finite elements based on the interpretation as differential-algebraic equations," *Comput. Methods Appl. Mech. Engrg.*, vol. 191, no. 13–14, pp. 1439–1470, 2002.

[90] F. Cortés and M. J. Elejabarrieta, "Finite element formulations for transient dynamic analysis in structural systems with viscoelastic treatments containing fractional derivative models," *Int. J. Numer. Meth. Engng*, vol. 69, no. 10, pp. 2173–2195, 2007.

[91] T.-M. Chen, "The hybrid Laplace transform/finite element method applied to the quasi-static and dynamic analysis of viscoelastic Timoshenko beams," *Int. J. Numer. Meth. Engng*, vol. 38, no. 3, pp. 509–522, 1995.

[92] Y. Aköz and F. Kadioğlu, "The mixed finite element method for the quasi-static and dynamic analysis of viscoelastic Timoshenko beams," *Int. J. Numer. Meth. Engng*, vol. 44, no. 12, pp. 1909–1932, 1999.

[93] B. Temel, F. F. Calim, and N. Tütüncü, "Quasi-static and dynamic response of viscoelastic helical rods," *J. Sound Vib.*, vol. 271, no. 3–5, pp. 921–935, 2004.

[94] Q. Chen and Y. W. Chan, "Integral finite element method for dynamical analysis of elastic-viscoelastic composite structures," *Comput. Struct.*, vol. 74, no. 1, pp. 51–64, 2000.

[95] M. A. Trindade, A. Benjeddou, and R. Ohayon, "Finite element modelling

of hybrid active-passive vibration damping of multilayer piezoelectric sandwich beams-Part I: Formulation," *Int. J. Numer. Meth. Engng*, vol. 51, no. 7, pp. 835–854, 2001.

[96] A. Pálfalvi, "A comparison of finite element formulations for dynamics of viscoelastic beams," *Finite Elem. Anal. Des.*, vol. 44, no. 14, pp. 814–818, 2008.

[97] D. J. McTavish and P. C. Hughes, "Finite element modeling of linear viscoelastic structures–the GHM method," in *AIAA/ASME/ASCE/AHS/ASC Structures*, Dallas, TX; United States, April 1992, Structural Dynamics and Materials Conference, pp. 1753–1763.

[98] D. J. McTavish and P. C. Hughes, "Modeling of linear viscoelastic space structures," *J. Vib. Acoust.*, vol. 115, no. 1, pp. 103–110, 1993.

[99] V. Balamurugan and S. Narayanan, "Finite element formulation and active vibration control study on beams using smart constrained layer damping (SCLD) treatment," *J. Sound Vib.*, vol. 249, no. 2, pp. 227–250, 2002.

[100] V. Balamurugan and S. Narayanan, "Active-passive hybrid damping in beams with enhanced smart constrained layer treatment," *Eng. Struct.*, vol. 24, no. 3, pp. 355–363, 2002.

[101] G. Ranzi and A. Zona, "A steel-concrete composite beam model with partial interaction including the shear deformability of the steel component," *Eng. Struct.*, vol. 29, no. 11, pp. 3026–3041, 2007.

[102] A. R. Johnson, A. Tessler, and M. Dambach, "Dynamics of thick viscoelastic beams," *J. Eng. Mater. Technol.*, vol. 119, no. 3, pp. 273–278, 1997.

[103] E. M. Austin and D. J. Inman, "Modeling of sandwich structures," *Smart Structures and Materials 1998: Passive Damping and Isolation*, vol. 3327, no. 1, pp. 316–327, 1998.

[104] A. C. Galucio, J.-F. Deü, and R. Ohayon, "Finite element formulation of viscoelastic sandwich beams using fractional derivative operators," *Comput. Mech.*, vol. 33, pp. 282–291, 2004.

[105] J. N. Reddy, "On locking-free shear deformable beam finite elements," *Comput. Methods Appl. Mech. Engrg.*, vol. 149, no. 1–4, pp. 113–132, 1997.

[106] J. N. Reddy, *Theory and Analysis of Elastic Plates and Shells*, CRC Press, Boca Raton, FL, 2nd edition, 2007.

[107] T. Belytschko, W. K. Liu, and B. Moran, *Nonlinear Finite Elements for Continua and Structures*, John Wiley and Sons, Ltd, New York, 2000.

[108] Y. Başar, Y. Ding, and R. Schultz, "Refined shear-deformation models for composite laminates with finite rotations," *Int. J. Solids Struct.*, vol. 30, no. 19, pp. 2611–2638, 1993.

[109] N. M. Newmark, "A method of computation for structural dynamics," *J. Eng. Mech.*, vol. 85, pp. 67–94, 1959.

[110] J. Chung and G. .M. Hulbert, "A time integration algorithm for structural dynamics with improved numerical dissipation: The generalized-α method," *J. Appl. Mech.*, vol. 60, pp. 371–375, 1993.

[111] J. C. Simo and T. J. R. Hughes, *Computational Inelasticity*, Springer-Verlag, Berlin, 1998.

[112] J. Lai and A. Bakker, "3-D Schapery representation for non-linear viscoelasticity and finite element implementation," *Comput. Mech.*, vol. 18, pp. 182–191, 1996.

[113] D.W. Van Krevelen, *Properties of Polymers*, Elsevier, Amsterdam, 3rd edition, 1990.

[114] D. Chapelle and K. J. Bathe, "Fundamental considerations for the finite element analysis of shell structures," *Comput. Struct.*, vol. 66, no. 1, pp. 19–36, 1998.

[115] J. N. Reddy, "A generalization of two-dimensional theories of laminated composite plates," *Communications Appl. Numer. Meth.*, vol. 3, no. 3, pp. 173–180, 1998.

[116] J. N. Reddy, *Mechanics of Laminated Composite Plates and Shells: Theory and Analysis*, CRC Press, Boca Raton, FL, 2nd edition, 2004.

[117] D. H. Robbins and J. N. Reddy, "Modelling of thick composites using a layerwise laminate theory," *Int. J. Numer. Meth. Engng*, vol. 36, no. 4, pp. 655–677, 1993.

[118] D. H. Robbins and J. N. Reddy, "Variable kinematic modeling of laminated composite plates," *Int. J. Numer. Meth. Engng*, vol. 39, no. 13, pp. 2283–2317, 1996.

[119] S. Ahmad, B. M. Irons, and O. C. Zienkiewicz, "Analysis of thick and thin shell structures by curved finite elements," *Int. J. Numer. Meth. Engng*, vol. 2, no. 3, pp. 419–451, 1970.

[120] J. C. Simo and D. D. Fox, "On stress resultant geometrically exact shell model. Part I: Formulation and optimal parametrization," *Comput. Methods Appl. Mech. Engrg.*, vol. 72, pp. 267–304, March 1989.

[121] C. Chinosi, L. Della Croce, and T. Scapolla, "Hierarchic finite elements for thin Naghdi shell model," *Int. J. Solids Struct.*, vol. 35, no. 16, pp. 1863–1880, 1998.

[122] M. Cho and H. Y. Roh, "Development of geometrically exact new shell elements based on general curvilinear co-ordinates," *Int. J. Numer. Meth. Engng*, vol. 56, no. 1, pp. 81–115, 2003.

[123] D. Chapelle, D. L. Oliveira, and M. L. Bucalem, "MITC elements for a classical shell model," *Comput. Struct.*, vol. 81, no. 811, pp. 523–533, 2003.

[124] R. A. Arciniega and J. N. Reddy, "Tensor-based finite element formulation for geometrically nonlinear analysis of shell structures," *Comput. Methods Appl. Mech. Engrg.*, vol. 196, no. 46, pp. 1048–1073, 2007.

[125] R. A. Arciniega and J. N. Reddy, "Large deformation analysis of functionally graded shells," *Int. J. Solids Struct.*, vol. 44, no. 6, pp. 2036–2052, 2007.

[126] N. Büechter and E. Ramm, "Shell theory versus degeneration–a comparison in large rotation finite element analysis," *Int. J. Numer. Meth. Engng*, vol. 34, no. 1, pp. 39–59, 1992.

[127] D. Chapelle and K. J. Bathe, *The Finite Element Analysis of Shells–Fundamentals*, Springer-Verlag, New York, 2003.

[128] P. Betsch, A. Menzel, and E. Stein, "On the parametrization of finite rotations in computational mechanics: A classification of concepts with application to

smooth shells," *Comput. Methods Appl. Mech. Engrg.*, vol. 155, no. 34, pp. 273–305, 1998.

[129] M. Bischoff and E. Ramm, "Shear deformable shell elements for large strains and rotations," *Int. J. Numer. Meth. Engng*, vol. 40, no. 23, pp. 4427–4449, 1997.

[130] C. Sansour, "A theory and finite element formulation of shells at finite deformations involving thickness change: Circumventing the use of a rotation tensor," *Arch. Appl. Mech.*, vol. 65, pp. 194–216, 1995.

[131] M. Bischoff and E. Ramm, "On the physical significance of higher order kinematic and static variables in a three-dimensional shell formulation," *Int. J. Solids Struct.*, vol. 37, no. 4647, pp. 6933–6960, 2000.

[132] E. N. Dvorkin and K. J. Bathe, "A continuum mechanics based four-node shell element for general non-linear analysis," *Eng. Computation*, vol. 1, no. 1, pp. 77–88, 1984.

[133] E. Hinton and H. C. Huang, "A family of quadrilateral Mindlin plate elements with substitute shear strain fields," *Comput. Struct.*, vol. 23, no. 3, pp. 409–431, 1986.

[134] J. C. Simo and M. S. Rifai, "A class of mixed assumed strain methods and the method of incompatible modes," *Int. J. Numer. Meth. Engng*, vol. 29, no. 8, pp. 1595–1638, 1990.

[135] J. P. Pontaza and J. N. Reddy, "Mixed plate bending elements based on least-squares formulation," *Int. J. Numer. Meth. Engng*, vol. 60, no. 5, pp. 891–922, 2004.

[136] J. P. Pontaza and J. N. Reddy, "Least-squares finite element formulation for shear-deformable shells," *Comput. Methods Appl. Mech. Engrg.*, vol. 194, no. 2124, pp. 2464–2493, 2005.

[137] F. Moleiro, C. M. Mota Soares, C. A. Mota Soares, and J. N. Reddy, "Mixed least-squares finite element models for static and free vibration analysis of laminated composite plates," *Comput. Methods Appl. Mech. Engrg.*, vol. 198, no. 21–26, pp. 1848–1856, 2009.

[138] F. Moleiro, C. M. Mota Soares, C. A. Mota Soares, and J. N. Reddy, "Layerwise mixed least-squares finite element models for static and free vibration analysis of multilayered composite plates," *Composite Struct.*, vol. 92, no. 9, pp. 2328–2338, 2010.

[139] E. Riks, "An incremental approach to the solution of snapping and buckling problems," *Int. J. Solids Struct.*, vol. 15, no. 7, pp. 529–551, 1979.

[140] E. Riks, "Some computational aspects of the stability analysis of nonlinear structures," *Comput. Methods Appl. Mech. Engrg.*, vol. 47, no. 3, pp. 219–259, 1984.

[141] M. A. Crisfield, "A fast incremental/iterative solution procedure that handles "snap-through"," *Comput. Struct.*, vol. 13, no. 13, pp. 55–62, 1981.

[142] J. H. Kweon and C. S. Hong, "An improved arc-length method for postbuckling analysis of composite cylindrical panels," *Comput. Struct.*, vol. 53, no. 3, pp. 541–549, 1994.

[143] H. Parisch, "Large displacements of shells including material nonlinearities," *Comput. Methods Appl. Mech. Engrg.*, vol. 27, no. 2, pp. 183–214, 1981.

[144] J. C. Simo, M. S. Rifai, and D. D. Fox, "On a stress resultant geometrically exact shell model. Part IV: Variable thickness shells with through-the-thickness stretching," *Comput. Methods Appl. Mech. Engrg.*, vol. 81, no. 1, pp. 91–126, 1990.

[145] A. F. Saleeb, T. Y. Chang, W. Graf, and S. Yingyeunyong, "Hybrid/mixed model for non-linear shell analysis and its applications to large-rotation problems," *Int. J. Numer. Meth. Engng*, vol. 29, no. 2, pp. 407–446, 1990.

[146] C. W. S. To and M. L. Liu, "Hybrid strain based three node flat triangular shell elements-II. Numerical investigation of nonlinear problems," *Comput. Struct.*, vol. 54, no. 6, pp. 1057–1076, 1995.

[147] K. Y. Sze and S. J. Zheng, "A stabilized hybrid-stress solid element for geometrically nonlinear homogeneous and laminated shell analyses," *Comput. Methods Appl. Mech. Engrg.*, vol. 191, no. 17–18, pp. 1945–1966, 2002.

[148] K. Y. Sze, W. K. Chan, and T. H. H. Pian, "An eight-node hybrid-stress solid-shell element for geometric non-linear analysis of elastic shells," *Int. J. Numer. Meth. Engng*, vol. 55, no. 7, pp. 853–878, 2002.

[149] K. Y. Sze, X. H. Liu, and S. H. Lo, "Popular benchmark problems for geometric nonlinear analysis of shells," *Finite Elem. Anal. Des.*, vol. 40, no. 11, pp. 1551–1569, 2004.

[150] R. Arciniega, "On a tensor-based finite element model for the analysis of shell structures," Ph.D. dissertation, Texas A&M University, College Station, TX, 2005.

[151] P. Massin and M. Al Mikdad, "Nine node and seven node thick shell elements

with large displacements and rotations," *Comput. Struct.*, vol. 80, no. 9–10, pp. 835–847, 2002.

[152] S. Timoshenko, *Theory of Elasticity*, McGraw-Hill, New York, NY, 1936.

[153] B. Brank, F. B. Damjanić, and D. Perić, "On implementation of a nonlinear four node shell finite element for thin multilayered elastic shells," *Comput. Mech.*, vol. 16, pp. 341–359, 1995.

[154] A. Barut, E. Madenci, and A. Tessler, "Nonlinear analysis of laminates through a Mindlin-type shear deformable shallow shell element," *Comput. Methods Appl. Mech. Engrg.*, vol. 143, no. 1–2, pp. 155–173, 1997.

[155] P. Mohan and R. K. Kapania, "Updated Lagrangian formulation of a flat triangular element for thin laminated shells," *AIAA Journal*, vol. 36, no. 2, pp. 273–281, 1998.

[156] C. Sansour and F. G. Kollmann, "Families of 4-node and 9-node finite elements for a finite deformation shell theory. An assesment of hybrid stress, hybrid strain and enhanced strain elements," *Comput. Mech.*, vol. 24, pp. 435–447, 2000.

[157] L. Jiang and M. W. Chernuka, "A simple four-noded corotational shell element for arbitrarily large rotations," *Comput. Struct.*, vol. 53, no. 5, pp. 1123–1132, 1994.

[158] W. Wagner and F. Gruttmann, "A simple finite rotation formulation for composite shell elements," *Eng. Computation*, vol. 11, no. 2, pp. 145–176, 1994.

[159] M. Balah and H. N. Al-Ghamedy, "Finite element formulation of a third order laminated finite rotation shell element," *Comput. Struct.*, vol. 80, no. 26, pp. 1975–1990, 2002.

[160] N. Bodard, R. Bouffanais, and M. O. Deville, "Solution of moving-boundary problems by the spectral element method," *Appl. Numer. Math.*, vol. 58, pp. 968–984, 2008.

VITA

Gregory Steven Payette was born in Bountiful, Utah. He spent the majority of his youth in Ras Tanura, Saudi Arabia, where his father was employed by Saudi Aramco. He graduated in 2004 with a bachelor's degree in Mechanical Engineering from the University of Idaho in Moscow, ID. In the fall of the same year he enrolled at Texas A&M University in the Department of Mechanical Engineering. He graduated with his M.S. degree in May of 2006. In May 2012 he received his Ph.D. from Texas A&M University, also through the Department of Mechanical Engineering. Throughout the summer months of his Ph.D. studies, he interned with ExxonMobil Upstream Research Company in Houston, TX and also with Sandia National Laboratories in Albuquerque, NM. During the initial stages of his dissertation research, he was supported by the Oscar Wyatt Endowed Chair of Professor J. N. Reddy and a MURI grant from the Air Force Office of Scientific Research. During the latter part of his Ph.D. research, he was supported by the Excellence in Engineering Graduate Fellowship awarded jointly by Sandia National Laboratories and the Dwight Look College of Engineering at Texas A&M University.

Greg Payette may be contacted through Professor J. N. Reddy at the Mechanical Engineering Department, Texas A&M University, College Station, TX 77843-3123.