The Dissertation Committee for Phillip Joseph Phillips
certifies that this is the approved version of the following dissertation:

Finite Element Methods in Linear Poroelasticity: Theoretical and Computational Results

Committee:

Mary F. Wheeler, Supervisor

Todd Arbogast

Clint Dawson

Irene Gamba

Mark Mear

Finite Element Methods in Linear Poroelasticity: Theoretical and Computational Results

by

Phillip Joseph Phillips, B.A., M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2005

Dedicated to my loving wife, Winnie.

Acknowledgments

I would like to express my appreciation for my advisor, Professor Mary F. Wheeler. She introduced me to an interesting topic and provided important insight throughout. I am also thankful for her gracious personal and financial support. It has been a privilege and honor to be her student.

I am also thankful for Professors Todd Arbogast, Clint Dawson, Irene Gamba and Mark Mear for serving on my committee and providing helpful comments and suggestions. Professor Arbogast has been a particularly inspirational teacher and mentor to me throughout all my years at ICES. Professor Gamba has been a source of encouragement and advice. I will fondly remember Professor Mear's enthusiasm and effectiveness in his teaching. I am also appreciative for Professor Dawson who has provided me with helpful advice.

I would like to thank some of my earlier teachers, especially Professors Ta-Keng Teng and Don Williams who first introduced me to the joy of studying mathematics and for inspiring my decision to pursue an undergraduate degree in mathematics. I would also like to thank Professors Terry Comingore and Bill Shaw. At Rice University, I am particularly thankful to Professor Richard Tapia for his wonderful teaching and personal support. Also at Rice, I enjoyed my courses under Dr. Kevin Scannell. Later, at UT-Austin, I would also like to thank Dr. Venkat Ganesan from the Chemical Engineering faculty.

I would also like to express my appreciation for three of my former graduate student colleagues: Dr. Ruijie Liu, Dr. Xuilie Gai and Dr. Shuyu Sun. Each individual personifies the traits of teamwork and personal warmth.

I am eternally grateful for my parents, Raymond and Betty Phillips, who have given me their unconditional and overflowing love all of my life. I would also like to express my appreciation for my sisters, Sherry L. Witt and Michelle S. Wittman, and for my brother, Dr. Steven R. Phillips.

Lastly, I am forever indebted to my wife, Winnie, for her everlasting love and support. Her steadfast encouragement and inspiration made this dissertation possible.

Finite Element Methods in Linear Poroelasticity: Theoretical and Computational Results

Publication No. _____

Phillip Joseph Phillips, Ph.D.
The University of Texas at Austin, 2005

Supervisor: Mary F. Wheeler

Linear Poroelasticity refers to fluid flow within a deformable porous medium under the assumption of relatively small deformations. Some of the areas that are being modeled with the equations of linear poroelasticity include reservoir engineering, soil mechanics and, more recently, biomedical engineering. The purpose of this dissertation is to present original results for the development, analysis and application of numerical finite element algorithms in the field of linear poroelasticity.

A fully coupled finite element method involving continuous elements for displacements and a mixed space for flow is developed (CG/Mixed). Existence, uniqueness and optimality results are provided. The norm measuring the pressure error, however, depends on the value of the *constrained specific storage coefficient*. For degenerate values, this leads to a slightly weaker optimality result.

For the not untypical case of a null constrained specific storage coefficient, the solution produced by the CG/Mixed scheme sometimes produces

non-physical pressure oscillations, a phenomenon referred to as locking. One potential remedy is to eliminate the continuity requirement for the elements approximating displacements. Therefore, a family of schemes which couples discontinuous elements for displacements and a mixed space for flow is introduced (DG/Mixed). Existence and uniqueness are established, optimal *a priori* error estimates are provided, and some success in the removal of locking is shown.

Direct verification of several benchmark analytical solutions shows that solutions in linear poroelasticity can lack regularity. This sometimes manifests in pressure boundary layers which might degrade the rate of convergence of numerical solutions. The situation can often be ameliorated with the development of adaptive grid refinement strategies. This motivates *a posteriori* estimates in terms of computable residual quantities. Interestingly, it is also shown that the CG/Mixed method can be combined with adaptive grid refinement as an alternative means to eliminate locking.

The produced algorithms are then applied to some interesting application areas. In one instance, they are used to analyze the deformation and pressure dynamics in a cantilever bracket. Additionally, a variety of miscellaneous problems ranging from subsidence and well placement to scuba suit design highlight intriguing applications.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Linear Poroelasticity[1] refers to fluid flow within a deformable porous medium under the assumption of relatively small deformations. Examples of poroelastic structures include soil, rock, cartilage, the brain, the heart and bone. The necessity of coupling deformations to flow in porous media was recognized in the early one-dimensional work of Terzaghi [1925] and others, but it was given a firm mathematical formulation in publication of M.A. Biot's 1941 paper. Since then, some analytical solutions have been derived for Biot's famous equations modeling linear poroelasticity and they are highlighted in Coussy [2004], Barry and Mercer [1999a]. Understandably, since analytical solutions are generally only solved for one-dimensional or axisymmetrical problems, or simple two-dimensional domains, they have rarely been useful in predicting the behavior of large-scale engineering problems. However, the advent of greater computational capabilities has been accompanied by the development of many numerical algorithms, and this has helped to make Biot's model a viable option for those seeking more accurate solutions to large problems.

This introduction begins with a brief overview of some of the ways linear poroelasticity modeling is being used today which, in turn, will provide a sense of its growing importance. Next, a survey is presented for some of the

[1] A term that was coined by J. Geertsma in 1966 [Wang, 2000].

highlights of current theoretical and computational work. Having done so, a synopsis of the original contributions of this dissertation is presented.

1.1 Overview of Modern Applications

Linear poroelasticity is of increasing value today in a diverse range of engineering fields. Notable contributions of poroelasticity have been made to areas that include reservoir engineering, biomechanics and environmental engineering, and promising possibilities hold for fields like earthquake engineering. Following is a brief survey of some of the work that is currently underway.

1.1.1 Reservoir Engineering

In reservoir engineering, terms like *consolidation* – the reduction in volume of a porous medium due to fluid extraction, *compaction* – volume reduction due to air removal, and *subsidence* – vertically-oriented consolidation or compaction, have entered the nomenclature as a result of the observed side effects of drilling. Sometimes the effects of fluid flow-induced deformations are harmless, but there are indeed vivid examples that remind that the fluid-solid coupling cannot be ignored . In Lubick [2002], Dean discusses one such case of an oil company's platforms in the North Sea at Ekofisk field:

> They subsided so much they had to go in and raise the platforms, costing them several billion dollars. If they'd known ahead of time, they could have built their platforms taller.

"Knowing ahead of time" is precisely why poroelastic modeling has come to be viewed as an important resource by which reservoir engineers attempt to

avert future financial calamities. Predicting drilling-induced subsidence can sometimes be accomplished by running computer simulations of various injection/extraction well configurations. Further examples illustrating the coupled effects of geomechanics and flow in a reservoir can be found in Settari and Walters [1999] and Behie et al. [2000].

Another problem confronting the petroleum industry is borehole damage caused by the shifting subsurface. As Bruno [2001] pointed out, "localized shear deformation at weak layers within the overburden appears to have occurred in almost every field investigated." Figure 1.1 shows this type of damage (from the Wilmington Field in California) is usually localized over only several feet. A more indirect type of damage caused to wellbores is buckling. Dusseault et al. [2001] states that this is "caused by axial loading and loss of lateral restraint resulting from solids production and reduction of lateral stress through depletion." Other related work can be found in Detournay and Cheng [1988], Garagash and Detournay [1997], Rajapakse [1993]; other work which also includes thermal effects is found in Wang and Dusseault [2003].

1.1.2 Bioengineering

In biomechanics, Roose et al. [2003] used a poroelastic model to estimate tumor-induced stress levels in the brain, and thus provided assistance in a clinical diagnostic setting. Smillie et al. [2004] likewise modeled the brain, but in this case, their goal was to study the pathological condition hydrocephalus, which induces an irregular cerebrospinal fluid flow that cause the brain to undergo potentially lethal deformations. Meanwhile, Bryne and Preziosi [2003] undertook an effort to understand and limit vascular cancerous tumours. Others, like Swan et al. [2003], use poroelastic modeling of bone in order to esti-

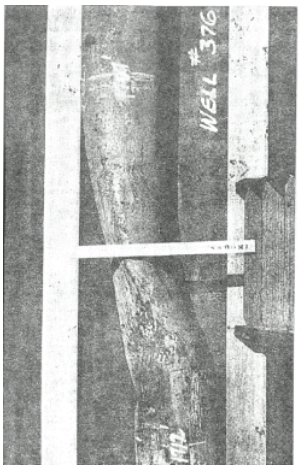

Figure 1.1: Damage to Wellbore Caused by Shearing Subsurface Forces (from Bruno [2001])

mate optimal (external) mechanical loading in order to produce effective bone adaptation. Still others use poroelasticity to assist in their development of prosthetic devices (for cartilage, bone, heart valves, etc.).

1.1.3 Environmental Engineering

In environmental engineering, researchers are often concerned with understanding and, if necessary, limiting the unintended effects from such activities as ground-water pumping or oil extraction on the environment. In Figure 1.2, one can see the effect of years of irrigation in a stretch of California farmland. The top of the telephone pole represents the height of the land in the year 1925, whereas the picture is taken in the year 1977.

Another well-known case is when excessive ground-water removal caused Venice to start sinking [Lubick, 2002]. Langford [1997] pointed to the same problem in Houston, TX, and discussed city plans for avoiding further subsi-

Figure 1.2: Land Subsidence

dence by converting to surface water usage. Another environmentally-related problem is the attempt to control seepage flow from buried hazardous waste sites to local groundwater. The poroelastic model of Kim and Parizek [1997] can be applied to this problem; the nuclear waste disposal problem is similarly tackled in Hudson et al. [2001], Rutqvist and Tsang [2003].

1.1.4 Earthquake Engineering

Within the field of earthquake engineering, a term from poroelasticity that is frequently used is *liquefaction* – the state in which the fluid pressure in a porous medium becomes greater than the forces holding the solid together, and thus converting the mostly solid-like structure to a more fluid-like structure. Tectonic shaking induces an increase in subsurface fluid pressure and

thus potentially devastating liquefaction. One such example was the 1964 Niigata earthquake when liquefaction caused many apartment buildings to be overturned. As the Civil Engineering Department at the University of Washington points out on their website, there are basically three ways to mitigate liquefaction hazards: 1) avoid liquefaction susceptible soils; 2) build liquefaction resistant structures; 3) improve the soil. The latter two methods can benefit from the use of poroelastic model-based computer simulations. In Lin [2003], the author points to poroelastic modeling as one of the reasons that he sees "... an exciting possibility that future models will have significantly improved predictive capability."

Figure 1.3: Damage Caused by Earthquake-Induced Liquefaction

1.2 Survey of Theoretical and Computational Results to Date

The equations governing linear poroelasticity form the following quasi-static, time-dependent, coupled system of partial differential equations,

$$-(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u})-\mu\nabla^2\mathbf{u}+\alpha\nabla p = \mathbf{f}, \quad (1.1)$$

$$\frac{\partial}{\partial t}(c_o p+\alpha\nabla\cdot\mathbf{u})-\frac{1}{\mu_f}\nabla\cdot\boldsymbol{\kappa}(\nabla p-\rho_f\mathbf{g}) = s_f. \quad (1.2)$$

Here, $\mathbf{u}$ refers to displacements, and p refers to the pressure. The momentum conservation equation (1.1) is very similar to the equation governing linear elasticity; the exception is the addition of the term involving pressure. In the mass conservation equation (1.2), the relative fluid velocity is modeled as *Darcy's law*, and the *fluid content*, $\eta = c_o p + \alpha\nabla\cdot\mathbf{u}$, measures the effect of both the pressure and material volume. Here, $\mathbf{f}$ and s_f are the body force and fluid source, respectively.

1.2.1 Analytical Solutions

Because of the somewhat complicated nature of (1.1)-(1.2), only a few analytical solutions have been found. Three problems with such solutions on finite domains are the problems of Terzaghi [1925], Mandel [1953], and Barry and Mercer [1999b]. These three problems and corresponding solutions will be discussed later, and each solution will serve as a benchmark for comparing numerical solutions. Below is a brief description of each:

- *Terzaghi's Problem.* This problem models a 1-D consolidation process. The analytical solution is historically important because it was the first to demonstrate the coupling effects of fluid and solid components in porous media, and because the solution closely matched experimental observations.

- *Mandel's Problem.* This problem involves a saturated porous slab sandwiched between two rigid plates in 2-D. The solution captures the pressure dynamics of the medium as the two plates are squeezed together. Aside from providing a useful test problem, a most important contribution was the solution's emphatic illustration that the coupling effects in poroelasticity cannot be ignored. Indeed, the solution shows that the pressure profile actually increases (for a period) above the initial pressure that is induced by the plate forces. This is because the deformation and rigid plate constraint act somewhat as a source term in a reduced equation for the pressure [see Coussy, 2004]. This unexpected behavior came to be known as the *Mandel-Creyer* effect, and was later experimentally verified by Gibson et al. [1963]. Importantly, Abousleiman et al. [1996] extended the analytical solution to include displacements and stresses.

- *The Problem of Barry and Mercer.* This problem involves a point source which oscillates between an injection and extraction state. The boundary conditions are chosen for the purpose of admitting an analytical solution. Thus, the problem doesn't correspond to one arising naturally, but is vaguely reminiscent on of oil injection/extraction problem.

Some additional problems with analytical solutions are available. For example, solutions to problems with point or line injection source terms are presented in Coussy [2004]. The same reference also provides a solution for the problem of drilling of a borehole. Another example is provided in Kaasschieter and Frijns [2003] where the authors provide an analytical solution for the 3-D squeezing sponge problem.

1.2.2 Numerical Algorithms and Error Estimates

In this subsection an overview of commonly employed numerical methods for modeling in linear poroelasticity is presented. The finite difference method and the finite element method are the most frequently used and are discussed below.

Finite Difference Methods

The finite difference method (FDM) is being used to solve a host of problems in poroelasticity. For example, Zoppou et al. [1997] used a FDM iteratively coupled scheme to study the dynamics of human milk extraction. Their goal was to devise a breast pumping machine that mimics infant suckling and thus to minimize bruising to the areola and nipple. However, one of the drawbacks of the finite difference method arises due to its inability to treat nontrivial geometries effectively. In another example, Steck et al. [2000] used a FDM to model the load-induced fluid displacements under mechanical loading. The authors correctly point out that since "direct flow measurements (especially in compact bone, *in vivo* and *in situ*) are not yet possible, theoretical modelling offers an alternative approach to determine the fluid flow velocities, displacements and effects of interstitial fluid flow." Another interesting effort using the FDM is in acoustic landmine detection. Zeng and Liu [2000, 2001] included the $\rho \mathbf{u}_{tt}$-term in the momentum equation (1.1) so that the problem is no longer quasi-static. Having done so, they used a staggered grid in three dimensions and a leap-frog time-stepping scheme. They posed their problem on an infinite domain which necessitated a truncated computational domain and absorbing boundary conditions. Also using staggered grids is Peter Gerstoft at the Marine Physical Laboratory, UCSD. His

algorithm incorporates a pseudo-spectral finite difference code for poroelastic media. He has made his code, **CABRILLO**, publicly available for download at `www.mpl.ucsd.edu/people/gerstoft/cabrillo/cabrillo.html`.

Convergence analysis for the finite difference method in poroelasticity is rare, although recent work by Ewing et al. [2004] has provided favorable theoretical convergence results for the 1-D case. The authors investigate the FDM on a staggered grid and include the possibility of discontinuous material coefficients.

Finite Element Methods

The finite element method (FEM) has proved popular in the field of poroelasticity for a variety of reasons, in particular because of its capability to handle general domains and the simplicity of its implementation.

Christian [1968] and Sandhu and Wilson [1969] provide examples of the earliest finite element work in poroelasticity. Further examples can be found in Ghaboussi and Wilson [1973], Booker and Small [1975], Zienkiewicz [1982], Cividini and Rossi [1983], Zienkiewicz and Shiomi [1984]. This early pioneering work is primarily focused on continuous Galerkin methods.

In Lewis and Schrefler [1987], the authors provide a comprehensive compilation of their work involving the FEM in poroelasticity. Their method is fully coupled, and they also use continuous Galerkin elements for the pressure and displacements. To complete their formulation, they differentiate with respect to time the momentum equation (1.1). They go on to demonstrate the usefulness of their algorithm by solving one- and two-dimensional problems in consolidation and to the problem of subsidence in Venice. They also expand their continuous Galerkin algorithm to incorporate thermal effects and to an

elastoplastic material skeleton assumption. However, one drawback of using equal low order elements for $\mathbf{u}$ and p is that it can result in an instability in the pressure calculations. Therefore, the equal low order spaces often need to be augmented with certain stabilization techniques such as the fractional step algorithm of Li et al. [2003]. Other finite element methods have developed around the choice of $\mathbf{u} - p$ stable spaces which satisfy the Ladyshenskaya-Babuska-Brezzi (LBB) condition [Babuska, 1971, Brezzi, 1974] or the patch test of Zienkiewicz and Taylor. Murad et al. [1996], Lipnikov [2002] addressed such algorithms. In addition, while most finite element algorithms use a finite difference scheme for time, Runesson et al. [2000] used space-time finite elements.

More recently, numerical work in the finite element method has been conducted at the *Center for Subsurface Modeling* (CSM), a research group at The University of Texas at Austin. There, Liu [2004] has implemented a scheme involving Taylor-Hood elements, and subsequently a DG variant based on the work of Phillips and Wheeler [2005b]. He has also implemented a model which includes thermal effects, and a plasticity model for the porous medium. Also at *CSM*, Gai [2004] has used continuous elements for displacements and a cell-centered finite difference method (CCFD) for pressure. The CCFD for flow is actually equivalent to a mixed formulation for the flow with reduced integration and the velocity eliminated [see Weiser and Wheeler, 1988]. To solve the resulting spatially discretized equations, she has implemented a set of iteratively coupled schemes. She has also tackled the multiphase flow version of the poroelasticity equations and has applied her algorithm to reservoir simulation problems. In addition, Phillips has independently implemented the continuous and discontinuous Galerkin schemes that were proposed by Phillips

and Wheeler [2005a,c,b], and has applied the software to numerous test problems.

With the progress that has been made in the computational realm, there has also been concurrent increase in the attention given to *a priori* error estimates. Murad et al. [1996] provide the first error estimates, but in a restricted sense. They assume a null *constrained specific storage coefficient* value, $c_o = 0$, and they consider only the case of zero Dirichlet conditions for both pressure and displacements and an initial condition of $\nabla \cdot \mathbf{u}(0) = 0$. They also do not allow for a mass source term. These are somewhat unnatural conditions and create a problem in the limit as $t \to 0$; in this case, the poroelasticity problem reduces to Stokes problem with homogenous Dirichlet pressure boundary conditions. This results in an ill-posed problem. Because of this, their error estimates deteriorate in the initial time limit. Later, Phillips and Wheeler [2005a] propose an alternative algorithm which models displacements with continuous elements and the flow with a mixed method. They then provide optimal continuous-in-time and discrete-in-time error estimates for general boundary and initial conditions but only for the case $c_o > 0$.

As one might deduce from the above paragraph, the value of the constrained specific storage coefficient provides a demarkation point between the two sets of error estimates, and its value can be somewhat problematic in a numerical setting. This problem is referred to as *locking*[2], and in the field of poroelasticity can lead to pressure oscillations. The problem and its remedies are discussed at length in Phillips and Wheeler [2005d] and a subsequent

[2]Locking generally refers to numerical problems resulting from extreme parameter values. This problem and its relevance in linear poroelasticity will be discussed extensively in Chapter 5.

chapter of this proposal. It suffices here to mention that Phillips and Wheeler [2005b] first proposed a scheme based on discontinuous Galerkin elements for the deformation and a mixed space for flow. This scheme has been shown to obtain optimal error estimates and to have some success in eliminating locking.

1.2.3 Theoretical Properties

The first theoretical results on the existence and uniqueness of a (weak) solution are by Zenisek [1984]. His investigation involves, however, only the case of a null constrained specific storage coefficient value, $c_o = 0$. As pointed out in Murad et al. [1996], Zenisek's method is to discretize in time the variational formulation of (1.1) -(1.2), and then use a compactness argument to conclude existence. Further theoretical work is seen in the work of Showalter [2000, 2001]. He considers the general case where $c_o \geq 0$, and derives existence, uniqueness and regularity theory for $\mathbf{u}$ and p as an application of the theory of evolution equations in Hilbert space. Showalter's regularity theorem will be discussed in greater detail in subsequent chapters.

1.3 Summary of Major Work

A fully coupled finite element method involving continuous elements for displacements and a mixed space for flow is developed (CG/Mixed). For general boundary and initial conditions, proofs of existence and uniqueness are provided. Tools from functional analysis, approximation theory and elliptic regularity are used to provide optimal *a priori* error estimates. For the special case where the constrained specific storage coefficient, c_o, is bounded below by a positive constant, optimality is demonstrated for the pressure solution when the error is measured in $L^\infty(L^2)$ norm. In the general case, the pressure

solution obtains optimality in the weaker $L^2(L^2)$ norm.

For situations with a null constrained specific storage coefficient value and other factors like a small permeability, the CG/Mixed scheme can produce non-physical pressure oscillations, one way in which locking manifests in linear poroelasticity. One potential way to remedy this occurrence is to eliminate the continuity requirement for the elements approximating displacements. Therefore, a family of schemes which couples discontinuous elements for displacements and a mixed space for flow is introduced (DG/Mixed). Existence and uniqueness are established, and optimal *a priori* error estimates are provided. Unfortunately, using discontinuous linear elements for displacements is shown to have only moderate success in removing locking.

Theoretical and computational results indicate that problems in linear poroelasticity often lack regularity. This sometimes manifests in pressure boundary layers which degrade the rate of convergence in numerical solutions. The situation can often be improved with the utilization of adaptive grid refinement strategies. This leads to the development of *a posteriori* estimates in terms of computable residual quantities. An heuristic analysis of locking also shows that the CG/Mixed method can be combined with adaptive grid refinement as an alternative means to eliminate locking.

The algorithms are applied to some interesting application areas related to the dissertation's theoretical content. In one instance, they are used to analyze the deformation and pressure dynamics in the cantilever bracket. This is an interesting problem because of its similarity to load-induced blood flow within bone. Additionally, a variety of miscellaneous problems as diverse as well-placement and scuba suit design highlight potential application areas.

1.3.1 Plan of Dissertation

Chapter 2 commences with a discussion of the governing equations of linear poroelasticity, and an overview of physical parameters definitions. Mathematical properties of the solutions are outlined and some potential numerical problems related to regularity are discussed.

Chapter 3 introduces a fully coupled finite element algorithm for producing numerical solutions. The algorithm uses continuous elements for approximating displacements and a mixed method for approximating the flow variables. Existence, uniqueness and optimal *a priori* error estimates are provided. When c_o is bounded below by a positive constant, the pressure solution is shown to be optimal in the $L^\infty(L^2)$ norm. For general values, $c_o \geq 0$, the pressure solution obtains optimality in the weaker $L^2(L^2)$ norm.

In Chapter 4, the problems of Terzaghi, Mandel and Barry and Mercer are introduced and their respective solutions and mathematical properties are discussed. Within the context of the regularity, the CG/Mixed method is shown to produce good convergence results. However, two problems are evident from the computational results, and this will eventually serve as motivation for the development of algorithmic variations.

Chapter 5 reviews the problem of locking as it arises in linear elasticity, and some techniques used in overcoming the problem are discussed. Next, an heuristic explanation is given to explain how locking becomes a problem in linear poroelasticity when a small constrained specific storage coefficient is combined with other conditions such as a small permeability. Lessons from linear elasticity are used to explain why the use of discontinuous elements is one way to possibly remedy the locking problem in poroelasticity. Additionally,

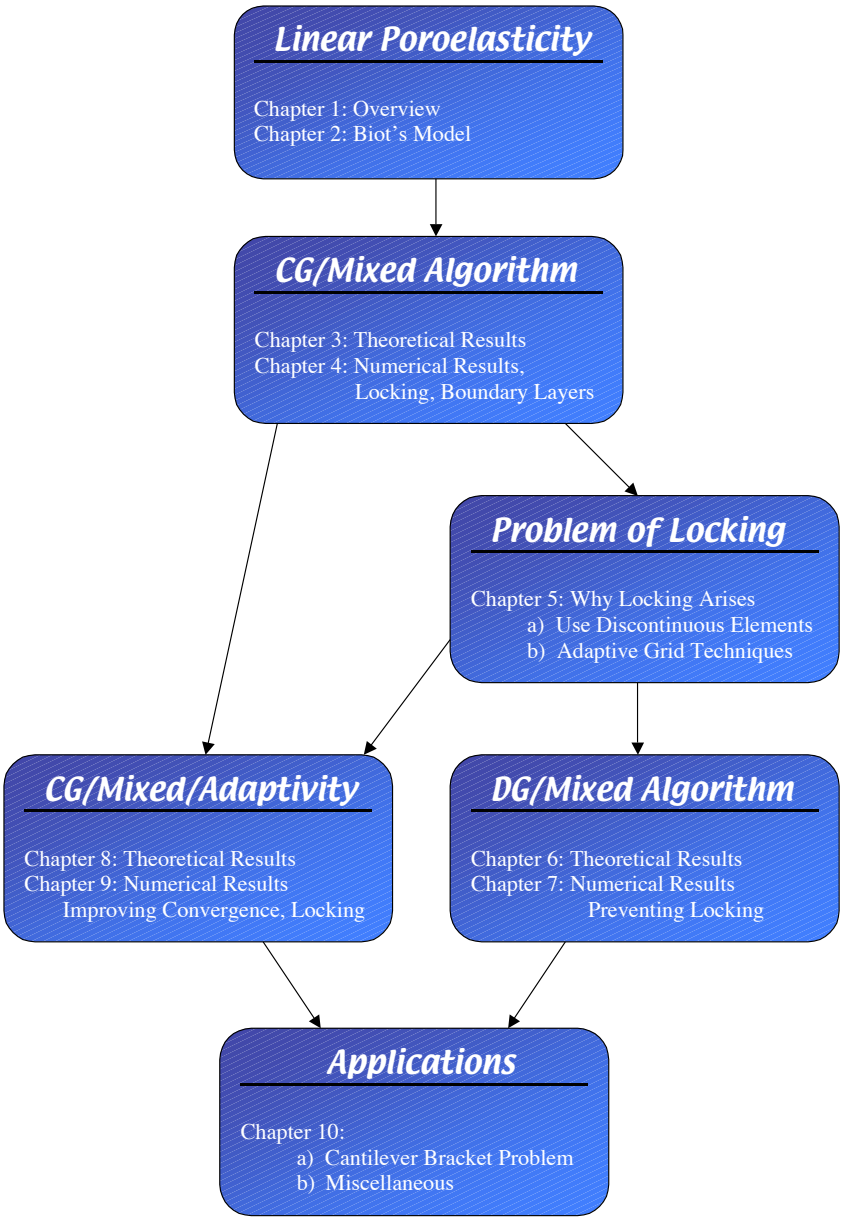

Figure 1.4: Outline of Dissertation

the adaptive grid refinement is shown to be an alternative means to handle locking.

Chapter 6 develops an algorithm for linear poroelasticity in which discontinuous elements are used for the displacement while a mixed space is retained for flow. Existence, uniqueness and optimal error estimates are proved.

Chapter 7 applies the DG/Mixed algorithm to the same problems as used with the CG/Mixed scheme. Again, the expected convergence rates are discovered. Furthermore, locking is shown to be eliminated in most cases, but with a potential problem regarding penalty parameter sensitivity.

Chapter 8 provides the development explicit theoretical *a posteriori* error estimates for the CG/Mixed scheme in terms of naturally defined and computable residual terms.

Chapter 9 provides a grid refinement strategy, and produces numerical examples to show the increases in convergence rate that are achieved. Additionally, numerical results indicate that the CG/Mixed when combined with adaptivity can sometimes be effectively used to eliminate locking.

Chapter 10 outlines some applications related to the work in this dissertation. The cantilever bracket problem with a null constrained specific storage coefficient value is tested, and a number of other interesting problems are examined.

Chapter 11 provides a summary of the content of the dissertation and discusses possibilities for future work related to computational and theoretical results in poroelasticity.

Chapter 2

Model Formulation

2.1 Derivation of the Model

The linear poroelasticity equations consist of a momentum and a mass conservation equation, and are derived at the macroscopic scale in the work of Terzaghi [1925] and M. A. Biot [1941, 1955]. In the derivation of the equations which follows, much of the presentation closely resembles that found in Showalter [2000].

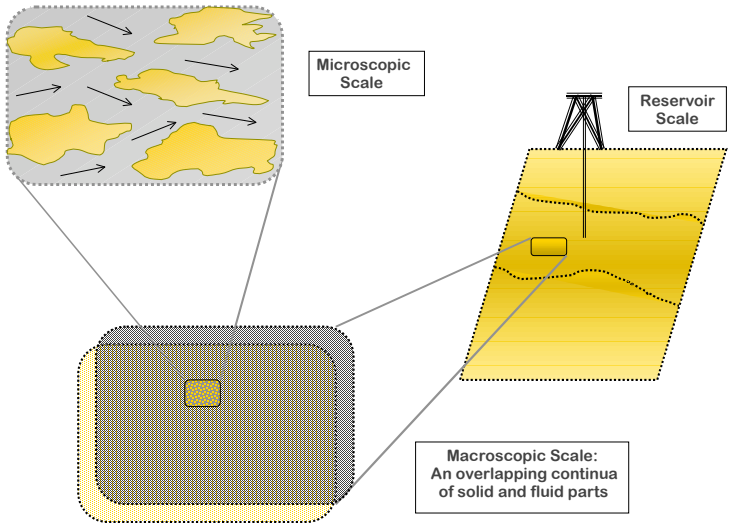

Figure 2.1: Spatial scales in linear poroelasticity.

The momentum conservation is similar to that found in linear elasticity; the exception being the addition of a fluid pressure term. Since the deformation of the material is usually much slower than the flow rate, a quasi-static assumption is made, so that the second time-derivative for displacements is ignored. In order to derive the momentum equation, let $V \subset \Omega$ be a fixed, arbitrary open subset of Ω. Then for the total stress tensor, $\tilde{\boldsymbol{\sigma}}$, and a body force, $\mathbf{f}$, one finds

$$-\int_{\partial V} \tilde{\boldsymbol{\sigma}} \boldsymbol{\nu} \, ds = \int_V \mathbf{f} \, dV,$$

where $\boldsymbol{\nu}$ is the outward normal. Using the divergence theorem on the left side allows one to conclude that

$$-\int_V \nabla \cdot \tilde{\boldsymbol{\sigma}} \, dV = \int_V \mathbf{f} \, dV.$$

Since V was chosen arbitrarily, it follows that $-\nabla \cdot \tilde{\boldsymbol{\sigma}} = \mathbf{f}$ over Ω.

Turning to the mass conservation equation, the variables of interest are the fluid content, η, the fluid flux, $\mathbf{v}_f$, and the volumetric fluid source, s_f. Then, with $V \subset \Omega$ as before, elementary conservation principles lead to

$$\frac{\partial}{\partial t} \int_V \eta \, dV = -\int_{\partial V} \mathbf{v}_f \cdot \boldsymbol{\nu} \, ds + \int_V s_f \, dV.$$

Using the divergence on the first term on the right side of the above equation (and the fact that V was chosen arbitrarily) implies that

$$\frac{\partial \eta}{\partial t} = -\nabla \cdot \mathbf{v}_f + s_f.$$

To close the model, constitutive relations must be formulated in order to relate the total stress, $\tilde{\boldsymbol{\sigma}}$, flux, $\mathbf{v}_f$, and fluid content, η, to the primary variables pressure, p, and deformation, $\mathbf{u}$.

The total stress must account for the usual material stress, as in solid mechanics, and for the fluid pressure; consequently, the assumption is made that $\tilde{\boldsymbol{\sigma}} = \boldsymbol{\sigma} - \alpha\nabla p$. Here, $\boldsymbol{\sigma}$ is referred to as the *effective stress* and measures the material properties of the solid matrix under drained conditions. In linear poroelasticity, it is assumed to be a linear function of the strain tensor, $\epsilon_{ij}(\mathbf{u}) = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$. The pressure term measures the effect of the fluid pressure on the material medium; an increase in pressure generally causes an expansion.

The standard assumption of *Darcy's* law from porous media holds for the flux: $\mathbf{v}_f = -\frac{1}{\mu_f}\boldsymbol{\kappa}(\nabla p - \rho_f \mathbf{g})$. The symmetric permeability tensor, $\boldsymbol{\kappa}$, is assumed to be uniformly bounded and uniformly elliptic; that is, there exist positive constants λ_{min} and λ_{max} such that for all $x \in \Omega$ the following relation holds for $\forall \xi \in \mathbb{R}^n$,

$$\lambda_{min}||\boldsymbol{\xi}||^2 \quad \leq \quad \boldsymbol{\xi}^t \boldsymbol{\kappa}(\mathbf{x})\boldsymbol{\xi} \quad \leq \quad \lambda_{max}||\boldsymbol{\xi}||^2. \tag{2.1}$$

Figure 2.2 describes the constitutive relation for the fluid content, η. It is natural to assume that the amount of fluid content depends on the fluid pressure, p, and the material volume which is measured locally by $\nabla \cdot \mathbf{u}$. More specifically, set $\eta = c_o p + \alpha \nabla \cdot \mathbf{u}$, and observe that $c_o p$ measures the amount of fluid that can be injected into a fixed material volume, and $\alpha \nabla \cdot \mathbf{u}$ measures the amount of fluid that can be squeezed out.

The momentum and mass conservation equations are coupled through the *Biot-Willis* constant, α, with a range of values

Table 2.1: Summary of Constitutive Relations

$\tilde{\sigma}_{ij}(\mathbf{u}, p)$	$=$	$\sigma_{ij}(\mathbf{u}) - \alpha\delta_{ij}p$	Total stress
$\sigma_{ij}(\mathbf{u})$	$=$	$\lambda\delta_{ij}\epsilon_{kk}(\mathbf{u}) + 2\mu\epsilon_{ij}(\mathbf{u})$	Effective stress
η	$=$	$c_o p + \alpha\nabla\cdot\mathbf{u}$	Fluid content
$\mathbf{v}_f$	$=$	$-\frac{1}{\mu_f}\boldsymbol{\kappa}(\nabla p - \rho_f\mathbf{g})$	Average fluid velocity

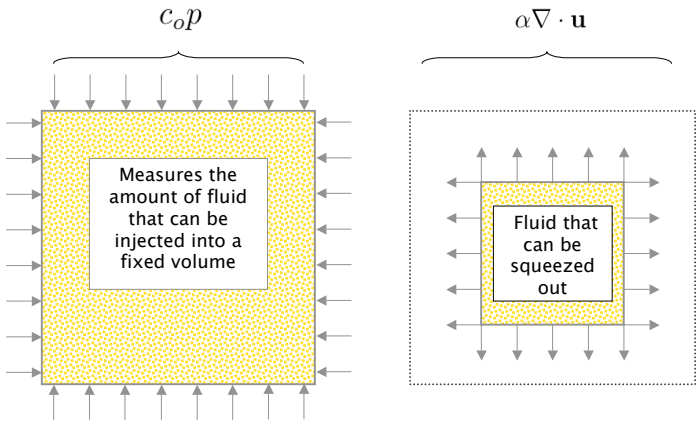

Figure 2.2: The components of the fluid content.

$$0 \;<\; \alpha \;\leq\; 1. \tag{2.2}$$

The nonnegative *constrained specific storage coefficient*[1], c_o, is also assumed to be bounded above by a positive constant M_c,

$$0 \;\leq c_o \;\leq M_c. \tag{2.3}$$

[1]This is discussed further in Wang [2000]; He defines the term as the change in fluid content per change in pressure constrained by a fixed strain, $c_o \equiv \frac{\partial\eta}{\partial p}|_{\delta\nabla\cdot\mathbf{u}=0}$.

The upper bound for c_o follows from an assumption of a reasonable limit on the compressibility on the solid and fluid constituents.

Table 2.2: Summary of Physical Parameters

Parameter	Description
λ, μ	Positive Lamé constants
$c_o \geq 0$	*Constrained specific storage coefficient*
$\alpha > 0$	*Biot-Willis*
$\boldsymbol{\kappa}$	Symmetric permeability tensor
μ_f	Fluid viscosity

For the case of a fully connected pore space and a microscopically homogeneous and isotropic material, Detournay and Cheng [1993] show that α and c_o are related to the *porosity*, ϕ, the *drained skeleton bulk modulus*, K, the *solid bulk modulus*, K_s and the *fluid bulk modulus*, K_f as follows:

$$\alpha = 1 - \frac{K}{K_s}, \tag{2.4}$$

$$c_o = \frac{\phi}{K_f} + \frac{\alpha - \phi}{K_s}. \tag{2.5}$$

From (2.4)-(2.5), the following conditions[2] hold in the event of an incompressible solid and/or fluid constituent:

$$K_s \to \infty \implies \alpha \approx 1, \tag{2.6}$$

$$K_s \text{ and } K_f \to \infty \implies c_o \approx 0. \tag{2.7}$$

[2]It will be seen that the case $c_o \approx 0$ poses a problem for certain numerical schemes, thus prompting several novel computational approaches.

The incompressible limit described by (2.6)-(2.7) may be valid for the entire domain or for only a subsection.

2.2 Summary of the Poroelasticity Equations

Summarized below are the governing equations, including the necessary boundary and initial conditions. There are two distinct sets of boundary conditions, one corresponding to the flow, and one corresponding to the deformation:

$$-(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u})-\mu\nabla^2\mathbf{u}+\alpha\nabla p = \mathbf{f}, \tag{2.8a}$$

$$\frac{\partial}{\partial t}(c_o p+\alpha\nabla\cdot\mathbf{u})-\frac{1}{\mu_f}\nabla\cdot\boldsymbol{\kappa}(\nabla p-\rho_f\mathbf{g}) = s_f, \tag{2.8b}$$

$$p(t) = p_o \quad \text{on} \quad \Gamma_p, \tag{2.9a}$$

$$-\frac{1}{\mu_f}\boldsymbol{\kappa}(\nabla p-\rho_f\mathbf{g})\cdot\boldsymbol{\nu} = q \quad \text{on} \quad \Gamma_f, \tag{2.9b}$$

$$\mathbf{u}(t) = \mathbf{u}_D \quad \text{on} \quad \Gamma_o, \tag{2.9c}$$

$$\tilde{\boldsymbol{\sigma}}\boldsymbol{\nu} = \mathbf{t}_N \quad \text{on} \quad \Gamma_t, \tag{2.9d}$$

$$p(0) = p^o, \tag{2.9e}$$

$$\mathbf{u}(0) = \mathbf{u}^o, \tag{2.9f}$$

where $\partial\Omega = \Gamma_p \bigcup \Gamma_f$ and $\partial\Omega = \Gamma_t \bigcup \Gamma_o$ (see Figure 2.3). In addition, $\boldsymbol{\nu}$ represents the outward normal to $\partial\Omega$.

Remark 2.2.1. Often in practice, initial conditions are found by first setting $p(0)$ equal to the hydrostatic pressure (which amounts to solving $\nabla p(0) = \rho_f\mathbf{g}$), and then using $p(0)$ in (2.8a) to solve for $\mathbf{u}(0)$.

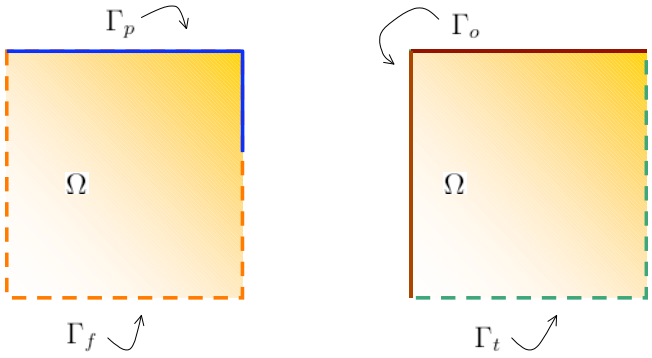

Figure 2.3: The boundary, $\partial\Omega$, is decomposed in two, unrelated ways: one corresponding to the pressure and flux, another related to the displacement and total stress.

2.3 Overview of Mathematical Properties

To provide a proper functional setting for understanding (2.8a)-(2.8b), the following *Sobolev* spaces are defined [Girault and Raviart, 1986, Brenner and Scott, 1994, Evans, 1998]. Set $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ to be a *multiindex* of order

$$|\alpha| = \sum_{j=1}^{n} \alpha_j.$$

Define

$$D^\alpha \equiv \frac{\partial^{|\alpha|} u(x)}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}.$$

Here the derivatives are understood in the distributional sense.

Then for $1 \leq p \leq \infty$ and r any nonnegative integer, let

$$W^{r,p}(\Omega) \equiv \{f \in L^p(\Omega) \mid D^s f \in L^p(\Omega) \;\; |s| \leq r\}.$$

For $1 \leq p < \infty$ the space is endowed with the norm

$$||f||_{r,p,\Omega} \equiv \left(\sum_{|s| \leq r} ||D^s f||^p_{L^p(\Omega)} \right)^{\frac{1}{p}},$$

and for $p = \infty$

$$||f||_{r,\infty,\Omega} \equiv \sum_{|s| \leq r} \operatorname*{ess\,sup}_{x \in \Omega} |D^s f(x)|.$$

For those instances when the context is obvious, the subscript Ω is omitted. In the special case that $p = 2$, the following Hilbert spaces are designated as $H^r(\Omega) \equiv W^{r,2}$ with the norm $|| \cdot ||_r \equiv || \cdot ||_{r,2}$. The inner product $(f, g)_r$ is defined to be

$$(f, g)_r \equiv \sum_{|s| \leq r} \int_\Omega D^s f \; D^s g \; d\Omega.$$

For $r = 0$, the subscript for the norm or inner product may be conveniently omitted.

The fractional order *Sobolev* spaces in $\mathbb{R}^d$ are also defined. Let m be a nonnegative integer and let $\sigma \in (0, 1)$ satisfy $s = m + \sigma$. Then, for $1 \leq p < \infty$, $W^{s,p}(\Omega)$ contains all functions $f \in W^{m,p}(\Omega)$ such that

$$\int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{||x - y||^{d+\sigma p}} \; dx \; dy \leq \infty \quad \forall |\alpha| = m,$$

and given the norm

$$||f||_{s,p,\Omega} \equiv \left\{ ||f||^p_{m,p,\Omega} + \sum_{|\alpha|=m} \int_\Omega \int_\Omega \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{||x-y||^{d+\sigma p}} \, dx \, dy \right\}^{\frac{1}{p}}.$$

For $p = \infty$, $W^{s,\infty}(\Omega)$ contains all functions $f \in W^{m,\infty}(\Omega)$ such that

$$\max_{|\alpha|=m} \operatorname*{ess\,sup}_{x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|^p}{||x-y||^\sigma} \leq \infty.$$

The norm for the case $p = \infty$ is analogous to the one for $1 \leq p < \infty$.

Additionally, since linear poroelasticity is time-dependent, the following spaces are defined Evans [1998]. Let $L^p(0,T;X)$ contain all functions $u : [0,T] \to X$ such that

$$||u||_{L^p(0,T;X)} \equiv \left(\int_0^T ||u(\tau)||^p_X \, d\tau \right)^{\frac{1}{p}} \leq \infty$$

for $1 \leq p < \infty$, and

$$||u||_{L^\infty(0,T;X)} \equiv \operatorname*{ess\,sup}_{0 \leq \tau \leq T} ||u(\tau)||_X \leq \infty.$$

Here, X is shorthand for any of the $W^{s,p}$ spaces. In addition, when no confusion may arise, the time integration domain will be omitted from the notation.

2.3.1 Existence, Uniqueness and Regularity

The most complete work to date on the mathematical properties of solutions to (2.8a)-(2.8b) is undertaken in Showalter [2000, 2001]. The author's theoretical work assumes that the Dirichlet boundary terms for the pressure and displacement are null, and his Neumann boundary conditions differ somewhat from those considered here. His initial conditions also differ slightly, and require a certain compatibility condition. With the further assumption that the all remaining boundary and source data are functions in $C^{\beta}([0,T];L^2)$ (Hölder continuous), the existence and uniqueness of a pressure, p, and displacement, $\mathbf{u}$, solution are proved. The regularity is shown to satisfy

$$||\mathbf{u}(t)||_3 \quad + \quad ||p(t)||_2 \quad \leq \quad \frac{C}{t}. \tag{2.10}$$

The noted differences in the boundary conditions and initial conditions found in Showalter's work do not seem to affect the regularity results. Indeed, both Terzaghi's consolidation problem and Mandel's problem satisfy boundary and initial conditions of the type used in this dissertation, and the pressure in each case is later shown directly to satisfy (2.10).

2.3.2 Potential Numerical Problems

In the best of circumstances, the regularity of the displacement rarely poses difficulties within a computational setting, but the same is not true of the pressure. Whatever finite element scheme is used to approximate the pressure and displacement, an $L^2(H^1)$ bound for the interpolation error for

p is frequently desired[3]. To see why this poses a problem, consider a linear interpolant, I. Then for any time t, and element size, h, the interpolation error is calculated using (2.10) to be

$$||\nabla(p(t) - Ip(t))||_0 \quad \leq \quad C||p(t) - Ip(t)||_1 \quad \leq \quad Ch||p(t)||_2 \quad \leq \quad \frac{Ch}{t}. \tag{2.11}$$

Since linear poroelasticity is time-dependent, in order to establish error estimates (2.11) is usually L^2-integrated in time from 0 to T. This is problematic as the resulting integral in infinite:

$$\int_0^T ||\nabla(p(t) - Ip(t))||_0^2 \, dt \quad \leq \quad (Ch)^2 \int_0^T \frac{1}{t^2} \, dt \quad = \quad \infty.$$

$||\nabla(p(t) - Ip(t))||_0$ needs to be $O(t^\omega)$ for some $\omega < \frac{1}{2}$ in order to produce a finite L^2 integral in time. This possibility is not precluded by Showalter's bound (2.10), but neither is it guaranteed.

It is possible that $p \in L^2(H^s)$ for some $s \in (1, 2)$. This leads to the interpolation estimate

$$||\nabla(p - Ip)||_{L^2(L^2)} \quad \leq \quad C||p - Ip||_{L^2(H^1)} \quad \leq \quad Ch^{s-1}||p||_{L^2(H^s)} \quad < \quad \infty$$

This means that the best convergence rate that could be expected for a finite element scheme with uniform refinement is $(s - 1) < 1$, irrespective of the order of the approximating spaces.

[3]Or, equivalently, an $L^2(L^2)$ bound for the interpolation error for the flux is desired as will be done in this dissertation later when mixed spaces are introduced.

Therefore, it is important to keep in mind the potential effect that lack of regularity may have on the performance of finite element schemes. There are some circumstances (such as when short time steps are used) when it would be fruitless to use a high order approximation space for the purpose of improving the order of convergence. On the other hand, since boundary layers are sometimes produced in these cases, adaptive grid refinement might lead to a significant increase in the rate of convergence. This will be explored in subsequent chapters.

Chapter 3

Continuous Galerkin for Displacements Coupled with a Mixed Formulation for Flow: Theoretical Results

This chapter presents a finite element algorithm for linear poroelasticity which couples continuous Galerkin elements for displacements with a mixed space formulation for flow (CG/Mixed). The primary motivation for doing so are threefold:

1. *Flux a primary variable.* This eliminates the need to use a post-processing techniques to produce the flux from the numerical pressure.

2. *Element-wise mass conservation.* By construction of the mixed space, the flux normal is continuous across element boundaries thus ensuring local mass conservation.

3. *Error estimates for CCFD.* Many commercial flow simulators use a cell-centered finite difference method (CCFD) to simulate flow. Since CCFD is equivalent to the mixed method with reduced integration [Weiser and Wheeler, 1988], there is the potential to re-use the time-tested CCFD code in the linear poroelasticity algorithm presented here and maintain the optimality of the error estimates.

The fully coupled algorithm presented here includes a mixed formulation for flow, so the computational variable for the flux is introduced, $\mathbf{z} = -\frac{1}{\mu_f}\boldsymbol{\kappa}(\nabla p - \rho_f \mathbf{g})$. Then, setting $\tilde{\boldsymbol{\kappa}} = \frac{1}{\mu_f}\boldsymbol{\kappa}$, (2.8a)-(2.8b) then becomes

$$
\begin{aligned}
-(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla^2\mathbf{u} + \alpha\nabla p &= \mathbf{f}, && \text{(3.1a)}\\
c_o\frac{\partial p}{\partial t} + \alpha\frac{\partial}{\partial t}\nabla\cdot\mathbf{u} + \nabla\cdot\mathbf{z} &= s_f, && \text{(3.1b)}\\
\tilde{\boldsymbol{\kappa}}^{-1}\mathbf{z} + \nabla p &= \rho_f\mathbf{g}. && \text{(3.1c)}
\end{aligned}
$$

This set of equations serves as the basis for the variational formulation.

Proofs of the existence, uniqueness and optimality are provided for the CG/Mixed scheme. For the restricted case where the constrained specific storage coefficient, c_o, is bounded below by a positive constant, the pressure solution attains optimality when measured in the $L^\infty(L^2)$ norm. For the more general case where $c_o \geq 0$, optimality can again be shown, but only when the pressure error is measured in the weaker $L^2(L^2)$ norm. For $c_o = 0$, this will help to explain why non-physical pressure oscillations sometimes exist but are always short-lived.

In practice, optimality is often a qualified term since problems in linear poroelasticity sometimes lack regularity. This can degrade convergence rates, and so care is taken to reflect regularity issues in the presentation of the error estimates.

3.1 Function Spaces and Bilinear Forms

Since the algorithm developed here utilizes a mixed space formulation for flow, the following space is defined for the flux in $\mathbb{R}^d$,

$$H(div;\Omega) \equiv \{\mathbf{v} \in L^2(\Omega)^d \mid \nabla \cdot \mathbf{v} \in L^2\},$$

and endowed with the norm,

$$||\mathbf{v}||_{div} \equiv \left(||\mathbf{v}||_0^2 + ||\nabla \cdot \mathbf{v}||_0^2\right)^{1/2}.$$

Therefore, the space $(L^2 \times H(\mathrm{div}))$ is relevant to the pressure and flux.

The space pertinent to the displacement is

$$\mathbf{V} \equiv \{\mathbf{v} \in (H^1(\Omega))^d \mid \mathbf{v}|_{\Gamma_o} = 0\}$$

for which the following bilinear form is defined

$$a_{\mathbf{u}}(\mathbf{u},\mathbf{v}) \equiv \int_\Omega \boldsymbol{\sigma}(\mathbf{u}) : \epsilon(\mathbf{v}) dx,$$

where the linearized *strain* tensor is given by $\epsilon_{kl} \equiv \frac{1}{2}(\partial_k u_l + \partial_l u_k)$. Evidently, one deduces that

$$a_{\mathbf{u}}(\mathbf{u},\mathbf{v}) = \int_\Omega (2\mu(\epsilon(\mathbf{u}) : \epsilon(\mathbf{v})) + \lambda(\nabla \cdot \mathbf{u})(\nabla \cdot \mathbf{v}))dx, \tag{3.2}$$

where $(\sigma : \tau) = \sum_i \sum_j \sigma_{ij}\tau_{ij}$. Define the energy norm by: $||\mathbf{u}||_{a_{\mathbf{u}}}^2 = a_{\mathbf{u}}(\mathbf{u},\mathbf{u})$. Clearly $a_{\mathbf{u}}(,)$ is symmetric, continuous, and given that $|\Gamma_o| > 0$ by *Korn's inequality* $a_{\mathbf{u}}(,)$ is coercive (see Brenner and Scott [1994], Showalter [2000]):

$$a_{\mathbf{u}}(\mathbf{u},\mathbf{v}) \leq C_{cont}||\mathbf{u}||_{H^1}||\mathbf{v}||_{H^1} \quad \forall \mathbf{u},\mathbf{v} \in (H^1)^d, \tag{3.3}$$

$$C_{coer}||\mathbf{u}||_{H^1}^2 \leq a_{\mathbf{u}}(\mathbf{u},\mathbf{u}) \quad \forall \mathbf{u} \in (H^1)^d. \tag{3.4}$$

3.2 Approximating Spaces

Provided now are some definitions important for the development of the scheme. Let $\mathcal{E}_h = \{E_1, E_2, ..., E_{N_h}\}$ be a nondegenerate subdivision of Ω, where E_j is a triangle or quadrilateral for $d = 2$, or a tetrahedron if $d = 3$. Let $h_j = \text{diam}(E_j)$; then nondegeneracy requires the existence of $\rho > 0$ such that E_j contains a ball of radius ρh_j. Set $h = \max\{h_j : j = 1, ..., N\}$.

The finite dimensional approximating spaces are defined as follows. Let $(W_h, \mathbf{S}_h) \subset (L^2 \times H(\text{div}))$ denote the mixed finite element space defined in Raviart and Thomas [1977]. Denote $\mathbf{S}_{h,c} = \mathbf{S}_h \cap \{\mathbf{s} : \mathbf{s} \cdot \boldsymbol{\nu}|_{\Gamma_f} = c\}$. Recall here the associated projection operators and their interpolation properties.

Let $k \geq 0$ refer to the index of the mixed finite element space. Let $P_h : L^2 \to W_h$ be the L^2 projection. For $\epsilon > 0$, let $\mathbf{S}^\epsilon = (H^{\frac{1}{2}+\epsilon})^d \cap H(div)$. Then there exists a projection operator

$$\Pi_h : \mathbf{S}^\epsilon \to \mathbf{S}_h$$

which commutes with P_h,

$$\nabla \cdot \Pi_h = P_h \nabla \cdot, \tag{3.5}$$

and so that P_h and Π_h satisfy the orthogonality properties

$$(\nabla \cdot \mathbf{s}_h, p - P_h p) = 0, \quad \forall \mathbf{s}_h \in \mathbf{S}_h, \tag{3.6a}$$

$$(\nabla \cdot (\mathbf{s} - \Pi_h \mathbf{s}), w) = 0, \quad \forall w \ \in W_h, \tag{3.6b}$$

$$< (\mathbf{s} - \Pi_h \mathbf{s}) \cdot \boldsymbol{\nu}^a, r >_{e_a} \quad \forall r \in \mathbb{P}^k(e_a). \tag{3.6c}$$

Here, e_a is an edge for any element $E \subset \mathcal{E}$ and $\boldsymbol{\nu}^a$ is the unit outward normal. Furthermore, for $1 \leq q \leq \infty$ the following important approximation properties hold for P_h and Π_h [Scholz, 1977, Milner, 1985, Park, 1995, Jiang, 1999]:

$$||p - P_h p||_{0,q} \leq Ch^s ||p||_{s,q}, \quad 0 \leq s \leq k+1, \quad \forall p \in L^2 \cap W^{s,q} \tag{3.7a}$$

$$||\mathbf{s} - \Pi_h \mathbf{s}||_{0,q} \leq Ch^s ||\mathbf{s}||_{s,q}, \quad \frac{1}{q} < s \leq k+1, \quad \forall \mathbf{s} \in \mathbf{S}^\epsilon \cap (W^{s,q})^d \tag{3.7b}$$

Remark 3.2.1. In particular, for $q = 2$, (3.7a)-(3.7b) become

$$||p - P_h p||_0 \leq Ch^s ||p||_s, \quad 0 \leq s \leq k+1, \quad \forall p \in H^s \tag{3.8a}$$

$$||\mathbf{s} - \Pi_h \mathbf{s}||_0 \leq Ch^s ||\mathbf{s}||_s, \quad \frac{1}{2} < s \leq k+1, \quad \forall \mathbf{s} \in (H^s)^d \cap H(div) \tag{3.8b}$$

Notice that the lower bound for s in (3.8b) is $\frac{1}{2}^+$ and not unity as in some literature on mixed methods. Since problems in linear poroelasticity sometimes lack regularity it is helpful to have the flexibility to reduce the bounds for s to strictly greater than a half. This will help theoretically explain less than optimal convergence rates later.

There is also an occasion when the following estimate [Ciarlet, 1976] is needed

$$||(\mathbf{s} - \Pi_h \mathbf{s}) \cdot \boldsymbol{\nu}^a||_{0,e_a} \leq Ch^r ||\mathbf{s}||_{r,e_a}, \quad 0 \leq r \leq k+1, \tag{3.9}$$

as well as the nonstandard trace theorem [Grisvard, 1985]

$$||\mathbf{s}||_{r,e_a} \leq C||\mathbf{s}||_{r+\frac{1}{2},E}. \tag{3.10}$$

Regarding the deformation space, let $\mathbf{V}_h$ be the space of continuous piecewise polynomials of degree r, and set $\mathbf{V}_{h,\mathbf{w}} = \mathbf{V}_h \cap \{\mathbf{v} : \mathbf{v}|_{\Gamma_o=\mathbf{w}}\}$. The elliptic projection operator $\tilde{P}$: $(H^{r+1})^d \rightarrow \mathbf{V}_h$ will be used in the following theorem [Wheeler, 1973]; $\tilde{P}$ satisfies

$$a_{\mathbf{u}}(\mathbf{u} - \tilde{P}\mathbf{u}, \mathbf{v}_h) = 0, \quad \forall \mathbf{v}_h \in \mathbf{V}_h. \tag{3.11}$$

The following inequality describes its approximation property:

$$||\mathbf{u} - \tilde{P}\mathbf{u}||_{a_{\mathbf{u}}} \leq Ch^{s-1}||\mathbf{u}||_{H^s}, \quad 1 \leq s \leq r+1. \tag{3.12}$$

3.3 Continuous in Time Formulation

In order to produce a variational formulation, (3.1a) is multiplied by $\mathbf{v} \in \mathbf{V}$, and (3.1b)-(3.1c) by $(w, \mathbf{s}) \in L^2 \times \{H(div) \cap \{\mathbf{s} : \mathbf{s} \cdot \nu|_{\Gamma_f} = 0\}\}$. Integrating each of the equations over Ω, and integrating by parts when necessary, one arrives at the following: Find $\mathbf{u} \in H^1 \cap \{\mathbf{u} : \mathbf{u}|_{\Gamma_o} = \mathbf{u}_D\}$ and $(p, \mathbf{z}) \in L^2 \times \{H(div) \cap \{\mathbf{s} : \mathbf{s} \cdot \nu|_{\Gamma_f} = q\}\}$ such that

$$a_{\mathbf{u}}(\mathbf{u}, \mathbf{v}) - \alpha(\nabla \cdot \mathbf{v}, p) = l_1(\mathbf{v}), \tag{3.13}$$

$$\Big(c_o \frac{\partial p}{\partial t}, w\Big) + \alpha\Big(\frac{\partial}{\partial t}\nabla \cdot \mathbf{u}, w\Big) + (\nabla \cdot \mathbf{z}, w) = l_2(w), \tag{3.14}$$

$$(\tilde{\boldsymbol{\kappa}}^{-1}\mathbf{z}, \mathbf{s}) - (p, \nabla \cdot \mathbf{s}) = l_3(\mathbf{s}). \tag{3.15}$$

Here, l_1, l_2 and l_3 are bounded linear functionals defined as

$$
\begin{aligned}
l_1(\mathbf{v}) &= \int_\Omega \mathbf{f}\cdot\mathbf{v} + \int_{\Gamma_t} \mathbf{t}_N\cdot\mathbf{v},\\
l_2(w) &= \int_\Omega s_f\ w,\\
l_3(\mathbf{s}) &= -\int_{\Gamma_p} p_o\mathbf{s}\cdot\boldsymbol{\nu} + \int_\Omega \rho_f\mathbf{g}\cdot\mathbf{s}.
\end{aligned}
$$

The fully coupled algorithm is based on the above variational formulation restricted to finite-dimensional spaces, and thus becomes: Find $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}}) \in (\mathbf{V}_{h,\mathbf{u}_D}, W_h, \mathbf{S}_{h,q})$ such that

$$
\begin{aligned}
a_{\mathbf{u}}(\bar{\mathbf{u}},\mathbf{v}) - \alpha(\nabla\cdot\mathbf{v},\bar{p}) &= l_1(\mathbf{v}), && (3.16)\\
\Big(c_o\frac{\partial\bar{p}}{\partial t}, w\Big) + \alpha\Big(\frac{\partial}{\partial t}\nabla\cdot\bar{\mathbf{u}}, w\Big) + (\nabla\cdot\bar{\mathbf{z}}, w) &= l_2(w), && (3.17)\\
(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}\bar{\mathbf{z}}, \mathbf{s}) - (\bar{p}, \nabla\cdot\mathbf{s}) &= l_3(\mathbf{s}), && (3.18)
\end{aligned}
$$

for all $(\mathbf{v}, w, \mathbf{s}) \in (\mathbf{V}_{h,0}, W_h, \mathbf{S}_{h,0})$.

3.4 Theoretical Properties of the CG/Mixed Scheme

In this section, existence and uniqueness of the CG/Mixed scheme are proved. For the special case where c_o is bounded below by a positive constant, optimality is achieved when the pressure solution error is measured in the $L^\infty(L^2)$ norm. For general values where $c_o \geq 0$, the pressure solution obtains optimality in the weaker $L^2(L^2)$ norm.

3.4.1 Existence and Uniqueness

Since the discussion here assumes the general situation in which c_o, for example, may be null on all or part of the domain, an equation like $||c_o^{1/2}\bar{p}||_0 = 0$

would not conclusively prove that $\bar{p} = 0$. By itself this would entail only that $\bar{p} = 0$ in the regions where $c_o > 0$. For this reason a parabolic lift argument is included which will assist in securing proofs of the existence and uniqueness for this general case.

For null boundary and initial conditions, (3.16)-(3.18) has at least one solution, namely the null solution. With this in mind, the following parabolic lift theorem is provided.

Lemma 3.4.1. *Let* $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ *be some solution to the scheme (3.16)-(3.18) with null boundary conditions and data. Then the following inequality holds for some* $C > 0$ *and bounded* h*:*

$$||\bar{p}||_0 \;\leq\; C||\bar{\mathbf{z}}||_0. \tag{3.19}$$

Proof. Let ψ solve the equation $-\nabla^2\psi = \bar{p}$ with boundary conditions compatible with the pressure boundary conditions. Elliptic theory [Gilbarg and Trudinger, 2001] guarantees that such a solution can be found. Now let $\phi = -\nabla\psi$. Then regularity shows that $||\phi||_1 \leq C||p||_0$. The following sequence of inequalities can now be established[1] with the aid of (3.6b) and (3.8b):

[1] As a reminder, throughout the dissertation, C represents the usual generic constant.

$$
\begin{aligned}
||\bar{p}||_0^2 &= (\bar{p}, \bar{p}) \\
&= (\nabla \cdot \phi, \bar{p}) \\
&= (\nabla \cdot \Pi_h \phi, \bar{p}) \\
&= (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}, \Pi_h \phi) \\
&= (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}, \Pi_h \phi - \phi) + (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}, \phi) \\
&\leq ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\Pi_h \phi - \phi||_0 + ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\phi||_0 \\
&\leq C(h||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\phi||_1 + ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\phi||_1) \\
&\leq C(h||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\bar{p}||_0 + ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\bar{p}||_0) \\
&= C(1+h)||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}||_0 ||\bar{p}||_0.
\end{aligned}
$$

From this and inequality (2.1) follows the important inequality,

$$
||\bar{p}||_0 \ \leq \ \frac{C}{\lambda_{min}}(1+h)||\bar{\mathbf{z}}||_0.
$$

Thus, for reasonably bounded h, (3.19) follows.

□

So to continue with the examination of existence and uniqueness, let $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ be some solution to the scheme (3.16)-(3.18) with null boundary conditions and data. Take $\mathbf{v} = \frac{\partial}{\partial t}\bar{\mathbf{u}}$, $w = \bar{p}$ and $\mathbf{s} = \bar{\mathbf{z}}$ in (3.16)-(3.18). By adding (3.16) and (3.17) together, one finds

$$
a_{\mathbf{u}}(\bar{\mathbf{u}}, \frac{\partial}{\partial t}\bar{\mathbf{u}}) + (\frac{\partial}{\partial t} c_o \bar{p}, \bar{p}) + (\nabla \cdot \bar{\mathbf{z}}, \bar{p}) = 0. \tag{3.20}
$$

Now, from (3.18),

$$(\tilde{\boldsymbol{\kappa}}^{-1}\bar{\mathbf{z}}, \bar{\mathbf{z}}) = (\nabla \cdot \bar{\mathbf{z}}, \bar{p}).$$

Substituting this into (3.20), and by using the chain rule from calculus,

$$\frac{1}{2}\frac{\partial}{\partial t}a_{\mathbf{u}}(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + \frac{1}{2}\frac{\partial}{\partial t}(c_o\bar{p}, \bar{p}) + (\tilde{\boldsymbol{\kappa}}^{-1}\bar{\mathbf{z}}, \bar{\mathbf{z}}) = 0. \tag{3.21}$$

Integrating above equation from 0 to T, the equation follows:

$$\frac{1}{2}[a_{\mathbf{u}}(\bar{\mathbf{u}}(T), \bar{\mathbf{u}}(T)) + (c_o\bar{p}(T), \bar{p}(T))] - \frac{1}{2}\overbrace{[a_{\mathbf{u}}(\bar{\mathbf{u}}(0), \bar{\mathbf{u}}(0)) + c_o(\bar{p}(0), \bar{p}(0))]}^{=\ 0\ by\ assumption} + \int_0^T (\tilde{\boldsymbol{\kappa}}^{-1}\bar{\mathbf{z}}(\tau), \bar{\mathbf{z}}(\tau))\ d\tau = 0.$$

From the coercivity of $a_{\mathbf{u}}(,)$, the assumed symmetric positive definiteness of $\boldsymbol{\kappa}$ and the nonnegativity of c_o, the uniqueness of $\bar{\mathbf{u}}$ and $\bar{\mathbf{z}}$ is established. Inequality (3.19) additionally implies uniqueness for $\bar{p}$. Then since uniqueness is equivalent to existence for finite dimensional square systems, the existence of the solution $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ is established automatically for general boundary and initial conditions.

3.4.2 Continuous in Time Error Estimates

This subsection begins with a useful lemma regarding time derivatives.

Lemma 3.4.2. *Let B be a continuous linear operator defined on a Banach space X, and let $f : [0, T] \to X$ be continuously differentiable in time. Then B and $\frac{\partial}{\partial t}$ are a commutative pair; that is, $B\frac{\partial f}{\partial t} = \frac{\partial}{\partial t}B(f)$.*

Proof.

$$\begin{aligned} B\Big(\frac{\partial}{\partial t}f\Big) &= B\Big(\lim_{h\to 0}\frac{f(t+h)-f(t)}{h}\Big) \\ &= \lim_{h\to 0} B\Big(\frac{f(t+h)-f(t)}{h}\Big) \\ &= \lim_{h\to 0}\frac{B(f(t+h))-B(f(t))}{h} \\ &= \frac{\partial}{\partial t}(B(f)). \end{aligned}$$

□

Remark 3.4.1. The above lemma is important when developing error estimates. In particular, interest here regards the cases $B = P_h$ or $B = R_h$.

For convenience some additional notation is now introduced. In particular, the time-dependent auxiliary and projection errors are written as:

$$\begin{aligned} E_p^A &= P_h p - \bar{p}, \qquad & E_p^I &= p - P_h p, \\ E_{\mathbf{z}}^A &= \Pi_h \mathbf{z} - \bar{\mathbf{z}}, \qquad & E_{\mathbf{z}}^I &= \mathbf{z} - \Pi_h \mathbf{z}, \\ E_{\mathbf{u}}^A &= \tilde{P}\mathbf{u} - \bar{\mathbf{u}}, \qquad & E_{\mathbf{u}}^I &= \mathbf{u} - \tilde{P}\mathbf{u}. \end{aligned}$$

In order to facilitate the proof of the following theorem, it is again useful to employ a parabolic lift, this time for the auxiliary pressure error.

Lemma 3.4.3. *Let $(\mathbf{u}, p, \mathbf{z})$ be the solution to (3.1a)-(3.1c), and let $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ be the solution to (3.16)-(3.18). Then there is a positive constant C such that for bounded h, the following holds:*

$$||\bar{p} - P_h p||_0 \;\leq\; C_{\mathbf{z}}||\mathbf{z} - \bar{\mathbf{z}}||_0. \tag{3.22}$$

Proof. From Galerkin orthogonality (3.18) the following equation holds for each $\mathbf{s}_h \in \mathbf{S}_h$:

$$\begin{aligned}
& (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \mathbf{s}_h) - (p-\bar{p}, \nabla\cdot\mathbf{s}_h) = 0 \\
\implies \quad & (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \mathbf{s}_h) - (E_p^A, \nabla\cdot\mathbf{s}_h) - (E_p^I, \nabla\cdot\mathbf{s}_h) = 0 \\
\implies \quad & (E_p^A, \nabla\cdot\mathbf{s}_h) = (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \mathbf{s}_h) - \overbrace{(E_p^I, \nabla\cdot\mathbf{s}_h)}^{=\,0} \\
\implies \quad & (E_p^A, \nabla\cdot\mathbf{s}_h) = (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \mathbf{s}_h). \qquad (3.23)
\end{aligned}$$

Now, let ψ solve the equation $-\Delta\psi = E_p^A$ with boundary conditions compatible with the pressure boundary conditions. Define $\phi = -\nabla\psi$; from elliptic regularity, it is known that $||\phi||_1 \leq C||E_p^A||_0$. Then,

$$\begin{aligned}
||E_p^A||_0^2 &= (E_p^A, E_p^A) \\
&= (E_p^A, \nabla\cdot\phi) \\
&= (E_p^A, \nabla\cdot\Pi_h\phi) \\
&= (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \Pi_h\phi) \\
&= (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \Pi_h\phi-\phi) + (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \phi) \\
&\leq ||\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}})||_0||\Pi_h\phi-\phi||_0 + ||\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}})||_0||\phi||_1 \\
&\leq (Ch+1)||\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}})||_0||\phi||_1 \\
&\leq \left(\frac{Ch+1}{\lambda_{min}}\right)||\mathbf{z}-\bar{\mathbf{z}}||_0||E_p^A||_0.
\end{aligned}$$

Assuming that h is bounded, inequality (3.22) follows.

□

A priori continuous in time error estimates are now examined[2]. It is necessary from the outset to assume that the linear poroelasticity solution belongs to the following spaces

$$p \in L^{\infty}(0,T;L^2) \cap L^2(0,T;H^1), \qquad \mathbf{u} \in L^{\infty}(0,T;H^1). \tag{3.24}$$

The above three norms are used to measure the finite element error produced by the CG/Mixed scheme.

The order of convergence is determined by the order of the approximating spaces, but is limited by the maximal regularity of the true solution and its time derivatives. Let q, s and t be the largest positive real numbers such that

$$p \in L^2(0,T;H^q), \quad p_t \in L^2(0,T;H^s), \quad \mathbf{u}_t \in L^2(0,T;H^t). \tag{3.25}$$

Remark 3.4.2. For the purpose of using the Π_h projection, it is also assumed that $\mathbf{z} \in H(div)\ \forall t \in [0,T]$ and $q > \frac{3}{2}$. As mentioned in the preceding chapter, it turns out that the value of q is the most problematic in obtaining optimal rates of convergence. For the conditions on permeability used here,

$$||\mathbf{z}(t)||_{\beta} \leq C||p(t)||_{\beta+1}$$

for all values of β up to the regularity of the solutions at a particular time t. Returning to the interpolation bounds for the flux, one finds for a mixed

[2]For simplicity, the initial time solutions for the pressure and displacement are assumed to be functions in their respective finite element spaces. In eliminating this assumption, the proof which follows can be easily modified in the usual way.

method of order k,

$$||z(t) - \Pi_h z(t)||_0 \quad \leq Ch^d ||z(t)||_d \quad \leq \quad Ch^d ||p(t)||_{d+1}, \quad \frac{1}{2} < d \leq k+1.$$

Since this bound will need to be L^2-integrated in time, the only suitable choice for d is to set $d = q - 1$. This produces the largest possible exponent for h, while maintaining integrability in time. Unfortunately, if, for example, $\frac{3}{2} < q < 2$, then $(q-1) < 1$ is the best rate of convergence that one may expect irrespective of the order of the approximation spaces. The following theorem and corollary reflect this limitation, something which numerical results later confirm. Fortunately, as will be shown later, adaptive grid techniques can improve the rate significantly in the degenerate case where $\frac{3}{2} < q < 2$.

The primary result of this subsection is summarized in the following theorem.

Theorem 3.4.4 (General Auxiliary Error Estimate)**.** *Let $r_1 \geq 0$ be the order of the mixed space $(W_h, \mathbf{S}_h)$, and let $r_2 > 0$ be the degree of the polynomials used in the displacement space $\mathbf{V}_h$. Then, assuming the regularity (3.24)-(3.25) of the exact solution, parameter constraints (2.1)-(2.3) and no initial time error,*

$$||E^A_{\mathbf{u}}||_{L^\infty(H^1)} + ||c_o^{1/2} E^A_p||_{L^\infty(L^2)} + ||E^A_{\mathbf{z}}||_{L^2(L^2)} \ \leq \ Ch^R. \tag{3.26}$$

Here, $C = C(T, \boldsymbol{\kappa}, C_{coer}, c_o, p, p_t, \mathbf{z}, \mathbf{u}_t)$, and

$$R = \min\{r_1 + 1, \ r_2, \ q - 1, \ s, \ t - 1\},$$

which reflects both the order of the approximation spaces and the regularity of the true solution.

Proof. Denote by $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ the approximate solution. Then by Galerkin orthogonality, the following equations hold for all $(\mathbf{v}, w, \mathbf{s}) \in (\mathbf{V}_{h,0}, W_h, \mathbf{S}_{h,0})$:

$$\begin{aligned} a_{\mathbf{u}}(\mathbf{u}-\bar{\mathbf{u}}, \mathbf{v}) - \alpha(p-\bar{p}, \nabla\cdot\mathbf{v}) &= 0, \quad (3.27)\\ (c_o(p-\bar{p})_t, w) + \alpha(\nabla\cdot(\mathbf{u}-\bar{\mathbf{u}})_t, w) + (\nabla\cdot(\mathbf{z}-\bar{\mathbf{z}}), w) &= 0, \quad (3.28)\\ (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z}-\bar{\mathbf{z}}), \mathbf{s}) - (p-\bar{p}, \nabla\cdot\mathbf{s}) &= 0. \quad (3.29) \end{aligned}$$

Now, by noting that $\mathbf{u}-\bar{\mathbf{u}} = E^A_{\mathbf{u}}+E^I_{\mathbf{u}},\ p-\bar{p} = E^A_p+E^I_p,\ \mathbf{z}-\bar{\mathbf{z}} = E^A_{\mathbf{z}}+E^I_{\mathbf{z}}$ and using the properties of the above projections, one finds that (3.28)-(3.29) becomes

$$\begin{aligned} &(c_o {E^A_p}_t, w) + (c_o {E^I_p}_t, w) + \alpha(\nabla\cdot {E^A_{\mathbf{u}}}_t, w) + \alpha(\nabla\cdot E^I_{\mathbf{u}t}, w)\\ &\qquad +(\nabla\cdot E^A_{\mathbf{z}}, w) + \overbrace{(\nabla\cdot E^I_{\mathbf{z}}, w)}^{=\ 0\ by\ (3.6b)} = 0, \quad (3.30)\\ &(\tilde{\boldsymbol{\kappa}}^{-1}E^A_{\mathbf{z}}, \mathbf{s}) + (\tilde{\boldsymbol{\kappa}}^{-1}E^I_{\mathbf{z}}, \mathbf{s}) - (E^A_p, \nabla\cdot\mathbf{s}) - \overbrace{(E^I_p, \nabla\cdot\mathbf{s})}^{=\ 0\ by\ (3.6a)} = 0. \quad (3.31) \end{aligned}$$

Notice that the properties of the L^2 projection cannot be taken here as c_o is not assumed to be constant. Then letting $w = E^A_p$ and $\mathbf{s} = E^A_{\mathbf{z}}$ in (3.30)-(3.31), after summing the non-zero terms in the above two equations and using the chain rule (in time),

$$\begin{aligned} &\frac{1}{2}\frac{\partial}{\partial t}(c_o E^A_p, E^A_p) + \alpha(\nabla\cdot {E^A_{\mathbf{u}}}_t, E^A_p) + (\tilde{\boldsymbol{\kappa}}^{-1}E^A_{\mathbf{z}}, E^A_{\mathbf{z}})\\ &= -(\tilde{\boldsymbol{\kappa}}^{-1}E^I_{\mathbf{z}}, E^A_{\mathbf{z}}) + (c_o {E^I_p}_t, E^A_p) - \alpha(\nabla\cdot E^I_{\mathbf{u}t}, E^A_p). \quad (3.32) \end{aligned}$$

Next, set $\mathbf{v} = E^A_{\mathbf{u}\,t}$ in (3.27). Then by using the symmetry of $a_{\mathbf{u}}$, the elliptic projection property (3.11), and the chain rule in time,

$$\frac{1}{2}\frac{\partial}{\partial t}a_{\mathbf{u}}(E^A_{\mathbf{u}}, E^A_{\mathbf{u}}) + \overbrace{a_{\mathbf{u}}(E^I_{\mathbf{u}}, E^A_{\mathbf{u}\,t})}^{=\,0\ by\ (3.11)} - \alpha(E^A_p, \nabla\cdot E^A_{\mathbf{u}\,t}) - \alpha(E^I_p, \nabla\cdot E^A_{\mathbf{u}\,t}) = 0. \quad (3.33)$$

Sum (3.32) and (3.33) and observe that the two appearances of the term $\alpha(\nabla\cdot E^A_{\mathbf{u}\,t}, E^A_p)$ conveniently cancel. Next, integrate from 0 to T, and after using the assumption that $E^A_{\mathbf{u}}(0) = E^A_p(0) = 0$, one finds

$$\begin{aligned}\frac{1}{2}[a_{\mathbf{u}}(E^A_{\mathbf{u}}, E^A_{\mathbf{u}}) + (c_o E^A_p, E^A_p)]|_{t=T} + \int_0^T (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E^A_{\mathbf{z}}(\tau), E^A_{\mathbf{z}}(\tau))\, d\tau \\ = \Phi_1 + \Phi_2 + \Phi_3 + \Phi_4, \qquad (3.34)\end{aligned}$$

where

$$\Phi_1 = -\int_0^T (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E^I_{\mathbf{z}}(\tau), E^A_{\mathbf{z}}(\tau))\, d\tau, \quad (3.35)$$

$$\Phi_2 = -\int_0^T \alpha(\nabla\cdot E^I_{\mathbf{u}t}, E^A_p)\, d\tau, \quad (3.36)$$

$$\Phi_3 = \int_0^T \alpha(E^I_p(\tau), \nabla\cdot E^A_{\mathbf{u}\,t}(\tau))\, d\tau, \quad (3.37)$$

$$\Phi_4 = \int_0^T (c_o E^I_{p\,t}(\tau), E^A_p(\tau))\, d\tau. \quad (3.38)$$

To bound the above Φ_1, the Cauchy-Schwarz and Young inequalities (i.e. for $a, b \geq 0, \epsilon \geq 0, ab \leq \frac{1}{2\epsilon}a^2 + \frac{\epsilon}{2}b^2$) are readily used to discover

$$\begin{aligned}
\Phi_1 &= -\int_0^T (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E_{\mathbf{z}}^I(\tau), E_{\mathbf{z}}^A(\tau))\ d\tau \\
&\leq \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E_{\mathbf{z}}^I(\tau)||_0 ||E_{\mathbf{z}}^A(\tau)||_0\ d\tau \\
&\leq C\int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E_{\mathbf{z}}^I(\tau)||_0^2\ d\tau + \frac{\epsilon}{2}\int_0^T ||E_{\mathbf{z}}^A(\tau)||_0^2\ d\tau \\
&\leq \frac{C}{\lambda_{min}^2}\int_0^T ||E_{\mathbf{z}}^I(\tau)||_0^2\ d\tau + \frac{\epsilon}{2}\int_0^T ||E_{\mathbf{z}}^A(\tau)||_0^2\ d\tau.
\end{aligned}$$

The bound for Φ_2 is discovered similarly, along with the parabolic lift lemma, to find

$$\begin{aligned}
\Phi_2 &= -\int_0^T \alpha(\nabla \cdot E_{\mathbf{u}t}^I, E_p^A)\ d\tau \\
&\leq C\int_0^T ||\nabla \cdot E_{\mathbf{u}t}^I(\tau)||_0 ||E_p^A(\tau)||_0\ d\tau \\
&\leq C\int_0^T ||E_{\mathbf{u}t}^I(\tau)||_1^2\ d\tau + \frac{\epsilon}{2}\int_0^T ||E_p^A(\tau)||_0^2\ d\tau \\
&\leq C\int_0^T ||E_{\mathbf{u}t}^I(\tau)||_1^2\ d\tau + \frac{\epsilon}{2}\int_0^T \left[||E_{\mathbf{z}}^A(\tau)||_0^2 + ||E_{\mathbf{z}}^I(\tau)||_0^2\right]\ d\tau.
\end{aligned}$$

To bound Φ_3, integrate by parts in time to deduce

$$
\begin{aligned}
\Phi_3 &= \int_0^T \alpha(E_p^I(\tau), \nabla \cdot E_{\mathbf{u}\,t}^A(\tau))\ d\tau \\
&= -\int_0^T \alpha(E_{p_t}^I(\tau), \nabla \cdot E_{\mathbf{u}}^A(\tau))\ d\tau + \alpha(E_p^I, \nabla \cdot E_{\mathbf{u}}^A)|_{\tau=T} \\
&\leq \int_0^T \alpha||E_{p_t}^I(\tau)||_0||\nabla \cdot E_{\mathbf{u}}^A(\tau)||_0\ d\tau + \alpha||E_p^I(T)||_0||\nabla \cdot E_{\mathbf{u}}^A(T)||_0 \\
&\leq \frac{1}{2}\int_0^T ||E_{p_t}^I(\tau)||_0^2\ d\tau + \frac{1}{2}\int_0^T ||E_{\mathbf{u}}^A(\tau)||_1^2\ d\tau \\
&\quad + C||E_p^I(T)||_0^2 + \epsilon||E_{\mathbf{u}}^A(T)||_1^2.
\end{aligned}
$$

The above inequality uses the bound $\alpha \leq 1$ from (2.4). The bound for Φ_4 is straightforward:

$$
\begin{aligned}
\Phi_4 &= \int_0^T (c_o E_{p_t}^I(\tau), E_p^A(\tau))\ d\tau \\
&= \int_0^T (c_o^{1/2} E_{p_t}^I(\tau), c_o^{1/2} E_p^A(\tau))\ d\tau \\
&\leq \int_0^T ||c_o^{1/2} E_{p_t}^I(\tau)||_0||c_o^{1/2} E_p^A(\tau)||_0\ d\tau \\
&\leq \frac{1}{2}\int_0^T ||c_o^{1/2} E_{p_t}^I(\tau)||_0^2\ d\tau + \frac{1}{2}\int_0^T ||c_o^{1/2} E_p^A(\tau)||_0^2\ d\tau.
\end{aligned}
$$

Now use the above bounds on Φ_1 to Φ_4 in (3.34); in addition, use the inequalities, $a_{\mathbf{u}}(E_{\mathbf{u}}^A, E_{\mathbf{u}}^A) \geq C_{coer}||E_{\mathbf{u}}^A||_1^2$ (coercivity) and $(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E_{\mathbf{z}}^A, E_{\mathbf{z}}^A) \geq \frac{1}{\lambda_{max}}||E_{\mathbf{z}}^A||_0^2$ (boundedness of $\boldsymbol{\kappa}$). So after rearrangement,

$$\Big(\frac{1}{2}C_{coer}-\epsilon\Big)||E^A_{\mathbf{u}}(T)||^2_1+\frac{1}{2}||c_o^{1/2}E^A_p(T)||^2_0+\Big(\frac{1}{\lambda_{max}}-\epsilon\Big)\int_0^T||E^A_{\mathbf{z}}(\tau)||^2_0\,d\tau$$
$$\leq\frac{1}{2}\Big[\int_0^T||c_o^{1/2}E^A_p(\tau)||^2_0\,d\tau+\int_0^T||E^A_{\mathbf{u}}(\tau)||^2_1\,d\tau\Big]$$
$$+C\Big[||E^I_p(T)||^2_0+\int_0^T||E^I_{\mathbf{z}}(\tau)||^2_0\,d\tau$$
$$+\int_0^T||E^I_{\mathbf{u}t}(\tau)||^2_1\,d\tau+\int_0^T||E^I_{p_t}(\tau)||^2_0\,d\tau\Big],$$

where ϵ and C are independent of T and h. Then provided ϵ is small enough so that the left hand side coefficients are all positive, the inequality is preserved by setting each coefficient on the left hand side to equal the smallest of the terms. After dividing the equation by that term,

$$||E^A_{\mathbf{u}}(T)||^2_1+||c_o^{1/2}E^A_p(T)||^2_0+\int_0^T||E^A_{\mathbf{z}}(\tau)||^2_0\,d\tau$$
$$\leq C(\boldsymbol{\kappa},C_{coer})\Big[\int_0^T||c_o^{1/2}E^A_p(\tau)||^2_0\,d\tau+\int_0^T||E^A_{\mathbf{u}}(\tau)||^2_1\,d\tau\Big]$$
$$+C(\boldsymbol{\kappa},C_{coer},c_o)\Big[||E^I_p(T)||^2_0+\int_0^T||E^I_{\mathbf{z}}(\tau)||^2_0\,d\tau$$
$$+\int_0^T||E^I_{\mathbf{u}t}(\tau)||^2_1\,d\tau+\int_0^T||E^I_{p_t}(\tau)||^2_0\,d\tau\Big].$$

Using Gronwall's inequality,

$$
\begin{aligned}
&||E^A_{\mathbf{u}}(T)||^2_1 + ||c_o^{1/2}E^A_p(T)||^2_0 + \int_0^T ||E^A_{\mathbf{z}}(\tau)||^2 \, d\tau \\
&\quad \leq C(T, \boldsymbol{\kappa}, C_{coer}, c_o)\Big[\int_0^T ||E^I_p(\tau)||^2_0 \, d\tau + \int_0^T ||E^I_{\mathbf{z}}(\tau)||^2_0 \, d\tau \\
&\quad + \int_0^T ||E^I_{\mathbf{u}t}(\tau)||^2_1 \, d\tau + \int_0^T ||E^I_{p_t}(\tau)||^2_0 \Big].
\end{aligned}
$$

The above estimate is true for all $0 \leq T$, so a supremum over time can be taken. Thereafter, an application of the approximation properties yields

$$
\begin{aligned}
&\sup_{0\leq\tau\leq T} ||E^A_{\mathbf{u}}(\tau)||^2_1 + \sup_{0\leq\tau\leq T} ||c_o^{1/2}E^A_p(\tau)||^2_0 + \int_0^T ||E^A_{\mathbf{z}}(\tau)||^2_0 \, d\tau \\
&\quad \leq C(T, \boldsymbol{\kappa}, C_{coer}, c_o)\Big[h^{\min\{2r_1+2,2q\}} \int_0^T ||p(\tau)||^2_{\min\{r_1+1,q\}} \, d\tau \\
&\qquad + h^{\min\{2r_1+2,2q-2\}} \int_0^T ||z(\tau)||^2_{\min\{r_1+1,q-1\}} \, d\tau \\
&\qquad + h^{min\{2r_2,2t-2\}} \int_0^T ||\mathbf{u}_t(\tau)||^2_{\min\{r_2+1,t\}} \, d\tau \\
&\qquad + h^{\min\{2r_1+2,2s\}} \int_0^T ||p_t(\tau)||^2_{\min\{r_1+1,s\}} \, d\tau \Big].
\end{aligned}
$$

Remembering that R represents the minimal exponent for the element size, h, the above can be written more compactly as

$$
\begin{aligned}
&||E^A_{\mathbf{u}}||^2_{L^\infty(H^1)} + ||c_o^{1/2}E^A_p(\tau)||^2_{L^\infty(L^2)} + ||E^A_{\mathbf{z}}||^2_{L^2(L^2)} \\
&\qquad \leq C(T, \boldsymbol{\kappa}, C_{coer}, c_o, p, p_{,t}, \mathbf{z}, \mathbf{u}_{,t}) h^{2R},
\end{aligned}
$$

from which the theorem easily follows.

□

The importance of auxiliary error estimates is found when coupled with the triangle inequality and interpolation/projection approximation properties, and is summarized in the following corollary.

Corollary 3.4.5 (General Finite Element Error Estimate)**.** *With the same conditions as in the preceding theorem, the following finite element error estimate holds:*

$$||\mathbf{u} - \bar{\mathbf{u}}||_{L^\infty(H^1)} + ||c_o^{1/2}(p - \bar{p})||_{L^\infty(L^2)} + ||\mathbf{z} - \bar{\mathbf{z}}||_{L^2(L^2)} \ \leq \ Ch^R. \qquad (3.39)$$

Here, $C = C(T, \boldsymbol{\kappa}, C_{coer}, c_o, p, p_t, \mathbf{z}, \mathbf{u}_t)$*, and*

$$R = \min\{r_1 + 1, \ r_2, \ q - 1, \ s, \ t - 1\}.$$

In those parts of the domain where both the fluid and solid constituents are incompressible, (2.5) shows that $c_o = 0$. Therefore, the weighted semi-norm $||c_o^{1/2}(p - \bar{p})||_{L^\infty(L^2)}$ provides useful information only on those regions where $c_o > 0$.

If, however, c_o is bounded below by a positive constant[3], then (3.39) implies

$$||p - \bar{p}||_{L^\infty(L^2)} \ \leq \ Ch^R. \qquad (3.40)$$

[3]That is, when $c_o(x) \geq \gamma_c > 0$ for $\forall x \in \Omega$ and for some positive constant γ_c.

For the situation where $c_o = 0$ on parts of the domain, one would ideally like to have some normed measure for the pressure error. Since the $L^\infty(L^2)$ norm appears too strong to provide information where $c_o = 0$, an alternative, weaker norm might be useful. The parabolic lift lemma for the auxiliary error suggests that the natural candidate for the weaker norm is the $L^2(L^2)$ norm. In this regard, squaring the terms in (3.22) and then integrating in time leads to the inequality

$$||E_p^A||_{L^2(L^2)} \leq C_{\mathbf{z}}||\mathbf{z} - \bar{\mathbf{z}}||_{L^2(L^2)}. \tag{3.41}$$

Therefore, the above inequality and straightforward applications of the triangle inequality, the flux finite element error bound in (3.39), and of interpolation estimates yield a bound for the finite element pressure error measured in the weaker $L^2(L^2)$ norm, valid even where $c_o = 0$. The conclusion is summarized in the following corollary.

Corollary 3.4.6 (Error Estimate for the Incompressible Limit Case with an $L^2(L^2)$ Estimate for the Pressure)**.** *Let $r_1 \geq 0$ be the order of the mixed space $(W_h, \mathbf{S}_h)$, and let $r_2 > 0$ be the degree of the polynomials used in the displacement space $\mathbf{V}_h$. Then, assuming the regularity (3.24)-(3.25) of the exact solution, parameter constraints (2.1)-(2.3) and no initial time error,*

$$||\mathbf{u} - \bar{\mathbf{u}}||_{L^\infty(H^1)} + ||p - \bar{p}||_{L^2(L^2)} + ||\mathbf{z} - \bar{\mathbf{z}}||_{L^2(L^2)} \ \leq \ Ch^R. \tag{3.42}$$

Here, $C = C(T, \boldsymbol{\kappa}, C_{coer}, c_o, p, p_t, \mathbf{z}, \mathbf{u}_t)$, and

$$R = \min\{r_1 + 1,\ r_2,\ q - 1,\ s,\ t - 1\},$$

which reflects both the order of the approximation spaces and the regularity of the true solution.

Remark 3.4.3. It should be noted that nothing in the auxiliary error estimate theorem or subsequent corollaries precludes the possibility that the displacement and flow variables be calculated on different grids. Indeed, in the proofs the coupling terms cancel. So the theorems hold in the asymptotic sense as the element size of each grid becomes arbitrarily small.

Remark 3.4.4. The above theorem and its corollaries can also easily be modified to include the term $\rho\frac{\partial^2 \mathbf{u}}{\partial t^2}$, and thus eliminating the quasi-static assumption. To do so, the additional initial condition, $\mathbf{u}_t(0) = \mathbf{u}_T$, is added. Then in the auxiliary estimates, one adds $\rho(E^A_{\mathbf{u},tt}, E^A_{\mathbf{u},t}) + \rho(E^I_{\mathbf{u},tt}, E^A_{\mathbf{u},t})$ to the momentum variational equation. The first term is re-written as $\frac{1}{2}\rho\frac{\partial}{\partial t}||E^A_{\mathbf{u},t}||^2_0$. The second term is moved to the rhs to be bounded with the Cauchy-Schwarz and Young inequalities. This leaves the term $C||E^I_{\mathbf{u},tt}||^2_{L^2(L^2)}$ to be estimated with interpolation results and the term $C\int_0^T ||E^A_{\mathbf{u},t}||^2_0$ which can be eliminated with Gronwall's inequality. The final result of the modification is an estimate of $||e_{\mathbf{u},t}||^2_{L^\infty(L^2)}$.

3.4.3 A Comparison with the Estimates from Murad et. al.

As mentioned in the Introduction, the paper by Murad et al. [1996] provides theoretical error estimates for the restricted case of a null constrained specific storage coefficient value, $c_o = 0$, and so offers a *partial* comparison with the above results. It is partial in the sense that the authors consider only

a very restricted[4] set of boundary and initial conditions. Namely, they allow neither for a traction boundary nor for a flow boundary condition; instead they assume the entire boundary in homogeneous Dirichlet for both the pressure and displacement. They also assume that there is no mass source term. So the entire dynamics are driven by the body force $\mathbf{f}$ which they assume is sufficiently smooth from the outset. These restrictions are important because a traction force or a mass source term, as will be verified later, are major sources of irregularity in linear poroelasticity solutions which can degrade the convergence rate of a numerical solution.

In context of the preceding comments, for the lowest order Arnold space (a stable $\mathbf{u} - p$ space augmented with a bubble function) on triangles, Murad et al. [1996] produce the error estimate

$$t^{1/2}||\mathbf{u}(t) - \bar{\mathbf{u}}(t)||_1 + t||p(t) - \bar{p}(t)||_1 \ \leq \ Ch. \tag{3.43}$$

The above weighted norm estimate is optimal in h, but is valid for only the very restricted and tame conditions mentioned earlier, and is *not* technically valid for the benchmark problems to be discussed later. For pressure, (3.43) implies that

$$||p(t) - \bar{p}(t)||_1 \ \leq \ \frac{Ch}{t}. \tag{3.44}$$

Therefore

[4]Their conditions *do not* hold for *any* of the problems which will be discussed later.

$$||p-\bar{p}||^2_{L^2(0,T;H^1)} \equiv \int_0^T ||p(t)-\bar{p}(t)||_1^2 \leq \int_0^T \left(\frac{Ch}{t}\right)^2 = \infty. \quad (3.45)$$

So the bound provided by Murad et al. [1996] is *inconclusive* in the $L^2(H^1)$ sense. The authors also produce similar results for other stable and unstable $\mathbf{u}-p$ spaces.

On the other hand, for the same case of a null constrained specific storage coefficient value, the error estimate in this dissertation produces an L^2 time integral of the pressure and flux variables. The results here also take into account the regularity of the true solution in order to determine the expected rate of convergence. More specifically, for the pressure and flux, it was shown that

$$\int_0^T ||p(t)-\bar{p}(t)||_0^2 + \int_0^T ||\mathbf{z}(t)-\bar{\mathbf{z}}(t)||_0^2 \leq Ch^{2R}. \quad (3.46)$$

This provides an error bound in the $L^2(L^2)$ sense for the flow variables. If one makes the assumption that $||p(t)-\bar{p}(t)||_0 \sim O(t^{-a})$ and $||\mathbf{z}(t)-\bar{\mathbf{z}}(t)||_0 \sim O(t^{-b})$ asymptotically as $t \to 0$, the finiteness of the integrals in (3.46) implies that $a < \frac{1}{2}$ and $b < \frac{1}{2}$. Using this assumption in (3.46) one deduces that

$$||p(t)-\bar{p}(t)||_0 < \frac{Ch^R}{t^a}, \quad ||\mathbf{z}(t)-\bar{\mathbf{z}}(t)||_0 < \frac{Ch^R}{t^b}, \quad (3.47)$$

as $t \to 0$. From (3.47) it follows that

$$t^a||p(t)-\bar{p}(t)||_0 + t^b||\mathbf{z}(t)-\bar{\mathbf{z}}(t)||_0 < Ch^R, \quad (3.48)$$

as $t \to 0$. Furthermore, the convergence rate R reflects both the order of the approximating spaces and the regularity of the exact solution, and therefore optimality is problem dependent.

Therefore, for the case of a null constrained specific storage coefficient value, $c_o = 0$, both Murad et al. [1996] in (3.43) and the dissertation's result expressed in (3.48) reflect a potential degeneracy in the early time numerical pressure solution. However, only the dissertation's results provide a finite value for the L^2 integral in time of the error. This allows for a smaller exponent for the time variable multiplying the error of the flow variables. Additionally, only the results presented here allow for a broad and natural set of initial and boundary conditions. Furthermore, only the results presented in this dissertation adequately reflect the potential irregularity of the exact solution and, therefore, the ability to explain the observed suboptimal convergence for certain numerical examples.

Chapter 4

Continuous Galerkin for Displacements Coupled with a Mixed Formulation for Flow: Numerical Results

The goal of this chapter is to examine the performance of the CG/Mixed algorithm when tested against several problems with known analytical solutions. The problems are grouped into two sets. The first set of problems highlights accuracy results for the continuous[1] in time scheme. Although convergence rates are shown to be less than optimal, it is, nonetheless, expected in light of the lack of regularity which is shown to exist in the true solutions. The second set of examples deals with the situation of a null constrained specific storage coefficient value. Since a weaker theoretical result for the pressure arises when $c_o = 0$ it is important to understand the behavior of the numerical solutions in this case.

The alternative purpose of this chapter is motivational. The first set of examples suggests the likely importance of incorporating adaptive grid refinement techniques as a means to resolve boundary layers, and, thus, to improve convergence rates. The second set of examples prods the search for a more ro-

[1]The methods are implemented discrete in time, and, unless otherwise specified, the examples use a time step small enough to be *effectively* continuous in time. A discrete in time error estimate is provided in Appendix C. Appendix D discusses the numerical effect of a larger time step and the potential improvements in convergence by doing so.

bust scheme in light of the effect of extreme parameter values. The remainder of this dissertation pivots around these two points.

4.1 Problems with Inadequate Regularity

Attention here is turned toward two problems typical in linear poroelasticity, *Terzaghi's consolidation problem* and *Mandel's problem*. The pressure and displacement dynamics are each driven solely by a traction force applied at time $t = 0^+$. The jump in the traction force means that the boundary data in each problem is not Hölder continuous, and so Showalter's theorem on regularity cannot be applied *a priori*. Interestingly, Showalter's bounds for pressure are still shown to hold for the analytical pressure solutions, $||p||_2 < \frac{C}{t}$.

But as mentioned earlier and as less than optimal convergence rates for the continuous in time numerical solution will underscore, Showalter's bounds are not tight enough to guarantee even first order convergence regardless of the order of approximation. This is not intended as a criticism of Showalter's regularity findings, but rather to point out that potential problems may be inherent in some solutions in linear poroelasticity, and these problems can have a deleterious effect of numerical performance.

Because of the potential degeneracy in regularity, it might be unproductive[2] to use higher order spaces as a means to improving convergence rates if a small time step is employed. Ironically, the lack of regularity which degrades performance for the continuous in time CG/Mixed method with uniform refinement may hold in itself the key to finding a way to improve convergence

[2]Higher order spaces may, however, provide a means to combat non-physical pressure oscillations.

rates. The pressure boundary layers that often ensue strongly suggest the potential benefits of adaptive grid refinement. Indeed, it will be shown later in work on *a posteriori* error estimates and grid refinement that this is the case.

4.1.1 Terzaghi's Problem

In 1923, K. Terzaghi [1923] provided one of the first solutions to the classic consolidation problem in poroelasticity[3]. The problem refers to a surface load being applied to a fully saturated poroelastic domain. As a result of the initial loading F at time $t = 0^+$, an instantaneous overpressure occurs throughout the domain; more specifically, theory predicts that $p^+ \equiv \lim_{t\to 0} p(y,t) = \frac{\alpha}{c_o}\frac{F}{K_u+\frac{4}{3}\mu}$. However because of the drainage at the surface, there is a diffusion of pressure back toward its initial state. Regarding the deformation, of primary interest is the surface settlement, $s(t) = u_y(y = H, t)$, where u_y is the vertical component of the displacement and H is the height of the domain. Like the pressure, the settlement responds instantaneously to the applied load, and is predicted to satisfy $s_0 \equiv \lim_{t\to 0} s(t) = \frac{FH}{K_u+\frac{4}{3}\mu}$. Moreover, theory also predict that asymptotic settlement to be $s_\infty \equiv \lim_{t\to\infty} s(t) = \frac{FH}{K+\frac{4}{3}\mu}$ This deformation is the very problem that produces so much of the costly damage that plagues some overland structures. So despite the simplicity of its 1-D formulation, Terzaghi's problem does illustrate some interesting aspects of solid-fluid interactions, and it is in fact applicable as a good approximation in a number of real world circumstances.

To the governing equations of linear poroelasticity are added the following boundary and initial conditions:

[3]The discussion here of the model and review of analytical results follow that found in Coussy [2004].

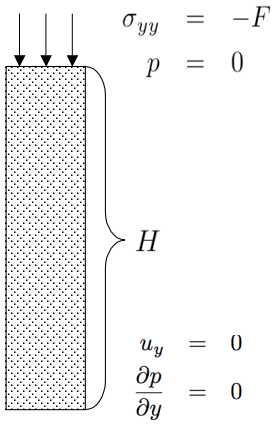

Figure 4.1: Terzaghi's Problem Formulation

$$
\begin{aligned}
p = 0, \quad y = 0; \quad \frac{\partial p}{\partial y} = 0, \quad y = -H; \\
\sigma_{yy} = t_N(t), \quad y = 0; \quad \mathbf{u} = 0, \quad y = -H; \\
p|_{t=0} \;=\; \mathbf{u}|_{t=0} \;=\; 0.
\end{aligned}
$$

Here $t_N(0) = 0,\ t_N(t) = -F,\ t > 0$. The analytical series solutions for the pressure and settlement are

$$
\begin{aligned}
p(y,t) &= \frac{\alpha F}{c_o(K_u + \frac{4}{3}\mu)} \sum_{n=0}^{\infty} \frac{4}{\pi(2n+1)} \sin\Big(\frac{(2n+1)\pi}{2H} y\Big) \\
&\quad \times \exp\Big(-\frac{(2n+1)^2\pi^2}{4H^2} c_f t\Big), \\
s(t) &= s_\infty + (s_0 - s_\infty) \sum_{n=0}^{\infty} \frac{8}{\pi^2(2n+1)^2} \times \exp\Big(-\frac{(2n+1)^2\pi^2}{4H^2} c_f t\Big).
\end{aligned}
$$

4.1.1.1 A Discussion of Regularity

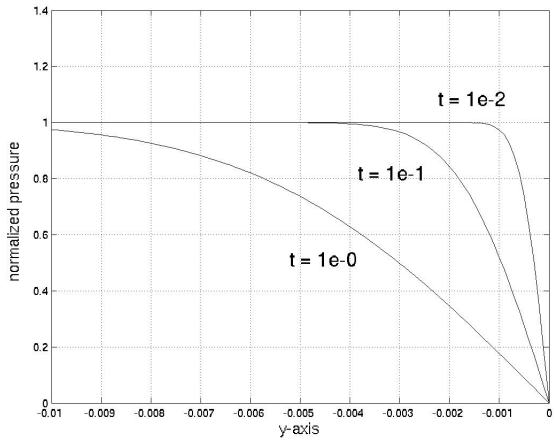

Figure 4.2: Terzaghi's pressure at various times. As $t \to 0$ the pressure solution begins to experience very large derivatives near the boundary at $y = 0$.

As discussed in the theorem on error estimates, the order of convergence of the CG/Mixed numerical solution is limited by the regularity of the true solution. Therefore, since the true pressure solution is available for Terzaghi's problem, it might be possible to get some idea regarding its regularity. In turn, this information might lead to a realistic expectation in regard to the best possible order of convergence.

Figure 4.2 shows the true pressure solution for a typical choice of physical parameters. As $t \to 0$ a very large derivative occurs. Naturally, this leads one to suspect that the regularity of the pressure might be limited.

Figure 4.3 shows the numerical approximation of $|\frac{\partial p}{\partial y}|^2$ and $|\frac{\partial^2 p}{\partial y^2}|^2$ at progressively smaller times (from top to bottom). The maximum value of each term increases for smaller times t, but the area also decrease over which each term is non-zero.

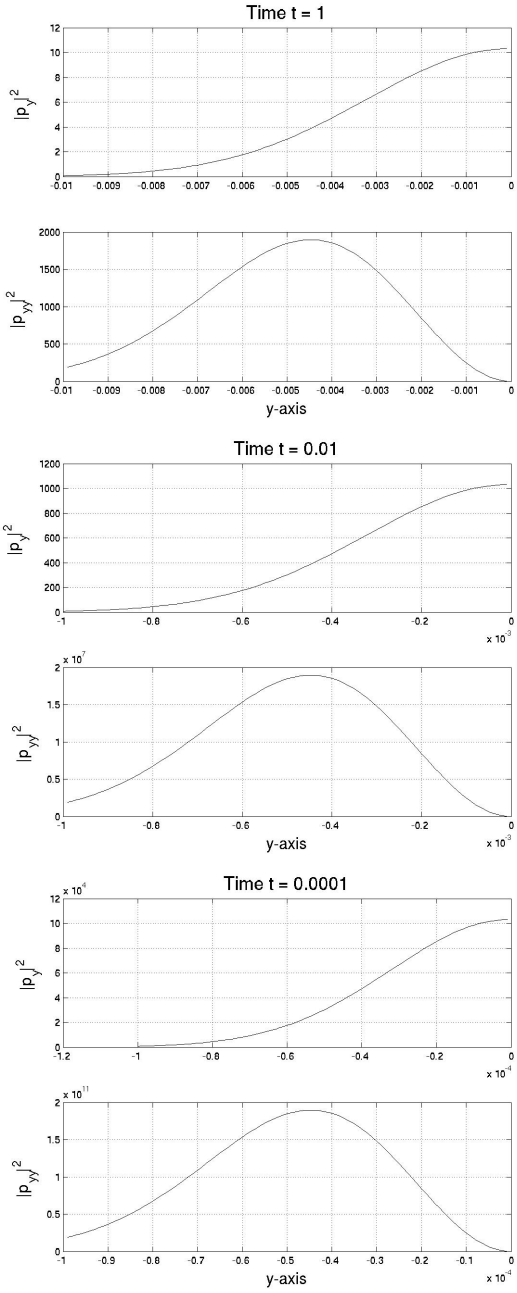

Figure 4.3: Terzaghi's Problem. $|\frac{\partial p}{\partial y}|^2$ and $|\frac{\partial^2 p}{\partial y^2}|^2$ at different times. $\sup(|\frac{\partial^2 p}{\partial y^2}|^2) \approx 2e3$ $(t = 1e0)$, $2e7$ $(t = 1e-2)$, and $2e11$ $(t = 1e-4)$.

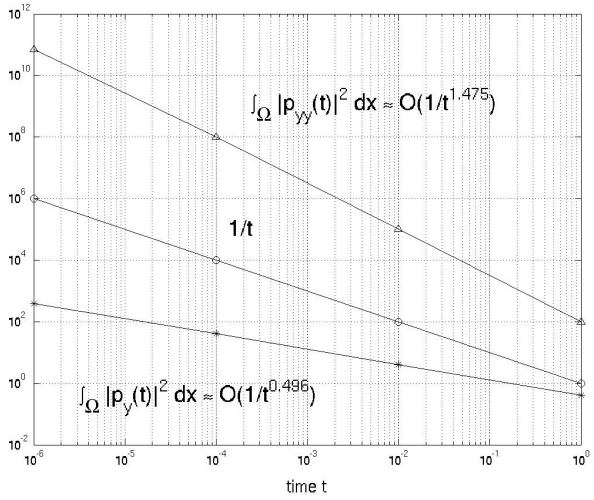

Figure 4.4: Terzaghi's Problem. The pressure integrals $||p_y(t)||_0^2$ and $||p_{yy}(t)||_0^2$.

To help determine the regularity of the pressure, $|\frac{\partial p(t)}{\partial y}|^2$ and $|\frac{\partial^2 p(t)}{\partial y^2}|^2$ are integrated over the spatial dimension to give values of $||p_y(t)||_0^2$ and $||p_{yy}(t)||_0^2$. Figure 4.4 shows their respective values as a function of time. It is discovered that

$$||p_y(t)||_0^2 \approx O(t^{-0.496}) \tag{4.1}$$

$$||p_{yy}(t)||_0^2 \approx O(t^{-1.475}) \tag{4.2}$$

The order approximation (4.1) implies that $p \in L^2(H^1)$, but (4.2) shows that $p \notin L^2(H^2)$.

One may then assume that $p \in L^2(H^{1+s})$ for some $s \in [0, 1)$. It is precisely this value of s, therefore, which limits the best rate of convergence which one might expect from the CG/Mixed algorithm for Terzaghi's problem.

Note that (4.1)-(4.2) imply satisfaction of Showalter's bound, $||p||_2 \leq \frac{C}{t}$, but already there are reasons to expect a less than optimal convergence rate.

4.1.1.2 Computational Results

For numerical verification of the CG/Mixed algorithm, two sets of parameters are chosen. For each case the unit height, $H = 1$, and force, $F = 1000$ are chosen. Standard continuous linear elements are used to approximate the displacement, and the lowest Raviart-Thomas space is for the flow variables.

Test Case I

The first test case uses the following set of physical parameters:

$$E = 1e + 5, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e - 1, \quad \boldsymbol{\kappa} = 1e - 6.$$

The left image in Figure 4.5 shows the computed (circles) and analytical normalized pressure profile, p/p^+, at various times. Despite using a fairly large time step, the true and the CG/Mixed approximate normalized pressure solutions are very close. Theory predicts that $p^+ \equiv \lim_{t \to 0} p(y,t) = \frac{\alpha}{c_o} \frac{F}{K_u + \frac{4}{3}\mu}$; for the chosen set of parameters $p^+ = .0899$.

The right image in Figure 4.5 shows the time-dependent settlement in Terzaghi's problem. There is good agreement between the computed (circles) and analytical expression for the settlement. The top line represents s_o, the instantaneous settlement. The bottom line represents s_∞, the predicted long time settlement.

Figure 4.6 shows how the computed pressure error, $||p - \bar{p}||_{l^\infty(L^2)}$, varies with element size, h. To minimize the effects of the error produced by the time

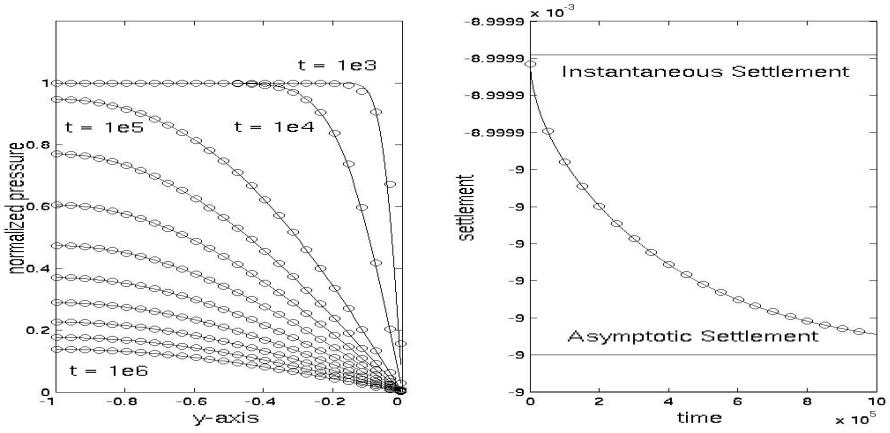

Figure 4.5: Pressure profile (left) and settlement (right) in Terzaghi's Problem. The circles represent the CG/Mixed solution; the straight line represents the true solution.

discretization, a small time step of $\Delta t = 1e-6$ is chosen. The convergence rate is determined by the slope of the logarithm of the error (≈ 0.50001). Although results are shown only for the backward Euler time-stepping scheme, the results for any other time-stepping scheme are indistinguishable.

Test Case II

Figure 4.7 depicts the results from a second series of tests using the following set of parameters:

$$E = 1e+7, \quad \nu = 0.4, \quad \alpha = 1, \quad c_o = 5e-1, \quad \boldsymbol{\kappa} = 1e-4.$$

Again, to minimize the effects of the error produced by the time discretization, a small time step of $\Delta t = 1e-6$ is chosen. The convergence rate for this set

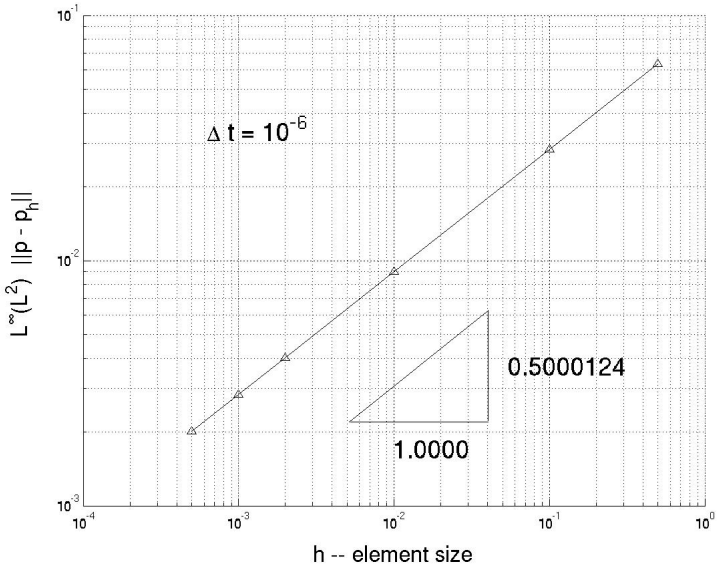

Figure 4.6: $L^{\infty}(L^2)$ pressure error in Terzaghi's problem. Physical parameters used: $E = 1e5;\ \nu = 0.2;\ c_o = 1e-1;\ \boldsymbol{\kappa} = 1e-6.$

of parameters is found to be about 0.50012. Once again, each time-stepping scheme produces identical results.

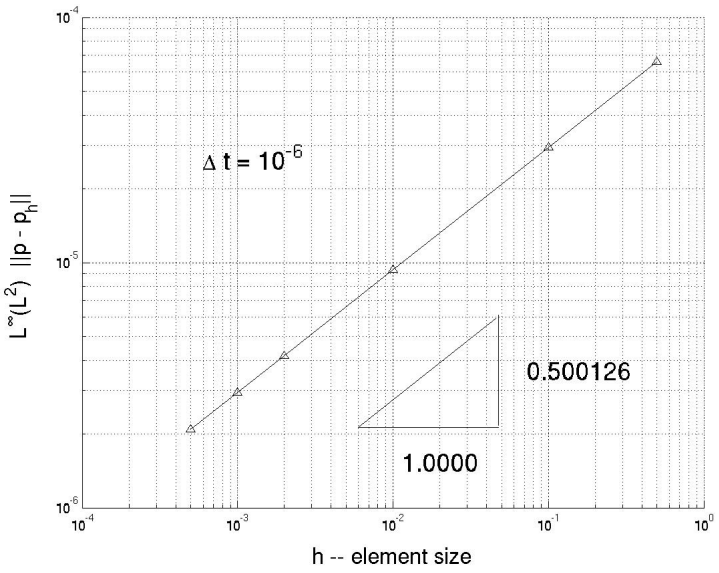

Figure 4.7: $L^{\infty}(L^2)$ pressure error in Terzaghi's problem. Physical parameters used: $E = 1e7;\ \nu = 0.4;\ c_o = 5e-1;\ \boldsymbol{\kappa} = 1e-4$.

4.1.1.3 Summary

The convergence rate ($\approx 0.5^{+}$) as $\Delta t \rightarrow 0$ for each set of parameters conforms to the expectation that the lack of regularity in the pressure solution would degrade performance. Although the rates of convergence are the same, the second set of parameters produces an overall smaller error than that found in the first.

From the convergence rate and the theorem on error estimates, one might be led to conclude that $p \in L^2(H^{\frac{3}{2}+\epsilon})$ for some small positive $\epsilon << 1$. In this way, the finite element error estimates has been combined with numerical

results to possibly shed some light onto the regularity of the pressure solution.

4.1.2 Mandel's Problem

Mandel's problem is important because it admits an analytical solution in two dimensions on a finite domain. The original paper by Mandel [1953] presented only an analytical form for the pressure; later Abousleiman et al. [1996] extended the results to include analytical expressions for the displacement and stress. It is therefore an excellent model to verify the accuracy of a poroelasticity algorithm. Mandel's problem is also fascinating from the point of view that it clearly illustrates that solid-fluid interactions can lead to unexpected behavior, and thus highlights the need for poroelasticity theory in practice.

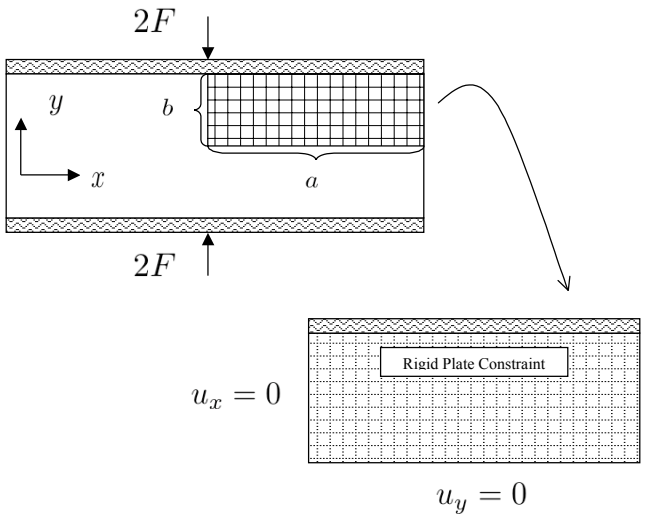

Figure 4.8: Mandel's Problem Formulation. Because of the quarter symmetry of the problem, the domain can be reduced to the upper right hand quadrant.

For a precise formulation of the problem, consider a poroelastic slab of extent $2a$ in the x-direction, $2b$ in the y-direction, and infinitely long in the z-direction. The slab is sandwiched in between two rigid plates as shown in Figure 4.8. A downward surface force of magnitude $2F$ is applied to the top plate at time $t = 0^+$, and an equal but upward force is applied to the bottom plate. Since the plates are considered rigid, there is the additional constraint that the vertical displacements at the top and bottom, respectively, are uniform. This requirement ensures that the slab remains in contact with the plates. This constraint is incorporated into the method by use of Lagrange multipliers. In addition, at all times the slab at $x = \pm a$ remains drained, thus leading to the boundary condition $p(\pm a, y, t) = 0$.

The application of load causes an instantaneous and uniform pressure increase throughout the domain; the theory predicts this to be $p^+ \equiv \lim_{t\to 0} p(\mathbf{x}, t) = \frac{1}{3a}B(1+\nu_u)F$. Also, as it turns out that $x-$displacement depends only on x and t, and the $y-$displacement depends on y and t. Thus, the following $x-$ and $y-$ settlements may be defined: $s_x(t) = u_x(x = a, t)$, and $s_y = u_y(y = b, t)$. It is predicted that there will be instantaneous and asymptotic settlements,

$$\begin{aligned}
s_{xo} &\equiv \lim_{t\to 0} s_x(t) = \frac{F\nu_u}{2\mu}, \\
s_{x\infty} &\equiv \lim_{t\to \infty} s_x(t) = \frac{F\nu}{2\mu}, \\
s_{yo} &\equiv \lim_{t\to 0} s_y(t) = \frac{-F(1-\nu_u)b}{2\mu a}, \\
s_{y\infty} &\equiv \lim_{t\to \infty} s_y(t) = \frac{-F(1-\nu)b}{2\mu a}.
\end{aligned}$$

At this point, it is useful to remark that the problem is symmetric about

the $x-$ and the $y-$ axes, and is independent of the $z-$direction. Therefore, it is possible to reduce the computational domain to only the upper right quadrant of the xy-plane as illustrated in Figure 4.8. Therefore, the following boundary and initial conditions on the domain $(0, a) \times (0, b)$ are incorporated into the computational routine:

$$\begin{aligned}
u_x &= 0, \quad x = 0; \\
u_y &= 0, \quad y = 0; \\
\sigma_{yy} &= t_N(t), \quad y = b, \quad = 0 \text{ otherwise}; \\
p &= 0, \quad x = a; \\
-\frac{1}{\mu_f}\boldsymbol{\kappa}\nabla p \cdot \boldsymbol{\nu} &= 0, \quad x = 0, \quad y = 0, \quad y = b; \\
u_y(x, y = b) &= \textit{constant} \quad \text{(Rigid Plate Constraint)}; \\
p|_{t=0} &= \mathbf{u}|_{t=0} \; = \; 0.
\end{aligned}$$

Here, u_x and u_y are the components of the displacement vector, $\mathbf{u}$, and $t_N(0) = 0,\ t_N(t) = -2F,\ t > 0$. It is also useful to remark that the constant in the rigid plate constraint *is* time-dependent.

The analytical series solutions for the pressure, displacement, and stress as found in Abousleiman et al. [1996]:

$$
\begin{aligned}
p &= \frac{2FB(1+\nu_u)}{3a}\sum_{n=1}^{\infty}\frac{\sin\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\Big(\cos\frac{\alpha_n x}{a}-\cos\alpha_n\Big)\exp(-\alpha_n^2 c_f t/a^2),\\
u_x &= \Big[\frac{F\nu}{2\mu a}-\frac{F\nu_u}{\mu a}\sum_{n=1}^{\infty}\frac{\sin\alpha_n\cos\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\exp(-\alpha_n^2 c_f t/a^2)\Big]x\\
&\quad+\frac{F}{\mu}\sum_{n=1}^{\infty}\frac{\cos\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\sin\frac{\alpha_n x}{a}\exp(-\alpha_n^2 c_f t/a^2),\\
u_y &= \Big[\frac{-F(1-\nu)}{2\mu a}+\frac{F(1-\nu_u)}{\mu a}\sum_{n=1}^{\infty}\frac{\sin\alpha_n\cos\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\exp(-\alpha_n^2 c_f t/a^2)\Big]y,\\
\sigma_{yy} &= -\frac{F}{a}-\frac{2FB(\nu_u-\nu)}{a(1-\nu)}\sum_{n=1}^{\infty}\frac{\sin\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\cos\frac{\alpha_n x}{a}\exp(-\alpha_n^2 c_f t/a^2)\\
&\quad+\frac{2F}{a}\sum_{n=1}^{\infty}\frac{\sin\alpha_n\cos\alpha_n}{\alpha_n-\sin\alpha_n\cos\alpha_n}\exp(-\alpha_n^2 c_f t/a^2).
\end{aligned}
$$

It should be noted that all other components of the stress tensor are zero, $\sigma_{xx} = \sigma_{xy} = 0$. Also, for the above equations α_n represents the positive solutions to the nonlinear equation

$$\tan\alpha_n = \frac{1-\nu}{\nu_u-\nu}\alpha_n,$$

and must be solved for numerically.

An interesting feature of Mandel's problem is the unexpected increase in pressure (above the initial increase) near the center of the medium. This is known as the *Mandel-Creyer* effect, and is a phenomenon particular to poroelastic material. The increase occurs because the deformation and rigid plate condition create an effect similar to that of a source term in the pressure equation. Indeed, Coussy [2004] shows that one can reduce the diffusion equation to one involving the normalized pressure and a source as

$$\frac{\partial \hat{p}}{\partial \hat{t}} - \frac{\partial^2 \hat{p}}{\partial \hat{x}^2} = 2 \sum_{n=1}^{\infty} \frac{\alpha_n^2 \sin \alpha_n \cos \alpha_n}{\alpha_n - \sin \alpha_n \cos \alpha_n} \exp(-\alpha_n^2 \hat{t}). \tag{4.3}$$

The source term is time dependent only (independent of $\mathbf{x}$), and it can be quite large at early times. The interesting behavior has been confirmed experimentally in the work of Gibson et al. [1963], Verruijt [1965].

4.1.2.1 A Discussion of Regularity

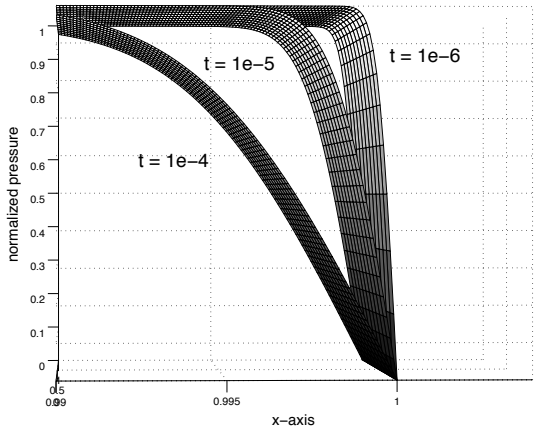

Figure 4.9: Mandel's pressure solution and boundary layer as $t \to 0$.

As demonstrated in the example problem of Terzaghi, the order of convergence of the CG/Mixed numerical solution is limited by the regularity of the true solution. For very similar reason, Mandel's problem suffers from a less than optimal convergence rate due to the lack of regularity of pressure solution.

Figure 4.9 shows the true pressure solution for a typical choice of physical parameters. As $t \to 0$ a very large gradient occurs. Naturally, this leads

one to suspect that the regularity of the pressure might be limited.

Figure 4.10 shows the numerical approximation of $|\frac{\partial p}{\partial x}|^2$ and $|\frac{\partial^2 p}{\partial x^2}|^2$ for a fixed value in the y-direction at progressively smaller times (from top to bottom). The maximum value of each term increases for smaller times t, but the area also decrease over which each term is non-zero.

To determine the regularity of the pressure, $|\frac{\partial p(t)}{\partial x}|^2$ and $|\frac{\partial^2 p(t)}{\partial x^2}|^2$ are integrated over the spatial dimension to give values of $||p_x(t)||_0^2$ and $||p_{xx}(t)||_0^2$. Figure 4.11 shows their respective values as a function of time. It is discovered that

$$||p_x(t)||_0^2 \approx O(t^{-0.488}) \tag{4.4}$$

$$||p_{xx}(t)||_0^2 \approx O(t^{-1.447}) \tag{4.5}$$

The order approximation (4.4) implies that $p \in L^2(H^1)$, but (4.5) shows that $p \notin L^2(H^2)$.

It is, thus, reasonable to assume that $p \in L^2(H^{1+s})$ for some $s \in [0, 1)$. Furthermore, it is precisely this value of s which limits the best rate of convergence which one might expect from the CG/Mixed algorithm for Mandel's problem. Similar to Terzaghi's problem, (4.4)-(4.5) imply satisfaction of Showalter's bound, $||p||_2 \leq \frac{C}{t}$, but it has already been shown a that less than optimal convergence rate is expected.

4.1.2.2 Computational Results

For numerical verification of our algorithms, a force, $F = 2e + 3$, and the following set of material parameters are chosen:

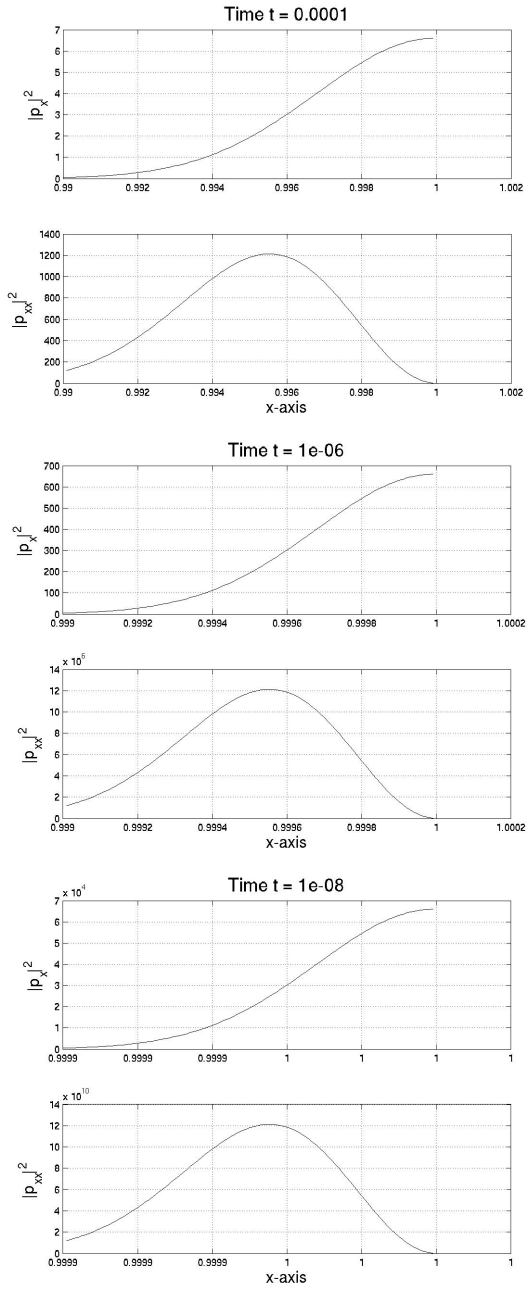

Figure 4.10: Mandel's Problem. $|\frac{\partial p}{\partial x}|^2$ and $|\frac{\partial^2 p}{\partial x^2}|^2$ at different times. $\sup(|\frac{\partial^2 p}{\partial x^2}|^2) \approx 1.4e3$ $(t = 1e-4)$, $1.4e7$ $(t = 1e-6)$, and $1.4e11$ $(t = 1e-8)$.

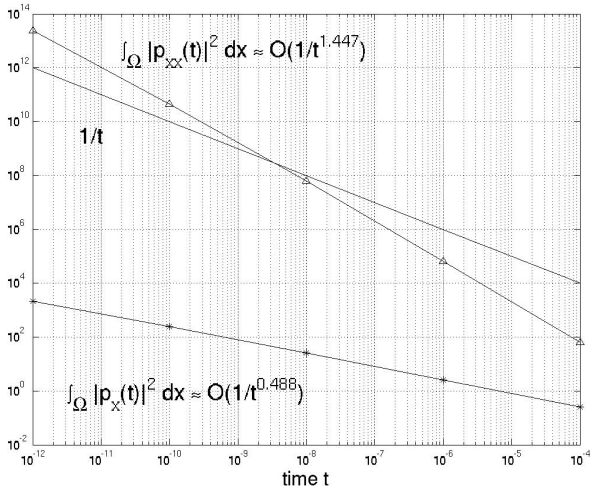

Figure 4.11: Mandel's Problem. The pressure integrals $||p_x(t)||_0^2$ and $||p_{xx}(t)||_0^2$.

$$E \;=\; 1e+4, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e-1, \quad \boldsymbol{\kappa} = 1e-2.$$

Standard continuous linear elements are used to approximate the displacement, and the lowest Raviart-Thomas space is for the flow variables.

The top image in Figure 4.12 shows how the computed pressure error, $||p - \bar{p}||_{L^\infty(L^2)}$, varies with element size, h. To minimize the effects of the error produced by the time discretization, a small time step of $\Delta t = 1e-6$ is chosen. The convergence rate is determined by the slope of the logarithm of the error (≈ 0.500117). Although results are shown only for the backward Euler time-tepping scheme, the results for any other time-stepping scheme are indistinguishable.

Because Mandel's problem offers a solution for the stress tensor, the displacement error can be measured in the $L^\infty(L^2)$ norm. The bottom image

in Figure 4.6 shows how the computed displacement error, $||\mathbf{u} - \bar{\mathbf{u}}||_{L^\infty(H^1)}$. The convergence rate is determined by the slope of the logarithm of the error (≈ 0.500202).

4.1.2.3 Summary

The convergence rate ($\approx 0.5^+$) for both the pressure and displacement conforms to the expectation that the lack of regularity in the pressure solution would degrade performance. From the convergence rate and the theorem on error estimates, one might be led to conclude that $p \in L^2(H^{\frac{3}{2}+\epsilon})$ for some small positive $\epsilon << 1$. Thus, the finite element error estimates has been combined with numerical results to possibly shed some light onto the regularity of the pressure solution.

4.2 Problems with a Null Constrained Specific Storage Coefficient

There are cases in linear poroelasticity when the fluid and solid constituents are both incompressible. In this case, (2.7) shows that $c_o \to 0$. As the previous chapter demonstrated, there is a theoretical distinction for results involving a null constrained specific storage coefficient value. The fact that the pressure error converges only in a weaker norm suggests the potential for some oscillations in the numerical pressure field.

Therefore, in this section the CG/Mixed numerical solution is tested against *Barry and Mercer's problem* which obtains an analytical solution for the case that $c_o = 0$. Fairly good results are found for one choice of parameters, but when $c_o = 0$ is paired with a small permeability and a short time step, non-physical pressure oscillation, or *locking* occurs.

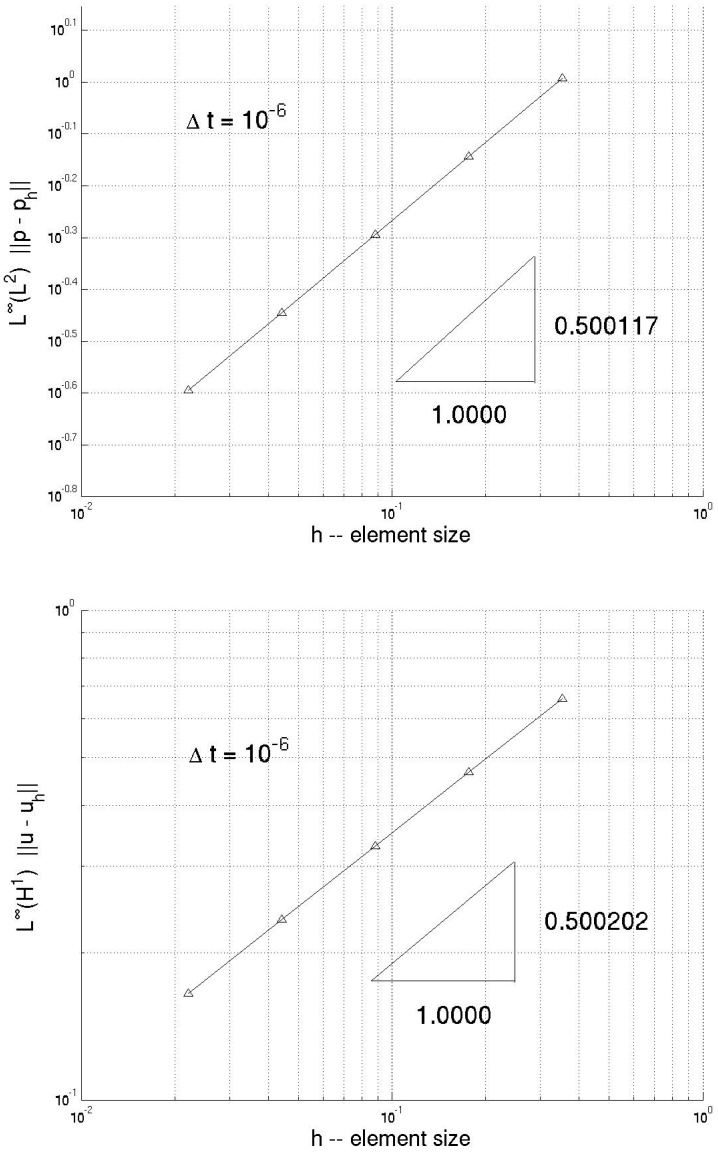

Figure 4.12: $L^\infty(L^2)$ pressure error (top) and $L^\infty(H^1)$ displacement error (bottom) in Mandel's problem. Physical parameters used: $E = 1e4$; $\nu = 0.2$; $c_o = 1e-1$; $\boldsymbol{\kappa} = 1e-2$.

In addition, the CG/Mixed scheme is re-tested against Mandel's problem, this time with $c_o = 0$. Although no analytical solution is available for this case, the numerical solution exhibits locking when paired with a certain value of Young's modulus, E.

These two examples will highlight the necessity for examining more robust numerical schemes. The fact that locking only occurs when $c_o = 0$ suggests the necessity of this condition for numerical degeneracy. Conversely, as Barry and Mercer's problem ($\boldsymbol{\kappa}$) and Mandel's problem (E) will illustrate that other physical parameter play a role in locking. The remaining results of this chapter, thus, are intended to provide a clue to the possible reason for locking, something which will be covered extensively in the subsequent chapter.

4.2.1 The Problem of Barry and Mercer

A somewhat recent problem in poroelasticity with an analytical solution on a finite two-dimensional domain is the point-source problem and analytical series solution of Barry and Mercer [1999b]. The problem corresponds to a point-source sine wave on a rectangular domain. The boundary conditions are contrived in order to admit a solution, and thus the problem does not correspond to a realistic relevant situation. However, the problem and solution have interesting mathematical properties which make it particularly suitable for numerical validation. For one thing, this problem offers a solution for the case when the constrained specific storage coefficient is zero, $c_o = 0$. A less important, but still interesting, feature is that the nondimensionalized deformation and pressure solution are each independent of the dimensionless parameter $m = 1 + \frac{\lambda}{\mu}$.

Figure 4.13 describes the problem formulation. The boundary of the domain $(0, a) \times (0, b)$ is drained on all sides, $p|_{\partial\Omega} = 0$. Along the boundaries $y = 0, y = b$ the deformation constraint is $\mathbf{u}_x = 0, \frac{\partial \mathbf{u}_y}{\partial y} = 0$, and along the boundaries $x = 0, x = a$, the deformation constraint is $\mathbf{u}_y = 0, \frac{\partial \mathbf{u}_x}{\partial x} = 0$. In addition, there is a source term,

$$s_f(t) = 2\beta F(\beta t),$$

where $\beta = a^{-1}b^{-1}(\lambda+2\mu)\tilde{\boldsymbol{\kappa}}$ [4] is the material dependent source strength. $\hat{t} = \beta t$ is referred to as the *normalized time*.

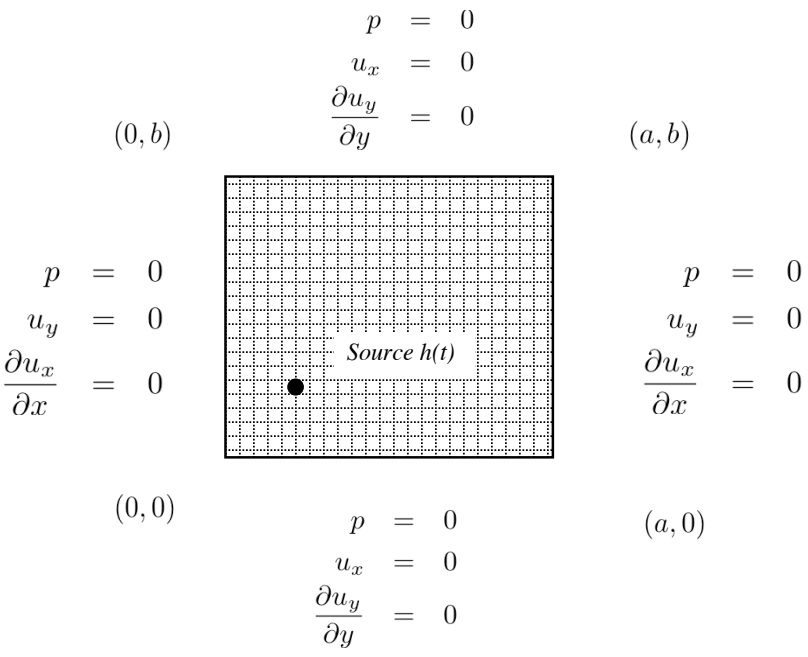

Figure 4.13: Boundary conditions for Barry and Mercer's source problem

[4] Here, $\tilde{\boldsymbol{\kappa}}$ is a constant, although the tensorial notation is retained for uniformity.

For simplicity, the test case here sets $a = b = 1$. Moreover, as in Barry and Mercer [1999b], $x_o = 0.25, y_o = 0.25$, and

$$F(s) = \delta(x - x_o)\delta(y - y_o)\sin(s),$$

an oscillating point source. Here, δ is the Dirac delta function To present the solution, one may refer to the following,

$$\gamma_n = n\pi, \quad \gamma_q = q\pi, \quad \gamma_{nq} = \gamma_n^2 + \gamma_q^2,$$

$$\hat{p}(n,q,t) = -\frac{2\sin(\gamma_n x_o)\sin(\gamma_q y_o)}{\gamma_{nq}^2 + 1}\big(\gamma_{nq}\sin(\beta t) - \cos(\beta t) + e^{-\gamma_{nq}\beta t}\big),$$
$$\hat{u}(n,q,t) = \frac{\gamma_n}{\gamma_{nq}}\hat{p}(n,q,t), \qquad \hat{w}(n,q,t) = \frac{\gamma_q}{\gamma_{nq}}\hat{p}(n,q,t).$$

The solution in then

$$\begin{aligned}
p(x,y,t) &= -4(\lambda + 2\mu)\sum_{n=1}^{\infty}\sum_{q=1}^{\infty}\hat{p}(n,q,t)\sin(\gamma_n x)\sin(\gamma_q y), \\
\mathbf{u}_x(x,y,t) &= 4\sum_{n=1}^{\infty}\sum_{q=1}^{\infty}\hat{u}(n,q,t)\cos(\gamma_n x)\sin(\gamma_q y), \\
\mathbf{u}_y(x,y,t) &= 4\sum_{n=1}^{\infty}\sum_{q=1}^{\infty}\hat{w}(n,q,t)\sin(\gamma_n x)\cos(\gamma_q y).
\end{aligned}$$

4.2.1.1 Computational Results

Unlike Terzaghi's and Mandel's solutions, the pressure and displacement solutions are double infinite summations and depend on both the x and y space variables. This makes it computationally cumbersome to demonstrate

convergence rates. In addition, the stress variable solution is not provided. Therefore, the results in this subsection are intended to only provide an indication of the agreement between the analytical and numerical solutions.

For numerical verification of the CG/Mixed algorithm, a problem on the unit domain is posed, and the following set of material parameters are chosen:

$$E = 1e + 5, \quad \nu = 0.1, \quad \alpha = 1, \quad c_o = 0, \quad \boldsymbol{\kappa} = 1e - 2.$$

Standard continuous linear elements are used to approximate the displacement, and the lowest Raviart-Thomas space is for the flow variables.

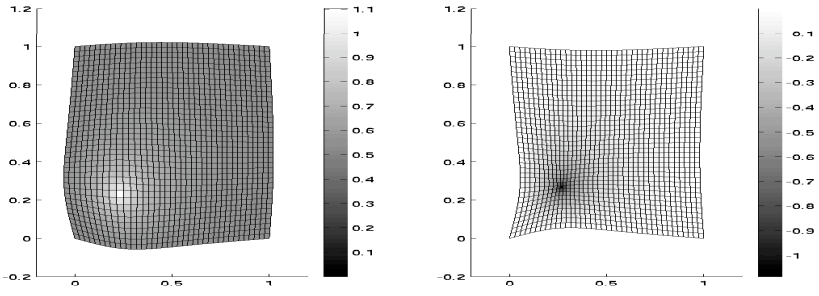

Figure 4.14: The problem of Barry and Mercer: CG/Mixed solution at times $\hat{t} = \frac{\pi}{2}$ (left), and $\hat{t} = \frac{3\pi}{2}$ (right).

Using the prescribed set of physical parameters, Figure 4.14 shows the numerical solution of the pressure (shaded field) and deformation produced by the CG/Mixed algorithm. At $\hat{t} = \frac{\pi}{2}$ (left), the source is positive; this fluid injection causes an expansion of the poroelastic medium. On the other hand, at $\hat{t} = \frac{3\pi}{2}$ (right), the source is negative; the resultant fluid withdrawal causes a contraction of the medium.

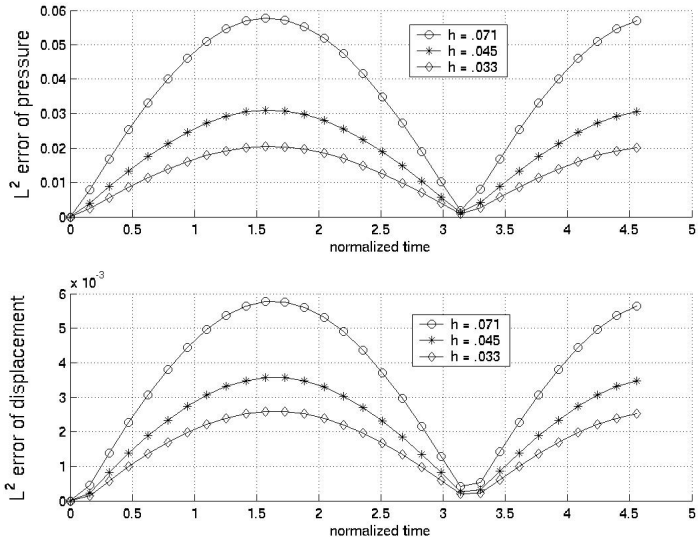

Figure 4.15: CG/Mixed error in Barry and Mercer's problem.

Figure 4.15 illustrates the effect of element size on the error of the pressure and displacement. The backward Euler time-stepping scheme is used and a relatively large normalized time step of $\Delta\hat{t} = \frac{0.1\pi}{2}$ is used for this test case. The graphic shows the error of the normalized pressure (top), $||\hat{p} - \hat{p_h}||_0$, and normalized displacement (bottom), $||\hat{\mathbf{u}} - \hat{\mathbf{u}}||_0$, using various element sizes. As one can see, a smaller step size leads to a reduction in error at most times. However, for a given time step, the error seems to be independent of element size (and minimal) at times $\hat{t} = n\pi$. This is precisely when the source term vanishes.

Independence of $1 + \frac{\lambda}{\mu}$

It was mentioned in the opening discussion of Barry and Mercer's problem that the normalized analytical pressure and deformation solutions are independent of $m = 1 + \frac{\lambda}{\mu}$. Furthermore, since $1 + \frac{\lambda}{\mu} = \frac{1}{1-2\nu}$, this implies an independence of Poisson's ratio, ν, over its acceptable range.

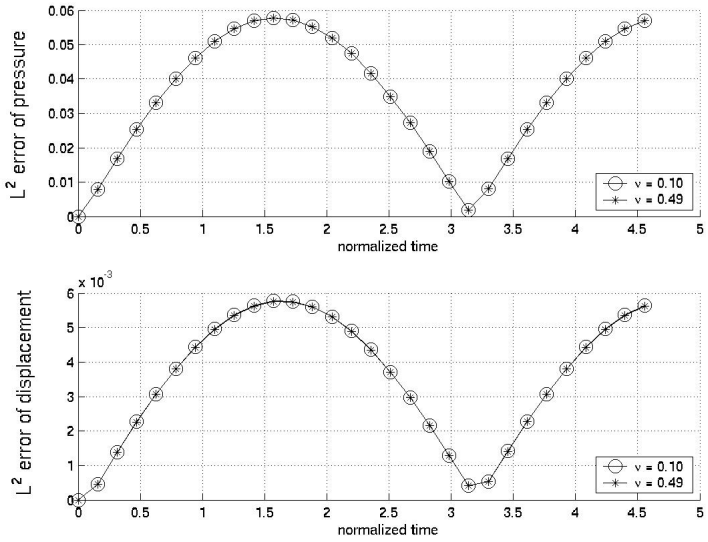

Figure 4.16: Independence of $1 + \frac{\lambda}{\mu} = \frac{1}{1-2\nu}$

For concreteness, Figure 4.16 is included which demonstrates that this independence carries over to the numerical algorithm. For the experiment, there are two cases: $\nu = 0.1$ and $\nu = 0.49$. The top graph shows the error of the normalized pressure, $||\hat{p_h}^{\nu} - \hat{p}||_0$, and shows identical error estimates for the normalized pressure errors. The bottom graph shows the normalized displacement, $||\hat{\mathbf{u}}^{\nu} - \hat{\mathbf{u}}||_0$, and shows very good agreement between the normalized deformation errors. Thus, since the analytical solution is independent

of ν, the triangle inequality implies,

$$||\hat{p_h}^{\nu=0.1} - \hat{p_h}^{\nu=0.49}||_0 \leq ||\hat{p} - \hat{p_h}^{\nu=0.1}||_0 + ||\hat{p} - \hat{p_h}^{\nu=0.49}||_0,$$

and a similar result for the deformation. Since each term is quite small on the rhs of the above equation, one concludes that $\hat{p_h}^{\nu=0.1} \approx \hat{p_h}^{\nu=0.49}$. Thus, the numerical solution for the normalized pressure (and deformation) retains the material parameter independence.

4.2.1.2 Non-Physical Pressure Oscillations

In the preceding example, good agreement is found between the analytical and CG/Mixed solutions despite the fact that $c_o = 0$. However, if the permeability is set to a smaller value (e.g. $\boldsymbol{\kappa} = 1e-6$), and if a smaller time step is used ($\Delta\tilde{t} = \frac{1e-5\pi}{2}$), problems with the numerical solution occur. Figure 4.17 shows non-physical pressure oscillations arising with this choice of parameters for the initial time step solution. However, the oscillations smooth out significantly as time progresses.

4.2.2 Mandel's Problem Revisited

The same time step, $\Delta t = 1e-6$ and physical parameters used earlier in this chapter are reused here except that the constrained specific storage coefficient has been set to zero, $c_o = 0$. Figure 4.18 shows that the early time numerical solutions experience some locking, whereas at later times the oscillations dissipate.

Alternatively, if the constrained specific storage coefficient is kept at a zero value and Young's modulus is increased from $1e4$ to $1e5$, Figure 4.19

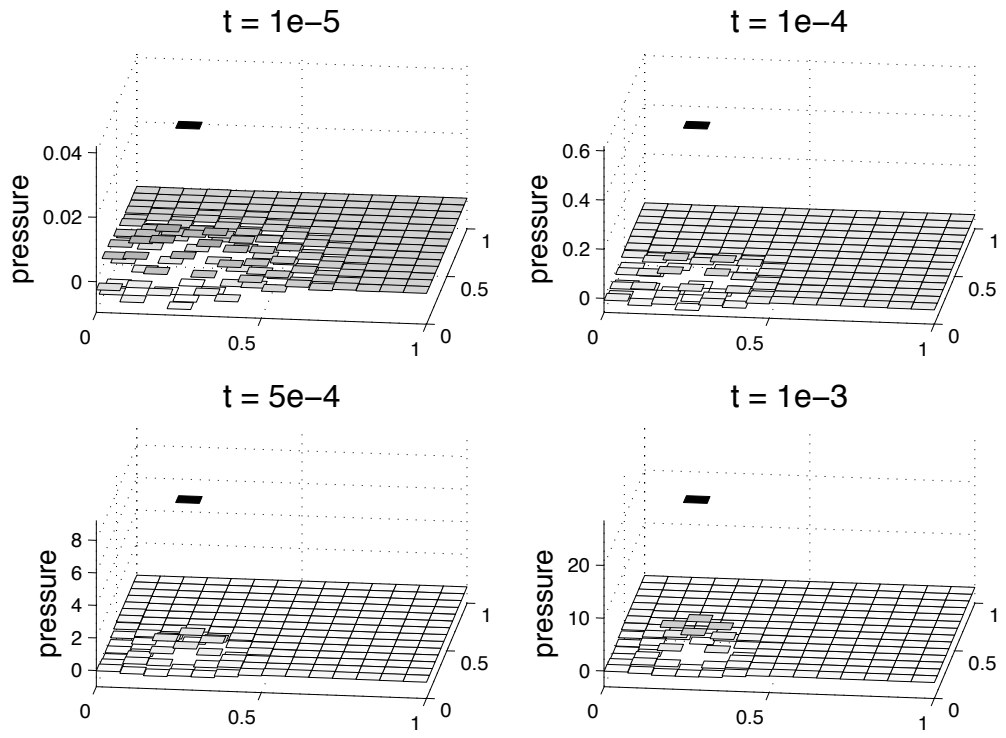

Figure 4.17: With a reduced time step and permeability, locking occurs in Barry and Mercer's problem for the very early time solutions, but dissipates later. Physical parameters include $c_o = 0$ and $\boldsymbol{\kappa} = 1e - 6$.

shows that no locking occurs even for the initial time step.

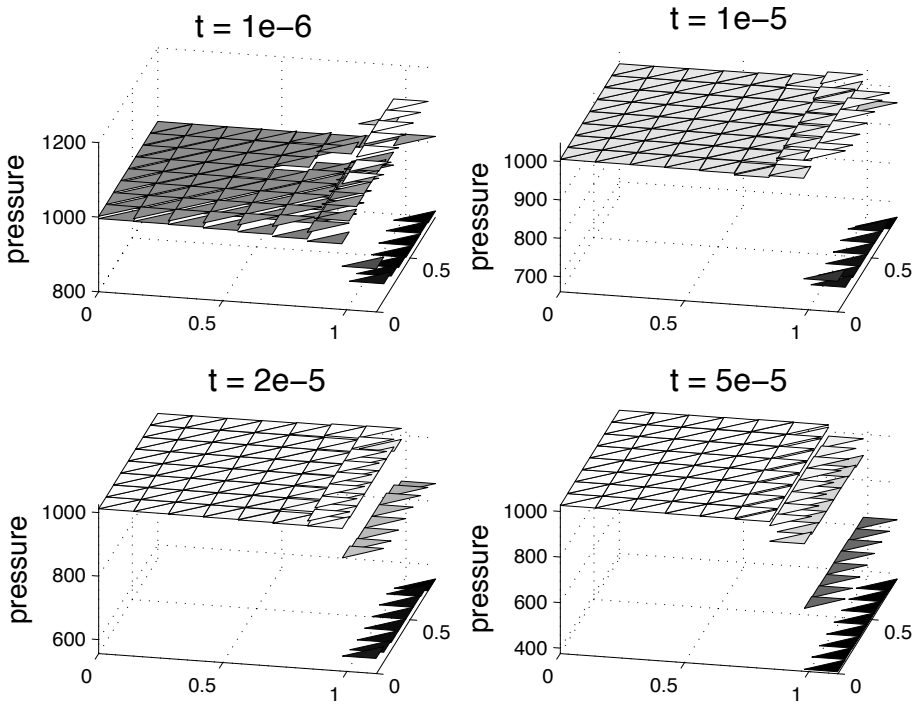

Figure 4.18: A null constrained specific storage coefficient causes initial non-physical pressure oscillations in Mandel's problem. However, the locking dissipates quite rapidly as time progresses. Physical parameters include $c_o = 0$ and $E = 1e4$.

4.3 Discussion

The results of this chapter confirm the theoretical predictions regarding accuracy if viewed in the context of regularity. But that caveat is in itself motivation for exploring other means to improve accuracy. Subsequent chapters will follow through with this idea by incorporating adaptive grid technique into

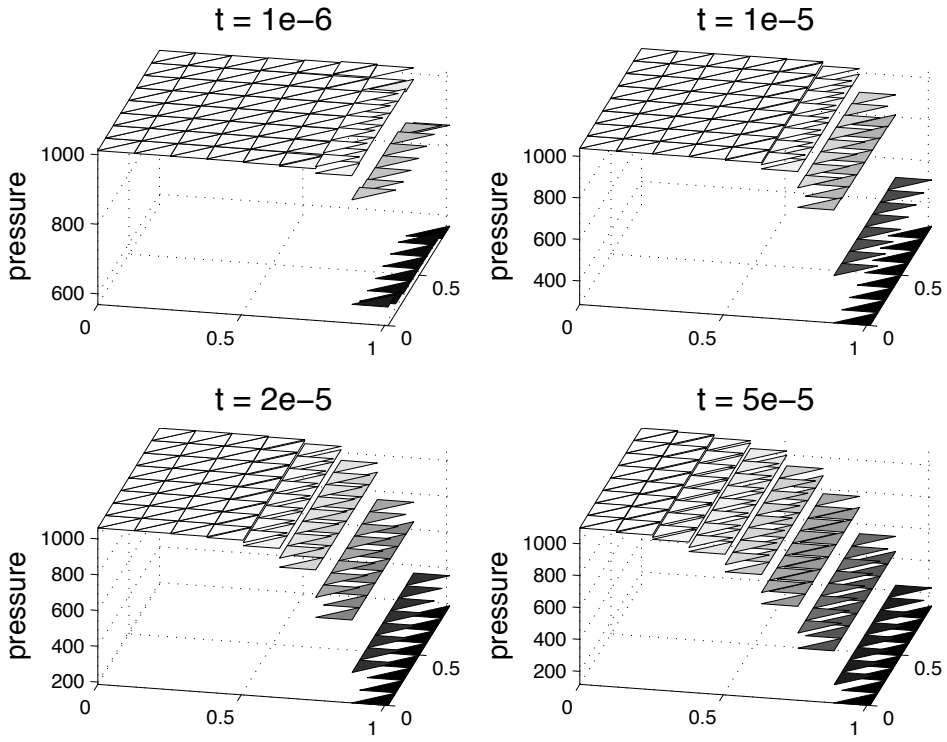

Figure 4.19: Despite a null constrained specific storage coefficient value, with a larger Young's modulus there is no locking in in Mandel's problem. Physical parameters include $c_o = 0$ and $E = 1e5$.

the CG/Mixed scheme, and demonstrate results which justify this approach.

An important theoretical distinction was made in the analysis of the CG/Mixed scheme. For a strictly positive constrained specific storage coefficient, the pressure attains optimality when the error measured in the $L^{\infty}(L^2)$ norm; for the general case only the weaker $L^2(L^2)$ norm endures as a theoretically satisfactory measurement. For the case $c_o = 0$, numerical examples show that non-physical pressure oscillations sometimes occur for short time steps. These oscillations obviously increase the $L^2(\Omega)$ for those early times. But as the oscillations dissipate through time so does the $L^2(\Omega)$ pressure error. This explains the necessity of using the weaker $L^2(L^2)$ to measure the pressure error when the constrained specific storage coefficient may be zero.

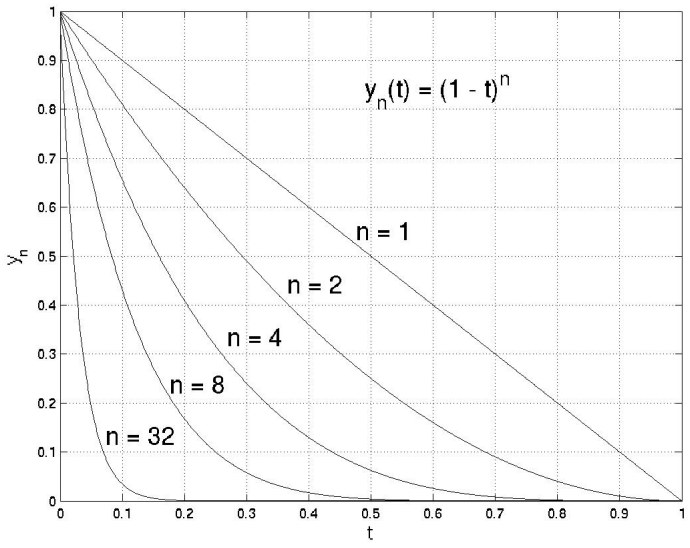

Figure 4.20: Functions converging in L^2 but not in L^{∞}.

Figure 4.20 illustrates the point. Let $y_n(t) = (1-t)^n$ for $t \in [0,1]$.

Then

$$\int_0^1 y_n^2(t)\ dt \ = \ \frac{1}{2n+1}.$$

So $y_n \to 0$ in L^2. In fact, $y_n \to 0$ pointwise a.e. However, since

$$\operatorname{ess\,sup} y_n \ = \ 1, \quad \forall n \geq 1,$$

y_n does not converge to the zero function in L^∞.

Let $e_p \equiv p - \bar{p}$ be the finite element error. Then for $c_o = 0$, $y_n(t)$ may be considered a rough approximate for $||e_p(t)||_{L^2(\Omega)}$ and n for the number of elements in a uniformly refined grid. In this analogy, the numerical pressure oscillations initiate a large $L^2(\Omega)$ error and persist for an interval of time, but this interval decreases precipitously as the grid is refined. So for a large time step or for a refined grid, the initial time step may miss this range and avoid the oscillations. In practice, this oscillation time interval is dependent on material parameters in addition to the constrained specific storage coefficient. This explains why locking occurred for only half of the test parameters for both Barry and Mercer's problem and Mandel's problem.

For those situations where a particular time step and material parameters leads to locking, the question remains as to whether or nor it can be eliminated. To better understand the problem, the next chapter examines more closely the problem of locking and proposes two potential remedies: using discontinuous elements for the deformation and adaptive grid refinement.

Chapter 5

The Problem of Locking

Sometimes too much reliance is placed in finite element error estimates. A typical estimate usually comes in the form

$$||q - q_h||_0 \leq Ch^{r+1}||q||_{r+1}.$$

Unfortunately, the constant C usually depends on certain physical parameters and can become quite large. This makes the error estimate practically worthless. For instance, for linear elasticity C depends strongly on the Lamé coefficient λ, and degrades as $\lambda \to \infty$. In poroelasticity, C depends on a combination of the constrained specific storage coefficient, c_o, the permeability, $\boldsymbol{\kappa}$, and the size of the time step. In the case of a zero constrained specific storage coefficient value, low permeability, and a short time step, the error estimates in poroelasticity (at least using continuous elements) can become unreliable. *Locking* is the general name for this problem, and can manifest itself in different ways. In linear elasticity, locking leads to oscillations in the stress field and some problems with the deformation field. In poroelasticity, the most common problem is spurious, non-physical pressure oscillations.

The question remains why this happens. To understand the problem of locking, an examination is first given of how the problem arises in linear elasticity from both a theoretical and heuristic viewpoint. Subsequent is a

short discussion of how locking has been overcome in that field.

Attention is then directed to the problem of locking in poroelasticity. It is found that similar methods from linear elasticity might be successfully applied to eliminate locking in linear poroelasticity. In particular, analogous reasoning suggests the use of discontinuous elements for displacements. Alternatively, it is also shown that the CG/Mixed method can sometimes be used in a locking-free manner *if* small enough elements are used near a boundary layer. This provides motivation for the development for adaptive grid refinement methods. This also points out a significant difference in the nature of locking between linear elasticity and linear poroelasticity. No matter how small the elements are made in linear elasticity, it is always possible to produce a numerical solution with stress oscillations when using continuous linear elements [Brenner and Scott, 1994].

5.1 Overcoming Locking in Elasticity

While the ultimate goal of this chapter is to understand and prevent locking in poroelasticity, it is understandable to begin the discussion with a look at the problem of locking in linear elasticity[1] for two reasons. First is that locking is well understood and techniques have been developed which have successfully tackled the problem. Second is the fact that there are similarities in the problems that arise in both linear elasticity and poroelasticity, and thus a quick look at locking in the elastic case might point toward a solution in the poroelastic case.

[1]Most of the following discussion of the analytical aspects of the problem of locking in linear elasticity is based on treatment found in Brenner and Scott [1994].

The primary reason for the potential similarities is the governing equations for the two problems. In examining the equation for linear elasticity, we shall see that it actually is embedded as part of the equations for poroelasticity, but without the pressure term. Specifically, the equation and boundary equations are: Find $\mathbf{u} \in H^1$ such that

$$\begin{aligned} -(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla^2\mathbf{u} &= \mathbf{f}, \\ \mathbf{u} &= \mathbf{g}_D, \quad \forall \mathbf{x} \in \Gamma_D, \\ \boldsymbol{\sigma}\nu &= \mathbf{g}_N, \quad \forall \mathbf{x} \in \Gamma_N, \end{aligned}$$

where Γ_D, and Γ_N are the Dirichlet and Neumann boundaries, respectively.

The variational form of the above problem in finite dimensions is cast in the form: Find $\bar{\mathbf{u}} \in \mathbf{V}_h$ such that for all $\mathbf{v} \in \mathbf{V}_h \cap \{\mathbf{v} : \mathbf{v}|_{\Gamma_D} = 0\}$

$$a_{elas}(\mathbf{u}, \mathbf{v}) = (\mathbf{f}, \mathbf{v}) + < \mathbf{g}_N, \mathbf{v} >_{\Gamma_N},$$

where $\mathbf{V}_h$ is some appropriate finite-dimensional space, and

$$a_{elas}(\mathbf{u}, \mathbf{v}) = (2\mu(\epsilon(\mathbf{u}) : \epsilon(\mathbf{v})) + \lambda(\nabla\cdot\mathbf{u}, \nabla\cdot\mathbf{v})$$

is the standard bilinear form arising in elasticity. By assuming[2] $|\Gamma_D| > 0$, Korn's inequality implies that a_{elas} is coercive and thus the Lax-Milgram theorem tell us that the above variational problem has a solution. Now, from the standard methods for deriving error estimates, it follows that

[2]This is something that will be assumed throughout this section for simplicity.

$$||\mathbf{u} - \bar{\mathbf{u}}||_1 \leq C(\lambda, \mu)h||\mathbf{u}||_2,$$

where it is emphasized that the constant C depends on λ and μ. The first hint of the trouble that leads to locking is the regularity estimate summarized in Brenner and Scott [1994]:

$$\begin{aligned} &||\mathbf{u}||_2 + \lambda||\nabla \cdot \mathbf{u}||_1 \\ &\quad \leq C(||\mathbf{f}||_o + ||w_1||_2 + ||w_2||_1), \end{aligned}$$

where C is independent of λ and μ, and $\mathbf{g}_D = w_1|_{\Gamma_o}$ and $\mathbf{g}_N = w_2|_{\Gamma_N}$.

5.1.1 The Problem

One can see from the above equation that as $\lambda \to \infty$ it is necessary that $\nabla \cdot \mathbf{u} \to 0$. Thus, the material becomes incompressible. This is why use of continuous linear elements on triangle, or continuous bilinear elements on quadrilaterals is problematic – because the only divergence-free vectors belonging to either of those spaces are constant vectors.

To see why this is the case, consider the quadrilateral case in two dimensions. A typical basis element is of the form

$$\Phi = \begin{pmatrix} a + bx + cy + dxy \\ e + fx + gy + hxy \end{pmatrix},$$

for constants a, b, c, d, e, f, g, h. Now the requirement that Φ be divergence-free demands that

$$0 = \nabla \cdot \Phi = b + dy + g + hx.$$

Since x and y are allowed to vary over the element domain, the above equation sets up the following constraints on the constants:

$$\begin{aligned} b &= -g, \\ d &= 0, \\ h &= 0. \end{aligned}$$

Hence, because of the divergence-free requirement, a basis which originally contained eight degrees of freedom has been reduced to five.

$$\Phi = \begin{pmatrix} a + bx + cy + dxy \\ e + fx + gy + hxy \end{pmatrix} \quad \overset{becomes}{\Longrightarrow} \quad \Phi = \begin{pmatrix} a + bx + cy \\ e + fx - by \end{pmatrix}. \tag{5.1}$$

If it is further demanded that the basis functions be continuous across the elements, then one is left with only two degrees of freedom for the entire domain; the only basis function available is the constant vector. In the case that $|\Gamma_D| > 0$, then there are actually no degrees of freedom left, as the solution must match the boundary.

So to summarize the conundrum: *In the case that $\lambda \to \infty$, the ensuing restriction is that the continuous (bi)linear basis is reduced to being a constant vector; on the other hand, the boundary condition may imply a non-constant deformation.* One cannot simply ameliorate this problem by ever-reducing the

element size. There is a fundamental problem that must be remedied by use of alternative strategies.

It should be noted that the discussion has been largely heuristic, but in fact a theoretical basis for the problem has been established in Brenner and Scott [1994]. There the authors in §9.3 prove that no matter how small the element size, if λ is large enough, there is a set of boundary condition such that the relative error $\frac{|\mathbf{u}-\bar{\mathbf{u}}|_1}{||\boldsymbol{\sigma}(\mathbf{u})||_0}$ is bounded below by a constant independent of the element size.

5.1.2 The Remedy

Having diagnosed the problem, the solution is easily spotted: *find a basis with non-constant, divergence-free vectors*. Indeed, there are bases available that satisfy that requirement. One such basis is the non-conforming method on triangles, where continuity at the midpoint is demanded [Brenner and Scott, 1994]. However, one possible drawback of this method is its restriction to triangles. In order to disscover another method which is applicable on both triangles and quadrilaterals, return to (5.1). It is the demand of continuity across elements that reduced the degrees of freedom from five to two (or worse). Thus, if continuity is no longer required, *five degrees of freedom are retained* on each element to approximate a non-constant divergence-free vector. Indeed, discontinuous methods have been developed in linear elasticity in Rivière and Wheeler [2000], and implemented in the work of Liu [2004].

5.1.3 Examples

Indeed, discontinuous methods have been developed in linear elasticity in Rivière and Wheeler [2000]. In order to demonstrate the effectiveness of

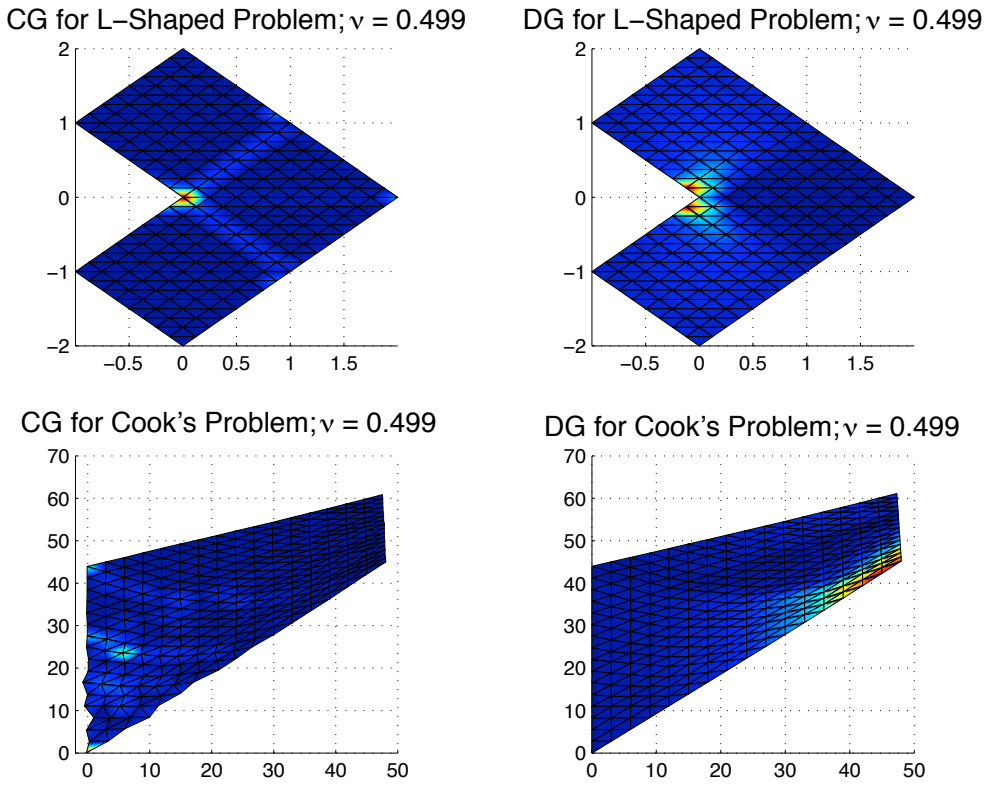

Figure 5.1: CG vs. DG in linear elasticity. The top row is on an L-shaped domain, the bottom row is Cook's problem. The figures above illustrate the problems encountered in a continuous Galerkin method (left column) when the material is nearly incompressible. For these cases where $\nu = .499...$, the CG method produces an inaccurate deviatoric stress field (the gray field) and deformation. On the other hand, by using discontinuous Galerkin (right column), the problems are eliminated.

the DG method in overcoming locking, one is refered[3] to Figure 5.1. In all of the numerical experiments, $\nu = .499...$, the equivalent of making λ very large. The top row represents an L-shaped domain with an applied deformation on the boundary; this causes a stress and deformation throughout the material. The colored field represents the deviatoric stress field. One can see that the CG method produces a non-physical stress field. On the other hand, using the DG method (NIPG version) one recovers a stress field resembling what one would expect in nature.

The bottom row illustrates the results from Cook's problem. This problem involves a material slab attached to a wall on the left. On the right, a traction exerts an upward pull. As seen, the CG method deteriorates badly in both the stress and deformation fields. Conversely, DG (IIPG version) completely solves the problem.

5.2 Overcoming Locking in Poroelasticity

As it turns out, one remedy for locking in poroelasticity is similar to that found in linear elasticity. That is, one needs a basis which contains non-constant divergence-free vectors. However, despite the fact that λ is also a material parameter in poroelasticity, it is not the source of the problem. To discover the problematic origin, one has to look more deeply.

5.2.1 The Problem

To begin, the flow equation (2.8b) is re-examined,

[3]Phillips' DG code for elasticity is compared with the code CG found in Alberty et al. [2002]. Their code works perfectly well for any compressible material, but like any method based on continuous linear elements, it suffers for the case $\lambda \to \infty$, or equivalently, $\nu \to 0.5$.

$$\frac{\partial}{\partial t}(c_o p + \alpha \nabla \cdot \mathbf{u}) - \frac{1}{\mu_f} \nabla \cdot \boldsymbol{\kappa}(\nabla p - \rho_f \; g) = \; s_f.$$

In the circumstance of no source term, $s_f = 0$, problems can occur when a small time step is combined with a null constrained specific storage coefficient value, $c_o = 0$, and a small permeability. To see why more completely, assume for simplicity that the permeability is constant and use the notation $k = \frac{1}{\mu_f}\boldsymbol{\kappa}$. If the above equation is discretized in time, one finds the approximate equation

$$\alpha \frac{\nabla \cdot \mathbf{u}^{n+1} - \nabla \cdot \mathbf{u}^n}{\Delta t} - k \; \nabla \cdot \nabla p^{n+1} \; = \; 0$$

Rearranging the above equation shows that

$$\alpha \nabla \cdot \mathbf{u}^{n+1} - \Delta t \; k \; \nabla \cdot \nabla \; p^{n+1} \; = \; \alpha \nabla \cdot \mathbf{u}^n \tag{5.2}$$

Assuming that $\mathbf{u}^0 = 0$, one finds the initial time step satisfies

$$\alpha \nabla \cdot \mathbf{u}^1 - \Delta t \; k \; \nabla \cdot \nabla p^1 \; = \; 0. \tag{5.3}$$

As discussed in the previous subsection, problems manifest when $\nabla \cdot \mathbf{u}^1 \to 0$ because continuous (bi)linear elements are incapable of approximating a non-constant, divergence-free deformation. To avoid this situation, (5.3) shows that $\Delta t \; k \; \nabla \cdot \nabla p^1$ needs to be non-negligible. There are two ways this might occur.

As shown earlier, problems such as Teraghi's and Mandel's problems admit pressure boundary layers, and problems like Barry and Mercer's also

have a large pressure gradient. In these cases $\nabla \cdot \nabla p^1$ becomes very large in certain regions so that $\Delta t\ k\ \nabla \cdot \nabla p^1$ would be sufficiently large to bound $\nabla \cdot \mathbf{u}^1$ away from zero. But for this to work in a numerical setting, the element size needs to be small enough so that the large flux is well approximated.

When the element size is not small enough to adequately resolve the boundary layer, the numerical approximation of the flux is most likely greatly undervalued. However, $\Delta t\ k\ \nabla \cdot \nabla p^1$ may still be adequately large *if* the time step and the permeability are large.

For later times, (5.2) shows that $\nabla \cdot \mathbf{u}^{n+1} \approx \nabla \cdot \mathbf{u}^n$ in those situations where the elements are too large for the numerical flux to approach the true size of the flux. This might sufficiently balance the divergence of the deformation so that it no longer is numerically divergence-free. This explains why the non-physical pressure oscillations seen in Barry and Mercer's and Mandel's problems are short-lived.

If the constrained specific storage coefficient is nonzero and $p^0 = 0$, then (5.3) becomes

$$c_o p^1 + \alpha \nabla \cdot \mathbf{u}^1 - \Delta t\ k\ \nabla \cdot \nabla p^1 \ = \ 0. \tag{5.4}$$

Therefore, $\nabla \cdot \mathbf{u}^1$ is likely to be balanced by $c_o p^1$ irrespective of the value of $\Delta t\ k\ \nabla \cdot \nabla p^1$. In this way, a null constrained specific storage coefficient value, $c_o \approx 0$, appears to be a necessary condition for numerical problems to occur. Sufficiency is generally associated with relatively unrefined grids, and with both a small time step and small permeability.

5.2.2 Two Potential Remedies for Early Time Locking

1. *Discontinuous Spaces.* For large elements one solution to the problem of locking in linear poroelasticity, like in linear elasticity, is to use a space for displacement which contains non-constant divergence-free vectors[4]. Any number of methods are available which satisfy this requirement. One obvious possibility is to use higher order elements to model the displacement. Alternatively, Murad et al. [1996] discussed the possibility of stable spaces which satisfy an inf-sup condition as a means to eliminate locking. This, for instance, might include Taylor-Hood elements for the displacement and pressure. Later, Wan [2002] elaborated the discussion of stabilized spaces as an effective means of eliminating pressure oscillations and applied her results to Terzaghi's 1-D consolidation problem. Also, Liu [2004] implemented a DG variant based on the work of Phillips and Wheeler [2005b] to eliminate locking in Mandel's problem.

   The first approach that is developed in this dissertation is to use a family of discontinuous spaces to approximate the displacement. Since this method leaves sufficient degrees of freedom on each element to approximate a non-constant divergence-free vector, it appears to be a natural solution to the problem of locking. The next chapter presents the derivation of the DG/Mixed algorithm for linear poroelasticity and proof of optimal theoretical *a priori* error estimates. Interestingly, subsequent chapters demonstrate only moderate effectiveness in eliminating locking on several test problems.

[4] In a finite difference setting, Ewing et al. [2004] indicate that use of staggered grids might effectively handle locking.

2. *Adaptive Grid Refinement.* Another approach can be discovered by re-examining (5.3). As pointed out, if pressure boundary layers are present in the analytical solution $\nabla \cdot \nabla p^1$ might be so large that $\nabla \cdot \mathbf{u}^1$ would be sufficiently balanced that it would not be effectively divergence free.

   The problem is that the boundary layer may be small relative to the element size. In this case, the size of the approximating flux on those elements touching the boundary layer would be highly undervalued. Conversely, uniform refinement would make the elements smaller near the boundary layer, and produce a larger flux in that region. Unfortunately, uniform refinement is sometimes very costly and often unneeded.

   This leads to a second alternative that is to be developed in this dissertation: adaptive grid refinement. A later chapter produces a theorem regarding *a posteriori* error estimates in terms of naturally defined and computable residual quantities. These residual quantities are then used to develop a grid refinement strategy. Results are presented which show the ability of CG/Mixed/Adaptivity to overcome locking. In addition, results will also demonstrate the improvement in accuracy and efficiency that can be gained from adaptive grid refinement techniques.

Chapter 6

Discontinuous Galerkin for Displacements Coupled with a Mixed Formulation for Flow: Theoretical Results

This chapter presents a finite element algorithm for linear poroelasticity which couples a family of discontinuous Galerkin elements for displacements with a mixed formulation for flow. The motivation for this extension of the CG/Mixed algorithm is twofold:

1. *Potential to remove locking.* The use of discontinuous elements may avoid the potential non-physical pressure oscillations which sometime occur with the CG/Mixed scheme when $c_o = 0$.

2. *Element-wise momentum conservation.* The use of discontinuous spaces has been shown to preserve element-wise momentum conservation in linear elasticity [Rivière and Wheeler, 2000], and the result naturally extends to the case of linear poroelasticity.

Existence and uniqueness are proved for the DG/Mixed algorithm. Like the results for the CG/Mixed scheme, the form for the *a priori* error estimates depends on the value of the constrained specific storage coefficient. When c_o is bounded below by a positive constant, the pressure is optimal when measured in the $L^\infty(L^2)$ norm. For $c_o \geq 0$, optimality can be obtained only when the pressure error is measured in the $L^2(L^2)$ norm.

The principal motivation for the DG/Mixed scheme is its potential to remove locking. However, the success of the scheme depends on the flexibility to make certain penalty parameters large. Since these parameters penalize inter-element discontinuity, large parameters may cause the DG/Mixed scheme to become effectively continuous and thus also experience locking. Therefore, care is taken to understand those situations which require large penalty parameters.

6.1 Discontinuous Space for Approximating Displacements

Provided now are some definitions important for the development of the discontinuous Galerkin component of the modified scheme. The notation here follows Rivière and Wheeler [2000]. Let $\mathcal{E}_h = \{E_1, E_2, ..., E_{N_h}\}$ be a non-degenerate subdivision of Ω, where E_j is a triangle or quadrilateral for $d = 2$, or a tetrahedron if $d = 3$. Let $h_j = \text{diam}(E_j)$; then nondegeneracy requires the existence of $\rho > 0$ such that E_j contains a ball of radius ρh_j. In addition, set $h = \max\{h_j : j = 1, ..., N\}$.

Denote the set of interior edges (faces for $d = 3$) of $\mathcal{E}_h$ by Γ_{int}. To each interior edge (or face), $e_a = E_- \cap E_+$, associate a fixed unit normal vector $\boldsymbol{\nu}^a$ pointing from E_- into E_+, and for each boundary edge (or face), $e_a = E_j \cap \partial\Omega$, let $\boldsymbol{\nu}^a$ be the outward pointing normal. Denote this set of fixed normals by $\mathcal{N}$. Alternatively, when a normal specifically refers to the *outward normal relative to a particular element* a subscript for that normal will be used.

For $s > 0$, define the following spaces:

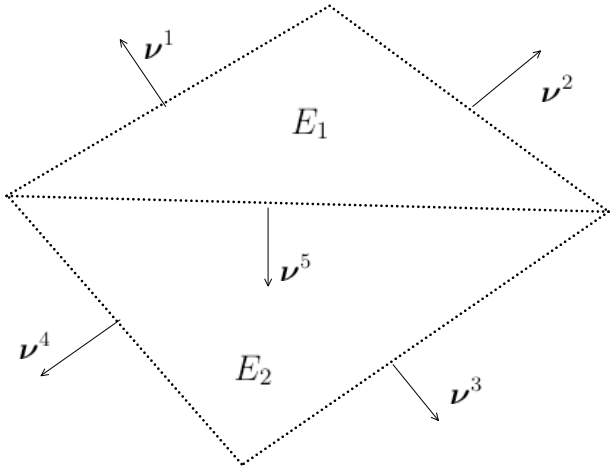

Figure 6.1: The unit normal, $\boldsymbol{\nu}^a$, for an edge intersecting $\partial\Omega$ is always chosen to be outwardly oriented; all other unit normals point from a lower numbered element into a higher numbered element (e.g., $\boldsymbol{\nu}^5$)

$$
\begin{aligned}
H^s(\mathcal{E}_h) &\equiv \{v \in L^2(\Omega) : v|_{E_j} \in H^{s+\epsilon}, \forall j \quad \exists \epsilon > 0\}, \\
\mathbf{H}^s(\mathcal{E}_h) &\equiv (H^s(\mathcal{E}_h))^d.
\end{aligned}
$$

The usual Sobolev norm on $H^m(E)$ for each $E \subset \mathcal{E}_h$ is denoted by $||\cdot||_{m,E}$, whereby the following broken norms for positive integer m are defined:

$$
\begin{aligned}
|||w|||_m &\equiv \left(\sum_{j=1}^{N_h} ||w||^2_{m,E_j} \right)^{\frac{1}{2}}, \quad \forall w \in H^m(\mathcal{E}_h), \\
|||\mathbf{w}|||_m &\equiv \left(\sum_{j=1}^{N_h} ||\mathbf{w}||^2_{m,E_j} \right)^{\frac{1}{2}}, \quad \forall \mathbf{w} \in \mathbf{H}^m(\mathcal{E}_h).
\end{aligned}
$$

Define the average and the jump for $w \in H^s(\mathcal{E}_h),\ s \geq \frac{1}{2}$, as

$$\begin{aligned} \{w\} &= \frac{1}{2}(w|_{E_a^1})|_{e_a} + \frac{1}{2}(w|_{E_a^2})|_{e_a}, \\ \big[w\big] &= (w|_{E_a^1})|_{e_a} - (w|_{E_a^2})|_{e_a}. \end{aligned}$$

The above spaces are then used to define the approximating broken space for displacements,

$$\mathbf{V}_{\mathbf{f}_o}^B = \mathbf{H}^s(\mathcal{E}) \cap \{\mathbf{u}|_{\Gamma_o} = \mathbf{f}_o\}.$$

Define the space of discontinuous piecewise polynomials:

$$\mathcal{D}_r^d(\mathcal{E}_h) \equiv \{\mathbf{v} : \mathbf{v}|_E \in (P_r(E))^d \quad \forall E \in \mathcal{E}_h\},$$

where $P_r(E)$ is the set of polynomials of degree less than or equal to r on E. Then the finite element space to be used to approximate displacements is

$$\mathbf{V}_{h,\mathbf{f}_o}^B = \mathcal{D}_r^d(\mathcal{E}_h) \cap \{\mathbf{u}|_{\Gamma_o} = \mathbf{f}_o\}.$$

6.1.1 A Different Projection Operator

A different projection operator outlined in work of Girault et al. [2003], $R_h : \mathbf{H}^{r+1} \to \mathcal{D}_r^n(\mathcal{E}_h)$, is associated with the above spaces for triangles in two dimensions (tetrahedra in three)[1]. It is constructed with the purpose to satisfy

[1]In addition to satisfying the described properties on triangle (or tetrahedra), Girault has relayed in private communication that the properties hold also on bricks. Currently, work is also being done on extending the above interpolation results to prisms and general quadrilaterals.

$$\sum_{E\in\mathcal{E}}\int_E w\nabla\cdot(\mathbf{u}-R_h\mathbf{u})=0,\quad \forall w\in\mathbb{P}^{r-1}. \tag{6.1}$$

Here, $\mathbb{P}^{r-1}$ is the space of piecewise polynomials of degree $r-1$. This property is shown to be valid for $r=1,\ 2$ or 3. Below is a list of its salient approximation properties for $m=0$ or 1:

$$|||\mathbf{u}-R_h\mathbf{u}|||_m \ \leq\ Ch^{s-m}|||\mathbf{u}|||_s,\qquad m\leq s\leq r+1. \tag{6.2}$$

In addition, the projection operator is assumed to satisfy

$$||\mathbf{u}-R_h\mathbf{u}||_{0,e}\leq Ch_e^{\mu-\frac{1}{2}}||\mathbf{u}||_{s,E}\quad s>\frac{1}{2}, \tag{6.3}$$

$$||\mathbf{u}-R_h\mathbf{u}||_{1,e}\leq Ch_e^{\mu-\frac{3}{2}}||\mathbf{u}||_{s,E}\quad s>\frac{3}{2}, \tag{6.4}$$

with $\mu=\min\{r+1,s\}$ and $e\in\partial E$ for any element E.

6.2 Deriving the Discontinuous Galerkin Momentum Equation

The development of a scheme coupling discontinuous elements for displacement with a mixed formulation for flow (DG/Mixed) differs from the development of the CG/Mixed only in the way the momentum equation (2.8a) is handled.

To derive the discontinuous variational formulation, take the solution $\mathbf{u}$ and p to (2.8a)-(2.8b) (assumed to be sufficiently regular) and write the

momentum equation in terms of the total stress, $\tilde{\boldsymbol{\sigma}}(\mathbf{u}, p)$. Next, multiply the equation by a test function, $\mathbf{v}$, from the broken space, and integrate over a *single element* E_a to find

$$-\int_{E_a} \mathbf{v} \cdot \nabla \cdot \tilde{\boldsymbol{\sigma}}(\mathbf{u}, p)\ dE = \int_{E_a} \mathbf{f} \cdot \mathbf{v}\ dE.$$

Integrating by parts yields

$$\int_{E_a} \tilde{\boldsymbol{\sigma}}(\mathbf{u}, p) : \epsilon(\mathbf{v})\ dE - \int_{\partial E_a} \mathbf{v} \cdot \tilde{\boldsymbol{\sigma}}(\mathbf{u}, p)\boldsymbol{\nu}_a\ ds = \int_{E_a} \mathbf{f} \cdot \mathbf{v}\ dE.$$

Recall that $\boldsymbol{\nu}_a$ (subscript) is the outward normal relative to element E_a. Also note that in the integration by parts, the substitution $\nabla \mathbf{v} \rightarrow \epsilon(\mathbf{v})$ can be made because of the symmetry of $\tilde{\boldsymbol{\sigma}}$.

Summing over all elements produces

$$\sum_{E_a \in \mathcal{E}} \int_{E_a} \tilde{\boldsymbol{\sigma}}(\mathbf{u}, p) : \epsilon(\mathbf{v})\ dE - \sum_{E_a \in \mathcal{E}} \int_{\partial E_a} \mathbf{v} \cdot \tilde{\boldsymbol{\sigma}}(\mathbf{u}, p)\boldsymbol{\nu}_a\ ds = \sum_{E_a \in \mathcal{E}} \int_{E_a} \mathbf{f} \cdot \mathbf{v}\ dE. \tag{6.5}$$

The second term on the left side of above equation can be usefully rewritten using the identity $[ab] = [a]\{b\} + \{a\}[b]$:

$$
\begin{aligned}
&-\sum_{E_a\in\mathcal{E}}\int_{\partial E_a}\mathbf{v}\cdot\tilde{\boldsymbol{\sigma}}(\mathbf{u},p)\boldsymbol{\nu}_a\;ds\\
&\quad=\quad-\sum_{e_a\in E_a\cap\Gamma_t}\int_{e_a}\mathbf{v}\cdot\mathbf{t}_N\;ds\\
&\qquad-\sum_{e_a\in\Gamma_{int}}\int_{e_a}\Big(\mathbf{v}|_{a-}\cdot\tilde{\boldsymbol{\sigma}}|_{a-}\boldsymbol{\nu}_{a-}+\mathbf{v}|_{a+}\cdot\tilde{\boldsymbol{\sigma}}|_{a+}\boldsymbol{\nu}_{a+}\Big)\;ds\\
&\quad=\quad-\sum_{e_a\in E_a\cap\Gamma_t}\int_{e_a}\mathbf{v}\cdot\mathbf{t}_N\;ds\\
&\qquad-\sum_{e_a\in\Gamma_{int}}\int_{e_a}\Big(\mathbf{v}|_{a-}\cdot\tilde{\boldsymbol{\sigma}}|_{a-}\boldsymbol{\nu}^a-\mathbf{v}|_{a+}\cdot\tilde{\boldsymbol{\sigma}}|_{a+}\boldsymbol{\nu}^a\Big)\;ds\\
&\quad=\quad-\sum_{e_a\in E_a\cap\Gamma_t}\int_{e_a}\mathbf{v}\cdot\mathbf{t}_N\;ds\\
&\qquad-\sum_{e_a\in\Gamma_{int}}\int_{e_a}\Big[\mathbf{v}\cdot\tilde{\boldsymbol{\sigma}}\boldsymbol{\nu}^a\Big]\;ds\\
&\quad=\quad-\sum_{e_a\in E_a\cap\Gamma_t}\int_{e_a}\mathbf{v}\cdot\mathbf{t}_N\;ds\\
&\qquad-\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{v}]\cdot\{\tilde{\boldsymbol{\sigma}}(\mathbf{u},p)\boldsymbol{\nu}^a\}\;ds-\sum_{e_a\in\Gamma_{int}}\underbrace{\int_{e_a}\{\mathbf{v}\}\cdot[\tilde{\boldsymbol{\sigma}}(\mathbf{u},p)\boldsymbol{\nu}^a]\;ds}_{=0\ by\ regularity}.
\end{aligned}
$$

Recall that $-\boldsymbol{\nu}_{a+}=\boldsymbol{\nu}_{a-}=\boldsymbol{\nu}^a$ by definition of the fixed set of normals, $\mathcal{N}$. Also note that the test function $\mathbf{v}$ is zero on the Dirichlet boundary.

Substitution the above equation into (6.5) and rearranging, one finds

$$\sum_{E_a\in\mathcal{E}}\int_{E_a}\tilde{\boldsymbol{\sigma}}(\mathbf{u},p):\epsilon(\mathbf{v})\;dE-\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{v}]\cdot\{\tilde{\boldsymbol{\sigma}}(\mathbf{u},p)\boldsymbol{\nu}^a\}\;ds$$
$$=\underbrace{\sum_{E_a\in\mathcal{E}}\int_{E_a}\mathbf{f}\cdot\mathbf{v}\;dE+\sum_{e_a\in E_a\cap\Gamma_t}\int_{e_a}\mathbf{t}_N\cdot\mathbf{v}\;ds}_{\equiv\;L_1(\mathbf{v})}. \tag{6.6}$$

Plugging $\tilde{\boldsymbol{\sigma}}(\mathbf{u},p)=\boldsymbol{\sigma}(\mathbf{u})-\alpha p\boldsymbol{\delta}$ into (6.6) produces

$$\sum_{E_a\in\mathcal{E}}\int_{E_a}\boldsymbol{\sigma}(\mathbf{u}):\epsilon(\mathbf{v})\;dE-\alpha\sum_{E_a\in\mathcal{E}}\int_{E_a}p\nabla\cdot\mathbf{v}\;dE+\alpha\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{v}]\cdot\{p\boldsymbol{\nu}^a\}\;ds$$
$$-\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{v}]\cdot\{\boldsymbol{\sigma}(\mathbf{u})\boldsymbol{\nu}^a\}\;ds$$
$$=L_1(\mathbf{v}). \tag{6.7}$$

To complete the formulation, three terms are added to the left hand side of (6.7) which are all equal to zero due to the assumed regularity of $\mathbf{u}$:

$$\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{u}]\cdot\{\boldsymbol{\sigma}(\mathbf{v})\boldsymbol{\nu}^a\}\;ds, \tag{6.8}$$
$$J_0^{\delta,\beta}(\mathbf{u},\mathbf{v})\equiv\sum_{e_a\in\Gamma_{int}}\frac{\delta_a}{|e_a|^\beta}\int_{e_a}[\mathbf{u}]\cdot[\mathbf{v}], \tag{6.9}$$
$$J_0^{\tilde{\delta},\beta}(\mathbf{u}_t,\mathbf{v})\equiv\sum_{e_a\in\Gamma_{int}}\frac{\tilde{\delta}_a}{|e_a|^\beta}\int_{e_a}[\mathbf{u}_t]\cdot[\mathbf{v}]. \tag{6.10}$$

Although the terms are zero for the true solution, they have an effect upon the discontinuous displacement approximation. The bilinear form in (6.8) characterizes the DG scheme by its value of θ_{DG} as shown in Table 6.1. The

Symmetric Interior Penalty Galerkin scheme (*SIPG*) results when $\theta_{DG} = -1$; the Incomplete Interior Penalty Galerkin scheme (*IIPG*) results when $\theta_{DG} = 0$; and the Non-symmetric Interior Penalty Galerkin scheme (*NIPG*) results when $\theta_{DG} = 1$. This term affects the overall symmetry of the formulation. The flexibility to choose different DG schemes may be important when choosing an existing solver, although this will not be elaborated further here. The penalty terms in (6.9)-(6.10) have the effect of determining the allowable amount of discontinuity across an edge. Parameters δ and $\tilde{\delta}$ are discrete, positive functions that take a constant value on the edge or face e_a (with $|e_a|$ its Lebesgue measure). As δ and $\tilde{\delta}$ become very large the DG scheme becomes asymptotically equivalent to a CG scheme[2].

The resulting DG variational formulation for the momentum equation is then found to be

$$\sum_{E_a \in \mathcal{E}} \int_{E_a} \boldsymbol{\sigma}(\mathbf{u}) : \epsilon(\mathbf{v}) \; dE - \alpha \sum_{E_a \in \mathcal{E}} \int_{E_a} p \nabla \cdot \mathbf{v} \; dE + \alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} [\mathbf{v}] \cdot \{p \boldsymbol{\nu}^a\} \; ds$$
$$+\theta_{DG} \sum_{e_a \in \Gamma_{int}} \int_{e_a} [\mathbf{u}] \cdot \{\boldsymbol{\sigma}(\mathbf{v}) \boldsymbol{\nu}^a\} \; ds - \sum_{e_a \in \Gamma_{int}} \int_{e_a} [\mathbf{v}] \cdot \{\boldsymbol{\sigma}(\mathbf{u}) \boldsymbol{\nu}^a\} \; ds$$
$$+J_0^{\delta,\beta}(\mathbf{u}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(\mathbf{u}_t, \mathbf{v}) = L_1(\mathbf{v}). \quad (6.11)$$

6.2.1 The DG Bilinear Form from Linear Elasticity

The notation in (6.11) can be simplified by using the DG bilinear form, $A_{\mathbf{u}}(,)$, for linear elasticity introduced in Rivière and Wheeler [2000]:

[2]However, large penalty parameters also lead to a potentially ill-conditioned matrix.

Table 6.1: Family of DG Schemes

θ_{DG}	=	1	*NIPG*
θ_{DG}	=	0	*IIPG*
θ_{DG}	=	-1	*SIPG*

$$\begin{aligned} A_{\mathbf{u}}(\mathbf{u},\mathbf{v}) \;\equiv\; & \sum_{E_a\in\mathcal{E}}\int_{E_a}\boldsymbol{\sigma}(\mathbf{u}):\epsilon(\mathbf{v})\;dE + J_0^{\delta,\beta}(\mathbf{u},\mathbf{v}) \\ & +\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{u}]\cdot\{\boldsymbol{\sigma}(\mathbf{v})\boldsymbol{\nu}^a\}\;ds \\ & -\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\mathbf{v}]\cdot\{\boldsymbol{\sigma}(\mathbf{u})\boldsymbol{\nu}^a\}\;ds. \end{aligned} \tag{6.12}$$

Continuity and coercivity are not available for $A_{\mathbf{u}}(,)$ and are not needed. However, certain component terms do satisfy the aforementioned properties:

$$\sum_{E_a\in\mathcal{E}}\int_{E_a}\boldsymbol{\sigma}(\mathbf{u}):\epsilon(\mathbf{v})\;dE \;\le\; C_{cont}|||\mathbf{u}|||_1|||\mathbf{v}|||_1, \quad \forall\;\mathbf{u},\;\mathbf{v}\in H^1(\mathcal{E}_h). \tag{6.13}$$

Coercivity is slightly more complicated. Firstly, since $A_{\mathbf{u}}$ is defined element-wise, the standard *Korn's* inequality cannot be applied. Secondly, $A_{\mathbf{u}}$ contains potentially negative terms (its last two terms). Nevertheless, the following *Korn's* inequality for piecewise $H^1(\mathcal{E}_h)$ vector fields [Brenner, 2002] can be usefully employed,

$$\gamma|||\mathbf{u}|||_1^2 \;\le\; \Big[\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\mathbf{u}):\epsilon(\mathbf{u})) + J_0^{\delta,\beta}(\mathbf{u},\mathbf{u})\Big] \quad \forall\mathbf{u}\in H^1, \tag{6.14}$$

for some $\gamma > 0$ and for the case[3] that $|\Gamma_0| > 0$.

6.2.2 The Fully Coupled DG/Mixed Scheme

With the above notation for $A_{\mathbf{u}}(,)$, (6.11) combines with the established mixed equations for flow to produce the full DG/Mixed variational formulation: Find $\mathbf{u} \in \mathbf{V}^B_{\mathbf{u}_D}$ and $(p, \mathbf{z}) \in (L^2 \times H(div) \cap \{\mathbf{s} : \mathbf{s} \cdot \nu|_{\Gamma_f} = q\})$ such that

$$\begin{aligned} A_{\mathbf{u}}(\mathbf{u}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(\mathbf{u}_t, \mathbf{v}) - \alpha(\nabla \cdot \mathbf{v}, p) & \\ +\alpha \sum_{e \in \Gamma_{int}} \int_{e_a} \{p\} \boldsymbol{\nu}^a \cdot [\mathbf{v}] &= L_1(\mathbf{v}), \qquad (6.15a) \\ (\frac{\partial}{\partial t}(c_o p + \alpha \nabla \cdot \mathbf{u}), w) + (\nabla \cdot \mathbf{z}, w) &= L_2(w), \qquad (6.15b) \\ (\tilde{\boldsymbol{\kappa}}^{-1}\mathbf{z}, \mathbf{s}) - (p, \nabla \cdot \mathbf{s}) &= L_3(\mathbf{s}), \qquad (6.15c) \end{aligned}$$

for all $\mathbf{v} \in \mathbf{V}^B_0$ and $(w, \mathbf{s}) \in (L^2 \times H(div) \cap \{\mathbf{s} : \mathbf{s} \cdot \nu|_{\Gamma_f} = 0\})$. Here, L_1 to L_3 are the bounded linear functional defined as

$$\begin{aligned} L_1(\mathbf{v}) &= \sum_{E \in \mathcal{E}_h} \int_E \mathbf{f} \cdot \mathbf{v} + \sum_{e \in \Gamma_t} \int_e \mathbf{t}_N \cdot \mathbf{v}, \\ L_2(w) &= \int_\Omega s_f w, \\ L_3(\mathbf{s}) &= -\int_{\Gamma_p} p_o \mathbf{s} \cdot \boldsymbol{\nu} + \int_\Omega \rho_f \mathbf{g} \cdot \mathbf{s}. \end{aligned}$$

The finite element algorithm is found by replacing the above spaces with

[3] An assumption that is held throughout this dissertation.

the appropriate finite dimensional spaces in (6.15a)-(6.15c). The DG/Mixed scheme becomes: Find $\bar{\mathbf{u}} \in \mathbf{V}^B_{h,\mathbf{u}_D}$ and $(\bar{p}, \bar{\mathbf{z}}) \in (W_h \times \mathbf{S}_{h,q})$ such that

$$\begin{aligned} A_{\mathbf{u}}(\bar{\mathbf{u}}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t, \mathbf{v}) - \alpha(\nabla \cdot \mathbf{v}, \bar{p}) & \\ +\alpha \sum_{e \in \Gamma_{int}} \int_{e_a} \{\bar{p}\} \boldsymbol{\nu}^a \cdot [\mathbf{v}] &= L_1(\mathbf{v}), \qquad (6.16a) \\ (\frac{\partial}{\partial t}(c_o \bar{p} + \alpha \nabla \cdot \bar{\mathbf{u}}), w) + (\nabla \cdot \bar{\mathbf{z}}, w) &= L_2(w), \qquad (6.16b) \\ (\tilde{\boldsymbol{\kappa}}^{-1} \bar{\mathbf{z}}, \mathbf{s}) - (\bar{p}, \nabla \cdot \mathbf{s}) &= L_3(\mathbf{s}), \qquad (6.16c) \end{aligned}$$

for all $\mathbf{v} \in \mathbf{V}^B_{h,0}$ and $(w, \mathbf{s}) \in (W_h \times \mathbf{S}_{h,0})$. Recall from the earlier development of the CG/Mixed algorithm that W_h and $\mathbf{S}_h$ form a standard finite dimensional mixed space pair.

Because of the way that the DG/Mixed scheme is derived, consistency is automatic, and, therefore, Galerkin orthogonality may be employed whenever necessary. The outstanding issues regarding (6.16a)-(6.16c) left to be addressed concern the existence, uniqueness and optimality of its solution.

Several useful trace inequalities will assist in the proofs of existence and optimality. Two which are valid for finite dimensional spaces are [Schwab, 1998]

$$||\boldsymbol{\sigma}(\mathbf{v})\boldsymbol{\nu}||^2_{L^2(e)} \leq C h_E^{-1} \int_E (\boldsymbol{\sigma}(\mathbf{v}) : \epsilon(\mathbf{v}))\, dx, \quad \forall \mathbf{v} \in \mathcal{D}^s_r(E_h), \qquad (6.17)$$

$$||w||_{L^2(e)} \leq C h_E^{-\frac{1}{2}} ||w||_{L^2(E)}, \quad \forall w \in W_h(E), \qquad (6.18)$$

where e is any edge of E, and h_E is the diameter of E.

For Sobolev spaces the following trace inequality is useful [Brenner and Scott, 1994]

$$||\phi||_{L^p(e)} \leq C||\phi||^{1-\frac{1}{p}}_{L^p(E)} \, ||\phi||^{\frac{1}{p}}_{W^{1,p}(E)} \quad \forall e \in \partial E. \tag{6.19}$$

For $p = 2$ the fact that Hilbert spaces are nested implies

$$||\phi||_{L^2(e)} \leq C||\phi||_{H^1(E)}. \tag{6.20}$$

On the other hand, by using the trivial inequality $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$ and a scaling argument, (6.19) becomes

$$||\phi||_{L^2(e)} \leq C\left(h_E^{-\frac{1}{2}}||\phi||_{L^2(E)} + h_E^{\frac{1}{2}}||\phi||_{H^1(E)}\right). \tag{6.21}$$

6.3 Theoretical Properties of the DG/Mixed Scheme

The theoretical properties of the DG/Mixed scheme in many ways mirror those of the CG/Mixed scheme. In particular, the proof of existence and uniqueness once again requires the utilization of parabolic lift theorems. Moreover, optimality for the pressure solution again depends on the assumptions one makes in regard to the admissible range of values of the constrained specific storage coefficient. One important difference for the DG/Mixed scheme is its dependence on the penalty parameters, δ and $\tilde{\delta}$, in order to ensure existence, uniqueness and optimality. Since large values for these penalty parameters

may impede the ability of the scheme to overcome locking, special care will be taken to understand those factors which influence the size of the penalty parameters.

6.3.1 Existence and Uniqueness

For the general case of a nonnegative constrained specific storage coefficient, $c_o \geq 0$, the existence and uniqueness of the DG/Mixed solution can be established. First, let all problem data be zero. Then there is at least one solution to the finite dimensional problem which is the null solution. Let $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ be some solution to (6.16a)-(6.16c). Then take $\mathbf{v} = \frac{\partial}{\partial t}\bar{\mathbf{u}}$, $w = \bar{p}$ and $\mathbf{s} = \bar{\mathbf{z}}$. By adding (6.16a) and (6.16b) together,

$$\begin{aligned} A_{\mathbf{u}}(\bar{\mathbf{u}}, \bar{\mathbf{u}}_t) + J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t, \bar{\mathbf{u}}_t) + (c_o \bar{p}_t, \bar{p}) \\ +\alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{\bar{p}\} \boldsymbol{\nu}^a \cdot [\bar{\mathbf{u}}_t] + (\nabla \cdot \bar{\mathbf{z}}, \bar{p}) = 0. \end{aligned}$$

From (6.16c) it is seen that $(\nabla \cdot \bar{\mathbf{z}}, \bar{p}) = (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} \bar{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} \bar{\mathbf{z}})$; substituting this into the above equation, and then using the chain rule from calculus,

$$\begin{aligned} A_{\mathbf{u}}(\bar{\mathbf{u}}, \bar{\mathbf{u}}_t) + J_0^{\tilde{\delta},\beta}(\mathbf{u}_t, \mathbf{u}_t) + \frac{1}{2}\frac{\partial}{\partial t} ||c_o^{1/2} \bar{p}||_0^2 + ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} \bar{\mathbf{z}}||_0^2 \\ +\alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{\bar{p}\} \boldsymbol{\nu}^a \cdot [\mathbf{u}_t] = 0. \end{aligned}$$

Notice that (except for the case $\theta_{DG} = -1$) $A_{\mathbf{u}}$ is not symmetric, and thus it is not true that $\frac{1}{2}\frac{\partial}{\partial t} A_{\mathbf{u}}(\bar{\mathbf{u}}, \bar{\mathbf{u}}) = A_{\mathbf{u}}(\bar{\mathbf{u}}, \bar{\mathbf{u}}_t)$. However, part of the constituent terms of $A_{\mathbf{u}}$ are always symmetric; if $A_{\mathbf{u}}$ decomposed accordingly the above equation becomes

$$\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))+\frac{1}{2}\frac{\partial}{\partial t}J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})+J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t)\\
&\quad+\frac{1}{2}\frac{\partial}{\partial t}||c_o^{1/2}\bar{p}||_0^2+||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||_0^2+\alpha\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\bar{p}\}\boldsymbol{\nu}^a\cdot[\bar{\mathbf{u}}_t]\\
-\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}\cdot[\bar{\mathbf{u}}_t]+\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(\bar{\mathbf{u}}_t)\boldsymbol{\nu}^a\}\cdot[\bar{\mathbf{u}}]&=0.
\end{aligned}$$

Integrating above equation from $t=0$ to $t=T$, and rearranging,

$$\begin{aligned}
&\frac{1}{2}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))_{t=T}+\frac{1}{2}||c_o^{1/2}\bar{p}(T)||_0^2+\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||_0^2\\
&\qquad+\frac{1}{2}J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T}+\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t)\\
=\int_0^T\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\nu^a\}\cdot[\bar{\mathbf{u}}_t]-\int_0^T\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(\bar{\mathbf{u}}_t)\nu^a\}\cdot[\bar{\mathbf{u}}]\\
-\alpha\int_0^T\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\bar{p}\}\boldsymbol{\nu}^a\cdot[\bar{\mathbf{u}}_t].
\end{aligned}$$

Note that the assumption $\mathbf{u}^o=p^o=0$ has been used. Next, integrate by parts (in time) the middle term on the rhs of the above equation, and then take absolute values to arrive at

$$\begin{aligned}
&\frac{1}{2}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))_{t=T}+\frac{1}{2}||c_o^{1/2}\bar{p}(T)||_0^2+\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||_0^2\\
&\qquad+\frac{1}{2}J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T}+\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t)\\
&\qquad\leq M_1+M_2+M_3, \qquad (6.22)
\end{aligned}$$

where

$$
\begin{aligned}
M_1 &= (1+\theta_{DG}) \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \left| \{\boldsymbol{\sigma}(\bar{\mathbf{u}})\nu^a\} \cdot [(\bar{\mathbf{u}}_t)] \right|, \\
M_2 &= \theta_{DG} \sum_{e_a \in \Gamma_{int}} \int_{e_a} \left| (\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\nu^a\} \cdot [\bar{\mathbf{u}}])_{t=T} \right|, \\
M_3 &= \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \left| \{\bar{p}\}\boldsymbol{\nu}^a \cdot [\bar{\mathbf{u}}_t] \right|.
\end{aligned}
$$

In order to bound the above quantities, make the following definition: for each interior edge e_a, assume that $e_a = \partial E_a^1 \cap \partial E_a^2$, where E_a^1 and E_a^2 are elements of $\mathcal{E}_h$, and set $E_a^{12} = E_a^1 \cup E_a^2$. The inequalities are derived primarily using Cauchy-Schwarz, the trace inequality (6.17), and Young's inequality with $\epsilon > 0$ (that is, for $a, b \geq 0, ab \leq \frac{1}{2\epsilon}a^2 + \frac{\epsilon}{2}b^2$). Also define $\delta_{min} = \min_{e \in \Gamma_{int}} \{\delta_e\}$ and $\tilde{\delta}_{min} = \min_{e \in \Gamma_{int}} \{\tilde{\delta}_e\}$.

$$
\begin{aligned}
M_1 &\leq C\int_0^T \sum_{e_a\in\Gamma_{int}}\int_{e_a} |\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\nu^a\}||[\bar{\mathbf{u}}_t]| \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\nu^a\}||_{e_a,0}||[\bar{\mathbf{u}}_t]||_{e_a,0} \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right)^{\frac{1}{2}} ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}||_{e_a,0} \left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right)^{\frac{1}{2}} ||[\bar{\mathbf{u}}_t]||_{e_a,0} \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right) ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}||^2_{e_a,0} + \frac{1}{4}\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right) ||[\bar{\mathbf{u}}_t]||^2_{e_a,0} \\
&\leq \frac{C}{\tilde{\delta}_{min}} h^{\beta(d-1)}\int_0^T \sum_{e_a\in\Gamma_{int}} ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}||^2_{e_a,0} + \frac{1}{4}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t) \\
&\leq \frac{C}{\tilde{\delta}_{min}} h^{\beta(d-1)-1}\int_0^T \sum_{e_a\in\Gamma_{int}} \int_{E_a^{12}} (\boldsymbol{\sigma}(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{u}})) + \frac{1}{4}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t) \\
&\leq \frac{C}{\tilde{\delta}_{min}} \int_0^T \sum_{E\in\mathcal{E}} \int_E (\boldsymbol{\sigma}(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{u}})) + \frac{1}{4}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t).
\end{aligned}
$$

$$
\begin{aligned}
M_2 &\leq C\sum_{e_a\in\Gamma_{int}}\int_{e_a} |\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}_{t=T}||[\bar{\mathbf{u}}]_{t=T}| \\
&\leq C\sum_{e_a\in\Gamma_{int}} ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}_{t=T}||_{e_a,0}||[\bar{\mathbf{u}}]_{t=T}||_{e_a,0} \\
&\leq C\sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\delta_a}\right)^{\frac{1}{2}} ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}_{t=T}||_{e_a,0} \left(\frac{\delta_a}{|e_a|^\beta}\right)^{\frac{1}{2}} ||[\bar{\mathbf{u}}]_{t=T}||_{e_a,0} \\
&\leq C\sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\delta_a}\right) ||\{\boldsymbol{\sigma}(\bar{\mathbf{u}})\boldsymbol{\nu}^a\}_{t=T}||^2_{e_a,0} + \frac{1}{4}\sum_{e_a\in\Gamma_{int}} \left(\frac{\delta_a}{|e_a|^\beta}\right) ||[\bar{\mathbf{u}}]_{t=T}||^2_{e_a,0} \\
&\leq \frac{C}{\delta_{min}} h^{\beta(d-1)-1}\sum_{e_a\in\Gamma_{int}} \int_{E_a^{12}} (\boldsymbol{\sigma}(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{u}}))_{t=T} + \frac{1}{4} J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T} \\
&\leq \frac{C}{\delta_{min}} \sum_{E\in\mathcal{E}} \int_E (\boldsymbol{\sigma}(\bar{\mathbf{u}}) : \epsilon(\bar{\mathbf{u}}))_{t=T} + \frac{1}{4} J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T}.
\end{aligned}
$$

In order to bound M_3, again utilize a trace inequality.

$$\begin{aligned}
M_3 &\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} |\{\bar{p}\}||[\bar{\mathbf{u}}_t]| \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} ||\{\bar{p}\}||_{e_a,0}||[\bar{\mathbf{u}}_t]||_{e_a,0} \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right)^{\frac{1}{2}} ||\{\bar{p}\}||_{e_a,0} \left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right)^{\frac{1}{2}} ||[\bar{\mathbf{u}}_t]||_{e_a,0} \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right) ||\{\bar{p}\}||^2_{e_a,0} + \frac{1}{4}\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right) ||[\bar{\mathbf{u}}_t]||^2_{e_a,0} \\
&\leq \frac{C}{\tilde{\delta}_{min}} h^{\beta(d-1)-1}\int_0^T \sum_{e_a\in\Gamma_{int}} ||\bar{p}||^2_{E_a^{12},0} + \frac{1}{4}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t) \\
&\leq \frac{C}{\tilde{\delta}_{min}}\int_0^T ||\bar{p}||_0^2 + \frac{1}{4}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t).
\end{aligned}$$

Rearranging the above results in the following inequality

$$\begin{aligned}
\left(\frac{1}{2}-\frac{C_1}{\delta_{min}}\right)\sum_{E\in\mathcal{E}_h}\int_E (\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))_{t=T} + \frac{1}{2}||c_o^{1/2}\bar{p}||^2_{t=T} + \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||^2 \\
+\frac{1}{4}J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T} + \frac{1}{2}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t) \\
\leq C\int_0^T \sum_{E\in\mathcal{E}_h}\int_E (\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}})) + \frac{C}{\tilde{\delta}_{min}}\int_0^T ||\bar{p}||^2.
\end{aligned}$$

Next, incorporate inequality (3.19) to find

$$\left(\frac{1}{2}-\frac{C_1}{\delta_{min}}\right)\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))_{t=T}+\frac{1}{2}||c_o^{1/2}\bar{p}||^2_{t=T}+\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||^2$$
$$+\frac{1}{4}J_0^{\delta,\beta}(\bar{\mathbf{u}},\bar{\mathbf{u}})_{t=T}+\frac{1}{2}\int_0^T J_0^{\tilde{\delta},\beta}(\bar{\mathbf{u}}_t,\bar{\mathbf{u}}_t)$$
$$\leq C\int_0^T\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))+\frac{C_2}{\tilde{\delta}_{min}}\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||^2.$$

Rearranging the above equation and dropping the inessential positive terms from the lhs, one arrives at

$$\left(\frac{1}{2}-\frac{C_1}{\delta_{min}}\right)\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}}))_{t=T}+\left(1-\frac{C_2}{\tilde{\delta}_{min}}\right)\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}\bar{\mathbf{z}}||^2$$
$$\leq C\int_0^T\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(\bar{\mathbf{u}}):\epsilon(\bar{\mathbf{u}})).$$

Thus, if δ_{min} and $\tilde{\delta}_{min}$ are chosen so that $\frac{C_1}{\delta_{min}}<\frac{1}{2}$ and $\frac{C_2}{\tilde{\delta}_{min}}<1$ then each of the above terms on the lhs is positive. Using the positive definiteness of $\boldsymbol{\sigma}$ and $\tilde{\boldsymbol{\kappa}}^{-1}$, Gronwall's inequality is employed to conclude that $\bar{\mathbf{u}}=\bar{\mathbf{z}}=0$. Moreover, using the parabolic lift theorem (3.19), it is deduced that $\bar{p}=0$. Hence uniqueness is established, and as the system of equations is square, existence follows provided δ_{min} and $\tilde{\delta}_{min}$ are sufficiently large.

6.3.2 Continuous in Time Error Estimates

A priori continuous in time error estimates for the DG/Mixed scheme are now examined[4]. The following auxiliary error estimate measures the aux-

[4]For simplicity, the initial time solutions for the pressure and displacement are assumed to be functions in their respective finite element spaces. In eliminating this assumption, the proof which follows can be easily modified in the usual way.

iliary errors of $\bar{\mathbf{u}}$ and $\bar{\mathbf{z}}$. The subsequent corollary uses inequality (3.22) from the parabolic lift theorem to incorporate the pressure, $\bar{p}$, in the full finite element error estimates when $c_o = 0$ on parts of the domain. To proceed, it is necessary to assume that

$$\mathbf{u} \in L^{\infty}(0,T;H^{\frac{3}{2}}), \quad p \in L^2(0,T;H^1). \tag{6.23}$$

The factor $\frac{3}{2}$ in the displacement regularity requirement is to ensure that the first order derivatives have well-defined traces along boundary edges.

As in the case of error estimates for the CG/Mixed scheme, the rate of convergence is related to the order of approximating spaces, but is restricted by the maximal regularity of the true solution and its time derivatives. To this end let $q,\ r,\ s$ and t be the largest positive real numbers such that

$$\begin{aligned} &p \in L^2(0,T;H^q), \quad p_t \in L^2(0,T;H^r), \\ &\mathbf{u} \in L^2(0,T;H^s), \quad \mathbf{u}_t \in L^2(0,T;H^t). \end{aligned} \tag{6.24}$$

The same remarks following the CG/Mixed schemes hold here, and thus the additional assumptions are made: $\mathbf{z} \in H(div)\ \forall t \in [0,T]$ and $q > \frac{3}{2}$.

Since the R_h operator is used here to facilitate the proof, the following notational modification is made: $E^A_{\mathbf{u}} \equiv R_h\mathbf{u} - \bar{\mathbf{u}}$ and $E^I_{\mathbf{u}} = \mathbf{u} \equiv R_h\mathbf{u}$. The corresponding notation for the flow variables remains the same as earlier defined in Chapter 3. With this, the results are summarized in the following theorem and corollaries.

Theorem 6.3.1 (General Auxiliary Error Estimate)**.** *Let $r_1 \geq 0$ be the order of the mixed space $(W_h, \mathbf{S}_h)$, and let $r_2 \in \{1, 2, 3\}$ be the degree of the polynomials used in the displacement space $\mathbf{V}_h$. Assume $r_1 < r_2$, the regularity (6.23)-(6.24) of the exact solution, parameter constraints (2.1)-(2.3) and no initial time error, Then, if $\beta = (d-1)^{-1}$, and the parameters δ and $\tilde{\delta}$ are large enough,*

$$|||E^A_{\mathbf{u}}|||_{L^\infty(H^1)} + ||c_o^{1/2} E^A_p||_{L^\infty(L^2)} + ||E^A_{\mathbf{z}}||_{L^2(L^2)} \;\leq\; Ch^R. \tag{6.25}$$

Here, $C = C(T, \boldsymbol{\kappa}, C_{coer}, C_{cont}, c_o, p, p_{,t}, \mathbf{z}, \mathbf{u}, \mathbf{u}_{,t})$, and

$$R = \min\{r_1 + 1,\ r_2,\ q - 1,\ r,\ s - 1,\ t - 1\},$$

which reflects both the order of the approximation spaces and the regularity of the true solution.

Proof. Denote by $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ the approximate solution. Then by Galerkin orthogonality, the following equations hold for all $(\mathbf{v}, w, \mathbf{s}) \in (\mathbf{V}_{h,0}, W_h, \mathbf{S}_{h,0})$,

$$\begin{aligned} A_{\mathbf{u}}(\mathbf{u} - \bar{\mathbf{u}}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(\mathbf{u}_t - \bar{\mathbf{u}}_t, \mathbf{v}) - \alpha(p - \bar{p}, \nabla \cdot \mathbf{v}) & \\ + \alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{p - \bar{p}\} \boldsymbol{\nu} \cdot [\mathbf{v}] &= 0, \quad \text{(6.26a)} \\ (c_o(p - \bar{p})_t, w) + \alpha(\nabla \cdot (\mathbf{u} - \bar{\mathbf{u}})_t, w) + (\nabla \cdot (\mathbf{z} - \bar{\mathbf{z}}), w) &= 0, \quad \text{(6.26b)} \\ (\tilde{\boldsymbol{\kappa}}^{-1}(\mathbf{z} - \bar{\mathbf{z}}), \mathbf{s}) - (p - \bar{p}, \nabla \cdot \mathbf{s}) &= 0. \quad \text{(6.26c)} \end{aligned}$$

Now by noting that $\mathbf{u} - \bar{\mathbf{u}} = E^A_{\mathbf{u}} + E^I_{\mathbf{u}}$, $p - \bar{p} = E^A_p + E^I_p$, $\mathbf{z} - \bar{\mathbf{z}} = E^A_{\mathbf{z}} + E^I_{\mathbf{z}}$ and using the properties of the mixed space and displacement projections, one finds that (6.26b)-(6.26c) becomes

$$\begin{aligned}
(c_o E^I_{p\,t}, w) + (c_o E^A_{p\,,t}, w) + \alpha \overbrace{(\nabla \cdot E^I_{\mathbf{u},t}, w)}^{=0 \ by \ (6.1)} & \\
+\alpha(\nabla \cdot E^A_{\mathbf{u}\,,t}, w) + \overbrace{(\nabla \cdot E^I_{\mathbf{z}}, w)}^{=0 \ by \ (3.6b)} + (\nabla \cdot E^A_{\mathbf{z}}, w) &= 0, \quad (6.27a) \\
(\tilde{\boldsymbol{\kappa}}^{-1} E^A_{\mathbf{z}}, \mathbf{s}) + (\tilde{\boldsymbol{\kappa}}^{-1} E^I_{\mathbf{z}}, \mathbf{s}) - \overbrace{(E^I_p, \nabla \cdot \mathbf{s})}^{=0 \ by \ (3.6a)} - (E^A_p, \nabla \cdot \mathbf{s}) &= 0. \quad (6.27b)
\end{aligned}$$

Notice that one cannot use the properties of the L^2-projection to eliminate the first term in equation (6.27a) because $c_o = c_o(x)$ is a function of the space variable and not necessarily a member of W_h.

Then letting $w = E^A_p$ and $\mathbf{s} = E^A_{\mathbf{z}}$ in (6.27a)-(6.27b), after summing the two equations and canceling like terms,

$$\begin{aligned}
(c_o E^A_{p\,,t}, E^A_p) + \alpha(\nabla \cdot E^A_{\mathbf{u}\,,t}, E^A_p) + (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}) & \\
= -(c_o E^I_{p\,,t}, E^A_p) - (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^I_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}). & \quad (6.28)
\end{aligned}$$

Similarly decomposing eq. (6.26a),

$$\begin{aligned}
A_{\mathbf{u}}(E^I_{\mathbf{u}}, \mathbf{v}) + A_{\mathbf{u}}(E^A_{\mathbf{u}}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, \mathbf{v}) + J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u}\,,t}, \mathbf{v}) - \alpha(E^I_p, \nabla \cdot \mathbf{v}) & \\
-\alpha(E^A_p, \nabla \cdot \mathbf{v}) + \alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^A_p + E^I_p\} \boldsymbol{\nu} \cdot [\mathbf{v}] = 0. &
\end{aligned}$$

Now, letting $\mathbf{v} = E^A_{\mathbf{u},t}$ in the above equation and rearranging,

$$\begin{aligned} & A_{\mathbf{u}}(E^A_{\mathbf{u}}, E^A_{\mathbf{u},t}) + J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}) \\ = \; & -A_{\mathbf{u}}(E^I_{\mathbf{u}}, E^A_{\mathbf{u},t}) - J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u},t}) + \alpha(E^A_p, \nabla \cdot E^A_{\mathbf{u},t}) \\ & +\alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^A_p + E^I_p\} \boldsymbol{\nu}^a \cdot [E^A_{\mathbf{u},t}] + \alpha(E^I_p, \nabla \cdot E^A_{\mathbf{u},t}). \end{aligned} \tag{6.29}$$

By summing (6.28) and (6.29), and observing that the $\alpha(E^A_p, \nabla \cdot E^A_{\mathbf{u},t})$-terms cancel, one discovers

$$\begin{aligned} & A_{\mathbf{u}}(E^A_{\mathbf{u}}, E^A_{\mathbf{u},t}) + J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}) + (c_o E^A_{p,t}, E^A_p) + (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}) \\ = \; & -A_{\mathbf{u}}(E^I_{\mathbf{u}}, E^A_{\mathbf{u},t}) - J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u},t}) + \alpha \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^A_p + E^I_p\} \boldsymbol{\nu}^a \cdot [E^A_{\mathbf{u},t}] \\ & -(c_o E^I_{p,t}, E^A_p) + \alpha(E^I_p, \nabla \cdot E^A_{\mathbf{u},t}) - (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^I_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}). \end{aligned}$$

Next integrate in time from 0 to T and use the chain rule to find

$$\begin{aligned} \int_0^T A_{\mathbf{u}}(E^A_{\mathbf{u}}, E^A_{\mathbf{u},t}) + \int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}) + \frac{1}{2}||c_o^{1/2} E^A_p(T)||_0^2 + \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}||_0^2 \\ = -\int_0^T A_{\mathbf{u}}(E^I_{\mathbf{u}}, E^A_{\mathbf{u},t}) - \int_0^T J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u},t}) \\ +\alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^A_p + E^I_p\} \boldsymbol{\nu}^a \cdot [E^A_{\mathbf{u},t}] - \int_0^T (c_o E^I_{p,t}, E^A_p) \\ +\alpha \int_0^T (E^I_p, \nabla \cdot E^A_{\mathbf{u},t}) - \int_0^T (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^I_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}). \end{aligned}$$

Notice the use of the assumption $E^A_p(0) = 0$. Likewise using the assumption that $E^A_{\mathbf{u}}(0) = 0$, if $A_{\mathbf{u}}$ is now decomposed into its symmetric and non-symmetric components (as was done in the proof of existence), one finds

$$\frac{1}{2}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(E^A_{\mathbf{u}})\epsilon(E^A_{\mathbf{u}}))_{t=T}+\frac{1}{2}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T}+\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u}\,,t},E^A_{\mathbf{u}\,,t})$$

$$+\frac{1}{2}||c_o^{1/2}E_p^A(T)||_0^2+\int_0^T||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}E^A_{\mathbf{z}}||_0^2$$

$$=-\int_0^T A_{\mathbf{u}}(E^I_{\mathbf{u}},E^A_{\mathbf{u}\,,t})-\int_0^T J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t},E^A_{\mathbf{u}\,,t})$$

$$+\alpha\int_0^T\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{E_p^A+E_p^I\}\boldsymbol{\nu}^a\cdot[E^A_{\mathbf{u}\,,t}]-\int_0^T(c_oE^I_{p\,,t},E_p^A)$$

$$+\alpha\int_0^T(E_p^I,\nabla\cdot E^A_{\mathbf{u}\,,t})-\int_0^T(\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}E^I_{\mathbf{z}},\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}}E^A_{\mathbf{z}})$$

$$+\int_0^T\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}\cdot[E^A_{\mathbf{u}\,,t}]-\theta_{DG}\int_0^T\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(E^A_{\mathbf{u}\,,t})\nu^a\}\cdot[E^A_{\mathbf{u}}].$$

Integrate by parts in time the first and last terms on the right side of the above equation. Then rearrangement of terms and the boundedness of $\boldsymbol{\kappa}$ (2.1) results in the following inequality which will serve as the basis for the error estimates:

$$\frac{1}{2}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(E^A_{\mathbf{u}}):\epsilon(E^A_{\mathbf{u}}))_{t=T}+\frac{1}{2}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T}$$

$$+\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u}\,,t},E^A_{\mathbf{u}\,,t})+\frac{1}{2}||c_o^{1/2}E_p^A(T)||_0^2+\frac{1}{\lambda_{max}}\int_0^T||E^A_{\mathbf{z}}||_0^2$$

$$\leq\Phi_1+\Phi_2+\Phi_3+\Phi_4+\Phi_5$$

$$+\Phi_6+\Phi_7+\Phi_8+\Phi_9+\Phi_{10},\qquad(6.30)$$

where

$$\Phi_1 = \int_0^T A_{\mathbf{u}}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u}}), \tag{6.31a}$$

$$\Phi_2 = -A_{\mathbf{u}}(E^I_{\mathbf{u}}, E^A_{\mathbf{u}})_{t=T}, \tag{6.31b}$$

$$\Phi_3 = -\int_0^T J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u},t}), \tag{6.31c}$$

$$\Phi_4 = \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^A_p\} \boldsymbol{\nu}^a \cdot [E^A_{\mathbf{u},t}], \tag{6.31d}$$

$$\Phi_5 = \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E^I_p\} \boldsymbol{\nu}^a \cdot [E^A_{\mathbf{u},t}], \tag{6.31e}$$

$$\Phi_6 = -\int_0^T (c_o E^I_{p,t}, E^A_p), \tag{6.31f}$$

$$\Phi_7 = \alpha \int_0^T (E^I_p, \nabla \cdot E^A_{\mathbf{u},t}), \tag{6.31g}$$

$$\Phi_8 = -\int_0^T (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^I_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E^A_{\mathbf{z}}), \tag{6.31h}$$

$$\Phi_9 = (1+\theta_{DG}) \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\} \cdot [E^A_{\mathbf{u},t}], \tag{6.31i}$$

$$\Phi_{10} = -\theta_{DG} \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\} \cdot [E^A_{\mathbf{u}}]_{t=T}. \tag{6.31j}$$

Now, Φ_1 and Φ_2 are bounded by using the projection properties (6.2)-(6.4), the trace inequality (6.17), the Cauchy-Schwarz and Young inequalities.

$$
\begin{aligned}
\Phi_1 &= \int_0^T \sum_{E\in\mathcal{E}_h}\int_E (\boldsymbol{\sigma}(E^I_{\mathbf{u},t}) : \epsilon(E^A_{\mathbf{u}})) + \int_0^T J_0^{\delta,\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u}}) \\
&\quad +\theta_{DG}\int_0^T \sum_{e_a\in\Gamma_{int}}\int_{e_a} [E^I_{\mathbf{u},t}]\cdot\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\boldsymbol{\nu}^a\}\, ds \\
&\quad -\int_0^T \sum_{e_a\in\Gamma_{int}}\int_{e_a} [E^A_{\mathbf{u}}]\cdot\{\boldsymbol{\sigma}(E^I_{\mathbf{u},t})\boldsymbol{\nu}^a\}\, ds \\
&\leq \int_0^T |||E^I_{\mathbf{u},t}|||_1\ |||E^A_{\mathbf{u}}|||_1 \\
&\quad +\int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}||[E^I_{\mathbf{u},t}]||_{0,e_a}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}||[E^A_{\mathbf{u}}]||_{0,e_a} \\
&\quad +\int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}||[E^I_{\mathbf{u},t}]||_{0,e_a}\left(\frac{|e_a|^\beta}{\delta}\right)^{1/2}||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\boldsymbol{\nu}^a\}||_{0,e_a} \\
&\quad +\int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}||[E^A_{\mathbf{u}}]||_{0,e_a}\left(\frac{|e_a|^\beta}{\delta}\right)^{1/2}||\{\boldsymbol{\sigma}(E^I_{\mathbf{u},t})\boldsymbol{\nu}^a\}||_{0,e_a} \\
&\leq C\int_0^T |||E^I_{\mathbf{u},t}|||_1^2 + C\int_0^T |||E^A_{\mathbf{u}}|||_1^2 \\
&\quad +C\int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{h_a}\right)||[E^I_{\mathbf{u},t}]||_{0,e_a} + \frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}}, E^A_{\mathbf{u}}) \\
&\quad +\int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{h_a}{\delta}\right)||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\boldsymbol{\nu}^a\}||^2_{0,e_a} + \int_0^T \sum_{e_a\in\Gamma_{int}}\left(\frac{h_a}{\delta}\right)||\{\boldsymbol{\sigma}(E^I_{\mathbf{u},t})\boldsymbol{\nu}^a\}||^2_{0,e_a} \\
&\leq Ch^{\min\{2r_2,2t-2\}}\int_0^T ||\mathbf{u}_t||^2_{\min\{r_2+1,t\}} + C\int_0^T |||E^A_{\mathbf{u}}|||_1^2 \\
&\quad +\frac{1}{8}\int_0^T J_0^{\delta,\beta}(E^A_{\mathbf{u}}, E^A_{\mathbf{u}}).
\end{aligned}
$$

$$
\begin{aligned}
\Phi_2 &= -\sum_{E\in\mathcal{E}_h}\int_E (\boldsymbol{\sigma}(E^I_{\mathbf{u}}) : \epsilon(E^A_{\mathbf{u}}))|_{t=T} - J_0^{\delta,\beta}(E^I_{\mathbf{u}}, E^A_{\mathbf{u}})|_{t=T} \\
&\quad -\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}[E^I_{\mathbf{u}}]\cdot\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\boldsymbol{\nu}^a\}|_{t=T}\,ds \\
&\quad +\sum_{e_a\in\Gamma_{int}}\int_{e_a}[E^A_{\mathbf{u}}]\cdot\{\boldsymbol{\sigma}(E^I_{\mathbf{u}})\boldsymbol{\nu}^a\}|_{t=T}\,ds \\
&\leq |||E^I_{\mathbf{u}}(T)|||_1\ |||E^A_{\mathbf{u}}(T)|||_1 \\
&\quad +\sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}|||[E^I_{\mathbf{u}}(T)]|||_{0,e_a}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}|||[E^A_{\mathbf{u}}(T)]|||_{0,e_a} \\
&\quad +\sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}|||[E^I_{\mathbf{u}}(T)]|||_{0,e_a}\left(\frac{|e_a|^\beta}{\delta}\right)^{1/2}||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}}(T))\boldsymbol{\nu}^a\}||_{0,e_a} \\
&\quad +\sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{|e_a|^\beta}\right)^{1/2}|||[E^A_{\mathbf{u}}(T)]|||_{0,e_a}\left(\frac{|e_a|^\beta}{\delta}\right)^{1/2}||\{\boldsymbol{\sigma}(E^I_{\mathbf{u}}(T))\boldsymbol{\nu}^a\}||_{0,e_a} \\
&\leq C|||E^I_{\mathbf{u}}(T)|||_1^2+\epsilon|||E^A_{\mathbf{u}}(T)|||_1^2 \\
&\quad +C\sum_{e_a\in\Gamma_{int}}\left(\frac{\delta}{h_a}\right)||[E^I_{\mathbf{u}}]||_{0,e_a}+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}}(T),E^A_{\mathbf{u}}(T)) \\
&\quad +\epsilon_1\sum_{e_a\in\Gamma_{int}}\left(\frac{h_a}{\delta}\right)||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}}(T))\boldsymbol{\nu}^a\}||^2_{0,e_a}+C\sum_{e_a\in\Gamma_{int}}\left(\frac{h_a}{\delta}\right)||\{\boldsymbol{\sigma}(E^I_{\mathbf{u}}(T))\boldsymbol{\nu}^a\}||^2_{0,e_a} \\
&\leq Ch^{\min\{2r_2,2s-2\}}||\mathbf{u}(T)||^2_{\min\{s_2+1,t\}}+\epsilon_1|||E^A_{\mathbf{u}}(T)|||_1^2+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}}(T),E^A_{\mathbf{u}}(T)),
\end{aligned}
$$

where $\epsilon_1 > 0$ is an arbitrarily small parameter. And Φ_3 is bounded by use of Cauchy-Schwarz, Young's inequality, the definition of $J_0^{\tilde{\delta},\beta}$ and (6.3):

$$
\begin{aligned}
\Phi_3 &= -\int_0^T J_0^{\tilde{\delta},\beta}(E^I_{\mathbf{u},t}, E^A_{\mathbf{u},t}) \\
&= -\int_0^T \sum_{e_a\in\Gamma_{int}} \int_{e_a} \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right)^{1/2} [E^I_{\mathbf{u},t}] \cdot \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right)^{1/2} [E^A_{\mathbf{u},t}] \\
&\leq \int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right)^{1/2} ||E^I_{\mathbf{u},t}||_{0,e_a} \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right)^{1/2} ||E^A_{\mathbf{u},t}||_{0,e_a} \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right) ||E^I_{\mathbf{u},t}||^2_{0,e_a} + \frac{1}{8}\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}}{|e_a|^\beta}\right) ||E^A_{\mathbf{u},t}||^2_{0,e_a} \\
&= C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}}{h_a}\right) ||E^I_{\mathbf{u},t}||^2_{0,e_a} + \frac{1}{8}\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}) \\
&\leq Ch^{\min\{2r_2,2t-2\}} \int_0^T ||\mathbf{u}_t||^2_{\min\{r_2+1,t\}} + \frac{1}{8}\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}).
\end{aligned}
$$

Using the trace inequality (6.18), and the inequality and the constant $C_{\mathbf{z}}$ from the parabolic lift theorem (3.22) leads to the following bound for Φ_4:

$$
\begin{aligned}
\Phi_4 &= \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E_p^A\} \boldsymbol{\nu}^a \cdot [E_{\mathbf{u},t}^A] \\
&\leq \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} |\{E_p^A\}||[E_{\mathbf{u},t}^A]| \\
&\leq \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}} \right)^{1/2} ||\{E_p^A\}||_{0,e_a} \left(\frac{\tilde{\delta}}{|e_a|^\beta} \right)^{1/2} |||[E_{\mathbf{u},t}^A]|||_{0,e_a} \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||\{E_p^A\}||_{0,e_a}^2 + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||E_p^A|_{E_1}||_{0,e_a}^2 + C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||E_p^A|_{E_2}||_{0,e_a}^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a^{-1}}{\tilde{\delta}} ||E_p^A||_{0,E_1}^2 + C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a^{-1}}{\tilde{\delta}} ||E_p^A||_{0,E_2}^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq \frac{C}{\tilde{\delta}_{min}} \int_0^T ||E_p^A||_0^2 + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq \frac{C_{\mathbf{z}}^2}{\tilde{\delta}_{min}} \int_0^T \left(||E_{\mathbf{z}}^I||_0^2 + ||E_{\mathbf{z}}^A||_0^2 \right) + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq C h^{\min\{2r_1+2, 2q-2\}} \int_0^T ||\mathbf{z}||_{\min\{r_1+1, q-1\}}^2 + \frac{C_{\mathbf{z}}^2}{\tilde{\delta}_{min}} \int_0^T ||E_{\mathbf{z}}^A||_0^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A).
\end{aligned}
$$

To bound Φ_5 the trace inequality (6.21) can be employed over each edge-element pair since $E_p^I|_{E_a}$ belongs to $H^1(E_a)$ for each element $E_a \in \mathcal{E}_h$. Hence,

$$
\begin{aligned}
\Phi_5 &= \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} \{E_p^I\} \boldsymbol{\nu}^a \cdot [E_{\mathbf{u},t}^A] \\
&\leq \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \int_{e_a} |\{E_p^I\}||[E_{\mathbf{u},t}^A]| \\
&\leq \alpha \int_0^T \sum_{e_a \in \Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}} \right)^{1/2} ||\{E_p^I\}||_{0,e_a} \left(\frac{\tilde{\delta}}{|e_a|^\beta} \right)^{1/2} ||E_{\mathbf{u},t}^A||_{0,e_a} \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||\{E_p^I\}||_{0,e_a}^2 + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||E_p^I|_{E_1}||_{0,e_a}^2 + C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta}{\tilde{\delta}} ||E_p^I|_{E_2}||_{0,e_a}^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a^{-1}}{\tilde{\delta}} ||E_p^I||_{0,E_1}^2 + C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a}{\tilde{\delta}} ||E_p^I||_{1,E_1}^2 \\
&\quad C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a^{-1}}{\tilde{\delta}} ||E_p^I||_{0,E_2}^2 + C \int_0^T \sum_{e_a \in \Gamma_{int}} \frac{|e_a|^\beta h_a}{\tilde{\delta}} ||E_p^I||_{1,E_2}^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq \frac{C}{\tilde{\delta}_{min}} \int_0^T \sum_{E_a \in \mathcal{E}_h} ||E_p^I||_{0,E_a}^2 + \frac{C}{\tilde{\delta}_{min}} \int_0^T \sum_{E_a \in \mathcal{E}_h} h_a^2 ||E_p^I||_{1,E_a}^2 \\
&\quad + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq \frac{C}{\tilde{\delta}_{min}} \int_0^T \sum_{E_a \in \mathcal{E}_h} h_a^{\min\{2r_1+2,2q\}} ||p||_{\min\{r_1+1,q\},E_a}^2 \\
&\quad + \frac{C}{\tilde{\delta}_{min}} h^2 \int_0^T \sum_{E_a \in \mathcal{E}_h} h_a^{\min\{2r_1,2q\}} ||p||_{\min\{r_1+1,q\},E_a}^2 + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A) \\
&\leq \frac{C}{\tilde{\delta}_{min}} h^{\min\{2r_1+2,2q\}} \int_0^T ||p||_{\min\{r_1+1,q\}}^2 + \frac{1}{8} \int_0^T J_0^{\tilde{\delta},\beta}(E_{\mathbf{u},t}^A, E_{\mathbf{u},t}^A).
\end{aligned}
$$

The bound for Φ_6 is a straightforward use of the Cauchy-Schwarz and Young inequalities:

$$\begin{aligned}
\Phi_6 &= -\int_0^T (c_o E^I_{p\,,t}, E^A_p) \\
&= -\int_0^T (c_o^{1/2} E^I_{p\,,t}, c_o^{1/2} E^A_p) \\
&\leq \int_0^T ||c_o^{1/2} E^I_{p\,,t}||_0 ||c_o^{1/2} E^A_p||_0 \\
&\leq \frac{1}{2}\int_0^T ||c_o^{1/2} E^I_{p\,,t}||_0^2 + \frac{1}{2}\int_0^T ||c_o^{1/2} E^A_p||_0^2 \\
&\leq C\int_0^T ||E^I_{p\,,t}||_0^2 + \frac{1}{2}\int_0^T ||c_o^{1/2} E^A_p||_0^2 \\
&\leq Ch^{\min\{2r_1+2,2r\}}\int_0^T ||p_t||^2_{\min\{r_1+1,r\}} + \frac{1}{2}\int_0^T ||c_o^{1/2} E^A_p||_0^2.
\end{aligned}$$

To bound Φ_7, integrate by parts and again use the initial time assumption, $\nabla \cdot E^A_{\mathbf{u}}|_{t=0} = 0$:

$$\begin{aligned}
\Phi_7 &= \alpha\int_0^T (E^I_p, \nabla\cdot E^A_{\mathbf{u}\,,t}) \\
&= -\alpha\int_0^T (E^I_{p\,,t}, \nabla\cdot E^A_{\mathbf{u}}) + \alpha(E^I_p, \nabla\cdot E^A_{\mathbf{u}})_{t=T} \\
&\leq \alpha\int_0^T \sum_{E\in\mathcal{E}} ||E^I_{p\,,t}||_{E,0} ||\nabla\cdot E^A_{\mathbf{u}}||_{E,0} + \alpha||E^I_p(T))||_{E,0}||\nabla\cdot E^A_{\mathbf{u}}(T)||_{E,0} \\
&\leq \frac{1}{2}\int_0^T ||E^I_{p\,,t}||_0^2 + \frac{1}{2}\int_0^T |||E^A_{\mathbf{u}}|||_1^2 \\
&\quad + C||E^I_p(T)||_0^2 + \epsilon_2|||E^A_{\mathbf{u}}(T)|||_1^2 \\
&\leq Ch^{\min\{2r_1+2,2r\}}\int_0^T ||p_t||^2_{\min\{r_1+1,r\}} + \frac{1}{2}\int_0^T |||E^A_{\mathbf{u}}|||_1^2 \\
&\quad + Ch^{\min\{2r_1+2,2q\}}||p(T)||^2_{\min\{r_1+1,q\}} + \epsilon_2|||E^A_{\mathbf{u}}(T)|||_1^2,
\end{aligned}$$

where $\epsilon_2 > 0$ is an arbitrarily small parameter.

The bound for Φ_8 is found by applying the Cauchy-Schwarz inequality and Young's inequality (with $\epsilon = \frac{\lambda_{min}}{2\lambda_{max}}$),

$$\begin{aligned}
\Phi_8 &= -\int_0^T (\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^I, \tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^A) \\
&\leq \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^I||_0 ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^A||_0 \\
&\leq C \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^I||_0^2 + \frac{\lambda_{min}}{2\lambda_{max}} \int_0^T ||\tilde{\boldsymbol{\kappa}}^{-\frac{1}{2}} E_{\mathbf{z}}^A||_0^2 \\
&\leq C \int_0^T ||E_{\mathbf{z}}^I||_0^2 + \frac{1}{2} \int_0^T ||E_{\mathbf{z}}^A||_0^2 \\
&\leq C h^{\min\{2r_1+2, 2q-2\}} \int_0^T ||\mathbf{z}||^2_{\min\{r_1+1, q-1\}} + \frac{1}{2\lambda_{max}} \int_0^T ||E_{\mathbf{z}}^A||_0^2.
\end{aligned}$$

In order to bound Φ_9 the Cauchy-Schwarz and Young inequalities are employed, the trace inequality (6.17) is used, and the continuity of the elasticity operator is recalled to deduce

$$
\begin{aligned}
\Phi_9 &= (1+\theta_{DG})\int_0^T \sum_{e_a\in\Gamma_{int}} \int_{e_a} \{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}\cdot[E^A_{\mathbf{u},t}] \\
&= (1+\theta_{DG})\int_0^T \sum_{e_a\in\Gamma_{int}} \int_{e_a} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right)^{\frac{1}{2}} \{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}\cdot\left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right)^{\frac{1}{2}}[E^A_{\mathbf{u},t}] \\
&\leq C\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{|e_a|^\beta}{\tilde{\delta}_a}\right)^{\frac{1}{2}} ||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}||_{0,e_a}\left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right)^{\frac{1}{2}} |||[E^A_{\mathbf{u},t}]|||_{0,e_a} \\
&\leq Ch^{\beta(d-1)}\int_0^T \sum_{e_a\in\Gamma_{int}} ||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}||^2_{0,e_a} + \frac{1}{8}\int_0^T \sum_{e_a\in\Gamma_{int}} \left(\frac{\tilde{\delta}_a}{|e_a|^\beta}\right)|||[E^A_{\mathbf{u},t}]|||^2_{0,e_a} \\
&\leq Ch^{\beta(d-1)-1}\int_0^T \sum_{e_a\in\Gamma_{int}} \int_{E^{12}_a} (\boldsymbol{\sigma}(E^A_{\mathbf{u}}) : \epsilon(E^A_{\mathbf{u}})) + \frac{1}{8}\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}) \\
&\leq C\int_0^T |||E^A_{\mathbf{u}}|||^2_1 + \frac{1}{8}\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t}, E^A_{\mathbf{u},t}).
\end{aligned}
$$

Finally, Φ_{10} is bounded similarly; however, since there is no time derivative in this term (and thus one cannot use Gronwall's inequality here), there will be some care made to keep the δ_{min}-term visible in the constant C:

$$
\begin{aligned}
\Phi_{10} &= -\theta_{DG}\sum_{e_a\in\Gamma_{int}}\int_{e_a}\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}\cdot[E^A_{\mathbf{u}}]_{t=T}\\
&\leq C\sum_{e_a\in\Gamma_{int}}\int_{e_a}\left(\frac{|e_a|^\beta}{\delta_a}\right)^{\frac{1}{2}}\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}_{t=T}\cdot\left(\frac{\delta_a}{|e_a|^\beta}\right)^{\frac{1}{2}}[E^A_{\mathbf{u}}]_{t=T}\\
&\leq C\sum_{e_a\in\Gamma_{int}}\left(\frac{|e_a|^\beta}{\delta_a}\right)^{\frac{1}{2}}||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}_{t=T}||_{0,e_a}\left(\frac{\delta_a}{|e_a|^\beta}\right)^{\frac{1}{2}}|||[E^A_{\mathbf{u}}(T)]|||_{0,e_a}\\
&\leq \frac{C}{\delta_{min}}h^{\beta(n-1)}\sum_{e_a\in\Gamma_{int}}||\{\boldsymbol{\sigma}(E^A_{\mathbf{u}})\nu^a\}_{t=T}||^2_{0,e_a}+\frac{1}{8}\sum_{e_a\in\Gamma_{int}}\left(\frac{\delta_a}{|e_a|^\beta}\right)|||[E^A_{\mathbf{u}}(T)]|||^2_{0,e_a}\\
&\leq \frac{C}{\delta_{min}}h^{\beta(n-1)-1}\sum_{e_a\in\Gamma_{int}}\int_{E^{12}_a}(\boldsymbol{\sigma}(E^A_{\mathbf{u}}):\epsilon(E^A_{\mathbf{u}}))_{t=T}+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T}\\
&\leq \frac{C}{\delta_{min}}\sum_{E\in\mathcal{E}}\int_E(\boldsymbol{\sigma}(E^A_{\mathbf{u}}):\epsilon(E^A_{\mathbf{u}}))_{t=T}+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T}.
\end{aligned}
$$

Having bounded the terms Φ_1 to Φ_{10}, (6.30) is revisited. Rearranging its terms now as an inequality,

$$
\begin{aligned}
&\left(\frac{1}{2}-\frac{C}{\delta_{min}}\right)\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(E^A_{\mathbf{u}}):\epsilon(E^A_{\mathbf{u}}))_{t=T}+\left(\frac{1}{2}-\frac{1}{8}-\frac{1}{8}\right)J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T}\\
&+\left(1-\frac{1}{8}-\frac{1}{8}-\frac{1}{8}-\frac{1}{8}\right)\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t},E^A_{\mathbf{u},t})+\frac{1}{2}||c_o^{1/2}E^A_p(T)||^2_0\\
&+\left(\frac{1}{\lambda_{max}}-\frac{1}{2\lambda_{max}}-\frac{C^2_{\mathbf{z}}}{\tilde{\delta}_{min}}\right)\int_0^T||E^A_{\mathbf{z}}||^2_0\\
&\qquad\leq I+II+III,
\end{aligned}
\tag{6.32}
$$

where

$$I = (\epsilon_1+\epsilon_2)|||E^A_{\mathbf{u}}(T)|||^2_1,$$

$$II = C\bigg(\int_0^T ||c_o^{1/2}E^A_p||^2_0 + \int_0^T |||E^A_{\mathbf{u}}|||^2_1 + \int_0^T J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})\bigg),$$

$$\begin{aligned} III = C\bigg(& h^{\min\{2r_1+2,2q\}}||p(T)||^2_{\min\{r_1+1,q\}} \\ &+h^{\min\{2r_1+2,2r\}}\int_0^T ||p_t(\tau)||^2_{\min\{r_1+1,r\}}\,d\tau \\ &+h^{\min\{2r_1+2,2q-2\}}\int_0^T ||\mathbf{z}(\tau)||_{\min\{r_1+1,q-1\}}\,d\tau \\ &+h^{\min\{2r_2+2,2s-2\}}||\mathbf{u}(T)||^2_{\min\{r_2+1,s\}} \\ &+h^{\min\{2r_2+2,2t-2\}}\int_0^T ||\mathbf{u}_t(\tau)||^2_{\min\{r_2+1,t\}}\,d\tau\bigg) \end{aligned}$$

The I-term can be hidden in the left side of eq. (6.32) after modifying the lhs by using the piecewise H^1 Korn's inequality; the II-term will be eliminated by use of Gronwall's inequality; the III-term consists of the interpolation error.

Specifically, in order to handle term I, use is made of the piecewise H^1 version of Korn's inequality (6.14) and the assumption that $|\Gamma_o|>0$. Let δ be chosen so that $\left(\frac{1}{2}-\frac{C}{\delta_{min}}\right)=\frac{1}{8}$. Then

$$\frac{1}{8}\sum_{E\in\mathcal{E}_h}\int_E(\boldsymbol{\sigma}(E^A_{\mathbf{u}}):\epsilon(E^A_{\mathbf{u}}))_{t=T}+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})_{t=T} \;\geq\; \frac{1}{8}C_{coer}|||E^A_{\mathbf{u}}(T)|||^2_1.$$

Using this inequality in eq. (6.32) allows one to hide the I-term in the left hand side as follows:

$$\begin{aligned}\left(\frac{1}{8}C_{coer}-\epsilon_1-\epsilon_2\right)|||E^A_{\mathbf{u}}(T)|||_1^2+\frac{1}{8}J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})|_{t=T}\\ +\frac{1}{2}\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t},E^A_{\mathbf{u},t})+\frac{1}{2}||c_o^{1/2}E^A_p(T)||_0^2\\ +\left(\frac{1}{2\lambda_{max}}-\frac{C_{\mathbf{z}}^2}{\tilde{\delta}_{min}}\right)\int_0^T||E^A_{\mathbf{z}}||_0^2\\ \leq II+III. \qquad (6.33)\end{aligned}$$

Choose ϵ_1 and ϵ_2 to be small enough, the coefficient of the $|||E^A_{\mathbf{u}}(T)|||_0^2$-term is positive. Also, let $\tilde{\delta}_{min}$ be large enough so that the coefficient of $\int_0^T||E^A_{\mathbf{z}}||_0^2$ is positive. Having done this, let C_{min} to be the positive minimum coefficient of the terms constituting the lhs of eq. (6.33) to find

$$\begin{aligned}C_{min}\Bigg(|||E^A_{\mathbf{u}}(T)|||_1^2+J_0^{\delta,\beta}(E^A_{\mathbf{u}},E^A_{\mathbf{u}})|_{t=T}+\int_0^T J_0^{\tilde{\delta},\beta}(E^A_{\mathbf{u},t},E^A_{\mathbf{u},t})\\ +||c_o^{1/2}E^A_p(T)||_0^2+\int_0^T||E^A_{\mathbf{z}}||_0^2\Bigg)\\ \leq II+III. \qquad (6.34)\end{aligned}$$

After dividing through by C_{min}, Gronwall's inequality eliminates the II-term from the rhs of eq. (6.34). At this point, the inessential positive terms on the left side of (6.34) may be dropped while preserving the inequality. Then by setting $2R$ to be the minimal exponent of the element size h in the term III the theorem results:

$$|||E^A_{\mathbf{u}}|||^2_{L^\infty(H^1)}+||c_o^{1/2}E^A_p||^2_{L^\infty(L^2)}+||E^A_{\mathbf{z}}||^2_{L^2(L^2)}\leq CM(p,p_t,\mathbf{z},\mathbf{u},\mathbf{u}_t)\ h^{2R},$$

where

$$M(p,p_t,\mathbf{z},\mathbf{u},\mathbf{u}_t) = \max\Big\{ ||p||^2_{L^2(H^{\min\{r_1+1,q\}})}, ||p_{,t}||^2_{L^2(H^{\min\{r_1+1,r\}})}, \\ ||\mathbf{z}||^2_{L^2(H^{\min\{r_1+1,q-1\}})}, \\ ||\mathbf{u}||^2_{L^2(H^{\min\{r_2+1,s\}})}, ||\mathbf{u}_{,t}||^2_{L^2(H^{\min\{r_2+1,t\}})} \Big\}.$$

The theorem readily follows from the above inequality.

□

Having obtained the above auxiliary error estimates and the parabolic lift theorem (3.22), it is a routine matter to obtain the general finite element error estimate which is summarized in the following corollary:

Corollary 6.3.2 (General Finite Element Error Estimate)**.** *With the same conditions for the above theorem, the following finite element error estimate holds:*

$$|||\mathbf{u}-\bar{\mathbf{u}}|||^2_{L^\infty(H^1)} + ||c_o^{1/2}(p-\bar{p})||_{L^\infty(L^2)} + ||\mathbf{z}-\bar{\mathbf{z}}||_{L^2(L^2)} \\ \leq\ C(T,\boldsymbol{\kappa},C_{cont},C_{coer},c_o,p,p_t,\mathbf{z},\mathbf{u},\mathbf{u}_t)\ h^R. \tag{6.35}$$

Here, $C = C(T,\boldsymbol{\kappa},C_{coer},C_{cont},c_o,p,p_{,t},\mathbf{z},\mathbf{u},\mathbf{u}_{,t})$*, and*

$$R = \min\{r_1+1,\ r_2,\ q-1,\ r,\ s-1,\ t-1\}.$$

As for the CG/Mixed theorem, for the situation where c_o is bounded below by a positive constant[5], one may deduce the following result for the pressure error:

$$||p - \bar{p}||_{L^\infty(L^2)} \ \leq \ Ch^R. \tag{6.36}$$

For those problems where $c_o = 0$ on measurable regions of the domain, one may use the same reasoning as that utilized for the CG/Mixed method to conclude the following corollary which admits optimality for the pressure in the weaker $L^2(L^2)$ norm, valid even where $c_o = 0$.

Corollary 6.3.3 (Error Estimate for the Incompressible Limit Case with an $L^2(L^2)$ Estimate for the Pressure)**.** *Let $r_1 \geq 0$ be the order of the mixed space $(W_h, \mathbf{S}_h)$, and let $r_2 \in \{1, 2, 3\}$ be the degree of the polynomials used in the displacement space $\mathbf{V}_h$. Assume $r_1 < r_2$, the regularity (6.23)-(6.24) of the exact solution, parameter constraints (2.1)-(2.3) and no initial time error, Then, if $\beta = (d-1)^{-1}$, and the parameters δ and $\tilde{\delta}$ are large enough,*

$$||\mathbf{u} - \bar{\mathbf{u}}||_{L^\infty(H^1)} + ||p - \bar{p}||_{L^2(L^2)} + ||\mathbf{z} - \bar{\mathbf{z}}||_{L^2(L^2)} \ \leq \ Ch^R. \tag{6.37}$$

Here, $C = C(T, \boldsymbol{\kappa}, C_{cont}, C_{coer}, c_o, p, p_t, \mathbf{z}, \mathbf{u}_t)$, and

$$R = \min\{r_1 + 1, \ r_2, \ q - 1, \ s, \ t - 1\},$$

which reflects both the order of the approximation spaces and the regularity of the true solution.

[5]That is, when $c_o(x) \geq \gamma_c > 0$ for $\forall x \in \Omega$ and for some positive constant γ_c.

6.3.2.1 Penalty Parameters and Locking

As pointed out earlier, the proof of the above error estimates hinges on the flexibility to make the penalty parameters, δ and $\tilde{\delta}$, as large as necessary. Unfortunately, since these parameters penalize inter-element discontinuity large parameters might cause the solution to be effectively continuous and produce locking.

Inequality (6.32) explains the reason for the size of δ as it must be chosen large enough so that $\delta_{min} > 2C$. In this case C comes from the trace inequality (6.17). On the other hand, (6.33) shows that $\tilde{\delta}$ must be large enough so that $\tilde{\delta}_{min} > 2\lambda_{max} C_{\mathbf{z}}^2$; $C_{\mathbf{z}}$ comes from the parabolic lift theorem (3.22) and is proportional to λ_{min}^{-1}. This implies that $\tilde{\delta}_{min} > \frac{2\lambda_{max}}{\lambda_{min}^2}$. For the typical case where $\lambda_{max} = \lambda_{min}$, then the requirement becomes $\tilde{\delta}_{min} > \frac{2}{\lambda_{min}}$. Thus, for a small permeability $\tilde{\delta}$ must be chosen to be very large.

The dependence here on penalty parameter differs from the DG case for linear elasticity where large penalty parameters are not required [Rivière and Wheeler, 2000]. The dependence comes from the pressure and deformation coupling terms and the time dependence in linear poroelasticity. Furthermore, the dependence is not merely an artifact of the proof. Computational results will demonstrate show the DG/Mixed scheme to be moderately successful in overcoming locking, but small permeabilities do require large penalties which lead to an effectively continuous scheme.

Another problem potentially exists as well. Relatively large deformations may have large jumps in value across element boundaries. This means that matrix entries for the $J_0^{\delta,\beta}$ and $J_0^{\tilde{\delta},\beta}$ terms may be large even for moderate values of the penalty parameters. This might lead to a poorly conditioned

matrix. However, circumstances with large deformations may be an indicator that the linear poroelasticity equations should not be used as a physical model.

Chapter 7

Discontinuous Galerkin for Displacements Coupled with a Mixed Formulation for Flow: Numerical Results

The first purpose of this chapter is to re-run[1] the same numerical experiments using the DG/Mixed method that are used to determine the accuracy of the CG/Mixed method. It is shown that the convergence rates are optimal in the context of regularity and very similar for both methods. The second purpose of this chapter is to apply the DG/Mixed scheme to problems which are known to cause locking. Several examples will show success in this area, but only for precisely chosen penalty parameters.

7.1 Accuracy

In an earlier chapter, the CG/Mixed method is tested on several problems with known analytical solutions. In context of regularity, good convergences results are found for those sets of parameter which do not cause locking. In this section, the DG/Mixed method is tested against the same problems with the same choice of parameters. Attention is restricted to the backward Euler time-stepping scheme, but emphasis will be given to understanding any

[1] As before, only standard continuous linear elements are used to approximate the displacement, and the lowest order Raviart-Thomas space is for the flow variables.

possible differences between the various DG schemes and to the effect of the size of the penalty parameters.

7.1.1 Terzaghi's Problem

Corresponding to the test cases using the CG/Mixed method, two sets of parameters are used. Each test case uses a small time step of $\Delta t = 1e-6$ in order to minimize errors with respect to the time discretization.

Test Case I

The first test case uses the following set of physical parameters:

$$E = 1e+5, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e-1, \quad \boldsymbol{\kappa} = 1e-6.$$

Since for Terzaghi's problem, $c_o > 0$, the stronger $L^\infty(L^2)$ is used to measure the error for the pressure. Figure 7.1 show the $L^\infty(L^2)$ convergence rates using *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom). Notice that all three sets of results are virtually identical. For the choice of each DG scheme, a wide range of penalty parameter choices are used. In addition the results depict the case where the deformation penalty and the deformation time derivative penalty are equal, $\delta = \tilde{\delta}$.

As in the results for the CG/Mixed method, the convergence rate for the DG/Mixed method applied to Terzaghi's problem is degraded due to the lack of regularity of the pressure solution. $L^\infty(L^2)$ error estimates for pressure occur for

$$\delta,\ \tilde{\delta}\ \in\ [10^{-8}, 10^{7}],$$

and within this range converge on the order $O(h^{0.50001})$, a result very similar to that found with the CG/Mixed algorithm.

As shown, only when the parameters fall outside this range do problems with convergence occur primarily due the poorly conditioned matrix that results. So the convergence results for each DG method are *very robust* with respect to choice of penalty parameters, varying over a magnitude of choices on the order of 10^{15}.

It is also remarked that the penalty parameters, $\delta,\ \tilde{\delta}$, need not be equal. Good convergence rates are found if both $\delta \in [10^{-8}, 10^{7}]$ and $\tilde{\delta} \in [10^{-8}, 10^{7}]$, but independent of the value of the other. In fact, it is found that either one penalty parameter or the other can actually be zero as long as the other falls within the acceptable range.

Test Case II

The second test case uses the following set of physical parameters:

$$E = 1e+7, \quad \nu = 0.4, \quad \alpha = 1, \quad c_o = 5e-1, \quad \boldsymbol{\kappa} = 1e-4.$$

FIgure 7.2 shows the convergence rate using the $L^{\infty}(L^2)$ norm for each of the DG schemes. Once again, there is virtually no difference between the various DG schemes. The acceptable range of penalty parameters is found to be

$$\delta,\ \tilde{\delta}\ \in\ [10^{-6}, 10^{9}],$$

so the scheme is very robust with respect to the choice of penalty parameters. As each figure shows, outside this large range, problems with convergence begin

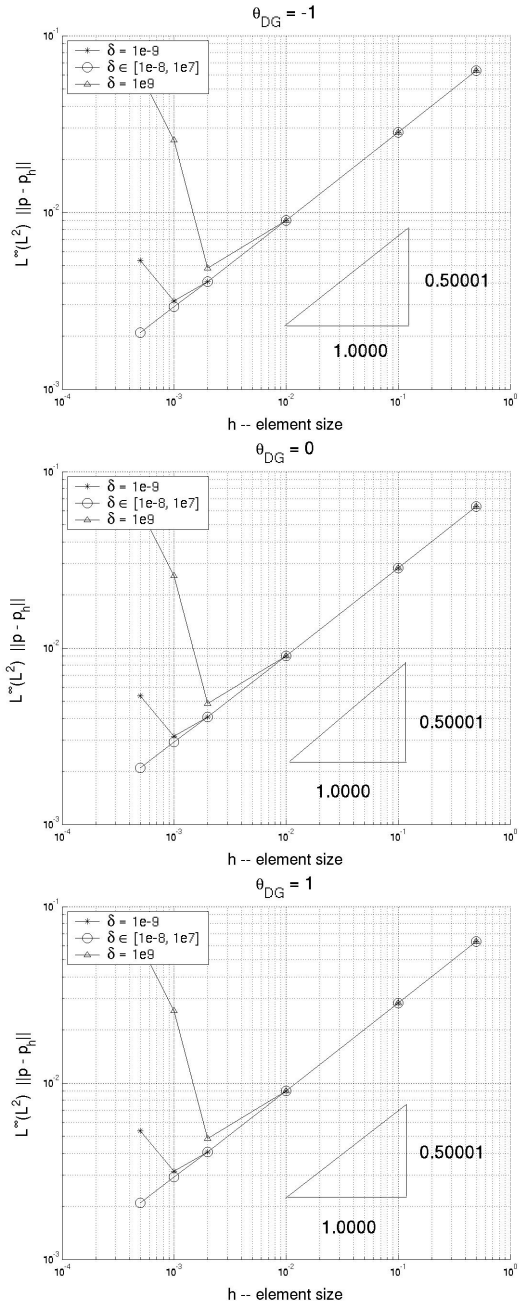

Figure 7.1: $L^{\infty}(L^2)$ pressure error in Terzaghi's problem using various penalty choices. Shown above are the *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom) variations of the DG scheme. Physical parameters used: $E = 1e5$; $\nu = 0.2$; $c_o = 1e-1$; $\boldsymbol{\kappa} = 1e-6$.

to occur. As in Test Case I, one or the other penalty parameters may actually be zero, as long as the other penalty parameter is within the appropriate range of values.

7.1.2 Mandel's Problem

The DG/Mixed algorithm is re-run using the same set of parameters when using the CG/Mixed method. A small time step of $\Delta t = 1e-6$ is used in order to minimize errors with respect to the time discretization. To remind, the following parameter values are chosen:

$$E = 1e+4, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e-1, \quad \boldsymbol{\kappa} = 1e-2.$$

Since for Mandel's problem, $c_o > 0$, the stronger $L^\infty(L^2)$ is used to measure the error for the pressure. Figure 7.3 shows the $L^\infty(L^2)$ convergence rates for the pressure using *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom). Figure 7.4 shows the $L^\infty(H^1)$ convergence rates for the displacement using *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom).

As in the results for the CG/Mixed method, the convergence rates for the pressure and displacement using the DG/Mixed method are degraded due to the lack of regularity of the pressure solution. $L^\infty(L^2)$ error estimates for pressure converge on the order $O(h^{0.500117})$, whereas the $L^\infty(H^1)$ error estimates for displacement converge on the order $O(h^{0.500196})$. The acceptable choice of penalty parameters are found to be

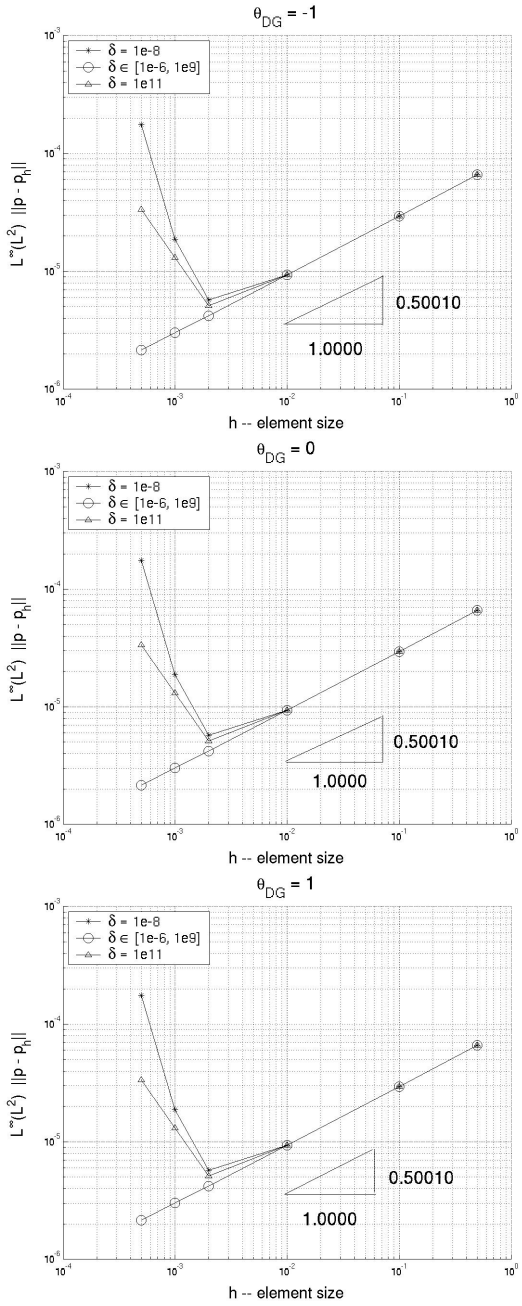

Figure 7.2: $L^{\infty}(L^2)$ pressure error in Terzaghi's problem using various penalty choices. Shown above are the *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom) variations of the DG scheme. This second set of physical parameters includes: $E = 1e7;\ \nu = 0.4;\ c_o = 5e-1;\ \boldsymbol{\kappa} = 1e-4.$

$$
\begin{aligned}
SIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [0, 10^4]; \\
IIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [10^{-6}, 10^4]; \\
NIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [0, 10^4].
\end{aligned}
$$

As shown, only when the parameters fall outside this range do problems with convergence occur primarily due the poorly conditioned matrix that results. So the convergence results for each DG method are *very robust* with respect to choice of penalty parameters, in particular the SIPG and NIPG routines which admit good numerical solutions when both penalty parameters are zero. In addition, either of the penalty parameters may be zero in the IIPG scheme as long as the other penalty parameter falls within the acceptable range of values.

7.2 Problems with a Null Constrained Specific Storage Coefficient

One of the guiding reasons for the development of the discontinuous variant of the CG/Mixed scheme is its potential to non-physical pressure oscillations when $c_o = 0$. Therefore, the DG/Mixed algorithm is applied to the same set of problems with a null constrained specific storage coefficient value that were used earlier. For the non-locking parameters, the DG/Mixed solution behaves similarly to that produced by the CG/Mixed algorithm. For the cases where locking is produced, the DG/Mixed scheme will prove to some success in removing locking

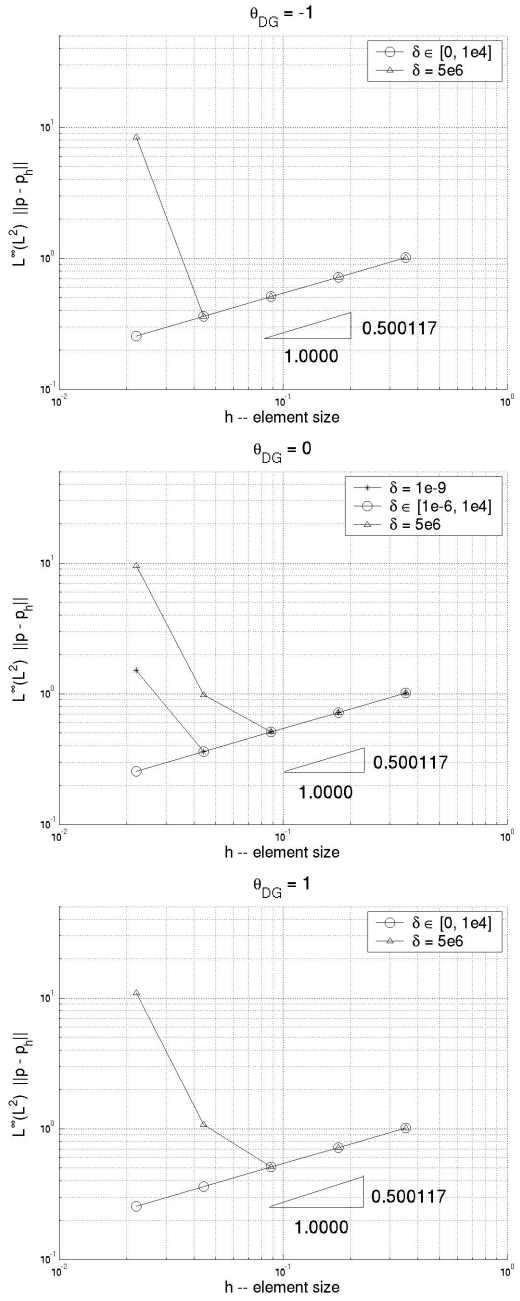

Figure 7.3: $L^\infty(L^2)$ pressure error in Mandel's problem using various penalty choices. Shown above are the *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom) variations of the DG scheme. Physical parameters used: $E = 1e4$; $\nu = 0.2$; $c_o = 1e-1$; $\boldsymbol{\kappa} = 1e-2$.

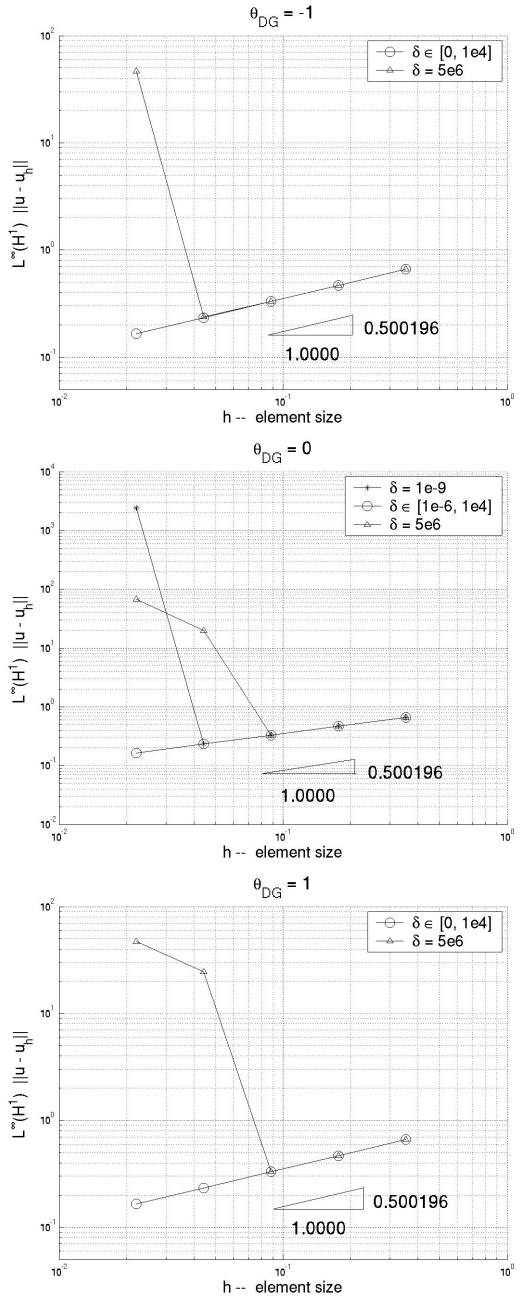

Figure 7.4: $L^{\infty}(H^1)$ displacement error in Mandel's problem using various penalty choices. Shown above are the *SIPG* (top), *IIPG* (middle) and *NIPG* (bottom) variations of the DG scheme. Physical parameters used: $E = 1e4;\ \nu = 0.2;\ c_o = 1e-1;\ \boldsymbol{\kappa} = 1e-2.$

7.2.1 Barry and Mercer's Problem

Barry and Mercer's problem was introduced earlier and the CG/Mixed algorithm was used to produce its numerical approximation. The first set of parameters did not induce locking in the CG/Mixed numerical solution, while the second set did. It will be seen that the DG/Mixed maintains good agreement with the analytical solution while overcoming the non-physical oscillations which can occur with continuous elements.

Non-Locking Parameters

To remind, the first non-locking set of parameters are

$$E = 1e+5, \quad \nu = 0.1, \quad \alpha = 1, \quad c_o = 0, \quad \boldsymbol{\kappa} = 1e-2,$$

and use a fairly large normalized time step of $\Delta\tilde{t} = \frac{1e-1\pi}{2}$.

Figure 7.5 shows the L^2 norm of the *SIPG* spatial error of the pressure and displacement at each time step. The top to bottom images shows, respectively, the results for the case where $\delta = \tilde{\delta} = 0,\ 1,\ 1e10$. For the case where $\delta = \tilde{\delta} \in [1e4, 1e10]$, very little inter-element discontinuity is allowed, and the discontinuous scheme becomes effectively continuous. This explains the similarity of the results with the CG/Mixed results. For the cases $\delta = \tilde{\delta} \in [0, 1e3]$, less continuity is demanded, and, as one might expect, the deformation error is not as small as when continuous elements are used.

Figures 7.6-7.7 show the results for the IIPG/Mixed and NIPG/MIxed schemes, respectively, for various choices of penalty parameters. The one point of significance for these two schemes is the need for the penalty parameters

to be fairly large. Any choice smaller than around $1e+3$ produces unstable numerical results.

In summary, good agreement between the various DG/Mixed numerical solutions and true solution is found for

$$\begin{aligned} SIPG: \quad \delta,\ \tilde{\delta} \quad &\in \quad [10^4, 10^{10}]; \\ IIPG: \quad \delta,\ \tilde{\delta} \quad &\in \quad [10^4, 10^{10}]; \\ NIPG: \quad \delta,\ \tilde{\delta} \quad &\in \quad [10^4, 10^{10}]. \end{aligned}$$

For the situation when $c_o = 0$ the *SIPG*-based scheme is the most robust of the family of DG alternatives provided, but *in a limited sense*. The *IIPG*- and *NIPG*-based schemes alternate between unstable and effectively continuous schemes, whereas the *SIPG*-based scheme has an intermediary stage for values $\delta = \tilde{\delta} \in [0, 1e3]$ where stable, but less accurate, solutions are admitted.

Locking Parameters

As demonstrated, the CG/Mixed scheme can sometimes suffer from numerical pressure oscillations. One of the principal motivations for the development of the DG/Mixed schemes is, in fact, to overcome this problem.

Recall the parameters which induce locking in the CG/Mixed scheme:

$$E = 1e+5, \quad \nu = 0.1, \quad \alpha = 1, \quad c_o = 0, \quad \boldsymbol{\kappa} = 1e-6,$$

and a smaller normalized time step of $\Delta\tilde{t} = \frac{1e-5\pi}{2}$.

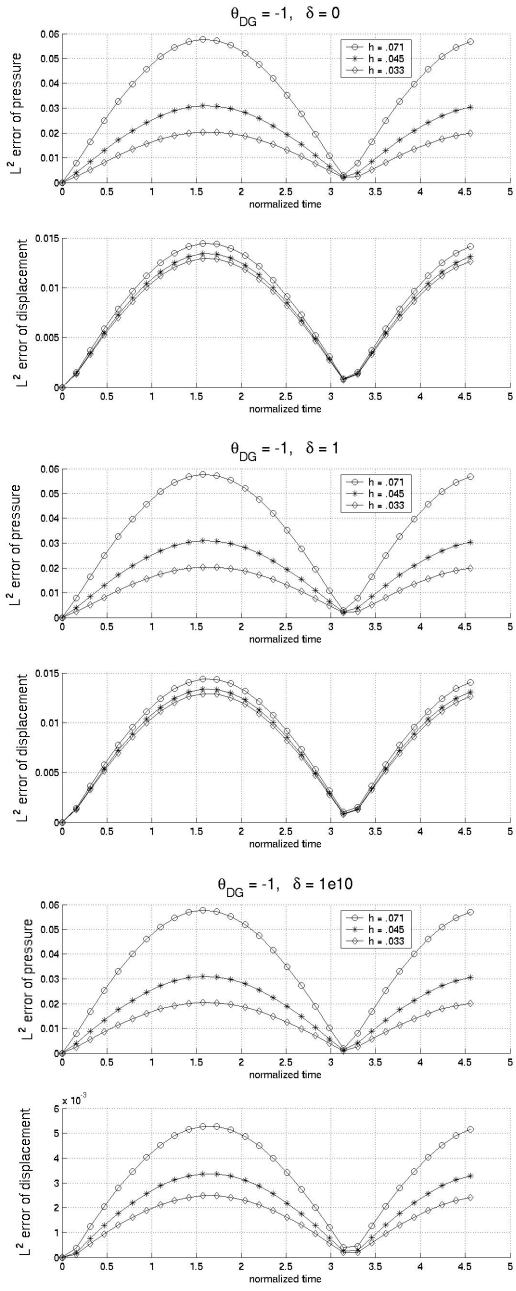

Figure 7.5: SIPG/Mixed error in Barry and Mercer's problem for various choices of penalty parameters.

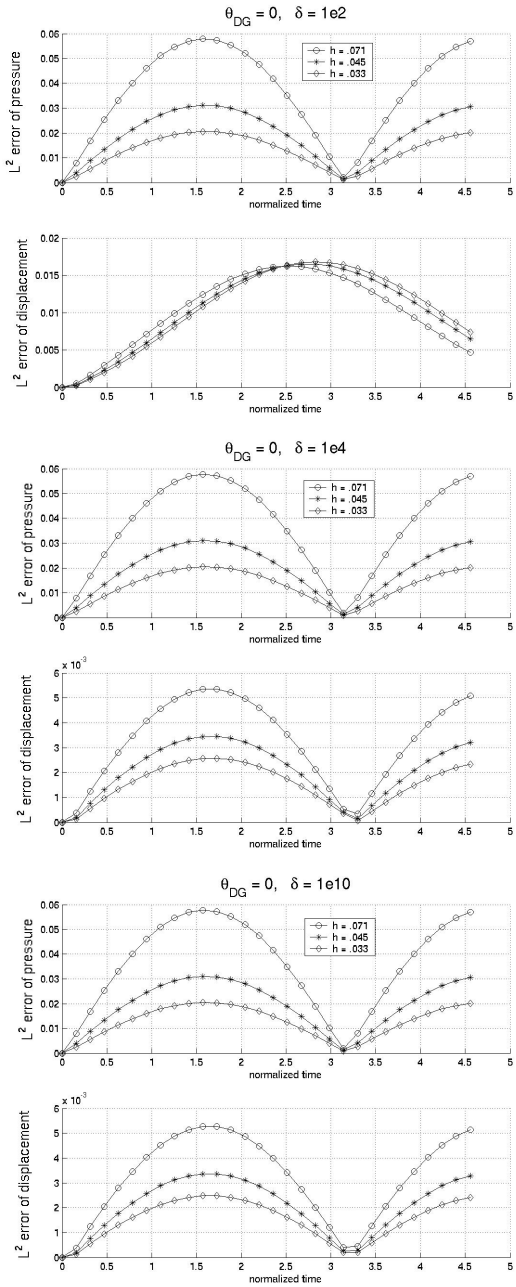

Figure 7.6: IIPG/Mixed error in Barry and Mercer's problem for various choices of penalty parameters.

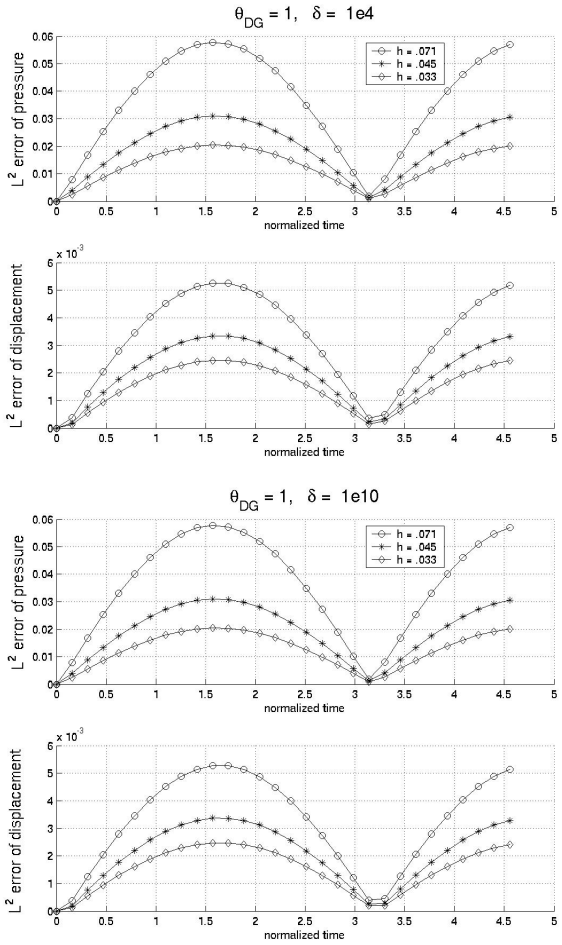

Figure 7.7: NIPG/Mixed error in Barry and Mercer's problem for various choices of penalty parameters.

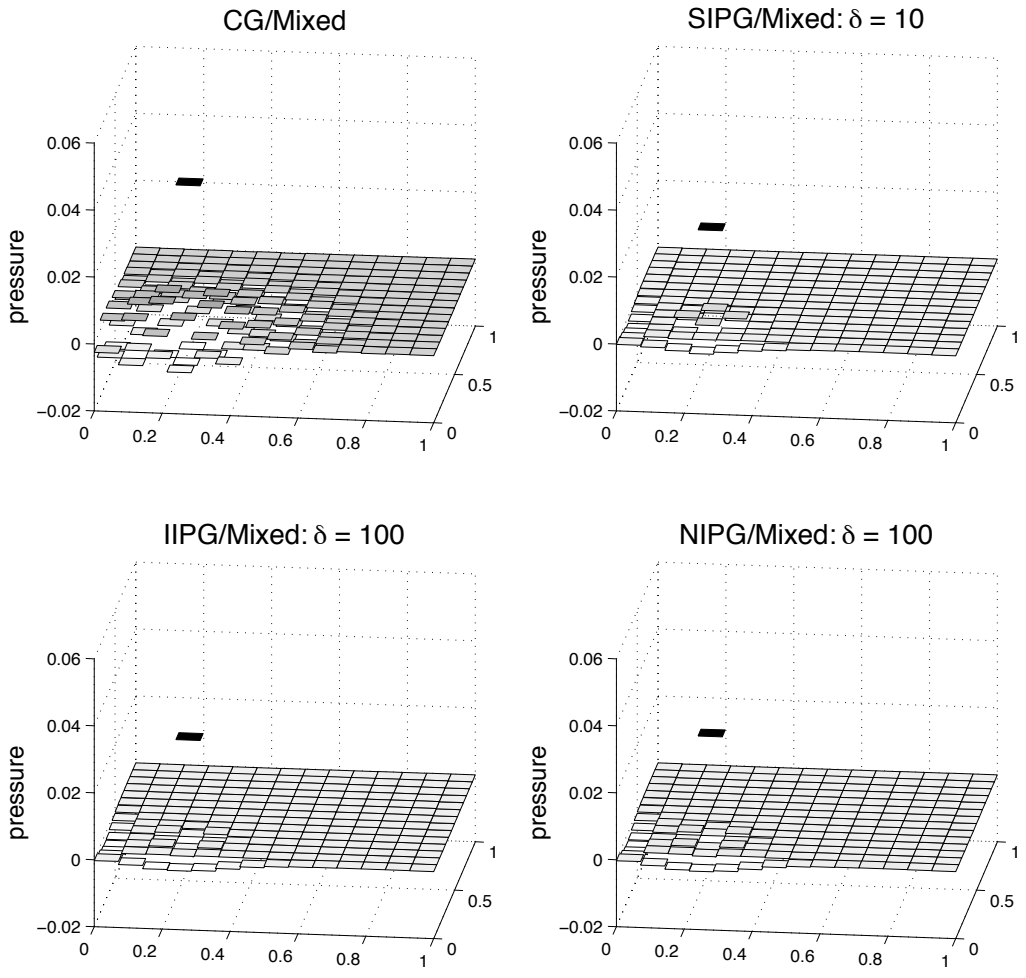

Figure 7.8: CG/Mixed Solution (upper left) vs. the DG/Mixed schemes. Using a short time interval and low permeability, spurious pressure oscillations are suppressed to a fair degree with the DG/Mixed methods.

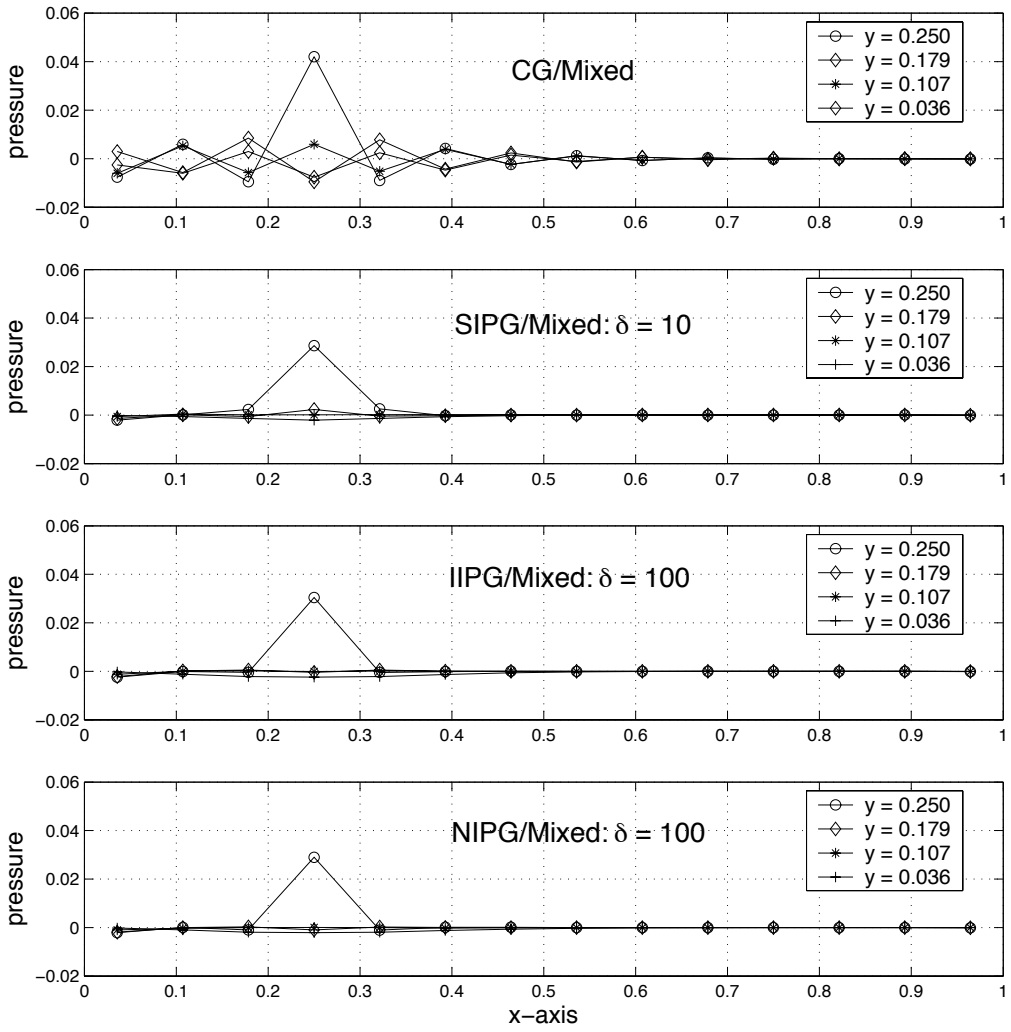

Figure 7.9: CG/Mixed solution (top) vs. the DG/Mixed schemes. Using a short time interval and low permeability, the spurious oscillations shown in the pressure cross-sections of the CG/Mixed solution are lessened with the DG/Mixed methods.

Figure 7.8 shows how each of the various DG-based schemes fares when confronted with the above parameters. The non-physical pressure oscillations have been removed, but there are some very slightly negative pressures that are produced. Figure 7.9 provides an alternate view with a look at the pressure cross-sections. Unfortunately, the penalty parameter choices which eliminate locking are very limited. In fact, the only suitable choices are

$$\begin{aligned} SIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [10^1, 10^2]; \\ IIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [10^2, 10^3]; \\ NIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [10^2, 10^3]. \end{aligned}$$

Any values smaller than the above ranges produces unstable solutions. Any values larger produces effectively continuous schemes which exhibit locking. For the larger values, the pressure oscillations, as expected, do decrease with progressive time steps.

7.2.2 Mandel's Problem and Locking

As discussed earlier, Mandel's problems admits solutions which contain spurious pressure oscillations when a null constrained specific storage coefficient value is combined with certain choices for the remaining parameters. For the example in the chapter which discusses the CG/Mixed numerical results, it was shown that Young's modulus had an important effect. For the value $E = 1e5$ the solution experienced no locking. With the value $E = 1e4$ a pressure oscillation occurred in the x-direction for each fixed value of y.

Figure 7.10 compares the numerical solutions of the CG/Mixed and the

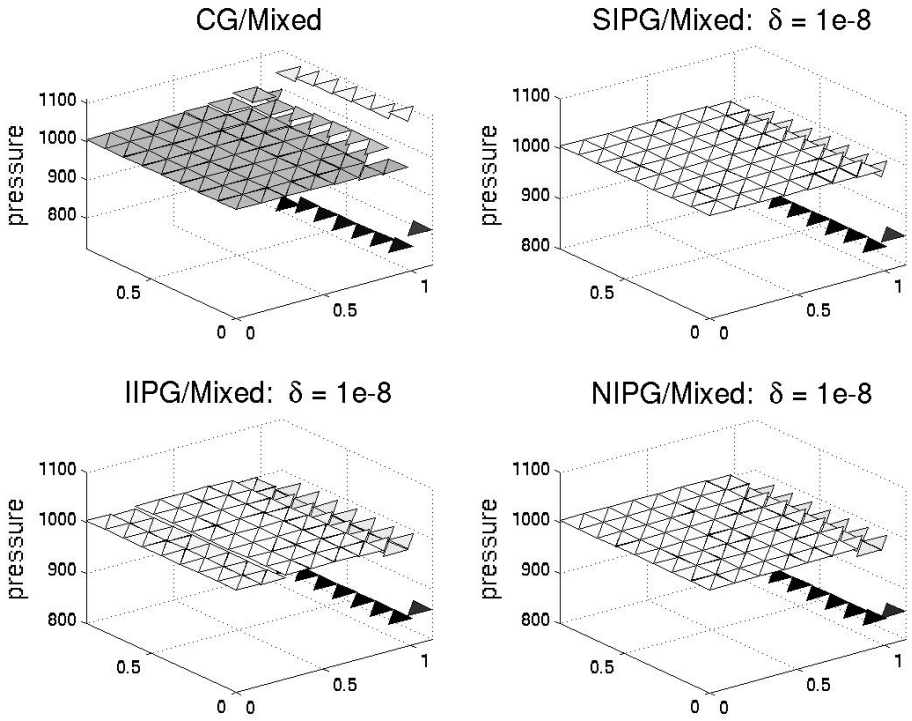

Figure 7.10: CG/Mixed Solution (upper left) vs. the DG/Mixed schemes. The true solution has non-zero divergence, but the CG/Mixed numerical solution is driven to a divergence-free state thus producing spurious pressure oscillations. These are overcome with the DG/Mixed methods.

various DG-based schemes. As one can see, for each fixed value of y, each of the pressure cross-sections decreases monotonically for the DG-based schemes.

This correct solution behavior differs from that of the CG/Mixed scheme, but once again the non-locking results for the DG-based algorithms are very sensitive with respect to the choice of penalty parameters. Suitable choices are found to be

$$\begin{aligned} SIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [0, 10^{-5}]; \\ IIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [10^{-10}, 10^{-4}]; \\ NIPG: \quad & \delta,\ \tilde{\delta} \quad \in \quad [0, 10^{-4}]. \end{aligned}$$

Values for the *IIPG*-based scheme smaller that $1e-10$ produce inaccurate results. For each scheme, larger values than the above ranges produce effectively continuous solutions, and thus experience locking if full quadrature for the flow variables is used. Again, the pressure oscillations do not persist in time, just as for the solution produced by the CG/Mixed method. As with the CG/Mixed scheme, if reduced quadrature for the flow variables is used no locking occurs. But this seems to be peculiar to Mandel's problem and not a general rule.

7.3 Discussion

For a strictly positive constrained specific storage coefficient value, the DG/Mixed method produces optimal convergence rates with respect to regularity. Furthermore, the quality of the approximation is very robust in regard to the selection of penalty parameters. However, the results are identical

to those produced by the CG/Mixed method but are more computationally expensive and difficult to implement[2]. It seems the only advantages of the DG/Mixed method in this case are that it is more capable for use with a non-conforming grid and that it provides element-wise momentum conservation.

The principal motivation of the DG/Mixed method was its potential to eliminate locking for those problems with a null constrained specific storage coefficient value. The hope was that the the extra degrees of freedom on each element would allow the DG/Mixed scheme using discontinuous linear elements to approximate a non-constant divergence-free vector. However, unlike the DG algorithm for linear elasticity [Rivière and Wheeler, 2000] large penalty parameters are sometimes required for stability and this effectively eliminates the extra degrees of freedom. The result is an effectively continuous scheme which suffers from locking.

Nevertheless, for most of the test problems some success is achieved in overcoming locking. For Barry and Mercer's problem each of the DG/Mixed schemes is able to prevent locking with fairly precisely chosen penalty parameters. The small range of acceptable parameters is a byproduct of the small permeability. For Mandel's problem posed on triangles each of the DG/Mixed schemes is able to prevent locking for penalty parameters within a small interval. The smallness of the range of acceptable values stems from the fact that a large deformation (which indicates that this problem with a small Young's modulus should not be modeled with the equations of linear poroelasticity). For rectangles there are no penalty parameters which eliminate the locking.

[2]The displacement degrees of freedom: DG/Mixed – 8× #elements; CG/Mixed – 6× #grid points. In addition, the several additional terms in the DG/Mixed algorithm add to the flop count.

For homogeneous physical parameters, one can visually examine the pressure solution to detect the checkerboard pattern which so often indicates locking. Adjusting the value of the penalty parameters might then lead to an appropriate choice. But for heterogeneous physical parameters, the task is more difficult as it might be difficult to discern whether or not pressure fluctuations are due to locking or a true reflection of the underlying physics. The examples is this chapter use constant penalty parameters for each test problem. For heterogeneous media, one might try varying the penalty values for each element edge. But a catch-22 situation ensues: those locking-prone elements which would benefit from extra degrees of freedom are the same elements which would require the larger penalty parameters.

The other problem affecting the CG/Mixed method is the degraded convergence rates due to a lack of regularity in the true solution, and the DG/Mixed method offers no help in this regard. Additionally, it was shown that the spurious pressure oscillations produced by the CG/Mixed scheme dissipate with time and lessen as the grid is refined. Therefore, since the sensitivity with respect to penalty parameters makes the DG/Mixed method less than ideal as a means to overcome locking in practice, the greater computational cost might make the discontinuous scheme a less attractive option.

Chapter 8

A Posteriori Error Estimates for the CG/Mixed Scheme

The purpose of this chapter is to develop a grid refinement strategy which builds on theoretical *a posteriori* error estimates developed herein. One goal of the grid refinement strategy is to use computable residual errors as an indicator for refining those areas of the grid which are responsible for producing the largest contribution to the finite element error. It is hoped that allocation of computational resources in those areas of the grid might lead to substantial improvements in accuracy and efficiency. A residual benefit of an adaptive grid based algorithm, as suggested in the chapter on locking, is the potential to remove the problematic non-physical pressure oscillation which sometimes occur.

The first *a posteriori* estimate is based on a semi-formal adjoint problem defined below. The estimate of the displacement finite element error is measured in the $L^2(H^1)$ norm. The second *a posteriori* estimate is based on more standard duality techniques. For this situation, the displacement finite element error is measured in the $L^\infty(L^2)$ norm. For both sets of *a posteriori* estimates, the pressure finite element error is measured in the standard $L^2(L^2)$ norm. For simplicity of implementation, only triangles and tetrahedra are considered.

The theoretical work included here covers several areas which are inde-

pendently discussed elsewhere. A good overall reference for time-independent methods is found in Ainsworth and Oden [1997]; work related specifically to the mixed method can be found in Kirby [2003], Wheeler and Yotov [2003]; also some work involving time-dependent transport problems can be found in Sun [2003].

8.1 Notation

The CG/Mixed finite element error to be analyzed is

$$e_{\mathbf{u}} = \mathbf{u} - \bar{\mathbf{u}}, \qquad e_p = p - \bar{p}, \qquad e_{\mathbf{z}} = \mathbf{z} - \bar{\mathbf{z}},$$

where $(\mathbf{u}, p, \mathbf{z})$ is the solution to (3.13)-(3.15), and $(\bar{\mathbf{u}}, \bar{p}, \bar{\mathbf{z}})$ is the finite element solution to (3.16)-(3.18).

8.1.1 Residual Quantities

In order to establish explicit *a posteriori* error estimates, the following residual quantities are defined:

$$\begin{aligned}
\mathbf{R}_{mom} &= \mathbf{f} + \nabla \cdot \boldsymbol{\sigma}(\bar{\mathbf{u}}) - \alpha \nabla \bar{p}, \qquad \text{in} \quad \Omega, \\
\mathbf{R}_{tr} &= t_N - \tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \boldsymbol{\nu}_a, \qquad \text{on} \quad \Gamma_t, \\
R_{int} &= [\tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p})], \qquad \text{on} \quad \Gamma_{int}, \\
R_{mass} &= s_f - c_o \frac{\partial}{\partial t} \bar{p} - \alpha \frac{\partial}{\partial t} \nabla \cdot \bar{\mathbf{u}} - \nabla \cdot \bar{\mathbf{z}}, \qquad \text{in} \quad \Omega, \\
\mathbf{R}_{flux} &= \tilde{\boldsymbol{\kappa}}^{-1} \bar{\mathbf{z}} + \nabla \bar{p} - \rho_f \mathbf{g}, \qquad \text{in} \quad \Omega, \\
R_{pres} &= p_o - \bar{p}, \qquad \text{on} \quad \Gamma_o.
\end{aligned}$$

Importantly, each residual term is computable elementwise or edgewise directly from the finite element solution and the problem data. These terms will form the basis for estimating the numerical error and determining the areas of the grid in need of refinement.

8.2 The Semi-Formal Adjoint and *A Posteriori* Error Estimate

One ingredient in the development of *a posteriori* error estimates is the existence and regularity of an adjoint-type problem. A simple variational problem involves a bilinear form $b(\cdot,\cdot)$ and the problem of finding a function u in an appropriate space such that for all suitable test functions v, $b(u,v) = l(v)$. The adjoint equation is discovered by substituting $v \to \phi$ and $u \to v$ in the bilinear form. The adjoint solution, ϕ, is found by solving:

$$b(v,\phi) = l(v).$$

8.2.1 The Semi-Formal Adjoint Equation and Solution

The nature of the infinite dimensional variational formulation underlying the CG/Mixed algorithm makes finding the adjoint slightly more complicated. To do so, first sum the three infinite dimensional variational equations (3.13)-(3.15) to find

$$
\begin{aligned}
a_{\mathbf{u}}(\mathbf{u},\mathbf{v}) - \alpha(\nabla\cdot\mathbf{v},p) \\
+\Big(c_o\frac{\partial p}{\partial t},w\Big) + \alpha\Big(\frac{\partial}{\partial t}\nabla\cdot\mathbf{u},w\Big) + (\nabla\cdot\mathbf{z},w) \\
+(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}\mathbf{z},\mathbf{s}) - (p,\nabla\cdot\mathbf{s}) \\
= l_1(\mathbf{v}) + l_2(w) + l_3(\mathbf{s}).
\end{aligned} \tag{8.1}
$$

This equation holds for each triplet $(\mathbf{v},w,\mathbf{s}) \in (H^1)^d \times L^2 \times H(div)$. Note that equation (3.13) is recovered by restricting to the test function triplet $(\mathbf{v},0,0)$, equation (3.14) by the restriction to $(0,w,0)$ and equation (3.15) by the restriction to $(0,0,\mathbf{s})$.

Then, like the simple example shown in the beginning of this subsection, make the following substitutions in the left side of (8.1)

$$
\begin{aligned}
\mathbf{v} &\rightarrow \Phi_{\mathbf{u}}, & \mathbf{u} &\rightarrow \mathbf{v}, \\
w &\rightarrow \Phi_p, & p &\rightarrow w, \\
\mathbf{s} &\rightarrow \Phi_{\mathbf{z}}, & \mathbf{z} &\rightarrow \mathbf{s},
\end{aligned}
$$

to find

$$
\begin{aligned}
a_{\mathbf{u}}(\mathbf{v},\Phi_{\mathbf{u}}) - \alpha(\nabla\cdot\Phi_{\mathbf{u}},w) \\
+\Big(c_o\frac{\partial w}{\partial t},\Phi_p\Big) + \alpha\Big(\frac{\partial}{\partial t}\nabla\cdot\mathbf{v},\Phi_p\Big) + (\nabla\cdot\mathbf{s},\Phi_p) \\
+(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}\mathbf{s},\Phi_{\mathbf{z}}) - (w,\nabla\cdot\Phi_{\mathbf{z}}) \\
= l_1(\mathbf{v}) + l_2(w) + l_3(\mathbf{s}).
\end{aligned} \tag{8.2}
$$

Next, use the assumed symmetry of $\tilde{\boldsymbol{\kappa}}^{-1}$ and take the formal adjoint in time for the terms involving time derivatives to find

$$\begin{aligned} a_{\mathbf{u}}(\mathbf{v}, \Phi_{\mathbf{u}}) - \alpha(\nabla \cdot \Phi_{\mathbf{u}}, w) \\ -\Big(c_o w, \frac{\partial \Phi_p}{\partial t}\Big) - \alpha\Big(\nabla \cdot \mathbf{v}, \frac{\partial \Phi_p}{\partial t}\Big) + (\nabla \cdot \mathbf{s}, \Phi_p) \\ +(\mathbf{s}, \tilde{\boldsymbol{\kappa}}^{-1}\Phi_{\mathbf{z}}) - (w, \nabla \cdot \Phi_{\mathbf{z}}) \\ = l_1(\mathbf{v}) + l_2(w) + l_3(\mathbf{s}). \end{aligned} \tag{8.3}$$

Make the substitutions $l_1(\mathbf{v}) \rightarrow (e_{\mathbf{u}}, \mathbf{v})_{H^1}$, $l_2(w) \rightarrow (e_p, w)$ and $l_3(\mathbf{s}) \rightarrow (e_{\mathbf{z}}, \mathbf{s})$. Then, restricting to the test functions $(\mathbf{v}, 0, 0)$, $(0, w, 0)$ and $(0, 0, \mathbf{s})$ leads to the following semi-formal adjoint equations for $\Phi_{\mathbf{u}}, \Phi_p$ and $\Phi_{\mathbf{z}}$:

$$a_{\mathbf{u}}(\mathbf{v}, \Phi_{\mathbf{u}}) - \alpha\Big(\nabla \cdot \mathbf{v}, \frac{\partial \Phi_p}{\partial t}\Big) = (e_{\mathbf{u}}, \mathbf{v})_{H^1}, \tag{8.4}$$

$$-\Big(c_o w, \frac{\partial \Phi_p}{\partial t}\Big) - \alpha(w, \nabla \cdot \Phi_{\mathbf{u}}) - (w, \nabla \cdot \Phi_{\mathbf{z}}) = (e_p, w), \tag{8.5}$$

$$(\mathbf{s}, \tilde{\boldsymbol{\kappa}}^{-1}\Phi_{\mathbf{z}}) + (\nabla \cdot \mathbf{s}, \Phi_p) = (e_{\mathbf{z}}, \mathbf{s}), \tag{8.6}$$

for appropriate test functions $\mathbf{v}$, w, and $\mathbf{s}$, respectively. In addition, $\Phi_{\mathbf{u}}$, Φ_p and $\Phi_{\mathbf{z}}$ are to satisfy null Dirichlet boundary conditions for the adjoint pressure and adjoint displacement variables and since the semi-formal adjoint equations are backward in time, a null *final* time condition, $\Phi_p|_{t=T} = 0$, is imposed.

Note that the semi-formal adjoint equations are set in infinite dimensions, but its finite dimensional approximation can be found using finite element techniques. If the substitution $t \rightarrow (T - t)$ is made, the time-derivatives are approximated with a backward Euler scheme and (8.4)-(8.6) is approximated using the spaces from the fully coupled CG/Mixed scheme, the resulting

finite dimensional matrix is exactly the transpose of the matrix found in the CG/Mixed scheme, (3.16)-(3.18). This motivates the definition of the semi-formal adjoint equation since the adjoint of a real finite dimensional matrix is its transpose.

The semi-formal adjoint solution is assumed to exist and satisfy the following regularity estimate

$$\begin{aligned} &||\Phi_{\mathbf{u}}||_{L^2(0,T;H^2)} + ||\Phi_p||_{L^2(0,T;H^2)} + ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)} \\ &\quad \leq \quad C\left(||e_{\mathbf{u}}||_{L^2(0,T;H^1)} + ||e_p||_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||_{L^2(0,T;L^2)} \right). \end{aligned} \tag{8.7}$$

Remark 8.2.1. There is nothing in the mathematical literature regarding the regularity of the semi-formal adjoint solution. Even Showalter's theorems on the regularity of the original linear poroelasticity problem are not as complete as that given in (8.7); his bounds only involve the spatial norms for the pressure and displacement at each time, and do not incorporate norms of the data.

In most of the problems of linear poroelasticity discussed earlier it was seen that $p \notin L^2(H^2)$, so if the properties of the adjoint solution bear any resemblance to to the properties of the original solution, (8.7) may be slightly ambitious. However, in those problems an instantaneous traction or point-source may have been the source of degraded regularity. That is not present in the adjoint formulation.

Another point is worth mentioning. Most theorems involving *a posteriori* error estimates either assume or, indeed, provide a theorem establishing sufficient regularity to complete a proof. However, the essential value of most grid refinement techniques derived from such theorems is to improve convergence rates on those problems that *lack* the same regularity that was assumed

in the theorem. A case in point is the aforementioned paper by Wheeler and Yotov [2003]. The authors assume the elliptic problem under consideration to be H^2-regular. This helps to facilitate the proof of their theoretical work, but their most striking computational results occur for a problem whose solution is in $H^{1+\alpha}$ for some $\alpha \in (0,1)$.

So while the following theorem is contingent on the preceding assumptions regarding the existence and regularity of a semi-formal adjoint solution, the assumptions are neither unreasonable nor overarching, and are posed well within the tradition of *a posteriori* error estimates. Ultimately, however, the adaptive grid techniques that follow from the assumptions and theorem will prove to be quite impressive in a series of computational tests, and this, in the absence of mathematical theory, imparts the sense that the assumptions are useful.

8.2.2 *A Posteriori* Error Estimate

Theorem 8.2.1 (*A Posteriori* Error Estimate)**.** *Assuming (8.7), the following computable bound holds for the finite element error incurring in the CG/Mixed scheme for linear poroelasticity for triangles or tetrahedra:*

$$
\begin{aligned}
&||e_{\mathbf{u}}||_{L^2(0,T;H^1)} + ||e_p||_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||_{L^2(0,T;L^2)} \\
&\quad \leq \quad C \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\},
\end{aligned}
$$

where

$$
\begin{aligned}
\eta^2_{mom,E} &= h_E^4 ||\mathbf{R}_{mom}||^2_{L^2(0,T;L^2(E))} \\
&\quad + \sum_{e_a \in \Gamma_t \cap E} h_{e_a}^3 ||\mathbf{R}_{tr}||^2_{L^2(0,T;L^2(e_a))} \\
&\quad + \frac{1}{2} \sum_{e_a \in \Gamma_{int} \cap E} \hat{h}_{e_a}^3 |||[R_{int}]|||^2_{L^2(0,T;L^2(e_a))}, \\
\eta^2_{mass,E} &= h_E^4 ||R_{mass}||^2_{L^2(0,T;L^2(E))}, \\
\eta^2_{flux,E} &= h_E^4 ||\mathbf{R}_{flux}||^2_{L^2(0,T;L^2(E))} + \sum_{e_a \in \Gamma_o \cap E} h_{e_a}^3 ||R_{pres}||^2_{L^2(0,T;L^2(e_a))},
\end{aligned}
$$

and $\hat{h}_{e_a} = \max(h_{E_i}, h_{E_j})$ *where* $e_a = E_i \cap E_j$.

Proof. The error equation for the momentum variational equation satisfies

$$
\begin{aligned}
& a_{\mathbf{u}}(e_{\mathbf{u}}, \mathbf{v}) - \alpha(e_p, \nabla \cdot \mathbf{v}) \\
&\quad = \; l_1(\mathbf{v}) - a_{\mathbf{u}}(\bar{\mathbf{u}}, \mathbf{v}) + \alpha(\bar{p}, \nabla \cdot \mathbf{v}) \\
&\quad = \; \sum_{E \in \mathcal{E}} \int_E \mathbf{f} \cdot \mathbf{v} \; dE + \sum_{e_a \in \Gamma_t} \int_{e_a} \mathbf{t}_N \cdot \mathbf{v} \; ds - \sum_{E \in \mathcal{E}} \int_E \tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) : \nabla \mathbf{v} \; dE \\
&\quad = \; \sum_{E \in \mathcal{E}} \int_E \mathbf{f} \cdot \mathbf{v} \; dE + \sum_{e_a \in \Gamma_t} \int_{e_a} \mathbf{t}_N \cdot \mathbf{v} \; ds \\
&\qquad + \sum_{E \in \mathcal{E}} \int_E (\nabla \cdot \tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p})) \cdot \mathbf{v} \; dE - \sum_{E \in \mathcal{E}} \int_{\partial E} (\tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \boldsymbol{\nu}_a) \cdot \mathbf{v} \; ds \\
&\quad = \; \sum_{E \in \mathcal{E}} \int_E \big(\mathbf{f} + \nabla \cdot \boldsymbol{\sigma}(\bar{\mathbf{u}}) - \alpha \nabla \bar{p} \big) \cdot \mathbf{v} \; dE \\
&\qquad + \sum_{e_a \in \Gamma_t} \int_{e_a} \big(t_N - \tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \boldsymbol{\nu}_a \big) \cdot \mathbf{v} \; ds \\
&\qquad - \sum_{e_a \in \Gamma_{int}} \int_{e_a} \big[\tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \big] \boldsymbol{\nu}_a \cdot \mathbf{v} \; ds. \qquad (8.8)
\end{aligned}
$$

for all $\mathbf{v} \in \mathbf{V}$. From Galerkin orthogonality $a_{\mathbf{u}}(e_{\mathbf{u}}, \bar{\mathbf{v}}) - \alpha(e_p, \nabla \cdot \bar{\mathbf{v}}) = 0$ for any $\bar{\mathbf{v}} \in \mathbf{V}_h$. Adding this to (8.8) yields

$$
\begin{aligned}
& a_{\mathbf{u}}(e_{\mathbf{u}}, \mathbf{v}) - \alpha(e_p, \nabla \cdot \mathbf{v}) \\
&= \sum_{E \in \mathcal{E}} \int_E \big(\underbrace{\mathbf{f} + \nabla \cdot \boldsymbol{\sigma}(\bar{\mathbf{u}}) - \alpha \nabla \bar{p}}_{\mathbf{R}_{mom}} \big) \cdot (\mathbf{v} - \bar{\mathbf{v}}) \, dE \\
&\quad + \sum_{e_a \in \Gamma_t} \int_{e_a} \big(\underbrace{t_N - \tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \boldsymbol{\nu}_a}_{\mathbf{R}_{tr}} \big) \cdot (\mathbf{v} - \bar{\mathbf{v}}) \, ds \\
&\quad - \sum_{e_a \in \Gamma_{int}} \int_{e_a} \underbrace{\big[\tilde{\boldsymbol{\sigma}}(\bar{\mathbf{u}}, \bar{p}) \big]}_{R_{int}} \boldsymbol{\nu}_a \cdot (\mathbf{v} - \bar{\mathbf{v}}) \, ds. \qquad (8.9)
\end{aligned}
$$

Now, let $\bar{\Phi}_{\mathbf{u}} \in \mathbf{V}_h$ be any continuous piecewise linear approximate to $\Phi_{\mathbf{u}}$ with optimal element and edge approximation properties (see, e.g., Babuska and Suri [1987] for triangles or tetrahedra). Set $\mathbf{v} = \Phi_{\mathbf{u}}$ and $\bar{\mathbf{v}} = \bar{\Phi}_{\mathbf{u}}$. Then, the following bounds are obtained:

$$
\begin{aligned}
& a_{\mathbf{u}}(e_{\mathbf{u}}, \Phi_{\mathbf{u}}) - \alpha(e_p, \nabla \cdot \Phi_{\mathbf{u}}) \\
&= \sum_{E \in \mathcal{E}} \int_E \mathbf{R}_{mom} \cdot (\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}) \, dE \\
&\quad + \sum_{e_a \in \Gamma_t} \int_{e_a} \mathbf{R}_{tr} \cdot (\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}) \, ds \\
&\quad - \sum_{e_a \in \Gamma_{int}} \int_{e_a} R_{int} \boldsymbol{\nu}_a \cdot (\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}) \, ds \\
&\leq \sum_{E \in \mathcal{E}} ||\mathbf{R}_{mom}||_{0,E} ||\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}||_{0,E} \\
&\quad + \sum_{e_a \in \Gamma_t} ||\mathbf{R}_{tr}||_{0,e_a} ||\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}||_{0,e_a} \\
&\quad + \sum_{e_a \in \Gamma_{int}} ||R_{int}||_{0,e_a} ||\Phi_{\mathbf{u}} - \bar{\Phi}_{\mathbf{u}}||_{0,e_a} \\
&\leq \sum_{E \in \mathcal{E}} h_E^2 ||\mathbf{R}_{mom}||_{0,E} ||\Phi_{\mathbf{u}}||_{2,E} \\
&\quad + \sum_{e_a \in \Gamma_t} h_{e_a}^{3/2} ||\mathbf{R}_{tr}||_{0,e_a} ||\Phi_{\mathbf{u}}||_{2,E} \\
&\quad + \sum_{e_a \in \Gamma_{int}} \hat{h}_{e_a}^{3/2} ||R_{int}||_{0,e_a} ||\Phi_{\mathbf{u}}||_{2,E}. \qquad (8.10)
\end{aligned}
$$

Next, the error equations for flow are similarly handled:

$$
\begin{aligned}
& \Big(c_o \frac{\partial e_p}{\partial t}, w\Big) + \alpha\Big(\frac{\partial}{\partial t} \nabla \cdot e_{\mathbf{u}}, w\Big) + (\nabla \cdot e_{\mathbf{z}}, w) \\
&= l_2(w) - c_o\Big(\frac{\partial}{\partial t} \bar{p}, w\Big) - \alpha\Big(\frac{\partial}{\partial t} \nabla \cdot \bar{\mathbf{u}}, w\Big) - (\nabla \cdot \bar{\mathbf{z}}, w) \\
&= \sum_{E \in \mathcal{E}} \int_E \Big(s_f - c_o \frac{\partial}{\partial t} \bar{p} - \alpha \frac{\partial}{\partial t} \nabla \cdot \bar{\mathbf{u}} - \nabla \cdot \bar{\mathbf{z}}\Big) w \, dE, \qquad (8.11)
\end{aligned}
$$

$$
\begin{aligned}
&(\tilde{\boldsymbol{\kappa}}^{-1} e_{\mathbf{z}}, \mathbf{s}) - (e_p, \nabla \cdot \mathbf{s}) \\
&\quad = \; l_3(\mathbf{s}) - (\tilde{\boldsymbol{\kappa}}^{-1} \bar{\mathbf{z}}, \mathbf{s}) + (\bar{p}, \nabla \cdot \mathbf{s}) \\
&\quad = \; -\sum_{e_a \in \Gamma_o} \int_{e_a} p_o \mathbf{s} \cdot \boldsymbol{\nu}_a \; ds + \sum_{E \in \mathcal{E}} \int_E \rho_f \mathbf{g} \cdot \mathbf{s} \; dE \\
&\qquad -\sum_{E \in \mathcal{E}} \int_E \tilde{\boldsymbol{\kappa}}^{-1} \bar{\mathbf{z}} \cdot \mathbf{s} \; dE + \sum_{E \in \mathcal{E}} \int_E \bar{p} \nabla \cdot \mathbf{s} \; dE \\
&\quad = \; -\sum_{E \in \mathcal{E}} \int_E \left(\tilde{\boldsymbol{\kappa}}^{-1} \bar{\mathbf{z}} + \nabla \bar{p} - \rho_f \mathbf{g} \right) \cdot \mathbf{s} \; dE \\
&\qquad - \sum_{e_a \in \Gamma_{int}} \int_{e_a} [\bar{p}] \mathbf{s} \cdot \boldsymbol{\nu}_a \; ds - \sum_{e_a \in \Gamma_o} \int_{e_a} (p_o - \bar{p}) \mathbf{s} \cdot \boldsymbol{\nu}_a \; ds. \qquad (8.12)
\end{aligned}
$$

The above equations hold for all $(w, \mathbf{s}) \in L^2 \times H(div)$. Again, from Galerkin orthogonality, the flow error equations are each zero for $(\bar{w}, \bar{\mathbf{s}})$ in a suitable finite dimensional mixed space (i.e., RT1). Utilizing this in (8.11) and (8.12), respectively, produces

$$
\begin{aligned}
&\Big(c_o \frac{\partial e_p}{\partial t}, w \Big) + \alpha \Big(\frac{\partial}{\partial t} \nabla \cdot e_{\mathbf{u}}, w \Big) + (\nabla \cdot e_{\mathbf{z}}, w) \\
&\quad = \; \sum_{E \in \mathcal{E}} \int_E \Big(\underbrace{s_f - c_o \frac{\partial}{\partial t} \bar{p} - \alpha \frac{\partial}{\partial t} \nabla \cdot \bar{\mathbf{u}} - \nabla \cdot \bar{\mathbf{z}}}_{R_{mass}} \Big) \cdot (w - \bar{w}) \; dE, \qquad (8.13)
\end{aligned}
$$

$$
\begin{aligned}
&(\tilde{\boldsymbol{\kappa}}^{-1}e_{\mathbf{z}},\mathbf{s})-(e_p,\nabla\cdot\mathbf{s})\\
&\quad= \quad -\sum_{E\in\mathcal{E}}\int_E \big(\underbrace{\tilde{\boldsymbol{\kappa}}^{-1}\bar{\mathbf{z}}+\nabla\bar{p}-\rho_f\mathbf{g}}_{\mathbf{R}_{flux}}\big)\cdot(\mathbf{s}-\bar{\mathbf{s}})\;dE\\
&\qquad\quad -\sum_{e_a\in\Gamma_{int}}\int_{e_a}[\bar{p}](\mathbf{s}-\bar{\mathbf{s}})\cdot\boldsymbol{\nu}_a\;ds\\
&\qquad\quad -\sum_{e_a\in\Gamma_o}\int_{e_a}(\underbrace{p_o-\bar{p}}_{R_{pres}})(\mathbf{s}-\bar{\mathbf{s}})\cdot\boldsymbol{\nu}_a\;ds. \qquad (8.14)
\end{aligned}
$$

Now, using the mixed space operators to set $w=\Phi_p$, $\bar{w}=\bar{\Phi}_p\equiv P_h\Phi_p$, $\mathbf{s}=\Phi_{\mathbf{z}}$, and $\bar{\mathbf{s}}=\bar{\Phi}_{\mathbf{z}}\equiv\Pi_h\Phi_{\mathbf{z}}$ the following two bounds are obtained using, in particular, (3.9) and (3.10),

$$
\begin{aligned}
&\Big(c_o\frac{\partial e_p}{\partial t},\Phi_p\Big)+\alpha\Big(\frac{\partial}{\partial t}\nabla\cdot e_{\mathbf{u}},\Phi_p\Big)+(\nabla\cdot e_{\mathbf{z}},\Phi_p)\\
&\quad= \quad\sum_{E\in\mathcal{E}}\int_E R_{mass}(\Phi_p-\bar{\Phi}_p)\;dE\\
&\quad\leq \quad\sum_{E\in\mathcal{E}}||R_{mass}||_{0,E}||\Phi_p-\bar{\Phi}_p||_{0,E}\\
&\quad\leq \quad\sum_{E\in\mathcal{E}}h_E^2||R_{mass}||_{0,E}||\Phi_p||_{2,E}, \qquad (8.15)
\end{aligned}
$$

$$
\begin{aligned}
&(\tilde{\boldsymbol{\kappa}}^{-1}\mathbf{e_z}, \Phi_{\mathbf{z}}) - (e_p, \nabla \cdot \Phi_{\mathbf{z}}) \\
&= -\sum_{E\in\mathcal{E}} \int_E \mathbf{R}_{flux} \cdot (\Phi_{\mathbf{z}} - \bar{\Phi}_{\mathbf{z}})\, dE \\
&\quad - \sum_{e_a\in\Gamma_{int}} \underbrace{\int_{e_a} [\bar{p}](\Phi_{\mathbf{z}} - \bar{\Phi}_{\mathbf{z}}) \cdot \boldsymbol{\nu}_a\, ds}_{=\ 0\ by\ (3.6c)} \\
&\quad - \sum_{e_a\in\Gamma_o} \int_{e_a} R_{pres}(\Phi_{\mathbf{z}} - \bar{\Phi}_{\mathbf{z}}) \cdot \boldsymbol{\nu}_a\, ds \\
&\leq \sum_{E\in\mathcal{E}} \int_E ||\mathbf{R}_{flux}||_{0,E} ||\Phi_{\mathbf{z}} - \bar{\Phi}_{\mathbf{z}}||_{0,E} \\
&\quad \sum_{e_a\in\Gamma_o} \int_{e_a} ||R_{pres}||_{0,e_a} ||(\Phi_{\mathbf{z}} - \bar{\Phi}_{\mathbf{z}}) \cdot \boldsymbol{\nu}_a||_{0,e_a} \\
&\leq \sum_{E\in\mathcal{E}} \int_E h_E^2 ||\mathbf{R}_{flux}||_{0,E} ||\Phi_{\mathbf{z}}||_{2,E} \\
&\quad \sum_{e_a\in\Gamma_o} \int_{e_a} h_{e_a}^{3/2} ||R_{pres}||_{0,e_a} ||\Phi_{\mathbf{z}}||_{3/2,e_a} \\
&\leq \sum_{E\in\mathcal{E}} \int_E h_E^2 ||\mathbf{R}_{flux}||_{0,E} ||\Phi_{\mathbf{z}}||_{2,E} \\
&\quad \sum_{e_a\in\Gamma_o} \int_{e_a} h_{e_a}^{3/2} ||R_{pres}||_{0,e_a} ||\Phi_{\mathbf{z}}||_{2,E}. \qquad (8.16)
\end{aligned}
$$

Turning attention now to the semi-formal adjoint equations, let $\mathbf{v} = e_{\mathbf{u}}$ in (8.4). Integrating form 0 to T, integrating by parts in time, and using the initial condition (null for $e_{\mathbf{u}}$) and final condition (null for Φ_p), one finds,

$$
||e_{\mathbf{u}}||^2_{L^2(0,T;H^1)} = \int_0^T a_{\mathbf{u}}(e_{\mathbf{u}}, \Phi_{\mathbf{u}}) + \alpha \int_0^T \Big(\frac{\partial \nabla \cdot e_{\mathbf{u}}}{\partial t}, \Phi_p\Big). \qquad (8.17)
$$

Likewise, letting $w = \Phi_p$ in (8.5), integrating form 0 to T, integrating by parts

in time, and using the initial and final conditions, the following holds,

$$\begin{aligned} ||e_p||^2_{L^2(0,T;L^2)} &= \int_0^T \Big(c_o \frac{\partial e_p}{\partial t}, \Phi_p\Big) - \alpha \int_0^T (e_p, \nabla \cdot \Phi_{\mathbf{u}}) \\ &\quad - \int_0^T (e_p, \nabla \cdot \Phi_{\mathbf{z}}). \end{aligned} \tag{8.18}$$

Finally, setting $\mathbf{s} = e_{\mathbf{z}}$ in (8.6), and integrating form 0 to T,

$$\begin{aligned} ||e_{\mathbf{z}}||^2_{L^2(0,T;L^2)} &= \int_0^T (e_{\mathbf{z}}, \tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \Phi_{\mathbf{z}}) + \int_0^T (\nabla \cdot e_{\mathbf{z}}, \Phi_p) \\ &= \int_0^T (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} e_{\mathbf{z}}, \Phi_{\mathbf{z}}) + \int_0^T (\nabla \cdot e_{\mathbf{z}}, \Phi_p). \end{aligned} \tag{8.19}$$

Then, by summing (8.17)-(8.19) and rearranging, one finds

$$\begin{aligned} &||e_{\mathbf{u}}||^2_{L^2(0,T;H^1)} + ||e_p||^2_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||^2_{L^2(0,T;L^2)} \\ &= \int_0^T \Big[a_{\mathbf{u}}(e_{\mathbf{u}}, \Phi_{\mathbf{u}}) - \alpha(e_p, \nabla \cdot \Phi_{\mathbf{u}})\Big] \\ &+ \int_0^T \Big[\Big(c_o \frac{\partial e_p}{\partial t}, \Phi_p\Big) + \alpha\Big(\frac{\partial \nabla \cdot e_{\mathbf{u}}}{\partial t}, \Phi_p\Big) + (\nabla \cdot e_{\mathbf{z}}, \Phi_p)\Big] \\ &+ \int_0^T \Big[(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} e_{\mathbf{z}}, \Phi_{\mathbf{z}}) - (e_p, \nabla \cdot \Phi_{\mathbf{z}})\Big]. \end{aligned} \tag{8.20}$$

The bracketed terms are recognized as the error equations. Substituting their respective bounds found in (8.10) and (8.15)-(8.16) one finds

$$
\begin{aligned}
&||e_{\mathbf{u}}||^2_{L^2(0,T;H^1)} + ||e_p||^2_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||^2_{L^2(0,T;L^2)} \\
&\quad \leq \quad ||\Phi_{\mathbf{u}}||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||\mathbf{R}_{mom}||^2_{L^2(0,T;L^2(E))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_t} h_{e_a}^3 ||\mathbf{R}_{tr}||^2_{L^2(0,T;L^2(e_a))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_{int}} \hat{h}_{e_a}^3 ||R_{int}||^2_{L^2(0,T;L^2(e_a))} \Bigg)^{\frac{1}{2}} \\
&\quad + \quad ||\Phi_p||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||R_{mass}||^2_{L^2(0,T;L^2(E))} \Bigg)^{\frac{1}{2}} \\
&\quad + \quad ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||\mathbf{R}_{flux}||^2_{L^2(0,T;L^2(E))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_o} h_{e_a}^3 ||\mathbf{R}_{tr}||^2_{L^2(0,T;L^2(e_a))} \Bigg)^{\frac{1}{2}}.
\end{aligned}
$$

With this one deduces

$$
\begin{aligned}
&||e_{\mathbf{u}}||^2_{L^2(0,T;H^1)} + ||e_p||^2_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||^2_{L^2(0,T;L^2)} \\
&\quad \leq \quad \Big(||\Phi_{\mathbf{u}}||_{L^2(0,T;H^2)} + ||\Phi_p||_{L^2(0,T;H^2)} + ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)} \Big) \\
&\qquad \times \max \left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\}.
\end{aligned}
$$

Using the assumed regularity of the adjoint solution (8.7), one finds

$$
\begin{aligned}
&||e_{\mathbf{u}}||^2_{L^2(0,T;H^1)} + ||e_p||^2_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||^2_{L^2(0,T;L^2)} \\
&\quad \leq \; C\Big(||e_{\mathbf{u}}||_{L^2(0,T;H^1)} + ||e_p||_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||_{L^2(0,T;L^2)}\Big) \\
&\qquad \times \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\}.
\end{aligned}
$$

By using the equivalence of finite dimensional norms, the above left side can be modified[1] to find

$$
\begin{aligned}
&||e_{\mathbf{u}}||_{L^2(0,T;H^1)} + ||e_p||_{L^2(0,T;L^2)} + ||e_{\mathbf{z}}||_{L^2(0,T;L^2)} \\
&\quad \leq \; C \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\},
\end{aligned}
$$

which completes the proof.

□

8.3 A Duality-Based *A Posteriori* Error Estimate

An alternative *a posteriori* error estimate is derived which is based on duality. This method is more direct than the previous theorem, and it results in an estimate of the displacement in a different norm.

8.3.1 The Dual Problem

To begin, let $\Phi_{\mathbf{u}}$, Φ_p and $\Phi_{\mathbf{z}}$ satisfy the following backward partial differential equation, boundary and final conditions:

[1]For example, $(a^2+b^2+c^2)^{\frac{1}{2}} \geq C(|a|+|b|+|c|)$.

$$
\begin{aligned}
-(\lambda+\mu)\nabla(\nabla\cdot\Phi_{\mathbf{u}}) - \mu\nabla^2\Phi_{\mathbf{u}} + \alpha\nabla\Phi_p &= e_{\mathbf{u}}, && \text{(8.21a)}\\
-c_o\frac{\partial\Phi_p}{\partial t} - \alpha\frac{\partial}{\partial t}\nabla\cdot\Phi_{\mathbf{u}} + \nabla\cdot\Phi_{\mathbf{z}} &= e_p, && \text{(8.21b)}\\
\tilde{\boldsymbol{\kappa}}^{-1}\Phi_{\mathbf{z}} + \nabla\Phi_p &= 0. && \text{(8.21c)}
\end{aligned}
$$

$$
\begin{aligned}
\Phi_p(t) &= 0 \quad \text{on} \quad \Gamma_p, && \text{(8.22a)}\\
-\frac{1}{\mu_f}\boldsymbol{\kappa}(\nabla\Phi_p - \rho_f\mathbf{g})\cdot\boldsymbol{\nu} &= 0 \quad \text{on} \quad \Gamma_f, && \text{(8.22b)}\\
\Phi_{\mathbf{u}}(t) &= 0 \quad \text{on} \quad \Gamma_o, && \text{(8.22c)}\\
\tilde{\boldsymbol{\sigma}}(\Phi_{\mathbf{u}},\Phi_p)\boldsymbol{\nu} &= 0 \quad \text{on} \quad \Gamma_t, && \text{(8.22d)}\\
\Phi_p(T) &= 0, && \text{(8.22e)}\\
\Phi_{\mathbf{u}}(T) &= 0. && \text{(8.22f)}
\end{aligned}
$$

If the substitution $t \rightarrow T - t$ is made, then the dual problem is of the same form as that for the equations of linear poroelasticity. Furthermore, since the null boundary and initial conditions in (8.22a)-(8.22f) are compatible with Showalter's boundary and initial conditions, the existence and uniqueness of the dual solution hold. In addition, the dual solution is assumed to bounded in terms of the data as follows:

$$
\begin{aligned}
&||\Phi_{\mathbf{u}}||_{L^2(0,T;H^2)} + ||\Phi_{\mathbf{u}t}||_{L^2(0,T;H^2)} + ||\Phi_p||_{L^2(0,T;H^2)} + ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)}\\
&\qquad\qquad \leq C\left(||e_{\mathbf{u}}||_{L^\infty(0,T;L^2)} + ||e_p||_{L^2(0,T;L^2)} \right). \qquad \text{(8.23)}
\end{aligned}
$$

The same remarks regarding regularity following the semi-formal adjoint solution hold here: (8.23) may be slightly ambitious, but given the absence of more complete mathematical results, its utility is that it helps complete the subsequent theorem on *a posteriori* error estimates, and is ultimately justified to an extent by computational results.

8.3.2 *A Posteriori* Error Estimate

Theorem 8.3.1 (*A Posteriori* Error Estimate)**.** *Assuming (8.23), the following computable bound holds for the finite element error incurring in the CG/Mixed scheme for linear poroelasticity for triangles or tetrahedra:*

$$
\begin{aligned}
&||e_{\mathbf{u}}||_{L^\infty(0,T;L^2)} + ||e_p||_{L^2(0,T;L^2)} \\
&\quad \leq \; C \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\},
\end{aligned}
$$

where η_{mom}, η_{mass} and η_{flux} are the same as in the previous theorem.

Proof. First differentiate with respect to time equation (8.21a), and then multiply (8.21a)-(8.21c) by $\mathbf{v}$, w and $\mathbf{s}$, respectively[2], and integrate over Ω to find

$$
\begin{aligned}
-(\nabla \cdot \tilde{\boldsymbol{\sigma}}(\Phi_{\mathbf{u}t}, \Phi_{p_t}), \mathbf{v}) &= \left(\frac{\partial e_{\mathbf{u}}}{\partial t}, \mathbf{v} \right), \\
-\left(c_o \frac{\partial \Phi_p}{\partial t}, w \right) - \alpha \left(\frac{\partial}{\partial t} \nabla \cdot \Phi_{\mathbf{u}}, w \right) + (\nabla \cdot \Phi_{\mathbf{z}}, w) &= (e_p, w), \\
(\tilde{\boldsymbol{\kappa}}^{-1} \Phi_{\mathbf{z}}, \mathbf{s}) + (\nabla \Phi_p, \mathbf{s}) &= 0.
\end{aligned}
$$

[2]Each compatible with the boundary conditions.

Next, integrate by parts in space and use the chain rule from calculus (on selected terms) to find

$$\begin{aligned} a_{\mathbf{u}}(\Phi_{\mathbf{u}t}, \mathbf{v}) + \alpha(\Phi_p, \nabla \cdot \mathbf{v}_t) - \alpha \frac{\partial}{\partial t}(\Phi_p, \nabla \cdot \mathbf{v}) &= \left(\frac{\partial e_{\mathbf{u}}}{\partial t}, \mathbf{v}\right), \\ \left(c_o \Phi_p, \frac{\partial w}{\partial t}\right) - \frac{\partial}{\partial t}(c_o \Phi_p, w) & \\ -\alpha\left(\frac{\partial}{\partial t}\nabla \cdot \Phi_{\mathbf{u}}, w\right) + (\nabla \cdot \Phi_{\mathbf{z}}, w) &= (e_p, w), \\ (\tilde{\boldsymbol{\kappa}}^{-1}\Phi_{\mathbf{z}}, \mathbf{s}) - (\Phi_p, \nabla \cdot \mathbf{s}) &= 0. \end{aligned}$$

Now, set $\mathbf{v} = e_{\mathbf{u}}$, $w = e_p$ and $\mathbf{s} = e_{\mathbf{z}}$ to arrive at

$$\begin{aligned} a_{\mathbf{u}}(\Phi_{\mathbf{u}t}, e_{\mathbf{u}}) + \alpha\left(\Phi_p, \frac{\partial}{\partial t}\nabla \cdot e_{\mathbf{u}}\right) - \alpha \frac{\partial}{\partial t}(\Phi_p, \nabla \cdot e_{\mathbf{u}}) &= \frac{1}{2}\frac{\partial}{\partial t}||e_{\mathbf{u}}||_0^2, \quad (8.24) \\ \left(c_o \Phi_p, \frac{\partial e_p}{\partial t}\right) - \frac{\partial}{\partial t}(c_o \Phi_p, e_p) & \\ -\alpha\left(\frac{\partial}{\partial t}\nabla \cdot \Phi_{\mathbf{u}}, e_p\right) + (\nabla \cdot \Phi_{\mathbf{z}}, e_p) &= ||e_p||_0^2, \quad (8.25) \\ (\tilde{\boldsymbol{\kappa}}^{-1}\Phi_{\mathbf{z}}, e_{\mathbf{z}}) - (\Phi_p, \nabla \cdot e_{\mathbf{z}}) &= 0. \quad (8.26) \end{aligned}$$

Next, integrate in time from 0 to T, and use the null initial and final conditions. Then, sum the equations resulting from (8.24)-(8.25) and subtract the result for (8.26). After grouping the appropriate terms, the equation follows:

$$
\begin{aligned}
&\frac{1}{2}||e_{\mathbf{u}}(T)||_0^2 + \int_0^T ||e_p||_0^2 \\
&= \int_0^T \left[a_{\mathbf{u}}(\Phi_{\mathbf{u}t}, e_{\mathbf{u}}) - \alpha(\nabla \cdot \Phi_{\mathbf{u}t}, e_p) \right] \\
&\quad + \int_0^T \left[\left(c_o \Phi_p, \frac{\partial e_p}{\partial t} \right) + \alpha \left(\Phi_p, \frac{\partial \nabla \cdot e_{\mathbf{u}}}{\partial t} \right) + (\Phi_p, \nabla \cdot e_{\mathbf{z}}) \right] \\
&\quad - \int_0^T \left[(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \Phi_{\mathbf{z}}, e_{\mathbf{z}}) - (\nabla \cdot \Phi_{\mathbf{z}}, e_p) \right].
\end{aligned}
$$

Then, using the symmetry of $a_{\mathbf{u}}$, $\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}$ and the L^2 inner product, one finds

$$
\begin{aligned}
&\frac{1}{2}||e_{\mathbf{u}}(T)||_0^2 + \int_0^T ||e_p||_0^2 \\
&= \int_0^T \left[a_{\mathbf{u}}(e_{\mathbf{u}}, \Phi_{\mathbf{u}t}) - \alpha(e_p, \nabla \cdot \Phi_{\mathbf{u}t}) \right] \\
&\quad + \int_0^T \left[\left(c_o \frac{\partial e_p}{\partial t}, \Phi_p \right) + \alpha \left(\frac{\partial \nabla \cdot e_{\mathbf{u}}}{\partial t}, \Phi_p \right) + (\nabla \cdot e_{\mathbf{z}}, \Phi_p) \right] \\
&\quad - \int_0^T \left[(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} e_{\mathbf{z}}, \Phi_{\mathbf{z}}) - (e_p, \nabla \cdot \Phi_{\mathbf{z}}) \right]. \qquad (8.27)
\end{aligned}
$$

The right side of the above equality is the same as (8.20) from the previous section with the exception that $\Phi_{\mathbf{u}}$ is replaced with its time derivative. Therefore, using the same reasoning as that following (8.20), one deduces the following inequality for (8.27):

$$
\begin{aligned}
&||e_{\mathbf{u}}(T)||_0^2 + ||e_p||^2_{L^2(0,T;L^2)} \\
&\quad\leq \quad ||\Phi_{\mathbf{u}t}||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||\mathbf{R}_{mom}||^2_{L^2(0,T;L^2(E))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_t} h_{e_a}^3 ||\mathbf{R}_{tr}||^2_{L^2(0,T;L^2(e_a))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_{int}} \hat{h}_{e_a}^3 ||R_{int}||^2_{L^2(0,T;L^2(e_a))} \Bigg)^{\frac{1}{2}} \\
&\quad + \quad ||\Phi_p||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||R_{mass}||^2_{L^2(0,T;L^2(E))} \Bigg)^{\frac{1}{2}} \\
&\quad + \quad ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)} \Bigg(\sum_{E\in\mathcal{E}} h_E^4 ||\mathbf{R}_{flux}||^2_{L^2(0,T;L^2(E))} \\
&\qquad\qquad + \sum_{e_a\in\Gamma_o} h_{e_a}^3 ||\mathbf{R}_{tr}||^2_{L^2(0,T;L^2(e_a))} \Bigg)^{\frac{1}{2}}.
\end{aligned}
$$

With this one deduces

$$
\begin{aligned}
&||e_{\mathbf{u}}(T)||_0^2 + ||e_p||^2_{L^2(0,T;L^2)} \\
&\quad\leq \quad \Big(||\Phi_{\mathbf{u}t}||_{L^2(0,T;H^2)} + ||\Phi_p||_{L^2(0,T;H^2)} + ||\Phi_{\mathbf{z}}||_{L^2(0,T;H^2)} \Big) \\
&\qquad \times \max \left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\}.
\end{aligned}
$$

Using the assumed regularity of the dual solution (8.23), one finds

$$
\begin{aligned}
&||e_{\mathbf{u}}(T)||_0^2 + ||e_p||^2_{L^2(0,T;L^2)} \\
&\quad \leq \; C\Big(||e_{\mathbf{u}}||_{L^\infty(0,T;L^2)} + ||e_p||_{L^2(0,T;L^2)}\Big) \\
&\qquad \times \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\}.
\end{aligned}
$$

Since the above inequality is true for all times T, the equivalence of finite dimensional norms allows one to deduce

$$
\begin{aligned}
&||e_{\mathbf{u}}||_{L^\infty(0,T;L^2)} + ||e_p||_{L^2(0,T;L^2)} \\
&\quad \leq \; C \max\left\{ \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mom,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{mass,E} \right)^{\frac{1}{2}}, \left(\sum_{E\in\mathcal{E}} \int_E \eta^2_{flux,E} \right)^{\frac{1}{2}} \right\},
\end{aligned}
$$

which completes the proof.

□

Remark 8.3.1. Note that the above proof could be easily modified in order to bound the flux in the $L^2(L^2)$ norm. To do so, simply make the substitution $e_p \to 0$ in (8.21b) and $0 \to e_{\mathbf{z}}$ in (8.21c), and consider a modified regularity assumption.

8.4 Grid Refinement Strategy

For simplicity, the implementation suggested here uses only triangles on conforming meshes. To maintain a conforming mesh, each element marked for refinement in bisected along the longest edge *as well as* the adjacent element sharing the longest edge. Figure 8.1 shows an example of this approach.

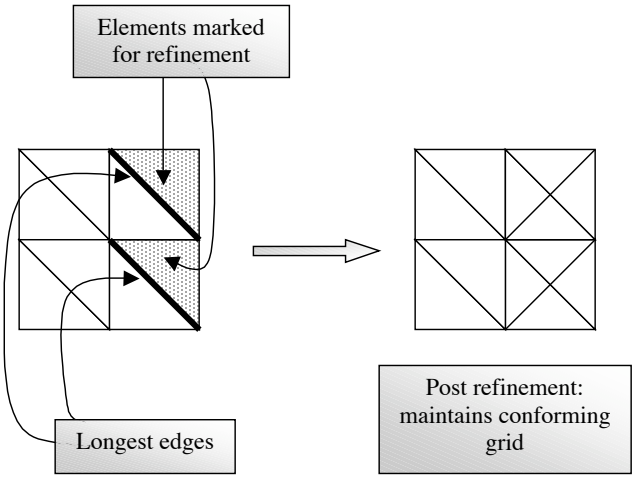

Figure 8.1: The grid refinement used here maintains a conforming grid through bisection of the longest edge of elements marked for refinement.

The basic strategy for adaptive refinement places emphasis on the computable residual quantities defined above and is summarized below:

1. Run the CG/Mixed algorithm;
2. Compute the quantities $\eta^2_{mom,E},\ \eta^2_{mass,E}$ and $\eta^2_{flux,E}$ for each element E;
3. Set $\eta^2_{max} = \max\left\{ \max_{E\in\mathcal{E}} \eta^2_{mom,E},\ \max_{E\in\mathcal{E}} \eta^2_{mass,E},\ \max_{E\in\mathcal{E}} \eta^2_{flux,E} \right\}$;
4. Mark for refinement all elements E_a such that:

$$\max\{\eta^2_{mom,E_a},\ \eta^2_{mass,E_a},\ \eta^2_{flux,E_a}\} \geq \gamma \eta^2_{max},$$

   for some *refinement factor* $\gamma \in (0,1)$;
5. Refine mesh in accordance with the marked elements, but maintain a conforming mesh;

6. Repeat (1) until all error indicators are below a given tolerance; that is, until $\eta_{max} \leq \epsilon$ for some prescribed, positive ϵ.

The refinement factor, γ, determines the aggressiveness of the refinement at each iteration. At one extreme, $\gamma = 0$ implies a uniform refinement at each stage. On the other hand, as γ becomes larger, fewer elements are marked for refinement. For every $\gamma \in (0, 1)$, η_{max} is likely to decrease with each iteration, so termination is probable for reasonable tolerances. However, there is the possibility that the error indicators may reach a lower limit due to time pollution error. Therefore, it is advisable to limit the number of iterations to a reasonable number[3].

[3]One reasonable suggestion is to terminate the iteration when η_{max} ceases to change much from one iteration to the next.

Chapter 9

Adaptive Grid Refinement Strategy: Numerical Results

Two advantages of adaptive grid refinement are demonstrated in this chapter with several numerical examples. The first advantage pertains to the issue of accuracy and efficiency. Two problems with analytical solutions are used to compare the effect of uniform refinement versus adaptive refinement. It is shown that adaptive refinement substantially improves accuracy at a lower computational cost. The second advantage is in reference to the problem of locking in linear poroelasticity. After briefly reexamining the way the problem arises, it is shown that how adaptive grid refinement might help. Two examples demonstrate its effectiveness.

9.1 Improvement in Accuracy Per Computational Cost

It is not unusual for a problem in linear poroelasticity to admit a solution with a pressure boundary layer. For these circumstances significant gains in accuracy might be gained from effective adaptive grid refinement. Both Terzaghi's and Mandel's solutions have been shown in this dissertation to exhibit a pressure boundary layer due to a traction force along a drained boundary. This makes each problem ideal for examining the effectiveness of the grid refinement strategy.

As a means of comparing the performance of uniform versus adaptive grid refinement, the *effective element size* is defined to be

$$h = \frac{1}{(n_{el})^d},$$

where n_{el} is the number of elements used in the finite element program and d is the dimension of the problem. The advantage of this definition is that for uniform refinement the rate of convergence with respect to both the element size and the effective element size is the same.

9.1.1 Terzaghi's Problem

Previously, numerical results for Terzaghi's problem showed less than optimal convergence rates for both the CG/Mixed and DG/Mixed algorithms when coupled to standard uniform grid refinement. In those discussions, it was also seen that since $p \notin L^2(H^2)$ optimal convergence rates could not be expected. From the theoretical error estimates, the order of convergence of the numerical solution ($\approx O(h^{0.500^+})$) provides a strong implication that $p \in L^2(H^{\frac{3}{2}+\epsilon})$ for some $|\epsilon| << 1$. Figure 9.1 is an example of the boundary layers produced as a result of a lack of regularity.

The evident boundary layer which provides difficulties for any scheme with uniform refinement makes Terzaghi's problem well-positioned to make use of adaptive grid refinement. The goal is to use the computable residual quantities to automatically identify only those areas of grid contributing to the error. By allocating computational resources in those area alone, impressive improvements in the rate of convergence might be discovered.

Just as in earlier experiments, two test cases are used to examine Terzaghi's problem. For each test case, a force, $F = 1e3$, is used.

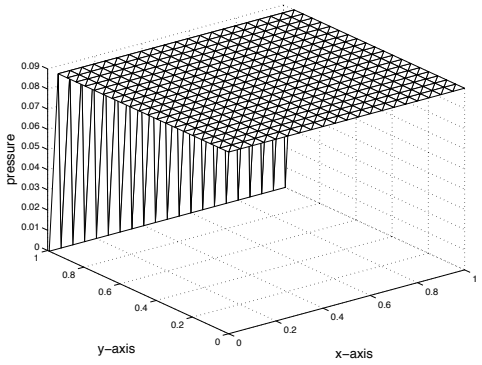

Figure 9.1: Early time analytical pressure solution for Terzaghi's problem and the evident boundary layer.

Test Case I

To remind, the model parameters for this example are the following:

$$E \;=\; 1e+5, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e-1, \quad \boldsymbol{\kappa} = 1e-6.$$

Using a refinement factor of $\gamma = 0.8$, Figure 9.2 shows the improvement in accuracy[1] per cost using adaptive refinement when compared to uniform refinement. The $L^2(L^2)$ pressure error is shown to exhibit exponential convergence. Not surprisingly, the largest residual turns out to be the one involving the flux, and this guides the refinement toward the drained surface.

Test Case II

The model parameters for the second example are the following:

[1]Only the analytical solution for pressure and settlement is given for this problem.

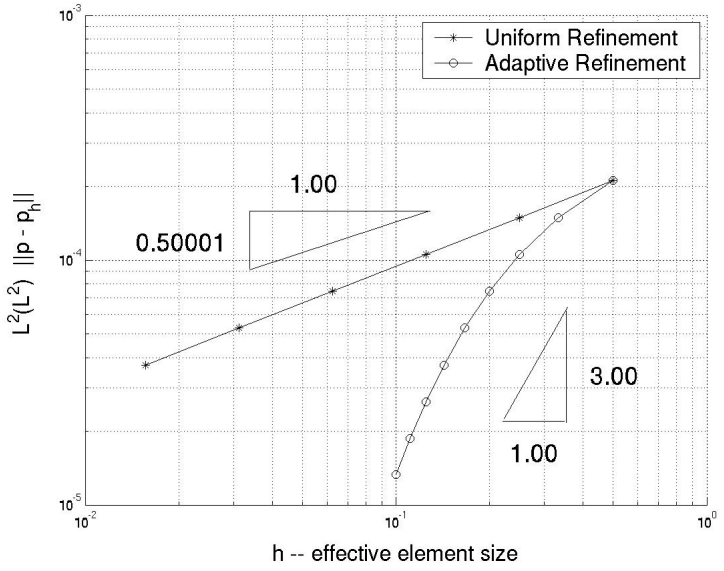

Figure 9.2: Uniform vs. adaptive refinement in Terzaghi's problem. $L^2(L^2)$ pressure error. Physical parameters used: $E = 1e5$; $\nu = 0.2$; $c_o = 1e-1$; $\boldsymbol{\kappa} = 1e-6$.

$$E \;=\; 1e+7, \quad \nu = 0.4, \quad \alpha = 1, \quad c_o = 5e-1, \quad \boldsymbol{\kappa} = 1e-4.$$

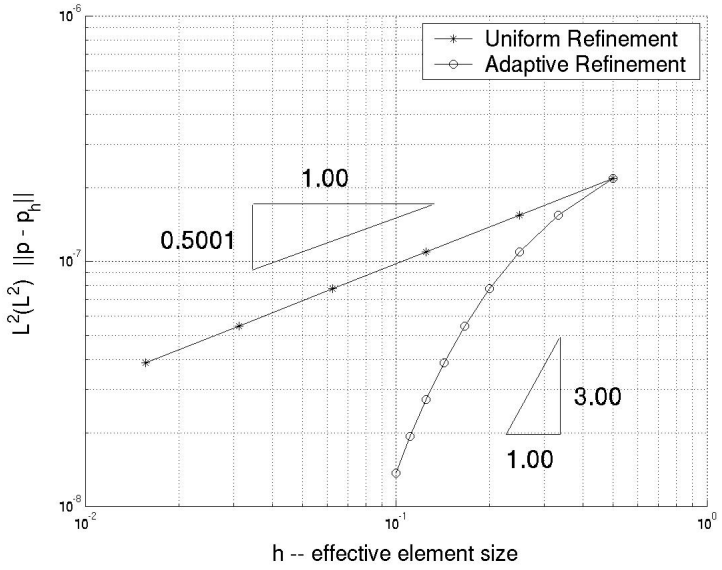

Figure 9.3: Uniform vs. adaptive refinement in Terzaghi's problem. $L^2(L^2)$ pressure error. Physical parameters used: $E = 1e7$; $\nu = 0.4$; $c_o = 5e-1$; $\boldsymbol{\kappa} = 1e-4$.

Figure 9.3 once again shows the improvement in accuracy per cost using adaptive refinement when compared to uniform refinement. The $L^2(L^2)$ pressure error is shown to exhibit exponential convergence, and represents a substantial performance increase over uniform refinement.

9.1.2 Mandel's Problem

Earlier chapters presented numerical results for Mandel's problem which demonstrated less than optimal convergence rates for both the CG/Mixed and

DG/Mixed algorithms when coupled to standard uniform grid refinement. However, it was also seen that since $p \notin L^2(H^2)$ optimal convergence rates could not be expected. In light of the theoretical error estimates, the order of convergence discovered ($\approx O(h^{0.500^+})$) provides a strong implication that $p \in L^2(H^{\frac{3}{2}+\epsilon})$ for some $|\epsilon| << 1$. Figure 9.4 provides a reminder that lack of regularity often manifests in boundary layers.

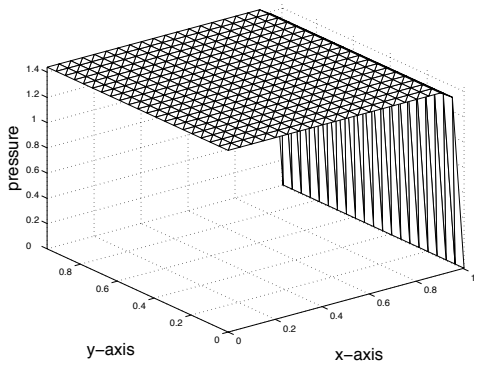

Figure 9.4: Early time analytical pressure solution for Mandel's problem.

However, the same boundary layer which provide difficulties for any scheme with uniform refinement makes Mandel's problem an ideal candidate for adaptive grid refinement. The goal is to use the computable residual quantities to automatically identify only those areas of grid contributing to the error. By allocating computational resources in those area alone, impressive improvements in the rate of convergence might be discovered.

As a reminder, the same choice of model parameters used earlier is employed to test the adaptive grid method:

$$E \;=\; 1e+4, \quad \nu = 0.2, \quad \alpha = 1, \quad c_o = 1e-1, \quad \boldsymbol{\kappa} = 1e-2.$$

Additionally, a force, $F = 2e3$, is used.

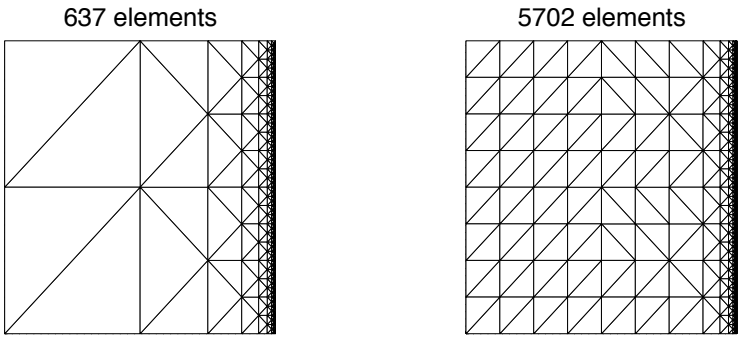

Figure 9.5: Different stages of refinement in Mandel's problem.

Figure 9.5 shows the nature of the refinement progression when using a refinement factor, $\gamma = 0.8$. Early in the procedure (left graph), the error in the flux predominates, and most refinement occurs on the right edge. For later iterations, the flux error indicator becomes smaller and begins to compare to the remaining error indicators. So although the principal area of refinement continues to be along the axis $x = 1$, there is some additional refinement within the interior.

Using the adaptive grid strategy, Figure 9.6 shows that both the $L^2(L^2)$ pressure error (top) and the $L^2(H^1)$ displacement error (bottom) exhibit an exponential rate of convergence. This represents a significant improvement over the results from a uniform refinement strategy. Figure 9.7 shows the improvement of the displacement error measured in the $L^\infty(L^2)$ norm. Although no error estimates were produced for uniformly refined grids in this norm, the duality-based *a posteriori* error estimates suggest the improvement in convergence seen in the figure.

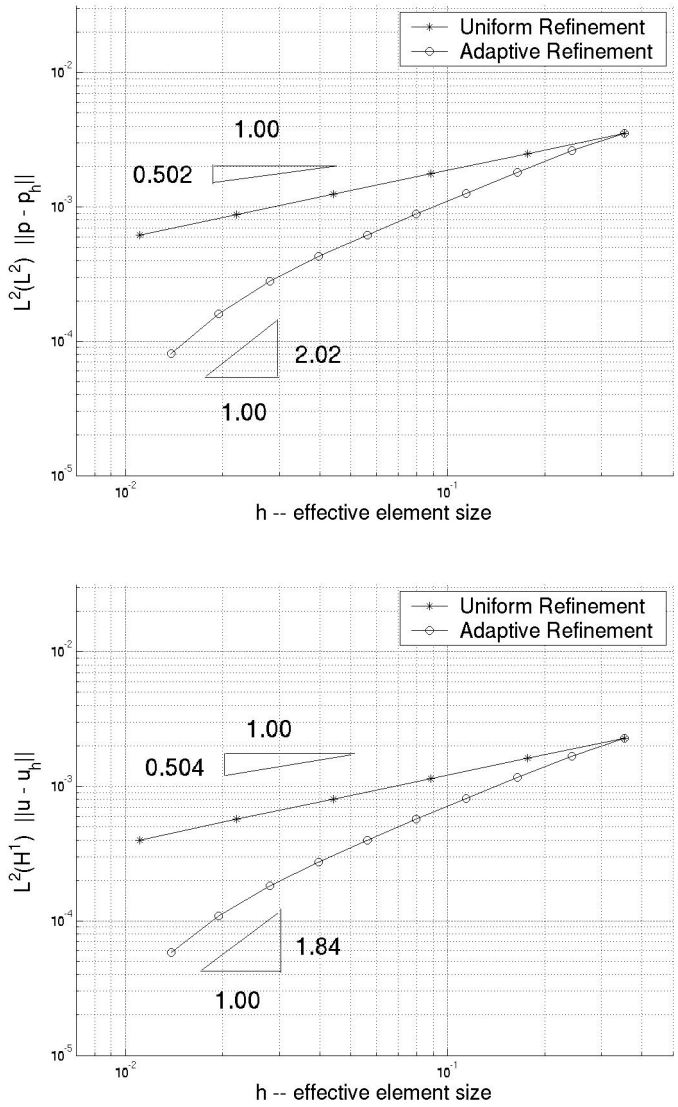

Figure 9.6: Uniform vs. adaptive refinement in Mandel's problem. $L^2(L^2)$ pressure error (top); $L^2(H^1)$ displacement error (bottom).

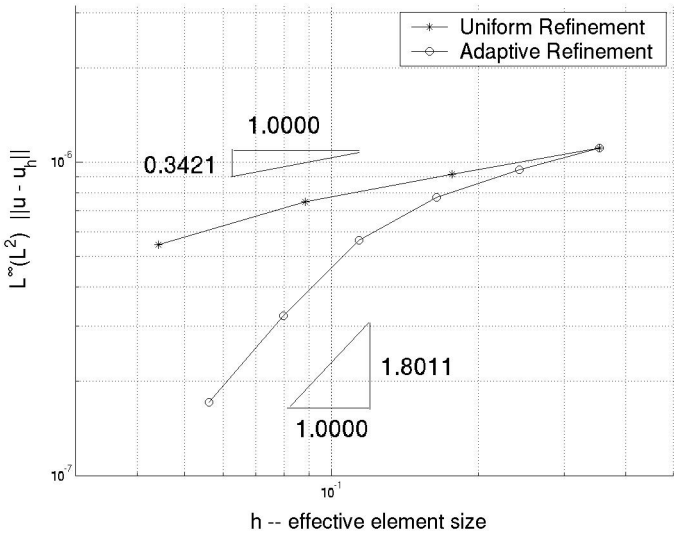

Figure 9.7: Uniform vs. adaptive refinement in Mandel's problem. $L^\infty(L^2)$ displacement error.

9.2 Controlling Locking

As shown in earlier chapters, locking, or non-physical pressure oscillations, is a problem that can sometimes plague numerical results in linear poroelasticity especially when a null constrained specific storage coefficient value, $c_o = 0$, is combined with continuous linear elements for displacements. To briefly recapitulate the idea: For the case that $c_o = 0$, $|\boldsymbol{\kappa}| \approx k << 1$ and the source term, $s_f(\mathbf{x}, t) = 0$, discretize in time the mass conservation equation (2.8b) to arrive at:

$$\nabla \cdot \mathbf{u}^{n+1} \approx \nabla \cdot \mathbf{u}^n + \Delta t \; k \; \nabla \cdot \nabla p^{n+1}.$$

Then for the typical case that $\mathbf{u}|_{t=0} = 0$, one finds

$$\nabla \cdot \mathbf{u}^1 \approx \Delta t \; k \; \nabla \cdot \nabla p^1. \tag{9.1}$$

For relatively large elements, the right side of (9.1) is likely to be small, and thus the initial deformation is driven to a divergence-free state, $\nabla \cdot \mathbf{u}^1 \rightarrow 0$. This can be problematic since the only continuous, linear *and* divergence-free vector is the constant vector. If boundary conditions imply a non-constant displacement field, problems will ensue.

One remedy proposed earlier was to use a discontinuous Galerkin approximation for the displacement field. The hope was that this would enable sufficient degrees of freedom to approximate a non-constant, divergence-free vector. Some success in removing non-physical pressure oscillations was achieved using discontinuous linear elements. Unfortunately, certain penalty parameters may need to be too large for this method to be a universal cure for the problem of locking.

The discontinuous Galerkin approach also does not specifically address the potentially useful boundary layers in the pressure field. If elements are small enough near a boundary layer to capture adequate detail in the solution, ∇p^1 may become very large. Thus $\nabla \cdot \mathbf{u}^1$ in (9.1) might be balanced by a sufficiently large term to avoid being effectively divergence-free.

Therefore, there are some instances when the CG/Mixed scheme might avoid locking *if* the elements are small enough to resolve a boundary layer. Since uniform refinement is computationally costly, a natural alternative is to use adaptive grid refinement.

The following two examples demonstrate the effectiveness of this approach in handling locking. The first example of Barry and Mercer contains

a useful large pressure gradient which is caused by a delta source function. The second example, the cantilever bracket problem, admits large pressure gradients along the corners of a domain edge due to a traction term.

9.2.1 The Problem of Barry and Mercer

Previous chapters showed the numerical results of the CG/Mixed and DG/Mixed schemes when applied to Barry and Mercer's problem. In those situations where locking did occur, a judicious choice of penalty parameters allowed the DG/Mixed scheme to eliminate the locking. The purpose here is to re-run the Barry and Mercer problem with the locking parameters using the adaptive grid technique.

To remind, the following set of material parameters are chosen:

$$E \;=\; 1e+5, \quad \nu = 0.1, \quad \alpha = 1, \quad c_o = 0, \quad \boldsymbol{\kappa} = 1e-6.$$

Furthermore, the refinement factor, $\gamma = 0.5$, is used.

Figure 9.8 shows the analytical (non-normalized) pressure solution in Barry and Mercer's problem at normalized time $\tau = 1e-1$. As time decreases, the peak in the analytical pressure solution decreases, and its width also decreases. Hence, one might expect large numerical fluxes in the region near the point source (here, $x = y = 0.25$). This leads to refinement indicators in the areas of the domain where the divergence is likely to be balanced by the flux, and thus limit locking.

Figure 9.9 shows how for an early time (normalized time $\tau = 1e-5$), locking occurs in the gray-shaded pressure field (left image) for a uniformly

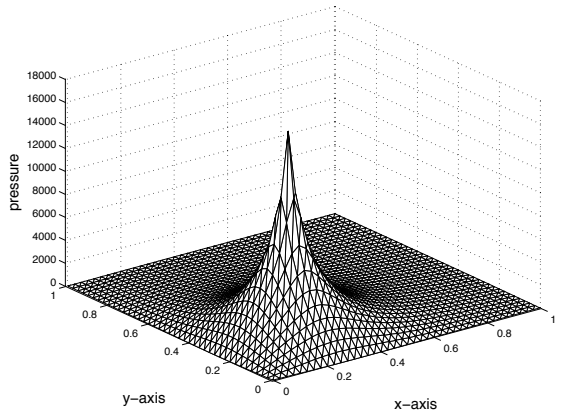

Figure 9.8: Early time analytical (non-normalized) pressure solution in Barry and Mercer's problem.

grid. Conversely, using less elements (right image), yet adaptively chosen, provides a correct physical pressure field. Figure 9.10 provides an alternative view of the pressure profile. In particular, the left image shows the negative and oscillating pressure field produced by a uniform grid. The right image shows no locking and the expected spike in the pressure field.

9.2.2 Mandel's Problem and Locking

The set of parameters which produce locking in Mandel's problem include a null constrained specific storage coefficient value and $E = 1e4$. The somewhat larger deformation produced through the smaller value for Young's modulus seems to magnify the effect of the penalty parameters[2] in the DG/Mixed scheme and make it difficult for the discontinuous scheme to

[2]The magnified effect essentially makes the DG/Mixed scheme effectively continuous. Note that the large deformation, however, means that the equations of *linear* poroelasticity are a poor model for this situation.

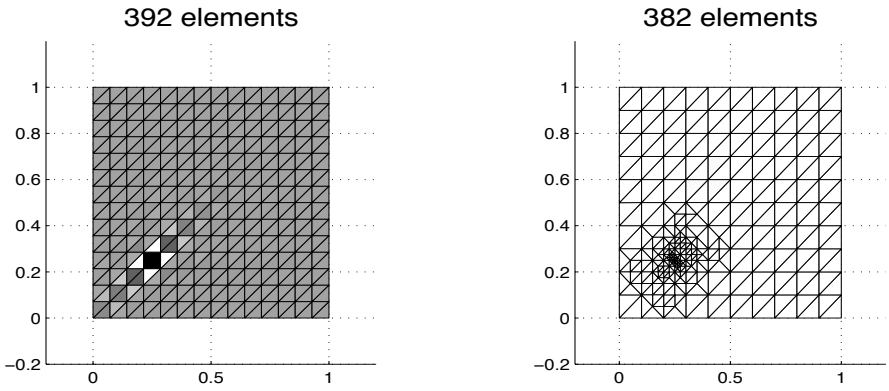

Figure 9.9: Non-physical pressure oscillations in Barry and Mercer's problem (left), whereas strategically placed and sized elements (right) eliminate locking.

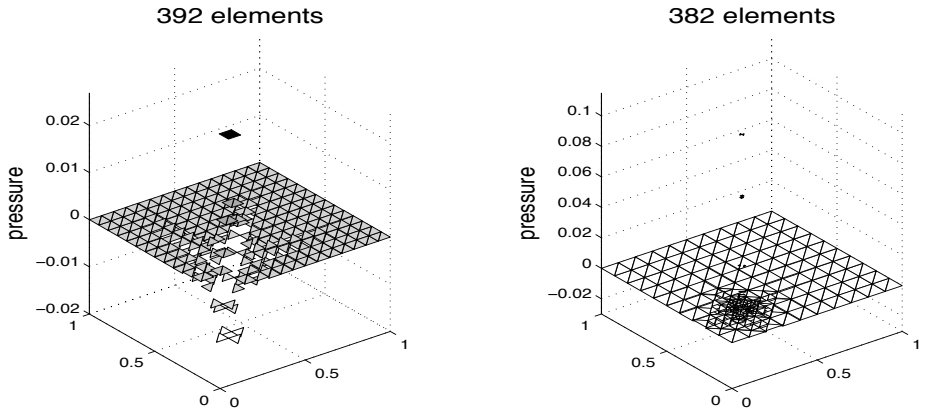

Figure 9.10: Locking produces negative and oscillating pressures in Barry and Mercer's problem (left), whereas strategically placed and sized elements (right) simulates accurately the delta function source term.

eliminate locking. In fact, for rectangles no suitable parameter choices were discovered.

On the other hand, Figure 9.11 shows that by resolving the pressure boundary layer the adaptive grid technique easily eliminates the spurious pressure oscillation present for the unrefined grid.

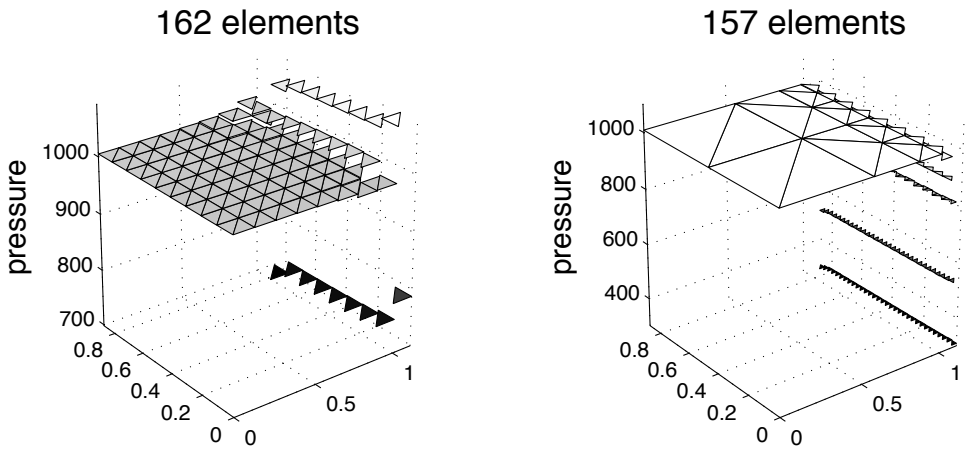

Figure 9.11: Locking in Mandel's problem is produced in a uniform grid with 162 elements using the CG/Mixed method. Fewer elements (157), but strategically placed, eliminate the locking quite effectively.

Chapter 10

Applications

The purpose of this chapter is to apply the numerical algorithms developed in this dissertation to some traditional and novel problems in linear poroelasticity. It is hoped that the results might provide some physical insight and potentially motivate engineering design choices. From a computational perspective, since there are no analytical solutions to any of the problems posed here, the intuition gathered from the numerical work of the previous chapters will indicate potential problems and point to the appropriate numerical strategies.

10.1 The Cantilever Bracket Problem

The example considered here is the cantilever bracket problem, pictured in Figure 10.1. Traditionally a problem in linear elasticity to model loads on beams, the loaded fluid-saturated medium may have some relevance in an industrial setting and was first studied as a problem in linear poroelasticity by Liu [2004]. The problem may also resemble and have applications to load-induced blood flow within bone [Steck et al., 2000, Swan et al., 2003]; this might help to estimate optimal (external) mechanical loading in order to produce effective bone adaptation.

The problem is posed on the unit square domain, $(0,1) \times (0,1)$. A

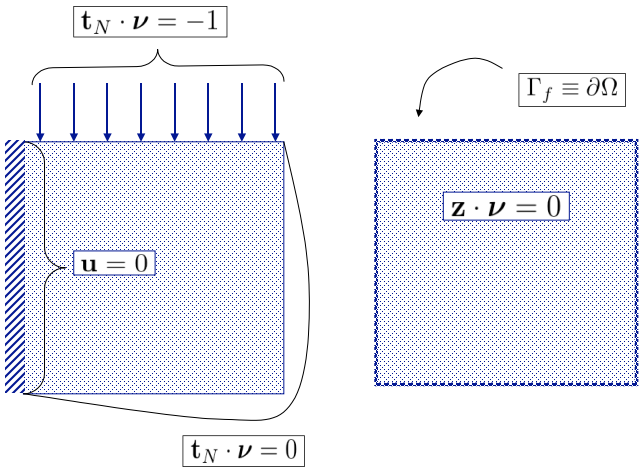

Figure 10.1: The Cantilever Bracket Problem.

no-flow boundary condition is imposed along all sides; the deformation is fixed along the left edge, $x = 0$, and a downward traction is applied at the top side, $y = 1$. The right and bottom sides are traction-free.

The material parameters are chosen so that the solid and fluid constituents are incompressible and a relatively small permeability is chosen. The material parameter values are summarized below:

$$E = 1e + 4, \quad \nu = 0.4, \quad c_o = 0, \quad \boldsymbol{\kappa} = 1e - 7.$$

A natural question concerns the evolution of the pressure field over time. There is interest in the initial response, but since $c_o = 0$, a short time step produces initial non-physical pressure oscillations as in Figure 10.2. Although the locking problem is ameliorated by each of the DG/Mixed schemes in Figures

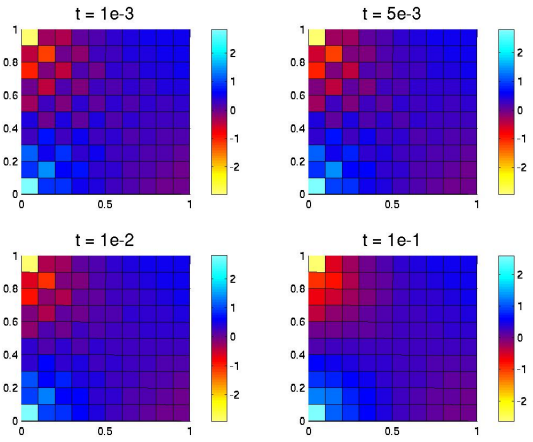

Figure 10.2: Locking in the cantilever bracket pressure field is observed for the CG/Mixed method when $c_o = 0$ at time $t = 1e-3$, but which dissipates at later times.

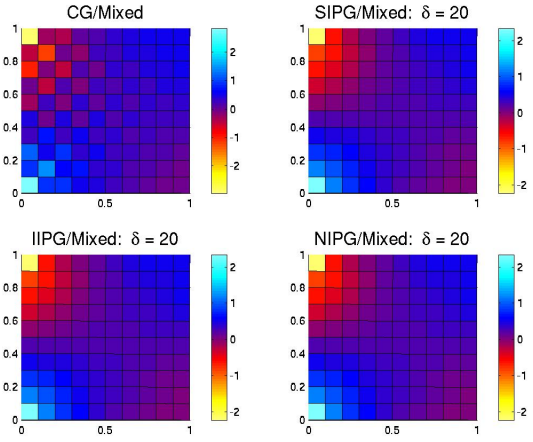

Figure 10.3: Locking in the pressure field is prevented by using the DG/Mixed method despite the fact that $c_o = 0$ and the permeability is small. Shown at time $t = 1e-3$.

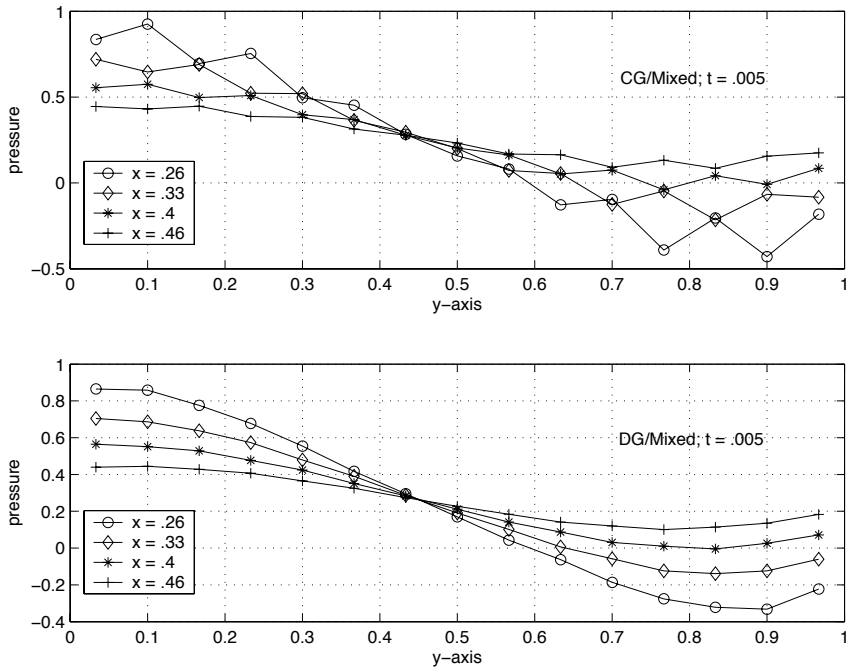

Figure 10.4: Shown in each graph is the pressure along different x-lines at time = .005: $p(x = .26, y, t = .005)$, $p(x = .33, y, t = .005)$, $p(x = .4, y, t = .005)$, $p(x = .46, y, t = .005)$.

10.3-10.4, the nature of the boundary conditions suggest that the pressure may experience larger variation than those produced by the DG/Mixed schemes.

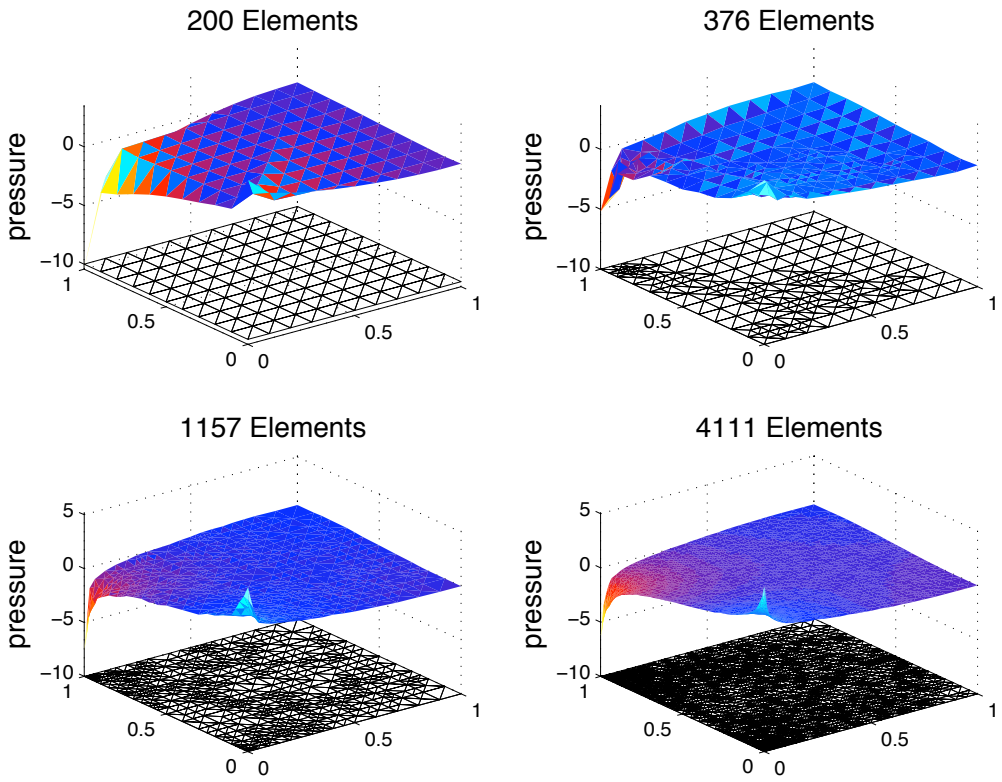

Figure 10.5: Adaptive refinement for the cantilever bracket problem with locking-susceptible parameters.

Therefore, a better numerical strategy may be couple the CG/Mixed scheme with the adaptive grid technique. Figure 10.5 shows different levels of the adaptive refinement iteration, and how locking is gradually removed. Notice that the entire domain is progressively refined, but the most dense location of elements are on the corners corresponding to the larger pressure values that were expected.

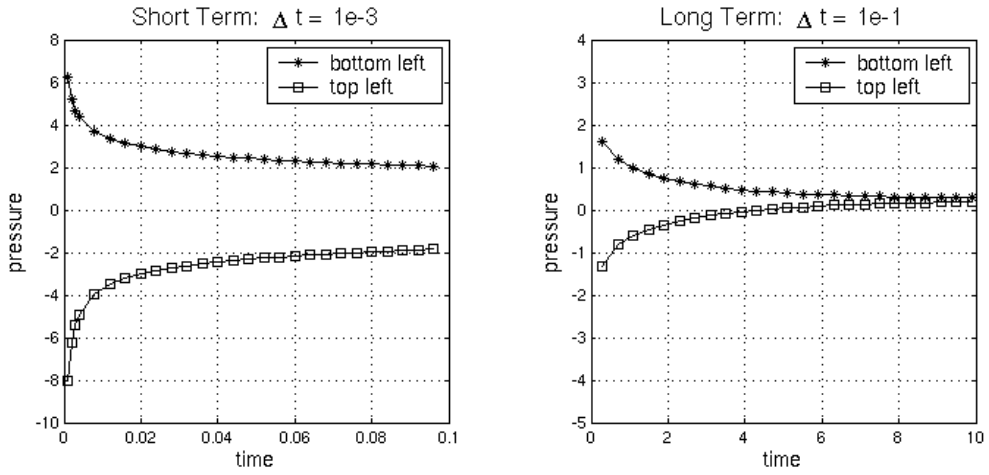

Figure 10.6: Extreme pressure at the corners in the cantilever bracket problem.

Using the refined grid some observations can be made. Using a short time step, Figure 10.6 shows that the short term solution experiences a large absolute initial pressure on the left corners. Depending on the materials being used, the large initial pressures might need to be taken into account in a design process. Alternatively, the same figure shows the longer term solution obtained from a longer time step. By approximately 10 seconds the pressure throughout the domain has reached an equilibrium value of $p \approx 0.2604$. Figure 10.7 shows the evolution of the pressure profile.

10.2 Well Placement and Subsidence

It is estimated that subsidence produces annual damages ranging into the hundreds of millions of dollars. In addition, subsidence can cause especially costly damage to coastal regions because it makes them more susceptible to flooding. Furthermore, as the following quote from *The Handbook of Texas*

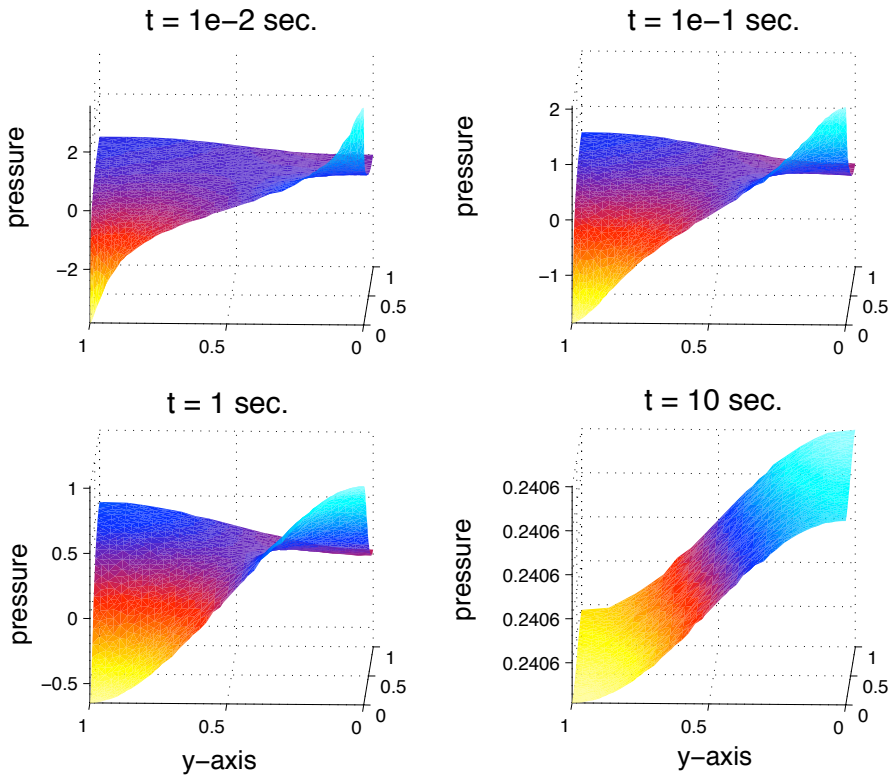

Figure 10.7: Evolution of the pressure field in the cantilever bracket problem.

Online [1999] illustrates, extraction of underground water is heavily utilized:

> Underground water is a principal water resource in Texas, and its importance as a source of supply for municipal, industrial, and irrigation uses, as well as domestic and livestock purposes, is immeasurable. In 1990 more than half of the water used in Texas was underground water; about 71 percent was for agricultural use and 21 percent for municipal use, with the remainder going to industrial needs.

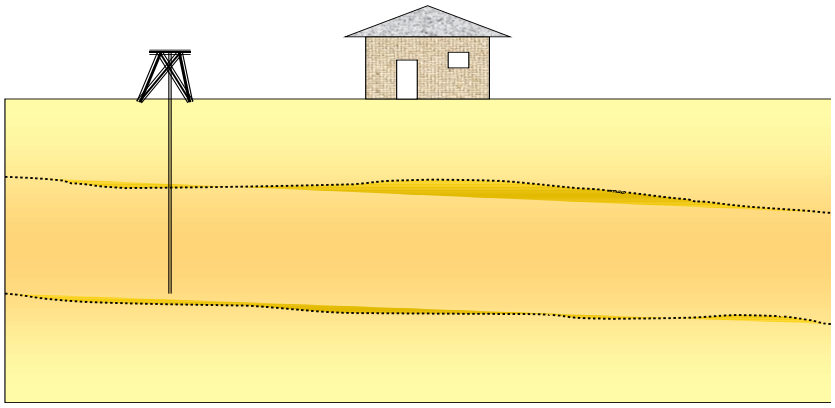

Figure 10.8: A simplified model of subsidence: The house exerts a downward force, and the well extracts the subsurface fluid.

Figure 10.8 represents a simplified model[1] of subsidence. The weight of the house exerts a downward force, and the extraction well extracts the

[1]The model is simplified in many ways, one of which is the lack of some sort of rigid plate constraint on the house-induced deformation. One would expect the foundation would compress in close to a linear way.

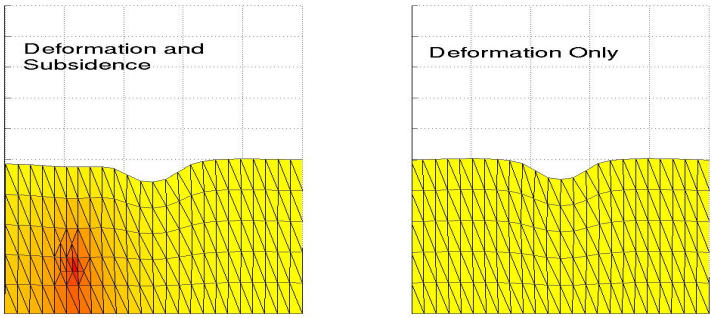

Figure 10.9: The subsidence and tilt produced by well extraction (left) and deformation produced by house alone (deformations magnified).

subsurface fluid. Figure 10.9 shows the magnified deformation induced by the house itself (right), and with the operating extraction well (left). As seen, the deformation on the right produces a tilt which might lead to foundation or other structural house damage.

Figure 10.10 illustrates one possible remedy for the tilt occurring at time $t = T_{tilt}$. At this time a second well is installed on the opposite side of the domain. Both wells are then set to half production to maintain the same overall production. By time $t = 2T_{tilt}$, the two-well system has removed the uneven foundation. Note that although the tilt has been addressed, subsidence continues as the domain continues to vertically compress.

If an extra well is economically infeasible, other options using ideas from linear poroelasticity are available to remove the tilt. If the field remains saturated, one could stop well 1 production altogether and the effects of subsidence eventually dissipate. A quicker solution is to convert well 1 to an injection well.

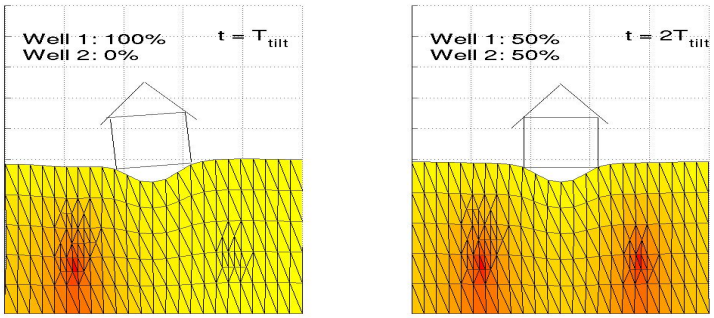

Figure 10.10: Foundation tilt mitigation. House foundation tilt is caused through well 1 production only (left). Subsequently, well 1 is taken to half production and well 2 is put online at an equal rate (right). In the time that it took for the tilt to occur, the two-well system has removed the tilt.

10.3 Sudden Pressure Changes in Porous Deformable Media

The third problem of interest regards small to medium scale porous deformable devices and the displacement and pressure dynamics of induced by sudden pressure changes. The potential exists for such devices to be used in a number of different mechanisms. A rapid pressure increase usually induces a quick expansion of the material, whereas a rapid pressure decrease induces a contraction.

10.3.1 An Application to Scuba Suit Design

There are times when deep sea scuba divers experience air loss or panic and must rapidly return to the surface. In doing so, the diver will experience a rapid pressure decrease which can lead to a life threatening condition known as Decompression Sickness. This causes nitrogen bubbles in the bloodstream

and bodily tissues which might lead to unconsciousness or death.

One possible remedy for this is to design a scuba suit with a thin poroelastic sheath embedded on its surface. The purpose of such a material addition would be to make for a slower change in pressure on the body's surface area. In effect, the poroelastic sheath would retain much of the deep sea pressure, and then slowly transfer the lower shallow water pressure to the body.

Figure 10.11 shows the pressure field and the deformation of a cross section of the scuba suit material after 1, 5, 20 and 40 seconds, respectively, using a permeability of $\boldsymbol{\kappa} = 1e-2$. Despite being exposed to a large pressure reduction during a rapid ascent, the diver has yet to experience a significant pressure change until about 40 seconds. Figure 10.12 provides another illustration of how the poroelastic sheath acts as a buffer slowing the effects of pressure reduction on the skin. The same figure also shows the effect of lowering the permeability.

One might observe that although pressure boundary layers might possibly be present for short times and at the upper boundary, the question of interest is the time at which the lower surface (i.e., the body) experiences pressure changes. So from a computational perspective, an unrefined grid and a relatively large time step $\Delta t = 1e-1$ are adequate to provide interesting answers.

Some open questions do, however, remain. One factor in the design would be diver flexibility. The poroelastic sheath should not be too bulky or rigid. Therefore, one might try models which incorporate different assumptions on the Lamé coefficients and material thicknesses. Another question pertains to the precise nature of physiological response to pressure drops. This would

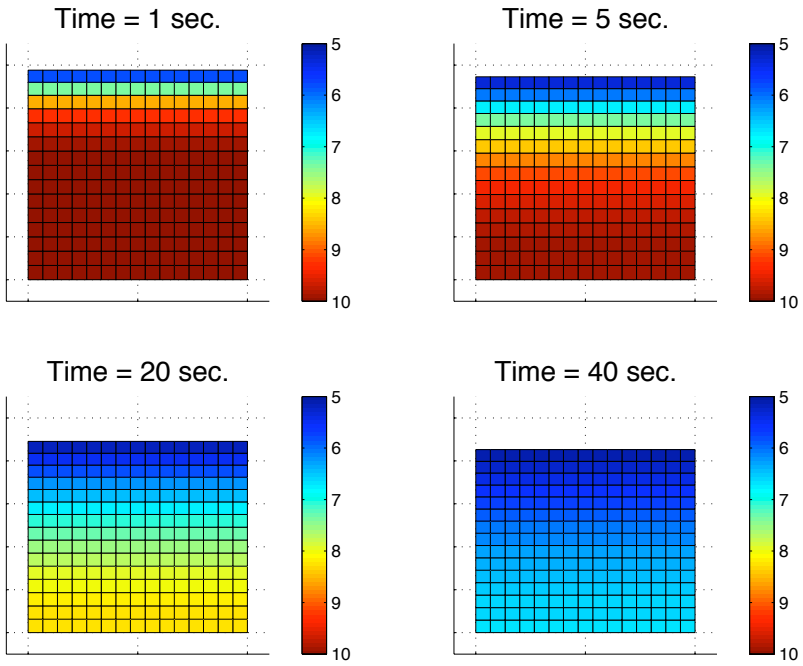

Figure 10.11: The pressure and deformation of a cross section of the scuba suit material at various times. The body surface (bottom of each image) is not immediately exposed to the pressure drop.

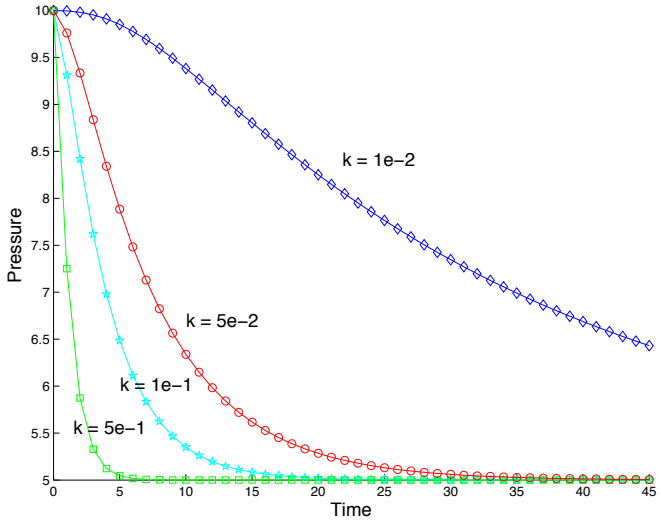

Figure 10.12: The effect of permeability on the pressure at the skin's surface

enable suit designers to choose the correct permeability which, as Figure 10.12 demonstrates, acts as a sort of timing mechanism.

10.3.2 Tubular Geometries

Tubular geometries may have applications as wide-ranging as wells and arteries. There are current research groups at the University of West Bohemia and the University of Arizona modeling the tunica media (the largest inner portion of an artery) as porous deformable media. The ability of the tunica media to expand with the pressure fluctuations helps to stabilize and reduce the blood pressure by the time the blood reaches the smaller and more fragile capillaries.

The example considered here is a simple tube exposed to a large but

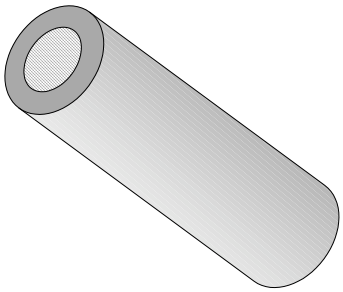

Figure 10.13: Tubular geometry exposed to a sudden pressure increase on the interior wall.

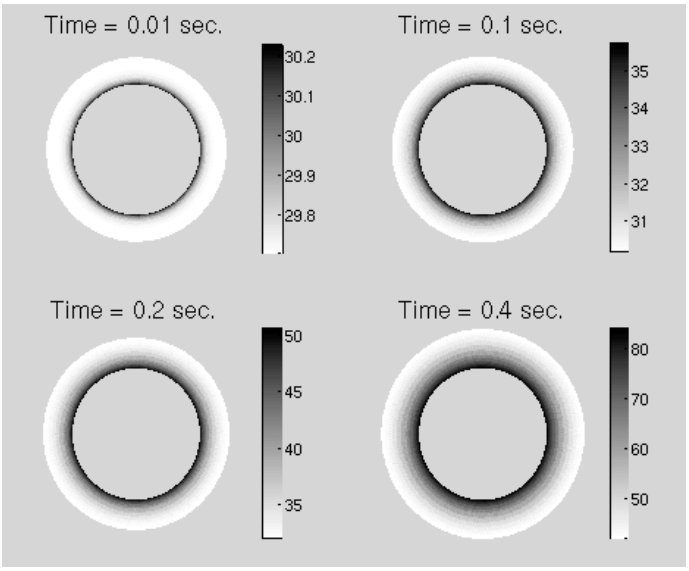

Figure 10.14: The pressure and deformation profile after being exposed to a sudden pressure increase on the interior wall (deformation magnified).

decreasing pressure[2] on its inner boundary. There is interest in the short term dynamics, so an adaptively refined grid is used. Figure 10.14 shows the pressure and deformation profile (with deformation magnified for easier visualization). By the time ($t \approx 0.4$) of the minimal pressure on the inner boundary, the tube has absorbed much of the pressure from the inner region of the tube partly through expansion.

[2]In the same way an artery is exposed to the larger systolic pressure followed by the smaller diastolic pressure.

Chapter 11

Conclusion

11.1 Summary

An algorithm for linear poroelasticity which couples a continuous space for displacements and a mixed space for flow was introduced and analyzed for its existence, uniqueness and optimality. From the theoretical and computational results for small time steps two principal issues emerged – non-physical pressure oscillations (locking) and regularity.

For the locking-prone situation where the constrained specific storage coefficient is null, $L^2(L^2)$ optimality for the pressure showed that the oscillations dissipate through time and gradually decrease with a refined grid. As one potential remedy for locking at early times and for unrefined grids, a discontinuous variation of the algorithm was proposed and analyzed. Existence, uniqueness and optimality were proved, but its ability to eliminate locking relied too heavily on certain penalty parameters to make it the universal solution.

The more prominent problem coming to the fore was the effect of regularity on convergence. It was shown directly that several example problems lack the regularity to achieve optimal convergence rates of even the lowest order. However, the typical pressure boundary layers which hinder convergence rates for uniformly refined grids suggested a fruitful opportunity to employ adaptive grid techniques. Therefore, *a posteriori* error estimates were proved

in terms of naturally defined residual terms. Using these residuals an indicators for grid refinement, exceptional improvements in convergence rates were discovered. Through resolution of boundary layers, the adaptive grid technique also proved to be a very effective means of eliminating locking.

11.2 Future Work

There are some unanswered questions within the scope of the theoretical and computational results of this work. For instance, higher order spaces were never implemented. For those problems which lack regularity, there is good reason to believe that higher order spaces would *not* lead to improved convergence when a small time step is utilized. However, it might be interesting to examine, for example, the impact of continuous quadratic elements for displacements on locking. The extra degrees of freedom might be sufficient to approximate the displacement in the locking-inducing situation where $\mathbf{u}$ is non-constant and divergence-free. Another interesting question regards using different grids for the displacement and flow variables. It was mentioned that the theoretical *a priori* error estimates remain unchanged for this situation in the asymptotic sense as the grids are refined. Since it was the pressure which often admitted boundary layers, it would be useful to adaptively refine only the pressure grid and to compare the improvements in convergence with those seen when one grid for both variables was used. Lastly, *a posteriori* estimates were not provided for the DG/Mixed scheme, but would represent a natural extension of the results developed for the CG/Mixed scheme. It would, in particular, be nice to examine the computational results for the DG/Mixed scheme on a refined, non-conforming grid.

There are also several areas which might pose interesting future research

problems. These areas fall outside the range of coverage here, but provide interesting challenges nonetheless.

Extended Model. The items listed below present opportunities for model extensions and theoretical analysis:

- Nonlinear poroelasticity resulting from more general constitutive relations;
- Incorporating electromechanical, chemical and/or thermal effects;
- Dropping the quasi-static assumption, multiporosity models, multiphase flow models, etc.

Algorithmic Enhancements. Below are some categories which might lead to useful upgrades of the algorithms presented herein:

- Time dependent grids – for larger times, the finite element error usually significantly dampens. This obviates the need for a costly refined grid.
- Developing implicit *a posteriori* estimates whereby local problems are solved;
- For lowest order mixed spaces, the addition of pressure edge penalty term $\sum_{e_a \in \Gamma_{int}} \frac{\delta_p}{|e_a|^\beta} \int_{e_a} [p][w]\, ds$ to the mass conservation equation (3.17) in the CG/Mixed method. Motivated by the chapter on locking, this added term moves the deformation away from a divergence-free state in areas with possible pressure oscillations.
- Development of a reliable indicator for locking.

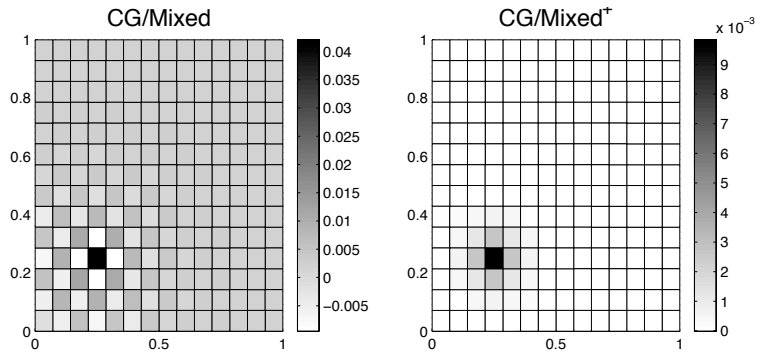

Figure 11.1: CG/Mixed solution (left) vs. the CG/Mixed$^+$ (right) schemes. The added pressure penalty eliminates locking as effectively as with the DG/Mixed methods.

Numerical Analysis. Several computational models have been proposed which extend either the physics or algorithms presented here, but are lacking theoretical examination:

- Analyzing the theoretical properties of the algorithms which incorporate thermal effects [Baggio et al., 1997, Liu, 2004];
- Analyzing the theoretical properties of the algorithms which incorporate a placticity assumption [Liu, 2004];
- Analyzing the theoretical properties of the algorithms which incorporate multiphase flow [Gai, 2004];
- Analyzing the theoretical properties of the CG/Mixed variant which uses an iteratively coupled time stepping scheme [Gai, 2004].

Appendices

Appendix A

Glossary of Terms

Variable	Description (SI units)
$\mathbf{u}$	solid displacement (m)
p	pressure (N/m^2)
$\mathbf{z}$	fluid flux (m/s)
$\tilde{\sigma}_{ij}$	total stress (N/m^2)
σ_{ij}	effective stress (N/m^2)
ϵ_{ij}	solid strain tensor (—)
$\boldsymbol{\kappa}$	permeability (m^2)
η	fluid content (—)
α	*Biot-Willis* constant (—)
c_o	constrained specific storage coefficient (m^2/N)
M	*Biot's* modulus (N/m^2)
$\lambda,\ \mu$	*Lamé* coefficients (N/m^2)
μ_f	fluid viscosity (Ns/m^2)
ϕ	porosity (—)
K	drained skeleton bulk modulus (N/m^2)
K_u	undrained skeleton bulk modulus (N/m^2)
K_s	solid bulk modulus (N/m^2)
K_f	fluid bulk modulus (N/m^2)
E	*Young's* modulus (N/m^2)
E_u	undrained *Young's* modulus (N/m^2)
ν	*Poisson's* coefficient (—)
ν_u	undrained *Poisson's* coefficient (—)
c_f	fluid diffusivity coefficient (1/s)
B	*Skempton's* coefficient (—)

A.1 Some Relations Between Terms

The first set of relations involve the Lamé coefficients μ, λ, and the *skeleton bulk modulus* – K, *Young's modulus* – E and *Poisson's coefficient* – ν. These relations are common in the field of linear elasticity.

$$\begin{aligned} K &= \lambda + \frac{2}{3}\mu, \\ E &= \mu\frac{9K}{3K+\mu}, \\ \nu &= \frac{3K-2\mu}{2(3K+\mu)}. \end{aligned}$$

Additionally, K_u has the following relation,

$$K_u = K + \frac{\alpha^2}{c_o}.$$

With the above, two useful relations are discovered for the *fluid diffusivity coefficient* – c_f, and *Skempton's coefficient* – B:

$$\begin{aligned} c_f &= \frac{1}{c_o}\tilde{\boldsymbol{\kappa}}\frac{K+\frac{4}{3}\mu}{K_u+\frac{4}{3}\mu}, \\ B &= \frac{\alpha}{c_o K_u}. \end{aligned}$$

Clearly, c_f is usefully defined only for the case where the permeability $\boldsymbol{\kappa}$ is a constant. Skempton's coefficient can be used to reveal the useful relation between Poisson's coefficient and its undrained version,

$$\frac{\alpha B(1-2\nu)}{3} = \frac{\nu_u - \nu}{1+2\nu_u}.$$

One last relation worth mentioning involves the *constrained specific storage coefficient*, c_o, which is sometimes expressed in terms of *Biot's modulus*, M, as $c_o = \frac{1}{M}$.

Appendix B

Nondimensionalisation

The full equations of linear poroelasticity are

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} - (\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla^2\mathbf{u} + \alpha\nabla p = \mathbf{f}, \tag{B.1}$$

$$\frac{\partial}{\partial t}(c_o p + \alpha\nabla\cdot\mathbf{u}) - \frac{1}{\mu_f}\nabla\cdot\boldsymbol{\kappa}(\nabla p - \rho_f\mathbf{g}) = s_f. \tag{B.2}$$

For the assumption of a positive, constant constrained specific storage coefficient, (B.1)-(B.2) can be nondimensionalised by setting $\tilde{x} = \frac{x}{x_o}$, $\tilde{t} = \frac{t}{t_o}$ and

$$\tilde{p} = \frac{p}{(\lambda+2\mu)}, \quad \tilde{\mathbf{u}} = \frac{\mathbf{u}}{x_o}, \quad \tilde{\mathbf{f}} = \frac{x_o}{\mu}\mathbf{f}, \quad \tilde{s_f} = \frac{t_o}{c_o(\lambda+2\mu)}s_f.$$

Here, $m = 1 + \frac{\lambda}{\mu}$, and x_o and t_o are characteristic length and time scales, respectively. Then assuming, for simplicity, that $k \equiv \frac{1}{\mu_f}\boldsymbol{\kappa}$ is constant, $\alpha = 1$ and the gravity term is negligible, the following equations result:

$$\left[\frac{\rho x_o^2}{\mu t_o^2}\right]\frac{\partial^2\tilde{\mathbf{u}}}{\partial\tilde{t}^2} - m\tilde{\nabla}(\tilde{\nabla}\cdot\tilde{\mathbf{u}}) - \tilde{\nabla}^2\tilde{\mathbf{u}} + (m+1)\tilde{\nabla}\tilde{p} = \tilde{\mathbf{f}}, \tag{B.3}$$

$$\frac{\partial\tilde{p}}{\partial\tilde{t}} + \left[\frac{1}{c_o(\lambda+2\mu)}\right]\frac{\partial\tilde{\nabla}\cdot\tilde{\mathbf{u}}}{\partial\tilde{t}} - \left[\frac{t_o}{x_o^2}\left(\frac{k}{c_o}\right)\right]\tilde{\nabla}^2\tilde{p} = \tilde{s_f}. \tag{B.4}$$

The flow equation is more likely to reveal a relationship between the time and space scales since a small deformation is assumed in linear poroelasticity. Therefore, the nondimensional coefficient of $\tilde{\nabla}\tilde{p}$ in (B.4) suggests the relation

$$x_o \approx \sqrt{\left(\frac{k}{c_o}\right) t_o}. \tag{B.5}$$

This may be helpful in predicting the location of boundary layers. Figure 4.2 illustrates this point for Terzaghi's problem, whereas Figure 4.9 does the same for Mandel's problem.

Alternatively, rewriting (B.5), one finds the following suggestive relation for the time scale

$$t_o \approx \left(\frac{c_o}{k}\right) x_o^2. \tag{B.6}$$

This shows that a small permeability induces a long time scale, whereas a small constrained specific storage coefficient value leads to a short time scale. This was seen when comparing the solutions for Mandel's problem with different values for the constrained specific storage coefficient.

Plugging in this value for t_o into (B.3) results in

$$\left[\frac{\rho k^2}{\mu c_o^2 x_o^2}\right] \frac{\partial^2 \tilde{\mathbf{u}}}{\partial \tilde{t}^2} - m\tilde{\nabla}(\tilde{\nabla} \cdot \tilde{\mathbf{u}}) - \tilde{\nabla}^2 \tilde{\mathbf{u}} + (m+1)\tilde{\nabla}\tilde{p} = \tilde{\mathbf{f}}.$$

For many situations, including most geomechanical settings, the permeability, k, is very small and so the nondimensional coefficient for $\frac{\partial^2 \tilde{\mathbf{u}}}{\partial \tilde{t}^2}$ can be quite

small. This justifies the quasi-static assumption in most cases. However, for a small constrained specific storage coefficient value and/or a small length scale, one might consider adding the $\frac{\partial^2 \tilde{\mathbf{u}}}{\partial \tilde{t}^2}$-term.

Remark B.0.1. The preceding analysis is somewhat general since it does not incorporate the effect boundary and initial conditions. This could be an important omission since those are factors which may indeed determine the dynamics of the problem. For a more accurate nondimensionalization, each problem should be analyzed separately.

Appendix C

CG/Mixed Discrete in Time Analysis

The purpose of this appendix entry is to provide an indication of the issues involved in the computational implementation of the CG/Mixed scheme and to provide[1] discrete in time error estimates. For brevity, only the case of positive and constant constrained specific storage coefficient value is provided.

C.1 Numerical Implementation

The numerical formulation is now provided by discretizing our algorithm in time. To do so the theta method is used from numerical differential equations.

Let $\Delta t = T/N$ where N is a positive integer and let $t_j = j\Delta t$. The following notation is valid for $\theta \in [0,1]$:

$$\begin{aligned} g_j &= g(x,t_j), \quad 0 \le j \le N, \\ g_{j,\theta} &= \frac{1}{2}(1+\theta)g_{j+1} + \frac{1}{2}(1-\theta)g_j, \quad 0 \le j \le N-1. \end{aligned}$$

With this, (3.16)-(3.18) becomes: Find $(\bar{\mathbf{u}}_j, \bar{\mathbf{z}}_j, \bar{p}_j)$ such that

[1] Although, for brevity, the results here are slightly less rigorous and general than for the continuous in time proofs. Some simplifying assumptions are made.

$$
\begin{aligned}
a_{\mathbf{u}}(\bar{\mathbf{u}}_{j,\theta}, \mathbf{v}_h) - \alpha(\nabla \cdot \mathbf{v}_h, \bar{p}_{j,\theta}) &= l_{1j,\theta}(\mathbf{v}_h), \qquad \text{(C.1)} \\
c_o\Big(\frac{\bar{p}_{j+1} - \bar{p}_j}{\Delta t}, w_h\Big) + \alpha\Big(\nabla \cdot \Big(\frac{\bar{\mathbf{u}}_{j+1} - \bar{\mathbf{u}}_j}{\Delta t}\Big), w_h\Big) & \\
+(\nabla \cdot \bar{\mathbf{z}}_{j,\theta}, w_h) &= l_{2j,\theta}(w_h), \qquad \text{(C.2)} \\
(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \bar{\mathbf{z}}_{j,\theta}, \mathbf{s}_h) - (\bar{p}_{j,\theta}, \nabla \cdot \mathbf{s}_h) &= l_{3j,\theta}(\mathbf{s}_h), \qquad \text{(C.3)}
\end{aligned}
$$

for all $(\mathbf{v}, w, \mathbf{s}) \in (\mathbf{V}_{h,0}, W, \mathbf{S}_{h,0})$. It is mentioned that $\theta = 1$ corresponds to the backward Euler method, $\theta = 0$ to the Crank-Nicolson scheme.

C.1.1 Matrix Formulation

Rearranging (C.1)-(C.3) and reverting to matrix form yields

$$
\begin{pmatrix}
\frac{1}{2}(1+\theta)A_{\mathbf{uu}} & -\alpha\frac{1}{2}(1+\theta)A_{\mathbf{u}p}^T & 0 \\
\alpha A_{\mathbf{u}p} & A_{pp} & \frac{\Delta t}{2}(1+\theta)A_{p\mathbf{z}}^T \\
0 & -\frac{1}{2}(1+\theta)A_{p\mathbf{z}} & \frac{1}{2}(1+\theta)A_{\mathbf{zz}}
\end{pmatrix}
\begin{pmatrix} \mathbf{u}_{j+1} \\ p_{j+1} \\ \mathbf{z}_{j+1} \end{pmatrix}
=
\begin{pmatrix} b_{1,j} \\ b_{2,j} \\ b_{3,j} \end{pmatrix},
$$

where the right side $\begin{pmatrix} b_{1,j} & b_{2,j} & b_{3,j} \end{pmatrix}^T$ equals

$$
\begin{pmatrix} l_{1j,\theta}(\mathbf{v}_h) \\ (\Delta t) l_{2j,\theta}(w_h) \\ l_{3j,\theta}(\mathbf{s}_h) \end{pmatrix}
-
\begin{pmatrix}
\frac{1}{2}(1-\theta)A_{\mathbf{uu}} & -\alpha\frac{1}{2}(1-\theta)A_{\mathbf{u}p}^T & 0 \\
-\alpha A_{\mathbf{u}p} & -A_{pp} & \frac{\Delta t}{2}(1-\theta)A_{p\mathbf{z}}^T \\
0 & -\frac{1}{2}(1-\theta)A_{p\mathbf{z}} & \frac{1}{2}(1-\theta)A_{\mathbf{zz}}
\end{pmatrix}
\begin{pmatrix} \mathbf{u}_j \\ p_j \\ \mathbf{z}_j \end{pmatrix}.
$$

It is remarked that the matrix $A_{\mathbf{uu}}$ can be found by using existing linear elasticity software. Likewise, the matrices $A_{\mathbf{zz}}$ and $A_{p\mathbf{z}}$ can be obtained using existing software for mixed formulations of elliptic equations. The only matrices that must be independently assembled are A_{pp} and the coupling matrix, $A_{\mathbf{u}p}$.

C.2 Existence and Uniqueness

Comment is made here regarding existence and uniqueness. Previously proved was the existence and uniqueness of the continuous in time scheme (3.16)-(3.18). For existence and uniqueness of (C.1)-(C.3), one examines the existence and uniqueness at each time step. Lipnikov [2002] proves the unique solvability of (C.1)-(C.3) at each time step by showing that the abstract bilinear operators in the problem satisfy Brezzi's first stability condition and the Ladyshenskaya-Babuska-Brezzi (LBB) $\inf - \sup$ condition. He also demonstrates that our choice of finite-dimensional spaces satisfies the discrete forms of the above conditions, and thus the problem has a unique solution.

C.3 Discrete in Time Error Estimates

To facilitate the development of discrete in time error estimates, the following definitions are made:

$$||g||_{l^\infty(L^2)} = \max_{j=0,\ldots,N} ||g_j||_{L^2}, \quad ||g||_{l^2(H^s)} = \Big(\sum_{j=0}^{N-1} ||g_{j,\theta}||^2_{H^s} \Delta t \Big)^{\frac{1}{2}}.$$

For any sufficiently smooth function g(t), by using the Taylor series expansion about $t = t_{j,\theta}$,

$$\begin{aligned} g_{j+1} &= g|_{t=t_{j,\theta}} + \frac{1}{2}(1-\theta)\Delta t \frac{\partial g}{\partial t}|_{t=t_{j,\theta}} + \frac{1}{8}(1-\theta)^2(\Delta t)^2 \frac{\partial^2 g}{\partial t^2}|_{t=t_{j,\theta}} + O(\Delta t^3), \\ g_j &= g|_{t=t_{j,\theta}} - \frac{1}{2}(1+\theta)\Delta t \frac{\partial g}{\partial t}|_{t=t_{j,\theta}} + \frac{1}{8}(1+\theta)^2(\Delta t)^2 \frac{\partial^2 g}{\partial t^2}|_{t=t_{j,\theta}} + O(\Delta t^3). \end{aligned}$$

Therefore, if the first equation is multiplied by $\frac{1}{2}(1+\theta)$ and the second equation by $\frac{1}{2}(1-\theta)$, summation produces

$$g_{j,\theta} = g|_{t=t_{j,\theta}} + \frac{1}{16}(\Delta t)^2(1-\theta)(1+\theta)\frac{\partial^2 g}{\partial t^2}|_{t=t_{j,\theta}} + O(\Delta t^3),$$

with the obvious exception that for $\theta = 1$, $g_{j,\theta} = g|_{t=t_{j,\theta}}$. The above expression allow one to approximate to order $O(\Delta t^2)$, $\mathbf{u}(\mathbf{x}, t_{j,\theta}) \approx \mathbf{u}_{j,\theta}$, $p(\mathbf{x}, t_{j,\theta}) \approx p_{j,\theta}$, and $\mathbf{z}(\mathbf{x}, t_{j,\theta}) \approx \mathbf{z}_{j,\theta}$.

The following relationships hold for $\theta \in (0, 1]$ as a result of Taylor's expansion (see Rivière and Wheeler [1999]):

$$\frac{p_{j+1} - p_j}{\Delta t} = p_t(x, t_{j,\theta}) + \Delta t\ \rho_{p,j,\theta}, \quad \forall x \in \Omega, \tag{C.4}$$

$$\frac{\mathbf{u}_{j+1} - \mathbf{u}_j}{\Delta t} = \mathbf{u}_t(x, t_{j,\theta}) + \Delta t\ \rho_{\mathbf{u},j,\theta}, \quad \forall x \in \Omega, \tag{C.5}$$

where $\rho_{p,j,\theta}$ and $\rho_{\mathbf{u},j,\theta}$ depend on time-derivatives of p and $\mathbf{u}$ respectively. For $\theta = 1$, the approximation would be one order higher in Δt.

A priori discrete in time error estimates are now examined. The estimates are required to hold asymptotically in the limit as $\Delta t \to 0$. As for the case of continuous in time estimates, it is necessary from the outset to assume that the linear poroelasticity solution belongs to the following spaces

$$p \in L^\infty(0, T; L^2) \cap L^2(0, T; H^1), \qquad \mathbf{u} \in L^\infty(0, T; H^1). \tag{C.6}$$

The discrete in time approximations of the above three norms are used to measure the finite element error.

The asymptotic order of convergence in the limit $\Delta t \to 0$ is determined by the order of the approximating spaces, but is limited by the maximal regularity of the true solution and its time derivatives. Let $q,\ r,\ s$ and t be largest positive real numbers such that

$$\begin{aligned} &p \in L^2(0,T;H^q), \quad p \in L^\infty(0,T;H^r), \\ &p_t \in L^2(0,T;H^s), \quad \mathbf{u}_t \in L^2(0,T;H^t). \end{aligned} \tag{C.7}$$

For the same reasons as in the continuous in time case, it is assumed that $\mathbf{z} \in H(div)\ \forall t \in [0,T]$ and $q > \frac{3}{2}$. Also, the simplifying time regularity assumption is made that the exact solutions satisfy the discrete in time analogues of the regularity in (C.6)-(C.7) for any time grid, and that the higher order time derivatives of the exact solution are sufficiently regular in order to take the Taylor expansion in time at will.

The elliptic projection is used here, so the auxiliary displacement error notation is $E^A_{\mathbf{u}} \equiv \tilde{P}\mathbf{u} - \bar{\mathbf{u}}$, and the interpolation displacement error is $E^I_{\mathbf{u}} \equiv \mathbf{u} - \tilde{P}\mathbf{u}$. The corresponding notation for the flow variables remains as in Chapter 3.

Theorem C.3.1 (Auxiliary Discrete Time Error Estimate)**.** *Let $r_1 \geq$ be associated with the degree of the mixed space $(W_h, \mathbf{S}_h)$, and let $r_2 > 0$ be the degree of the polynomials used in the displacement space $\mathbf{V}_h$. Let $\Delta t > 0$ be a constant, given time-step. Then, assuming c_o is a positive constant, the regularity in (C.6)-(C.7), sufficient time regularity, and $\theta \in (0,1]$, the following holds*

$$||E^A_{\mathbf{u}}||_{l^\infty(H^1)} + ||E^A_p||_{l^\infty(L^2)} + ||E^A_{\mathbf{z}}||_{l^2(L^2)} \leq Ch^R + O(\Delta t).$$

Here $C = C(T, \alpha, c_o, \boldsymbol{\kappa}, C_{coer}, p, p_{,t}, \mathbf{z}, \mathbf{u}_{,t})$ *and*

$$R = \min\{r_1 + 1,\ r_2,\ q - 1,\ r,\ s,\ t - 1\}.$$

Remark C.3.1. The case where $\theta = 0$ is not considered here. Corresponding to the Crank-Nicolson scheme, one might expect the usual second order convergence in time for this time-stepping scheme.

Proof. Plug the true solution $(\mathbf{u}, \mathbf{z}, p)$ at time $t = t_{j,\theta}$ into the continuous in time, spatially discrete scheme (3.16)-(3.18). Utilize (C.4)-(C.5) and the other remarks concerning the Taylor series expansion. The following then holds up to order $O(\Delta t^2)$

$$a_{\mathbf{u}}(\mathbf{u}_{j,\theta}, \mathbf{v}_h) - \alpha(\nabla \cdot \mathbf{v}_h, p_{j,\theta}) = l_{1j,\theta}(\mathbf{v}_h), \tag{C.8}$$

$$\begin{aligned} c_o\Big(\frac{p_{j+1} - p_j}{\Delta t}, w_h\Big) + \alpha\Big(\nabla \cdot \Big(\frac{\mathbf{u}_{j+1} - \mathbf{u}_j}{\Delta t}\Big), w_h\Big) \\ +(\nabla \cdot \mathbf{z}_{j,\theta}, w_h) \end{aligned}$$

$$= l_{2j\theta}(w_h) + c_o \Delta t(\rho_{p,j,\theta}, w_h) + \alpha \Delta t(\nabla \cdot \rho_{\mathbf{u},j,\theta}, w_h), \tag{C.9}$$

$$(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} \mathbf{z}_{j,\theta}, \mathbf{s}_h) - (p_{j,\theta}, \nabla \cdot \mathbf{s}_h) = l_{3j,\theta}(\mathbf{s}_h). \tag{C.10}$$

Next, subtracting the first set of equations from the second, one finds a set of equations reminiscent of Galerkin orthogonality (except for the time-derivative errors):

$$a_{\mathbf{u}}(\mathbf{u}_{j,\theta} - \bar{\mathbf{u}}_{j,\theta}, \mathbf{v}_h) - \alpha(\nabla \cdot \mathbf{v}_h, p_{j,\theta} - \bar{p}_{j,\theta}) = 0, \tag{C.11}$$

$$c_o\Big(\frac{(p_{j+1} - \bar{p}_{j+1}) - (p_j - \bar{p}_j)}{\Delta t}, w_h\Big)$$
$$+\alpha\Big(\nabla \cdot \Big(\frac{(\mathbf{u}_{j+1} - \bar{\mathbf{u}}_{j+1}) - (\mathbf{u}_j - \bar{\mathbf{u}}_j)}{\Delta t}\Big), w_h\Big)$$
$$+(\nabla \cdot (\mathbf{z}_{j,\theta} - \bar{\mathbf{z}}_{j,\theta}), w_h)$$
$$= c_o \Delta t(\rho_{p,j,\theta}, w_h) + \alpha \Delta t(\nabla \cdot \rho_{\mathbf{u},j,\theta}, w_h), \tag{C.12}$$

$$(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}(\mathbf{z}_{j,\theta} - \bar{\mathbf{z}}_{j,\theta}), \mathbf{s}_h) - (p_{j,\theta} - \bar{p}_{j,\theta}, \nabla \cdot \mathbf{s}_h) = 0. \tag{C.13}$$

To continue, set $w_h = E^A_{p\ j,\theta}$ and $\mathbf{s}_h = E^A_{\mathbf{z}\ j,\theta}$ in (C.12)-(C.13). Then, after adding and subtracting the appropriate projection operators, (C.12) and (C.13) become

$$c_o\left(\frac{(E^I_{p\,j+1} + E^A_{p\ j+1}) - (E^I_{p\,j} - E^A_{p\ j})}{\Delta t}, E^A_{p\ j,\theta}\right)$$
$$+\alpha\Big(\nabla \cdot \left(\frac{(E^I_{\mathbf{u}\,j+1} + E^A_{\mathbf{u}\ j+1}) - (E^I_{\mathbf{u}\,j} + E^A_{\mathbf{u}\ j})}{\Delta t}\right), E^A_{p\ j,\theta}\Big)$$
$$+(\nabla \cdot (E^I_{\mathbf{z}\,j,\theta} + E^A_{\mathbf{z}\ j,\theta}), E^A_{p\ j,\theta})$$
$$= c_o \Delta t(\rho_{p,j,\theta}, E^A_{p\ j,\theta}) + \alpha \Delta t(\nabla \cdot \rho_{\mathbf{u},j,\theta}, E^A_{p\ j,\theta}), \tag{C.14}$$

$$(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}(E^I_{\mathbf{z}\,j,\theta} + E^A_{\mathbf{z}\ j,\theta}), E^A_{\mathbf{z}\ j,\theta}) - (E^I_{p\,j,\theta} + E^A_{p\ j,\theta}, \nabla \cdot E^A_{\mathbf{z}\ j,\theta}) = 0. \tag{C.15}$$

By first noting orthogonality relationships of the projection operators, sum (C.14) and (C.15) to get

$$
\begin{aligned}
&c_o\left(\frac{(E^A_{p\,j+1}-E^A_{p\,j})}{\Delta t},E^A_{p\,j,\theta}\right)\\
+\alpha\left(\nabla\cdot\left(\frac{(E^I_{\mathbf{u}j+1}+E^A_{\mathbf{u}\,j+1})-(E^I_{\mathbf{u}j}+E^A_{\mathbf{u}\,j})}{\Delta t}\right),E^A_{p\,j,\theta}\right)&\\
+(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}(E^I_{\mathbf{z}\,j,\theta}+E^A_{\mathbf{z}\,j,\theta}),E^A_{\mathbf{z}\,j,\theta})&\\
=c_o\Delta t(\rho_{p,j,\theta},E^A_{p\,j,\theta})+\alpha\Delta t(\nabla\cdot\rho_{\mathbf{u},j,\theta},E^A_{p\,j,\theta}).&\qquad\text{(C.16)}
\end{aligned}
$$

Similarly, let $\mathbf{v}_h=\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}$, and use the properties of the elliptic projection, equation (C.11) becomes

$$
\begin{aligned}
&a_{\mathbf{u}}\left(E^A_{\mathbf{u}\,j,\theta},\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}\right)-\alpha\left(\nabla\cdot\left(\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}\right),E^I_{p\,j,\theta}\right)\\
&\qquad-\alpha\left(\nabla\cdot\left(\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}\right),E^A_{p\,j,\theta}\right)=0.\qquad\text{(C.17)}
\end{aligned}
$$

Summing equations (C.16) and (C.17) and rearranging, one finds an equation for the auxiliary error,

$$
\begin{aligned}
&\left[a_{\mathbf{u}}\left(E^A_{\mathbf{u}\,j,\theta},\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}\right)+c_o\left(E^A_{p\,j,\theta},\frac{E^A_{p\,j+1}-E^A_{p\,j}}{\Delta t}\right)\right.\\
&\qquad\left.(\tilde{\boldsymbol{\kappa}}^{-\frac{\mathbf{1}}{\mathbf{2}}}E^A_{\mathbf{z}\,j,\theta},\tilde{\boldsymbol{\kappa}}^{-\frac{\mathbf{1}}{\mathbf{2}}}E^A_{\mathbf{z}\,j,\theta})\right]\\
=\alpha\left(\nabla\cdot\left(\frac{E^A_{\mathbf{u}\,j+1}-E^A_{\mathbf{u}\,j}}{\Delta t}\right),E^I_{p\,j,\theta}\right)-\alpha\left(\nabla\cdot\left(\frac{E^I_{\mathbf{u}j+1}-E^I_{\mathbf{u}j}}{\Delta t}\right),E^A_{p\,j,\theta}\right)&\\
-(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}E^I_{\mathbf{z}\,j,\theta},E^A_{\mathbf{z}\,j,\theta})+c_o\Delta t(\rho_{p,j,\theta},E^A_{p\,j,\theta})&\\
+\alpha\Delta t(\nabla\cdot\rho_{\mathbf{u},j,\theta},E^A_{p\,j,\theta}).&
\end{aligned}
$$

To provide bounds for the above, start with the following easily proved inequalities:

$$\begin{aligned}
a_{\mathbf{u}}\left(E^A_{\mathbf{u}\,j,\theta}, \frac{E^A_{\mathbf{u}\,j+1} - E^A_{\mathbf{u}\,j}}{\Delta t}\right) &= a_{\mathbf{u}}\left(\frac{1+\theta}{2}E^A_{\mathbf{u}\,j+1} - \frac{1-\theta}{2}E^A_{\mathbf{u}\,j}, \frac{E^A_{\mathbf{u}\,j+1} - E^A_{\mathbf{u}\,j}}{\Delta t}\right) \\
&= \frac{1+\theta}{2\Delta t}a_{\mathbf{u}}(E^A_{\mathbf{u}\,j+1}, E^A_{\mathbf{u}\,j+1}) - \frac{1+\theta}{2\Delta t}a_{\mathbf{u}}(E^A_{\mathbf{u}\,j+1}, E^A_{\mathbf{u}\,j}) \\
&\quad + \frac{1-\theta}{2\Delta t}a_{\mathbf{u}}(E^A_{\mathbf{u}\,j}, E^A_{\mathbf{u}\,j+1}) - \frac{1-\theta}{2\Delta t}a(E^A_{\mathbf{u}\,j}, E^A_{\mathbf{u}\,j}) \\
&= \frac{1}{2\Delta t}||E^A_{\mathbf{u}\,j+1}||^2_{a_{\mathbf{u}}} - \frac{1}{2\Delta t}||E^A_{\mathbf{u}\,j}||^2_{a_{\mathbf{u}}} \\
&\quad + \frac{1}{2\Delta t}\theta(||E^A_{\mathbf{u}\,j+1}||^2_{a_{\mathbf{u}}} - 2a(E^A_{\mathbf{u}\,j+1}, E^A_{\mathbf{u}\,j}) \\
&\quad + ||E^A_{\mathbf{u}\,j}||^2_{a_{\mathbf{u}}}) \\
&= \frac{1}{2\Delta t}||E^A_{\mathbf{u}\,j+1}||^2_{a_{\mathbf{u}}} - \frac{1}{2\Delta t}||E^A_{\mathbf{u}\,j}||^2_{a_{\mathbf{u}}} \\
&\quad + \frac{1}{2\Delta t}\theta||E^A_{\mathbf{u}\,j+1} - E^A_{\mathbf{u}\,j}||^2_{a_{\mathbf{u}}} \\
&\geq \frac{1}{2\Delta t}(||E^A_{\mathbf{u}\,j+1}||^2_a - ||E^A_{\mathbf{u}\,j}||^2_{a_{\mathbf{u}}}).
\end{aligned}$$

Likewise,

$$\left(E^A_{p\,j,\theta}, \frac{E^A_{p\,j+1} - E^A_{p\,j}}{\Delta t}\right) \geq \frac{1}{2\Delta t}(||E^A_{p\,j+1}||^2_0 - ||E^A_{p\,j}||^2_0).$$

With the above identities, multiply the auxiliary error by $2\Delta t$ and then sum from 0 to $N-1$ to get:

$$(||E^A_{\mathbf{u}\,N}||^2_{a_\mathbf{u}} - \overbrace{||E^A_{\mathbf{u}\,0}||^2_{a_\mathbf{u}}}^{=0}) + c_o(||E^A_{p\,N}||^2_0 - \overbrace{||E^A_{p\,0}||^2_0}^{=0})$$
$$+2\sum_{j=0}^{N-1} ||\tilde{\boldsymbol{\kappa}}^{-\frac{\mathbf{1}}{\mathbf{2}}} E^A_{\mathbf{z}\,j,\theta}||^2_0 \Delta t$$
$$\leq \Psi_1 + \Psi_2 + \Psi_3 + \Psi_4 + \Psi_5, \tag{C.18}$$

where

$$\Psi_1 = 2\alpha \sum_{j=0}^{N-1} \left(\nabla \cdot \left(\frac{E^A_{\mathbf{u}\,j+1} - E^A_{\mathbf{u}\,j}}{\Delta t} \right), E^I_{p\,j,\theta} \right) \Delta t, \tag{C.19}$$

$$\Psi_2 = -2\alpha \sum_{j=0}^{N-1} \left(\nabla \cdot \left(\frac{E^I_{\mathbf{u}j+1} - E^I_{\mathbf{u}j}}{\Delta t} \right), E^A_{p\,j,\theta} \right) \Delta t, \tag{C.20}$$

$$\Psi_3 = -2 \sum_{j=0}^{N-1} (\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}} E^I_{\mathbf{z}\,j,\theta}, E^A_{\mathbf{z}\,j,\theta}) \Delta t, \tag{C.21}$$

$$\Psi_4 = 2c_o \sum_{j=0}^{N-1} (\Delta t \rho_{p,j,\theta}, E^A_{p\,j,\theta}) \Delta t, \tag{C.22}$$

$$\Psi_5 = 2\alpha \sum_{j=0}^{N-1} (\Delta t \nabla \cdot \rho_{\mathbf{u},j,\theta}, E^A_{p\,j,\theta}) \Delta t, \tag{C.23}$$

and reminding, as in the proof of continuous in time error estimates, it is assumed that $||E^A_{\mathbf{u}\,0}||^2_{a_\mathbf{u}} = ||E^A_{p\,0}||^2_0 = 0$.

In order to bound Ψ_1 utilize discrete integration by parts formula for grid functions f_j and g_j, $\sum_{j=0}^{N-1} g_j \frac{f_{j+1}-f_j}{\Delta t} \Delta t = (f_N g_N - f_0 g_0) - \sum_{j=0}^{N-1} f_{j+1} \frac{g_{j+1}-g_j}{\Delta t} \Delta t$. Also, use the approximation, $\frac{E^I_{p\,j+1,\theta} - E^I_{p\,j,\theta}}{\Delta t} \approx (E^I_{p\,j,\theta})_t + O(\Delta t)$.

$$
\begin{aligned}
\Psi_1 &= 2\alpha \sum_{j=0}^{N-1} \left(\nabla \cdot \left(\frac{E^A_{\mathbf{u}\,j+1} - E^A_{\mathbf{u}\,j}}{\Delta t} \right), E^I_{p\,j,\theta} \right) \Delta t \\
&= 2\alpha \Big[(\nabla \cdot E^A_{\mathbf{u}\,N}, E^I_{p\,N,\theta}) - (\overbrace{\nabla \cdot E^A_{\mathbf{u}\,0}}^{=\,0}, E^I_{p\,0,\theta}) \Big] \\
&\quad -2\alpha \sum_{j=0}^{N-1} \left(\nabla \cdot E^A_{\mathbf{u}\,j+1}, \frac{E^I_{p\,j+1,\theta} - E^I_{p\,j,\theta}}{\Delta t} \right) \Delta t \\
&\leq 2\alpha ||\nabla \cdot E^A_{\mathbf{u}\,N}||_0 ||E^I_{p\,N.\theta}||_0 - 2\alpha \sum_{j=0}^{N-1} ||\nabla \cdot E^A_{\mathbf{u}\,j+1}||_0 ||(E^I_{p\,j,\theta})_t||_0 \Delta t + O(\Delta t^2) \\
&\leq \epsilon ||E^A_{\mathbf{u}\,N}||_1^2 + C||E^I_{p\,N,\theta}||_0^2 \\
&\quad +\alpha \sum_{j=0}^{N} ||E^A_{\mathbf{u}\,j}||_1^2 \Delta t + \alpha \sum_{j=0}^{N-1} ||(E^I_{p\,j,\theta})_t||_0^2 \Delta t + O(\Delta t^2) \\
&\leq \epsilon ||E^A_{\mathbf{u}\,N}||_1^2 + C||E^I_{p\,N,\theta}||_0^2 \\
&\quad +\alpha \sum_{j=0}^{N} ||E^A_{\mathbf{u}\,j}||_1^2 \Delta t + \alpha \sum_{j=0}^{N} ||(E^I_{p_j})_t||_0^2 \Delta t + O(\Delta t^2).
\end{aligned}
$$

For (C.20), likewise approximate the difference with a derivative to discover the bound

$$
\begin{aligned}
\Psi_2 &= -2\alpha \sum_{j=0}^{N-1} \left(\nabla \cdot \left(\frac{E^I_{\mathbf{u}j+1} - E^I_{\mathbf{u}j}}{\Delta t} \right), E^A_{p\,j,\theta} \right) \Delta t \\
&\quad 2\alpha \sum_{j=0}^{N-1} ||(\nabla \cdot E^I_{\mathbf{u}j})_t||_0 ||E^A_{p\,j,\theta}||_0 \Delta t + O(\Delta t^2) \\
&\leq \alpha \sum_{j=0}^{N} ||(E^I_{\mathbf{u}j})_t||_1^2 \Delta t + \alpha \sum_{j=0}^{N-1} ||E^A_{p\,j,\theta}||_0^2 \Delta t + O(\Delta t^2) \\
&\leq \alpha \sum_{j=0}^{N} ||(E^I_{\mathbf{u}j})_t||_1^2 \Delta t + \alpha \sum_{j=0}^{N} ||E^A_{p\,j}||_0^2 \Delta t + O(\Delta t^2).
\end{aligned}
$$

To bound (C.21), use the assumption that $\boldsymbol{\kappa}$ is bounded below to discover

$$\begin{aligned}
\Psi_3 &= -2\sum_{j=0}^{N-1}(\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}E^I_{\mathbf{z}\,j,\theta}, E^A_{\mathbf{z}\ j,\theta})\Delta t \\
&\leq 2\sum_{j=0}^{N-1}||\tilde{\boldsymbol{\kappa}}^{-\mathbf{1}}E^I_{\mathbf{z}\,j,\theta}||_0||E^A_{\mathbf{z}\ j,\theta}||_0\Delta t \\
&\leq \frac{C}{\lambda^2_{min}}\sum_{j=0}^{N}||E^I_{\mathbf{z}\,j}||_0^2\Delta t + \epsilon\sum_{j=0}^{N-1}||E^A_{\mathbf{z}\ j,\theta}||_0^2\Delta t.
\end{aligned}$$

The bounds for (C.22) and (C.23) are found by simple use of the Cauchy-Schwarz and Young inequalities:

$$\begin{aligned}
\Psi_4 &= 2c_o\sum_{j=0}^{N-1}(\Delta t\rho_{p,j,\theta}, E^A_{p\ j,\theta})\Delta t \\
&\leq 2c_o\sum_{j=0}^{N-1}||\Delta t\rho_{p,j,\theta}||_0||E^A_{p\ j,\theta}||_0\Delta t \\
&\leq c_o\Delta t^2\sum_{j=0}^{N-1}||\rho_{p,j,\theta}||_0^2\Delta t + c_o\sum_{j=0}^{N}||E^A_{p\ j}||_0^2\Delta t,
\end{aligned}$$

$$\begin{aligned}
\Psi_5 &= 2\alpha\sum_{j=0}^{N-1}(\Delta t\nabla\cdot\rho_{\mathbf{u},j,\theta}, E^A_{p\ j,\theta})\Delta t \\
&\leq 2\alpha\sum_{j=0}^{N-1}||\Delta t\nabla\cdot\rho_{\mathbf{u},j,\theta}||_0||E^A_{p\ j,\theta}||_0\Delta t \\
&\leq \alpha\Delta t^2\sum_{j=0}^{N-1}||\rho_{\mathbf{u},j,\theta}||_1^2\Delta t + \alpha\sum_{j=0}^{N}||E^A_{p\ j}||_0^2\Delta t.
\end{aligned}$$

Utilizing the bounds for Ψ_1 to Ψ_5, the coercivity of $a_{\mathbf{u}}$ and the boundedness of $\boldsymbol{\kappa}$ allows one to rewrite (C.18) as

$$
\begin{aligned}
&(C_{coer}-\epsilon)||E^A_{\mathbf{u}\,N}||_1^2 + c_o||E^A_{p\,N}||_0^2 + \Big(\frac{1}{\lambda_{max}}-\epsilon\Big)\sum_{j=0}^{N}||E^A_{\mathbf{z}\,j,\theta}||_0^2\Delta t \\
&\quad\le C\Delta t\Bigg(\sum_{j=0}^{N}||E^A_{\mathbf{u}\,j}||_1^2 + \sum_{j=0}^{N}||E^A_{p\,j}||_0^2\Bigg) \\
&\qquad +C\Bigg(||E^I_{p\,N,\theta}||_0^2 + \sum_{j=0}^{N}||(E^I_{p\,j})_t||_0^2\Delta t + \sum_{j=0}^{N}||(E^I_{\mathbf{u}j})_t||_1^2\Delta t \\
&\qquad\quad + \sum_{j=0}^{N}||E^I_{\mathbf{z}\,j,\theta}||_0^2\Delta t\Bigg) + O(\Delta t^2),
\end{aligned}
$$

where $C = C(T, \alpha, c_o, \boldsymbol{\kappa})$. Then, if ϵ is small enough, all terms on the left side of the above equation are positive. Let c_{min} to be the minimum value of the left side coefficients, and set the value of each left side coefficient to c_{min}; the inequality is then preserved. Then divide both sides by c_{min} to find

$$
\begin{aligned}
&||E^A_{\mathbf{u}\,N}||_1^2 + ||E^A_{p\,N}||_0^2 + \sum_{j=0}^{N}||E^A_{\mathbf{z}\,j,\theta}||_0^2\Delta t \\
&\quad\le C\Delta t\Bigg(\sum_{j=0}^{N}||E^A_{\mathbf{u}\,j}||_1^2 + \sum_{j=0}^{N}||E^A_{p\,j}||_0^2\Bigg) \\
&\qquad +C\Bigg(||E^I_{p\,N,\theta}||_0^2 + \sum_{j=0}^{N}||(E^I_{p\,j})_t||_0^2\Delta t + \sum_{j=0}^{N}||(E^I_{\mathbf{u}j})_t||_1^2\Delta t \\
&\qquad\quad + \sum_{j=0}^{N}||E^I_{\mathbf{z}\,j,\theta}||_0^2\Delta t\Bigg) + O(\Delta t^2).
\end{aligned}
$$

Apply now the discrete version of Gronwall's lemma (see Gautschi [1997], de Bustamante Simas and Cuevas [2003]) to find

$$
\begin{aligned}
&||E^A_{\mathbf{u}\,N}||^2_1 + ||E^A_{p\;N}||^2_0 + \sum_{j=0}^{N} ||E^A_{\mathbf{z}\;j,\theta}||^2_0 \Delta t \\
&\quad \leq C(T, \alpha, c_o, \boldsymbol{\kappa}, C_{coer}) \Bigg(||E^I_{p\,N,\theta}||^2_0 + \sum_{j=0}^{N} ||(E^I_{p\,j})_t||^2_0 \Delta t \\
&\qquad + \sum_{j=0}^{N} ||(E^I_{\mathbf{u}j})_t||^2_1 \Delta t + \sum_{j=0}^{N} ||E^I_{\mathbf{z}\,j,\theta}||^2_0 \Delta t \Bigg) + O(\Delta t^2). \qquad \text{(C.24)}
\end{aligned}
$$

Since the above holds for all $j = 0, 1, ..., N$, an application of approximation results up to regularity yields the theorem. □

The auxiliary discrete in time error estimate and the triangle inequality allow one to conclude the following corollary.

Corollary C.3.2 (Discrete Time Finite Element Error Estimate)**.** *With the same conditions as in the preceeding theorem, to leading order in time the following estimate holds,*

$$||\mathbf{u} - \bar{\mathbf{u}}||_{l^\infty(H^1)} + ||p - \bar{p}||_{l^\infty(L^2)} + ||\mathbf{z} - \bar{\mathbf{z}}||_{l^2(L^2)} \leq Ch^R + O(\Delta t).$$

Here $C = C(T, \boldsymbol{\kappa}, c_o, C_{coer}, p, p_{,t}, \mathbf{z}, \mathbf{u}_{,t})$ *and*

$$R = \min\{r_1 + 1,\ r_2,\ q - 1,\ r,\ s,\ t - 1\}.$$

Remark C.3.2. The error estimate provided above is designed to be true in the asymptotic limit as $\Delta t \to 0$. Therefore, regularity needed to be reflected in the theorem because as $\Delta t \to 0$ the summations in (C.24) must be able to converge finitely to their respective Riemann integrals. Furthermore, the computational results presented earlier show that for small time steps, the discrete in time solution approximates well the continuous in time scheme, and thus suffers from irregularity which was shown to exist for certain problems.

On the other hand, with a large time step the aforementioned summations might significantly undervalue the approximation of the potentially infinite integrals by avoiding the early time large pressure gradients. Thus, better convergence rates are possible if a sufficiently large time step is used. Appendix D discusses the effect of the time step size in more detail, and presents some illuminative computational results.

Appendix D

Effect of Time Step

D.1 Time Step Error Pollution

The general error equation from the discrete in time estimates is

$$error \ \leq \ C_1 h^m + C_2 \Delta t.$$

Taking the logarithm of both sides of the inequality yields

$$\log(error) \ \leq \ \log(C_1 h^m + C_2 \Delta t).$$

When Δt is negligible (as in the results presented for Test Cases I and II with $\Delta t = 1e-6$) the continuous in time error estimates can be expected to hold and the error equation becomes

$$\log(error) \ \leq \ \log(C_1) + m \log(h).$$

An example of the effect of increasing the size of the time step is shown in Terzaghi's problem and demonstrates two effects. First, for a larger time step, the initial solution at time $t = \Delta t$ is further away from the boundary layer, and thus numerical solution does not suffer as much from any lack of regularity. As Figure D.1 shows, solutions for reasonably large time steps have greater rates of convergence (as large as 0.7198) for the larger element sizes.

A second effect of larger time steps, however, is found to negate some of the benefits of moving away from the boundary layer. For $C_1 h^r \ << \ C_2 \Delta t$,

one finds that

$$\log(error) \;\approx\; \log(C_2 \Delta t).$$

This is precisely what is seen in Figure D.1 for those cases when the time step error becomes large.

Remark D.1.1. It is also mentioned here that Mandel's problem experiences the same effects of an increased time step as in Terzaghi's problem. An increased time step temporarily increases the rate of convergence as it is influenced less from the solution's initial lack of regularity. But as the grid is further refined, the error gradually begins to reflect only the time discretization error.

D.2 Discrete versus Continuous in Time Optimality

The theoretical results presented in the main body of this dissertation are continuous in time, but spatially discrete. As such, computational results were presented for only very small time steps. In this way, the discrete in time approximations were *effectively* continuous in time, and, therefore, were affected by the boundary layers and lack of regularity. As demonstrated, this led in practice to suboptimal convergence.

For larger time steps, the discrete in time approximation is not as seriously degraded by a boundary layer. So for this situation, better convergence rates may be discovered. As an example, the CG/Mixed method is tested on Terzaghi's consolidation for the case of a null constrained specific storage coefficient value, $c_o = 0$, with a time step of $\Delta t = 1e - 6$. This was the same time step used in the results presented earlier, but as Appendix B shows, a small constrained specific storage coefficient leads to very fast time scales. Figure D.2 shows how the pressure boundary layers have dissipated by this

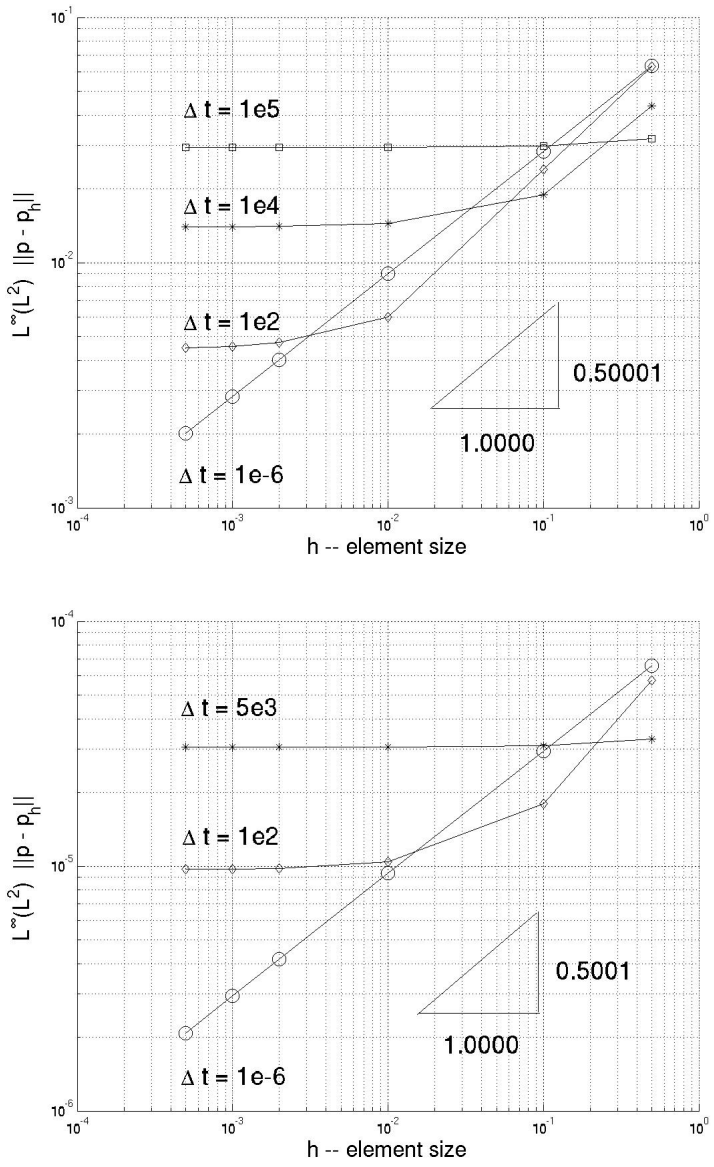

Figure D.1: Effect of the choice of time step on the $L^{\infty}(L^2)$ pressure error in Terzaghi's problem. The top graph represents the first test case and the bottom graph represents the second test case.

time. Taking advantage of this insight, Figure D.3 illustrates that this time step leads to full optimality of the CG/Mixed solution of order $O(h^1)$. Only after more refinement is the convergence polluted by the time step error.

Murad and Loula [1992] provide similar results for the same problem. They use continuous, quadratic elements for the displacement and continuous, linear elements for the pressure. They are able to show optimality (with respect to the orders of the spaces for the pressure and displacement). Their optimal convergence results are shown for a time step of $\Delta t = 5e - 5$. Additionally, their results also reflect the time error pollution which begins to occur as the grid is refined.

Alternatively, Figure D.4 shows that boundary layers are present for a much shorter time interval. A second test of the CG/Mixed method using a much smaller time step, $\Delta t = 1e - 12$, is conducted. The time step used here is small enough to be influenced by the large pressure gradients, and is, therefore, also small enough to reflect the continuous in time error estimates. Indeed, Figure D.3 shows a convergence rate of $O(h^{0.5^+})$ which conforms to theoretical predictions because $p \notin L^2(H^2)$.

The aforementioned paper by Murad and Loula [1992] does not provide results for a sufficiently small time step. So their results provide information only for the convergence rate of their numerical approximation in a time domain unaffected by irregularity.

D.3 Observations

The time error pollution demonstrated in Figure D.1 and Figure D.3 suggests that a higher order time-stepping scheme may may useful. Addition-

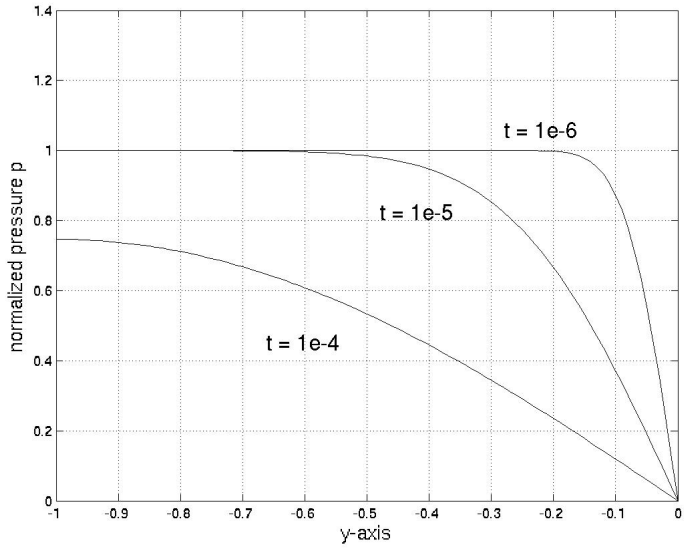

Figure D.2: Although $\Delta t = 1e-6$ appears to be a short time scale, because of the very fast time scales when $c_o = 0$, it is not. There are no large pressure gradients by this time. The domain shown above is the entire interval $[-1, 0]$.

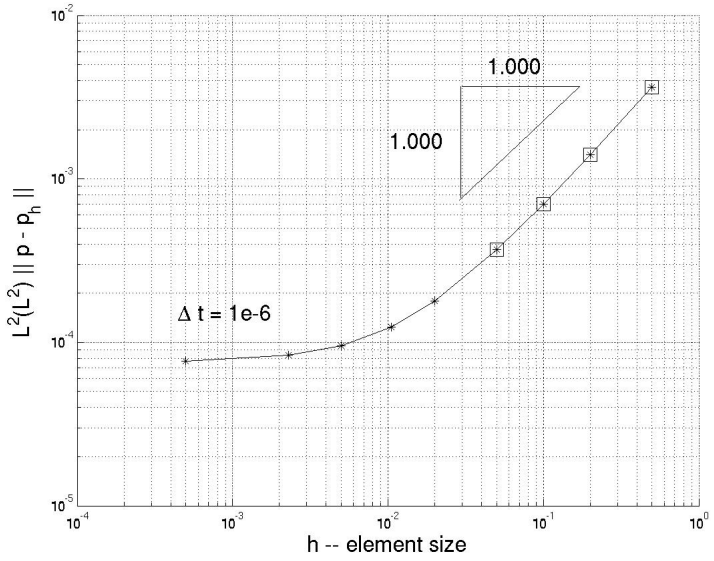

Figure D.3: Since a large time step is used, the discrete time CG/Mixed solution is not affected by the initial boundary layer. So full optimality is achieved. But for more refined grids, the time step error pollutes the convergence.

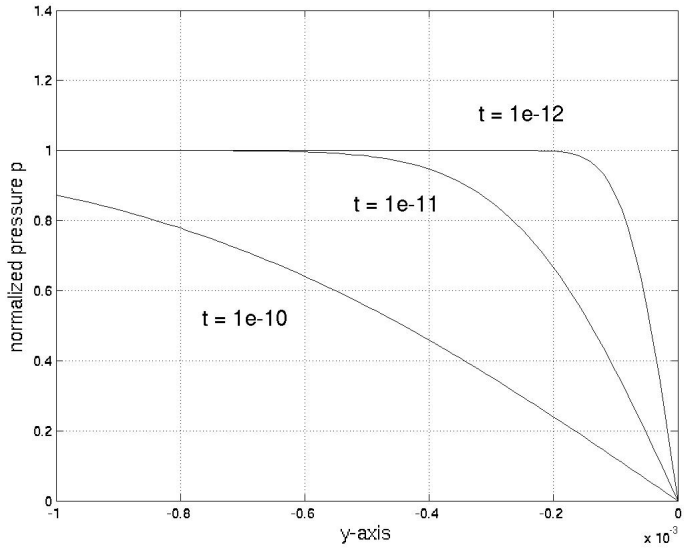

Figure D.4: The time step $\Delta t = 1e-12$ is small enough to capture the pressure boundary layer. The size of domain above is around 10^{-3}.

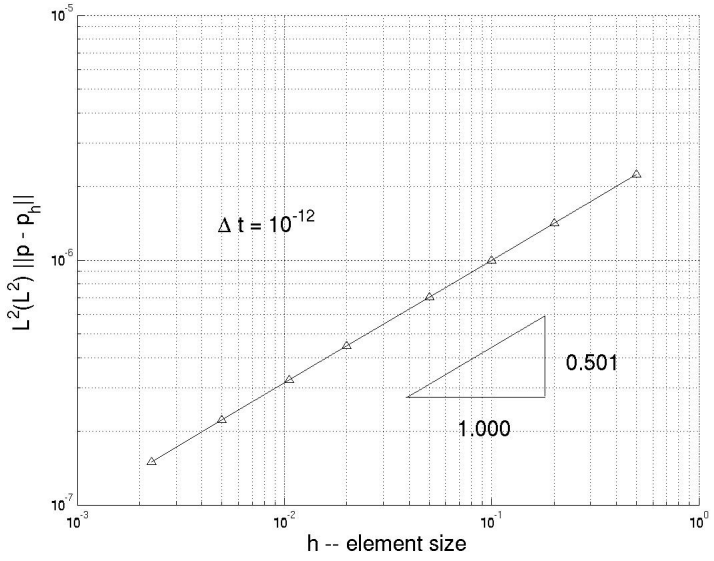

Figure D.5: For a small enough time step, the discrete-in-time CG/Mixed solution reasonably approximates the continuous-in-time solution, and, thus, is affected by the initial lack of regularity. In this case, the expected convergence rate of $O(.5^+)$ is found.

ally, if one is uninterested in a solution near a boundary layer or singularity, improvements in convergence may be possible with higher order elements. Indeed, for a time step large enough to avoid extreme pressure gradients, Murad and Loula [1992] demonstrate second order convergence in the displacement strain.

Appendix E

Effect of Reduced Quadrature

It was mentioned earlier that one of the benefits of using a mixed formulation for flow is that the flux can be eliminated through reduced integration [Weiser and Wheeler, 1988]. This produces an equivalence between the mixed formulation and a cell-centered finite difference method for the pressure.

For the problems tested earlier, there is little difference in the numerical solution whether reduced or full integration is used. However, Figure (E.1) shows that when using reduced integration the initial L^2 displacement error contains a kink unlike the case with full integration (which is always used unless specified). Otherwise, the error profiles for the pressure and displacement are quite similar irrespective of the quadrature used for the flow variable.

Certainly, this kink would stifle the convergence rate of displacement error in the $L^\infty(L^2)$ norm. But it is important to remember that error estimates are provided for the displacement error only when measured in the larger $L^\infty(H^1)$ norm.

Interestingly, Figure E.2 shows that for rectangles by using reduced quadrature the locking is eliminated, but with full quadrature the locking is present as with triangles. This differs from Barry and Mercer's problem where full and reduced quadrature both produce locking, so one cannot make generalizations with respect to quadrature and locking.

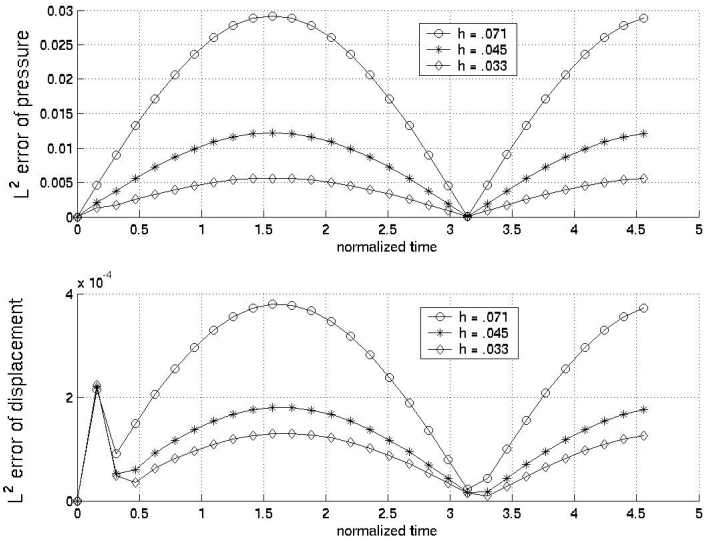

Figure E.1: CG/Mixed error in Barry and Mercer's problem with reduced quadrature. The displacement error contains a kink in the initial time step.

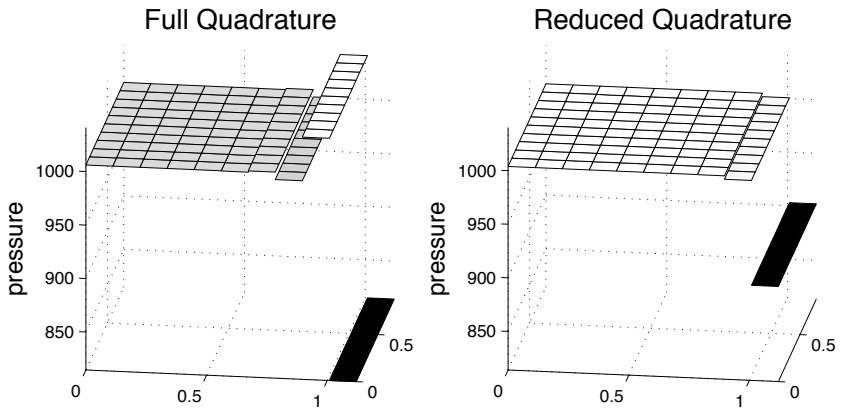

Figure E.2: For rectangles the use of full quadrature for the flow variables results in locking, while reduced quadrature eliminates it. Physical parameters include $c_o = 0$ and $E = 1e4$. Shown at time $t = 5e - 6$.

Appendix F

Existence, Uniqueness and Regularity of the Analytical Solution: A Sketch

As discussed in the Introduction, theoretical properties of solutions in linear poroelasticity have been developed in Zenisek [1984] for the case $c_o = 0$, and for the general case in Showalter [2000], who uses a semigroup approach. Existence, uniqueness and regularity are indicated here for the displacement solution, $\mathbf{u}$, and the pressure solution, p, for the equations of linear poroelasticity. This original sketch is meant to supplement the prior work in a straightforward way.

Recall the equations of linear poroelasticity,

$$-(\lambda+\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla^2\mathbf{u} + \alpha\nabla p = \mathbf{f}, \tag{F.1a}$$

$$\frac{\partial}{\partial t}(c_o p + \alpha\nabla\cdot\mathbf{u}) - \frac{1}{\mu_f}\nabla\cdot\boldsymbol{\kappa}\nabla p = s_f, \tag{F.1b}$$

with the simplified boundary and initial conditions,

$$p(t) = 0, \quad \text{on} \quad \partial\Omega, \tag{F.2a}$$

$$\mathbf{u}(t) = 0, \quad \text{on} \quad \partial\Omega, \tag{F.2b}$$

$$p(0) = 0, \quad \text{on} \quad \Omega, \tag{F.2c}$$

$$\mathbf{u}(0) = 0, \quad \text{on} \quad \Omega. \tag{F.2d}$$

Assume for the moment that c_o is a positive constant. Also assume that

$$\mathbf{f} \in L^\infty(L^2),$$
$$\mathbf{f}_t \in L^2(L^2),$$
$$s_f \in L^2(L^2).$$

1. To begin, let $\{\tilde{\mathbf{u}}_m\}$ and $\{\tilde{p}_m\}$ be orthogonal bases for $(H^1)^d$ and H^1, respectively. Then define the following Galerkin approximations:

$$\mathbf{u}_m = \sum_{k=1}^{m} a_k(t)\tilde{\mathbf{u}}_k,$$
$$p_m = \sum_{k=1}^{m} b_k(t)\tilde{p}_k.$$

2. Define the bilinear operator B_p by

$$B_p(p, w) \equiv \frac{1}{\mu_f}(\boldsymbol{\kappa}^{1/2}\nabla p, \boldsymbol{\kappa}^{1/2}\nabla w).$$

The multiplying (F.1a)-(F.1b) by appropriate test functions $\mathbf{v}$ and w, respectively, and integrating by parts, one finds the variational problem for $\mathbf{u}_m$ and p_m:

$$a_{\mathbf{u}}(\mathbf{u}_m, v) - \alpha(p_m, \nabla \cdot \mathbf{v}) = (\mathbf{f}, \mathbf{v}), \tag{F.3}$$
$$c_o(p_{m,t}, w) + \alpha(\nabla \cdot \mathbf{u}_{m,t}, w) + B_p(p_m, w) = (s_f, w). \tag{F.4}$$

Let $\mathbf{v} = \mathbf{u}_{m,t}$ and $w = p_m$. Then sum (F.3)-(F.4) to find:

$$\frac{1}{2}\frac{\partial}{\partial t}a_{\mathbf{u}}(\mathbf{u}_m,\mathbf{u}_m)+\frac{1}{2}c_o\frac{\partial}{\partial t}||p_m||_0^2+B_p(p_m,p_m)$$
$$=-(\mathbf{f}_t,u_m)+\frac{\partial}{\partial t}(\mathbf{f},u_m)+(s_f,p_m). \tag{F.5}$$

Then integrate in time from 0 to T, use Korn's inequality and the ellipticity of B_p to deduce

$$\frac{1}{2}||\mathbf{u}_m(T)||_1^2+\frac{1}{2}c_o||p_m(T)||_0^2+\int_0^T||p_m||_1^2$$
$$\leq \ -\int_0^T(\mathbf{f}_t,u_m)+(\mathbf{f}(T),u_m(T))+\int_0^T(s_f,p_m).$$

With Cauchy-Schwarz and Young's inequalities one then discovers

$$\frac{1}{2}||\mathbf{u}_m(T)||_1^2+\frac{1}{2}c_o||p_m(T)||_0^2+\int_0^T||p_m||_1^2$$
$$\leq \frac{1}{2}\int_0^T||\mathbf{f}_t||_0^2+\frac{1}{2}\int_0^T||u_m||_1^2$$
$$+\frac{C}{\epsilon}||\mathbf{f}(T)||_0^2+\epsilon||u_m(T)||_1^2$$
$$+\frac{1}{2}\int_0^T||s_f||_0^2+\frac{1}{2}\int_0^T||p_m||_0^2. \tag{F.6}$$

Then after hiding the $\epsilon||u_m(T)||_1^2$ term in the left side of the above inequality, using Gronwall's inequality helps one deduce that

$$||\mathbf{u}_m||_{L^\infty(H^1)}^2+||p_m||_{L^\infty(L^2)}^2+||p_m||_{L^2(H^1)}^2$$
$$\leq \ C\left[||\mathbf{f}_t||_{L^2(L^2)}^2+||\mathbf{f}||_{L^\infty(L^2)}^2+||s_f||_{L^2(L^2)}^2\right], \tag{F.7}$$

which holds for all $m \geq 1$.

3. Now for any finite interval $(0, T)$, it is true that $L^\infty(H^1) \subset L^2(H^1)$. Thus (F.7) implies that $\{\mathbf{u}_m\}$ and $\{p_m\}$ are bounded sequences in $L^2(H^1)$, and as such there exists subsequences $\{\mathbf{u}_{ml}\}$ and $\{p_{ml}\}$ and functions $\mathbf{u} \in L^2(H^1)$ and $p \in L^2(H^1)$ such that

$$\mathbf{u}_{ml} \rightharpoonup \mathbf{u},$$
$$p_{ml} \rightharpoonup p,$$

weakly in $L^2(H^1)$. One then deduces that inequality (F.7) holds for $\mathbf{u}$ and p

$$\begin{aligned} &||\mathbf{u}||^2_{L^\infty(H^1)} + ||p||^2_{L^\infty(L^2)} + ||p||^2_{L^2(H^1)} \\ &\quad \leq \quad C\Big[||\mathbf{f}_t||^2_{L^2(L^2)} + ||\mathbf{f}||^2_{L^\infty(L^2)} + ||s_f||^2_{L^2(L^2)}\Big]. \end{aligned} \tag{F.8}$$

Then, provided existence and uniqueness, (F.8) provides a regularity estimate in terms of the data.

4. To complete the sketch, one would now show that $\mathbf{u}$ and p satisfy the variational inequality for any $\mathbf{v} = \mathbf{u}_m$ and $w = p_m$, and use a density argument. One would also show that u and p satisfy the boundary and initial conditions. This would establish existence.

Uniqueness is established simply by assuming that $\mathbf{f} = h = 0$ in (F.3)-(F.4). Then set $\mathbf{v} = \frac{\partial \mathbf{u}}{\partial t}$ and $w = p$, and add the two equations to find

$$\frac{1}{2}\frac{\partial}{\partial t} a_{\mathbf{u}}(\mathbf{u}, \mathbf{u}) + \frac{1}{2} c_o \frac{\partial}{\partial t} ||p||^2_0 + B_p(p, p) = 0.$$

Integrating from 0 to T yields

$$\frac{1}{2}a_{\mathbf{u}}(\mathbf{u}(T),\mathbf{u}(T)) + \frac{1}{2}c_o||p(T)||_0^2 + \int_0^T B_p(p,p) = 0. \tag{F.9}$$

Then from Korn's inequality and the coercivity of B_p, (F.9) implies that $\mathbf{u} = p = 0$ at each time T.

5. With $c_o = 0$, the proof goes through somewhat similarly. However, instead of (F.6), Young's inequality with epsilon is used for the term involving the source.

$$\begin{aligned}
&\frac{1}{2}||\mathbf{u}_m(T)||_1^2 + \int_0^T ||p_m||_1^2 \\
&\quad\leq \frac{1}{2}\int_0^T ||\mathbf{f}_t||_0^2 + \frac{1}{2}\int_0^T ||u_m||_1^2 \\
&\qquad + \frac{C}{\epsilon}||\mathbf{f}(T)||_0^2 + \epsilon||u_m(T)||_1^2 \\
&\qquad + \frac{C}{\epsilon}\int_0^T ||s_f||_0^2 + \epsilon\int_0^T ||p_m||_1^2,
\end{aligned}$$

which leads to the regularity estimate

$$\begin{aligned}
&||\mathbf{u}||_{L^\infty(H^1)}^2 + ||p||_{L^2(H^1)}^2 \\
&\quad\leq C\left[||\mathbf{f}_t||_{L^2(L^2)}^2 + ||\mathbf{f}||_{L^\infty(L^2)}^2 + ||s_f||_{L^2(L^2)}^2\right].
\end{aligned} \tag{F.10}$$

6. If the quasi-static assumption is dropped so that the $\rho\mathbf{u}_{tt}$ is present in the momentum equation (F.1a), a similar result ensues. Assuming a null

initial condition for $\mathbf{u}_t$, the foregoing analysis is modified by adding $\frac{1}{2}\frac{\partial}{\partial t}||\mathbf{u}_{mt}||_0^2$ to equation (F.5).

This would result in the regularity estimate

$$\begin{aligned} &||\mathbf{u}||^2_{L^\infty(H^1)} + ||\mathbf{u}_t||^2_{L^\infty(L^2)} + ||p||^2_{L^\infty(L^2)} + ||p||^2_{L^2(H^1)} \\ &\leq \quad C\Bigg[||\mathbf{f}_t||^2_{L^2(L^2)} + ||\mathbf{f}||^2_{L^\infty(L^2)} + ||s_f||^2_{L^2(L^2)}\Bigg]. \end{aligned} \tag{F.11}$$

Remark F.0.1. The preceding sketch is missing some aspects in the proof of existence and it assumes only null Dirichlet boundary and initial conditions, but it nonetheless provides some confirmation of the regularity difficulties encountered earlier. In particular, the suggestion, again, that $p \in L^2(H^1)$ but no better reminds one of the regularity properties shown of Terzaghi's and Mandel's problem.

Alternatively, the sketch also shows more clearly the type of degeneracy which may be caused in the case of a null constrained specific storage coefficient value, $c_o = 0$. The lack of an $L^\infty(L^2)$ in (F.10) reflects the possibility that $||p(t)||_0$ may be infinite for short times, and thus lead to numerical problems as well.

Bibliography

Y. Abousleiman, A.H.-D. Cheng, L. Cui, E. Detournay, and J.C. Roegiers. Mandel's problem revisited. *Géotechnique*, 46:187–195, 1996.

M. Ainsworth and J. T. Oden. A posteriori error estimation in finite element analysis. *Comput. Methods Appl. Mech. Engrg.*, 142:1–88, 1997.

J. Alberty, C. Carstensen, S.A. Funken, and R. Klose. *Matlab Implementation of the Finite Element Method in Elasticity*, 2002.

I. Babuska. Error-bounds for the finite element method. *Numer. Math.*, 16: 322–333, 1971.

I. Babuska and M. Suri. The h-p version of the finite element method with quasi-uniform meshes. *Mathematical Modeling and Numerical Analysis*, 21 (2):199 –238, 1987.

P. Baggio, C. Bonacina, and B.A. Schrefler. Some considerations on modeling heat and mass transfer in porous media. *Transport in Porous Media*, 28: 233–251, 1997.

S.I. Barry and G.N. Mercer. Flow and deformation in poroelasticity - I unusual exact solutions. *Mathematical and Computer Modeling*, 30:23–29, 1999a.

S.I. Barry and G.N. Mercer. *Exact Solutions for Two-Dimensional Time Dependent Flow and Deformation Within a Poroelastic Medium*, July 1999b.

G.A. Behie, A. Settari, and D.A. Walters. Use of coupled reservoir and geomechanical modeling for integrated reservoir analysis and management. Technical report, Canadian International Petroleum Conference 2000, Calgary, Alberta, Canada, June 4-8 2000.

M.A. Biot. General theory of three-dimensional consolidation. *Journal of Applied Physics*, 2:155–164, 1941.

M.A. Biot. Theory of elasticity and consolidation for a porous anisotropic media. *Journal of Applied Physics*, 26(2):182–185, 1955.

J.R. Booker and J.C. Small. An investigation of the stability of numerical solution of Biot's equations of consolidation. *Int. J. Solids Structures*, 2: 907–917, 1975.

S.C. Brenner. Korn's inequality for piecewise H^1 vector fields. *Industrial Mathematics Institute Preprint Series*, 2002.

S.C. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer-Verlag, New York, 1994.

F. Brezzi. On the existence, uniqueness and approximation of saddle-point problems arising from Langrangian multipliers. *RAIRO Anal. Numer.*, 8: 129–151, 1974.

M.S. Bruno. Geomechanical analysis and decision analysis for mitigating compaction related casing damage. *Society of Petroleum Engineers*, 2001.

H. Bryne and L. Preziosi. Modeling tumour growth using the theory of mixtures. *Mathematical Medicine and Biology*, 2003.

J.T. Christian. Undrained stress distribution by numerical methods. *J. Soil Mech. Found. Div., ASCE*, 94:1333–1345, 1968.

P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North-Holland, Amsterdam, 1976.

A. Cividini and A.Z. Rossi. The consolidation problem treated by a consistent (static) finite element approach. *Int. J. Numer. Anal. Methods Geomech.*, 7:435–455, 1983.

O. Coussy. *Poromechanics*. Wiley, 2004.

Alexandre de Bustamante Simas and Claudio Cuevas. Gronwalls inequalities on differential and difference equations. 2003.

E. Detournay and A.H.-D. Cheng. *Chapter 5 in Comprehensive Rock Engineering: Priciples, Practice and Projects, Vol. II, Analysis and Design Method*, chapter Fundamentals of Poroelasticity, pages 113–171. Pergamon Press, 1993.

E. Detournay and A.H.-D. Cheng. Poroelastic response of a borehole in non-hydrostatic stress field. *International Journal of Rock Mechanics and Mining Sciences*, 1988.

M.B. Dusseault, M.S. Bruno, and J. Barrera. Casing shear: Causes, cases, cures. *Society of Petroleum Engineers*, 2001.

L. Tuffy Ellis. The Handbook of Texas Onliine, 1999. URL `www.tsha.utexas.edu/handbook/online/`.

L.C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, R.I., 1998.

R. Ewing, O. Iliev, R. Lazarov, and A. Naumovich. On convergence of certain finite difference discretizations for 1-D poroelasticity interface problems. *Fraunhofer Institut Techno- und Wirtschaftsmathematik*, 69, 2004.

Xiuli Gai. *A Coupled Geomechanics and Reservoir Flow Model on Parallel Computers*. PhD thesis, University of Texas at Austin, 2004.

D. Garagash and E. Detournay. An analysis of the influence of the pressurization rate on the borehole breakdown pressure. *Journal of Solids and Structures*, 1997.

Walter Gautschi. *Numerical Analysis: An Introduction*. Birkhauser, Boston, 1997.

J. Ghaboussi and E.L. Wilson. Flow of compressible fluid in porous elastic solids. *Int. J. Numer. Meth. Engng.*, 5:419–442, 1973.

R.E. Gibson, K. Knight, and P.W. Taylor. A critical experiment to examine theories of three-dimensional consolidation. *Eur. Conf. Soil Mech.*, 1:69–73, 1963.

D. Gilbarg and N. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer, 2001.

V. Girault and P.-A. Raviart. *Finite Element Methods for Navier-Stokes Equations*. Springer-Verlag, Berlin, 1986.

V. Girault, B. Riviere, and M. Wheeler. A discontinuous Galerkin method with non-overlapping domain decompsition for the Stokes and Navier-Stokes problems. *Mathematics of Computation*, 2003.

P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, Boston, MA, 1985.

J. Hudson, O. Stephansson, J. Andersson, C.-F. Tsang, and L. Ling. Coupled T-H-M issues related to radioactive waste repository design and performance. *International Journal of Rock Mechanics and Mining Sciences*, 38:143–161, 2001.

Ziwen Jiang. $L^\infty(L^2)$ and $L^\infty(L^\infty)$ error estimates for mixed methods for integro-differential equations of parabolic type. *Mathematical Modeling and Numerical Analysis*, 33(3):531–546, 1999.

E.F. Kaasschieter and A.J.H. Frijns. Squeezing a sponge: A three-dimensional solution in poroelasticity. *Computational Geosciences*, 7:49–59, 2003.

J.-M. Kim and R. Parizek. Numerical simulation of the Noordbergum effect resulting from groundwater pumping in a layered aquifer system. *Journal of Hydrology*, 202:231–243, 1997.

R. Kirby. Residual a posteriori error estimate for the mixed finite element method. 2003.

Terri Langford. Northwest Houston sinking faster than coastal areas. *Reporter-News.com*, Aug. 28 1997. URL `http://texnews.com/texas97/sink082897.html`.

R.W. Lewis and B.A. Schrefler. *The Finite Element Method in Static and Dynamic Deformation and Consolidation of Porous Media*. Wiley, Chichester, 1987.

X. Li, X. Han, and M. Pastor. An iterative stabilized fractional step algorithm for finite element analysis in saturated soil dynamics. *Computer Methods in Applied Mechanics and Engineering*, 192:38453859, 2003.

Jian Lin. An nterview with Jian Lin, Ph. D., Nov. 2003. URL `http://www.esi-topics.com/earthquakes/interviews/JianLin.html`.

Konstantin Lipnikov. *Numerical Methods for the Biot Model in Poroelasticity*. PhD thesis, University of Houston, 2002.

Ruijie Liu. *Discontinuous Galerkin Finite Element Solution for Poromechanics*. PhD thesis, University of Texas at Austin, 2004.

N. Lubick. Modeling complex, multiphase porous media systems. *Siam News*, 5(3), April 2002. URL `http://www.siam.org/siamnews/04-02/pmedia.htm`.

J. Mandel. Consolidation des sols (étude mathématique)‘. *Géotechnique*, 30: 287–299, 1953.

F.A. Milner. Mixed finite element methods for quasilinear second order elliptic problems. *Mathematics of Computation*, 44(170):302–320, April 1985.

M.A. Murad and A.F.D. Loula. Improved accuracy in finite element analysis of Biot's consolidation problem. *Computer Methods in Applied Mechanics and Engineering*, 95:359–382, 1992.

M.A. Murad, V. Thomée, and A.F.D. Loula. Asymptotic behavior of semidiscrete finite element approximations of biot's consolidation problem. *SIAM J. Numer. Anal.*, 33:1065–1083, 1996.

E.-J. Park. Mixed finite element methods for nonlinear second-order elliptic problems. *SIAM J. Numer. Anal.*, 32(3):865–885, 1995.

P .J. Phillips and Mary F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelesticity I: The continuous in time case. *submitted: Computational Geosciences*, 2005a.

P. J. Phillips and Mary F. Wheeler. A coupling of mixed and discontinuous Galerkin finite element methods for poroelesticity. *submitted: Computational Geosciences*, 2005b.

P. J. Phillips and Mary F. Wheeler. A coupling of mixed and continuous Galerkin finite element methods for poroelesticity II: The discrete in time case. *submitted: Computational Geosciences*, 2005c.

P. J. Phillips and Mary F. Wheeler. Overcoming the problem of locking in linear elasticity and poroelasticity: An heuristic approach. *submitted: Computational Geosciences*, 2005d.

R. Rajapakse. Stress analysis of borehole in poroelastic medium. *Journal of Engineering Mechanics*, 119(6):1205–1227, June 1993.

R.A. Raviart and J.M. Thomas. Mixed finite element method for second order elliptic problems. *Mathematical Aspects of the Finite Element Method, Lecture Notes in Mathematics 606*, pages 292–315, 1977.

B. Rivière and M.F. Wheeler. Discontinuous galerkin method applied to nonlinear parabolic problems. In B. Cockburn, G.E. Karniadakis, and C.-W. Shu, editors, *Discontinuous Galerkin Methods: Theory, Computation, and Applications*, volume 11. Springer, lecture notes in computational science and engineering edition, 1999.

B. Rivière and M.F. Wheeler. Optimal error estimates applied to linear elasticity. Technical report, ICES Report, 2000. URL `www.ices.utexas.edu/reports/2000/index.html`.

T. Roose, P.A. Netti, L. Munn, Y. Boucher, and R. Jain. Solid stress generated by spheroid growth estimated using a linear poroelastic model. *Microvascular Research*, 66:204–212, 2003. URL `www.elsevier.com/locate/ymvre`.

K. Runesson, F. Larsson, and P. Hansbo. Space-time finite elements and adaptive strategy for the coupled poroelasticity problem. In Smith and Carter, editors, *Developments in Theoretical Geomechanics*, pages 193–213, Rotterdam, 2000. Balkema.

J. Rutqvist and C.-F. Tsang. Analysis of thermal-hydrologic-mechanical behavior near an emplacement drift at Yucca mountain. *Journal of Contaminant Hydrology*, 62-63:637–652, 2003.

R.S. Sandhu and E.L. Wilson. Finite element analysis of seepage in elastic media. *J. Eng. Mech. Div., ASCE*, 95:641–652, 1969.

R. Scholz. L_∞-convergence of saddle-point approximations for second order problems. *RAIRO Model. Math. Anal. Numer.*, 119:209–216, 1977.

C. Schwab. p- and hp-finite element. *Numerical Mathematics and Scientific Computation*, 1998.

A. Settari and D.A. Walters. Advances in coupled geomechanical and reservoir modeling with applications to reservoir compaction. Technical report, 1999 SPE Reservoir Simulation Symposium, Houston, TX, Feb. 14-17 1999.

R. E. Showalter. Diffusion in poro-elastic media. *Jour. Math. Anal. Appl.*, 251:310–340, 2000.

R. E. Showalter. Partially saturated flow in a poroelastic medium. Technical report, University of Texas; ICES, 2001.

A. Smillie, I. Sobey, and Z. Molnar. A hydro-elastic model of hydrocephalus. Technical report, Oxford University Computing Laboratory: Numerical Analysis Group, 2004.

R. Steck, P. Niederer, and M.L. Knothe Tate. A finite difference model of load-induced fluid displacements within bone under mechanical loading. *Med Eng Phys.*, 22(2):117–25, 2000.

Shuyu Sun. *Discontinuous Galerkin Methods for Reactive Transport in Porous Media*. PhD thesis, University of Texas at Austin, 2003.

C.C. Swan, R.S. Lakes, R.A. Brand, and K.J. Stewart. Micromechanically based poroelastic modeling of fluid flow in haversian bone. *Journal of Biomechanical Engineering*, 125(1):25–37, Feb. 2003.

K. Terzaghi. Die berechnung der durchlassigkeitsziffer des tones aus dem verlauf der hydrodynamischen spannungsercheinungen. *Sitzung Berichte. Akadamie der Wissenschaften, Wien Mathematiesch-Naturwissenschaftliche Klasse, Abteilung IIa.*, 132:105–124, 1923.

K. Terzaghi. Principle of soil mechanics. *Eng. News Record, A Series of Articles*, 1925.

A. Verruijt. Discussion. *Proc. 6th Int. Conf. Soil Mech.*, 3:401–402, 1965.

Jing Wan. *Stabalized Finite Element Methods for Coupled Geomechanics and Multiphase Flow*. PhD thesis, Stanford University, Petroleum Eng. Dept., November 2002.

H. F. Wang. *Theory of Linear Poroelasticity with Applications to Geomechanics and Hydrogeology*. Princeton University Press, Princeton, N.J., 2000.

Y. Wang and M. Dusseault. A coupled conductive-convective thermo-poroelastic solution and implications for wellbore stability. *Journal of Petroleum Science and Engineering*, 38:187–198, 2003.

A. Weiser and M.F. Wheeler. On convergence of block-centered finite differences and elliptic equations. *SIAM J. Numer. Anal.*, 251(2):351–375, 1988.

M. F. Wheeler. A priori L^2 error estimates for Galerkin approximations to parabolic partial differential equations. *SIAM J. Numer. Anal.*, 10:723–759, 1973.

M.F. Wheeler and I. Yotov. A posteriori error estimates for the mortar mixed finite element method. Technical report, Institute for Computational Engineering and Sciences (ICES), 2003.

Y.Q. Zeng and Q.H. Liu. A poroelastic model for acoustic landmine detection. Technical report, Duke Deminig MURI, 2000.

Y.Q. Zeng and Q.H. Liu. Acoustic landmine detection: a 3D poroelastic model. Technical report, Duke Deminig MURI, 2001.

A. Zenisek. The existence and uniqueness theorem in Biot's consolidation theory. *Aplik. Matem.*, 29:194–210, 1984.

O.C. Zienkiewicz. *Numerical Methods in Geomechanics,*, chapter Basic formulation of static and dynamic behaviour of soil and other porous media, pages 39–56. 1982.

O.C. Zienkiewicz and T. Shiomi. Dynamic behavior of saturated porous media: the generalized Biot formulation and its numerical solution. *Int. J. Numer. Anal. Methods Geomech.*, 8:71–96, 1984.

C. Zoppou, S.I. Barry, and G.N. Mercer. Dynamics of human milk extraction: A comparative study of breast-feeding and breast pumping. Feb. 24 1997.

Vita

Phillip Joseph Phillips was born in Plainview, Texas, the son of Raymond J. Phillips and Betty J. Phillips. He received the Bachelor of Arts degree *summa cum laude* in Mathematics from Rice University in Houston, Texas on May 9$^{\text{th}}$, 1998. Two years later he applied to the University of Texas at Austin for enrollment in their Computational and Applied Mathematics program. He was accepted and was awarded the departmental CAM Fellowship and the University Preemptive Fellowship. He enrolled in August, 2000, and was later awarded the Master of Science degree on May 18$^{\text{th}}$, 2002. He currently resides in Austin, Texas with his wife, Rowena B. Phillips, R.N.

Permanent address: 5702 Libyan Dr.
Austin, Texas 78745

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.