5-2009

Numerical methods for fully nonlinear second order partial differential equations

Michael Joseph Neilan
University of Tennessee

To the Graduate Council:

I am submitting herewith a dissertation written by Michael Joseph Neilan entitled "Numerical methods for fully nonlinear second order partial differential equations." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Xiaobing Feng, Major Professor

We have read this dissertation and recommend its acceptance:

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a dissertation written by Michael Joseph Neilan entitled "Numerical Methods for Fully Nonlinear Second Order Partial Differential Equations." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Mathematics.

Xiaobing Feng, Major Professor

We have read this dissertation
and recommend its acceptance:

Ohannes Karakashian

Suzanne Lenhart

Christopher Pionke

Accepted for the Council:

Carolyn R. Hodges
Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Numerical Methods for Fully Nonlinear Second Order Partial Differential Equations

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Michael Joseph Neilan
May 2009

Acknowledgments

I would like to express my sincere gratitude to my advisor, professor Xiaobing Feng, who introduced me to the subject of numerical analysis many years ago. He has provided me with many years of guidance and motivation, and if it were not for his expertise and patience, I would not have made it this far.

I would also like to thank my dissertation committee, Dr. Ohannes Karakashian, Dr. Suzanne Lenhart, and Dr. Christopher Pionke who were always willing to answer any questions I have had.

Finally, I would like to thank my family and especially my wife, Rachael, who has provided me with constant love, encouragement, and understanding.

Abstract

This dissertation concerns the numerical approximations of solutions of fully nonlinear second order partial differential equations (PDEs). The numerical methods and analysis are based on a new concept of weak solutions called moment solutions, which unlike viscosity solutions, are defined by a constructive method called the vanishing moment method. The main idea of the vanishing moment method is to approximate fully nonlinear second order PDEs by a family of fourth order quasi-linear PDEs. Because the method is constructive, we can develop a wealth of convergent numerical discretization methods to approximate fully nonlinear second order PDEs. We first study the numerical approximation of the prototypical second order fully nonlinear PDE, the Monge-Ampère equation, $\det(D^2u) = f\ (> 0)$, using C^1 finite element methods, spectral Galerkin methods, mixed finite element methods, and a nonconforming Morley finite element method. We then generalize the analysis to other fully nonlinear second order PDEs including the nonlinear balance equation, a nonlinear formulation of the semigeostrophic flow equations, and the equation of prescribed Gauss curvature.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Prelude

Fully nonlinear partial differential equations (PDEs) are those PDEs which depend nonlinearly on the highest order derivatives of unknown functions. These PDEs arise in many areas of science and engineering such as kinetic theory, materials science, differential geometry, general relativity, optimal control, mass transportation, image processing, computer vision, meteorology, and semigeostrophic fluid dynamics. In the case of second order equations, the general form is given by

$$F(D^2u, Du, u, x) = 0, \tag{1.1}$$

where $D^2u(x)$ and $Du(x)$ denote the Hessian and gradient of u at x, respectively.

Examples of such equations include (cf. [57])

- *The Monge-Ampère equation*
$$\det(D^2u) = f. \tag{1.2}$$

- *The equation of prescribed Gauss curvature*
$$\det(D^2u) = K(1 + |Du|^2)^{\frac{n+2}{2}}. \tag{1.3}$$

- *The Bellman equation*
$$\inf_{\nu \in V}(L_\nu u - f_\nu) = 0. \tag{1.4}$$

The goal of this dissertation is to develop and analyze various numerical methods to approximate the viscosity solutions of (1.1) whenever such solutions exist (cf. Definition 1.2.2). Specifically, we use the Monge-Ampère equation to develop our ideas and methods, and then generalize these results to other nonlinear PDEs.

1.2 Viscosity Solutions

Because of the full nonlinearity in (1.1), the standard weak solution theory based on the integration by parts approach does not work and other notions of weak solutions must be sought. Much progress has been made in the latter half of the 20th century concerning this issue after the introduction of viscosity solutions. In 1983, Crandall and Lions introduced the notion of viscosity solutions and used the vanishing viscosity method to show existence of a solution for the Hamilton-Jacobi equation:

$$u_t + F(Du, u, x) = 0 \qquad (x, t) \in \mathbf{R}^n \times (0, \infty). \tag{1.5}$$

The vanishing viscosity method is defined by approximating the Hamilton-Jacobi equation by the following regularized, second-order quasi-linear PDE:

$$u_t^\epsilon + F(Du^\epsilon, u^\epsilon, x) - \epsilon \Delta u^\epsilon = 0 \qquad (x, t) \in \mathbf{R}^n \times (0, \infty). \tag{1.6}$$

It was shown that there exists a unique solution u^ϵ to the regularized Cauchy problem that converges locally and uniformly to a continuous function u which is defined to be a viscosity solution of the Hamilton-Jacobi equation [30]. To establish uniqueness, the following intrinsic definition of viscosity solutions was also proposed [31]:

Definition 1.2.1. *Let $F : \mathbf{R}^n \times \mathbf{R} \times \Omega \to \mathbf{R}$ and $g : \partial\Omega \to \mathbf{R}$ be continuous functions, and consider the following problem:*

$$F(Du, u, x) = 0 \qquad \text{in } \Omega, \tag{1.7}$$

$$u = g \qquad \text{on } \partial\Omega. \tag{1.8}$$

(i) $u \in C^0(\Omega)$ is called a viscosity subsolution *of (1.7)–(1.8) if $u\big|_{\partial\Omega} = g$, and for every C^1 function $\varphi(x)$ such that $u - \varphi$ has a local maximum at $x_0 \in \Omega$, there holds*

$$F(D\varphi(x_0), \varphi(x_0), x_0) \le 0.$$

(ii) $u \in C^0(\Omega)$ is called a viscosity supersolution *of (1.7)–(1.8) if $u\big|_{\partial\Omega} = g$, and for every C^1 function $\varphi(x)$ such that $u - \varphi$ has a local minimum at $x_0 \in \Omega$, there holds*

$$F(D\varphi(x_0), \varphi(x_0), x_0) \ge 0.$$

(iii) $u \in C^0(\Omega)$ is called a viscosity solution *of (1.9)–(1.10) if it is both a viscosity subsolution and supersolution.*

Clearly, the above definition is not variational as it is based on a "differentiation by

parts" approach. In addition, the word "viscosity" loses its original meaning in the definition. However, it was shown [30, 31] that every viscosity solution constructed by the vanishing viscosity method is an intrinsic viscosity solution. A reason to favor the intrinsic differentiation by parts definition is that the definition and the notion of viscosity solutions can be readily extended to fully nonlinear second order PDEs as follows:

Definition 1.2.2. *Let $F : \mathbf{R}^{n\times n} \times \mathbf{R}^n \times \mathbf{R} \times \Omega \to \mathbf{R}$ and $g : \partial\Omega \to \mathbf{R}$ be continuous functions, and consider the following problem:*

$$F(D^2u, Du, u, x) = 0 \qquad \text{in } \Omega, \tag{1.9}$$

$$u = g \qquad \text{on } \partial\Omega. \tag{1.10}$$

(i) $u \in C^0(\Omega)$ *is called a* viscosity subsolution *of* (1.9)–(1.10) *if* $u\big|_{\partial\Omega} = g$*, and for every* C^2 *function* $\varphi(x)$ *such that* $u - \varphi$ *has a local maximum at* $x_0 \in \Omega$*, there holds*

$$F(D^2\varphi(x_0), D\varphi(x_0), \varphi(x_0), x_0) \le 0.$$

(ii) $u \in C^0(\Omega)$ *is called a* viscosity supersolution *of* (1.9)–(1.10) *if* $u\big|_{\partial\Omega} = g$*, and for every* C^2 *function* $\varphi(x)$ *such that* $u - \varphi$ *has a local minimum at* $x_0 \in \Omega$*, there holds*

$$F(D^2\varphi(x_0), D\varphi(x_0), \varphi(x_0), x_0) \ge 0.$$

(iii) $u \in C^0(\Omega)$ *is called a* viscosity solution *of* (1.9)–(1.10) *if it is both a viscosity subsolution and supersolution.*

Remark 1.2.3. *Without loss of generality, we may assume that $u(x_0) = \varphi(x_0)$ whenever $u - \varphi$ achieves a local maximum or local minimum at $x_0 \in \Omega$ in Definition 1.2.2. Therefore in an informally setting, u is a viscosity solution if for every smooth function φ that "touches" the graph of u from above at x_0, $F(D^2\varphi(x_0), D\varphi(x_0), \varphi(x_0), x_0) \le 0$, and if φ "touches" the graph of u from below at x_0, then $F(D^2\varphi(x_0), D\varphi(x_0), \varphi(x_0), x_0) \ge 0$ (see Figure 1.1).*

In the case of fully nonlinear *first* oder PDEs, tremendous progress has been made in the past three decades in both PDE theory and numerical methods. A rich PDE viscosity solution theory has been established, and a wealth of efficient and robust numerical methods and algorithms have been developed and implemented [9, 22, 28, 33, 81, 82, 83, 87]. However, in the case of fully nonlinear *second* order PDEs, the setting is remarkably different. On the one hand, viscosity solution theory has been extended to second order PDEs with great success [32, 64, 65], but on the other hand, numerical solutions for general fully nonlinear second order PDEs is mostly an untouched area.

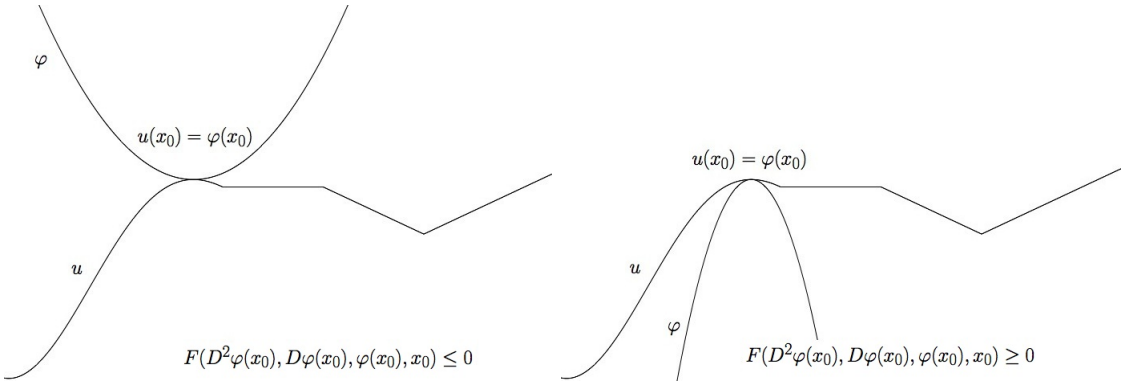


Figure 1.1: A Geometric interpretation of viscosity solutions.

There are several reasons for this lack of progress in numerical methods. First, the most obvious difficulty is the full nonlinearity in the equation. Second, solutions to fully nonlinear second order equations are often only unique in a certain class of functions, and this conditional uniqueness is very difficult to handle numerically. Lastly, the non-variational nature of viscosity solution theory poses a daunting challenge for computing these solutions because it is impossible to directly approximate viscosity solutions using any Galerkin-type numerical methods including finite element methods, spectral Galerkin methods, and discontinuous Galerkin methods, which are all based on variational formulations of PDEs. In addition, it is extremely difficult (if all possible) to mimic the differentiation by parts approach at the discrete level, so there is little hope to develop a discrete viscosity solution theory.

1.3 The Monge-Ampère Equation

The research presented in this dissertation will focus mainly on the Dirichlet problem for a prototypical fully nonlinear second order PDE, namely the Monge-Ampère equation:

$$\det(D^2u(x)) = f(x) \qquad \text{in } \Omega, \tag{1.11}$$

$$u = g \qquad \text{on } \partial\Omega, \tag{1.12}$$

where $\det(D^2u(x))$ denotes the determinant of D^2u at x. It is known that for non-strictly convex domain, Ω, the above problem does not have classical solutions in general even if f, g, and $\partial\Omega$ are smooth [57]. However, classical results of A. D. Aleksandrov state that the Dirichlet problem with $f > 0$ has a unique generalized solution in the class of convex functions [2, 60]. We note that other nonconvex solutions of (1.11)–(1.12) might exist even when $f > 0$. We also note that Monge-Ampère (1.11) is only elliptic in the class of convex

functions [57].

It is clear that equation (1.11) is of the form (1.1) with

$$F(D^2u(x), Du(x), u(x), x) = f(x) - \det(D^2u(x)).$$

Since the solution of (1.11) is only unique in the class of convex functions, the definition of the viscosity solution of (1.11) reads as follows [32]:

Definition 1.3.1. *Suppose $f \in C^0(\Omega)$ with $f > 0$ in Ω.*

(i) A convex function $u \in C^0(\Omega)$ is called a viscosity subsolution *of* (1.11)–(1.12) *if $u\big|_{\partial\Omega} = g$, and for every C^2 function $\varphi(x)$ such that $u - \varphi$ has a local maximum at $x_0 \in \Omega$, there holds*

$$\det(D^2\varphi(x_0) \geq f(x_0).$$

(i) A convex function $u \in C^0(\Omega)$ is called a viscosity supersolution *of* (1.11)-(1.12) *if $u\big|_{\partial\Omega} = g$, and for every C^2 function $\varphi(x)$ such that $u - \varphi$ has a local minimum at $x_0 \in \Omega$, there holds*

$$\det(D^2\varphi(x_0) \leq f(x_0).$$

(iii) A convex function $u \in C^0(\Omega)$ is called a viscosity solution *of* (1.11)–(1.12) *if it is both a viscosity subsolution and supersolution.*

Remark 1.3.2. *It has been shown that Aleksandrov's generalized solution of the Monge-Ampère is equivalent to the convex viscosity solution (cf. [60]).*

Noting $\det(D^2(-u)) = \det(D^2u)$ in the case n is even, we give the following notion of concave viscosity solutions for the Monge-Ampère equation.

Definition 1.3.3. *Suppose n is even, $f \in C^0(\Omega)$, and $f > 0$ in Ω. A concave function $\tilde{u} \in C^0(\Omega)$ is called a* concave viscosity solution *of* (1.11) *if $\tilde{u} := -u$, where u is the convex viscosity solution of*

$$\begin{aligned} \det(D^2u) &= f \qquad \text{in } \Omega, \\ u &= -g \qquad \text{on } \partial\Omega. \end{aligned}$$

1.4 Contributions and Related Works

The research presented in this dissertation mainly consists of results reported in [50]–[53],[78]. It also contains some new results which have not yet been reported. In [50], the vanishing moment method is introduced (cf. Chapter 2) as a platform to solve fully

nonlinear second order PDEs using Galerkin type methods. Various formulations are presented. [52] is devoted to studying both finite element and spectral Galerkin methods for the Monge-Ampère equation, [51] analyzes the Monge-Ampère equation using mixed finite elements, and [78] studies the finite element approximation of the Monge-Ampère equation using the nonconforming Morley element. [53] develops a fully discrete modified characteristic finite element method for a nonlinear formulation of the semigeostrophic flow equations. We give a detailed account of the analysis of these papers as well as give additional numerical examples, especially in three dimensions. We also develop finite element formulations to approximate other second-order fully nonlinear PDEs such as the nonlinear balance equation, and the equation of prescribed Gauss curvature.

A few results on numerical approximations of second order nonlinear PDEs have recently been reported which we now summarize. Oliker and Prussner [80] constructed a finite difference scheme for computing Aleksandrov measure induced by D^2u in 2-D and obtained the solution of problem (1.11)–(1.12) as a by-product. The scheme is very geometric and difficult to use and to generalize to other fully nonlinear second order PDEs. Baginski and Whitaker [6] introduced a finite difference scheme for the Gauss curvature equation (1.3) in 2-D by mimicking the continuation method (which is used to prove existence of the PDE) at the discrete level. Dean and Glowinski [39] presented an augmented Lagrange multiplier method and a least squares method for the Monge-Ampère equation and Pucci's equation in 2-D by treating the nonlinear equations as a constraint and using a variational principle to select a particular solution. However, it is unclear how the solutions of their methods relate to viscosity solutions. In [8], Barles and Souganidis showed that any monotone, stable, and consistent finite difference scheme converges to the viscosity solution provided that there exists a comparison principle for the limiting equation. Their results provide a guideline for constructing convergent finite difference methods, but it did not address how to construct such a scheme. Oberman [79] constructed a wide stencil difference scheme for nonlinear elliptic PDEs which can be written as functions of eigenvalues of the Hessian matrix. It was proved that the finite difference scheme satisfies the convergence criterion established by Barles and Souganidis. Finally, Böhmer [13] recently introduced a projection method using C^1 finite element functions for a certain class of fully nonlinear second order elliptic PDEs. Numerical experiments were reported in [80, 6, 79], but convergence analysis was not addressed except in [79, 13].

1.5 Applications and Impacts

Fully nonlinear second order PDEs arise in many areas of science including astrophysics, economics, shape optimization, meteorology, general relativity, and biomedical computing. Advancements in the applications of these areas largely depends on solving their underlying

equations. Despite their importance, little progress has been made in numerically solving these PDEs. Therefore, the results presented here are expected to have a significant impact on advancing many of these application areas.

Previous numerical algorithms for fully nonlinear second order PDEs are mostly heuristic methods that are tailored for specific equations. In contrast, we give a general framework to construct and numerically solve fully nonlinear second order PDEs (cf. Chapters 2, 9, and 10). We provide a new notion of weak solutions and then give a detailed exposition in constructing and analyzing various type of Galerkin-type methods. We especially give considerable attention to the Monge-Ampère equation (Chapters 3–6), the prototypical fully nonlinear second order PDE. However, we expect that the results presented in this dissertation can be generalized to a large class of fully nonlinear PDEs.

1.6 Dissertation Organization

The dissertation is organized as follows. In Chapter 2, we introduce the vanishing moment method and the notion of moment solutions for fully nonlinear second order PDEs. We summarize the findings of [49] which analyzes this method applied to the Monge-Ampère equation. In Chapters 3–6, we approximate the Monge-Ampère equation via its vanishing moment regularization (2.8)–(2.10). In Chapter 3, we study conforming finite element methods in 2-D and 3-D. The Argyris finite element is specifically considered although the analysis applies to any C^1 element. In Chapter 4, we extend the analysis of Chapter 3 to spectral Galerkin methods. In Chapter 5, we derive a Hermann-Myoshi type mixed method formulation for (2.8)–(2.10) and analyze the error of the numerical solution. In Chapter 6 we consider the numerical approximation of (2.8)–(2.10) using the nonconforming Morley element. In Chapter 7, we consider the numerical approximation of the nonlinear balance equation which arises in meteorology. Chapter 8 is devoted to studying the numerical approximation of the semigeostrophic flow equations in a fully nonlinear formulation which consists of the Monge-Ampère equation and the transport equation. Chapter 9 builds upon Chapter 3, where the analysis of C^1 finite elements is extended to general fully nonlinear second order PDEs satisfying certain structure conditions. Numerical approximations of the equation of prescribed Gauss curvature (1.3) is given as a specific example. Finally, in Chapter 10, we comment on further applications of the vanishing moment method and future directions we will pursue.

1.7 Mathematical Software and Implementation

We used the finite element method software package COMSOL Multiphysics [29] to run all of the numerical tests in Chapters 3,5,7,8, and 9, and used the programming language MATLAB [72] to develop code for the numerical experiments found in Chapter 6. For more information on these software packages see http://www.comsol.com/ and http://www.mathworks.com/. To solve the resulting nonlinear algebraic system for each test, we used a damped Newton method and used the direct linear solver UMFPACK [38] within each Newton iteration. We ran the experiments on a workstation with an Intel Core 2 Duo rated at 2.4 GHz.

1.8 General Notation

Standard space notations are adopted in this dissertation. n denotes the spatial dimension which will be restricted to the cases $n = 2$ and $n = 3$. Ω and U will denote open, bounded, convex domains in $\mathbf{R}^n$ unless otherwise stated. The L^2-inner product is defined by

$$(v, w) := \int_\Omega v \diamond w dx \qquad \forall v, w \in L^2(\Omega),$$

where '$\diamond$' refers to either multiplication, dot product, or tensor product. We define the L^2-inner product over the boundary $\partial\Omega$ as

$$\langle v, w \rangle_{\partial\Omega} := \int_{\partial\Omega} v \diamond w ds \quad \forall \varphi, \omega \in L^2(\partial\Omega).$$

We use $\langle \cdot, \cdot \rangle$ to denote the pairing between a Banach space X and its duel X^* (except in the case $X = L^2(\Omega)$). We denote the L^2-norm by

$$\|\varphi\|_{L^2} := \|\varphi\|_{L^2(\Omega)} := (\varphi, \varphi)^{\frac{1}{2}}.$$

For $m \geq 0$, $p \geq 1$, let $W^{m,p}(\Omega)$ denote the Sobolev space

$$W^{m,p}(\Omega) := \{\varphi \in L^p(\Omega);\ D^\alpha \varphi \in L^p(\Omega),\ |\alpha| \leq m\}$$

endowed with the norm

$$\|\varphi\|_{W^{m,p}} := \|\varphi\|_{W^{m,p}(\Omega)} := \left(\sum_{|\alpha| \leq m} \|D^\alpha \varphi\|_{L^p}^p \right)^{\frac{1}{p}}.$$

We denote the Hilbert spaces $W^{m,2}(\Omega)$ by $H^m(\Omega)$ and often write $H^m = H^m(\Omega)$. In particular $\|\cdot\|_{H^m} := \|\cdot\|_{H^m(\Omega)}$. We also define the Sobolev semi-norms

$$|\varphi|_{W^{m,p}} := |\varphi|_{W^{m,p}(\Omega)} := \left(\sum_{|\alpha|=m} \|D^\alpha \varphi\|_{L^p}^p \right)^{\frac{1}{p}}.$$

When $p = 2$, we write $|\varphi|_{H^m} := |\varphi|_{W^{m,p}}$. Finally, C is used to denote a generic ϵ and h-independent positive constant, and all constants are chapter-independent unless otherwise specified.

Chapter 2

The Vanishing Moment Method

2.1 Motivation

In contrast with the enormous advances in PDE theory for second order fully nonlinear PDEs, numerical approximations of this class of PDEs is essentially an untouched area. There are three main reasons for the lack of progress. The obvious difficulty is the nonlinearity of the PDE. Second is the conditional uniqueness. Recall that solutions to fully nonlinear PDEs are usually only unique in a certain class of functions. Thus, regardless of what numerical scheme is used to approximate (1.1), the resulting algebraic system would not only be difficult to solve, it would also be difficult to determine which solution one is approximating. Finally and most importantly, the notion of viscosity solutions is not variational and difficult to mimic at the discrete level.

To overcome the above difficulties, we introduce a new notion of solutions for fully nonlinear second order PDEs called *moment solutions*, and a constructive method called *the vanishing moment method* which mimics the vanishing viscosity method. Recall that the existence of the viscosity solution was first proved by Crandall and Lions [30] using the vanishing viscosity method for the Hamilton-Jacobi equations. A simple but crucial observation is that the essence of the the vanishing viscosity method involves approximating a lower order fully nonlinear PDE by a family of quasilinear higher order PDEs. This observation motivates us to apply the above principle to second order PDEs (1.1). That is, we approximate fully nonlinear second order PDEs

$$F(D^2u, Du, u, x) = 0 \qquad \text{in } \Omega, \tag{2.1}$$

$$u = g \qquad \text{on } \partial\Omega, \tag{2.2}$$

by the following higher order quasilinear PDEs:

$$G_\epsilon(D^r u^\epsilon) + F(D^2 u^\epsilon, Du^\epsilon, u^\epsilon, x) = 0 \qquad \text{in } \Omega, \tag{2.3}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{2.4}$$

where $r \geq 3$, $\epsilon > 0$, and $\{G_\epsilon\}$ is a family of suitably chosen linear or quasilinear differential operators of order r [50] .

Definition 2.1.1. *Suppose that $u^\epsilon \in H^2(\Omega) \cap C^0(\overline{\Omega})$ solves problem* (2.3)–(2.7). *$\lim_{\epsilon \to 0^+} u^\epsilon$, if it exists, is called a* weak *(resp.* strong*)* moment solution *to problem* (2.1)–(2.2) *if the convergence holds in H^1-weak (resp. H^2-weak) topology. We call this limiting process the* vanishing moment method.

All of the second order PDEs considered in this dissertation are elliptic, and thus, it is intuitively better to choose $G_\epsilon(D^r u^\epsilon)$ to be elliptic. Since an elliptic operator is necessarily of even order, the lowest order of (2.3) is $r = 4$. When thinking of fourth order elliptic operators, the biharmonic operator stands out immediately. Furthermore, we require $G_\epsilon \to 0$ in some reasonable sense as $\epsilon \to 0^+$. Making use of these observations, for the continuation of the dissertation, we set

$$G_\epsilon(D^r u^\epsilon) := \epsilon \Delta^2 u^\epsilon, \tag{2.5}$$

and (2.3) becomes

$$\epsilon \Delta^2 u^\epsilon + F(D^2 u^\epsilon, Du^\epsilon, u^\epsilon, x) = 0. \tag{2.6}$$

Noting Dirichlet boundary condition (2.4) is not sufficient for well-posedness, an additional boundary condition must be imposed. Several boundary conditions could be used for this purpose. Physically, any additional boundary condition will introduce a boundary layer, so a better choice would be one which minimizes the boundary layer. Thus, in addition to (2.4), we propose the use of one of the following boundary conditions:

$$\Delta u^\epsilon = c_\epsilon, \quad \text{or} \qquad \frac{\partial \Delta u^\epsilon}{\partial \eta} = c_\epsilon, \quad \text{or} \qquad D^2 u^\epsilon \eta \cdot \eta = c_\epsilon \quad \text{on } \partial\Omega, \tag{2.7}$$

where η denotes the outward unit normal to $\partial\Omega$. In summary, the vanishing moment method consists of approximating the Dirichlet problem (2.1)–(2.2) by the quasilinear fourth order boundary value problem (2.6),(2.4),(2.7).

Remark 2.1.2. *We note that the first two boundary conditions in* (2.7)*, which are natural boundary conditions, have an advantage in PDE convergence analysis. Also, the first*

boundary condition in (2.7) *is better suited for conforming and nonconforming finite element methods, where as the last boundary condition in* (2.7) *fits naturally with the mixed finite element formulation. Also, we comment that* $c_\epsilon = \epsilon$ *will be used in most parts of this dissertation.*

Remark 2.1.3. *When* $n = 2$ *in mechanical applications,* u^ϵ *often stands for the vertical displacement of a plate, and* D^2u^ϵ *is the moment tensor. In the weak formulation, the biharmonic term becomes* $\epsilon(D^2u^\epsilon, D^2v)$ *which should vanish as* $\epsilon \to 0^+$*. This is the reason we call* $\lim_{\epsilon\to 0^+} u^\epsilon$ *a moment solution and call the limiting process the vanishing moment method.*

2.2 Vanishing Moment Approximation for the Monge-Ampère Equation

Applying the vanishing moment method to the Monge-Ampère equation and choosing the first boundary condition in (2.7), we approximate (1.11)–(1.12) by the following fourth order quasilinear problem:

$$-\epsilon\Delta^2 u^\epsilon + \det(D^2u^\epsilon) = f\ (>0) \qquad \text{in } \Omega, \tag{2.8}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{2.9}$$

$$\Delta u^\epsilon = \epsilon \qquad \text{on } \partial\Omega. \tag{2.10}$$

Multiplying equation (2.8) by a function $v \in H^2(\Omega)\cap H_0^1(\Omega)$, integrating by parts, and using Green's formula we have

$$\begin{aligned}(f,v) &= -\epsilon(\Delta^2u^\epsilon, v) + (\det(D^2u^\epsilon), v)\\ &= \epsilon(D(\Delta u^\epsilon), Dv) + (\det(D^2u^\epsilon), v)\\ &= -\epsilon(\Delta u^\epsilon, \Delta v) + (\det(D^2u^\epsilon), v) + \epsilon\left\langle \Delta u^\epsilon, \frac{\partial v}{\partial \eta}\right\rangle_{\partial\Omega}\\ &= -\epsilon(\Delta u^\epsilon, \Delta v) + (\det(D^2u^\epsilon), v) + \left\langle \epsilon^2, \frac{\partial v}{\partial \eta}\right\rangle_{\partial\Omega}.\end{aligned}$$

Using this identity, we give the following definition.

Definition 2.2.1. *We define* $u^\epsilon \in H^2(\Omega)$ *to be a solution of* (2.8)–(2.10) *if* $u^\epsilon = g$ *on* $\partial\Omega$ *and*

$$-\epsilon(\Delta u^\epsilon, \Delta v) + (\det(D^2u^\epsilon), v) = (f,v) - \left\langle \epsilon^2, \frac{\partial v}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall v \in H^2(\Omega)\cap H_0^1(\Omega).$$

We remark that for a Lipschitz domain Ω, the embedding $H^2(\Omega) \hookrightarrow C^0(\overline{\Omega})$ holds, and hence, $u^\epsilon = g$ on $\partial\Omega$ makes sense.

2.2.1 PDE Results and Assumptions

We now summarize the results of [49] which concerns the well-posedness of (2.8)–(2.10).

Theorem 2.2.2. *In the case $n = 2$ there exists a unique solution u^ϵ to (2.8)–(2.10) for all $\epsilon > 0$. Furthermore, u^ϵ converges to u pointwise and H^1-weakly as $\epsilon \to 0^+$, where u denotes the unique viscosity solution of (1.11)–(1.12). Moreover, the following bounds hold:*

$$\begin{aligned} &\|u^\epsilon\|_{H^j} = O(\epsilon^{\frac{1-j}{2}}) \ (j = 1, 2, 3), \quad &&\|u^\epsilon\|_{W^{j,\infty}} = O(\epsilon^{1-j}) \ (j = 1, 2), \\ &\|\Phi^\epsilon\|_{L^\infty} = O(\epsilon^{-1}), &&\|\Phi^\epsilon\|_{L^2} = O(\epsilon^{-\frac{1}{2}}), \end{aligned} \tag{2.11}$$

where $\Phi^\epsilon = cof(D^2u^\epsilon)$, denotes the cofactor matrix of D^2u^ϵ.

Remark 2.2.3. *We note that (strong) convergence in H^1 and H^2 has not been proven, nor have any rates of convergence been shown. We address all of these issues in Sections 3.4, 5.5, and 6.5.*

Remark 2.2.4. *We note that the results of Theorem 2.2.2 were proved in the case $n = 2$ and $\Delta u^\epsilon = \epsilon$ a.e on $\partial\Omega$. However, throughout this dissertation, we will assume that Theorem 2.2.2 hold for $n = 3$ and the boundary condition replaced by $D^2u^\epsilon\eta \cdot \eta = \epsilon$. That is for every $\epsilon > 0$, there exists a unique solution u^ϵ that solves*

$$-\epsilon\Delta^2 u^\epsilon + \det(D^2u^\epsilon) = f \ (> 0) \qquad \text{in } \Omega, \tag{2.12}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{2.13}$$

$$D^2u^\epsilon\eta \cdot \eta = \epsilon \qquad \text{on } \partial\Omega, \tag{2.14}$$

and u^ϵ converges pointwise and H^1-weakly to u, the unique convex viscosity solution of (1.11)-(1.12). In Section 5.4, we will find that this assumption (which will be used often in Chapter 5) is most likely to be true.

Next, we extend the methodology of the vanishing moment method to approximate the other solution of (1.11)–(1.12) in the case $n = 2$. Recall, in the case n is even, we can define the notion of a concave solution of the Monge-Ampère equation, which we denote

by $\tilde{u}$ (cf. Definition 1.3.3). We approximate $\tilde{u}$ by $\tilde{u}^\epsilon$, where $\tilde{u}^\epsilon$ solves

$$\epsilon\Delta^2\tilde{u}^\epsilon + \det(D^2\tilde{u}^\epsilon) = f \qquad \text{in } \Omega, \tag{2.15}$$

$$\tilde{u}^\epsilon = g \qquad \text{on } \partial\Omega, \tag{2.16}$$

$$\Delta\tilde{u}^\epsilon = -\epsilon \qquad \text{on } \partial\Omega. \tag{2.17}$$

Definition 2.2.5. *We define $\tilde{u}^\epsilon \in H^2(\Omega)$ to be a solution of (2.15)–(2.17) if $\tilde{u}^\epsilon = g$ on $\partial\Omega$ and*

$$\epsilon(\Delta\tilde{u}^\epsilon, \Delta v) + (\det(D^2\tilde{u}^\epsilon), v) = (f, v) - \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall v \in H^2(\Omega) \cap H_0^1(\Omega).$$

We have the following result.

Theorem 2.2.6. *When $n = 2$, there exists a unique solution $\tilde{u}^\epsilon$ to (2.15)–(2.17). Furthermore, $\tilde{u}^\epsilon$ is strictly concave for each $\epsilon > 0$, and $\tilde{u}^\epsilon$ converges to $\tilde{u}$ pointwise and H^1-weakly as $\epsilon \to 0^+$, where $\tilde{u}$ is the concave viscosity solution of (1.11)–(1.12).*

Proof. It is easy to see that $\tilde{u}^\epsilon$ is the solution to (2.15)–(2.17) if and only if $u^\epsilon := -\tilde{u}^\epsilon$ is the unique convex solution to (2.8)–(2.10) with g replaced by $-g$. That is, u^ϵ satisfies

$$-\epsilon\Delta^2 u^\epsilon + \det(D^2 u^\epsilon) = f \qquad \text{in } \Omega,$$

$$u^\epsilon = -g \qquad \text{on } \partial\Omega,$$

$$\Delta u^\epsilon = \epsilon \qquad \text{on } \partial\Omega.$$

Thus, the first assertion holds by Theorem 2.2.2. Also, u^ϵ is strictly convex implies that $\tilde{u}^\epsilon$ is strictly concave.

Finally, let u denote the unique convex viscosity solution of (1.11)–(1.12) with g replaced by $-g$. We note that $\tilde{u} = -u$ and u^ϵ converges to u pointwise and H^1-weakly as $\epsilon \to 0^+$ by Theorem 2.2.2. Since $\tilde{u}^\epsilon = -u^\epsilon$, then $\tilde{u}^\epsilon$ converges to $\tilde{u}$ pointwise and H^1-weakly, and the proof is complete. □

With these results in place, we can now use existing numerical discretization methods devoted for the biharmonic problem to approximate the second order fully nonlinear Monge-Ampère equation. Chapters 3–6 are concerned with this analysis. In Chapter 3, we approximate (2.8)–(2.10) using C^1 finite elements. The Argyris finite element is specifically considered, although the analysis applies to any C^1 element. Chapter 4 extends the work of Chapter 3 to spectral Galerkin methods. In Chapter 5, we approximate (2.12)–(2.14) using Hermann-Myoshi mixed finite element methods. Next, in Chapter 6, we solve (2.8)–(2.10) using the n-dimensional Morley element which was recently introduced in [74]. Finally, we note by Theorems 2.2.2 and 2.2.6, we have an easy way to choose which solution of the

Monge-Ampère equation we are approximating by computing (2.8)–(2.10) or (2.15)–(2.17). We briefly touch upon this issue again in Section 3.5.

Chapter 3

C^1 Finite Element Methods for the Monge-Ampère Equation

The goal of this chapter is to develop and analyze conforming finite element methods that approximate the solution of (2.8)–(2.10) in 2-D and 3-D. As a result, we will obtain convergent methods to approximate the convex viscosity solution of the Monge-Ampère problem (1.11)–(1.12). When deriving error estimates, we are particularly interested in obtaining error bounds that show explicit dependence on ϵ. Argyris finite element methods are specifically considered in this chapter, although our analysis applies to any C^1 finite element method such as Bogner-Fox-Schmit and Hsieh-Clough-Tocher elements (cf. [27, 17]) when $n = 2$.

We note that finite element approximations of fourth order PDEs, in particular, the biharmonic equation, were carried out extensively in 1970's in the two-dimensional case [27], and have attracted renewed interests lately for generalizing the well-known 2-D finite elements to the 3-D case (cf. [90, 93, 94]). All these methods can be readily adapted to discretize problem (2.8)–(2.10) although their convergence analysis do not come easy because of the strong nonlinearity of the PDE (2.8).

To overcome this difficulty, we use a fixed point technique which makes strong use of the stability property of the linearized problem which is analyzed in Section 3.2. By doing so, in Section 3.3, we obtain optimal order error estimates in the energy norm as well as in the L^2-norm and H^1-norm. Section 3.4 studies the approximation results when the data is slightly changed in the discretization. We may think of this perturbation of data as the effect of numerical quadrature, but the analysis will also be useful in Chapter 8 when we study the semigeostrophic flow equations. Section 3.5 studies the finite element approximation of (2.15)–(2.17) in $n = 2$ which in turn approximates the concave solution of (1.11)–(1.12). The results in this section will follow directly from the analysis of Section 3.3. Finally, Section 3.6 presents a number of numerical experiments to validate the theoretical error

estimate results in 2-D. We then present a detailed computational study for determining the "best" choice of mesh size h in terms of ϵ in order to achieve the optimal rates of convergence and for estimating the rate of convergence for both $u - u_h^\epsilon$ and $u - u^\epsilon$ in terms of powers of ϵ.

3.1 Formulation of Finite Element Methods

We first introduce the following space notation:

$$V := H^2(\Omega), \qquad V_0 := H^2(\Omega) \cap H_0^1(\Omega), \qquad V_g := \{v \in V;\ v|_{\partial\Omega} = g\}.$$

Based on Definition 2.2.1, we define the variational formulation of (2.8)–(2.10) as follows: Find $u^\epsilon \in V_g$ such that

$$-\epsilon(\Delta u^\epsilon, \Delta v) + (\det(D^2 u^\epsilon), v) = (f, v) - \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall v \in V_0. \tag{3.1}$$

Remark 3.1.1. *We note*

$$\det(D^2 u^\epsilon) = \frac{1}{n}\Phi^\epsilon : D^2 u^\epsilon = \frac{1}{n}\sum_{i=1}^{n} \Phi_{ij}^\epsilon \frac{\partial^2 u}{\partial x_i \partial x_j} \qquad j = 1, 2, ...n,$$

where $\Phi^\epsilon = \mathrm{cof}(D^2 u^\epsilon)$ is the cofactor matrix of $D^2 u^\epsilon$. Thus, using the divergence free property of cofactor matrices (cf. Lemma A.0.1), we can define the following alternative variational formulation for problem (2.8)–(2.10)*:*

$$-\epsilon(\Delta u^\epsilon, \Delta v) - \frac{1}{n}(\Phi^\epsilon D u^\epsilon, Dv) = (f, v) - \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall v \in V_0. \tag{3.2}$$

However, we shall not use the above weak formulation in this dissertation.

To formulate the finite element method, let $\mathcal{T}_h$ be a quasiuniform triangular or rectangular partition of Ω if $n = 2$ or a quasiuniform tetrahedral or 3-D rectangular mesh if $n = 3$ with mesh size $h \in (0, 1)$. Let $V^h \subset V$ denote a conforming finite element space consisting of piecewise polynomial functions of degree $r(\geq 5)$ such that for any $v \in V \cap H^s(\Omega)$ $(s \geq 3)$

$$\inf_{v_h \in V^h} \|v - v_h\|_{H^j} \leq h^{\ell - j} \|v\|_{H^s}, \quad j = 0, 1, 2, \quad \ell = \min\{r + 1, s\}. \tag{3.3}$$

We recall that $r = 5$ in the case of the Argyris element (cf. [27, 17]). Let

$$V_g^h = \{v_h \in V^h;\ v_h|_{\partial\Omega} = g\}, \qquad V_0^h = \{v_h \in V^h;\ v_h|_{\partial\Omega} = 0\}.$$

Based on the weak formulation (3.1), we define our finite element method as follows: Find $u_h^\epsilon \in V_g^h$ such that

$$-\epsilon\big(\Delta u_h^\epsilon, \Delta v_h\big) + \big(\det(D^2 u_h^\epsilon), v_h\big) = (f, v_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta} \right\rangle_{\partial\Omega} \quad \forall v_h \in V_0^h. \tag{3.4}$$

Let u^ϵ be the solution to (3.1) and u_h^ϵ the solution of (3.4). The main task of this chapter is derive optimal order error estimates for $u^\epsilon - u_h^\epsilon$. To this end, we first need to prove existence and uniqueness of u_h^ϵ. Both tasks are not easy due to the strong nonlinearity in (3.4). Unlike the continuous PDE case where u^ϵ is assumed to be convex for all ϵ, it is not clear whether u_h^ϵ preserves the convexity even for small h and ϵ. Without a guarantee of convexity for u_h^ϵ, it would be difficult to establish any stability result for u_h^ϵ. This is the main obstacle for proving existence and uniqueness for (3.4). In addition, again due to the strong nonlinearity, the standard perturbation technique for deriving error estimate for numerical approximations of mildly nonlinear problems does work. To overcome these difficulties, we use a combined linearization and fixed point technique. We note that by using this technique, we are able to simultaneously prove existence and uniqueness for u_h^ϵ and also derive the desired error estimates. In the next two sections, we shall give a detailed account of this technique and apply it to problem (3.4).

3.2 Linearization and its Finite Element Approximation

To analyze (3.4), we shall study the linearization of (2.8) to build the required technical tools.

3.2.1 Linearization

For a given smooth function w and $t \in \mathbf{R}$, there holds (cf. [24])

$$\det(D^2(u^\epsilon + tw)) = \det(D^2 u^\epsilon) + t\mathrm{tr}(\Phi^\epsilon D^2 w) + \cdots + t^n \det(D^2 w). \tag{3.5}$$

Thus, the linearization of

$$M^\epsilon(u^\epsilon) := \epsilon\Delta^2 u^\epsilon - \det(D^2 u^\epsilon)$$

at the solution u^ϵ is given by

$$\begin{aligned} L_{u^\epsilon}(w) :&= \lim_{t\to 0} \frac{M^\epsilon(u^\epsilon + tw) - M^\epsilon(u^\epsilon)}{t} \\ &= \epsilon\Delta^2 w - \Phi^\epsilon : D^2 w \\ &= \epsilon\Delta^2 w - \mathrm{div}(\Phi^\epsilon Dw), \end{aligned} \tag{3.6}$$

where we have used Lemma A.0.1 to get the last equality. Next, we have for any $v, w \in H_0^2(\Omega)$

$$\begin{aligned}\langle L_{u^\epsilon}(w), v\rangle &= \langle \epsilon\Delta^2 w - \operatorname{div}(\Phi^\epsilon Dw), v\rangle \\ &= \langle -\epsilon D\Delta w + \Phi^\epsilon Dw, Dv\rangle \\ &= \epsilon(\Delta w, \Delta v) + (\Phi^\epsilon Dw, Dv) \\ &= -\epsilon\langle Dw, D(\Delta v)\rangle + (Dw, \Phi^\epsilon Dv) \\ &= \langle w, \epsilon\Delta^2 v - \operatorname{div}(\Phi^\epsilon Dv)\rangle \\ &= \langle w, L_{u^\epsilon}(v)\rangle.\end{aligned}$$

Hence L_{u^ϵ} is self-adjoint, that is, $\langle L_{u^\epsilon}(w), v\rangle = \langle w, L^*_{u^\epsilon}(v)\rangle = \langle w, L_{u^\epsilon}(v)\rangle \ \forall v, w \in H_0^2(\Omega)$, where $L^*_{u^\epsilon}$ is the adjoint operator of L_{u^ϵ}.

Given $\varphi \in V_0^*$, we now consider the following linear problem:

$$L_{u^\epsilon}(v) = \varphi \quad \text{in } \Omega, \tag{3.7}$$

$$v = 0 \quad \text{on } \partial\Omega, \tag{3.8}$$

$$\Delta v = \psi \quad \text{on } \partial\Omega. \tag{3.9}$$

Multiplying (3.7) by a test function $w \in V_0$ and integrating over Ω we get the following weak formulation for (3.7)–(3.9): Find $v \in V_0$ such that

$$B[v, w] = \langle \varphi, w\rangle - \epsilon \left\langle \psi, \frac{\partial w}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w \in V_0, \tag{3.10}$$

where

$$B[v, w] := \epsilon \int_\Omega \Delta v \Delta w \, dx + \int_\Omega \Phi^\epsilon Dv \cdot Dw \, dx. \tag{3.11}$$

The next theorem ensures the well-posedness of the above variational problem.

Theorem 3.2.1. *Suppose $\partial\Omega \in C^{0,1}$. Then for every $\varphi \in V_0^*$ and $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$ there exists a unique $v \in V_0$ such that*

$$B[v, w] = \langle \varphi, w\rangle - \epsilon \left\langle \psi, \frac{\partial w}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w \in V_0. \tag{3.12}$$

Furthermore, there exists a constant $C_1(\epsilon) = O(\epsilon^{-1})$ such that

$$\|v\|_{H^2} \le C_1(\epsilon) \left(\|\varphi\|_{(H_0^1 \cap H^2)^*} + \epsilon \|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} \right). \tag{3.13}$$

Proof. By (2.11), we have $\|\Phi^\epsilon\|_{L^2} \le C\epsilon^{-\frac{1}{2}}$. Thus, we bound $B[v,w]$ as follows:

$$|B[v,w]| \le \epsilon\|\Delta v\|_{L^2}\|\Delta w\|_{L^2} + C\epsilon^{-\frac{1}{2}}\|Dv\|_{L^4}\|Dw\|_{L^4} \le C\epsilon^{-\frac{1}{2}}\|v\|_{H^2}\|w\|_{H^2},$$

where we have used Sobolev's inequality, noting $H^1(\Omega) \hookrightarrow L^4(\Omega)$ for $1 \le n \le 4$.

To obtain coercivity, we note that since u^ϵ is strictly convex, D^2u^ϵ is positive definite. Thus, Φ^ϵ is positive definite. Therefore, there exists $\theta > 0$ such that

$$B[v,v] \ge \epsilon\|\Delta v\|_{L^2}^2 + \theta\|Dv\|_{L^2}^2 \ge \epsilon\|\Delta v\|_{L^2}^2 + \frac{\theta}{2}\|Dv\|_{L^2}^2 + \frac{\theta}{2C}\|v\|_{L^2}^2,$$

where we have used Poincaré's inequality. Since Ω is convex, we have $\|v\|_{H^2} \le C\|\Delta v\|_{L^2}$ [17]. Thus,

$$B[v,v] \ge C_2(\epsilon)\|v\|_{H^2}^2, \tag{3.14}$$

where $C_2(\epsilon) := C\min\{\epsilon,\theta\} = O(\epsilon)$.

Next, we confirm that $G(w) := \langle\varphi, w\rangle - \left\langle\psi, \frac{\partial w}{\partial\eta}\right\rangle_{\partial\Omega}$ is a bounded linear functional. Clearly, G is linear. Also,

$$\begin{aligned} |G(w)| &\le \|\varphi\|_{(H_0^1\cap H^2)^*}\|w\|_{H^2} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\left\|\frac{\partial w}{\partial\eta}\right\|_{H^{\frac{1}{2}}(\partial\Omega)} \\ &\le \|\varphi\|_{(H_0^1\cap H^2)^*}\|w\|_{H^2} + \epsilon C\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\|w\|_{H^2}, \end{aligned} \tag{3.15}$$

where we have used the trace inequality [44, p.258]. Noting $\varphi \in V_0^*$, $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$, G is bounded.

Thus, by the Lax-Milgram Theorem [44, p. 297], for every $\varphi \in V_0^*$ and $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$, there exists a unique $v \in V_0$ such that

$$B[v,w] = \langle\varphi, w\rangle - \epsilon\left\langle\psi, \frac{\partial w}{\partial\eta}\right\rangle_{\partial\Omega} \qquad \forall w \in V_0.$$

To obtain (3.13), we use (3.14) and (3.15) to get

$$\|v\|_{H^2} \le C_1(\epsilon)\left(\|\varphi\|_{(H^1\cap H^2)^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\right), \tag{3.16}$$

where $C_1(\epsilon) = CC_2^{-1}(\epsilon) = O(\epsilon^{-1})$. □

We now improve the above results in the case when the data is smoother.

Theorem 3.2.2. *Suppose $\varphi \in H^{-1}(\Omega)$, $\psi \in H^{\frac{1}{2}}(\partial\Omega)$, and $\partial\Omega \in C^1$. Then $v \in H^3(\Omega) \cap H_0^1(\Omega)$, and if $\psi \equiv 0$, the following bound holds:*

$$\|v\|_{H^3} \le C\epsilon^{-2}\|\varphi\|_{H^{-1}}. \tag{3.17}$$

Furthermore, if $\varphi \in L^2(\Omega)$, $\psi \in H^{\frac{3}{2}}(\partial\Omega)$, *and* $\partial\Omega \in C^2$, *then* $v \in H^4(\Omega) \cap H_0^1(\Omega)$, *and if* $\psi \equiv 0$, *we have the following bound:*

$$\|v\|_{H^4} \leq C\epsilon^{-3}\|\varphi\|_{L^2}. \tag{3.18}$$

Proof. The assertions that $v \in H^3(\Omega) \cap H_0^1(\Omega)$ and $v \in H^4(\Omega) \cap H_0^1(\Omega)$ follow from standard elliptic theory [44, 57, 58]. Thus, we only have to show bounds (3.17) and (3.18).

Suppose $\varphi \in H^{-1}(\Omega)$ and $\psi \equiv 0$. Multiplying (3.7) by v, integrating over Ω, and using Lemma A.0.1, we have

$$\epsilon(\Delta v, \Delta v) + (\Phi^\epsilon Dv, Dv) = \langle \varphi, v \rangle.$$

Next, we use Poincaré's inequality to obtain

$$\begin{aligned} \epsilon\|\Delta v\|_{L^2}^2 + \theta\|Dv\|_{L^2}^2 &\leq \|\varphi\|_{H^{-1}}\|v\|_{H^1} \\ &\leq C\|\varphi\|_{H^{-1}}\|Dv\|_{L^2}. \end{aligned}$$

Thus,

$$\|Dv\|_{L^2} \leq C\|\varphi\|_{H^{-1}}. \tag{3.19}$$

Multiplying (3.7) by Δv, integrating, and using (2.11),(3.19), we have

$$\begin{aligned} \epsilon\|D(\Delta v)\|_{L^2}^2 &= -\langle \varphi, \Delta v \rangle + (\Phi^\epsilon Dv, D(\Delta v)) \\ &\leq \|\varphi\|_{H^{-1}}\|\Delta v\|_{H^1} + \|\Phi^\epsilon\|_{L^\infty}\|Dv\|_{L^2}\|D(\Delta v)\|_{L^2} \\ &\leq C\big(\|\varphi\|_{H^{-1}} + \epsilon^{-1}\|Dv\|_{L^2}\big)\|D(\Delta v)\|_{L^2} \\ &\leq C\epsilon^{-1}\|\varphi\|_{H^{-1}}\|D(\Delta v)\|_{L^2}. \end{aligned}$$

Thus,

$$\|D(\Delta v)\|_{L^2} \leq C\epsilon^{-2}\|\varphi\|_{H^{-1}},$$

and (3.17) follows from Poincaré's inequality.

Next, suppose $\varphi \in L^2(\Omega)$. Multiplying (3.7) by $\Delta^2 v$, integrating, and using Theorem 3.2.1, we get

$$\begin{aligned} \epsilon\|\Delta^2 v\|_{L^2}^2 &= (\varphi, \Delta^2 v) + (\Phi^\epsilon : D^2 v, \Delta^2 v) \\ &\leq C\big(\|\varphi\|_{L^2} + \|\Phi^\epsilon\|_{L^\infty}\|D^2 v\|_{L^2}\big)\|\Delta^2 v\|_{L^2} \\ &\leq C\big(\|\varphi\|_{L^2} + \epsilon^{-1}\|D^2 v\|_{L^2}\big)\|\Delta^2 v\|_{L^2} \end{aligned}$$

$$\leq CC_1(\epsilon)\epsilon^{-1}\|\varphi\|_{L^2}\|\Delta^2 v\|_{L^2}.$$

Thus,

$$\|\Delta^2 v\|_{L^2} \leq CC_1(\epsilon)\epsilon^{-2}\|\varphi\|_{L^2} \leq C\epsilon^{-3}\|\varphi\|_{L^2}.$$

□

Next, we note that Theorem 3.2.1 can be extended to the case of nonhomogeneous boundary data which is shown in the following theorem.

Theorem 3.2.3. *For every $\varphi \in V^*$, $\xi \in H^{\frac{3}{2}}(\partial\Omega)$, $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$, there exists a unique $v \in V$ such that*

$$\begin{aligned} L_{u^\epsilon}(v) &= \varphi \quad \text{in } \Omega, \\ v &= \xi \quad \text{on } \partial\Omega, \\ \Delta v &= \psi \quad \text{on } \partial\Omega. \end{aligned} \tag{3.20}$$

Furthermore, we have the estimate

$$\|v\|_{H^2} \leq C_3(\epsilon)\left(\|\varphi\|_{(H^2)^*} + \epsilon^{-\frac{1}{2}}\|\xi\|_{H^{\frac{3}{2}}(\partial\Omega)} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\right), \tag{3.21}$$

where $C_3(\epsilon) = O(\epsilon^{-1})$.

Proof. Let $w = v - \hat{v}$, where $\hat{v} \in V$ is an extension of ξ. We then seek $w \in V_0$ such that

$$B[w,z] = \langle\varphi, z\rangle - \epsilon\left\langle\psi, \frac{\partial z}{\partial\eta}\right\rangle_{\partial\Omega} - B[\hat{v}, z] \qquad \forall z \in V_0. \tag{3.22}$$

From the proof of Theorem 3.2.1, it suffices to show $H(z) := \langle\varphi, z\rangle - \epsilon\langle\psi, \frac{\partial z}{\partial\eta}\rangle - B[\hat{v}, z]$ is a bounded linear functional. Trivially, H is linear and

$$\begin{aligned} |H(z)| &\leq \|\varphi\|_{(H^2)^*}\|z\|_{H^2} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\|Dz\|_{H^{\frac{1}{2}}(\partial\Omega)} + C\epsilon^{-\frac{1}{2}}\|\hat{v}\|_{H^2}\|z\|_{H^2} \\ &\leq C\left(\|\varphi\|_{(H^2)^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} + \epsilon^{-\frac{1}{2}}\|\xi\|_{H^{\frac{3}{2}}(\partial\Omega)}\right)\|z\|_{H^2}. \end{aligned} \tag{3.23}$$

Thus, $H(\cdot)$ is bounded. Therefore, by the Lax-Milgram Theorem for every $\varphi \in V^*$, $\xi \in H^{\frac{3}{2}}(\partial\Omega)$, $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$ there exists a unique solution w solving (3.22). It follows that there exists a unique solution, v, solving (3.20).

To get (3.21), we use (3.23) with $z = w$ and the coercivity of $B[\cdot,\cdot]$ to get

$$C_2(\epsilon)\|w\|_{H^2}^2 \leq C\left(\|\varphi\|_{(H^2)^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} + \epsilon^{-\frac{1}{2}}\|\xi\|_{H^{\frac{3}{2}}(\partial\Omega)}\right)\|w\|_{H^2}.$$

Thus,

$$C_2(\epsilon)\|v\|_{H^2} \le C\left((\|\varphi\|_{(H^2)^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} + \epsilon^{-\frac{1}{2}}\|\xi\|_{H^{\frac{3}{2}}(\partial\Omega)}\right) + C_2(\epsilon)\|\hat{v}\|_{H^2}$$
$$\le C\left(\|\varphi\|_{(H^2)^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} + \epsilon^{-\frac{1}{2}}\|\xi\|_{H^{\frac{3}{2}}(\partial\Omega)}\right).$$

From this inequality, we get (3.21). □

3.2.2 Finite Element Approximation of Linearized Problem

Let $V_0^h \subset V_0$ be one of the finite-dimensional subspaces of V_0 as defined in Section 3.1 (e.g. Argyris finite element), and $v \in V_0$ denote the solution of (3.10). Based on the variational equation (3.10), our finite element method for (3.7) is defined as seeking $v_h \in V_0^h$ such that

$$B[v_h, w_h] = \langle \varphi, w_h \rangle - \epsilon \left\langle \psi, \frac{\partial w_h}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w_h \in V_0^h. \tag{3.24}$$

Our objective in this section is to first prove existence and uniqueness for problem (3.24) and then derive optimal order error estimates in various norms.

Theorem 3.2.4. *Suppose $v \in V_0 \cap H^s(\Omega)$ $(s \ge 3)$. Then there exists a unique $v_h \in V_0^h$ satisfying* (3.24)*. Furthermore, we have the following estimates:*

$$\|v_h\|_{H^2(\Omega)} \le C_3(\epsilon)\left(\|\varphi\|_{(H^1\cap H^2)^*} + \|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\right), \tag{3.25}$$
$$\|v - v_h\|_{H^2(\Omega)} \le C_4(\epsilon)h^{\ell-2}\|v\|_{H^\ell(\Omega)}, \tag{3.26}$$
$$\|v - v_h\|_{H^1(\Omega)} \le C_5(\epsilon)h^{\ell-1}\|v\|_{H^\ell(\Omega)}, \tag{3.27}$$
$$\|v - v_h\|_{L^2(\Omega)} \le C_6(\epsilon)h^{\ell}\|v\|_{H^\ell(\Omega)}, \tag{3.28}$$

where $\ell := \min\{r+1, s\}$, $C_3(\epsilon) = (\epsilon^{-1})$, $C_4(\epsilon) = O(\epsilon^{-\frac{3}{2}})$, $C_5(\epsilon) = O(\epsilon^{-4})$, and $C_6(\epsilon) = O(\epsilon^{-5})$

Proof. Estimate (3.25) follows immediately by setting $w_h = v_h$ in (3.24) and using the coercivity of the bilinear form $B[\cdot,\cdot]$.

To derive the error in the H^2-norm, we use the error equation,

$$B[v - v_h, w_h] = 0 \qquad \forall w_h \in V_0^h.$$

Using the coercivity of $B[\cdot,\cdot]$, we have

$$C_2(\epsilon)\|v - v_h\|_{H^2}^2 \le B[v - v_h, v - v_h] = B[v - v_h, v] - B[v - v_h, v_h] \tag{3.29}$$
$$= B[v - v_h, v] = B[v - v_h, v - I_h v],$$

where $I_h v$ denotes the finite element interpolant of v onto V_0^h. Noting

$$B(v - v_h, v - I_h v) \leq C\epsilon^{-\frac{1}{2}} \|v - v_h\|_{H^2} \|v - I_h v\|_{H^2},$$

we use standard interpolation theory (cf. Theorems A.0.2 and A.0.3) to get

$$\|v - v_h\|_{H^2} \leq C C_2^{-1}(\epsilon) \epsilon^{-\frac{1}{2}} \|v - I_h v\|_{H^2} \leq C_4(\epsilon) h^{\ell-2} \|v\|_{H^\ell}.$$

Thus, (3.26) holds.

Next, we derive the error in H^1-norm using a standard duality argument. Define $e_h := v - v_h$ and consider the following auxillary problem:

$$\begin{aligned} L_{u^\varepsilon}(\phi) &= \Delta e_h \qquad && \text{in } \Omega, \\ \phi &= 0 && \text{on } \partial\Omega, \\ \Delta\phi &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Using (3.17), we have

$$\|\phi\|_{H^3} \leq C\epsilon^{-2} \|\Delta e_h\|_{H^{-1}}.$$

Since $\|\Delta e_h\|_{H^{-1}} = \sup\{\langle \Delta e_h, w\rangle |\ w \in H_0^1(\Omega),\ \|w\|_{H^1} \leq 1\}$, we have

$$\langle \Delta e_h, w\rangle = (De_h, Dw) \leq \|De_h\|_{L^2} \|Dw\|_{L^2} \leq \|De_h\|_{L^2} \|w\|_{H^1} = \|De_h\|_{L^2}.$$

It follows that

$$\|\phi\|_{H^3} \leq C\epsilon^{-2} \|\Delta e_h\|_{H^{-1}} \leq C\epsilon^{-2} \|De_h\|_{L^2}.$$

Thus,

$$\begin{aligned} \|De_h\|_{L^2}^2 = \langle \Delta e_h, e_h\rangle = B[\phi, e_h] = B[e_h, \phi] = B[e_h, \phi - I_h\phi] \\ &\leq C\epsilon^{-\frac{1}{2}} \|\phi - I_h\phi\|_{H^2} \|e_h\|_{H^2} \\ &\leq C\epsilon^{-\frac{1}{2}} h \|\phi\|_{H^3} \|e_h\|_{H^2} \\ &\leq C\epsilon^{-\frac{5}{2}} h \|De_h\|_{L^2} \|e_h\|_{H^2}. \end{aligned}$$

Hence,

$$\|De_h\|_{L^2} \leq C\epsilon^{-\frac{5}{2}} h \|e_h\|_{H^2}.$$

Combining the above inequality with (3.26), we get (3.27).

To derive the error in the L^2-norm, we consider the following problem:

$$\begin{aligned} L_{u^\varepsilon}(\phi) &= e_h \qquad \text{in } \Omega, \\ \phi &= 0 \qquad \text{on } \partial\Omega, \\ \Delta\phi &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

Using (3.18), we have

$$\|\phi\|_{H^4} \leq C\epsilon^{-3}\|e_h\|_{L^2}.$$

Thus,

$$\begin{aligned} \|e_h\|_{L^2}^2 = B[\phi, e_h] = B[e_h, \phi] &= B[e_h, \phi - I_h\phi] \\ &\leq C\epsilon^{-\frac{1}{2}}\|e_h\|_{H^2}\|\phi - I_h\phi\|_{H^2} \\ &\leq C\epsilon^{-\frac{1}{2}}h^2\|e_h\|_{H^2}\|\psi\|_{H^4} \\ &\leq C\epsilon^{-\frac{7}{2}}h^2\|e_h\|_{H^2}\|e_h\|_{L^2}. \end{aligned}$$

Dividing by $\|e_h\|_{L^2}$, we get (3.28). The proof is complete. □

3.3 Finite Element Method for Problem (3.4)

The goal of this section is to derive optimal order error estimates for the finite element method (3.4). Because of the small parameter ϵ in (3.4), we cannot absorb the strong nonlinearity in the biharmonic term, and as a result, we cannot derive error estimates directly. Furthermore, there is no guarantee that the solution (if it exists) is convex even for small ϵ and h, which makes it difficult to obtain any type of stability result.

To circumvent these difficulties, we use a fixed point technique which relies on the stability properties of the linearized problem studied in the previous section. To this end, we define a linear operator $T_h : V_g^h \mapsto V_g^h$ as follows: For any $v_h \in V_g^h$, define $T_h(v_h) \in V_g^h$ to be the solution of following problem:

$$\begin{aligned} B[v_h - T_h(v_h), w_h] = \epsilon(\Delta v_h, \Delta w_h) - (\det(D^2 v_h), w_h) \\ + (f, w_h) - \left\langle \epsilon^2, \frac{\partial w_h}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w_h \in V_0^h. \end{aligned} \tag{3.30}$$

By Theorem 3.2.1, $T_h(v_h)$ is uniquely defined for all $v_h \in V_g^h$. Notice that the right-hand side of (3.30) is the residual of v_h to equation (3.4), and hence, any fixed point v_h of the mapping T_h (i.e. $T_h(v_h) = v_h$) is a solution to problem (3.4) and vice-versa. In what follows, we shall show that the mapping T_h has a unique fixed point in a small neighborhood

of $I_h u^\epsilon$, the finite element interpolant of u^ϵ.

Next, we set

$$\mathbb{B}_h(\rho) := \big\{ v_h \in V_g^h;\ \|v_h - I_h u^\epsilon\|_{H^2} \le \rho \big\},$$

and for the continuation of the chapter, we assume $u^\epsilon \in H^s(\Omega)$ $(s \ge 3)$ and set $\ell = \min\{r+1, s\}$.

Lemma 3.3.1. *There exists a constant $C_7(\epsilon) = O\big(\epsilon^{\frac{3}{2}(1-n)}\big)$ $(n = 2, 3)$ such that*

$$\|I_h u^\epsilon - T_h(I_h u^\epsilon)\|_{H^2} \le C_7(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell}. \tag{3.31}$$

Proof. To simplify notation, let $\omega_h^\epsilon := I_h u^\epsilon - T_h(I_h u^\epsilon)$, $\alpha^\epsilon := I_h u^\epsilon - u^\epsilon$, and denote $I_h u^{\epsilon,\mu}$ to be the standard mollification of $I_h u^\epsilon$. Then using the mean value theorem and Lemma A.0.1, we have

$$\begin{aligned}
B[\omega_h^\epsilon, \omega_h^\epsilon] &= \epsilon(\Delta(I_h u^\epsilon), \Delta\omega_h^\epsilon) - (\det(D^2(I_h u^\epsilon)) - f, \omega_h^\epsilon) - \left\langle \epsilon^2, \frac{\partial \omega_h^\epsilon}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) + (\det(D^2 u^\epsilon) - \det(D^2(I_h u^\epsilon)), \omega_h^\epsilon) \\
&= \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) + (\det(D^2 u^\epsilon) - \det(D^2(I_h u^{\epsilon,\mu})), \omega_h^\epsilon) \\
&\qquad - (\det(D^2(I_h u^\epsilon)) - \det(D^2(I_h u^{\epsilon,\mu})), \omega_h^\epsilon) \\
&= \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) + (\tilde{\Phi}^\epsilon : (D^2 u^\epsilon - D^2(I_h u^{\epsilon,\mu})), \omega_h^\epsilon) \\
&\qquad - (\det(D^2(I_h u^\epsilon)) - \det(D^2(I_h u^{\epsilon,\mu})), \omega_h^\epsilon) \\
&= \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) - (\tilde{\Phi}^\epsilon D(u^\epsilon - I_h u^{\epsilon,\mu}), D\omega_h^\epsilon) \\
&\qquad - (\det(D^2(I_h u^\epsilon)) - \det(D^2(I_h u^{\epsilon,\mu})), \omega_h^\epsilon) \\
&\le \epsilon \|\Delta\alpha^\epsilon\|_{L^2} \|\Delta\omega_h^\epsilon\|_{L^2} + C\|\tilde{\Phi}^\epsilon\|_{L^2} \|u^\epsilon - I_h u^{\epsilon,\mu}\|_{H^2} \|\omega_h^\epsilon\|_{H^2} \\
&\qquad + \|\det(D^2(I_h u^\epsilon)) - \det(D^2(I_h u^{\epsilon,\mu})\|_{L^2} \|\omega_h^\epsilon\|_{L^2},
\end{aligned} \tag{3.32}$$

where $\tilde{\Phi}^\epsilon = \mathrm{cof}(\tau D^2(I_h u^{\epsilon,\mu}) + (1-\tau) D^2 u^\epsilon)$ for some $\tau \in [0,1]$, and we have used a Sobolev inequality.

Next, when $n = 2$, we bound $\|\tilde{\Phi}^\epsilon\|_{L^2}$ as follows:

$$\begin{aligned}
\|\tilde{\Phi}^\epsilon\|_{L^2} &= \|\mathrm{cof}(\tau D^2(I_h u^{\epsilon,\mu}) + (1-\tau) D^2 u^\epsilon)\|_{L^2} \\
&= \|\tau D^2(I_h u^{\epsilon,\mu}) + (1-\tau) D^2 u^\epsilon\|_{L^2} \\
&\le \|D^2(I_h u^\epsilon)\|_{L^2} + \|D^2 u^\epsilon\|_{L^2} + \|D^2(I_h u^\epsilon) - D^2(I_h u^{\epsilon,\mu})\|_{L^2} \\
&\le C\|D^2 u^\epsilon\|_{L^2} + \|D^2(I_h u^\epsilon) - D^2(I_h u^{\epsilon,\mu})\|_{L^2} \\
&\le C\epsilon^{-\frac{1}{2}} + \|D^2(I_h u^\epsilon) - D^2(I_h u^{\epsilon,\mu})\|_{L^2},
\end{aligned} \tag{3.33}$$

where we have used (2.11).

When $n = 3$, we note that

$$|\tilde{\Phi}^{\epsilon}_{ij}| = |\mathrm{cof}(\tau D^2(I_h u^{\epsilon,\mu}) + (1-\tau)D^2 u^{\epsilon})_{ij}| = |\det\big(\tau D^2(I_h u^{\epsilon,\mu})\big|_{ij} + (1-\tau)D^2 u^{\epsilon}\big|_{ij}\big)|,$$

where $D^2 u^{\epsilon}\big|_{ij}$ denotes the resulting 2×2 matrix after deleting the i^{th} row and j^{th} column of $D^2 u^{\epsilon}$. We can thus conclude that

$$\begin{aligned}
|(\tilde{\Phi}^{\epsilon})_{ij}| &\le 2 \max_{s\neq i, t\neq j} \big(|\tau(D^2(I_h u^{\epsilon,\mu}))_{st} + (1-\tau)(D^2 u^{\epsilon})_{st}|\big)^2 \\
&\le C \max_{s\neq i, t\neq j} \big(|(D^2 u^{\epsilon})_{st}|^2 + |D^2(I_h u^{\epsilon})_{st} - D^2(I_h u^{\epsilon,\mu})_{st}|^2\big) \\
&\le C\left(\|D^2 u^{\epsilon}\|^2_{L^\infty} + \|D^2(I_h u^{\epsilon}) - D^2(I_h u^{\epsilon,\mu})\|^2_{L^\infty}\right).
\end{aligned}$$

Hence, by (2.11)

$$\|\tilde{\Phi}^{\epsilon}\|_{L^2} \le C\epsilon^{-2} + \|D^2(I_h u^{\epsilon}) - D^2(I_h u^{\epsilon,\mu})\|^2_{L^\infty}. \tag{3.34}$$

Using bounds (3.33)–(3.34) into (3.32) and setting $\mu \to 0$ we obtain

$$\begin{aligned}
B[\omega^{\epsilon}_h, \omega^{\epsilon}_h] &\le \epsilon \|\Delta\alpha^{\epsilon}\|_{L^2}\|\Delta\omega^{\epsilon}_h\|_{L^2} + C\epsilon^{\frac{5-3n}{2}}\|\alpha^{\epsilon}\|_{L^2}\|\omega^{\epsilon}_h\|_{L^2} \\
&\le C\epsilon^{\frac{5-3n}{2}}\|\alpha^{\epsilon}\|_{H^2}\|\omega^{\epsilon}_h\|_{H^2}.
\end{aligned}$$

Using the coercivity of the bilinear form $B[\cdot,\cdot]$ we obtain

$$\|\omega^{\epsilon}_h\|_{H^2} \le CC_2^{-1}(\epsilon)\epsilon^{\frac{5-3n}{2}}\|\alpha^{\epsilon}\|_{H^2} \le C\epsilon^{\frac{3}{2}(1-n)}h^{\ell-2}\|u^{\epsilon}\|_{H^\ell}.$$

The proof is complete. □

Lemma 3.3.2. *There exists an $h_0 > 0$ such that for $h \le h_0$ there exists $\rho_0 \in (0,1)$ such that the mapping T_h is a contracting mapping in the ball $\mathbb{B}_h(\rho_0)$ with a contraction factor $\frac{1}{2}$. That is, for any $v_h, w_h \in \mathbb{B}_h(\rho_0)$, there holds*

$$\|T_h(v_h) - T_h(w_h)\|_{H^2} \le \frac{1}{2}\|v_h - w_h\|_{H^2}. \tag{3.35}$$

Proof. Let $v_h, w_h \in \mathbb{B}_h(\rho_0)$ and $z_h \in V_0^h$. Using the definition of $T_h(v_h)$ and $T_h(w_h)$, we have

$$\begin{aligned}
B[T_h(v_h) - T_h(w_h), z_h] &= B[v_h - w_h, z_h] + \epsilon(\Delta(w_h - v_h), \Delta z_h) \\
&\quad - (\det(D^2 w_h) + \det(D^2 v_h), z_h) \\
&= (\Phi^{\epsilon} D(v_h - w_h), Dz_h) + \big(\det(D^2 v_h) - \det(D^2 w_h), z_h\big).
\end{aligned}$$

Adding and subtracting $\det(D^2 v_h^\mu)$, $\det(D^2 w_h^\mu)$, where v_h^μ, w_h^μ denote the standard mollifications of v_h and w_h, respectively, yields

$$\begin{aligned}
&B[T_h(v_h) - T_h(w_h), z_h] \\
&\quad = (\Phi^\epsilon(Dv_h - Dw_h), Dz_h) + (\det(D^2 v_h) - \det(D^2 w_h), z_h) \\
&\quad = (\Phi^\epsilon(Dv_h - Dw_h), Dz_h) + (\det(D^2 v_h^\mu) - \det(D^2 w_h^\mu), z_h) \\
&\qquad\quad + (\det(D^2 v_h) - \det(D^2 v_h^\mu), z_h) + (\det(D^2 w_h^\mu) - \det(D^2 w_h), z_h) \\
&\quad = (\Phi^\epsilon(Dv_h - Dw_h), Dz_h) + (\Psi_h^\mu : (D^2 v_h^\mu - D^2 w_h^\mu), z_h) \\
&\qquad\quad + (\det(D^2 v_h) - \det(D^2 v_h^\mu), z_h) + (\det(D^2 w_h^\mu) - \det(D^2 w_h), z_h),
\end{aligned}$$

where $\Psi_h^\mu = \mathrm{cof}(D^2 v_h^\mu + \tau(D^2 w_h^\mu - D^2 v_h^\mu))$, $\tau \in [0,1]$.

Using Lemma $A.0.1$ we have

$$\begin{aligned}
&B[T_h(v_h) - T_h(w_h), z_h] \qquad (3.36)\\
&\quad = ((\Phi^\epsilon - \Psi_h^\mu)(Dv_h - Dw_h), Dz_h) + (\Psi_h^\mu(Dv_h - Dv_h^\mu), Dz_h) \\
&\qquad + (\Psi_h^\mu(Dw_h^\mu - Dw_h), z_h) + (\det(D^2 v_h) - \det(D^2 v_h^\mu), z_h) \\
&\qquad + (\det(D^2 w_h^\mu) - \det(D^2 w_h), z_h) \\
&\quad \le C\Big\{ \|\Phi^\epsilon - \Psi_h^\mu\|_{L^2} \|v_h - w_h\|_{H^2} \|z_h\|_{H^2} + \|\Psi_h^\mu\|_{L^2} \|z_h\|_{H^2} \big[\|v_h - v_h^\mu\|_{H^2} \\
&\qquad + \|w_h - w_h^\mu\|_{H^2} \big] + \big[\|\det(D^2 v_h) - \det(D^2 v_h^\mu)\|_{L^2} \\
&\qquad + \|\det(D^2 w_h) - \det(D^2 w_h^\mu)\|_{L^2} \big] \|z_h\|_{L^2} \Big\},
\end{aligned}$$

where we have used a Sobolev inequality.

Next, we derive an upper bound for $\|\Phi^\epsilon - \Psi_h^\mu\|_{L^2}$ when $n = 2$ as follows:

$$\begin{aligned}
\|\Phi^\epsilon - \Psi_h^\mu\|_{L^2} &= \|\mathrm{cof}(D^2 u^\epsilon) - \mathrm{cof}(D^2 v_h^\mu + \tau(D^2 w_h^\mu - D^2 v_h^\mu))\|_{L^2} \\
&= \|D^2 u^\epsilon - (D^2 v_h^\mu + \tau(D^2 w_h^\mu - D^2 v_h^\mu))\|_{L^2} \\
&\le \|D^2 u^\epsilon - D^2(I_h u^\epsilon)\|_{L^2} + \|D^2(I_h u^\epsilon) - D^2 v_h\|_{L^2} \\
&\quad + 2\|D^2 v_h - D^2 v_h^\mu\|_{L^2} + \|D^2 w_h - D^2 w_h^\mu\|_{L^2} + \|D^2 v_h - D^2 w_h\|_{L^2} \\
&\le C\big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0 + \|D^2 v_h - D^2 v_h^\mu\|_{L^2} + \|D^2 w_h - D^2 w_h^\mu\|_{L^2}\big).
\end{aligned}$$

When $n = 3$, we note

$$\begin{aligned}
\|(\Phi^\epsilon - \Psi_h^\mu)_{ij}\|_{L^2} &= \|\mathrm{cof}(D^2 u^\epsilon) - \mathrm{cof}(D^2 v_h^\mu + \tau(D^2 w_h^\mu - D^2 v_h^\mu))\|_{L^2} \\
&= \|\det(D^2 u^\epsilon\big|_{ij}) - \det(D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij}))\|_{L^2},
\end{aligned}$$

where we have used the same notation as in Lemma 3.3.1. Thus, using the mean value

theorem,

$$\|(\Phi^\epsilon - \Psi_h^\mu)_{ij}\|_{L^2} = \|\Lambda^{ij} : (D^2 u^\epsilon\big|_{ij} - (D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij})))\|_{L^2},$$

where $\Lambda^{ij} = \mathrm{cof}(D^2 u^\epsilon\big|_{ij} + \lambda(D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij}))) \in \mathbf{R}^{2\times 2}$, $\lambda \in [0,1]$. Bounding $\|\Lambda^{ij}\|_{L^\infty}$, we have

$$\begin{aligned}
\|\Lambda^{ij}\|_{L^\infty} &= \|\mathrm{cof}(D^2 u^\epsilon\big|_{ij} + \lambda(D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij})))\|_{L^\infty} \\
&= \|D^2 u^\epsilon\big|_{ij} + \lambda(D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij}))\|_{L^\infty} \\
&\le C\big(\|D^2 u^\epsilon\|_{L^\infty} + \|D^2 v_h^\mu - D^2 v_h\|_{L^\infty} + \|D^2 w_h^\mu - D^2 w_h\|_{L^\infty} + h^{-\frac{3}{2}}\rho_0\big),
\end{aligned}$$

where we have used the triangle inequality followed by the inverse inequality (A.21). Continuing,

$$\begin{aligned}
&\|(\Phi^\epsilon - \Psi_h^\mu)_{ij}\|_{L^2} \\
&\quad\le \|\Lambda^{ij}\|_{L^\infty}\|D^2 u^\epsilon\big|_{ij} - (D^2 v_h^\mu\big|_{ij} + \tau(D^2 w_h^\mu\big|_{ij} - D^2 v_h^\mu\big|_{ij}))\|_{L^2} \\
&\quad\le C\Big(\|D^2 u^\epsilon\|_{L^\infty} + \|D^2 v_h^\mu - D^2 v_h\|_{L^\infty} + \|D^2 w_h^\mu - D^2 w_h\|_{L^\infty} + h^{-\frac{3}{2}}\rho_0\Big) \\
&\qquad\times \Big(\|D^2 u^\epsilon - D^2 v_h^\mu\|_{L^2} + \|D^2 w_h^\mu - D^2 v_h^\mu\|_{L^2}\Big) \\
&\quad\le C\Big(\|D^2 u^\epsilon\|_{L^\infty} + \|D^2 v_h^\mu - D^2 v_h\|_{L^\infty} + \|D^2 w_h^\mu - D^2 w_h\|_{L^\infty} + h^{-\frac{3}{2}}\rho_0\Big) \\
&\qquad\times \Big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0 + \|D^2 v_h - D^2 v_h^\mu\|_{L^2} + \|D^2 w_h^\mu - D^2 w_h\|_{L^2}\Big).
\end{aligned}$$

It follows that

$$\begin{aligned}
\|\Phi^\epsilon - \Psi_h^\mu\|_{L^2} \le C\Big(&\|D^2 u^\epsilon\|_{L^\infty} + \|D^2 v_h^\mu - D^2 v_h\|_{L^\infty} + \|D^2 w_h^\mu - D^2 w_h\|_{L^\infty} \\
&+ h^{-\frac{3}{2}}\rho_0\Big)\Big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0 + \|D^2 v_h - D^2 v_h^\mu\|_{L^2} + \|D^2 w_h^\mu - D^2 w_h\|_{L^2}\Big).
\end{aligned}$$

Applying these bounds of $\|\Phi^\epsilon - \Psi_h^\mu\|_{L^2}$ to (3.36), setting $\mu \to 0$, and noting (2.11) we obtain

$$B[T_h(v_h) - T_h(w_h), z_h] \le C\big(\epsilon^{2-n} + (n-2)h^{-\frac{3}{2}}\rho_0\big)\big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0)\|v_h - w_h\|_{H^2}\|z_h\|_{H^2}.$$

Using the coercivity of $B[\cdot,\cdot]$, we have

$$\begin{aligned}
&\|T_h(v_h) - T_h(w_h)\|_{H^2} \qquad (3.37)\\
&\qquad\le \left(\frac{\epsilon^{2-n} + (n-2)h^{-\frac{3}{2}}\rho_0}{C_2(\epsilon)}\right)\big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0)\|v_h - w_h\|_{H^2}.
\end{aligned}$$

Set $h_0 = O\left(\frac{C_2(\epsilon)}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}$ in the $n=2$ case and $h_0 = O\left(\frac{C_2(\epsilon)\epsilon}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}$ in the $n=3$ case. Fix $h \le h_0$, and set $\rho_0 = O\big(C_2(\epsilon)\big)$ when $n=2$ and $\rho_0 = O\big(\min\{\epsilon C_2(\epsilon), \epsilon h^{\frac{3}{2}}\}\big)$ when $n=3$. Then it follows from (3.37)

$$\|T_h(v_h) - T_h(w_h)\|_{H^2} \le \frac{1}{2}\|v_h - w_h\|_{H^2} \qquad \forall v_h, w_h \in \mathbb{B}_h(\rho_0).$$

□

Remark 3.3.3. *Noting in the two dimensional case that ρ_0 does not depend on h, we can strengthen Lemma 3.3.2 by stating that there exists $h_0 > 0$ and $\rho_0 > 0$ such that for all $h \le h_0$, T_h is a contracting mapping in the ball $\mathbb{B}_h(\rho_0)$.*

We are now in position to state the first main theorem in this chapter.

Theorem 3.3.4. *There exists $h_1 > 0$ such that for $h \le h_1$, there exists a unique solution u_h^ϵ of (3.4) in the ball $\mathbb{B}_h(\rho_1)$, where $\rho_1 = 2C_7(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell}$. Moreover, there exists a constant $C_8(\epsilon) = O\big(\epsilon^{\frac{3}{2}(1-n)}\big)$ $(n=2,3)$ such that*

$$\|u^\epsilon - u_h^\epsilon\|_{H^2} \le C_8(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell}. \tag{3.38}$$

Proof. In the case $n=2$, set $h_1 = O\left(\frac{\epsilon^{\frac{5}{2}}}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}$. It follows that for $h \le h_1$,

$$\rho_1 = 2C_7(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \le C\epsilon^{-\frac{3}{2}}h_1^{\ell-2}\|u^\epsilon\|_{H^\ell} \le C\epsilon.$$

For the $n=3$ case, set $h_1 = O\left(\min\left\{\left(\frac{\epsilon^5}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}, \left(\frac{\epsilon^4}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{2}{2\ell-7}}\right\}\right)$. Then for $h \le h_1$

$$\rho_1 = 2C_7(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \le C\epsilon^{-3}h_1^{\ell-2}\|u^\epsilon\|_{H^\ell} \le C\epsilon^2,$$

and

$$\rho_1 = 2C_7(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \le Ch^{\frac{3}{2}}\big(\epsilon^{-3}h_1^{\frac{2\ell-7}{2}}\|u^\epsilon\|_{H^\ell}\big) \le Ch^{\frac{3}{2}}\epsilon.$$

Thus, we conclude that $\rho_1 \le \rho_0$ in both cases. We also note that $h_1 \le h_0$ for $n=2,\ n=3$.

Let $v_h \in \mathbb{B}_h(\rho_1)$. Using the triangle inequality and Lemmas 3.3.1 and 3.3.2 we have

$$\begin{aligned}\|I_h u^\epsilon - T_h(v_h)\|_{H^2} &\le \|I_h u^\epsilon - T_h(I_h u^\epsilon)\|_{H^2} + \|T_h(I_h u^\epsilon) - T_h(v_h)\|_{H^2} \\ &\le C_7(\epsilon)h^{\ell-2}\|u\|_{H^\ell} + \frac{1}{2}\|I_h u^\epsilon - v_h\|_{H^2} \le \frac{\rho_1}{2} + \frac{\rho_1}{2} = \rho_1.\end{aligned}$$

Hence, $T_h(v_h) \in \mathbb{B}_h(\rho_1)$. In addition, from Lemma 3.3.2 we know that T_h is a contracting

mapping. Thus, the Brouwer Fixed Point Theorem [57] guarantees that T_h has a unique fixed point $u_h^\epsilon \in \mathbb{B}_h(\rho_1)$, which is the unique solution to (3.4).

To derive the error estimate (3.38), we use the triangle inequality to get

$$\begin{aligned}\|u^\epsilon - u_h^\epsilon\|_{H^2} &\le \|u^\epsilon - I_h u^\epsilon\|_{H^2} + \|I_h u^\epsilon - u_h^\epsilon\|_{H^2} \\ &\le Ch^{\ell-2}\|u\|_{H^\ell} + \rho_1 = C_8(\epsilon)h^{\ell-2}\|u\|_{H^\ell},\end{aligned}$$

where $C_8(\epsilon) := CC_7(\epsilon)$. The proof is complete. □

Theorem 3.3.5. *In addition to the hypothesis of Theorem 3.3.4, assume that the linearized equation is H^4-regular. Then there holds*

$$\|u^\epsilon - u_h^\epsilon\|_{L^2} \le C_9(\epsilon)\Big(\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + \epsilon^{2-n}h^{2\ell-1-\frac{3}{2}n}C_8(\epsilon)\|u^\epsilon\|_{H^\ell}^2\Big), \tag{3.39}$$

where $C_9(\epsilon) = C_8(\epsilon)\epsilon^{-3} = O(\epsilon^{-\frac{3}{2}(1+n)})$.

Proof. Let $e_h^\epsilon := u^\epsilon - u_h^\epsilon$ and $u_h^{\epsilon,\mu}$ denote a standard mollification of u_h^ϵ. We note that e_h^ϵ satisfies the following error equation:

$$\epsilon(\Delta e_h^\epsilon, \Delta v_h) + (\det(D^2 u_h^\epsilon) - \det(D^2 u^\epsilon), v_h) = 0 \qquad \forall v_h \in V_0^h. \tag{3.40}$$

Using (3.40), the mean value theorem, and Lemma A.0.1 we have

$$\begin{aligned}0 &= \epsilon(\Delta e_h^\epsilon, \Delta v_h) + (\det(D^2 u_h^{\epsilon,\mu}) - \det(D^2 u^\epsilon), v_h) \\ &\quad + (\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu}), v_h) \\ &= \epsilon(\Delta e_h^\epsilon, \Delta v_h) - (\tilde{\Phi}^\epsilon D(u_h^{\epsilon,\mu} - u^\epsilon), Dv_h) + (\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu}), v_h),\end{aligned} \tag{3.41}$$

where $\tilde{\Phi}^\epsilon = \mathrm{cof}(D^2 u_h^{\epsilon,\mu} + \tau(D^2 u^\epsilon - D^2 u_h^{\epsilon,\mu}))$, $\tau \in [0,1]$. We note that we have abused the notation of $\tilde{\Phi}^\epsilon$ by using different definitions in different proofs.

Next, The H^4-regular assumption implies that there exists a unique solution to the following problem (cf. Theorem 3.2.1) :

$$\begin{aligned} L_{u^\epsilon}(\psi) &= e_h^\epsilon \qquad \text{in } \Omega, \\ \psi &= 0 \qquad \text{on } \partial\Omega, \\ \Delta\psi &= 0 \qquad \text{on } \partial\Omega. \end{aligned} \tag{3.42}$$

Moreover, there holds

$$\|\psi\|_{H^4} \le C\epsilon^{-3}\|e_h^\epsilon\|_{L^2}. \tag{3.43}$$

Thus, using (3.42) and (3.41), we have

$$\begin{aligned}
\|e_h^\epsilon\|_{L^2}^2 &= (e_h^\epsilon, e_h^\epsilon) = \epsilon(\Delta e_h^\epsilon, \Delta\psi) + (\Phi^\epsilon D\psi, De_h^\epsilon) \\
&= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi) + \epsilon(\Delta e_h^\epsilon, \Delta(I_h\psi)) \\
&\quad + (\Phi^\epsilon De_h^\epsilon, D(I_h\psi)) - \epsilon(\Delta e_h^\epsilon, \Delta(I_h\psi_h)) - (\tilde{\Phi}^\epsilon D(u^\epsilon - u_h^{\epsilon,\mu}), D(I_h\psi)) \\
&\quad - (\det(D^2u_h^\epsilon) - \det(D^2u_h^{\epsilon,\mu}), I_h\psi) \\
&= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi)) \\
&\quad + (\Phi^\epsilon De_h^\epsilon - \tilde{\Phi}^\epsilon D(u^\epsilon - u_h^{\epsilon,\mu}), D(I_h\psi)) - (\det(D^2u_h^\epsilon) - \det(D^2u_h^{\epsilon,\mu}), I_h\psi) \\
&= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi)) + ((\Phi^\epsilon - \tilde{\Phi}^\epsilon)De_h^\epsilon, D(I_h\psi)) \\
&\quad + (\tilde{\Phi}^\epsilon D(u_h^{\epsilon,\mu} - u_h^\epsilon), D(I_h\psi)) + (\det(D^2u_h^{\epsilon,\mu}) - \det(D^2u_h^\epsilon), I_h\psi) \\
&\leq \epsilon\|\Delta e_h^\epsilon\|_{L^2}\|\Delta(\psi - I_h\psi)\|_{L^2} + C\|\Phi^\epsilon\|_{L^2}\|e_h^\epsilon\|_{H^2}\|\psi - I_h\psi\|_{H^2} \\
&\quad + C\|\Phi^\epsilon - \tilde{\Phi}^\epsilon\|_{L^2}\|e_h^\epsilon\|_{H^2}\|I_h\psi\|_{H^2} + C\|\tilde{\Phi}^\epsilon\|_{L^2}\|u_h^{\epsilon,\mu} - u_h^\epsilon\|_{H^2}\|I_h\psi\|_{H^2} \\
&\quad + \|\det(D^2u_h^\epsilon) - \det(D^2u_h^{\epsilon,\mu})\|_{L^2}\|I_h\psi\|_{L^2},
\end{aligned}$$

where we have used Sobolev's inequality.

Next, using (3.3) we obtain

$$\begin{aligned}
\|e_h^\epsilon\|_{L^2}^2 \leq C\Big\{&\epsilon h^2\|e_h^\epsilon\|_{H^2} + h^2\|\Phi^\epsilon\|_{L^2}\|e_h^\epsilon\|_{H^2} + \|\Phi^\epsilon - \tilde{\Phi}^\epsilon\|_{L^2}\|e_h^\epsilon\|_{H^2} \\
&+ \|\tilde{\Phi}^\epsilon\|_{L^2}\|u_h^{\epsilon,\mu} - u_h^\epsilon\|_{L^2} + \|\det(D^2u_h^\epsilon) - \det(D^2u_h^{\epsilon,\mu})\|_{L^2}\Big\}\|\psi\|_{H^4}.
\end{aligned} \tag{3.44}$$

We now bound $\|\Phi^\epsilon - \tilde{\Phi}^\epsilon\|_{L^2}$ when $n = 2$ as follows:

$$\begin{aligned}
\|\Phi^\epsilon - \tilde{\Phi}^\epsilon\|_{L^2} &= \|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2u_h^{\epsilon,\mu} + \tau(D^2u^\epsilon - D^2u_h^{\epsilon,\mu}))\|_{L^2} \\
&= \|D^2u^\epsilon - D^2u_h^{\epsilon,\mu} + \tau(D^2u_h^{\epsilon,\mu} - D^2u^\epsilon)\|_{L^2} \\
&\leq 2\|D^2u^\epsilon - D^2u_h^\epsilon\|_{L^2} + 2\|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2} \\
&\leq 2C_8(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + 2\|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2}.
\end{aligned} \tag{3.45}$$

When $n = 3$, we have (using similar techniques found in Lemma 3.3.2)

$$\begin{aligned}
\|(\Phi^\epsilon - \tilde{\Phi}^\epsilon)_{ij}\|_{L^2} &= \|\mathrm{cof}(D^2u_{ij}^\epsilon) - \mathrm{cof}((D^2u_h^{\epsilon,\mu} + \tau(D^2u^\epsilon - D^2u_h^{\epsilon,\mu}))_{ij})\|_{L^2} \\
&= \|\det(D^2u^\epsilon\big|_{ij}) - \det(D^2u_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2u^\epsilon\big|_{ij} - D^2u_h^{\epsilon,\mu}\big|_{ij}))\|_{L^2} \\
&= \|\Lambda^{ij} : (D^2u^\epsilon\big|_{ij} - (D^2u_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2u^\epsilon\big|_{ij} - D^2u_h^{\epsilon,\mu}\big|_{ij})))\|_{L^2} \\
&\leq 2\|\Lambda^{ij}\|_{L^\infty}\|D^2u^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2} \\
&\leq 2\|\Lambda^{ij}\|_{L^\infty}\big(\|D^2u^\epsilon - D^2u_h^\epsilon\|_{L^2} + \|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2}\big),
\end{aligned}$$

where $\Lambda^{ij} = \mathrm{cof}(D^2u^\epsilon\big|_{ij} + \lambda(D^2u_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2u^\epsilon\big|_{ij} - D^2u_h^{\epsilon,\mu}\big|_{ij})))$. We note that we have abused the notation of Λ^{ij} by defining it differently in two proofs.

Bounding $\|\Lambda^{ij}\|_{L^\infty}$, we note $\Lambda^{ij} \in \mathbf{R}^{2\times 2}$. Thus,

$$\begin{aligned}
\|\Lambda^{ij}\|_{L^\infty} &= \|\mathrm{cof}(D^2u^\epsilon\big|_{ij} + \lambda(D^2u_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2u^\epsilon\big|_{ij} - D^2u_h^{\epsilon,\mu}\big|_{ij})))\|_{L^\infty} \\
&\le \|D^2u^\epsilon\|_{L^\infty} + \|D^2u_h^{\epsilon,\mu}\|_{L^\infty} \\
&\le \|D^2u^\epsilon\|_{L^\infty} + h^{-\frac{3}{2}}\|D^2u_h^\epsilon\|_{L^2} + \|D^2u_h^{\epsilon,\mu} - D^2u_h^\epsilon\|_{L^\infty} \\
&\le C\Big(\epsilon^{-1}h^{-\frac{3}{2}} + C_8(\epsilon)h^{\ell-\frac{7}{2}}\|u^\epsilon\|_{H^\ell} + \|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^\infty}\Big) \\
&\le C\big(\epsilon^{-1}h^{-\frac{3}{2}} + \|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^\infty}\big).
\end{aligned}$$

where we have used the triangle inequality, the inverse inequality, (2.11), and the fact for $h \le h_1$, $C_8(\epsilon)h^{\ell-\frac{7}{2}}\|u^\epsilon\|_{H^\ell} = O(\epsilon^{-1})$.

It follows that

$$\begin{aligned}
\|\Phi^\epsilon - \tilde{\Phi}^\epsilon\|_{L^2} &\le C\big(\epsilon^{-1}h^{-\frac{3}{2}} + \|D^2u_h^{\epsilon,\mu} - D^2u_h^\epsilon\|_{L^\infty}\big) \\
&\qquad \times \big(\|D^2u^\epsilon - D^2u_h^\epsilon\|_{L^2} + \|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2}\big) \\
&\le C\big(\epsilon^{-1}h^{-\frac{3}{2}} + \|D^2u_h^{\epsilon,\mu} - D^2u_h^\epsilon\|_{L^\infty}\big) \\
&\qquad \times \big(C_8(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \|D^2u_h^\epsilon - D^2u_h^{\epsilon,\mu}\|_{L^2}\big).
\end{aligned} \tag{3.46}$$

Substituting (3.45)–(3.46) into (3.44), setting $\mu \to 0$, and applying Theorem 3.3.4, we obtain

$$\begin{aligned}
\|e_h^\epsilon\|_{L^2}^2 &\le C\Big\{\epsilon h^2 + h^2\|\Phi^\epsilon\|_{L^2} + \epsilon^{2-n}h^{\ell+1-\frac{3}{2}n}C_8(\epsilon)\|u^\epsilon\|_{H^\ell}\Big\}\|e_h^\epsilon\|_{H^2}\|\psi\|_{H^4} \\
&\le C\Big\{\epsilon^{-\frac{1}{2}}C_8(\epsilon)h^\ell\|u^\epsilon\|_{H^\ell} + \epsilon^{2-n}h^{2\ell-1-\frac{3}{2}n}C_8^2(\epsilon)\|u^\epsilon\|_{H^\ell}^2\Big\}\|\psi\|_{H^4}.
\end{aligned}$$

Finally, using (3.43) yields

$$\|e_h^\epsilon\|_{L^2}^2 \le CC_8(\epsilon)\epsilon^{-3}\big(\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + \epsilon^{2-n}h^{2\ell-1-\frac{3}{2}n}C_8(\epsilon)\|u^\epsilon\|_{H^\ell}^2\big)\|e_h^\epsilon\|_{L^2}.$$

Dividing by $\|e_h^\epsilon\|_{L^2}$ gives (3.39) with $C_9 := CC_8(\epsilon)\epsilon^{-3}$. □

Remark 3.3.6. *We note that $2\ell - 1 - \frac{3}{2}n \ge \ell$ provided that $r \ge \frac{3}{2}n$ and $u^\epsilon \in H^{1+\frac{3}{2}n}(\Omega)$. Thus, we obtain optimal error estimates in the L^2-norm provided that the exact solution is regular enough.*

We end this section by showing optimal error estimates in the H^1-norm.

Theorem 3.3.7. *Assume that the linearized equation is H^3-regular. Then there exists an*

$h_2 > 0$ *such that for* $h \leq \min\{h_1, h_2\}$*, there holds*

$$\|e_h^\epsilon\|_{H^1} \leq h^{\ell-1} C_{10}(\epsilon) \|u^\epsilon\|_{H^\ell}, \tag{3.47}$$

where $C_{10}(\epsilon) := C C_8(\epsilon) \epsilon^{-\frac{5}{2}} = O(\epsilon^{-(1+\frac{3}{2}n)})$.

Proof. Let ψ be the unique solution to the following problem:

$$\begin{aligned} L_{u^\epsilon}(\psi) &= -\Delta e_h^\epsilon && \text{in } \Omega, \\ \psi &= 0 && \text{on } \partial\Omega, \\ \Delta\psi &= 0 && \text{on } \partial\Omega. \end{aligned}$$

such that

$$\|\psi\|_{H^3} \leq C\epsilon^{-2} \|De_h^\epsilon\|_{L^2}. \tag{3.48}$$

Using the same techniques and notation as in Theorem 3.3.5, we have for $h \leq h_1$

$$\begin{aligned} \|De_h^\epsilon\|_{L^2}^2 &= \epsilon(\Delta e_h^\epsilon, \Delta\psi) + (\Phi^\epsilon D\psi, De_h^\epsilon) \\ &= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi) + \epsilon(\Delta e_h^\epsilon, \Delta(I_h\psi)) \\ &\quad + (\Phi^\epsilon De_h^\epsilon, D(I_h\psi)) - \epsilon(\Delta e_h^\epsilon, \Delta(I_h\psi)) - (\tilde{\Phi} D(u^\epsilon - u_h^{\epsilon,\mu}), D(I_h\psi)) \\ &\quad - (\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu}), I_h\psi) \\ &= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi)) \\ &\quad + (\Phi^\epsilon De_h^\epsilon - \tilde{\Phi} D(u^\epsilon - u_h^{\epsilon,\mu}), D(I_h\psi)) \\ &\quad - (\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu}), I_h\psi) \\ &= \epsilon(\Delta e_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon De_h^\epsilon, D(\psi - I_h\psi)) + ((\Phi^\epsilon - \tilde{\Phi}) De_h^\epsilon, D(I_h\psi)) \\ &\quad + (\tilde{\Phi} D(u_h^{\epsilon,\mu} - u_h^\epsilon), D(I_h\psi)) + (\det(D^2 u_h^{\epsilon,\mu}) - \det(D^2 u_h^\epsilon), I_h\psi) \\ &\leq \epsilon \|\Delta e_h^\epsilon\|_{L^2} \|\Delta(\psi - I_h\psi)\|_{L^2} + C\|\Phi^\epsilon\|_{L^2} \|e_h^\epsilon\|_{H^2} \|\psi - I_h\psi\|_{H^2} \\ &\quad + C\|\Phi^\epsilon - \tilde{\Phi}\|_{L^2} \|De_h^\epsilon\|_{L^2} \|D(I_h\psi)\|_{L^\infty} \\ &\quad + C\|\tilde{\Phi}\|_{L^2} \|u_h^{\epsilon,\mu} - u_h^\epsilon\|_{H^2} \|I_h\psi\|_{H^2} \\ &\quad + \|\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu})\|_{L^2} \|I_h\psi\|_{L^2} \\ &\leq C\Big\{ \epsilon h \|e_h^\epsilon\|_{H^2} + h\|\Phi^\epsilon\|_{L^2} \|e_h^\epsilon\|_{H^2} + \|\Phi^\epsilon - \tilde{\Phi}\|_{L^2} \|De_h^\epsilon\|_{L^2} \\ &\quad + \|\tilde{\Phi}\|_{L^2} \|u_h^{\epsilon,\mu} - u_h^\epsilon\|_{L^2} + \|\det(D^2 u_h^\epsilon) - \det(D^2 u_h^{\epsilon,\mu})\|_{L^2} \Big\} \|\psi\|_{H^3}. \end{aligned}$$

Using bounds (3.45)–(3.46), (2.11), Theorem 3.3.4, and setting $\mu \to 0$, yields

$$\begin{aligned}\|De_h^\epsilon\|_{L^2}^2 &\le C\big(\epsilon h\|e_h^\epsilon\|_{H^2} + \epsilon^{-\frac{1}{2}}h\|e_h^\epsilon\|_{H^2} + C_8(\epsilon)\epsilon^{2-n}h^{\ell+1-\frac{3}{2}n}\|u^\epsilon\|_{H^\ell}\|De_h^\epsilon\|_{L^2}\big)\|\psi\|_{H^3}\\ &\le CC_8(\epsilon)\big(\epsilon^{-\frac{1}{2}}h^{\ell-1} + \epsilon^{2-n}h^{\ell+1-\frac{3}{2}n}\|De_h^\epsilon\|_{L^2}\big)\|u^\epsilon\|_{H^\ell}\|\psi\|_{H^3}.\end{aligned}$$

Using (3.48), we have

$$\|De_h^\epsilon\|_{L^2} \le CC_8(\epsilon)\epsilon^{-2}\Big(\epsilon^{-\frac{1}{2}}h^{\ell-1} + \epsilon^{2-n}h^{\ell+1-\frac{3}{2}n}\|De_h^\epsilon\|_{L^2}\Big)\|u^\epsilon\|_{H^\ell}.$$

Let $h_2 = O\left(\frac{\epsilon^{\frac{1}{2}(5n-3)}}{\|u^\epsilon\|_{H^\ell}}\right)^{\frac{2}{2\ell+2-3n}}$. From the definition of $C_8(\epsilon)$, it follows that for $h \le \min\{h_1, h_2\}$

$$\|De_h^\epsilon\|_{L^2} \le h^{\ell-1}C_8(\epsilon)\epsilon^{-\frac{5}{2}}\|u^\epsilon\|_{H^\ell}.$$

Using Poincaré's inequality, we obtain (3.47). □

3.4 Finite Element Method with Data Perturbations

In this section, we analyze the problem of finding $\hat{u}_h^\epsilon \in V_g^h$ such that

$$-\epsilon(\Delta\hat{u}_h^\epsilon, \Delta v_h) + (\det(D^2\hat{u}_h^\epsilon), v_h) = (\hat{f}, v_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta}\right\rangle_{\partial\Omega}, \tag{3.49}$$

where $\hat{f} = f + \delta f$ and δf is some small perturbation of f.

The reason to study such a problem is twofold. First, we note that (f, v_h) in (3.4) is never computed exactly, but rather some numerical quadrature is used. Thus, we may think of δf as some quadrature error, and we must determine whether this error will affect the convergence rate of $u^\epsilon - \hat{u}_h^\epsilon$. Second, we will find this analysis useful when we study the semigeostrophic equations in Chapter 8.

To study (3.49), we use similar techniques in the previous section. First, we define the linear operator $T_{\hat{f}}$ as follows. Given $v_h \in V_g^h$, define $T_{\hat{f}}(v_h) : V_g^h \mapsto V_g^h$ such that

$$\begin{aligned}B[v_h - T_{\hat{f}}(v_h), w_h] = \epsilon(\Delta v_h, \Delta w_h) - (\det(D^2 v_h), w_h) \\ + (\hat{f}, w_h) - \left\langle \epsilon^2, \frac{\partial w_h}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall w_h \in V_0^h.\end{aligned} \tag{3.50}$$

We then have the following lemma.

Lemma 3.4.1. *There exists constants $C_{11}(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)})$, $C_{12}(\epsilon) = O(\epsilon^{-1})$, such that*

$$\|I_h u^\epsilon - T_{\hat{f}}(I_h u^\epsilon)\|_{H^2} \leq C_{11}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{12}(\epsilon) \|\delta f\|_{H^{-2}}. \tag{3.51}$$

Proof. Let $\omega_h^\epsilon = I_h u^\epsilon - T_{\hat{f}}(I_h u^\epsilon)^*$ and $\alpha^\epsilon = I_h u^\epsilon - u^\epsilon$. By definition, we have

$$\begin{aligned} B[\omega_h^\epsilon, \omega_h^\epsilon] &= \epsilon(\Delta(I_h u^\epsilon), \Delta\omega_h^\epsilon) - (\det(D^2(I_h u^\epsilon)) - \hat{f}, \omega_h^\epsilon) - \left\langle \epsilon^2, \frac{\partial \omega_h^\epsilon}{\partial \eta} \right\rangle_{\partial\Omega} \\ &= \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) + (\det(D^2 u^\epsilon) - \det(D^2(I_h u^\epsilon)), \omega_h^\epsilon) + (\hat{f} - f, \omega_h^\epsilon). \end{aligned}$$

By the proof of Lemma 3.3.1, we have

$$\left| \epsilon(\Delta\alpha^\epsilon, \Delta\omega_h^\epsilon) + (\det(D^2 u^\epsilon) - \det(D^2(I_h u^\epsilon)), \omega_h^\epsilon) \right| \leq C\epsilon^{\frac{5-3n}{2}} \|\alpha^\epsilon\|_{H^2} \|\omega_h^\epsilon\|_{H^2}.$$

Using this bound, and the coercivity of $B[\cdot,\cdot]$, we get

$$\begin{aligned} \|\omega_h^\epsilon\|_{H^2} &\leq C\epsilon^{-1}\big(\epsilon^{\frac{5-3n}{2}} \|\alpha^\epsilon\|_{H^2} + \|\delta f\|_{H^{-2}}\big) \\ &\leq C_{11}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{12}(\epsilon) \|\delta f\|_{H^{-2}}. \end{aligned}$$

□

Lemma 3.4.2. *For $h \leq h_0$, $T_{\hat{f}}$ is a a contracting mapping in the ball $\mathbb{B}_h(\rho_0)$ with a contraction factor $\frac{1}{2}$, where h_0, ρ_0 are defined in Lemma 3.3.2.*

Proof. The proof immediately follows from Lemma 3.3.2 using the fact that for any $v_h, w_h \in V_g^h$,

$$B[T_{\hat{f}}(v_h) - T_{\hat{f}}(w_h), z_h] = B[T_h(v_h) - T_h(w_h), z_h] \qquad \forall z_h \in V_0^h.$$

□

With these two lemmas in hand, we have the following result.

Theorem 3.4.3. *Suppose $\|\delta f\|_{H^{-2}} = O(\epsilon^2)$ in the case $n = 2$, and $\|\delta f\|_{H^{-2}} = O(\min\{\epsilon^3, \epsilon^2 h^{\frac{3}{2}}\})$ when $n = 3$. Then for $h \leq h_1$, there exists a unique solution, $\hat{u}_h^\epsilon$ to (3.49), where h_1 is defined in Theorem 3.3.4. Moreover, there exists constants $C_{13}(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)})$, $C_{14}(\epsilon) = O(\epsilon^{-1})$ such that*

$$\|u^\epsilon - \hat{u}_h^\epsilon\|_{H^2} \leq C_{13}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon) \|\delta f\|_{H^{-2}}. \tag{3.52}$$

*We note, we have abused the notation of ω_h^ϵ, defining it differently in two different proofs of this chapter.

Proof. Let $\rho_2 = 2\big(C_{11}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{12}(\epsilon)\|\delta f\|_{H^{-2}}\big)$. Then from the hypotheses and the definition of h_1, ρ_0, we have $\rho_2 \leq \rho_0$ if $h \leq h_1$. Thus, using Lemmas 3.4.1 and 3.4.2, we have for any $v_h \in \mathbb{B}_h(\rho_2)$

$$\begin{aligned}\|I_h u^\epsilon - T_{\hat{f}}(v_h)\|_{H^2} &\leq \|I_h u^\epsilon - T_{\hat{f}}(I_h u^\epsilon)\|_{H^2} + \|T_{\hat{f}}(I_h u^\epsilon) - T_{\hat{f}}(v_h)\|_{H^2} \\ &\leq C_{11}h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{12}(\epsilon)\|\delta f\|_{H^{-2}} + \frac{1}{2}\|I_h u^\epsilon - v_h\|_{H^2} \\ &\leq \frac{\rho_2}{2} + \frac{\rho_2}{2} = \rho_2.\end{aligned}$$

Thus, $T_{\hat{f}}(v_h) \in \mathbb{B}_h(\rho_2)$, and it follows that $T_{\hat{f}}$ has a unique fixed point, $\hat{u}_h^\epsilon$, which is a solution to (3.49). Furthermore, we have

$$\begin{aligned}\|u^\epsilon - \hat{u}_h^\epsilon\|_{H^2} &\leq \|u^\epsilon - I_h u^\epsilon\|_{H^2} + \|I_h u^\epsilon - \hat{u}_h^\epsilon\|_{H^2} \\ &\leq Ch^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_2 \\ &\leq C_{13}(\epsilon)\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}}.\end{aligned}$$

□

Theorem 3.4.4. *In addition to the hypotheses of Theorem 3.4.3, assume that the linearized equation is H^4-regular. Then the following bound holds:*

$$\|u^\epsilon - \hat{u}_h^\epsilon\|_{L^2} \leq C\epsilon^{-3}\Big\{C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + (\epsilon^{-\frac{1}{2}}C_{14}(\epsilon)h^2 + 1)\|\delta f\|_{H^{-2}} \tag{3.53}$$

$$+ \epsilon^{2-n}h^{\frac{6-3n}{2}}\big[C_{13}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}}\big]^2\Big\}. \tag{3.54}$$

Proof. To derive the L^2-error, we use techniques similar to that of Theorem 3.3.5. Let $\hat{e}_h^\epsilon := u^\epsilon - \hat{u}_h^\epsilon$. We find that $\hat{e}_h^\epsilon$ satisfies the following error equation:

$$\epsilon(\Delta\hat{e}_h^\epsilon, \Delta v_h) + (\det(D^2\hat{u}_h^\epsilon) - \det(D^2u^\epsilon), v_h) + (f - \hat{f}, v_h) = 0 \qquad \forall v_h \in V_0^h.$$

Letting $\hat{u}_h^{\epsilon,\mu}$ denote a standard mollification of $\hat{u}_h^\epsilon$, we have using the mean value theorem and Lemma A.0.1.

$$\begin{aligned}&\epsilon(\Delta\hat{e}_h^\epsilon, \Delta v_h) - (\hat{\Phi}^\epsilon D(\hat{u}_h^{\epsilon,\mu} - u^\epsilon), Dv_h) \\ &\qquad + (\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu}), v_h) + (f - \hat{f}, v_h) = 0 \quad \forall v_h \in V_0^h,\end{aligned} \tag{3.55}$$

where $\hat{\Phi}^\epsilon = \mathrm{cof}\big(D^2\hat{u}_h^{\epsilon,\mu} + \tau(D^2u^\epsilon - D^2u_h^{\epsilon,\mu})\big),\ \tau \in [0,1]$.

Next, we let $\psi \in H^4$ be the solution to the following problem:

$$\begin{aligned} L_{u^\epsilon}(\psi) &= \hat{e}_h^\epsilon \qquad && \text{in } \Omega, \\ \psi &= 0 && \text{on } \partial\Omega, \\ \Delta\psi &= 0 && \text{on } \partial\Omega, \end{aligned}$$

with

$$\|\psi\|_{H^4} \le C\epsilon^{-3}\|\hat{e}_h^\epsilon\|_{L^2}. \tag{3.56}$$

Using (3.55), we have

$$\begin{aligned} \|\hat{e}_h^\epsilon\|_{L^2}^2 &= \epsilon(\Delta\hat{e}_h^\epsilon, \Delta\psi) + (\Phi^\epsilon D\psi, D\hat{e}_h^\epsilon) \qquad (3.57)\\ &= \epsilon(\Delta\hat{e}_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon D\hat{e}_h^\epsilon, D(\psi - I_h\psi)) + \epsilon(\Delta\hat{e}_h^\epsilon, \Delta(I_h\psi)) \\ &\quad + (\Phi^\epsilon D\hat{e}_h^\epsilon, D(I_h\psi)) - \epsilon(\Delta\hat{e}_h^\epsilon, \Delta(I_h\psi_h)) - (\hat{\Phi}^\epsilon D(u^\epsilon - \hat{u}_h^{\epsilon,\mu}), D(I_h\psi)) \\ &\quad - (\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu}), I_h\psi) - (f - \hat{f}, I_h\psi) \\ &= \epsilon(\Delta\hat{e}_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon D\hat{e}_h^\epsilon, D(\psi - I_h\psi)) + (\Phi^\epsilon D\hat{e}_h^\epsilon - \hat{\Phi}^\epsilon D(u^\epsilon - \hat{u}_h^{\epsilon,\mu}), D(I_h\psi)) \\ &\quad - (\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu}), I_h\psi) - (f - \hat{f}, I_h\psi) \\ &= \epsilon(\Delta\hat{e}_h^\epsilon, \Delta(\psi - I_h\psi)) + (\Phi^\epsilon D\hat{e}_h^\epsilon, D(\psi - I_h\psi)) + ((\Phi^\epsilon - \hat{\Phi}^\epsilon)D\hat{e}_h^\epsilon, D(I_h\psi)) \\ &\quad + (\hat{\Phi}^\epsilon D(\hat{u}_h^{\epsilon,\mu} - \hat{u}_h^\epsilon), D(I_h\psi)) + (\det(D^2\hat{u}_h^{\epsilon,\mu}) - \det(D^2\hat{u}_h^\epsilon), I_h\psi) - (f - \hat{f}, I_h\psi) \\ &\le \epsilon\|\Delta\hat{e}_h^\epsilon\|_{L^2}\|\Delta(\psi - I_h\psi)\|_{L^2} + C\|\Phi^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2}\|\psi - I_h\psi\|_{H^2} \\ &\quad + C\|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2}\|I_h\psi\|_{H^2} + C\|\hat{\Phi}^\epsilon\|_{L^2}\|\hat{u}_h^{\epsilon,\mu} - \hat{u}_h^\epsilon\|_{H^2}\|I_h\psi\|_{H^2} \\ &\quad + \|\det(D^2u_h^\epsilon) - \det(D^2u_h^{\epsilon,\mu})\|_{L^2}\|I_h\psi\|_{L^2} + \|\delta f\|_{H^{-2}}\|I_h\psi\|_{H^2} \\ &\le C\Big\{\epsilon h^2\|\hat{e}_h^\epsilon\|_{H^2} + h^2\|\Phi^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2} + \|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2} \\ &\quad + \|\hat{\Phi}^\epsilon\|_{L^2}\|\hat{u}^{\epsilon,\mu} - \hat{u}_h^\epsilon\|_{L^2} + \|\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu})\|_{L^2} + \|\delta f\|_{H^{-2}}\Big\}\|\psi\|_{H^4} \\ &\le C\Big\{\epsilon^{-\frac{1}{2}}h^2\|\hat{e}_h^\epsilon\|_{H^2} + \|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2} + \|\hat{\Phi}^\epsilon\|_{L^2}\|\hat{u}^{\epsilon,\mu} - \hat{u}_h^\epsilon\|_{L^2} \\ &\quad + \|\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu})\|_{L^2} + \|\delta f\|_{H^{-2}}\Big\}\|\psi\|_{H^4}. \end{aligned}$$

For the case $n = 2$, we have

$$\begin{aligned} \|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2} &= \|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}\big(D^2\hat{u}_h^{\epsilon,\mu} + \tau(D^2\hat{u}^\epsilon - D^2\hat{u}_h^{\epsilon,\mu})\big)\|_{L^2} \qquad (3.58)\\ &= \|D^2u^\epsilon - \big(D^2\hat{u}_h^{\epsilon,\mu} + \tau(D^2\hat{u}^\epsilon - D^2\hat{u}_h^{\epsilon,\mu})\big)\|_{L^2} \\ &\le 2\big(\|D^2u^\epsilon - D^2\hat{u}_h^\epsilon\|_{L^2} + \|D^2\hat{u}_h^\epsilon - \hat{u}_h^{\epsilon,\mu}\|_{L^2}\big) \\ &\le 2\Big(C_{13}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}} + \|D^2\hat{u}_h^\epsilon - \hat{u}_h^{\epsilon,\mu}\|_{L^2}\Big). \end{aligned}$$

When $n = 3$, we have

$$\begin{aligned}
\|(\Phi^\epsilon - \tilde{\Phi}^\epsilon)_{ij}\|_{L^2} &= \|\mathrm{cof}(D^2 u^\epsilon_{ij}) - \mathrm{cof}((D^2 \hat{u}_h^{\epsilon,\mu} + \tau(D^2 u^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}))_{ij})\|_{L^2} \\
&= \|\det(D^2 u^\epsilon\big|_{ij}) - \det(D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2 u^\epsilon\big|_{ij} - D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij}))\|_{L^2} \\
&= \|\hat{\Lambda}^{ij} : (D^2 u^\epsilon\big|_{ij} - (D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2 u^\epsilon\big|_{ij} - D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij})))\|_{L^2} \\
&\le 2\|\hat{\Lambda}^{ij}\|_{L^\infty} \|D^2 u^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^2} \\
&\le 2\|\hat{\Lambda}^{ij}\|_{L^\infty} \big(\|D^2 u^\epsilon - D^2 \hat{u}_h^\epsilon\|_{L^2} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^2}\big),
\end{aligned}$$

where $\hat{\Lambda}^{ij} = \mathrm{cof}(D^2 u^\epsilon\big|_{ij} + \lambda(D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2 u^\epsilon\big|_{ij} - D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij})))$.

Bounding $\|\hat{\Lambda}^{ij}\|_{L^\infty}$, we note $\hat{\Lambda}^{ij} \in \mathbf{R}^{2\times 2}$. Thus,

$$\begin{aligned}
\|\hat{\Lambda}^{ij}\|_{L^\infty} &= \|\mathrm{cof}(D^2 u^\epsilon\big|_{ij} + \lambda(D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij} + \tau(D^2 u^\epsilon\big|_{ij} - D^2 \hat{u}_h^{\epsilon,\mu}\big|_{ij})))\|_{L^\infty} \\
&\le \|D^2 u^\epsilon\|_{L^\infty} + \|D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^\infty} \\
&\le \|Du^\epsilon\|_{L^\infty} + Ch^{-\frac{3}{2}} \|D^2 u_h^\epsilon\|_{L^2} + \|D^2 u_h^{\epsilon,\mu} - D^2 u_h^\epsilon\|_{L^\infty} \\
&\le C\Big(\epsilon^{-1} h^{-\frac{3}{2}} + C_{13}(\epsilon) h^{\ell - \frac{7}{2}} \|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon) h^{-\frac{3}{2}} \|\delta f\|_{H^{-2}} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^\infty}\Big) \\
&\le C\big(\epsilon^{-1} h^{-\frac{3}{2}} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^\infty}\big),
\end{aligned}$$

where we have used $\|\delta f\|_{H^{-2}} = O(\epsilon^2 h^{\frac{3}{2}})$ and $C_{13}(\epsilon) h^{\ell - \frac{7}{2}} \|u^\epsilon\|_{H^\ell} = O(\epsilon^{-1})$ for $h \le h_1$.

Using these bounds, we have

$$\begin{aligned}
\|\Phi - \hat{\Phi}^\epsilon\|_{L^2} &\le C\big(\epsilon^{-1} h^{-\frac{3}{2}} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^\infty}\big) \\
&\qquad \times \big(\|D^2 u^\epsilon - D^2 \hat{u}_h^\epsilon\|_{L^2} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^2}\big) \\
&\le C\Big(\epsilon^{-1} h^{-\frac{3}{2}} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^\infty}\Big) \\
&\qquad \times \Big(C_{13}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon) \|\delta f\|_{H^{-2}} + \|D^2 \hat{u}_h^\epsilon - D^2 \hat{u}_h^{\epsilon,\mu}\|_{L^2}\Big).
\end{aligned} \tag{3.59}$$

Using (3.58)–(3.59) into (3.57), setting $\mu \to 0$, and applying (3.56), we obtain

$$\begin{aligned}
\|\hat{e}_h^\epsilon\|_{L^2}^2 &\le C\Big\{\epsilon^{-\frac{1}{2}} h^2 \|\hat{e}_h^\epsilon\|_{H^2} + \|\delta f\|_{H^{-2}} \\
&\qquad + \epsilon^{2-n} h^{\frac{6-3n}{2}} \big(C_{13}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon) \|\delta f\|_{H^{-2}}\big) \|\hat{e}_h^\epsilon\|_{H^2}\Big\} \|\psi\|_{H^4} \\
&\le C\epsilon^{-3}\Big\{\epsilon^{-\frac{1}{2}} h^2 \|\hat{e}_h^\epsilon\|_{H^2} + \|\delta f\|_{H^{-2}} \\
&\qquad + \epsilon^{2-n} h^{\frac{6-3n}{2}} \big(C_{13}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon) \|\delta f\|_{H^{-2}}\big) \|\hat{e}_h^\epsilon\|_{H^2}\Big\} \|\hat{e}_h^\epsilon\|_{L^2}.
\end{aligned}$$

Dividing by $\|\hat{e}_h^\epsilon\|_{L^2}$ and using (3.52) yields

$$\|\hat{e}_h^\epsilon\|_{L^2} \le C\epsilon^{-3}\Big\{C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + (\epsilon^{-\frac{1}{2}}C_{14}(\epsilon)h^2+1)\|\delta f\|_{H^{-2}} \\ + \epsilon^{2-n}h^{\frac{6-3n}{2}}\big[C_{13}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}}\big]^2\Big\}.$$

The proof is complete. □

We end this section with a bound in the H^1-norm.

Theorem 3.4.5. *In addition to the hypothesis of Theorem 3.4.3, assume that the linearized problem is H^3-regular. Moreover, assume $\|\delta f\|_{H^{-2}} = O(\epsilon^{n+1}h^{\frac{3n-6}{2}})$. Then for $h \le \min\{h_1, h_2\}$, the following bound holds:*

$$\|u^\epsilon - \hat{u}_h^\epsilon\|_{H^1} \le C\epsilon^{-2}\Big(C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^{\ell-1}\|u^\epsilon\|_{H^\ell} + (C_{14}(\epsilon)\epsilon^{-\frac{1}{2}}h+1)\|\delta f\|_{H^{-2}}\Big). \tag{3.60}$$

Proof. We use the same technique of the proof of Theorem 3.4.4. Let $\psi \in H^3$ be the solution to the following problem:

$$\begin{aligned} L_{u^\epsilon}(\psi) &= -\Delta\hat{e}_h^\epsilon && \text{in } \Omega, \\ \psi &= 0 && \text{on } \partial\Omega, \\ \Delta\psi &= 0 && \text{on } \partial\Omega, \end{aligned}$$

with

$$\|\psi\|_{H^3} \le C\epsilon^{-2}\|D\hat{e}_h^\epsilon\|_{L^2}. \tag{3.61}$$

Using a similar calculation as in Theorem 3.4.4, we have

$$\begin{aligned} \|D\hat{e}_h^\epsilon\|_{L^2}^2 &\le \epsilon\|\Delta\hat{e}_h^\epsilon\|_{L^2}\|\Delta(\psi - I_h\psi)\|_{L^2} + C\|\Phi^\epsilon\|_{L^2}\|\hat{e}_h^\epsilon\|_{H^2}\|\psi - I_h\psi\|_{H^2} \\ &\quad + C\|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2}\|D\hat{e}_h^\epsilon\|_{L^2}\|D(I_h\psi)\|_{L^\infty} + C\|\hat{\Phi}^\epsilon\|_{L^2}\|\hat{u}_h^{\epsilon,\mu} - \hat{u}_h^\epsilon\|_{H^2}\|I_h\psi\|_{H^2} \\ &\quad + \|\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu})\|_{L^2}\|I_h\psi\|_{L^2} + \|\delta f\|_{H^{-2}}\|I_h\psi\|_{H^2} \\ &\le C\Big\{\epsilon^{-\frac{1}{2}}h\|\hat{e}_h^\epsilon\|_{H^2} + \|\Phi^\epsilon - \hat{\Phi}^\epsilon\|_{L^2}\|D\hat{e}_h^\epsilon\|_{L^2} + \|\hat{\Phi}^\epsilon\|_{L^2}\|\hat{u}_h^{\epsilon,\mu} - \hat{u}_h^\epsilon\|_{H^2} \\ &\quad + \|\det(D^2\hat{u}_h^\epsilon) - \det(D^2\hat{u}_h^{\epsilon,\mu})\|_{L^2} + \|\delta f\|_{H^{-2}}\Big\}\|\psi\|_{H^3}. \end{aligned}$$

Using bounds (3.58)–(3.59), setting $\mu \to 0$, using Theorem 3.4.3, and (3.61), we get

$$\begin{aligned}\|D\hat{e}_h^\epsilon\|_{L^2}^2 &\le C\Big\{C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^{\ell-1}\|u^\epsilon\|_{H^\ell} + (C_{14}(\epsilon)\epsilon^{-\frac{1}{2}}h+1)\|\delta f\|_{H^{-2}} \\ &\qquad + \epsilon^{2-n}h^{\frac{6-3n}{2}}\big(C_{13}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}}\big)\|D\hat{e}_h^\epsilon\|_{L^2}\Big\}\|\psi\|_{H^3} \\ &\le C\epsilon^{-2}\Big\{C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^{\ell-1}\|u^\epsilon\|_{H^\ell} + (C_{14}(\epsilon)\epsilon^{-\frac{1}{2}}h+1)\|\delta f\|_{H^{-2}} \\ &\qquad + \epsilon^{2-n}h^{\frac{6-3n}{2}}\big(C_{13}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + C_{14}(\epsilon)\|\delta f\|_{H^{-2}}\big)\|D\hat{e}_h^\epsilon\|_{L^2}\Big\}\|D\hat{e}_h^\epsilon\|_{L^2}\end{aligned}$$

Using the definition of h_2, it follows that for $h \le \min\{h_1, h_2\}$ and $\|\delta f\|_{H^{-2}} = O(\epsilon^{n+1}h^{\frac{3n-6}{2}})$, we have

$$\|D\hat{e}_h^\epsilon\|_{L^2} \le C\epsilon^{-2}\Big(C_{13}(\epsilon)\epsilon^{-\frac{1}{2}}h^{\ell-1}\|u^\epsilon\|_{H^\ell} + (C_{14}(\epsilon)\epsilon^{-\frac{1}{2}}h+1)\|\delta f\|_{H^{-2}}\Big),$$

and (3.60) follows from Poincaré's inequality. □

3.5 Comments on the Finite Element Approximation of Concave Viscosity Solutions

All of the analysis above was devoted to the existence of a solution to (3.4) and derive error estimates of $\|u^\epsilon - u_h^\epsilon\|$ in various norms, where u^ϵ is the unique solution of (2.8)–(2.10) which converges to the unique convex viscosity solution of (1.11)–(1.12) as $\epsilon \to 0^+$.

We comment on the finite element approximation of the solution of (2.15)–(2.17) in 2-D which approximates the concave viscosity solution of (1.11)–(1.12) (cf. Definition 1.3.3). We denote by $\tilde{u}^\epsilon$ the solution of (2.15)-(2.17). First, we use Definition 2.2.5 to define the variational formulation of (2.15)–(2.17) as follows:

Find $\tilde{u}^\epsilon \in V_g$ such that

$$\epsilon(\Delta\tilde{u}^\epsilon, \Delta v) + (\det(D^2\tilde{u}^\epsilon), v) = (f, v) - \left\langle \epsilon^2, \frac{\partial v}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall v \in V_0. \tag{3.62}$$

Based on (3.62), we define the finite element formulation as to find $\tilde{u}_h^\epsilon \in V_g^h$ such that

$$\epsilon(\Delta\tilde{u}_h^\epsilon, \Delta v_h) + (\det(D^2\tilde{u}_h^\epsilon), v_h) = (f, v_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall v_h \in V_0^h. \tag{3.63}$$

Before we state our results, we introduce the following additional space notation.

$$V_{-g} := \{v \in V;\ v\big|_{\partial\Omega} = -g\}, \qquad V_{-g}^h := \{v_h \in V^h;\ v_h\big|_{\partial\Omega} = -g\}.$$

We then have the following theorem.

Theorem 3.5.1. *Assume $n = 2$ and the hypothesis in Theorems 3.3.4,3.3.5, and 3.3.7 hold. Then for $h \leq \min\{h_1, h_2\}$ there exists a unique solution solving* (3.63)*. Moreover, we have the following error estimates.*

$$\|\tilde{u}^\epsilon - \tilde{u}_h^\epsilon\|_{H^2} \leq C_8(\epsilon) h^{\ell-2} \|\tilde{u}^\epsilon\|_{H^\ell}, \tag{3.64}$$

$$\|\tilde{u}^\epsilon - \tilde{u}_h^\epsilon\|_{H^1} \leq C_{10} h^{\ell-1} \|\tilde{u}^\epsilon\|_{H^\ell}, \tag{3.65}$$

$$\|\tilde{u}^\epsilon - \tilde{u}_h^\epsilon\|_{L^2} \leq C_9(\epsilon) \Big(\epsilon^{-\frac{1}{2}} h^\ell \|\tilde{u}^\epsilon\|_{H^\ell} + C_8(\epsilon) h^{2\ell-4} \|\tilde{u}^\epsilon\|_{H^\ell}^2 \Big). \tag{3.66}$$

Proof. We first note,

$$\det(D^2 v_h) = \det(D^2(-v_h)) \ \forall v_h \in V^h, \ \forall K \in \mathcal{T}_h.$$

Thus, it is clear that $\tilde{u}_h^\epsilon$ is a solution to (3.63) if and only if $\tilde{u}_h^\epsilon = -u_h^\epsilon$, where $u_h^\epsilon \in V_{-g}^h$ is a solution to (3.4). Thus, existence and uniqueness of $\tilde{u}_h^\epsilon$ follows from Theorem (3.3.4).

Next, we let $u^\epsilon \in V_{-g}$ be the solution to (2.8)-(2.10) (with g replaced by $-g$). We then have

$$\|\tilde{u}^\epsilon - \tilde{u}_h^\epsilon\|_{H^2} = \| - u^\epsilon + u_h^\epsilon\|_{H^2} \leq C_8(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell} = C_8(\epsilon) h^{\ell-2} \|\tilde{u}^\epsilon\|_{H^\ell}.$$

Thus, (3.64) holds. (3.65) and (3.66) are obtained similarly. □

3.6 Numerical Experiments and Rates of Convergence

In this section, we provide several 2-D numerical experiments to gauge the efficiency of the finite element method developed in the previous sections. We numerically find the "best" choice of the mesh size h in terms of ϵ, and rates of convergence for both $u - u^\epsilon$ and $u^\epsilon - u_h^\epsilon$. All tests given below are done on the domain $\Omega = (0,1)^2$.

Test 3.1

For this test, we calculate $\|u - u_h^\epsilon\|$ for fixed $h = 0.009$, while varying ϵ in order to approximate $\|u - u^\epsilon\|$, where u is the viscosity solution of (1.11)–(1.12). We use the Argyris element and set to solve problem (3.4) with the following test functions and data:

(a)	$u = e^{(x_1^2+x_2^2)/2}$,	$f = (1 + x_1^2 + x_2^2)e^{(x_1^2+x_2^2)/2}$,	$g = e^{(x_1^2+x_2^2)/2}$.
(b)	$u = x_1^4 + x_2^2$,	$f = 24x_1^2$,	$g = x_1^4 + x_2^2$.
(c)	$u = -\sqrt{(2 - (x_1^2 + x_2^2))}$,	$f = \dfrac{2}{(2 - (x_1^2 + x_2^2))^2}$,	$g = -\sqrt{(2 - (x_1^2 + x_2^2))}$.

The computed solution is compared to the exact solution in Figures 3.1–3.3. We also compute the error $\|u - u_h^\epsilon\|$ and plot the results in Figure 3.4. The figure shows that $\|u - u_h^\epsilon\|_{H^2} = O(\epsilon^{\frac{1}{4}})$. Since we have fixed h very small, we then argue $\|u - u^\epsilon\|_{H^2} \approx \|u - u_h^\epsilon\|_{H^2} = O(\epsilon^{\frac{1}{4}})$. Based on this heuristic argument, we predict that $\|u - u^\epsilon\|_{H^2} = O(\epsilon^{\frac{1}{4}})$. Similarly, from Figure 3.4, we see that $\|u - u^\epsilon\|_{L^2} \approx O(\epsilon)$ and $\|u - u^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$.

We note that the test function in (c) was also used by the authors in [39]. Because u lacks H^2 regularity ($u \in W^{1,p}(\Omega)$, where $p \in [1, 4)$), the method presented in [39] fails to produce accurate approximations. However, as seen from these tests, it appears that this regularity is not needed using the vanishing moment method, and the convergence rate of $\|u - u_h^\epsilon\|$ is unaffected.

Test 3.2

This test is exactly the same as Test 3.1, but now we use a test function that is not convex.

$$u = \sqrt{(2 - (x_1^2 + x_2^2))}, \quad f = \frac{2}{(2 - (x_1^2 + x_2^2))^2}, \quad g = \sqrt{(2 - (x_1^2 + x_2^2))}.$$

As we can see from Figures 3.5–3.7, the solution diverges for positive ϵ and converges for negative ϵ as expected (cf. Theorems 2.2.6, 3.5.1). Also, from Figure 3.6, we see that the computed solution converges at the same rate as in Test 1.

Test 3.3

The purpose of this test is to calculate the rate of convergence of $\|u^\epsilon - u_h^\epsilon\|$ for fixed ϵ in various norms. As in Test 3.1, we use the Argyris element and solve problem (3.4) with boundary condition $\Delta u^\epsilon = \epsilon$ on $\partial\Omega$ being replaced by $\Delta u^\epsilon = \phi^\epsilon$ on $\partial\Omega$. We use the following test functions:

$$\begin{aligned}
&\text{(a) } u^\epsilon = 20x_1^6 + x_2^6, && f^\epsilon = 18000x_1^4x_2^4 - \epsilon(7200x_1^2 + 360x_2^2),\\
&\quad g^\epsilon = 20x_1^6 + x_2^6, && \phi^\epsilon = 600x_1^4 + 30x_2^4.\\
&\text{(b) } u^\epsilon = x_1\sin x_1 + x_2\sin x_2, && f^\epsilon = (2\cos x_1 - x_1\sin x_1)(2\cos x_2 - x_2\sin x_2)\\
&&& \quad - \epsilon(x_1\sin x_1 - 4\cos x_1 + x_2\sin x_2 - 4\cos x_2),\\
&\quad g^\epsilon = x_1\sin x_1 + x_2\sin x_2, && \phi^\epsilon = 2\cos x_1 - x_1\sin x_1 + 2\cos x_2 - x_2\sin x_2.
\end{aligned}$$

After recording the error, we divided each norm by the power of h expected to be the convergence rate by the analysis in Section 3.3. As seen by Table 3.1, the error converges faster than anticipated in all the norms.

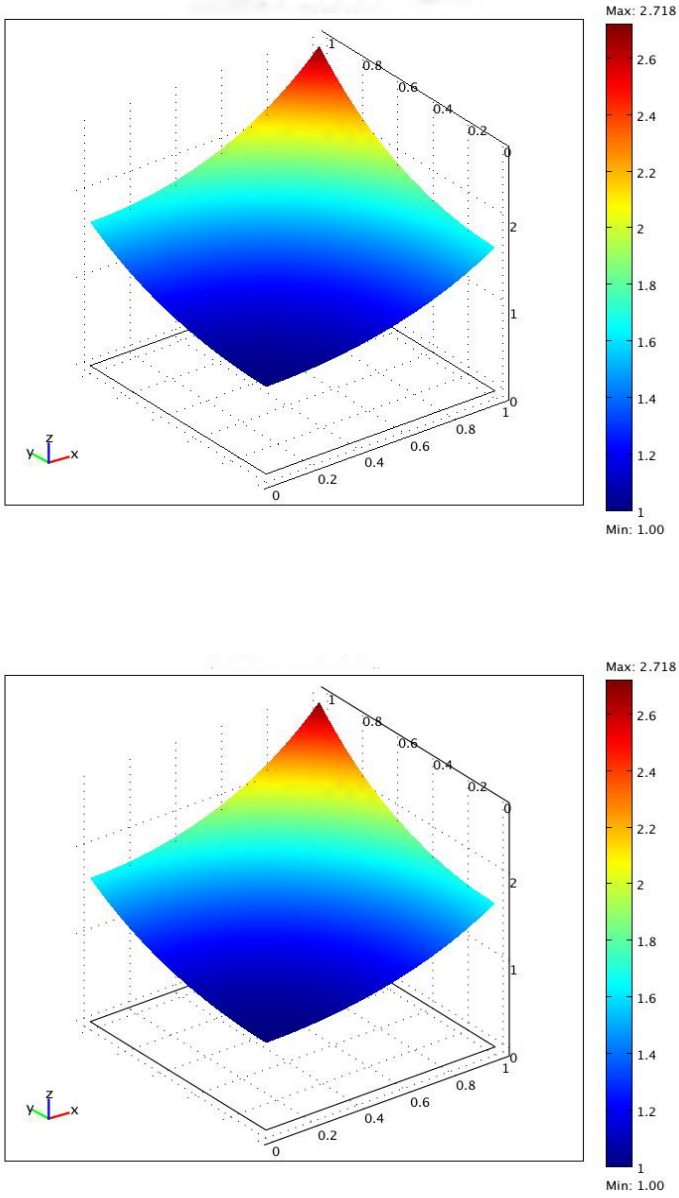


Figure 3.1: Test 3.1a. Computed solution (top) and exact solution (bottom). $\epsilon = 0.0125,\ h = 0.009$

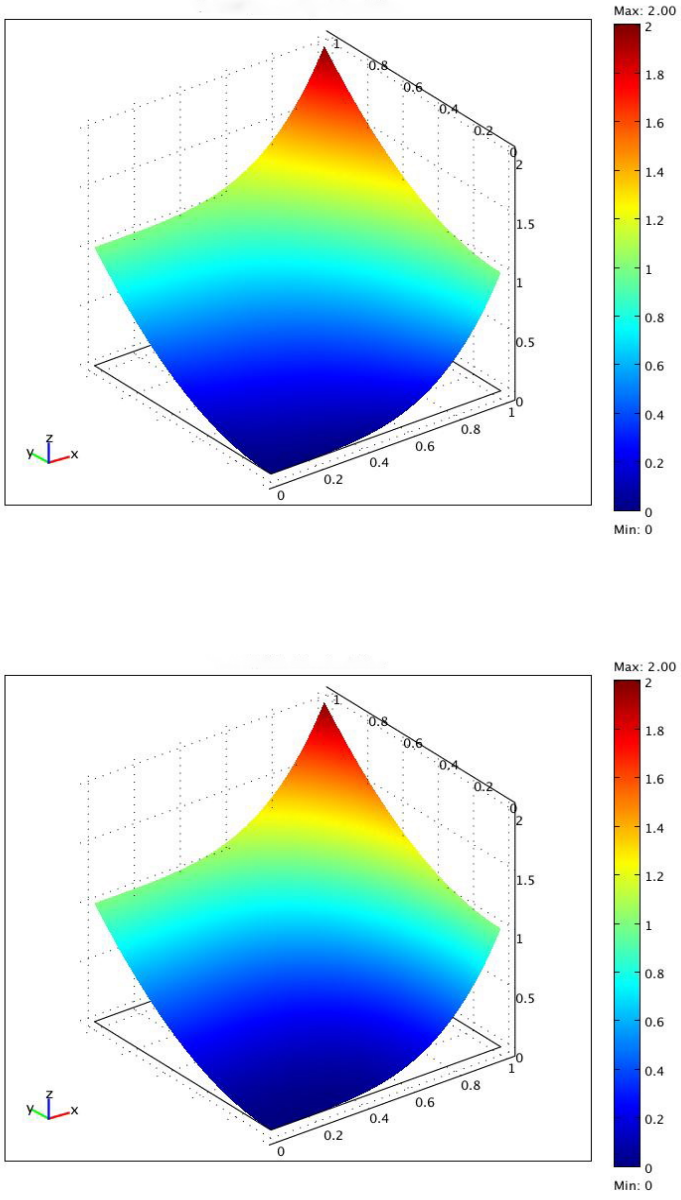


Figure 3.2: Test 3.1b. Computed solution (top) and exact solution (bottom). $\epsilon = 0.0125,\ h = 0.009$

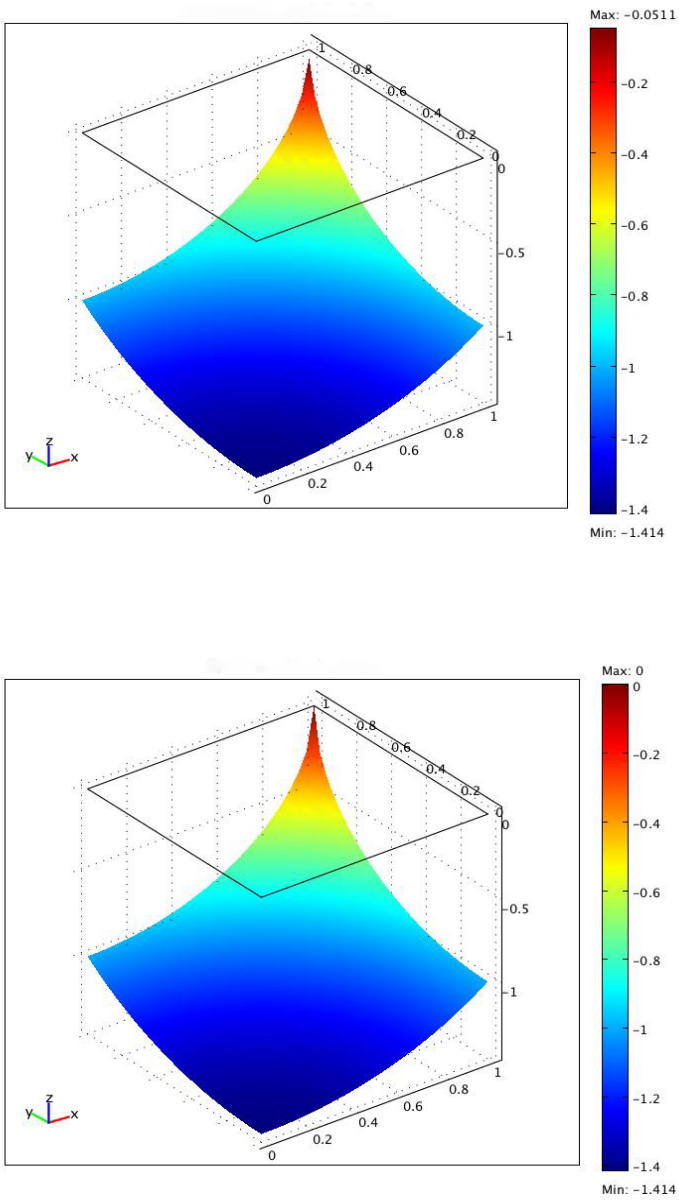


Figure 3.3: Test 3.1c. Computed solution (top) and exact solution (bottom). $\epsilon = 0.0125,\ h = 0.009$

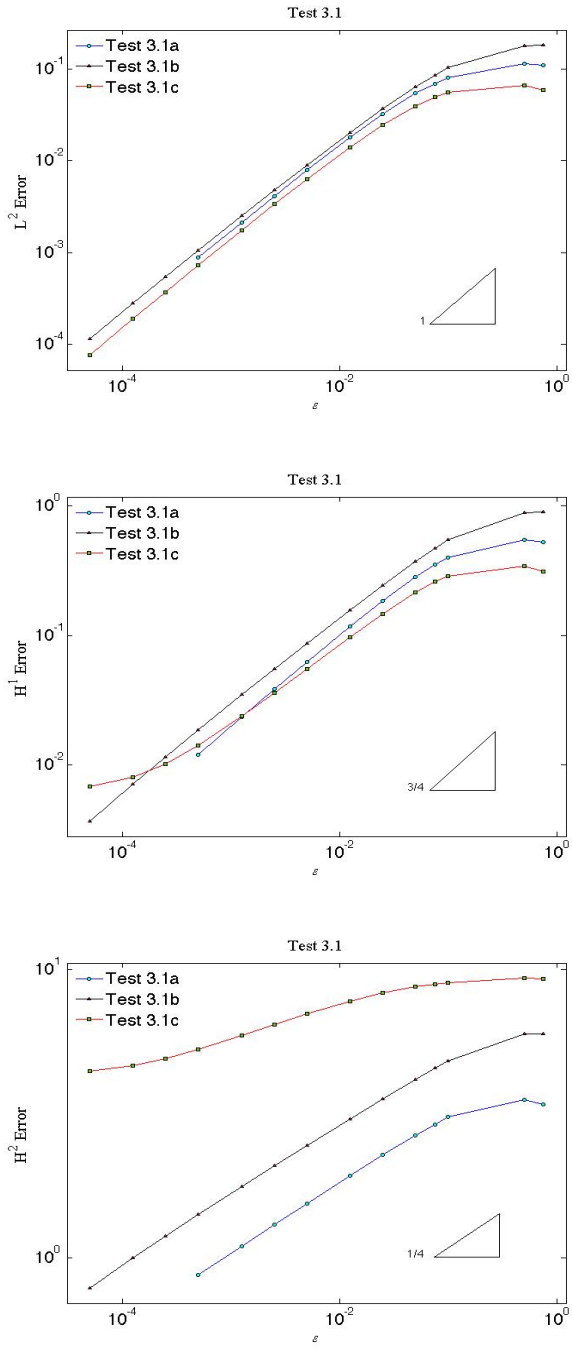


Figure 3.4: Test 3.1. Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ ($h = 0.009$)

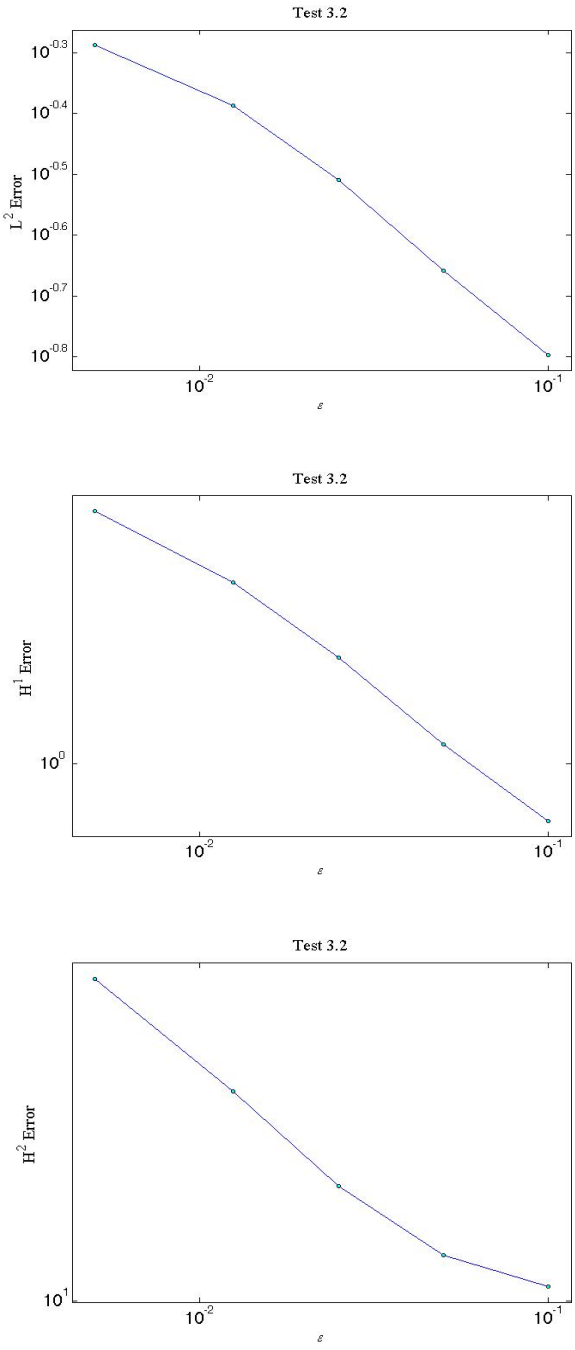


Figure 3.5: Test 3.2. Diverging L^2-error (top) H^1-error (middle) and H^2-error (bottom). ($\epsilon > 0$).

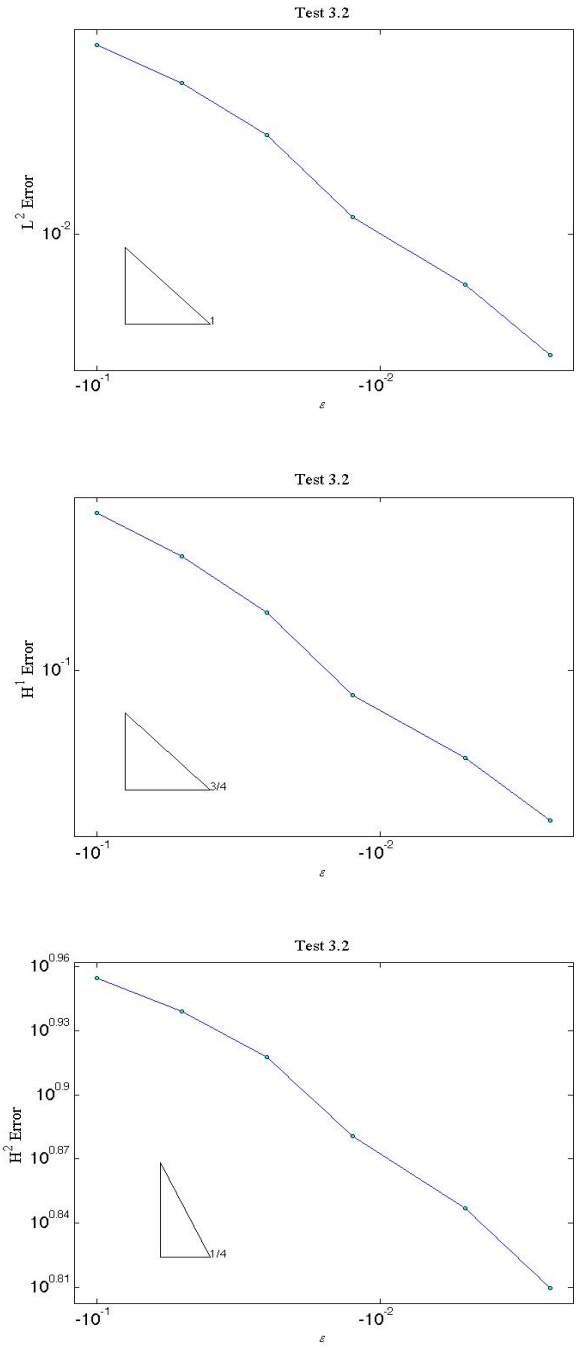


Figure 3.6: Test 3.2: Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ ($h = 0.009$, $\epsilon < 0$).

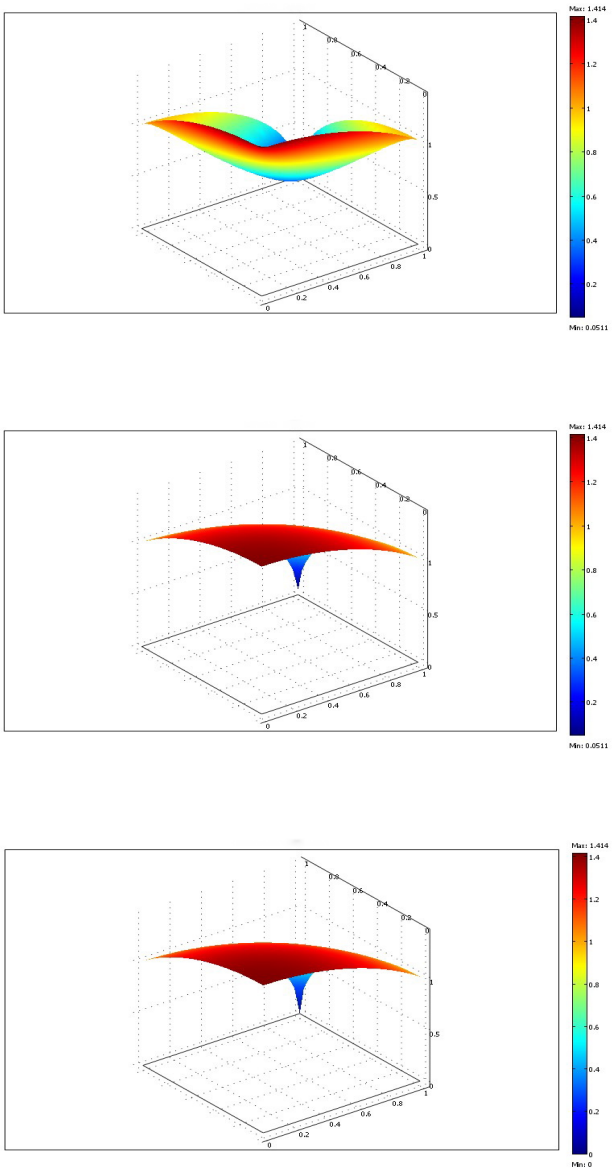


Figure 3.7: Test 3.2. Computed solution using $\epsilon = 0.05$ (top), $\epsilon = -0.05$ (middle) and exact solution (bottom)

Table 3.1: Test 3.3. Change of $\|u^\epsilon - u_h^\epsilon\|$ w.r.t. h ($\epsilon = 0.001$)

	h	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^2}}{h^6}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^1}}{h^5}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^2}}{h^4}$
Test 3.3a	0.083	122.4232319	668.897675	3522.069268
	0.050	69.34721174	317.3313846	1872.077947
	0.031	43.96086573	200.4928789	1116.396482
	0.023	41.81926563	167.8666007	969.5028297
	0.015	27.01059961	104.3140517	618.4873284
	0.012	19.88119861	70.07682598	438.4809442
Test 3.3b	0.083	0.062935746	0.122290283	0.863654524
	0.050	0.033106867	0.06091104	0.435387754
	0.031	0.021321831	0.038607272	0.271545609
	0.023	0.019597137	0.034099981	0.232558269
	0.015	0.012157901	0.021431653	0.145647654
	0.012	0.008152235	0.014239078	0.099470776

Test 3.4

In this section, we fix a relation between ϵ and h to determine a "best" choice for h in terms of ϵ such that the global error $u - u_h^\epsilon$ is the same convergence rate as that of $u - u^\epsilon$. We solve problem (3.4) with the following test functions and parameters.

$$\begin{aligned} &\text{(a) } u = x_1^4 + x_2^2, && f = 24x_1^2, && g = x_1^4 + x_2^2. \\ &\text{(b) } u = 20x_1^6 + x_2^6, && f = 18000x_1^4x_2^4, && g = 20x_1^6 + x_2^6. \end{aligned}$$

To see which relation gives the sought-after convergence rate, we compare the data with a function, $y = \beta x^\alpha$, where $\alpha = 1$ in the L^2-case, $\alpha = \frac{3}{4}$ in the H^1-case and $\alpha = \frac{1}{4}$ in the H^2-case. The constant, β, is determined using a least squares fitting algorithm based on the data.

Figures B.1–B.2 and B.5–B.6 (in Appendix) show that when $h = \epsilon^{\frac{1}{2}}$, $\|u - u_h^\epsilon\|_{L^2} \approx O(\epsilon)$ and $\|u - u_h^\epsilon\|_{H^2} \approx O(\epsilon^{\frac{1}{4}})$. We can conclude from the data that the relation $h = \epsilon^{\frac{1}{2}}$ is the "best choice" for h in terms of ϵ. It can also be seen from Figures B.3–B.4 that when $h = \epsilon$, $\|u - u_h^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$.

Chapter 4

Spectral Methods for the Monge-Ampère Equation

The goal of this chapter is to construct and analyze spectral Galerkin methods to approximate the solution to (2.8)–(2.10) in 2-D and 3-D. As a result, we will obtain convergent spectral methods to approximate the unique convex viscosity solution of the Monge-Ampère equation (1.11)–(1.12). We note that spectral Galerkin methods, as the name implies, are based on the variational formulation of the PDE. However, unlike finite element methods which use low-degree piecewise polynomials with small support as basis functions, spectral methods use high-degree global polynomials as basis functions. As a result, spectral Galerkin methods have considerable advantages and disadvantages compared to standard finite elements.

One advantage of spectral methods is the possibility of exponential convergence given that the function which is being approximated is smooth. These methods are also appealing due to their ability and ease to compute high order (global) derivatives. However, we note that spectral methods can only be practically used on rectangular domains. Also, the basis functions must be chosen carefully to minimize round-off errors, and because evaluation of polynomials of high degree is an unstable numerical procedure.

As in Chapter 3, we are interested in obtaining optimal error bounds of the computed solution that show explicit dependence on the parameter ϵ. We mention that the strategy and analysis presented in this chapter mirrors the work done in Chapter 3 (Sections 3.1–3.3). That is, we employ a combined linearization and fixed point strategy to handle the strong nonlinearity in (2.8). To this end, we study the spectral Galerkin approximation of the linearized problem in Section 4.2. Using the stability property of the linearization, in Section 4.3 we derive optimal error estimates in the energy norm, as well as in the H^1 and L^2-norms.

4.1 Formulation of Spectral Galerkin Method

We adopt the same space notation as in Chapter 3, that is,

$$V := H^2(\Omega), \qquad V_0 := H^2(\Omega) \cap H_0^1(\Omega), \qquad V_g := \{v \in V;\ v|_{\partial\Omega} = g\}.$$

To formulate the spectral Galerkin method, we assume Ω is a rectangular domain. Let L_j denote the j^{th} order Legendre polynomial of a single variable and define the following finite dimensional spaces: For $N_{x_\ell} \geq 2$ ($\ell = 1, 2, 3$), let $N = \sum_{\ell=1}^{n} N_{x_\ell}$ and define

$$\begin{aligned}
V^N &:= \text{span}\{L_0(x), L_2(x), ..., L_N(x)\} && \text{when } n = 1,\\
V^N &:= \text{span}\{L_i(x)L_j(y);\ 1 \leq i \leq N_{x_1},\ 1 \leq j \leq N_{x_2}\} && \text{when } n = 2,\\
V^N &:= \text{span}\{L_i(x)L_j(y)L_k(z);\ 1 \leq i \leq N_{x_1},\ 1 \leq j \leq N_{x_2},\ 1 \leq k \leq N_{x_3}\} && \text{when } n = 3.
\end{aligned}$$

Next, we give the following additional space notation:

$$V_0^N := \{v_N \in V^N;\ v_N\big|_{\partial\Omega} = 0\}, \qquad V_g^N := \{v_N \in V^N;\ v_N\big|_{\partial\Omega} = g\}.$$

It is well-known that V^N has the following approximation property (cf. [12]):

$$\inf_{v_N \in V^N} \|v - v_N\|_{H^j} \leq CN^{j-t}\|v\|_{H^t}, \quad 0 \leq j \leq \min\{t, N\}, \quad t = \min\{s, N+1\}. \tag{4.1}$$

for any $v \in V \cap H^s(\Omega)$.

Based on the weak formulation (3.1), our spectral Galerkin method is defined as seeking $u_N^\epsilon \in V_g^N$ such that for any $v_N \in V_0^N$

$$-\epsilon\big(\Delta u_N^\epsilon, \Delta v_N\big) + \big(\det(D^2 u_N^\epsilon), v_N\big) = (f, v_N) - \left\langle \epsilon^2, \frac{\partial v_N}{\partial \eta} \right\rangle_{\partial\Omega}. \tag{4.2}$$

Remark 4.1.1. *We note that the Galerkin methods* (3.4) *and* (4.2) *have the exact same structure. The key difference is the definition of the finite dimensional spaces V^h and V^N. However, as seen from* (3.3) *and* (4.1)*, both of these spaces have similar approximation properties if we use the relation $h = \frac{1}{N}$. Because of these similarities, the strategy to show optimal error estimates of $u^\epsilon - u_N^\epsilon$ will be similar to that of Chapter 3.*

4.2 Linearization and its Spectral Galerkin Approximation

As in Chapter 3, we first study the linearization of (2.8) in order to analyze equation (4.2). Since derivation and existence of the linearized problem (3.7)–(3.9) was already established in Chapter 3 (cf. Theorems 3.2.1, 3.2.2, and 3.2.3), we only have to study its

spectral Galerkin approximation. That is, we study the spectral Galerkin approximation of the following problem:

$$L_{u^\varepsilon}(v) = \varphi \quad \text{in } \Omega, \tag{4.3}$$

$$v = 0 \quad \text{on } \partial\Omega, \tag{4.4}$$

$$\Delta v = \psi \quad \text{on } \partial\Omega. \tag{4.5}$$

Based on the variational equation (3.10), we define the spectral Galerkin method for (4.3)–(4.5) as seeking $v_N \in V_0^N$ such that

$$B[v_N, w_N] = \langle \varphi, w_N \rangle + \epsilon \left\langle \psi, \frac{\partial w_N}{\partial \eta} \right\rangle \qquad \forall w_N \in V_0^N, \tag{4.6}$$

where $B[\cdot,\cdot]$ is defined by (3.11).

It is clear from the proof of Theorem 3.2.4 and (3.3), (4.1) (with the relation $h = \frac{1}{N}$), that the following error estimates hold.

Theorem 4.2.1. *Suppose $v \in V_0 \cap H^s(\Omega)$ $(s \geq 3)$ is the solution to (4.3)–(4.5). Then there exists a unique $v_N \in V_0^N$ satisfying (4.6). Furthermore, we have the following estimates:*

$$\|v_N\|_{H^2(\Omega)} \leq C_3(\epsilon) \left(\|\varphi\|_{(H^1 \cap H^2)^*} + \|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} \right), \tag{4.7}$$

$$\|v - v_N\|_{H^2(\Omega)} \leq C_4(\epsilon) h^{\ell-2} \|v\|_{H^\ell(\Omega)}, \tag{4.8}$$

$$\|v - v_N\|_{H^1(\Omega)} \leq C_5(\epsilon) h^{\ell-1} \|v\|_{H^\ell(\Omega)}, \tag{4.9}$$

$$\|v - v_N\|_{L^2(\Omega)} \leq C_6(\epsilon) h^{\ell} \|v\|_{H^\ell(\Omega)}, \tag{4.10}$$

where $\ell = \min\{N+1, s\}$, and the constants $C_i(\epsilon)$ $(i = 3,4,5,6)$ have the same order as in Theorem 3.2.4, that is,

$$C_3(\epsilon) = O(\epsilon^{-1}), \qquad C_4(\epsilon) = O(\epsilon^{-\frac{3}{2}}),$$
$$C_5(\epsilon) = O(\epsilon^{-4}), \qquad C_6(\epsilon) = O(\epsilon^{-5}).$$

4.3 Error Analysis for Spectral Galerkin Method (4.2)

The goal of this section is to derive optimal order error estimates for the spectral Galerkin method (4.2). As in Chapter 3, we employ a fixed point technique which will simultaneously provide existence, uniqueness and optimal error estimates in the energy norm.

We first define the linear operator $T_N : V_g^N \to V_g^N$ such that for any $v_N \in V_g^N$, $T_N(v_N)$

denotes the solution to the following problem:

$$B[v_N - T_N(v_N), w_N] = \epsilon(\Delta v_N, \Delta w_N) - (\det(D^2 v_N), w_N) + (f, w_N) - \left\langle \epsilon^2, \frac{\partial w_N}{\partial \eta} \right\rangle_{\partial\Omega} \quad \forall w_N \in V_0^N. \tag{4.11}$$

It is easy to see that the right hand side of (4.11) is the residual of (4.2). Thus, the specific goal of this section is show that T_N has a unique fixed point in a small neighborhood of $I_N u^\epsilon$, where $I_N u^\epsilon$ denotes the spectral Galerkin interpolant of u^ϵ. Next, we set

$$\mathbb{B}_N(\rho) := \{v_N \in V^N \cap V_g;\ \|v_N - I_N u^\epsilon\|_{H^2} \le \rho\}.$$

For the continuation of the chapter, we assume $u^\epsilon \in H^s(\Omega)$ $(s \ge 3)$ and set $\ell = \min\{N+1, s\}$. Our first result measures the effect of the mapping T_N applied towards the center of the ball $\mathbb{B}_h$.

Lemma 4.3.1. *There exists a constant $C_7(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)}) > 0$ $(n = 2, 3)$ such that*

$$\|I_N u^\epsilon - T_N(I_N u^\epsilon)\|_{H^2} \le C_7(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell}. \tag{4.12}$$

Proof. Following the proof of Lemma 3.3.1, we can derive

$$\begin{aligned} B[I_N u^\epsilon - T_N(I_N u^\epsilon), I_N u^\epsilon - T_N(I_N u^\epsilon)] &\le \epsilon \|\Delta(I_N u^\epsilon - u^\epsilon)\|_{L^2} \|\Delta(I_N u^\epsilon - T_N(I_N u^\epsilon))\|_{L^2} \\ &\quad + C\epsilon^{\frac{5-3n}{2}} \|I_N u^\epsilon - T_N(I_N u^\epsilon)\|_{H^2} \|I_N u^\epsilon - u^\epsilon\|_{H^2} \\ &\le C\epsilon^{\frac{5-3n}{2}} \|I_N u^\epsilon - T_N(I_N u^\epsilon)\|_{H^2} \|I_N u^\epsilon - u^\epsilon\|_{H^2}. \end{aligned}$$

Thus, using the coercivity of $B[\cdot,\cdot]$, we have

$$\|I_N u^\epsilon - T_N(I_N u^\epsilon)\|_{H^2} \le C C_2^{-1}(\epsilon) \epsilon^{\frac{5-3n}{2}} \|I_N u^\epsilon - u^\epsilon\|_{H^2} \le C C_2^{-1}(\epsilon) \epsilon^{\frac{5-3n}{2}} N^{\ell-2} \|u^\epsilon\|_{H^\ell},$$

where we have used (4.1), and $C_2(\epsilon)$ is defined in Section 3.2, that is $C_2(\epsilon) = C\min\{\epsilon, \theta\}$. Thus, (4.12) holds with $C_7(\epsilon) = C C_2^{-1}(\epsilon) \epsilon^{\frac{5-3n}{2}} = O(\epsilon^{\frac{3}{2}(1-n)})$. □

Lemma 4.3.2. *There exists an $N_0 > 0$ such that for $N \ge N_0$, there exists a ρ_0 such that $0 < \rho_0 < 1$ and for any $v_N, w_N \in \mathbb{B}_N(\rho_0)$, there holds*

$$\|T_N(v_N) - T_N(w_N)\|_{H^1} \le \frac{1}{2} \|v_N - w_N\|_{H^2}. \tag{4.13}$$

Proof. Using the same techniques used in the proof of Lemma 3.3.2, we can derive that for

any $v_N, w_N \in \mathbb{B}_N(\rho_0),\ z \in V_0^N$

$$B[T_N(v_N) - T_N(w_N), z_N] \\ \leq C\big(\epsilon^{2-n} + (n-2)N^{\frac{3}{2}}\rho_0\big)\big(N^{2-\ell}\|u^\epsilon\|_{H^\ell} + \rho_0\big)\|v_N - w_N\|_{H^2}\|z_N\|_{H^2}.$$

Thus, using the coercivity of the bilinear form $B[\cdot,\cdot]$, we get

$$\|T_N(v_N) - T_N(w_N)\|_{H^2} \leq \Big(\frac{\epsilon^{2-n} + (n-2)N^{\frac{3}{2}}\rho_0}{C_2(\epsilon)}\Big)\big(N^{2-\ell}\|u^\epsilon\|_{H^\ell} + \rho_0\big)\|v_h - w_h\|_{H^2}.$$

When $n = 2$, we set $N_0 = O\left(\frac{\|u^\epsilon\|_{H^\ell}}{C_2(\epsilon)}\right)^{\frac{1}{\ell-2}} = O\left(\frac{\|u^\epsilon\|_{H^\ell}}{\epsilon}\right)^{\frac{1}{\ell-2}}$ and set $N_0 = O\left(\frac{\|u^\epsilon\|_{H^\ell}}{\epsilon^2}\right)^{\frac{1}{\ell-2}}$ in the three dimensional case. Fix $N \geq N_0$ and set $\rho_0 = O(C_2(\epsilon))$ when $n = 2$ and $\rho_0 = O\left(\min\{\epsilon C_2(\epsilon), \epsilon N^{-\frac{3}{2}}\}\right)$ when $n = 3$.

It follows that

$$\|T(v_N) - T(w_N)\|_{H^2} \leq \frac{1}{2}\|v_N - w_N\|_{H^2}.$$

□

We now state the first main result of this chapter.

Theorem 4.3.3. *There exists an $N_1 > 0$ such that for $N \geq N_1$, there exists a unique solution u_N^ϵ to (4.2) in the ball $\mathbb{B}_N(\rho_1)$ where $\rho_1 = 2C_7(\epsilon)N^{2-\ell}\|u^\epsilon\|_{H^\ell}$. Moreover, there exists a constant $C_8(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)})$ such that*

$$\|u^\epsilon - u_N^\epsilon\|_{H^2} \leq C_8(\epsilon)N^{2-\ell}\|u^\epsilon\|_{H^\ell}. \tag{4.14}$$

Proof. In the two dimensional case, we set $N_1 = O\left(\epsilon^{-\frac{5}{2}}\|u^\epsilon\|_{H^\ell}\right)^{\frac{1}{\ell-2}}$. Then for $N \geq N_1$,

$$\rho_1 = 2C_7(\epsilon)N^{2-\ell}\|u^\epsilon\|_{H^\ell} \leq C\epsilon^{-\frac{3}{2}}N_1^{2-\ell}\|u^\epsilon\|_{H^\ell} \leq C\epsilon.$$

In the three dimensional case, we set $N_1 = \left(\max\left\{\left(\epsilon^{-5}\|u^\epsilon\|_{H^\ell}\right)^{\frac{1}{\ell-2}}, \left(\epsilon^{-4}\|u^\epsilon\|_{H^\ell}\right)^{\frac{2}{2\ell-7}}\right\}\right)$. Then for $N \geq N_1$,

$$\rho_1 = 2C_7(\epsilon)N^{2-\ell}\|u^\epsilon\|_{H^\ell} \leq C\epsilon^{-3}N_1^{2-\ell}\|u^\epsilon\|_{H^\ell} \leq C\epsilon^2,$$
$$\rho_1 = 2C_7(\epsilon)N^{2-\ell}\|u^\epsilon\|_{H^\ell} \leq CN^{-\frac{3}{2}}\left(\epsilon^{-3}N_1^{\frac{2}{2\ell-7}}\|u^\epsilon\|_{H^\ell}\right) \leq CN^{-\frac{3}{2}}\epsilon.$$

Thus, we conclude $\rho_1 \leq \rho_0$ for these choices of N_1. We also note that $N_1 \geq N_0$. Next, let

$v_N \in \mathbb{B}_N(\rho_1)$ and use Lemmas 4.3.1 and 4.3.2 to obtain

$$\begin{aligned}\|I_N u^\epsilon - T(v_N)\|_{H^2} &\le \|I_N u^\epsilon - T(I_N u^\epsilon)\|_{H^2} + \|T(I_N u^\epsilon) - T(v_N)\|_{H^2} \\ &\le C_7(\epsilon) N^{2-\ell} \|u^\epsilon\|_{H^\ell} + \frac{1}{2}\|I_N u^\epsilon - v_N\|_{H^2} \\ &\le \frac{\rho_1}{2} + \frac{\rho_1}{2} = \rho_1.\end{aligned}$$

Thus, $T_N(v_N) \in \mathbb{B}_N(\rho_1)$. Using the Brouwer Fixed Point Theorem, we conclude T_N has a unique fixed point, $u_N^\epsilon \in \mathbb{B}_N(\rho_1)$, which is the unique solution to (4.2). To derive (4.14), we use the triangle inequality.

$$\begin{aligned}\|u^\epsilon - u_N^\epsilon\|_{H^2} &\le \|u^\epsilon - I_N u^\epsilon\|_{H^2} + \|I_N u^\epsilon - u_N^\epsilon\|_{H^2} \\ &\le C N^{\ell-2} \|u^\epsilon\|_{H^\ell} + \rho_1 \le C_8(\epsilon) N^{2-\ell} \|u^\epsilon\|_{H^\ell},\end{aligned}$$

where $C_8(\epsilon) = C C_7(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)})$. □

Theorem 4.3.4. *Assuming that the linearized equation is H^4-regular, we have for $N \ge N_1$*

$$\|u^\epsilon - u_N^\epsilon\|_{L^2} \le C_9(\epsilon) \left(\epsilon^{-\frac{1}{2}} N^{-\ell} \|u^\epsilon\|_{H^\ell} + \epsilon^{2-n} C_8(\epsilon) N^{1+\frac{3}{2}n-2\ell} \|u^\epsilon\|_{H^\ell}^2 \right), \tag{4.15}$$

where $C_9(\epsilon) = C\epsilon^{-3} C_8(\epsilon)$.

Proof. The proof is exactly as the one presented in Theorem 3.3.5 with the relation $h = \frac{1}{N}$ and using (3.3),(4.1). □

Theorem 4.3.5. *Assume that the linearized equation is H^3-regular. Then there exists an $N_2 > 0$ such that for $N \ge \max\{N_1, N_2\}$, there holds*

$$\|u^\epsilon - u_N^\epsilon\|_{H^1} \le N^{1-\ell} C_{10}(\epsilon) \|u^\epsilon\|_{H^\ell}. \tag{4.16}$$

where $C_{10}(\epsilon) = C_8(\epsilon)\epsilon^{-\frac{5}{2}}$.

Proof. Using the same methods used in the proof of Theorem 3.3.7, we can derive

$$\|D(u^\epsilon - u_N^\epsilon)\|_{L^2} \le C C_8(\epsilon)\epsilon^{-2} \big(\epsilon^{-\frac{1}{2}} N^{1-\ell} + \epsilon^{2-n} N^{\frac{3}{2}n-\ell-1} \|D(u^\epsilon - u_N^\epsilon)\|_{L^2} \big) \|u^\epsilon\|_{H^\ell}.$$

Set $N_2 = O\left(\|u^\epsilon\|_{H^\ell} \epsilon^{\frac{1}{2}(3-5n)} \right)^{\frac{2}{2\ell+2-3n}}$. Then for $N \ge \max\{N_1, N_2\}$ and using the Poincaré's inequality, we have

$$\|u^\epsilon - u_N^\epsilon\|_{H^1} \le N^{1-\ell} C C_8(\epsilon)\epsilon^{-\frac{5}{2}} \|u^\epsilon\|_{H^\ell}.$$

□

Chapter 5

Mixed Finite Element Methods for the Monge-Ampère Equation

The goal of this chapter is to approximate the Monge-Ampère equation by constructing a family of Hermann-Myoshi mixed finite elements that approximate the solution of (2.12)–(2.14). The mixed formulation is based on rewriting (2.12) as a system of two second-order PDEs by introducing an additional variable that we call σ^ϵ. By breaking equation (2.12) as a system, we are able to approximate (2.12) using only C^0-elements, opposed to C^1-elements (used in Chapter 3) which are computationally expensive.

We note that the theory of mixed finite element methods have been extensively developed in the seventies and eighties for biharmonic problems in 2-D (cf. [27], [17]). We generalize these well-known results to three-dimensional biharmonic problems and other fourth order quasilinear PDEs.

The chapter is organized as follows. In Section 5.1, we derive the mixed formulation for problem (2.12) and propose a family of Hermmann-Myoshi type mixed finite element methods for approximating (2.12). In Section 5.2, we analyze he mixed finite element approximations of the linearized problem (3.7)–(3.8), which will play an important role for the error analysis in Section 5.3. In Section 5.3, we derive optimal order error estimates in the H^1 norm. Our main ideas are to adapt a fixed point argument and to make strong use of the stability property of the linearized problem and its finite element approximations. In Section 5.4, we present computational experiments which confirm the theory presented in the previous sections and also compare the numerical results with the results in Chapter 3. We give a numerical study for determining the "best" choice of mesh size, h, in terms of ϵ, and estimate rates of convergence for both $u - u_h^\epsilon$ and $u - u^\epsilon$ in terms of powers of ϵ. Finally, in Section 5.5, we comment on possible ways to improve the theory presented in Section 5.3.

5.1 Formulation

We note the Hessian matrix, D^2u^ϵ, appears in (2.12) in a nonlinear fashion. Thus, we cannot use Δu^ϵ alone as our additional variable, but rather, we are forced to use $\sigma^\epsilon := D^2u^\epsilon$ as a new variable. Because of this, we rule out the family of Ciarlet-Raviart mixed finite element methods (which use Δu^ϵ as the new variable). On the other hand, this observation suggests we try Hermann-Miyoshi mixed elements.

To define the mixed variational formulation for problem $(2.12) - (2.14)$, we rewrite the PDE into a system of two second order equations.

$$\sigma^\epsilon - D^2u^\epsilon = 0, \tag{5.1}$$

$$-\epsilon\Delta\mathrm{tr}(\sigma^\epsilon) + \det(\sigma^\epsilon) = f. \tag{5.2}$$

Next, we define the following function spaces:

$$\begin{aligned} &V := \{v \in H^1(\Omega)\}, && V_g := \{v \in H^1(\Omega); v|_{\partial\Omega} = g\},\\ &V_0 := \{v \in H^1(\Omega); v|_{\partial\Omega} = 0\}, && W := \{\mu \in V^{n\times n},\ \mu_{ij} = \mu_{ji}\},\\ &W_\epsilon := \{\mu \in W\ \mu\eta\cdot\eta|_{\partial\Omega} = \epsilon\}, && W_0 := \{\mu \in W,\ \mu\eta\cdot\eta|_{\partial\Omega} = 0\}. \end{aligned}$$

We have abused the definition of V, for we have defined it differently in Chapters 3 and 4.

To derive a weak formulation for $(5.1) - (5.2)$, we multiply (5.2) by $v \in V_0$, integrate over Ω, and integrate by parts to get

$$\epsilon\int_\Omega D(\mathrm{tr}(\sigma^\epsilon))\cdot Dv dx + \int_\Omega \det(\sigma^\epsilon)v dx = \int_\Omega fv dx. \tag{5.3}$$

Next, we note

$$\begin{aligned} D(\mathrm{tr}(\sigma^\epsilon))\cdot Dv &= \sum_{i=1}^n\sum_{j=1}^n \frac{\partial\sigma^\epsilon_{jj}}{\partial x_i}\frac{\partial v}{\partial x_i} = \sum_{i=1}^n\sum_{j=1}^n \frac{\partial^3 u^\epsilon}{\partial x_j^2\partial x_i}\frac{\partial v}{\partial x_i}\\ &= \sum_{i=1}^n \mathrm{div}\Big(\frac{\partial^2u^\epsilon}{\partial x_i\partial x_1}, \frac{\partial^2u^\epsilon}{\partial x_i\partial x_2}, ..., \frac{\partial^2u^\epsilon}{\partial x_i\partial x_n}\Big)\frac{\partial v}{\partial x_i} = \sum_{i=1}^n \mathrm{div}((\sigma^\epsilon)^i)\frac{\partial v}{\partial x_i}\\ &= \mathrm{div}(\sigma^\epsilon)\cdot Dv, \end{aligned}$$

where $(\sigma^\epsilon)^i$ denotes the i^{th} row of σ^ϵ.

Using this identity into (5.3), we obtain

$$\epsilon\int_\Omega \mathrm{div}(\sigma^\epsilon)\cdot Dv dx + \int_\Omega \det(\sigma^\epsilon)v dx = \int_\Omega fv dx. \tag{5.4}$$

Let $\mu \in W_0$. Multiplying (5.1) by μ, integrating over Ω and integrating by parts yields

$$\int_\Omega \sigma^\epsilon : \mu dx + \int_\Omega Du^\epsilon \cdot \mathrm{div}(\mu) dx - \int_{\partial\Omega} \mu Du^\epsilon \cdot \eta ds = 0, \tag{5.5}$$

where $\sigma^\epsilon : \mu := \sum_{i,j=1}^{n} \sigma_{ij}^\epsilon \mu_{ij}$.

Letting $\{\tau^{(i)}\}_{i=1}^{n-1}$ denote the standard basis for the tangent space of $\partial\Omega$, we have

$$\begin{aligned} \int_{\partial\Omega} \mu Du^\epsilon \cdot \eta ds &= \int_\Omega \mu\big(\frac{\partial u^\epsilon}{\partial \eta}\eta + \sum_{i=1}^{n-1} \frac{\partial u^\epsilon}{\partial \tau^{(i)}}\tau^{(i)}\big) \cdot \eta \\ &= \sum_{i=1}^{n-1} \int_{\partial\Omega} \mu\eta \cdot \tau^{(i)} \frac{\partial g}{\partial \tau^{(i)}}, \end{aligned} \tag{5.6}$$

where we have used the boundary condition $\mu\eta \cdot \eta = 0$ on $\partial\Omega$. Using (5.6) in (5.5), we have

$$\int_\Omega \sigma^\epsilon : \mu dx + \int_\Omega Du^\epsilon \cdot \mathrm{div}(\mu) dx = \sum_{i=1}^{n-1} \int_{\partial\Omega} \mu\eta \cdot \tau^{(i)} \frac{\partial g}{\partial \tau^{(i)}} ds. \tag{5.7}$$

Based on (5.4) and (5.7) we define the variational formulation for (5.1) − (5.2) as follows: Find $(u^\epsilon, \sigma^\epsilon) \in V_g \times W_\epsilon$ such that

$$(\sigma^\epsilon, \mu) + (\mathrm{div}(\mu), Du^\epsilon) = \langle \tilde{g}, \mu \rangle_{\partial\Omega} \qquad \forall \mu \in W_0, \tag{5.8}$$

$$(\mathrm{div}(\sigma^\epsilon), v) + \epsilon^{-1}(\det(\sigma^\epsilon), v) = (f^\epsilon, v) \qquad \forall v \in V_0, \tag{5.9}$$

where

$$\langle \tilde{g}, \mu \rangle_{\partial\Omega} := \sum_{i=1}^{n-1} \langle \frac{\partial g}{\partial \tau^{(i)}}, \mu n \cdot \tau_i \rangle_{\partial\Omega}, \qquad f^\epsilon := \frac{1}{\epsilon} f.$$

Remark 5.1.1. *Using the identities in Remark 3.1.1, we can define the following alternative variational formulation for* (5.1)-(5.2)*:*

$$\begin{aligned} (\sigma^\varepsilon, \mu) + \big(\mathrm{div}(\mu), Du^\epsilon\big) &= \langle \tilde{g}, \mu \rangle \qquad \forall \mu \in W_0, \\ (\mathrm{div}(\sigma^\epsilon), Dv) - \frac{1}{\epsilon n}\big(\Phi^\epsilon Du^\epsilon, Dv\big) &= (f^\epsilon, v) \qquad \forall v \in V_0, \end{aligned}$$

where again, Φ^ϵ denotes the cofactor matrix of $\sigma^\epsilon = D^2 u^\epsilon$.

Let $\mathcal{T}_h$ be a quasiuniform triangular or rectangular mesh if $n = 2$ and be a quasiuniform tetrahedral or 3-D rectangular mesh if $n = 3$ in the domain Ω. Let $V^h \subset H^1(\Omega)$ be the Lagrange finite element space consisting of continuous piecewise polynomials of degree

k (≥ 2) associated with the mesh $\mathcal{T}_h$. Let

$$V_g^h := V^h \cap V_g, \qquad V_0^h := V^h \cap V_0,$$

$$W_\epsilon^h := [V^h]^{n\times n} \cap W_\epsilon, \qquad W_0^h := [V^h]^{n\times n} \cap W_0.$$

The Hermann-Miyoshi type mixed finite element methods is as follows: Find $(u_h^\epsilon, \sigma_h^\epsilon) \in V_g^h \times W_\epsilon^h$ such that

$$(\sigma_h^\epsilon, \mu_h) + (\mathrm{div}(\mu_h), Du_h^\epsilon) = \langle \tilde{g}, \mu_h\rangle_{\partial\Omega} \qquad \forall \mu_h \in W_0^h, \tag{5.10}$$

$$(\mathrm{div}(\sigma_h^\epsilon), Dv_h) + \epsilon^{-1}(\det(\sigma_h^\epsilon), v_h) = (f^\epsilon, v_h) \qquad \forall v_h \in V_0^h. \tag{5.11}$$

Throughout this chapter, we assume $(\sigma^\epsilon, u^\epsilon)$ is the solution to $(5.8) - (5.9)$. Our goal is to prove there exists a solution, $(\sigma_h^\epsilon, u_h^\epsilon)$, to $(5.10) - (5.11)$ and then estimate $\|\sigma^\epsilon - \sigma_h^\epsilon\|$ and $\|u^\epsilon - u_h^\epsilon\|$ in various norms. To do this, we first analyze the linearization of $(5.8) - (5.9)$ and its mixed finite element approximation.

5.2 Linearized Problem and its Mixed Finite Element Approximations

To build the necessary machinery and technical tools, in this section we shall derive and study the linearization of (5.8)-(5.9) and its mixed finite element approximations.

5.2.1 Derivation of Linearized Problem

As in Chapter 3, we consider the linear problem (3.7)-(3.8), but with an alternative boundary condition.

$$L_{u^\epsilon}(w) = \varphi \qquad \text{in } \Omega, \tag{5.12}$$

$$w = 0 \qquad \text{on } \partial\Omega, \tag{5.13}$$

$$D^2 w\eta \cdot \eta = 0 \qquad \text{on } \partial\Omega, \tag{5.14}$$

where

$$L_{u^\epsilon}(w) = -\epsilon\Delta^2 w + \mathrm{div}(\Phi^\epsilon Dw).$$

To introduce a mixed formulation for (5.12)-(5.14), we rewrite the PDE as

$$\chi - D^2 w = 0, \tag{5.15}$$

$$-\epsilon\Delta \mathrm{tr}(\chi) + \mathrm{div}(\Phi^\epsilon Dw) = \varphi. \tag{5.16}$$

Its variational formulation is then defined as follows: Given $\varphi \in H^{-1}(\Omega)$, find $(\chi, w) \in W_0 \times V_0$ such that

$$(\chi, \mu) + (\mathrm{div}(\mu), Dw) = 0 \qquad \forall \mu \in W_0, \tag{5.17}$$

$$(\mathrm{div}(\chi), Dv) - \epsilon^{-1}(\Phi^\epsilon Dw, Dv) = \epsilon^{-1}(\varphi, v) \qquad \forall v \in V_0. \tag{5.18}$$

It is not hard to show that if (χ, w) solves (5.17)-(5.18) then $w \in H_0^2(\Omega)$ should be a (weak) solution of the fourth order linear PDE

$$-\epsilon \Delta^2 w + \mathrm{div}(\Phi^\epsilon Dw) = \varphi. \tag{5.19}$$

On the other hand, by standard elliptic theory for linear PDEs (cf. [67, 44, 57]), we know that if $\varphi \in H^{-1}(\Omega)$, then the solution $w \in H^3(\Omega)$ so that $\chi = D^2 w \in \left[H^1(\Omega)\right]^{n\times n}$. It is then direct to verify that (χ, w) is a solution to (5.17)-(5.18).

5.2.2 Mixed Finite Element Approximations of the Linearized Problem

Our finite element method for (5.17)-(5.18) is defined as seeking $(\chi_h, w_h) \in W_0^h \times V_0^h$ such that

$$(\chi_h, \mu_h) + (\mathrm{div}(\mu_h), Dw_h) = 0 \qquad \forall \mu_h \in W_0^h, \tag{5.20}$$

$$(\mathrm{div}(\chi_h), Dv_h) - \epsilon^{-1}(\Phi^\epsilon Dw_h, Dv_h) = \epsilon^{-1}(\varphi, v_h) \qquad \forall v_h \in V_0^h. \tag{5.21}$$

Our objective in this section is to first prove existence and uniqueness for problem (5.20)-(5.21) and then derive error estimates in various norms. First, we prove the following inf-sup condition for the finite element pair (W_0^h, V_0^h).

Lemma 5.2.1. *For every $v_h \in V_0^h$, there exists a constant $\beta > 0$, independent of h, such that*

$$\sup_{\mu_h \in W_0^h} \frac{\big(\mathrm{div}(\mu_h), Dv_h\big)}{\|\mu_h\|_{H^1}} \geq \beta \|v_h\|_{H^1}. \tag{5.22}$$

Proof. Given $v_h \in V_0^h$, set $\mu_h = I_{n\times n} v_h$, where $I_{n\times n}$ denotes the $n \times n$ identity matrix. Then $\big(\mathrm{div}(\mu_h), Dv_h\big) = \|Dv_h\|_{L^2}^2 \geq \beta \|v_h\|_{H^1}^2 = \beta \|v_h\|_{H^1} \|\mu_h\|_{H^1}$, where we have used the Poincaré inequality. □

Remark 5.2.2. *By [47, Proposition 1],* (5.22) *implies that there exists a linear operator $\Pi_h : W \to W^h$ such that*

$$\big(\mathrm{div}(\mu - \Pi_h \mu), Dv_h\big) = 0 \qquad \forall v_h \in V_0^h, \tag{5.23}$$

and for $\mu \in W \cap [H^s(\Omega)]^{n\times n},\ s \geq 1$, *there holds*

$$\|\mu - \Pi_h \mu\|_{H^j} \leq C h^{\ell - j} \|\mu\|_{H^\ell} \quad j = 0, 1, \quad 1 \leq \ell \leq \min\{k+1, s\}. \tag{5.24}$$

We note that the above results were proved in the 2*-D case in [47]. However, they also hold in the* 3*-D case as* (5.22) *holds in* 3*-D.*

Theorem 5.2.3. *For any* $\varphi \in V_0^*$, *there exists a unique solution* $(\chi_h, w_h) \in W_0^h \times V_0^h$ *to problem* (5.20)-(5.21).

Proof. Since we are in the finite dimensional case and the problem is linear, it suffices to show uniqueness. Thus, suppose $(\chi_h, w_h) \in W_0^h \times V_0^h$ solves

$$\begin{aligned} (\chi_h, \mu_h) + (\operatorname{div}(\mu_h), D w_h) &= 0 \qquad \forall \mu_h \in W_0^h, \\ (\operatorname{div}(\chi_h), D v_h) - \epsilon^{-1}(\Phi^\epsilon D w_h, D v_h) &= 0 \qquad \forall v_h \in V_0^h. \end{aligned}$$

Let $\mu_h = \chi_h$, $v_h = w_h$ and subtract the two equations to obtain

$$(\chi_h, \chi_h) + \epsilon^{-1}(\Phi^\epsilon D w_h, D w_h) = 0.$$

Since u^ϵ is strictly convex, then Φ^ϵ is positive definite. Thus, there exists $\theta > 0$ such that

$$\|\chi_h\|_{L^2}^2 + \epsilon^{-1}\theta \|D w_h\|_{L^2}^2 \leq 0.$$

Hence, $\chi_h = 0$, $D w_h = 0$, and since $w_h\big|_{\partial\Omega} = 0$, we conclude that $w_h \equiv 0$. The result follows. □

Theorem 5.2.4. *Let* $(\chi, w) \in [H^s(\Omega)]^{n\times n} \cap W_0 \times H^s(\Omega) \cap V_0$ *(*$s \geq 3$*) be the solution to* (5.17)-(5.18) *and* $(\chi_h, w_h) \in W_0^h \times V_0^h$ *solve* (5.20)–(5.21). *Let* $\ell = \min\{k+1, s\}$, *and assume that the linearized problem* (5.12)–(5.14) *is* H^4*-regular. Then the following bounds hold:*

$$\|\chi - \chi_h\|_{L^2} \leq C\epsilon^{-\frac{3}{2}} h^{\ell-2} \big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big) \tag{5.25}$$

$$\|\chi - \chi_h\|_{H^1} \leq C\epsilon^{-\frac{3}{2}} h^{\ell-3} \big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big) \tag{5.26}$$

$$\|w - w_h\|_{H^1} \leq C\epsilon^{-4} h^{\ell-1} \big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big). \tag{5.27}$$

Moreover, for $k \geq 3$ *there also holds*

$$\|w - w_h\|_{L^2} \leq C\epsilon^{-9} h^{\ell} \big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big). \tag{5.28}$$

Proof. Let $I_h w$ denote the standard finite element interpolant of w in V_0^h. Then

$$\begin{aligned}(\Pi_h\chi - \chi_h, \mu_h) + (\operatorname{div}(\mu_h), D(I_h w - w_h)) \\ = (\Pi_h\chi - \chi, \mu_h) + (\operatorname{div}(\mu_h), D(I_h w - w)),\end{aligned} \tag{5.29}$$

$$\begin{aligned}(\operatorname{div}(\Pi_h\chi - \chi_h), Dv_h) - \epsilon^{-1}\big(\Phi^\epsilon D(I_h w - w_h), Dv_h\big) \\ = -\epsilon^{-1}\big(\Phi^\epsilon D(I_h w - w), Dv_h\big).\end{aligned} \tag{5.30}$$

Let $\mu_h = \Pi_h - \chi_h$ and $v_h = I_h w - w_h$ and subtract (5.30) from (5.29) to get

$$\begin{aligned}(\Pi_h\chi - \chi_h, \Pi_h\chi - \chi_h) + \epsilon^{-1}\big(\Phi^\epsilon D(I_h w - w_h), D(I_h w - w_h)\big) \\ = (\Pi_h\chi - \chi, \Pi_h\chi - \chi_h) + \big(\operatorname{div}(\Pi_h\chi - \chi_h), D(I_h w - w)\big) \\ + \epsilon^{-1}\big(\Phi^\epsilon D(I_h w - w), D(I_h w - w_h)\big).\end{aligned}$$

Thus,

$$\begin{aligned}&\|\Pi_h\chi - \chi_h\|_{L^2}^2 + \epsilon^{-1}\theta\|D(I_h w - w_h)\|_{L^2}^2 \\ &\quad\le \|\Pi_h\chi - \chi\|_{L^2}\|\Pi_h\chi - \chi_h\|_{L^2} + \|\Pi_h\chi - \chi_h\|_{H^1}\|D(I_h w - w)\|_{L^2} \\ &\qquad + C\epsilon^{-2}|D(I_h w - w)\|_{L^2}\|D(I_h w - w_h)\|_{L^2} \\ &\quad\le \|\Pi_h\chi - \chi\|_{L^2}\|\Pi_h\chi - \chi_h\|_{L^2} + Ch^{-1}\|\Pi_h\chi - \chi_h\|_{L^2}\|D(I_h w - w)\|_{L^2} \\ &\qquad + C\epsilon^{-2}\|D(I_h w - w)\|_{L^2}\|D(I_h w - w_h)\|_{L^2},\end{aligned}$$

where we have used the inverse inequality (A.21).

Using Cauchy's inequality and rearranging terms yields

$$\begin{aligned}&\|\Pi_h\chi - \chi_h\|_{L^2}^2 + \epsilon^{-1}\|D(I_h w - w_h)\|_{L^2}^2 \\ &\qquad \le C\big(\|\Pi_h\chi - \chi\|_{L^2}^2 + h^{-2}\|I_h w - w\|_{H^1}^2 + \epsilon^{-3}\|I_h w - w\|_{H^1}^2\big).\end{aligned} \tag{5.31}$$

Hence, by the standard interpolation results (cf. Theorem A.0.2), we have

$$\begin{aligned}\|\Pi_h\chi - \chi_h\|_{L^2} &\le C\big(\|\Pi_h\chi - \chi\|_{L^2} + h^{-1}\|I_h w - w\|_{H^1} + \epsilon^{-\frac{3}{2}}\|I_h w - w\|_{H^1}\big) \\ &\le C\epsilon^{-\frac{3}{2}}h^{\ell-2}\big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big),\end{aligned}$$

which by the triangle inequality gets

$$\|\chi - \chi_h\|_{L^2} \le C\epsilon^{-\frac{3}{2}}h^{\ell-2}\big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big).$$

The above estimate and the inverse inequality yield

$$\begin{aligned}\|\chi-\chi_h\|_{H^1} &\le \|\chi-\Pi_h\chi\|_{H^1}+\|\Pi_h\chi-\chi_h\|_{H^1}\\ &\le \|\chi-\Pi_h\chi\|_{H^1}+h^{-1}\|\Pi_h\chi-\chi_h\|_{L^2}\\ &\le C\epsilon^{-\frac{3}{2}}h^{\ell-3}\big(\|\chi\|_{H^\ell}+\|w\|_{H^\ell}\big).\end{aligned}$$

Next, from (5.31) we have

$$\begin{aligned}\|D(I_hw-w_h)\|_{L^2} &\le \sqrt{\epsilon}C\big(\|\chi-\Pi_h\chi\|_{L^2}+h^{-1}\|w-I_hw\|_{H^1}+\epsilon^{-\frac{3}{2}}\|w-I_hw\|_{H^1}\big)\\ &\le C\epsilon^{-1}h^{\ell-2}\big(\|\chi\|_{H^\ell}+\|w\|_{H^\ell}\big).\end{aligned}\tag{5.32}$$

To derive (5.27), we consider the following auxiliary problem: Find $z\in H^2(\Omega)\cap H_0^1(\Omega)$ such that

$$\begin{aligned}-\epsilon\Delta^2 z+\operatorname{div}(\Phi^\epsilon Dz) &= -\Delta(w-w_h) && \text{in } \Omega,\\ D^2z\eta\cdot\eta &= 0 && \text{on } \partial\Omega.\end{aligned}$$

By hypothesis, there exists a unique solution $z\in H_0^1(\Omega)\cap H^3(\Omega)$ and (cf. Theorem 3.2.2)

$$\|z\|_{H^3}\le C\epsilon^{-2}\|\Delta(w-w_h)\|_{H^{-1}}\le C\epsilon^{-2}\|D(w-w_h)\|_{L^2}.\tag{5.33}$$

Setting $\kappa=D^2z$, it is easy to verify that $(\kappa,z)\in W_0\times V_0$ satisfy

$$\begin{aligned}(\kappa,\mu)+\big(\operatorname{div}(\mu),Dz\big) &= 0 && \forall\mu\in W_0,\\ \big(\operatorname{div}(\kappa),Dv\big)-\epsilon^{-1}\big(\Phi^\epsilon Dz,Dv\big) &= \epsilon^{-1}(D(w-w_h),Dv) && \forall v\in V_0.\end{aligned}$$

We also see that (5.20)–(5.21) produce the following error equations:

$$(\chi-\chi_h,\mu_h)+(\operatorname{div}(\mu_h),D(w-w_h))=0 \qquad \forall\mu_h\in W_0^h,\tag{5.34}$$

$$(\operatorname{div}(\chi-\chi_h),Dv_h)-\epsilon^{-1}(\Phi^\epsilon D(w-w_h),Dv_h)=0 \qquad \forall v_h\in V_0^h.\tag{5.35}$$

Thus,

$$\begin{aligned}\epsilon^{-1}\|D(w-w_h)\|_{L^2}^2 &= \big(\operatorname{div}(\kappa),D(w-w_h)\big)-\epsilon^{-1}\big(\Phi^\epsilon Dz,D(w-w_h)\big)\\ &= \big(\operatorname{div}(\kappa-\Pi_h\kappa),D(w-w_h)\big)-\epsilon^{-1}\big(\Phi^\epsilon Dz,D(w-w_h)\big)\\ &\quad+\big(\operatorname{div}(\Pi_h\kappa),D(w-w_h)\big)\\ &= \big(\operatorname{div}(\kappa-\Pi_h\kappa),D(w-I_hw)\big)-\epsilon^{-1}\big(\Phi^\epsilon Dz,D(w-w_h)\big)\\ &\quad+\big(\chi_h-\chi,\Pi_h\kappa\big)\end{aligned}$$

$$
\begin{aligned}
&= \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)\big) - \epsilon^{-1}\big(\Phi^\epsilon Dz, D(w - w_h)\big) \\
&\quad + \big(\chi_h - \chi, \Pi_h\kappa - \kappa\big) + \big(\chi_h - \chi, \kappa\big) \\
&= \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)\big) - \epsilon^{-1}\big(\Phi^\epsilon Dz, D(w - w_h)\big) \\
&\quad + \big(\chi_h - \chi, \Pi_h\kappa - \kappa\big) + \big(\mathrm{div}(\chi - \chi_h), Dz\big) \\
&= \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)\big) + (\chi_h - \chi, \Pi_h\kappa - \kappa) \\
&\quad + (\mathrm{div}(\chi - \chi_h), D(z - I_h z)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(w - w_h), D(z - I_h z)\big) \\
&\le \|\mathrm{div}(\kappa - \Pi_h\kappa)\|_{L^2}\|D(w - I_h w)\|_{L^2} + \|\chi_h - \chi\|_{L^2}\|\Pi_h\kappa - \kappa\|_{L^2} \\
&\quad + \|\mathrm{div}(\chi - \chi_h)\|_{L^2}\|D(z - I_h z)\|_{L^2} \\
&\quad + C\epsilon^{-2}\|D(z - I_h z)\|_{L^2}\|D(w - w_h)\|_{L^2} \\
&\le C\Big(\|D(w - I_h w)\|_{L^2} + h\|\chi_h - \chi\|_{L^2} + h^2\|\mathrm{div}(\chi - \chi_h)\|_{L^2} \\
&\quad + \epsilon^{-2}h^2\|D(w - w_h)\|_{L^2}\Big)\|z\|_{H^3}.
\end{aligned}
$$

Then, by (5.25),(5.26),(5.32), and (5.33), we have

$$
\|D(w - w_h)\|_{L^2} \le C\epsilon^{-4}h^{\ell-1}\big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big).
$$

Thus, (5.27) holds.

To derive the L^2-norm estimate for $w - w_h$, we consider the following auxiliary problem: Find $(\kappa, z) \in W_0 \times V_0$ such that

$$
\begin{aligned}
(\kappa, \mu) + \big(\mathrm{div}(\mu), Dz\big) &= 0 && \forall \mu \in W_0, \\
\big(\mathrm{div}(\kappa), Dv\big) - \epsilon^{-1}\big(\Phi^\epsilon Dz, Dv\big) &= \epsilon^{-1}(w - w_h, v) && \forall v \in V_0.
\end{aligned}
$$

By hypothesis, we have (cf. Theorem 3.2.2)

$$
\|z\|_{H^4} \le C\epsilon^{-3}\|w - w_h\|_{L^2}. \tag{5.36}
$$

We then have

$$
\begin{aligned}
\epsilon^{-1}\|w - w_h\|_{L^2}^2 &= \big(\mathrm{div}(\kappa), D(w - w_h)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(w - w_h), Dz\big) \\
&= \big(\mathrm{div}(\Pi_h\kappa), D(w - w_h)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(w - w_h), Dz\big) \\
&\qquad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(w - w_h\big) \\
&= (\chi_h - \chi, \Pi_h\kappa) - \epsilon^{-1}(\Phi^\epsilon Dz, D(w - w_h)) \\
&\qquad + (\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w))
\end{aligned}
$$

$$
\begin{aligned}
&= (\chi_h - \chi, \kappa) + (\chi_h - \chi, \Pi_h\kappa - \kappa) \\
&\qquad - \epsilon^{-1}(\Phi^\epsilon Dz, D(w - w_h)) + (\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)) \\
&= (\mathrm{div}(\chi - \chi_h), Dz) - \epsilon^{-1}(\Phi^\epsilon D(w - w_h), Dz) \\
&\qquad + (\chi_h - \chi, \Pi_h\kappa - \kappa) + (\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)) \\
&= (\mathrm{div}(\chi - \chi_h), D(z - I_h z)) - \epsilon^{-1}(\Phi^\epsilon D(w - w_h), D(z - I_h z)) \\
&\qquad + (\chi_h - \chi, \Pi_h\kappa - \kappa) + (\mathrm{div}(\kappa - \Pi_h\kappa), D(w - I_h w)) \\
&\le \big(\|\mathrm{div}(\chi - \chi_h)\|_{L^2} + C\epsilon^{-2}\|D(w - w_h)\|_{L^2}\big)\|D(z - I_h z)\|_{L^2} \\
&\quad + \|\chi_h - \chi\|_{L^2}\|\Pi_h\kappa - \kappa\|_{L^2} + \|\mathrm{div}(\kappa - \Pi_h\kappa)\|_{L^2}\|D(w - I_h w)\|_{L^2} \\
&\le Ch^3\big(\|\chi - \chi_h\|_{H^1} + \epsilon^{-2}\|w - w_h\|_{H^1}\big)\|z\|_{H^4} \\
&\qquad + Ch^2\|\chi_h - \chi\|_{L^2}\|\kappa\|_{H^2} + Ch\|w - I_h w\|_{H^1}\|\kappa\|_{H^2} \\
&\le C\epsilon^{-6}h^\ell\big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big)\|z\|_{H^4} \\
&\le C\epsilon^{-9}h^\ell\big(\|\chi\|_{H^\ell} + \|w\|_{H^\ell}\big)\|w - w_h\|_{L^2},
\end{aligned}
$$

where we have used (5.25),(5.26),(5.27), (5.36), and the assumption $k \ge 3$. Dividing the above inequality by $\|w - w_h\|_{L^2}$, we get (5.28). The proof is complete. □

Remark 5.2.5. *By Theorem 3.2.2, the hypothesis concerning the regularity of the linearized problem in Theorem 5.2.4 holds provided $\partial\Omega \in C^4$.*

5.3 Error Analysis for Finite Element Method (5.10)–(5.11)

The goal of this section is to derive error estimates for the finite element method (5.10)–(5.11). Our first step is to define the following mapping.

Definition 5.3.1. *Let $T : W_\epsilon^h \times V_g^h \to W_\epsilon^h \times V_g^h$ be a linear mapping such that for any $(\mu_h, v_h) \in W_\epsilon^h \times V_g^h$, $T(\mu_h, v_h) = (T^{(1)}(\mu_h, v_h), T^{(2)}(\mu_h, v_h))$ satisfies*

$$
\begin{aligned}
&\big(\mu_h - T^{(1)}(\mu_h, v_h), \kappa_h\big) + \big(\mathrm{div}(\kappa_h), D(v_h - T^{(2)}(\mu_h, v_h))\big) \qquad (5.37)\\
&\qquad = (\mu_h, \kappa_h) + \big(\mathrm{div}(\kappa_h), Dv_h\big) - \langle \tilde{g}, \kappa_h\rangle_{\partial\Omega} \quad \forall \kappa_h \in W_0^h, \\
&\big(\mathrm{div}(\mu_h - T^{(1)}(\mu_h, v_h)), Dz_h\big) - \epsilon^{-1}\big(\Phi^\epsilon D(v_h - T^{(2)}(\mu_h, v_h)), Dz_h\big) \qquad (5.38)\\
&\qquad = \big(\mathrm{div}(\mu_h), Dz_h\big) + \epsilon^{-1}\big(\det(\mu_h), z_h\big) - (f^\epsilon, z_h) \quad \forall z_h \in V_0.
\end{aligned}
$$

By Theorem 5.2.3, we conclude that $T(\mu_h, v_h)$ is well-defined. Clearly, any fixed point (χ_h, w_h) of the mapping T (i.e., $T(\chi_h, w_h) = (\chi_h, w_h)$) is a solution to problem (5.10)–(5.11), and vice-versa. Similar to Chapters 3 and 4, we show that the mapping T has a unique fixed point in a small neighborhood of $(I_h\sigma^\epsilon, I_h u^\epsilon)$. However, the analysis of the mixed finite element method proves to be more difficult than the aforementioned chapters.

The most obvious additional difficulty is that the mapping consists of two functions; one a scalar function, the other a symmetric tensor function. Moreover, for any pairing (μ_h, v_h) near the fixed point, the Hessian of the scalar function v_h has to be in some sense close to the tensor function μ_h in order for T to be a contracting mapping. To this end, we define

$$\begin{aligned}
\mathbb{S}_h(\rho) &:= \{(\mu_h, v_h) \in W^h_\epsilon \times V^h_g;\ \|\mu_h - I_h\sigma^\epsilon\|_{L^2} + \epsilon^{-\frac{1}{2}}\|v_h - I_h u^\epsilon\|_{H^1} \le \rho\},\\
\mathbb{Z}_h &:= \{(\mu_h, v_h) \in W^h_\epsilon \times V^h_g;\ (\mu_h, \kappa_h) + (\mathrm{div}(\kappa_h), Dv_h) = \langle \tilde{g}, \kappa_h\rangle_{\partial\Omega}\ \forall \kappa_h \in W^h_0\},\\
\mathbb{B}_h(\rho) &:= \mathbb{S}_h(\rho) \cap \mathbb{Z}_h.
\end{aligned}$$

We also assume $\sigma^\epsilon \in H^s(\Omega)$ and set $\ell := \min\{k+1, s\}$.

The next lemma measures the distance between the center of $\mathbb{S}_h(\rho)$ and its image under the mapping T (compare to Lemmas 3.3.1 and 4.3.1).

Lemma 5.3.2. *Suppose that the linearized problem* (5.12)–(5.14) *is H^3-regular. Then the mapping T satisfies the following estimates:*

$$\|I_h\sigma^\epsilon - T^{(1)}(I_h\sigma^\epsilon, I_h u^\epsilon)\|_{H^1} \le C_1(\epsilon)h^{\ell-3}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.39}$$

$$\|I_h\sigma^\epsilon - T^{(1)}(I_h\sigma^\epsilon, I_h u^\epsilon)\|_{L^2} \le C_2(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.40}$$

$$\|I_h u^\epsilon - T^{(2)}(I_h\sigma^\epsilon, I_h u^\epsilon)\|_{H^1} \le C_3(\epsilon)h^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.41}$$

where $C_1(\epsilon), C_2(\epsilon) = O(\epsilon^{\frac{4-3n}{2}})$, and $C_3(\epsilon) = O(\epsilon^{\frac{-2-3n}{2}})$, $(n = 2, 3)$.

Proof. We divide the proof into four steps.

Step 1: To ease notation we set $\omega^\epsilon_h = I_h\sigma^\epsilon - T^{(1)}(I_h\sigma^\epsilon, I_h u^\epsilon)$, $s^\epsilon_h = I_h u^\epsilon - T^{(2)}(I_h\sigma^\epsilon, I_h u^\epsilon)$. By the definition of T we have for any $(\mu_h, v_h) \in W^h_0 \times V^h_0$

$$\begin{aligned}
&(\omega^\epsilon_h, \mu_h) + \big(\mathrm{div}(\mu_h), Ds^\epsilon_h\big) = (I_h\sigma^\epsilon, \mu_h) + \big(\mathrm{div}(\mu_h), D(I_h u^\epsilon)\big) - \langle \tilde{g}, \mu_h\rangle_{\partial\Omega},\\
&\big(\mathrm{div}(\omega^\epsilon_h), Dv_h\big) - \epsilon^{-1}\big(\Phi^\epsilon Ds^\epsilon_h, Dv_h\big) = \big(\mathrm{div}(I_h\sigma^\epsilon), Dv_h\big) + \epsilon^{-1}\big(\det(I_h\sigma^\epsilon), v_h\big) - (f^\epsilon, v_h).
\end{aligned}$$

It follows from (5.8)–(5.9) that for any $(\mu_h, v_h) \in W^h_0 \times V^h_0$

$$(\omega^\epsilon_h, \mu_h) + (\mathrm{div}(\mu_h), Ds^\epsilon_h) = (I_h\sigma^\epsilon - \sigma^\epsilon, \mu_h) + \big(\mathrm{div}(\mu_h), D(I_h u^\epsilon - u^\epsilon)\big), \tag{5.42}$$

$$\begin{aligned}
\big(\mathrm{div}(\omega^\epsilon_h), Dv_h\big) - \epsilon^{-1}\big(\Phi^\epsilon Ds_h, Dv_h\big) &= \big(\mathrm{div}(I_h\sigma^\epsilon - \sigma^\epsilon), Dv_h\big)\\
&\quad + \epsilon^{-1}\big(\det(I_h\sigma^\epsilon) - \det(\sigma^\epsilon), v_h\big).
\end{aligned} \tag{5.43}$$

Letting $v_h = s^\epsilon_h$, $\mu_h = \omega^\epsilon_h$ in (5.42)-(5.43), subtracting the two equations and using the

mean value theorem, we get

$$\begin{aligned}(\omega_h^\epsilon, \omega_h^\epsilon) + \epsilon^{-1}\big(\Phi^\epsilon Ds_h^\epsilon, Ds_h^\epsilon\big) &= (I_h\sigma^\epsilon - \sigma^\epsilon, \omega_h^\epsilon) + \big(\mathrm{div}(\omega_h^\epsilon), D(I_h u^\epsilon - u^\epsilon)\big) \\ &\quad + \big(\mathrm{div}(\sigma^\epsilon - I_h\sigma^\epsilon), Ds_h^\epsilon\big) + \epsilon^{-1}\big(\det(\sigma^\epsilon) - \det(I_h\sigma^\epsilon), s_h^\epsilon\big) \\ &= (I_h\sigma^\epsilon - \sigma^\epsilon, \omega_h^\epsilon) + \big(\mathrm{div}(\omega_h^\epsilon), D(I_h u^\epsilon - u^\epsilon)\big) \\ &\quad + \big(\mathrm{div}(\sigma^\epsilon - I_h\sigma^\epsilon), Ds_h^\epsilon\big) + \epsilon^{-1}\big(\Psi^\epsilon : (\sigma^\epsilon - I_h\sigma^\epsilon), s_h^\epsilon\big),\end{aligned}$$

where $\Psi^\epsilon = \mathrm{cof}(\tau I_h\sigma^\epsilon + (1-\tau)\sigma^\epsilon)$ for $\tau \in [0,1]$.

Step 2: The case $n = 2$. Since Ψ^ϵ is a 2×2 matrix whose entries are same as those of $\tau I_h\sigma^\epsilon + (1-\tau)\sigma^\epsilon$ (up to sign), we have by (2.11)

$$\begin{aligned}\|\Psi^\epsilon\|_{L^2} &= \|\mathrm{cof}(\tau I_h\sigma^\epsilon + (1-\tau)\sigma^\epsilon)\|_{L^2} = \|\tau I_h\sigma^\epsilon + (1-\tau)\sigma^\epsilon\|_{L^2} \\ &\le \|I_h\sigma^\epsilon\|_{L^2} + \|\sigma^\epsilon\|_{L^2} \le C\|\sigma^\epsilon\|_{L^2} = O(\epsilon^{-\frac{1}{2}}).\end{aligned}$$

Step 3: The case $n = 3$. Note that $(\Psi^\epsilon)_{ij} = (\mathrm{cof}(\tau I_h\sigma^\epsilon + (1-\tau)\sigma^\epsilon))_{ij} = \det(\tau I_h\sigma^\epsilon|_{ij} + (1-\tau)\sigma^\epsilon|_{ij})$, where $\sigma^\epsilon|_{ij}$ denotes the 2×2 matrix after deleting the ith row and jth column of σ^ϵ. We can thus conclude that

$$\begin{aligned}|(\Psi^\epsilon)_{ij}| &\le 2\max_{s\neq i, t\neq j}\big(|\tau(I_h\sigma^\epsilon)_{st} + (1-\tau)(\sigma^\epsilon)_{st}|\big)^2 \\ &\le C\max_{s\neq i, t\neq j}|(\sigma^\epsilon)_{st}|^2 \le C\|\sigma^\epsilon\|_{L^\infty}^2.\end{aligned}$$

Thus, (2.11) implies that

$$\|\Psi^\epsilon\|_{L^2} \le C\|\sigma^\epsilon\|_{L^\infty}^2 = O(\epsilon^{-2}).$$

Step 4: Using the estimates of $\|\Psi^\epsilon\|_{L^2}$, we have

$$\begin{aligned}\|\omega_h^\epsilon\|_{L^2}^2 + \epsilon^{-1}\theta\|Ds_h^\epsilon\|_{L^2}^2 &\le \|I_h\sigma^\epsilon - \sigma^\epsilon\|_{L^2}\|\omega_h^\epsilon\|_{L^2} + \|\omega_h^\epsilon\|_{H^1}\|D(I_h u^\epsilon - u^\epsilon)\|_{L^2} \\ &\quad + \|I_h\sigma^\epsilon - \sigma^\epsilon\|_{H^1}\|Ds_h^\epsilon\|_{L^2} + C\epsilon^{\frac{3}{2}(1-n)}\|\sigma^\epsilon - I_h\sigma^\epsilon\|_{H^1}\|s_h^\epsilon\|_{H^1},\end{aligned}$$

where we have used Sobolev inequality. It follows from Poincaré's inequality, Cauchy's inequality, and the inverse inequality that

$$\begin{aligned}\|\omega_h^\epsilon\|_{L^2}^2 + \epsilon^{-1}\theta\|s_h^\epsilon\|_{H^1}^2 &\le C\epsilon^{(4-3n)}\|I_h\sigma^\epsilon - \sigma^\epsilon\|_{H^1}^2 + C\|\omega_h^\epsilon\|_{H^1}\|I_h u^\epsilon - u^\epsilon\|_{H^1} \\ &\le C\epsilon^{(4-3n)}h^{2\ell-2}\|\sigma^\epsilon\|_{H^\ell}^2 + Ch^{-1}\|\omega_h^\epsilon\|_{L^2}\|I_h u^\epsilon - u^\epsilon\|_{H^1}.\end{aligned} \tag{5.44}$$

Hence,

$$\|\omega_h^\epsilon\|_{L^2}^2 + \epsilon^{-1}\|s_h^\epsilon\|_{H^1}^2 \le C\epsilon^{(4-3n)}h^{2\ell-2}\|\sigma^\epsilon\|_{H^\ell}^2 + Ch^{2\ell-4}\|u^\epsilon\|_{H^\ell}^2.$$

Therefore,

$$\|\omega_h^\epsilon\|_{L^2} \le C_2(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^l}\big),$$

which by the inverse inequality yields

$$\|\omega_h^\epsilon\|_{H^1} \le C_1(\epsilon)h^{\ell-3}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big).$$

Next, from (5.42) we have

$$\begin{aligned}(\mathrm{div}(\mu_h), Ds_h^\epsilon) &\le \|\omega_h^\epsilon\|_{L^2}\|\mu_h\|_{L^2} + \|I_h\sigma^\epsilon - \sigma^\epsilon\|_{L^2}\|\mu_h\|_{L^2} \\ &\quad + \|\mathrm{div}(\mu_h)\|_{L^2}\|D(I_hu^\epsilon - u^\epsilon)\|_{L^2} \\ &\le CC_2(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big)\|\mu_h\|_{H^1}.\end{aligned}$$

It follows from (5.22) that

$$\|Ds_h^\epsilon\|_{L^2} \le CC_2(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big). \tag{5.45}$$

To prove (5.41), let (κ, z) be the solution to the following problem:

$$\begin{aligned}(\kappa,\mu) + (\mathrm{div}(\mu), Dz) &= 0 && \forall \mu \in W_0, \\ (\mathrm{div}(\kappa), Dv) - \epsilon^{-1}(\Phi^\epsilon Dz, Dv) &= \epsilon^{-1}(Ds_h^\epsilon, Dv) && \forall v \in V_0.\end{aligned}$$

By hypothesis, there exists such a z satisfying (cf. Theorem 3.2.2)

$$\|z\|_{H^3} \le C\epsilon^{-2}\|Ds_h^\epsilon\|_{L^2}.$$

We then have

$$\begin{aligned}\epsilon^{-1}\|Ds_h^\epsilon\|_{L^2}^2 &= (\mathrm{div}(\kappa), Ds_h^\epsilon) - \epsilon^{-1}(\Phi^\epsilon Dz, Ds_h^\epsilon) \\ &= (\mathrm{div}(\Pi_h\kappa), Ds_h^\epsilon) - \epsilon^{-1}(\Phi^\epsilon Dz, Ds_h^\epsilon) \\ &= -(\omega_h^\epsilon, \Pi_h\kappa) - \epsilon^{-1}(\Phi^\epsilon Dz, Ds_h^\epsilon) + (I_h\sigma^\epsilon - \sigma^\epsilon, \Pi_h\kappa) \\ &\quad + (\mathrm{div}(\Pi_h\kappa), D(I_hu^\epsilon - u^\epsilon)) \\ &= -(\omega_h^\epsilon, \kappa) + (\omega_h^\epsilon, \kappa - \Pi_h\kappa) - \epsilon^{-1}(\Phi^\epsilon Dz, Ds_h^\epsilon) \\ &\quad + (I_h\sigma^\epsilon - \sigma^\epsilon, \Pi_h\kappa) + (\mathrm{div}(\Pi_h\kappa), D(I_hu^\epsilon - u^\epsilon)) \\ &= (\mathrm{div}(\omega_h^\epsilon), Dz) - \epsilon^{-1}(\Phi^\epsilon Ds_h^\epsilon, Dz) + (\omega_h^\epsilon, \kappa - \Pi_h\kappa) \\ &\quad + (I_h\sigma^\epsilon - \sigma^\epsilon, \Pi_h\kappa) + (\mathrm{div}(\Pi_h\kappa), D(I_hu^\epsilon - u^\epsilon))\end{aligned}$$

$$
\begin{aligned}
&= (\mathrm{div}(\omega_h^\epsilon), D(z - I_h z)) - \epsilon^{-1}(\Phi^\epsilon D s_h^\epsilon, D(z - I_h z)) + (\omega_h, \kappa - \Pi_h \kappa) \\
&\quad + (I_h \sigma^\epsilon - \sigma^\epsilon, \Pi_h \kappa) + (\mathrm{div}(\Pi_h \kappa), D(I_h u^\epsilon - u^\epsilon)) \\
&\quad + (\mathrm{div}(\sigma^\epsilon - I_h \sigma^\epsilon), I_h z) + \epsilon^{-1}(\det(\sigma^\epsilon) - \det(I_h \sigma^\epsilon), I_h z) \\
&\leq \|\mathrm{div}(\omega_h^\epsilon)\|_{L^2} \|D(z - I_h z)\|_{L^2} + \epsilon^{-1} \|\Phi^\epsilon\|_{L^\infty} \|D s_h^\epsilon\|_{L^2} \|D(z - I_h z)\|_{L^2} \\
&\quad + \|\omega_h^\epsilon\|_{L^2} \|\kappa - \Pi_h \kappa\|_{L^2} + \|I_h \sigma^\epsilon - \sigma^\epsilon\|_{L^2} \|\Pi_h \kappa\|_{L^2} \\
&\quad + \|\mathrm{div}(\Pi_h \kappa)\|_{L^2} \|D(I_h u^\epsilon - u^\epsilon)\|_{L^2} \\
&\quad + \|\mathrm{div}(\sigma^\epsilon - I_h \sigma^\epsilon)\|_{L^2} \|I_h z\|_{L^2} + C\epsilon^{-1} \|\Psi^\epsilon\|_{L^2} \|\sigma^\epsilon - I_h \sigma^\epsilon\|_{H^1} \|I_h z\|_{H^1} \\
&\leq Ch^2 \big(\|\omega_h^\epsilon\|_{H^1} + \epsilon^{-2} \|D s_h^\epsilon\|_{L^2} \big) \|z\|_{H^3} \\
&\quad + C\epsilon^{\frac{3}{2}(1-n)} h^{\ell - 1} \big(\|I_h z\|_{L^2} + \|I_h z\|_{H^1} \big) \|\sigma^\epsilon\|_{H^\ell} \\
&\quad + Ch \|\omega_h^\epsilon\|_{L^2} \|\kappa\|_{H^1} + Ch^\ell \|\sigma^\epsilon\|_{H^\ell} \|\Pi_h \kappa\|_{L^2} + Ch^{\ell - 1} \|\Pi_h \kappa\|_{H^1} \|u^\epsilon\|_{H^\ell} \\
&\leq CC_2(\epsilon) \epsilon^{-2} h^{\ell - 1} \big(\|u^\epsilon\|_{H^\ell} + \|\sigma^\epsilon\|_{H^\ell} \big) \|z\|_{H^3} \\
&\leq CC_2(\epsilon) \epsilon^{-4} h^{\ell - 1} \big(\|u^\epsilon\|_{H^\ell} + \|\sigma^\epsilon\|_{H^\ell} \big) \|D s_h^\epsilon\|_{L^2}.
\end{aligned}
$$

Dividing by $\|D s_h^\epsilon\|_{L^2}$ and applying the Poincaré inequality, we get (5.41). The proof is complete. □

The next lemma shows the contractive property of the mapping T (compare to Lemmas 3.3.2 and 4.3.2).

Lemma 5.3.3. *There exists an $h_0 > 0$ such that for $h \leq h_0$, there exists a $\rho_0 \in (0, 1)$ such that T is a contracting mapping in the ball $\mathbb{B}_h(\rho_0)$. That is, for any (μ_h, v_h), $(\chi_h, w_h) \in \mathbb{B}_h(\rho_0)$ there holds*

$$
\begin{aligned}
\|T^{(1)}(\mu_h, v_h) - T^{(1)}(\chi_h, w_h)\|_{L^2} + \epsilon^{-\frac{1}{2}} \|T^{(2)}(\mu_h, v_h) - T^{(2)}(\chi_h, w_h)\|_{H^1} \\
\leq \frac{1}{2} \big(\|\mu_h - \chi_h\|_{L^2} + \epsilon^{-\frac{1}{2}} \|v_h - w_h\|_{H^1} \big).
\end{aligned} \tag{5.46}
$$

Proof. We divide the proof into five steps.

Step 1: To ease notation, let

$$
T^{(1)} := T^{(1)}(\mu_h, v_h) - T^{(1)}(\chi_h, w_h), \quad T^{(2)} := T^{(2)}(\mu_h, v_h) - T^{(2)}(\chi_h, w_h).
$$

By the definition of $T^{(i)}$ we get

$$
\big(T^{(1)}, \kappa_h\big) + \big(\mathrm{div}(\kappa_h), D(T^{(2)})\big) = 0 \quad \forall \kappa_h \in W_0^h, \tag{5.47}
$$

$$
\begin{aligned}
&\big(\mathrm{div}(T^{(1)}), D z_h\big) - \epsilon^{-1} \big(\Phi^\epsilon D(T^{(2)}), D z_h\big) \\
&\qquad = \epsilon^{-1} \big(\big(\Phi^\epsilon D(w_h - v_h), D z_h\big) + \big(\det(\chi_h) - \det(\mu_h), z_h\big) \big) \quad \forall z_h \in V_0^h.
\end{aligned} \tag{5.48}
$$

Letting $z_h = T^{(2)}$ and $\kappa_h = T^{(1)}$, subtracting (5.48) from (5.47), and using the mean value theorem we have $(n = 2, 3)$

$$
\begin{aligned}
&(T^{(1)}, T^{(1)}) + \epsilon^{-1}(\Phi^\epsilon DT^{(2)}, DT^{(2)}) \\
&\quad = \epsilon^{-1}\big((\Phi^\epsilon D(v_h - w_h), DT^{(2)}) + (\det(\mu_h) - \det(\chi_h), T^{(2)})\big) \\
&\quad = \epsilon^{-1}\big((\Phi^\epsilon D(v_h - w_h), DT^{(2)}) + (\Psi_h : (\mu_h - \chi_h), T^{(2)})\big) \\
&\quad = \epsilon^{-1}\big((\Phi^\epsilon D(v_h - w_h), DT^{(2)}) + (\Phi^\epsilon : (\mu_h - \chi_h), T^{(2)}) \\
&\qquad + ((\Psi_h - \Phi^\epsilon) : (\mu_h - \chi_h), T^{(2)})\big) \\
&\quad = \epsilon^{-1}\big((\mathrm{div}(\Phi^\epsilon T^{(2)}), D(v_h - w_h)) + (\mu_h - \chi_h, \Phi^\epsilon T^{(2)}) \\
&\qquad + ((\Psi_h - \Phi^\epsilon) : (\mu_h - \chi_h), T^{(2)})\big) \\
&\quad = \epsilon^{-1}\big((\mathrm{div}(\Pi_h(\Phi^\epsilon T^{(2)})), D(v_h - w_h)) + (\mu_h - \chi_h, \Phi^\epsilon T^{(2)}) \\
&\qquad + ((\Psi_h - \Phi^\epsilon) : (\mu_h - \chi_h), T^{(2)})\big) \\
&\quad = \epsilon^{-1}\big((\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)}), \mu_h - \chi_h) + ((\Psi_h - \Phi^\epsilon) : (\mu_h - \chi_h), T^{(2)})\big) \\
&\quad \le \epsilon^{-1}\big(\|\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)})\|_{L^2}\|\mu_h - \chi_h\|_{L^2} \\
&\qquad\quad + C\|\Psi_h - \Phi^\epsilon\|_{L^2}\|\mu_h - \chi_h\|_{L^2}\|T^{(2)}\|_{L^\infty}\big) \\
&\quad \le \epsilon^{-1}\big(\|\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)})\|_{L^2}\|\mu_h - \chi_h\|_{L^2} \\
&\qquad\quad + |\log h|^{\frac{3-n}{2}} h^{\frac{2-n}{2}}\|\Psi_h - \Phi^\epsilon\|_{L^2}\|\mu_h - \chi_h\|_{L^2}\|T^{(2)}\|_{H^1}\big),
\end{aligned}
$$

where $\Psi_h = \mathrm{cof}(\mu_h + \tau(\chi_h - \mu_h))$, $\tau \in [0, 1]$. We have used the inverse inequality to get the last inequality above.

Step 2: The case of $n = 2$. We bound $\|\Phi^\epsilon - \Psi_h\|_{L^2}$ as follows:

$$
\begin{aligned}
\|\Phi^\epsilon - \Psi_h\|_{L^2} &= \|\mathrm{cof}(\sigma^\epsilon) - \mathrm{cof}(\mu_h + \tau(\chi_h - \mu_h))\|_{L^2} \\
&= \|\sigma^\epsilon - \mu_h - \tau(\chi_h - \mu_h)\|_{L^2} \\
&\le \|\sigma^\epsilon - I_h\sigma^\epsilon\|_{L^2} + \|I_h\sigma^\epsilon - \mu_h\|_{L^2} + \|\chi_h - \mu_h\|_{L^2} \\
&\le C\big(h^\ell\|\sigma^\epsilon\|_{H^\ell} + \rho_0\big).
\end{aligned}
$$

Step 3: The case of $n = 3$. To bound $\|\Phi^\epsilon - \Psi_h\|_{L^2}$ in this case, we first write

$$
\begin{aligned}
\|(\Phi^\epsilon - \Psi_h)_{ij}\|_{L^2} &= \|(\mathrm{cof}(\sigma^\epsilon)_{ij}) - \mathrm{cof}(\mu_h + \tau(\chi_h - \mu_h))_{ij}\|_{L^2} \\
&= \|\det(\sigma^\epsilon|_{ij}) - \det(\mu_h|_{ij} + \tau(\chi_h|_{ij} - \mu_h|_{ij}))\|_{L^2},
\end{aligned}
$$

where we have used the same notation found in Lemma 5.3.2. Using the mean value

theorem,

$$\begin{aligned}\|(\Phi^\epsilon - \Psi_h)_{ij}\|_{L^2} &= \|\det(\sigma^\epsilon|_{ij}) - \det(\mu_h|_{ij} + \tau(\chi_h|_{ij} - \mu_h|_{ij}))\|_{L^2}\\ &= \|\Lambda^{ij} : (\sigma^\epsilon|_{ij} - \mu_h|_{ij} - \tau(\chi_h|_{ij} - \mu_h|_{ij}))\|_{L^2}\\ &\leq \|\Lambda^{ij}\|_{L^\infty}\|\sigma^\epsilon|_{ij} - \mu_h|_{ij} - \tau(\chi_h|_{ij} - \mu_h|_{ij})\|_{L^2},\end{aligned}$$

where $\Lambda^{ij} = \mathrm{cof}(\sigma^\epsilon|_{ij} + \lambda(\mu|_{ij} - \tau(\chi_h|_{ij} - \mu|_{ij}) - \sigma^\epsilon|_{ij}))$, $\lambda \in [0,1]$.

On noting that $\Lambda^{ij} \in \mathbf{R}^{2\times 2}$, we have

$$\begin{aligned}\|\Lambda^{ij}\|_{L^\infty} &= \|\mathrm{cof}(\sigma^\epsilon|_{ij} + \lambda(\mu|_{ij} - \tau(\chi_h|_{ij} - \mu|_{ij}) - \sigma^\epsilon|_{ij}))\|_{L^\infty}\\ &= \|\sigma^\epsilon|_{ij} + \lambda(\mu|_{ij} - \tau(\chi_h|_{ij} - \mu|_{ij}) - \sigma^\epsilon|_{ij})\|_{L^\infty}\\ &\leq C\big(\|\sigma^\epsilon\|_{L^\infty} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)\\ &\leq \big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big),\end{aligned}$$

where we have used the inverse inequality and (2.11). Combining the above estimates gives

$$\begin{aligned}\|(\Phi^\epsilon - \Psi_h)_{ij}\|_{L^2} &\leq C\big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)\|\sigma^\epsilon|_{ij} - \mu_h|_{ij} - \tau(\chi_h|_{ij} - \mu_h|_{ij})\|_{L^2}\\ &\leq C\big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)\left(h^\ell\|\sigma^\epsilon\|_{H^\ell} + \rho_0\right).\end{aligned}$$

Step 4: We now bound $\|\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)})\|_{L^2}$ as follows:

$$\begin{aligned}\|\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)})\|_{L^2}^2 &\leq Ch^2\|\Phi^\epsilon T^{(2)}\|_{H^1}^2\\ &= Ch^2\big(\|\Phi^\epsilon T^{(2)}\|_{L^2}^2 + \|D(\Phi^\epsilon T^{(2)})\|_{L^2}^2\big)\\ &\leq Ch^2\big(\|\Phi^\epsilon T^{(2)}\|_{L^2}^2 + \|\Phi^\epsilon DT^{(2)}\|_{L^2}^2 + \|D\Phi^\epsilon T^{(2)}\|_{L^2}^2\big)\\ &\leq Ch^2\big(\|\Phi^\epsilon\|_{L^4}^2\|T^{(2)}\|_{L^4}^2 + \|\Phi^\epsilon\|_{L^\infty}\|DT^{(2)}\|_{L^2}^2 + \|D\Phi^\epsilon\|_{L^3}^2\|T^{(2)}\|_{L^6}^2\big)\\ &\leq Ch^2\big(\|\Phi^\epsilon\|_{L^4}^2\|T^{(2)}\|_{H^1}^2 + \|\Phi^\epsilon\|_{L^\infty}^2\|DT^{(2)}\|_{L^2}^2 + \|D\Phi^\epsilon\|_{L^3}^2\|T^{(2)}\|_{H^1}^2\big)\\ &\leq Ch^2\big(\|\Phi^\epsilon\|_{L^\infty}^2 + \|D\Phi^\epsilon\|_{L^3}^2\big)\|DT^{(2)}\|_{L^2}^2\\ &\leq C\epsilon^{-\frac{13}{6}}h^2\|DT^{(2)}\|_{L^2}^2,\end{aligned}$$

where we have used Sobolev's inequality followed by Poincaré's inequality. Thus,

$$\|\Phi^\epsilon T^{(2)} - \Pi_h(\Phi^\epsilon T^{(2)})\|_{L^2} \leq C\epsilon^{-\frac{13}{12}}\|DT^{(2)}\|_{L^2}.$$

Step 5: Substituting all estimates from Steps 2–4 into Step 1, and using the fact that

Φ^ϵ is positive definite, we obtain for $n = 2, 3$

$$\begin{aligned}
&\|T^{(1)}\|_{L^2}^2 + \epsilon^{-1}\theta\|DT^{(2)}\|_{L^2}^2 \\
&\quad \le C\epsilon^{-1}\Big(\epsilon^{-\frac{13}{12}}h + |\log h|^{\frac{3-n}{2}}h^{\frac{2-n}{2}}\big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)^{n-2}(h^\ell\|\sigma^\epsilon\|_{H^\ell} + \rho_0)\Big) \\
&\qquad \times \Big(\|\mu_h - \chi_h\|_{L^2}\|T^{(2)}\|_{H^1}\Big) \\
&\quad \le C\Big(\epsilon^{-\frac{25}{12}} + \epsilon^{-1}|\log h|^{\frac{3-n}{2}}h^{\frac{2-n}{2}}\big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)^{n-2}(h^\ell\|\sigma^\epsilon\|_{H^\ell} + \rho_0)\Big) \\
&\qquad \times \Big(\|\mu_h - \chi_h\|_{L^2}\|DT^{(2)}\|_{L^2}\Big)
\end{aligned}$$

Using Cauchy's inequality we get

$$\begin{aligned}
&\|T^{(1)}\|_{L^2} + \epsilon^{-\frac{1}{2}}\|T^{(2)}\|_{H^1} \\
&\quad \le C\Big\{\epsilon^{-\frac{19}{12}}h + \epsilon^{-\frac{1}{2}}|\log h|^{3-n}h^{\frac{2-n}{2}}\big(\epsilon^{-1} + h^\ell\|\sigma^\epsilon\|_{H^\ell} + h^{-\frac{3}{2}}\rho_0\big)^{n-2} \\
&\qquad \times \big(h^\ell\|\sigma^\epsilon\|_{H^\ell} + \rho_0\big)\Big\}\|\mu_h - \chi_h\|_{L^2}.
\end{aligned}$$

For the $n = 2$ case, we choose $h_0 = O(\epsilon^{\frac{19}{12}})$ such that $|\log h_0|^{\frac{1}{2\ell}}h_0 \le \Big(\frac{\sqrt{\epsilon}}{\|\sigma^\epsilon\|_{H^\ell}}\Big)^{\frac{1}{\ell}}$. For the $n = 3$ case, we choose $h_0 = O\left(\min\Big\{\epsilon^{\frac{19}{12}}, \Big(\frac{\epsilon^{\frac{3}{2}}}{\|\sigma^\epsilon\|_{H^\ell}}\Big)^{\frac{2}{2\ell-1}}\Big\}\right)$. Fixing $h \le h_0$, we set $\rho_0 = O(\epsilon^{\frac{1}{2}}|\log h|^{-\frac{1}{2}})$ in the $n = 2$ case and $\rho_0 = O(\epsilon^{\frac{3}{2}}h^{\frac{1}{2}})$ in the $n = 3$ case. We then have the following estimate:

$$\begin{aligned}
\|T^{(1)}\|_{L^2} + \epsilon^{-\frac{1}{2}}\|T^{(2)}\|_{H^1} &\le \frac{1}{2}\|\mu_h - \chi_h\|_{L^2} \\
&\le \frac{1}{2}\big(\|\mu_h - \chi_h\|_{L^2} + \epsilon^{-\frac{1}{2}}\|v_h - w_h\|_{H^1}\big).
\end{aligned}$$

The proof is complete. □

We now state the first main theorem of this chapter.

Theorem 5.3.4. *Suppose that the linearized problem is H^3-regular. Then there exists an $h_1 > 0$ such that for $h \le \min\{h_0, h_1\}$, there exists a unique solution $(\sigma_h^\epsilon, u_h^\epsilon)$ to (5.10)-(5.11) in the ball $\mathbb{B}_h(\rho_1)$, where $\rho_1 = 2\big(C_2(\epsilon)h^{\ell-2} + \epsilon^{-\frac{1}{2}}C_3(\epsilon)h^{\ell-1}\big)\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big)$. Moreover,*

$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} + \epsilon^{-\frac{1}{2}}\|u^\epsilon - u_h^\epsilon\|_{H^1} \le C_4(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.49}$$

$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1} \le C_5(\epsilon)h^{\ell-3}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.50}$$

where $C_4(\epsilon), C_5(\epsilon) = O(\epsilon^{-\frac{1}{2}}C_3(\epsilon)) = O(\epsilon^{\frac{(-3-3n)}{2}})$.

Proof. Let $(\mu_h, v_h) \in \mathbb{B}_h(\rho_1)$. In the two dimensional case, choose h_1 such that

$$h_1|\log h|^{\frac{1}{2(\ell-2)}} \leq C\left(\frac{\sqrt{\epsilon}}{C_2(\epsilon)(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell})}\right)^{\frac{1}{\ell-2}}$$

and

$$h_1|\log h|^{\frac{1}{2(\ell-1)}} \leq C\left(\frac{\epsilon}{C_3(\epsilon)(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell})}\right)^{\frac{1}{\ell-1}}.$$

For the three dimensional case, we choose h_1 such that

$$h_1 \leq C\left(\min\Big\{\left(\frac{\epsilon^{\frac{3}{2}}}{C_2(\epsilon)(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell})}\right)^{\frac{2}{2\ell-5}}, \left(\frac{\epsilon^2}{C_3(\epsilon)(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell})}\right)^{\frac{2}{2\ell-3}}\Big\}\right).$$

Then for $h \leq h_1$, we have $\rho_1 \leq \rho_0$. Thus, using the triangle inequality and Lemmas 5.3.2 and 5.3.3 we get

$$\begin{aligned}
\|I_h\sigma^\epsilon - T^{(1)}(\mu_h, v_h)\|_{L^2} + \epsilon^{-\frac{1}{2}}\|I_h u^\epsilon - T^{(2)}(\mu_h, v_h)\|_{H^1} &\leq \|I_h\sigma^\epsilon - T^{(1)}(I_h\sigma^\epsilon, I_h u^\epsilon)\|_{L^2} \\
&\quad + \|T^{(1)}(I_h\sigma^\epsilon, I_h u^\epsilon) - T^{(1)}(\mu_h, v_h)\|_{L^2} + \epsilon^{-\frac{1}{2}}\|I_h u^\epsilon - T^{(2)}(I_h\sigma^\epsilon, I_h u^\epsilon)\|_{H^1} \\
&\quad + \epsilon^{-\frac{1}{2}}\|T^{(2)}(I_h\sigma^\epsilon, I_h u^\epsilon) - T^{(2)}(\mu_h, v_h)\|_{H^1} \\
&\leq \big(C_2(\epsilon)h^{\ell-2} + \epsilon^{-\frac{1}{2}}C_3(\epsilon)h^{\ell-1}\big)\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big) \\
&\quad + \frac{1}{2}\big(\|I_h\sigma^\epsilon - \mu_h\|_{L^2} + \epsilon^{-\frac{1}{2}}\|I_h u^\epsilon - v_h\|_{H^1}\big) \\
&\leq \frac{\rho_1}{2} + \frac{\rho_1}{2} = \rho_1 < 1.
\end{aligned}$$

So $T(\mu_h, v_h) \in \mathbb{B}_h(\rho_1)$. Clearly, T is a continuous mapping. Thus, T has a unique fixed point $(\sigma_h^\epsilon, u_h^\epsilon) \in \mathbb{B}_h(\rho_1)$ which is the unique solution to (5.10)-(5.11).

Next, we use the triangle inequality to get

$$\begin{aligned}
\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} + \epsilon^{-\frac{1}{2}}\|u^\epsilon - u_h^\epsilon\|_{H^1} &\leq \|\sigma^\epsilon - I_h\sigma^\epsilon\|_{L^2} + \|I_h\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} \\
&\quad + \epsilon^{-\frac{1}{2}}\big(\|u^\epsilon - I_h u^\epsilon\|_{H^1} + \|I_h u^\epsilon - u_h^\epsilon\|_{H^1}\big) \\
&\leq \rho_1 + Ch^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big) \\
&\leq C_4(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big).
\end{aligned}$$

Finally, using the inverse inequality we have

$$\begin{aligned}\|\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1} &\le \|\sigma^\epsilon - I_h\sigma^\epsilon\|_{H^1} + \|I_h\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1}\\ &\le \|\sigma^\epsilon - I_h\sigma^\epsilon\|_{H^1} + Ch^{-1}\|I_h\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2}\\ &\le Ch^{\ell-1}\|\sigma^\epsilon\|_{H^\ell} + Ch^{-1}\rho_1\\ &\le C_5(\epsilon)h^{\ell-3}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big).\end{aligned}$$

□

Remark 5.3.5. *Comparing with error estimates for the linearized problem in Theorem 5.2.4, we see that the above H^1-error for the scalar variable is not optimal. Next, we shall employ a similar duality argument as used in the proof of Theorem 5.2.4 to show that the estimate can be improved to optimal order.*

Theorem 5.3.6. *Under the same hypothesis of Theorem 5.3.4 there holds*

$$\|u^\epsilon - u_h^\epsilon\|_{H^1} \le C\epsilon^{-2}\Big(\epsilon^{-\frac{1}{2}}(C_4(\epsilon)+C_5(\epsilon))h^{\ell-1} + C_4(\epsilon)\epsilon^{1-n}h^{2\ell-4}\Big)\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big). \quad (5.51)$$

Proof. The regularity assumption implies that there exists $(\kappa, z) \in W_0 \times V_0 \cap H^3(\Omega)$ such that

$$(\kappa,\mu) + (\mathrm{div}(\mu), Dz) = 0 \qquad \forall \mu \in W_0, \quad (5.52)$$

$$(\mathrm{div}(\kappa), Dv) - \epsilon^{-1}(\Phi^\epsilon Dz, Dv) = \epsilon^{-1}(D(u^\epsilon - u_h^\epsilon), Dv) \qquad \forall v \in V_0, \quad (5.53)$$

with

$$\|z\|_{H^3} \le C\epsilon^{-2}\|D(u^\epsilon - u_h^\epsilon)\|_{L^2}. \quad (5.54)$$

It is easy to check that $\sigma^\epsilon - \sigma_h^\epsilon$ and $u^\epsilon - u_h^\epsilon$ satisfy the following error equations:

$$(\sigma^\epsilon - \sigma_h^\epsilon, \mu_h) + (\mathrm{div}(\mu_h), D(u^\epsilon - u_h^\epsilon)) = 0 \qquad \forall \mu_h \in W_0^h, \quad (5.55)$$

$$(\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon), Dv_h) + \epsilon^{-1}(\det(\sigma^\epsilon) - \det(\sigma_h^\epsilon), v_h) = 0 \qquad \forall v_h \in V_0^h. \quad (5.56)$$

By (5.52)-(5.56) and the mean value theorem we get

$$\begin{aligned}\epsilon^{-1}\|D(u^\epsilon - u_h^\epsilon)\|_{L^2}^2 &= \big(\mathrm{div}(\kappa), D(u^\epsilon - u_h^\epsilon)\big) - \epsilon^{-1}\big(\Phi^\epsilon Dz, D(u^\epsilon - u_h^\epsilon)\big)\\ &= \big(\mathrm{div}(\Pi_h\kappa), D(u^\epsilon - u_h^\epsilon)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), Dz\big) + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - u_h^\epsilon)\big)\\ &= \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), Dz\big) + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - u_h^\epsilon)\big)\end{aligned}$$

$$
\begin{aligned}
&= \big(\sigma_h^\epsilon - \sigma^\epsilon, \kappa\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), Dz\big) \\
&\qquad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - I_h u^\epsilon)\big) + \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa - \kappa\big) \\
&= \big(\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon), Dz\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), Dz\big) \\
&\qquad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - I_h u^\epsilon)\big) + \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa - \kappa\big) \\
&= \big(\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon), D(z - I_h z)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(z - I_h z)\big) \\
&\qquad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - I_h u^\epsilon)\big) + \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa - \kappa\big) \\
&\qquad - \epsilon^{-1}\big(\det(\sigma^\epsilon) - \det(\sigma_h^\epsilon), I_h z\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(I_h z)\big) \\
&= \big(\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon), D(z - I_h z)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(z - I_h z)\big) \\
&\qquad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - I_h u^\epsilon)\big) + \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa - \kappa\big) \\
&\qquad - \epsilon^{-1}\big(\Psi^\epsilon : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(I_h z)\big),
\end{aligned}
$$

where $\Psi^\epsilon = \mathrm{cof}(\sigma^\epsilon + \tau(\sigma_h^\epsilon - \sigma^\epsilon))$ for $\tau \in [0,1]$. We note we have abused the notation of Ψ^ϵ defining it differently in two different proofs.

Next, we note that

$$
\begin{aligned}
&\big(\Psi^\epsilon : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) + \big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(I_h z)\big) \\
&\quad = \big(\Phi^\epsilon : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) + \big(\mathrm{div}(\Phi^\epsilon I_h z), D(u^\epsilon - u_h^\epsilon)\big) + \big((\Psi^\epsilon - \Phi^\epsilon) : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) \\
&\quad = \big(\sigma^\epsilon - \sigma_h^\epsilon), \Phi^\epsilon I_h z\big) + \big(\mathrm{div}(\Pi_h(\Phi^\epsilon I_h z)), D(u^\epsilon - u_h^\epsilon)\big) + \big((\Psi^\epsilon - \Phi^\epsilon) : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) \\
&\qquad\quad + \big(\mathrm{div}(\Phi^\epsilon I_h z - \Pi_h(\Phi^\epsilon I_h z)), D(u^\epsilon - I_h u^\epsilon)\big) \\
&\quad = \big(\sigma^\epsilon - \sigma_h^\epsilon, \Phi^\epsilon I_h z - \Pi_h(\Phi^\epsilon I_h z)\big) + \big((\Psi^\epsilon - \Phi^\epsilon) : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) \\
&\qquad\quad + \big(\mathrm{div}(\Phi^\epsilon I_h z - \Pi_h(\Phi^\epsilon I_h z)), D(u^\epsilon - I_h u^\epsilon)\big).
\end{aligned}
$$

Using this identity and using the same technique used in Step 4 of Lemma 5.3.3, we have

$$
\begin{aligned}
\epsilon^{-1}\|D(u^\epsilon - u_h^\epsilon)\|_{L^2}^2 &= \big(\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon), D(z - I_h z)\big) - \epsilon^{-1}\big(\Phi^\epsilon D(u^\epsilon - u_h^\epsilon), D(z - I_h z)\big) \\
&\quad + \epsilon^{-1}\Big\{\big((\Phi^\epsilon - \Psi^\epsilon) : (\sigma^\epsilon - \sigma_h^\epsilon), I_h z\big) + \big(\sigma^\epsilon - \sigma_h^\epsilon, \Pi_h(\Phi^\epsilon I_h z) - \Phi^\epsilon I_h z\big) \\
&\quad + \big(\mathrm{div}(\Pi_h(\Phi^\epsilon I_h z) - \Phi^\epsilon I_h z), D(u^\epsilon - I_h u^\epsilon)\big)\Big\} + \big(\sigma_h^\epsilon - \sigma^\epsilon, \Pi_h\kappa - \kappa\big) \\
&\quad + \big(\mathrm{div}(\kappa - \Pi_h\kappa), D(u^\epsilon - I_h u^\epsilon)\big) \\
&\le \big(\|\mathrm{div}(\sigma^\epsilon - \sigma_h^\epsilon)\|_{L^2} + C\epsilon^{-2}\|D(u^\epsilon - u_h^\epsilon)\|_{L^2}\big)\|D(z - I_h z)\|_{L^2} \\
&\quad + C\epsilon^{-1}\big(\|\Phi^\epsilon - \Psi^\epsilon\|_{L^2}\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2}\|I_h z\|_{L^\infty} + \|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2}\|\Pi_h(\Phi^\epsilon I_h z) - \Phi^\epsilon I_h z\|_{L^2} \\
&\quad + \|\mathrm{div}(\Pi_h(\Phi^\epsilon I_h z) - \Phi^\epsilon I_h z)\|_{L^2}\|D(u^\epsilon - I_h u^\epsilon)\|_{L^2}\big) + \|\kappa - \Pi_h\kappa\|_{L^2}\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} \\
&\quad + \|\mathrm{div}(\kappa - \Pi_h\kappa)\|_{L^2}\|D(u^\epsilon - I_h u^\epsilon)\|_{L^2}
\end{aligned}
$$

$$
\begin{aligned}
&\le Ch^2\big(\|\sigma^\epsilon-\sigma_h^\epsilon\|_{H^1}+\epsilon^{-2}\|u^\epsilon-u_h^\epsilon\|_{H^1}\big)\|z\|_{H^3}\\
&\quad+C\epsilon^{-2}\big(\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}\|\sigma^\epsilon-\sigma_h^\epsilon\|_{L^2}+h\|\sigma^\epsilon-\sigma_h^\epsilon\|_{L^2}+\|u^\epsilon-I_hu^\epsilon\|_{H^1}\big)\|z\|_{H^3}\\
&\quad+Ch\|\sigma^\epsilon-\sigma_h^\epsilon\|_{L^2}\|\kappa\|_{H^1}+C\|u^\epsilon-I_hu^\epsilon\|_{H^1}\|\kappa\|_{H^1}\\
&\le\Big\{\epsilon^{-\frac{3}{2}}\big(C_4(\epsilon)+C_5(\epsilon)\big)h^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell}+\|u^\epsilon\|_{H^\ell}\big)+\epsilon^{-2}C_4(\epsilon)h^{\ell-2}\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}\Big\}\|z\|_{H^3}\\
&\le C\epsilon^{-2}\Big\{\epsilon^{-\frac{3}{2}}\big(C_4(\epsilon)+C_5(\epsilon)\big)h^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell}+\|u^\epsilon\|_{H^\ell}\big)\\
&\quad+\epsilon^{-2}C_4(\epsilon)h^{\ell-2}\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}\Big\}\|D(u^\epsilon-u_h^\epsilon)\|_{L^2}.
\end{aligned}
$$

We now bound $\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}$ separately for the cases $n=2$ and $n=3$. First, when $n=2$ we have

$$
\begin{aligned}
\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}&=\|\mathrm{cof}(\sigma^\epsilon)-\mathrm{cof}(\sigma_h^\epsilon+\tau(\sigma^\epsilon-\sigma_h^\epsilon))\|_{L^2}\\
&=\|\sigma^\epsilon-(\sigma_h^\epsilon+\tau(\sigma^\epsilon-\sigma_h^\epsilon))\|_{L^2}\\
&\le C_4(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell}+\|u^\epsilon\|_{H^\ell}\big).
\end{aligned}
$$

Second, when $n=3$, on noting that

$$
\begin{aligned}
|(\Phi^\epsilon-\Psi^\epsilon)_{ij}|&=|(\mathrm{cof}(\sigma^\epsilon))_{ij}-(\mathrm{cof}(\sigma_h^\epsilon+\tau(\sigma^\epsilon-\sigma_h^\epsilon)))_{ij}|\\
&=|\det(\sigma^\epsilon|_{ij})-\det(\sigma^\epsilon|_{ij}+\tau(\sigma^\epsilon|_{ij}-\sigma_h^\epsilon|_{ij}))|,
\end{aligned}
$$

and using the mean value theorem we get

$$
\begin{aligned}
\|(\Psi^\epsilon)_{ij}-(\Phi^\epsilon)_{ij}\|_{L^2}&=(1-\tau)\|\Lambda^{ij}:(\sigma^\epsilon|_{ij}-\sigma_h^\epsilon|_{ij})\|_{L^2}\\
&\le\|\Lambda^{ij}\|_{L^\infty}\|\sigma^\epsilon|_{ij}-\sigma_h^\epsilon|_{ij}\|_{L^2},
\end{aligned}
$$

where $\Lambda^{ij}=\mathrm{cof}(\sigma^\epsilon|_{ij}+\lambda(\sigma_h^\epsilon|_{ij}-\sigma^\epsilon|_{ij}))$ for $\lambda\in[0,1]$. Since $\Lambda^{ij}\in\mathbf{R}^{2\times2}$, then

$$
\|\Lambda^{ij}\|_{L^\infty}=\|\sigma^\epsilon|_{ij}+\lambda(\sigma_h^\epsilon|_{ij}-\sigma^\epsilon|_{ij})\|_{L^\infty}\le C\|\sigma^\epsilon\|_{L^\infty}=O(\epsilon^{-1}).
$$

Thus,

$$
\|\Phi^\epsilon-\Psi^\epsilon\|_{L^2}\le C_4(\epsilon)\epsilon^{2-n}h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell}+\|u^\epsilon\|_{H^\ell}\big).
$$

Finally, combining the above estimates, we obtain

$$
\|D(u^\epsilon-u_h^\epsilon)\|_{L^2}\le C^{-2}\Big(\epsilon^{-\frac{1}{2}}(C_4(\epsilon)+C_5(\epsilon))h^{\ell-1}+\epsilon^{1-n}C_4(\epsilon)h^{2\ell-4}\Big)\big(\|\sigma^\epsilon\|_{H^\ell}+\|u^\epsilon\|_{H^\ell}\big).
$$

□

Remark 5.3.7. *We note that* $2(\ell-2) \geq \ell-1$ *for* $k \geq 2$. *Thus, we have obtained optimal error estimates in the* H^1*-norm.*

5.4 Numerical Experiments and Rates of Convergence

In this section, we provide several 2-D and 3-D numerical experiments to gauge the efficiency of the mixed finite element method developed in the previous sections, and we also compare the results with those found in Chapter 3. We numerically determine the "best" choice of the mesh size h in terms of ϵ, and rates of convergence for both $u-u^\epsilon$ and $u^\epsilon-u_h^\epsilon$. All tests given below are done on domain $\Omega=(0,1)^n$ $(n=2,3)$. We like to remark that the 2-D mixed finite element methods we tested are often 10–20 times faster than the Argyris finite element method studied in Chapter 3.

Test 5.1

For this test, we calculate $\|u-u_h^\epsilon\|$ for fixed $h=0.009$, while varying ϵ in order to estimate $\|u-u^\epsilon\|$. We the use quadratic Lagrange element, and set to solve problem (5.8)–(5.9) with the following test functions:

(a) $u=e^{\frac{x_1^2+x_2^2}{2}}$,	$f=(1+x_1^2+x_2^2)e^{\frac{x_1^2+x_2^2}{2}}$,	$g=e^{\frac{x_1^2+x_2^2}{2}}$,
(b) $u=x_1^4+x_2^2$,	$f=24x_1^2$,	$g=x_1^4+x_2^2$,
(c) $u=e^{\frac{x_1^2+x_2^2+x_3^2}{2}}$,	$f=(1+x_1^2+x_2^2+x_3^2)e^{\frac{3}{2}(x^2+y^2+z^2)}$,	$g=e^{\frac{x_1^2+x_2^2+x_3^2}{2}}$,
(d) $u=x_1^2+x_2^2+x_3^2$,	$f=8$,	$g=x_1^2+x_2^2+x_3^2$,
(e) $u=\frac{1}{2}(x_1^4+x_2^2+x_3^4)$,	$f=36x_1^2x_3^2$,	$g=\frac{1}{2}(x_1^4+x_2^2+x_3^4)$.

After having computed the error, we plot the results in Figure 5.1 for the two dimensional tests and Figure 5.2 for the three dimensional tests. We also plot the computed solution and corresponding errors in Figures 5.3–5.4. The figures show that $\|\sigma-\sigma_h^\epsilon\|_{L^2}=O(\epsilon^{\frac{1}{4}})$ in both the two dimensional and three dimensional case. Since h is very small, we expect $\|u-u^\epsilon\|_{H^2} \approx \|\sigma-\sigma_h^\epsilon\|_{L^2}=O(\epsilon^{\frac{1}{4}})$. Based on this heuristic argument, we predict that $\|u-u^\epsilon\|_{H^2}=O(\epsilon^{\frac{1}{4}})$. Similarly, from Figures 5.1 and 5.2 we see that $\|u-u^\epsilon\|_{L^2} \approx O(\epsilon)$ and $\|u-u^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$. We note that these are the same rates of convergence found in Test 3.1 in Section 3.6

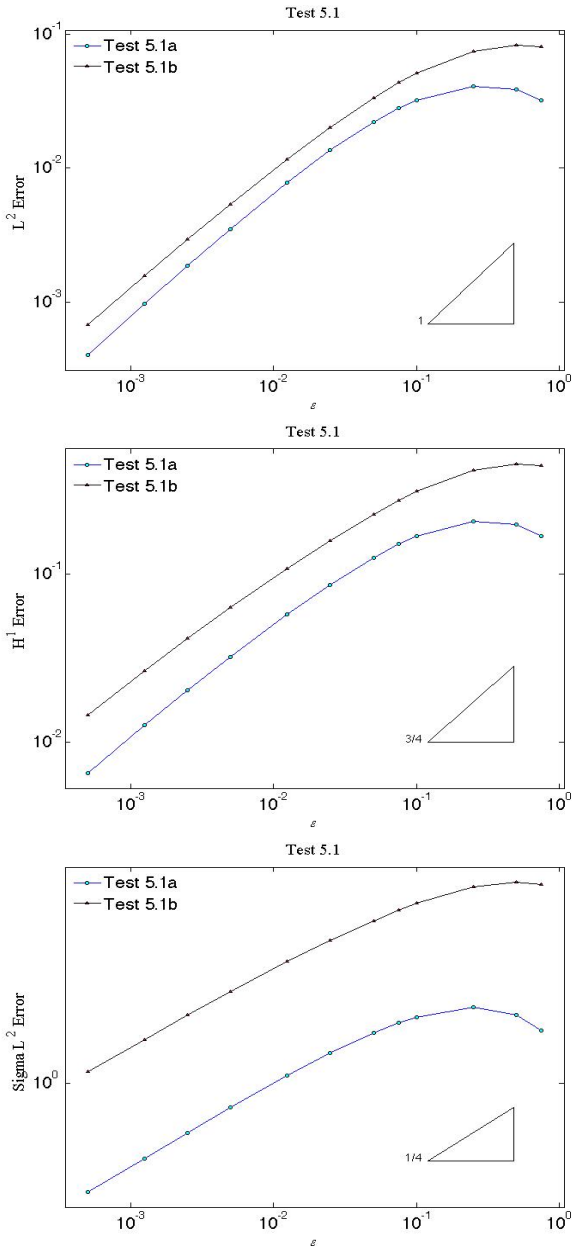


Figure 5.1: Test 5.1 (2-D). Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ.

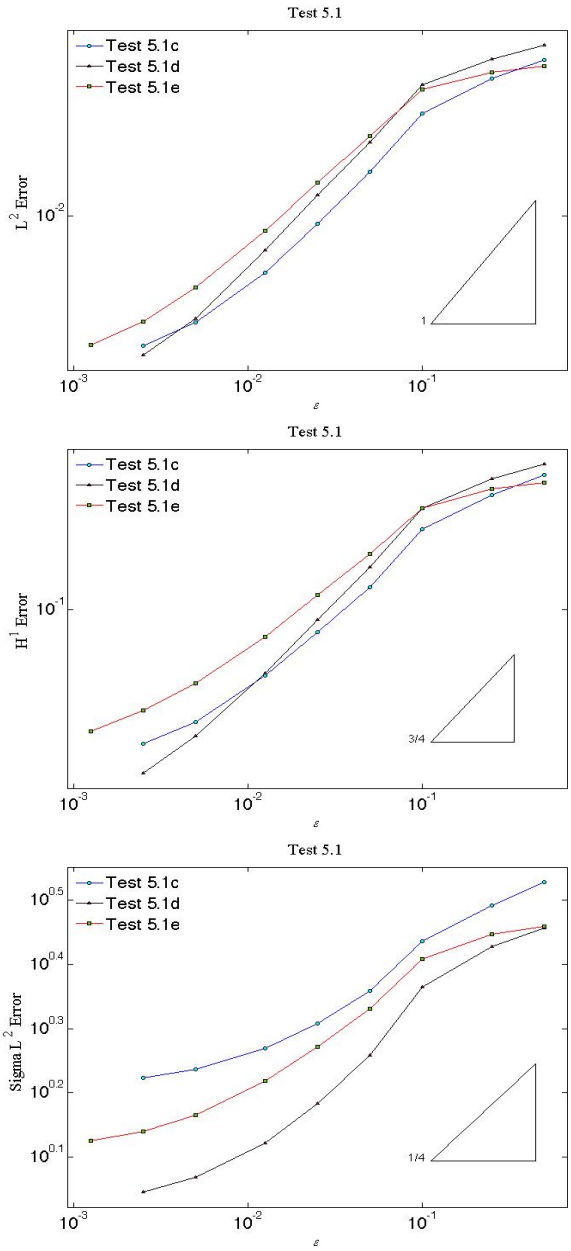


Figure 5.2: Test 5.1 (3-D). Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ.

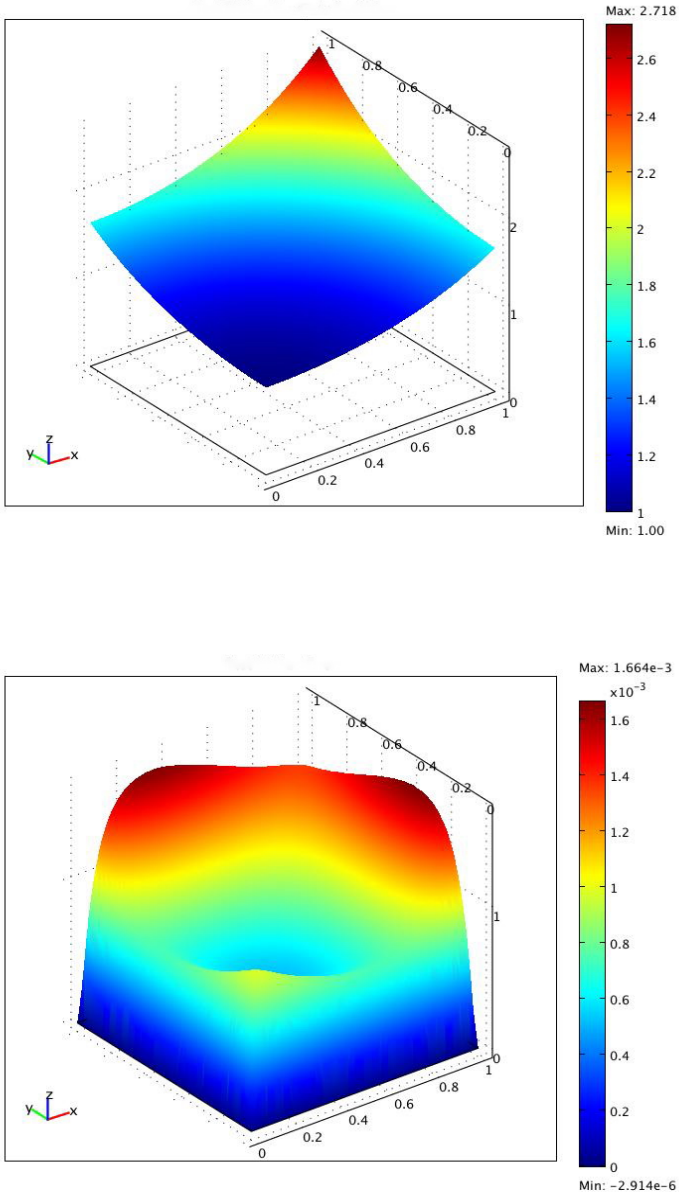


Figure 5.3: Test 5.1a. Computed solution (top) and error (bottom). $\epsilon = 0.0125,\ h = 0.009$

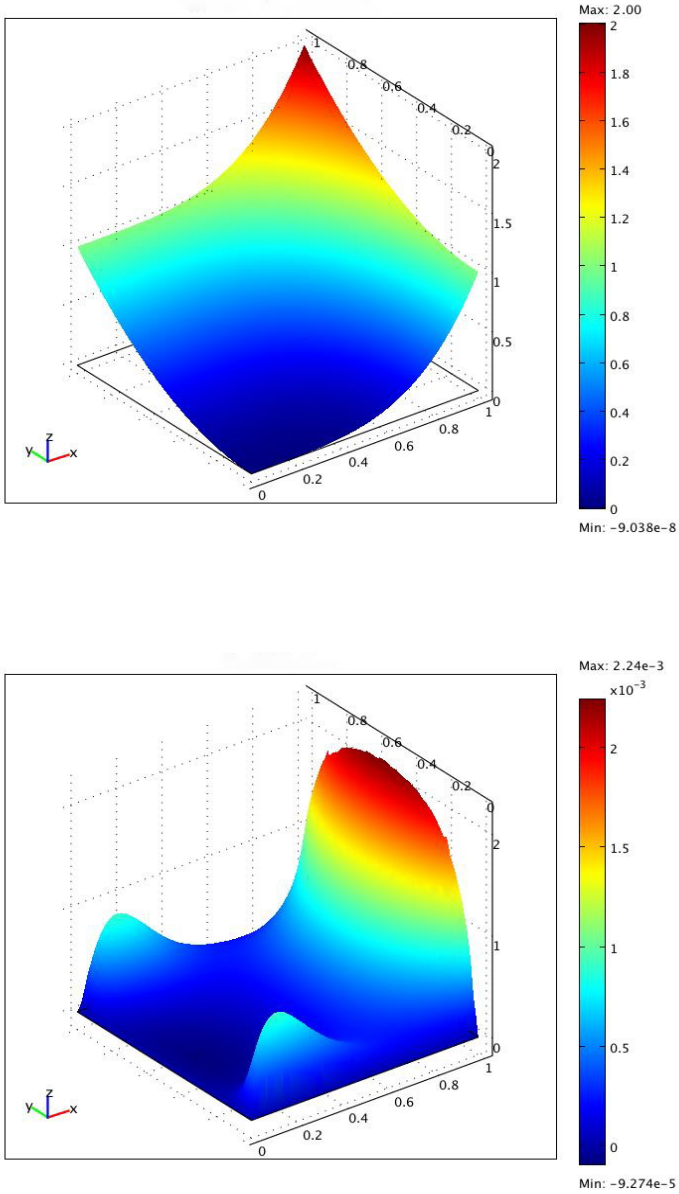


Figure 5.4: Test 5.1b. Computed solution (top) and error (bottom). $\epsilon = 0.0125,\ h = 0.009$

Test 5.2

The purpose of this test is to calculate the rate of convergence of $\|u^\epsilon - u_h^\epsilon\|$ for fixed ϵ in various norms. We use the quadratic Lagrange element for both variables and solve problem (5.8)–(5.9) with boundary condition $D^2u^\epsilon\eta\cdot\eta = \epsilon$ on $\partial\Omega$ being replaced by $D^2u^\epsilon\eta\cdot\eta = \phi^\epsilon$ on $\partial\Omega$ and using the following test functions:

$$
\begin{aligned}
&\text{(a) } u^\epsilon = 20x_1^6 + x_2^6, && f^\epsilon = 18000x_1^4x_2^4 - \epsilon(7200x_1^2 + 360x_2^2),\\
&\quad g^\epsilon = 20x_1^6 + x_2^6, && \phi^\epsilon = 600x_1^4\eta_1^2 + 30x_2^4\eta_2^2,\\
&\text{(b) } u^\epsilon = x_1\sin x_1 + x_2\sin x_2, && f^\epsilon = (2\cos x_1 - x_1\sin x_1)(2\cos x_2 - x_2\sin x_2)\\
&&& \quad - \epsilon(x_1\sin x_1 - 4\cos x_1 + x_2\sin x_2 - 4\cos x_2),\\
&\quad g^\epsilon = x_1\sin x_1 + x_2\sin x_2, && \phi^\epsilon = (2\cos x_1 - x_1\sin x_1)\eta_1^2 + (2\cos x_2 - x_2\sin x_2)\eta_2^2,\\
&\text{(c) } u^\epsilon = x_1^2 + x_2^2 + x_3^2, && f^\epsilon = 8,\\
&\quad g^\epsilon = x_1^2 + x_2^2 + x_3^2, && \phi^\epsilon = 2\eta_1^2 + 2\eta_2^2 + 2\eta_3^2,\\
&\text{(d) } u^\epsilon = x_1^4 + x_2^2 + x_3^6, && f^\epsilon = -\epsilon 8640x_3^2 + 720x_1^2x_3^4,\\
&\quad g^\epsilon = x_1^4 + x_2^2 + x_3^6, && \phi^\epsilon = 12x_1^2\eta_1^2 + 2\eta_2^2 + 30\eta_3^2.
\end{aligned}
$$

After having computed the error in different norms, we divided each value by a power of h expected to be the convergence rate by the analysis in Section 5.3. As seen from Tables 5.1 and 5.2 the error converges exactly as expected in the H^1-norm, but σ_h^ϵ appears to converge one order of h better than the analysis shows (in 2-D/quadratic case). In addition, the error seems to converge optimally in the L^2-norm although a theoretical proof of such a result has not yet been proved. We talk about these findings in the next section, and discuss ways to improve the analysis in Section 5.3 to correspond to the numerical tests.

Table 5.1: Test 5.2 (2-D): Change of $\|u^\epsilon - u_h^\epsilon\|$ w.r.t. h $(\epsilon = 0.001)$

	h	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^2}}{h^3}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^1}}{h^2}$	$\frac{\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2}}{h}$
Test 5.2a	0.1	4.3348494	33.591367	0.836958
	0.05	4.3607192	33.782835	0.2378385
	0.033	4.3657346	33.818929	0.1152246
	0.025	4.3675102	33.832290	0.0699180
	0.02	4.3683359	33.838088	0.0479970
Test 5.2b	0.1	0.0134917	0.1045140	0.0006866
	0.05	0.0134978	0.1045561	0.0002399
	0.033	0.013499	0.1045638	0.0001301
	0.025	0.0134994	0.1045666	8.436E-05
	0.02	0.0134996	0.1045678	6.029E-05

Table 5.2: Test 5.2 (3-D): Change of $\|u^\epsilon - u_h^\epsilon\|$ w.r.t. h $(\epsilon = 0.001)$

	h	$\|u^\epsilon - u_h^\epsilon\|_{L^2}$	$\|u^\epsilon - u_h^\epsilon\|_{H^1}$	$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^2}}{h^3}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^1}}{h^2}$
Test 5.2c	0.17	0.04647	0.2456175	0.7568817	1.5467769	1.4169674
	0.12	0.0224646	0.1715080	0.8748965	1.5659655	1.4319419
	0.074	0.0079533	0.1036339	0.8388694	1.4536629	1.401069
	0.059	0.0051257	0.080715	0.6612057	1.4935865	1.3778199
	0.039	0.0019731	0.0528109	0.5852905	1.2968491	1.3539134
	0.031	0.0011300	0.0416893	0.5282357	1.1770106	1.3454285
	0.021	0.0004444	0.0284367	0.4817416	0.9851921	1.3388245
Test 5.2d	0.17	0.104110518	0.8719690	3.9052875	3.4649376	5.0303878
	0.12	0.0545618	0.6803014	3.9245544	3.8033947	5.6799218
	0.074	0.0197350	0.426235	3.7512680	3.6070523	5.762455
	0.059	0.013048	0.3398852	3.330496	3.8020875	5.8018648
	0.039	0.0075652	0.2291658	3.2466388	4.9723099	5.8751179
	0.031	0.0084335	0.1848377	3.0434380	8.7837988	5.9652200
	0.02	0.0090064	0.1356318	3.0204549	19.963717	6.3856484

Test 5.3

In this test, we fix a relation between ϵ and h, and then determine the "best" choice for h in terms of ϵ such that the global error $u - u_h^\epsilon$ has the same convergence rate as that of $u - u^\epsilon$. We solve problem (5.8)–(5.9) with the following test functions:

$$\begin{aligned} &\text{(a) } u = x_1^4 + x_2^2, && f = 24x_1^2, && g = x_1^4 + x_2^2. \\ &\text{(b) } u = 20x_1^6 + x_2^6, && f = 18000x_1^4x_2^4, && g = 20x_1^6 + x_2^6. \end{aligned}$$

To see which relation gives the sought-after convergence rate, we compare the data with a function, $y = \beta x^\alpha$, where $\alpha = 1$ in the L^2-case, $\alpha = \frac{3}{4}$ in the H^1-case, and $\alpha = \frac{1}{4}$ in the H^2-case. The constant, β is determined using a least squares fitting algorithm based on the data.

As seen in the figures in the Appendix, the best $h-\epsilon$ relation depends on which norm one considers. Figures B.7–B.8 and B.11–B.12 indicate that when $h = \epsilon^{\frac{1}{2}}$, $\|u - u_h^\epsilon\|_{L^2} \approx O(\epsilon)$ and $\|\sigma - \sigma_h^\epsilon\|_{L^2} \approx O(\epsilon^{\frac{1}{4}})$. It can also be seen from Figures B.9–B.10 that when $h = \epsilon$, $\|u - u_h^\epsilon\|_{H^1} = O(\epsilon^{\frac{3}{4}})$.

5.5 Concluding Remarks

In Section 5.4 (Test 5.2), the rate of convergence of $\|\sigma^\epsilon - \sigma_h^\epsilon\|$ in the L^2 and H^1 norms appears to converge faster than what is proved in Section 5.3 which is a very interesting phenomenon considering we have obtained error estimates similar to the Hermann-Myoshi method applied to the linear biharmonic equation [47]; estimates that are sharp in practice. We now wish to explain how the analysis in Section 5.3 could be improved upon to correlate to the numerical tests.

A careful study of Theorem 5.3.4 shows that the error bound of $\|\sigma^\epsilon - \sigma_h^\epsilon\|$ depends on the definition of ρ_1. Moreover, the definition of ρ_1 depends on the results proved in Lemma 5.3.2. Thus, in order to get a better estimate on the error of σ_h^ϵ, we need to improve the results of this Lemma.

To do so, we notice that $(I_h\sigma^\epsilon, I_h u^\epsilon) \notin \mathbb{Z}_h$. It then seems plausible that we could obtain better estimates if we choose functions in $\mathbb{Z}_h$ as the center of the ball, $\mathbb{S}_h(\rho)$. This in fact proves to be the case. However, we require the following assumption that has yet to be shown.

Suppose there exists $(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon) \in \mathbb{Z}_h$ such that

$$\|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{L^2} + h\|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{H^1} \leq Ch^\ell\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big). \tag{5.57}$$

Next, we redefine $\mathbb{S}_h(\rho)$ and $\mathbb{B}_h(\rho)$ as follows:

$$\mathbb{S}_h(\rho) = \{(\mu_h, v_h) \in W_\epsilon^h \times V_g^h;\ \|\mu_h - \tilde{\sigma}_h^\epsilon\|_{L^2} + \epsilon^{-\frac{1}{2}}\|v_h - \tilde{u}_h^\epsilon\|_{H^1} \leq \rho\},$$
$$\mathbb{B}_h(\rho) = \mathbb{S}_h(\rho) \cap \mathbb{Z}_h.$$

We then have the following result.

Proposition 5.5.1. *Assume that* (5.57) *holds. Then under the same hypothesis of Theorem 5.3.4, we have*

$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} \leq C_6(\epsilon)h^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big), \tag{5.58}$$
$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1} \leq C_7(\epsilon)h^{\ell-2}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big). \tag{5.59}$$

Proof. We divide the proof into three steps.

Step 1: Image of Center of Ball

As in Lemma 5.3.2, we ease notation by setting $\omega_h = \tilde{\sigma}_h^\epsilon - T^{(1)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon)$, $s_h = \tilde{u}_h^\epsilon - T^{(2)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon)$. By the definition of T and $\tilde{Z}_h$, we have for any $\mu_h \in W_0^h$, $v_h \in V_0^h$

$$(\omega_h, \mu_h) + (\mathrm{div}(\mu_h), Ds_h) = 0$$
$$\big(\mathrm{div}(\omega_h), Dv_h\big) - \epsilon^{-1}\big(\Phi^\epsilon Ds_h, Dv_h\big) = \big(\mathrm{div}(\tilde{\sigma}_h^\epsilon - \sigma^\epsilon), Dv_h\big) + \epsilon^{-1}\big(\det(\tilde{\sigma}_h^\epsilon) - \det(\sigma^\epsilon), v_h\big).$$

Set $v_h = s_h$, $\mu_h = \omega_h$, subtract the two equations, and use the mean value theorem to get

$$(\omega_h, \omega_h) + \epsilon^{-1}(\Phi^\epsilon Ds_h, Ds_h) = (\mathrm{div}(\sigma^\epsilon - \tilde{\sigma}_h^\epsilon), Ds_h) + \epsilon^{-1}(\Psi^\epsilon : (\sigma^\epsilon - \tilde{\sigma}_h^\epsilon), s_h),$$

where Ψ^ϵ is the same as in Lemma 5.3.2, replacing $I_h\sigma^\epsilon$ by $\tilde{\sigma}_h^\epsilon$. Considering (5.57), the bounds of Ψ^ϵ are the same as in Lemma 5.3.2. Thus, we have

$$\begin{aligned}\|\omega_h\|_{L^2}^2 + \epsilon^{-1}\theta\|Ds_h\|_{L^2}^2 &\leq \|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{H^1}\|Ds_h\|_{L^2} + C\epsilon^{\frac{3}{2}(1-n)}\|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{H^1}\|s_h\|_{H^1} \\ &\leq C\epsilon^{\frac{3}{2}(1-n)}\|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{H^1}\|Ds_h\|_{L^2}.\end{aligned}$$

Using (5.57), Cauchy-Schwarz, and Poincare's inequality, we have

$$\|\omega_h\|_{L^2} + \epsilon^{-\frac{1}{2}}\|s_h\|_{H^1} \leq C\epsilon^{\frac{4-3n}{2}}h^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big). \tag{5.60}$$

Step 2: Contraction Property

We claim that the results of Lemma 5.3.3 hold for the new definition of $\mathbb{B}_h$. Indeed, the proof is exactly the same except in bounding $\|\Phi^\epsilon - \Lambda_h\|_{L^2}$ in Steps 2-3. However, considering (5.57), the same bounds hold, and thus, we have that there exists an $h_0 > 0$ that for $h \leq h_0$, T is a contracting mapping in the ball $\mathbb{B}_h(\rho_0)$ with a contraction factor $\frac{1}{2}$, where h_0 and ρ_0 are defined in Lemma 5.3.3.

Step 3: Finishing up

Let $\rho_2 = 2C\epsilon^{\frac{4-3n}{2}} h^{\ell-1}$. For $n = 2$, choose $h_2 > 0$ such that

$$h_2 |\log h_2|^{\frac{1}{1-\ell}} \leq C \left(\frac{\epsilon^{\frac{3}{2}(n-1)}}{\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}} \right)^{\frac{1}{\ell-1}}.$$

For the case $n = 3$, choose h_2 such that

$$h_2 \leq C \left(\frac{\epsilon^{\frac{3n-1}{2}}}{\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}} \right)^{\frac{2}{2\ell-3}}.$$

Then $h \leq \min\{h_0, h_2\}$ implies $\rho_2 \leq \rho_0$.

Thus, using (5.46) and (5.60), we have for any $(\mu_h, v_h) \in \mathbb{B}_h(\rho_2)$,

$$\begin{aligned}
\|\tilde{\sigma}_h^\epsilon - T^{(1)}(\mu_h, v_h)\|_{L^2} + \epsilon^{-\frac{1}{2}} \|\tilde{u}_h^\epsilon - T^{(2)}(\mu_h, v_h)\|_{H^1} &\leq \|\tilde{\sigma}^\epsilon - T^{(1)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon)\|_{L^2} \\
&\quad + \|T^{(1)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon) - T^{(1)}(\mu_h, v_h)\|_{L^2} + \epsilon^{-\frac{1}{2}} \|\tilde{u}_h^\epsilon - T^{(2)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon)\|_{H^1} \\
&\quad + \epsilon^{-\frac{1}{2}} \|T^{(2)}(\tilde{\sigma}_h^\epsilon, \tilde{u}_h^\epsilon) - T^{(2)}(\mu_h, v_h)\|_{H^1} \\
&\leq C h^{\ell-1} \epsilon^{\frac{4-3n}{2}} \big(\|\sigma^\epsilon\|_{H^l} + \|u^\epsilon\|_{H^l} \big) \\
&\quad + \frac{1}{2} \big(\|\tilde{\sigma}_h^\epsilon - \mu_h\|_{L^2} + \epsilon^{-\frac{1}{2}} \|\tilde{u}_h^\epsilon - v_h\|_{H^1} \big) \\
&\leq \frac{\rho_2}{2} + \frac{\rho_2}{2} = \rho_2 < 1.
\end{aligned}$$

From this result, we can conclude that that $(\sigma_h^\epsilon, u_h^\epsilon) \in \mathbb{B}_h(\rho_2)$.

Thus,

$$\begin{aligned}
\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} &\leq \|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{L^2} + \|\tilde{\sigma}_h^\epsilon - \sigma_h^\epsilon\|_{L^2} \\
&\leq C h^\ell \big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell} \big) + \rho_2 \\
&\leq C h^{\ell-1} C_6(\epsilon) \big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell} \big).
\end{aligned}$$

Finally, using the inverse inequality, we have

$$\begin{aligned}\|\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1} &\le \|\sigma^\epsilon - \tilde{\sigma}_h^\epsilon\|_{H^1} + \|\tilde{\sigma}_h^\epsilon - \sigma_h^\epsilon\|_{H^1} \\ &\le Ch^{\ell-1}\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big) + h^{-1}\rho_2 \\ &\le Ch^{\ell-2}C_7(\epsilon)\big(\|\sigma^\epsilon\|_{H^\ell} + \|u^\epsilon\|_{H^\ell}\big).\end{aligned}$$

We note by the definition of ρ_2 that $C_6(\epsilon) = C_7(\epsilon) = O(\epsilon^{\frac{4-3n}{2}})$. □

Chapter 6

A Nonconforming Morley Finite Element Method for the Monge-Ampère Equation

The specific goal of this chapter is to develop and analyze a nonconforming Morley finite element method to approximate the solution of (2.8)–(2.10), which in turn will approximate the unique convex viscosity solution of (1.11)–(1.12). As in Chapters 3–5, when deriving error estimates of the proposed numerical method, we are particularly interested in obtaining error bounds that show explicit dependence on ϵ.

The motivation to use the Morley finite element to approximate (2.8)–(2.10) is that it has the least number of degrees of freedom on each element for fourth order problems, as its basis functions consist of only quadratic polynomials [74, 76, 89]. As a result, using Morley elements results in only a third of the amount of degrees of freedom compared to fifth order Argyris elements. Therefore, the total computation time is considerably shorter.

The rest of the chapter is organized as follows. In Section 6.1, we introduce the Morley finite element as defined in [74], where the well-known two dimensional Morley element is generalized to any dimension in a canonical fashion. We then define the variational formulation of (2.8)–(2.10) and the nonconforming finite element method based upon the weak formulation. In Section 6.2, we state certain properties of the Morley finite element which will play a crucial role in the analysis in Sections 6.3 and 6.4. In Section 6.3, we study the approximation of linearization of problem (2.8) using the Morley finite element. The results of this section are used in the error analysis for the numerical method presented in Section 6.1. The main results of the chapter are presented in Section 6.4, where we prove existence, uniqueness, and optimal error estimates in the energy norm for the proposed nonconforming finite element scheme. Our main idea is to use a fixed point technique

using the strong stability properties of the linearized problem established in Section 6.3. From this result, we are also able to employ a duality argument to obtain optimal order error estimates in the broken H^1-norm. Finally, we end the chapter with several numerical tests, comparing the results of our tests with those in Chapters 3 and 5 and validating the analysis presented in the previous sections.

6.1 The Morely Element and Finite Element Formulation

In this section, we give the precise definition of our finite element formulation. First, we adopt the space notation of Chapter 3, that is,

$$V := H^2(\Omega), \quad V_0 := H^2(\Omega) \cap H_0^1(\Omega), \quad V_g := \{v \in V;\ v\big|_{\partial\Omega} = g\}.$$

Before defining the variational formulation of (2.8)–(2.10), we first provide the following technical lemma.

Lemma 6.1.1. *Let η denote the unit outward normal of $\partial\Omega$. Then, there exists an orthogonal frame of the tangent space of $\partial\Omega$ such that for u, v sufficiently smooth and $v\big|_{\partial\Omega} = 0$, there holds the following identity ($n = 2, 3$):*

$$(\Delta u, \Delta v) = (D^2 u, D^2 v) - \sum_{i=1}^{n-1} \left\langle \frac{\partial^2 u}{\partial (\tau^{(i)})^2}, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}$$

Proof. First, we write [48]

$$(\Delta u, \Delta v) = (D^2 u, D^2 v) + \left\langle \frac{\partial}{\partial \eta} Du, Dv \right\rangle_{\partial\Omega} - \left\langle \Delta u, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega},$$

$$\begin{aligned}
\frac{\partial}{\partial \eta} Du &= D^2 u \eta \\
&= \sum_{i=1}^{n-1} \left(\frac{\partial^2 u}{\partial (\tau^{(i)})^2} (\tau^{(i)})(\tau^{(i)})^t + \frac{\partial^2 u}{\partial \tau^{(i)} \partial \eta} ((\tau^{(i)})(\eta)^t + (\eta)(\tau^{(i)})^t) + \frac{\partial^2 u}{\partial \eta^2} (\eta)(\eta)^t \right) \eta \\
&= \sum_{i=1}^{n-1} \left(\frac{\partial^2 u}{\partial \tau^{(i)} \partial \eta} \tau^{(i)} + \frac{\partial^2 u}{\partial \eta^2} \eta \right)
\end{aligned}$$

Since $v\big|_{\partial\Omega} = 0$

$$Dv\big|_{\partial\Omega} = \frac{\partial v}{\partial \eta} \eta\big|_{\partial\Omega},$$

and we have

$$(\Delta u, \Delta v) = (D^2 u, D^2 v) + \left\langle \frac{\partial^2 u}{\partial \eta^2} - \Delta u, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}.$$

The case $n = 2$: In the two dimensional case, we take $\tau = (\eta_2, -\eta_1)$. We then have

$$\begin{aligned}
(\Delta u, \Delta v) &= (D^2 u, D^2 v) + \left\langle \frac{\partial^2 u}{\partial \eta^2} - \Delta u, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= (D^2 u, D^2 v) + \left\langle \frac{\partial^2 u}{\partial x_1^2}(\eta_1^2 - 1) + \frac{\partial^2 u}{\partial x_2^2}(\eta_2^2 - 1) + 2\frac{\partial^2 u}{\partial x_1 \partial x_2}\eta_1\eta_2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}, \\
&= (D^2 u, D^2 v) - \left\langle \frac{\partial^2 u}{\partial x_1^2}\eta_2^2 + \frac{\partial^2 u}{\partial x_2^2}\eta_1^2 - 2\frac{\partial^2 u}{\partial x_1 \partial x_2}\eta_1\eta_2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}, \\
&= (D^2 u, D^2 v) - \left\langle \frac{\partial^2 u}{\partial \tau^2}, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}.
\end{aligned}$$

The case $n = 3$: If $\eta_2^2 + \eta_3^2 > 0$ take

$$\begin{aligned}
\tau^{(1)} &= \left(0, \frac{\eta_3}{\sqrt{\eta_2^2 + \eta_3^2}}, -\frac{\eta_2}{\sqrt{\eta_2^2 + \eta_3^2}}\right), \\
\tau^{(2)} &= \left(\sqrt{\eta_2^2 + \eta_3^2}, -\frac{\eta_1\eta_2}{\sqrt{\eta_2^2 + \eta_3^2}}, -\frac{\eta_1\eta_3}{\sqrt{\eta_2^2 + \eta_3^2}}\right).
\end{aligned}$$

Otherwise let

$$\begin{aligned}
\tau^{(1)} &= (0, 1, 0), \\
\tau^{(2)} &= (0, 0, 1).
\end{aligned}$$

We note that with this choice of tangental vectors, the following identities hold:

$$\begin{aligned}
\left(\tau_1^{(1)}\right)^2 + \left(\tau_1^{(2)}\right)^2 &= \eta_2^2 + \eta_3^2, \\
\left(\tau_2^{(1)}\right)^2 + \left(\tau_2^{(2)}\right)^2 &= \eta_1^2 + \eta_3^2, \\
\left(\tau_3^{(1)}\right)^2 + \left(\tau_3^{(2)}\right)^2 &= \eta_1^2 + \eta_2^2, \\
\tau_1^{(1)}\tau_2^{(1)} + \tau_1^{(2)}\tau_2^{(2)} &= -\eta_1\eta_2, \\
\tau_1^{(1)}\tau_3^{(1)} + \tau_1^{(2)}\tau_3^{(2)} &= -\eta_1\eta_3, \\
\tau_2^{(1)}\tau_3^{(1)} + \tau_2^{(2)}\tau_3^{(2)} &= -\eta_2\eta_3.
\end{aligned}$$

Thus,

$$\begin{aligned}
(\Delta u, \Delta v) &= (D^2u, D^2v) + \left\langle \frac{\partial^2 u}{\partial \eta^2} - \Delta u, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= (D^2u, D^2v) + \left\langle \frac{\partial^2 u}{\partial x_1^2}(\eta_1^2 - 1) + \frac{\partial^2 u}{\partial x_2^2}(\eta_2^2 - 1) + \frac{\partial^2 u}{\partial x_3^2}(\eta_3^2 - 1), \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&\quad + 2\left\langle \frac{\partial^2 u}{\partial x_1 \partial x_2}\eta_1\eta_2 + \frac{\partial^2 u}{\partial x_1 \partial x_3}\eta_1\eta_3 + \frac{\partial^2 u}{\partial x_2 \partial x_3}\eta_2\eta_3, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= (D^2u, D^2v) - \left\langle \frac{\partial^2 u}{\partial x_1^2}(\eta_2^2 + \eta_3^2) + \frac{\partial^2 u}{\partial x_2^2}(\eta_1^2 + \eta_3^2) + \frac{\partial^2 u}{\partial x_3^2}(\eta_1^2 + \eta_2^2), \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&\quad + 2\left\langle \frac{\partial^2 u}{\partial x_1 \partial x_2}\eta_1\eta_2 + \frac{\partial^2 u}{\partial x_1 \partial x_3}\eta_1\eta_3 + \frac{\partial^2 u}{\partial x_2 \partial x_3}\eta_2\eta_3, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= (D^2u, D^2v) - \sum_{i=1}^{n-1} \left\langle \frac{\partial^2 u}{\partial (\tau^{(i)})^2}, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}.
\end{aligned}$$

□

Multiplying (2.8) by $v \in V_0$, integrating over Ω, and integrating by parts, and using Lemma 6.1.1 we make the following identity:

$$\begin{aligned}
(f, v) &= -\epsilon(\Delta^2 u^\epsilon, v) + (\det(D^2 u^\epsilon), v) \\
&= \epsilon(D(\Delta u^\epsilon), Dv) + (\det(D^2 u^\epsilon), v) \\
&= -\epsilon(\Delta u^\epsilon, \Delta v) + (\det(D^2 u^\epsilon), v) + \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \\
&= -\epsilon(D^2 u^\epsilon, D^2 v) + (\det(D^2 u^\epsilon), v) + \epsilon\Big\langle \frac{\partial^2 g}{\partial \tau^2} + \epsilon, \frac{\partial v}{\partial \eta} \Big\rangle_{\partial\Omega},
\end{aligned} \tag{6.1}$$

Here and for the continuation of the chapter, we shall omit the summation of tangential derivatives for notation convenience.

Based on (6.1), we define the variational formulation of (2.8)–(2.10) as to find $u^\epsilon \in V_g$ such that

$$-\epsilon(D^2 u^\epsilon, D^2 v) + (\det(D^2 u^\epsilon), v) = (f, v) - \epsilon\Big\langle \frac{\partial^2 g}{\partial \tau^2} + \epsilon, \frac{\partial v}{\partial \eta} \Big\rangle_{\partial\Omega} \qquad \forall v \in V_0. \tag{6.2}$$

Let $\mathcal{T}_h$ be a quasiuniform triangular mesh of Ω if $n = 2$ or a quasiuniform tetrahedral mesh if $n = 3$ with mesh size $h \in (0, 1)$. Let K be a $n-$simplex with $n + 1$ vertices which we will denote by a_i $(1 \le i \le n)$. Let F_j $(1 \le j \le n + 1)$ denote the $(n - 1)$-dimensional subsimplex of K (without vertices), b_j denote the barycenter of F_j, and S_{ij} $(1 \le i < j \le n + 1)$ be the $(n - 2)$-dimensional subsimplex without a_i and a_j as its vertices. Next, we let $\mathcal{E}_h$ denote the set of all $(n - 1)$-dimensional subsimplexes in the mesh $\mathcal{T}_h$, and define

the set of interior and boundary $(n-1)$-dimensional subsimplexes as follows:

$$\mathcal{E}_h^i := \{F;\ F \cap \partial\Omega = \emptyset\},$$
$$\mathcal{E}_h^b := \{F;\ F \cap \partial\Omega \neq \emptyset\}.$$

For given $K \in \mathcal{T}_h$, set

$$\mathcal{E}_h(K) := \{F \in \mathcal{E}_h;\ F \subset \partial K\},$$
$$\mathcal{E}_h^b(K) := \mathcal{E}_h(K) \cap \mathcal{E}_h^b$$
$$\mathcal{E}_h^i(K) := \mathcal{E}_h(K) \cap \mathcal{E}_h^i = \mathcal{E}_h(K) \backslash \mathcal{E}_h^b(K).$$

Let (K, P_K, Σ_K) be the n-dimensional Morley element defined in [74], that is,

1. K, an $n-$dimensional simplex,
2. $P_K = \mathbb{P}_2(K)$, the space of quadratic polynomials on K,
3. Σ_K, the linear independent functionals, $\{\phi_{ij}^K, \psi_j^K\}$, such that for $v_h \in P_K$,

$$\phi_{ij}^K(v_h) := \frac{1}{|S_{ij}|}\int_{S_{ij}} v_h d\sigma \qquad 1 \le i < j \le n+1,$$
$$\psi_j^K(v_h) := \frac{1}{|F_j|}\int_{F_j} \frac{\partial v_h}{\partial \eta_{F_j}} ds = \frac{\partial v_h}{\partial \eta_{F_j}}(b_j) \qquad 1 \le j \le n+1,$$

where ds and $d\sigma$ are the $(n-1)$ and $(n-2)$ dimensional Lebesgue measures respectively, and η_{F_j} denotes the outward normal of the $(n-1)$-dimensional subsimplex, F_j. We have used the fact that v_h is quadratic to obtain the last equality. The two and three dimensional Morley element are depicted in Figure 6.1.

Remark 6.1.2. *In the two dimensional case,*

$$\phi_{ij}^K(v_h) = v_h(a_k) \qquad k \neq i,\ k \neq j \qquad \forall v_h \in P_K.$$

Remark 6.1.3. *The Morley element of $n-$dimension is P_K-unisolvent [74, Lemma 2], and the following bound holds for all $v \in H^3(K)$:*

$$|v - I_K v|_{H^m(K)} \le C h^{3-m} |v|_{H^3(K)} \qquad m = 0, 1, 2, 3, \tag{6.3}$$

where I_K is the standard interpolation operator.

Let V^h be the finite element space corresponding to the Morley element defined above,

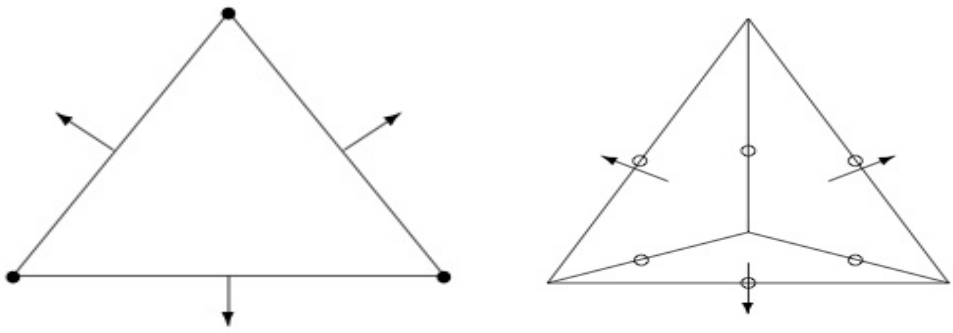


Figure 6.1: The two (left) and three (right) dimensional Morley element. Solid circles indicate function value evaluation, arrows indicate normal derivative evaluation, and open circles indicate function average evaluation.

and let

$$V_0^h := \{v_h \in V^h;\ \phi_{ij}^K(v_h) = 0\ \forall (i,j) \text{ with } \partial\Omega \cap S_{ij} \neq \emptyset\},$$
$$V_g^h := \{v_h \in V^h;\ \phi_{ij}^K(v_h) = \phi_{ij}^K(g)\ \forall (i,j) \text{ with } \partial\Omega \cap S_{ij} \neq \emptyset\}.$$

Based on (6.2), we define our finite element method as follows: Find $u_h^\epsilon \in V_g^h$ such that

$$\sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D^2 u_h^\epsilon, D^2 v_h)_K + (\det(D^2 u^\epsilon), v_h)_K \Big\} \tag{6.4}$$
$$= (f, v_h) - \epsilon \sum_{F\in\mathcal{E}_h^b} \Big\langle \frac{\partial^2 g}{\partial \tau_F^2} + \epsilon, \frac{\partial v_h}{\partial \eta_F} \Big\rangle_F \qquad \forall v_h \in V_0^h.$$

where

$$(D^2 v, D^2 w)_K := \int_K D^2 v : D^2 w dx = \sum_{i,j=1}^n \int_K \frac{\partial^2 v}{\partial x_i \partial x_j} \frac{\partial^2 w}{\partial x_i \partial x_j} dx,$$
$$\langle v, w\rangle_F := \int_F vw ds, \qquad \frac{\partial^2 g}{\partial \tau_F^2} := \sum_{i=1}^{n-1} \frac{\partial^2 g}{(\partial \tau_F^{(i)})^2},$$

and $\{\tau_F^{(i)}\}_{i=1}^{n-1}$ denotes an orthogonal frame of the tangent space of F.

Let u^ϵ be the solution to (6.2) and u_h^ϵ the solution to (6.4). The main goals of this chapter are to prove existence and uniqueness of u_h^ϵ and to also show optimal order error estimates of $u^\epsilon - u_h^\epsilon$. To realize these goals, we use a combined linearization and fixed point technique similar to the one employed in Chapters 3 and 4.

6.2 Properties of the Morley Element

Before studying the finite approximation of the linearization of (2.8), we first state certain properties of the Morley element which will used in both Sections 6.3 and 6.4.

For $v_h \in V^h$ and $K \in \mathcal{T}_h$, let v_h^K be the continuous extension of v_h from the interior of K to ∂K. Given any $(n-1)$-dimensional subsimplex F, define the jumps of v_h across F as follows:

$$\begin{aligned} [v_h] &= v_h^{K^+} - v_h^{K^-} \qquad && F = \partial K^+ \cap \partial K^-, \\ [v_h] &= v_h^{K^+} && F = \partial K^+ \cap \partial\Omega. \end{aligned}$$

We then have the following lemma [74].

Lemma 6.2.1. *If F is a common $(n-1)$-dimensional subsimplex of K^+ and K^-, then for all $v_h \in V^h$, $\alpha \in \mathbf{R}^n$*

$$\int_F [Dv_h \cdot \alpha]\, ds = 0. \tag{6.5}$$

We note that (6.5) does not hold for $F \in \mathcal{E}_h^b$ even for $v_h \in V_0^h$. However, we do have the following useful result.

Lemma 6.2.2. *For any $v_h \in V_0^h$ and $F \in \mathcal{E}_h^b$, we have*

$$\int_F \frac{\partial v_h}{\partial \tau_F^{(i)}} ds = 0 \qquad i = 1, ..., n-1. \tag{6.6}$$

Proof. Given $F \in \mathcal{E}_h^b$, let $S_1, ..., S_n$ denote the $(n-2)$-dimensional subsimplexes of F, and η_{S_j} the unit outward normal of S_j. We then have for any $v_h \in V_0^h$, $i = 1, ..., n-1$

$$\int_F \frac{\partial v_h}{\partial \tau_F^{(i)}} = \tau_F^{(i)} \cdot \eta_F \int_F \frac{\partial v_h}{\partial \eta_F} ds + \sum_{j=1}^{n} \tau_F^{(i)} \cdot \eta_{S_j} \int_{S_j} v_h d\sigma = 0.$$

□

Next, we introduce the mesh-dependent norms and semi-norms for $v\big|_K \in H^m(K)\ \forall K \in \mathcal{T}_h$:

$$\|v\|_{m,h} := \left(\sum_{K \in \mathcal{T}_h} \|v\|_{H^m(K)}^2 \right)^{\frac{1}{2}}, \qquad |v|_{m,h} := \left(\sum_{K \in \mathcal{T}_h} |v|_{H^m(K)}^2 \right)^{\frac{1}{2}}.$$

The next lemma is a consequence of the proof of [74, Lemma 6].

Lemma 6.2.3. *For any $v_h \in V_0^h$, there exists $v_0 \in H_0^1(\Omega)$ with $v_0\big|_K \in \mathbb{P}_1(K)$ $\forall K \in \mathcal{T}_h$ such that*

$$|v_h - v_0|_{m,h} \leq Ch^{2-m}|v_h|_{2,h} \qquad m = 0, 1, 2, \tag{6.7}$$

where C is independent of the mesh parameter h.

Next, with Lemma 6.2.3 in hand, we are immediately able to derive a Poincaré-type inequality in the mesh-dependent norm.

Lemma 6.2.4. *There exists a constant $C > 0$ independent of h such that*

$$|v_h|_{1,h} \leq \|v_h\|_{1,h} \leq C|v_h|_{1,h} \qquad \forall v_h \in V_0^h. \tag{6.8}$$

Proof. Given $v_h \in V_0^h$, let $v_0 \in H_0^1(\Omega)$ be the linear interpolant of v_h such that (6.7) holds. We then have

$$\begin{aligned}
\|v_h\|_{L^2} &\leq \|v_0\|_{L^2} + \|v_h - v_0\|_{L^2} \\
&\leq C\big(|v_0|_{H^1} + h^2|v_h|_{2,h}\big) \\
&\leq C\big(|v_h|_{1,h} + |v_h - v_0|_{1,h} + h^2|v_h|_{2,h}\big) \\
&\leq C\big(|v_h|_{1,h} + h|v_h|_{2,h}\big) \\
&\leq C|v_h|_{1,h},
\end{aligned}$$

where we have used the inverse inequality. □

6.3 Finite Element Approximation of the Linearized Problem

To prove existence, uniqueness, and error estimates of the finite element method (6.4), we first study the linearization of (2.8) at the solution u^ϵ. That is, for given $\varphi \in L^2(\Omega)$, $\psi \in H^{\frac{3}{2}}(\partial\Omega)$, we now consider the following problem:

$$L_{u^\epsilon}(v) = \varphi \qquad \text{in } \Omega, \tag{6.9}$$

$$v = 0 \qquad \text{on } \partial\Omega, \tag{6.10}$$

$$\Delta v = \psi \qquad \text{on } \partial\Omega, \tag{6.11}$$

where $L_{u^\epsilon}(v) = \epsilon\Delta^2 v - \Phi^\epsilon : D^2 v$, and $\Phi^\epsilon = \mathrm{cof}(D^2 u^\epsilon)$.

Existence and uniqueness for problem (6.9)–(6.11) were shown in Chapter 3 (cf. Theorems 3.2.1, 3.2.2, and 3.2.3). Thus, in this section, we are only concerned with the finite

element approximation of (6.9)–(6.11) using the Morley finite element.

To define the variational formulation of (6.9)–(6.11), we first define the following bilinear form:

$$a^\epsilon(v,w) := \epsilon(D^2v, D^2w) + (\Phi^\epsilon Dv, Dw).$$

The variational formulation is then defined as seeking $v \in V_0$ such that

$$a^\epsilon(v,w) = (\varphi, w) + \epsilon \left\langle \psi, \frac{\partial w}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w \in V_0$$

Based on the variational formulation, we define the finite element method of (6.9)–(6.11) as to find $v_h \in V_0^h$ such that

$$a_h^\epsilon(v_h, w_h) = (\varphi, w_h) + \epsilon \sum_{F\in\mathcal{E}_h^b} \left\langle \psi, \frac{\partial w_h}{\partial \eta_F} \right\rangle_F \qquad \forall w_h \in V_0^h, \tag{6.12}$$

where

$$a_h^\epsilon(v_h, w_h) := \sum_{K\in\mathcal{T}_h} \left\{ \epsilon(D^2v_h, D^2w_h)_K + (\Phi^\epsilon Dv_h, Dw_h)_K \right\}.$$

We then have the following theorem.

Theorem 6.3.1. *There exists a unique solution to* (6.12). *Moreover, if* $\varphi \in L^2(\Omega)$, $v \in H^3(\Omega)$, *where* v *is unique solution to* (6.9)–(6.11), *then there exists a constant* C_1 *independent of* ϵ *and* h *such that the following bound holds:*

$$\|v - v_h\|_{2,h} \le C_1\epsilon^{-1}h\Big\{\epsilon^{-1}\|v\|_{H^1} + \epsilon\|v\|_{H^3} + h\|\varphi\|_{L^2}\Big\}. \tag{6.13}$$

Proof. Noting u^ϵ is strictly convex, Φ^ϵ is positive definite, and therefore there exists a constant $\theta > 0$ such that $(\Phi^\epsilon Dw, Dw)_K \ge \theta\|Dw\|^2_{L^2(K)} \quad \forall w \in H^1(K),\ K \in \mathcal{T}_h$. Thus, by Lemma 6.2.4, we have

$$a_h^\epsilon(w_h, w_h) \ge C\epsilon\|w_h\|^2_{2,h},$$

and it follows that there exists a unique $v_h \in V_0^h$ solving (6.12).

To derive (6.13), we use Strang's Second Lemma [27] to conclude

$$\|v - v_h\|_{2,h} \le C\epsilon^{-1}\Big\{ \inf_{w_h\in V_0^h} \|v - w_h\|_{2,h} + \sup_{w_h\in V_0^h} \frac{|E^\epsilon(w_h)|}{\|w_h\|_{2,h}} \Big\}, \tag{6.14}$$

where

$$E^\epsilon(w_h) := a_h^\epsilon(v, w_h) - (\varphi, w_h) - \epsilon \sum_{F\in\mathcal{E}_h^b} \langle \psi, \frac{\partial w_h}{\partial \eta_F}\rangle_F.$$

For $w_h \in V_0^h$, let $w_0 \in H_0^1(\Omega)$ be the linear interpolant such that (6.7) holds. Integrating by parts, we obtain

$$\begin{aligned}
(\varphi, w_h) &= (\epsilon\Delta^2 v - \Phi^\epsilon : D^2 v, w_0) + (\varphi, w_h - w_0)\\
&= -\epsilon(D(\Delta v), Dw_0) + (\Phi^\epsilon Dv, Dw_0) + (\varphi, w_h - w_0)\\
&= \sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D(\Delta v), Dw_h)_K + (\Phi^\epsilon Dv, Dw_h)_K\\
&\qquad\qquad - (\epsilon D(\Delta v) - \Phi^\epsilon Dv, D(w_0 - w_h))_K\Big\} + (\varphi, w_h - w_0)\\
&= a_h^\epsilon(v, w_h) - \sum_{K\in\mathcal{T}_h} \Big\{(\epsilon D(\Delta v) - \Phi^\epsilon Dv, D(w_0 - w_h))_K\\
&- \epsilon \sum_{F\in\mathcal{E}_h(K)} \Big(\Big\langle \Delta v + \frac{\partial^2 v}{\partial \tau_F^2}, \frac{\partial w_h}{\partial \eta_F}\Big\rangle_F - \Big\langle \frac{\partial^2 v}{\partial \eta_F \partial \tau_F}, \frac{\partial w_h}{\partial \tau_F}\Big\rangle_F\Big)\Big\}\\
&+ (\varphi, w_h - w_0).
\end{aligned}$$

Thus,

$$E^\epsilon(w_h) = G^\epsilon(v, w_h) + H^\epsilon(v, w_h, w_0), \tag{6.15}$$

where

$$H^\epsilon(v, w_h, w_0) := (\varphi, w_0 - w_h) + \sum_{K\in\mathcal{T}_h} (\epsilon D(\Delta v) - \Phi^\epsilon Dv, D(w_0 - w_h))_K,$$

$$\begin{aligned}
G^\epsilon(v, w_h) := \epsilon \sum_{K\in\mathcal{T}_h} \Big\{ &\sum_{F\in\mathcal{E}_h^i(K)} \Big\langle \Delta v + \frac{\partial^2 v}{\partial \tau_F^2}, \frac{\partial w_h}{\partial \eta_F}\Big\rangle_F\\
&- \sum_{F\in\mathcal{E}_h(K)} \Big\langle \frac{\partial^2 v}{\partial \eta_F \tau_F}, \frac{\partial w_h}{\partial \tau_F}\Big\rangle_F\Big\}.
\end{aligned}$$

Bounding $H^\epsilon(v, w_h, w_0)$, we use (2.11) to derive

$$\begin{aligned}
\big|H^\epsilon(v, w_h, w_0)\big| &\le C\Big\{\|\varphi\|_{L^2}\|w_0 - w_h\|_{L^2} + \big(\epsilon\|v\|_{H^3} + \epsilon^{-1}\|v\|_{H^1}\big)\|w_0 - w_h\|_{1,h}\Big\} \qquad (6.16)\\
&\le C\Big\{h\big(\epsilon\|v\|_{H^3} + \epsilon^{-1}\|v\|_{H^1}\big) + h^2\|\varphi\|_{L^2}\Big\}\|w_h\|_{2,h}.
\end{aligned}$$

To bound $G^\epsilon(v, w_h)$, we let $P_F : L^2(F) \to \mathbb{P}_0(F)$ denote the constant L^2-projection. Using Lemmas 6.2.1 and 6.2.2, we make the following identity:

$$
\begin{aligned}
&G^\epsilon(v, w_h) = \\
&\epsilon \sum_{K \in \mathcal{T}_h} \Big\{ \sum_{F \in \mathcal{E}_h^i(K)} \Big\langle \Delta v + \frac{\partial^2 v}{\partial \tau_F^2} - P_F\big(\Delta v + \frac{\partial^2 v}{\partial \tau_F^2}\big), \frac{\partial w_h}{\partial \eta_F} - P_F\big(\frac{\partial w_h}{\partial \eta_F}\big) \Big\rangle_F, \\
&- \sum_{F \in \mathcal{E}_h(K)} \Big\langle \frac{\partial^2 v}{\partial \eta_F \partial \tau_F} - P_F\big(\frac{\partial^2 v}{\partial \eta_F \partial \tau_F}\big), \frac{\partial w_h}{\partial \tau_F} - P_F\big(\frac{\partial w_h}{\partial \tau_F}\big) \Big\rangle_F \Big\}.
\end{aligned} \tag{6.17}
$$

Thus,

$$
\big|G^\epsilon(v, w_h)\big| \leq C\epsilon h \|v\|_{H^3} \|w_h\|_{2,h}. \tag{6.18}
$$

Combining (6.15), (6.16), and (6.18)

$$
|E^\epsilon(w_h)| \leq Ch\big(\epsilon^{-1}\|v\|_{H^1} + \epsilon\big(\|v\|_{H^3} + h\|\varphi\|_{L^2}\big)\Big\}\|w_h\|_{2,h}. \tag{6.19}
$$

Completing the proof, we use (6.14), (6.3), and (6.19) to obtain

$$
\|v - v_h\|_{2,h} \leq C\epsilon^{-1}h\Big\{\epsilon^{-1}\|v\|_{H^1} + \epsilon\|v\|_{H^3} + h\|\varphi\|_{L^2}\Big\}.
$$

□

6.4 Finite Element Approximation of (6.4)

In this section, we provide our main results, where we show existence and error estimates of the solution to (6.4). As in Chapters 3–5, the strategy to prove these results is to use a fixed point argument that relies on the stability properties of the linearized problem. However, there are subtile but important differences between the analysis presented here and the fixed point arguments in the aforementioned chapters. Namely, the finite element space V^h is not part of the energy space, and as a result, interior edge integrals appear when integrating by parts. To overcome this additional difficulty, we use the approximation properties established in Lemmas 6.2.3, 6.2.2, and 6.2.1 to bound these extra terms appropriately.

As a first step in proving existence, we define a linear operator $T_M : V_g^h \mapsto V_g^h$ such that for given $v_h \in V_g^h$, $T_M(v_h)$ is the solution to the following problem:

$$a_h^\epsilon(v_h - T_M(v_h), w_h) = \sum_{K\in\mathcal{T}_h} \big\{\epsilon(D^2 v_h, D^2 w_h)_K - (\det(D^2 v_h), w_h)_K\big\} + (f, w_h) - \epsilon \sum_{F\in\mathcal{E}_h^b} \Big\langle \frac{\partial^2 g}{\partial \tau_F^2} + \epsilon, \frac{\partial w_h}{\partial \eta_F}\Big\rangle_F.$$

By Theorem 6.3.1, T_M is well-defined. We also see that any fixed point of T_M (i.e. $T_M(v_h) = v_h$) will be a solution to (6.4) and vice-versa. The main task of this section is to show T_M has a unique fixed point in a small neighborhood of u^ϵ. To prove this, we first give the following definition:

$$\mathbb{B}_h(\rho) := \{v_h \in V_g^h;\ \|I_h u^\epsilon - v_h\|_{2,h} \le \rho\},$$

where $I_h u^\epsilon \in V_g^h$ is defined such that $I_h u^\epsilon\big|_K := I_K u^\epsilon\big|_K\ \forall K \in \mathcal{T}_h$.

We then have the following lemma.

Lemma 6.4.1. *The following bound holds* $(n = 2, 3)$*:*

$$\|I_h u^\epsilon - T_M(I_h u^\epsilon)\|_{2,h} \le C_2\big(\epsilon^{\frac{3}{2}(1-n)} h \|u^\epsilon\|_{H^3} + h^2 \|\Delta^2 u^\epsilon\|_{L^2}\big), \tag{6.20}$$

where the constant, C_2*, is independent of* ϵ *and the mesh parameter* h*.*

Proof. Let $\omega_h^\epsilon := I_h u^\epsilon - T_M(I_h u^\epsilon)$ and $\alpha^\epsilon := u^\epsilon - I_h u^\epsilon$. By the definition of T_M, we have for any $w_h \in V_0^h$,

$$a_h^\epsilon(\omega_h^\epsilon, w_h) = \sum_{K\in\mathcal{T}_h} \big\{\epsilon(D^2(I_h u^\epsilon), D^2 w_h)_K - (\det(D^2(I_h u^\epsilon)), w_h)_K\big\} + (f, w_h) - \epsilon \sum_{F\in\mathcal{E}_h^b} \Big\langle \frac{\partial^2 g}{\partial \tau_F^2} + \epsilon, \frac{\partial w_h}{\partial \eta_F}\Big\rangle_F.$$

Let $w_0 \in H_0^1(\Omega)$ be the linear interpolant of w_h such that (6.7) holds. We then have

$$\begin{aligned}(f, w_h) &= -\epsilon(\Delta^2 u^\epsilon, w_h) + (\det(D^2 u^\epsilon), w_h)\\ &= \sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D^2 u^\epsilon, D^2 w_h)_K + (\det(D^2 u^\epsilon), w_h)_K\\ &+ \epsilon \sum_{F\in\mathcal{E}(K)} \Big(\Big\langle \Delta u^\epsilon + \frac{\partial^2 u^\epsilon}{\partial \tau_F^2}, \frac{\partial w_h}{\partial \eta_F}\Big\rangle_F - \Big\langle \frac{\partial^2 u^\epsilon}{\partial \eta_F \partial \tau_F}, \frac{\partial w_h}{\partial \tau_F}\Big\rangle_F\Big)\Big\}\\ &+ F^\epsilon(u^\epsilon, w_h, w_0),\end{aligned}$$

where

$$F^\epsilon(u^\epsilon, w_h, w_0) := \epsilon(\Delta^2 u^\epsilon, w_0 - w_h) + \epsilon \sum_{K\in\mathcal{T}_h} (D(\Delta u^\epsilon), D(w_0 - w_h))_K.$$

Thus,

$$\begin{aligned} &a_h^\epsilon(\omega_h^\epsilon, w_h) \\ &\quad = \sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D^2\alpha^\epsilon, D^2 w_h)_K + (\det(D^2 u^\epsilon) - \det(D^2(I_h u^\epsilon)), w_h)_K \Big\} \\ &\qquad + G^\epsilon(u^\epsilon, w_h) + F^\epsilon(u^\epsilon, w_h, w_0), \end{aligned} \tag{6.21}$$

where $G^\epsilon(\cdot,\cdot)$ is defined in Theorem 6.3.1.

We bound $F^\epsilon(u^\epsilon, w_h, w_0)$ as follows:

$$\begin{aligned} \big|F^\epsilon(u^\epsilon, w_h, w_0)\big| &\le C\Big\{\epsilon\|\Delta^2 u^\epsilon\|_{L^2}\|w_0 - w_h\|_{L^2} + \epsilon\|u^\epsilon\|_{H^3}\|w_0 - w_h\|_{1,h}\Big\} \\ &\le C\epsilon\Big\{h\|u^\epsilon\|_{H^3} + h^2\|\Delta^2 u^\epsilon\|_{L^2}\Big\}\|w_h\|_{2,h}. \end{aligned} \tag{6.22}$$

Next, we use the mean value theorem to conclude that for any $K \in \mathcal{T}_h$

$$(\det(D^2 u^\epsilon) - \det(D^2(I_h u^\epsilon)), w_h)_K = (\tilde{\Phi}_K^\epsilon : D^2\alpha^\epsilon, w_h)_K,$$

where $\tilde{\Phi}_K^\epsilon := \mathrm{cof}(D^2 u^\epsilon - \tau_K \alpha^\epsilon)$ for some $\tau_K \in [0,1]$.

For $n = 2$, we have

$$\|\tilde{\Phi}_K^\epsilon\|_{L^2(K)} \le C\|D^2 u^\epsilon\|_{L^2(K)} \le C\epsilon^{-\frac{1}{2}}. \tag{6.23}$$

In the case $n = 3$, we have

$$\|\tilde{\Phi}_K^\epsilon\|_{L^2(K)} \le C\|D^2 u^\epsilon\|_{L^\infty(K)}^2 \le C\epsilon^{-2}. \tag{6.24}$$

Thus, using (6.21)–(6.24), (6.3), (6.18), and a Sobolev inequality, we can derive the following inequality:

$$\big|a_h^\epsilon(\omega_h^\epsilon, w_h)\big| \le C\big(\epsilon^{\frac{1}{2}(5-3n)} h\|u^\epsilon\|_{H^3} + \epsilon h^2\|\Delta^2 u^\epsilon\|_{L^2}\big)\|w_h\|_{2,h}.$$

Finally, using the coercivity of $a_h^\epsilon(\cdot,\cdot)$, we obtain (6.20). □

Remark 6.4.2. *By* (2.11)*, we have* $\|u^\epsilon\|_{H^3} = O(\epsilon^{-1})$*, and we can bound* $\|\Delta^2 u^\epsilon\|_{L^2}$ *as follows:*

$$\|\Delta^2 u^\epsilon\|_{L^2}^2 \le C\epsilon^{-2}\big(\|f\|_{L^2}^2 + \|\det(D^2 u^\epsilon)\|_{L^2}^2\big) \le C\epsilon^{-2}\big(\|f\|_{L^2}^2 + \|D^2 u^\epsilon\|_{L^\infty}^{2n}\big) \le C\epsilon^{-2n-2}.$$

Thus,

$$\begin{aligned}\|I_h u^\epsilon - T_M(I_h u^\epsilon)\|_{2,h} &\le C_2\big(\epsilon^{\frac{3}{2}(1-n)}h\|u^\epsilon\|_{H^3} + h^2\|\Delta^2 u^\epsilon\|_{L^2}\big)\\ &\le C_3\big(\epsilon^{\frac{1}{2}(1-3n)}h + \epsilon^{-(n+1)}h^2\big),\end{aligned}$$

with $C_3 = CC_2$.

Lemma 6.4.3. *There exists a constant* C_4 *independent of* h *and* ϵ *such that the following estimate holds for all* $v_h, w_h \in \mathbb{B}_h(\rho)$ $(n = 2, 3)$*:*

$$\begin{aligned}&\|T_M(w_h) - T_M(v_h)\|_{2,h}\\ &\quad\le C_4\epsilon^{-1}\big(\epsilon^{-1} + h^{-\frac{3}{2}}\rho\big)^{n-2}\big(\epsilon^{-1}h + \rho\big)\|w_h - v_h\|_{2,h}.\end{aligned} \tag{6.25}$$

Proof. Let $v_h, w_h \in \mathbb{B}_h(\rho)$, and to ease notation, let $\sigma_h := w_h - v_h$. Using the definition of T_M and the mean value theorem, we have for any $z_h \in V_0^h$,

$$\begin{aligned}&a_h^\epsilon(T_M(w_h) - T_M(v_h), z_h)\\ &\qquad= \sum_{K\in\mathcal{T}_h}\Big\{(\Phi^\epsilon D\sigma_h, Dz_h)_K + (\det(D^2 w_h) - \det(D^2 v_h), z_h)_K\Big\}\\ &\qquad= \sum_{K\in\mathcal{T}_h}\Big\{(\Phi^\epsilon D\sigma_h, Dz_h)_K + (\Psi_K : D^2\sigma_h, z_h)_K\Big\},\end{aligned}$$

where $\Psi_K := \operatorname{cof}(D^2 w_h - \tau_K D^2\sigma_h),\ \tau_K \in [0,1]$.

For fixed $z_h \in V_0^h$, let $z_0 \in H_0^1(\Omega)$ be its linear interpolant defined in Lemma 6.2.3. Using Lemma A.0.1 and (2.11), we integrate by parts to obtain

$$\begin{aligned}&a_h^\epsilon(T_M(w_h) - T_M(v_h), z_h)\\ &\quad= \sum_{K\in\mathcal{T}_h}\Big\{(\Phi^\epsilon D\sigma_h, Dz_0)_K + (\Psi_K : D^2\sigma_h, z_h)_K + (\Phi^\epsilon D\sigma_h, D(z_h - z_0))_K\Big\}\\ &\quad= \sum_{K\in\mathcal{T}_h}\Big\{((\Psi_K - \Phi^\epsilon) : D^2\sigma_h, z_h)_K + (\Phi^\epsilon : D^2\sigma_h, z_h - z_0)_K\\ &\qquad\quad + (\Phi^\epsilon D\sigma_h, D(z_h - z_0))_K + \sum_{F\in\mathcal{E}_h^i(K)}\big\langle\Phi^\epsilon D\sigma_h\cdot\eta_F, z_0\big\rangle_F\Big\}\\ &\quad\le C\Big(\sum_{K\in\mathcal{T}_h}\|\Phi^\epsilon - \Psi_K\|_{L^2(K)} + \epsilon^{-1}h\Big)\|\sigma_h\|_{2,h}\|z_h\|_{2,h}\\ &\qquad\quad + \Big|\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)}\big\langle\Phi^\epsilon D\sigma_h\cdot\eta_F, z_0\big\rangle_F\Big|.\end{aligned} \tag{6.26}$$

For the case $n = 2$, we have

$$\begin{aligned}\|\Phi^\epsilon - \Psi_K\|_{L^2(K)} &= \|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2w_h - \tau_K D^2\sigma_h)\|_{L^2(K)} \\ &= \|D^2u^\epsilon - (D^2w_h - \tau_K D^2\sigma_h)\|_{L^2(K)}.\end{aligned}$$

Thus,

$$\sum_{K\in\mathcal{T}_h} \|\Phi^\epsilon - \Psi_K\|_{L^2(K)} \le C\big(h\|u^\epsilon\|_{H^3} + \rho\big) \le C\big(\epsilon^{-1}h + \rho\big). \tag{6.27}$$

For the case $n = 3$, let $D^2u^\epsilon\big|_{ij}$ denote the resulting 2×2 matrix after deleting the i^{th} row and j^{th} column of D^2u^ϵ. Then for $i, j = 1, 2, 3$, $K \in \mathcal{T}_h$, we use the mean value theorem to obtain

$$\begin{aligned}\|(\Phi^\epsilon - \Psi_K)_{ij}\|_{L^2(K)} &= \|\det(D^2u^\epsilon\big|_{ij}) - \det(D^2w_h\big|_{ij} - \tau_K D^2\sigma_h\big|_{ij})\|_{L^2(K)} \\ &= \|\Lambda_K^{ij} : (D^2u^\epsilon\big|_{ij} - (D^2w_h\big|_{ij} - \tau_K D^2\sigma_h))\|_{L^2(K)},\end{aligned}$$

where $\Lambda_K^{ij} = \mathrm{cof}(D^2u^\epsilon\big|_{ij} + \lambda_K^{ij}(D^2w_h\big|_{ij} - \tau_K D^2\sigma_h\big|_{ij}))$, $\lambda_K^{ij} \in [0,1]$. Since $\Lambda_K^{ij} \in \mathbf{R}^{2\times 2}$, we have

$$\sum_{K\in\mathcal{T}_h}\sum_{i,j=1}^{n} \|\Lambda_K^{ij}\|_{L^\infty(K)} \le C\big(\epsilon^{-1} + h^{-\frac{3}{2}}\rho\big),$$

where we have used the triangle inequality, inverse inequality, and (2.11). Thus,

$$\begin{aligned}\sum_{K\in\mathcal{T}_h} \|\Phi^\epsilon - \Psi_K\|_{L^2(K)} &\le C\big(\epsilon^{-1} + h^{-\frac{3}{2}}\rho\big)\big(h\|u^\epsilon\|_{H^3} + \rho\big) \\ &\le C\big(\epsilon^{-1} + h^{-\frac{3}{2}}\rho\big)\big(\epsilon^{-1}h + \rho\big).\end{aligned} \tag{6.28}$$

To bound the last term in (6.26), we denote $P_F(Dv_h) \in \mathbf{R}^n$, $P_F(\Phi^\epsilon) \in \mathbf{R}^{n\times n}$ such that

$$\begin{aligned}\Big(P_F(Dv_h)\Big)_k &= P_F\Big(\frac{\partial v_h}{\partial x_k}\Big) \qquad k = 1, ..., n, \\ \Big(P_F(\Phi^\epsilon)\Big)_{k\ell} &= P_F\big(\Phi^\epsilon_{k\ell}\big) \qquad k, \ell = 1, ..., n.\end{aligned}$$

Using Lemma 6.2.1, we make the following identity:

$$\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)}\big\langle \Phi^\epsilon D\sigma_h\cdot\eta_F, z_0\big\rangle_F$$
$$=\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)}\Big\{\big\langle(\Phi^\epsilon-P_F(\Phi^\epsilon))(D\sigma_h-P_F(D\sigma_h))\cdot\eta_F,z_0\big\rangle_F$$
$$+\big\langle P_F(\Phi^\epsilon)(D\sigma_h-P_F(D\sigma_h)\cdot\eta_F, z_0-P_F(z_0)\big\rangle_F\Big\}.$$

Thus,

$$\Big|\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)}\big\langle\Phi^\epsilon D\sigma_h\cdot\eta_F,z_0\big\rangle_F\Big|\le Ch\|\Phi^\epsilon\|_{H^1}\|\sigma_h\|_{2,h}\|z_h\|_{2,h}. \tag{6.29}$$

In the case $n=2$, we have

$$\|\Phi^\epsilon\|_{H^1}=\|u^\epsilon\|_{H^3}\le C\epsilon^{-1}, \tag{6.30}$$

and for the case $n=3$,

$$\|\Phi^\epsilon\|_{H^1}\le C\|D^2u^\epsilon\|_{L^\infty}\|u^\epsilon\|_{H^3}\le C\epsilon^{-2}. \tag{6.31}$$

Finally, using (6.26)–(6.31) and the coercivity of $a_h^\epsilon(\cdot,\cdot)$, we have

$$\|T(v_h)-T(w_h)\|_{2,h}\le C\epsilon^{-1}\Big\{(\epsilon^{-1}+h^{-\frac{3}{2}}\rho)^{n-2}(\epsilon^{-1}h+\rho)+h\epsilon^{1-n}\Big\}\|\sigma_h\|_{2,h}$$
$$\le C\epsilon^{-1}\big(\epsilon^{-1}+h^{-\frac{3}{2}}\rho\big)^{n-2}\big(\epsilon^{-1}h+\rho\big)\|\sigma_h\|_{2,h}.$$

□

Theorem 6.4.4. *There exists an $h_1>0$ such that for $h\le h_1$, there exists a unique solution to* (6.4)*. Furthermore, we have the following estimate* ($n=2,3$)*:*

$$\|u^\epsilon-u_h^\epsilon\|_{2,h}\le C_5\big(\epsilon^{\frac{3}{2}(1-n)}h\|u^\epsilon\|_{H^3}+h^2\|\Delta^2u^\epsilon\|_{L^2}\big). \tag{6.32}$$

Proof. Let $h_1=\min\Big\{\frac{\epsilon^{\frac{7}{2}}}{12C_3C_4},\frac{\epsilon^2}{2\sqrt{C_3C_4}}\Big\}$ when $n=2$, $h_1=\Big(\frac{\epsilon^9}{20C_3^2C_4}\Big)^2$ when $n=3$, and set $\rho_0=2C_2\big(\epsilon^{\frac{3}{2}(1-n)}h\|u^\epsilon\|_{H^3}+h^2\|\Delta^2u^\epsilon\|_{L^2}\big)$.

Then for $h\le h_1$, $v_h\in\mathbb{B}_h(\rho_0)$, we use Lemma 6.4.3, and Remark 6.4.2 to obtain

$$
\begin{aligned}
&\|T_M(I_h u^\epsilon) - T_M(v_h)\|_{2,h} \\
&\quad \le C_4\epsilon^{-1}(\epsilon^{-1} + h^{-\frac{3}{2}}\rho_0)^{n-2}(\epsilon^{-1}h + \rho_0)\|I_h u^\epsilon - v_h\|_{2,h} \\
&\quad \le C_4\epsilon^{-1}\big(\epsilon^{-1} + 4C_3\epsilon^{-4}h^{-\frac{1}{2}}\big)^{n-2} \\
&\qquad\quad \times \big(\epsilon^{-1}h + 2C_3(\epsilon^{\frac{1}{2}(1-3n)}h + \epsilon^{-(n+1)}h^2)\big)\|I_h u^\epsilon - v_h\|_{2,h} \\
&\quad \le C_3C_4\epsilon^{-1}\big(5C_3\epsilon^{-4}h^{-\frac{1}{2}}\big)^{n-2}\big(3\epsilon^{\frac{1}{2}(1-3n)}h + \epsilon^{-(n+1)}h^2\big)\|I_h u^\epsilon - v_h\|_{2,h} \\
&\quad \le \frac{1}{2}\|I_h u^\epsilon - v_h\|_{2,h}.
\end{aligned}
$$

Hence, using Lemma 6.4.1,

$$
\begin{aligned}
\|I_h u^\epsilon - T_M(v_h)\|_{2,h} &\le \|I_h u^\epsilon - T_M(I_h u^\epsilon)\|_{2,h} + \|T_M(I_h u^\epsilon) - T_M(v_h)\|_{2,h} \\
&\le C_2\big(\epsilon^{\frac{3}{2}(1-n)}h\|u^\epsilon\|_{H^3} + h^2\|\Delta^2 u^\epsilon\|_{L^2}\big) + \frac{1}{2}\|I_h u^\epsilon - v_h\|_{2,h} \\
&\le \frac{\rho_0}{2} + \frac{\rho_0}{2} = \rho_0.
\end{aligned}
$$

Thus, T_M maps $\mathbb{B}_h(\rho_0)$ into $\mathbb{B}_h(\rho_0)$, and therefore, T_M has a unique fixed point in $\mathbb{B}_h(\rho_0)$ which is a solution to (6.4). To derive the error estimate, we use the triangle inequality to get

$$
\begin{aligned}
\|u^\epsilon - u_h^\epsilon\|_{2,h} &\le \|u^\epsilon - I_h u^\epsilon\|_{2,h} + \|I_h u^\epsilon - u_h^\epsilon\|_{2,h} \\
&\le Ch\|u^\epsilon\|_{H^3} + \rho_0 \\
&\le C\big(\epsilon^{\frac{3}{2}(1-n)}h\|u^\epsilon\|_{H^3} + h^2\|\Delta^2 u^\epsilon\|_{L^2}\big).
\end{aligned}
$$

The proof is complete □

Theorem 6.4.5. *Under the same hypotheses of Theorem 6.4.4, there exists constants $C_6(\epsilon) > 0,\ C_7(\epsilon) > 0$ such that the following estimate hold in the case $n = 2$:*

$$
\begin{aligned}
\|u^\epsilon - u_h^\epsilon\|_{1,h} \le C_6(\epsilon)\big(\epsilon^{-\frac{3}{2}}h^2\|u^\epsilon\|_{H^3} + h^3\|\Delta^2 u^\epsilon\|_{L^2}\big) \\
+ C_7(\epsilon)\big(\epsilon^{-\frac{3}{2}}h\|u^\epsilon\|_{H^3} + h^2\|\Delta^2 u^\epsilon\|_{L^2}\big)^2.
\end{aligned} \tag{6.33}
$$

Proof. We break the proof into five separate steps.
Step 1. The error equation:

Let $e^\epsilon := u^\epsilon - u_h^\epsilon$, and note for any $v_h \in V_0^h,\ w \in H_0^1(\Omega)$ we have

$$\begin{aligned}(f, v_h) &= -\epsilon(\Delta^2 u^\epsilon, v_h) + (\det(D^2 u^\epsilon), v_h) \\ &= \epsilon(\Delta^2 u^\epsilon, w) + (\det(D^2 u^\epsilon), v_h) - \epsilon(\Delta^2 u^\epsilon, v_h - w) \\ &= \sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D^2 u^\epsilon, D^2 v_h)_K + (\det(D^2 u^\epsilon), v_h)_K \\ &\qquad + \epsilon \sum_{F\in\mathcal{E}_h(K)} \Big(\Big\langle \Delta u^\epsilon + \frac{\partial^2 u^\epsilon}{\partial \tau_F^2}, \frac{\partial v_h}{\partial \eta_F} \Big\rangle_F - \Big\langle \frac{\partial^2 u^\epsilon}{\partial \eta_F \partial \tau_F}, \frac{\partial v_h}{\partial \tau_F} \Big\rangle_F \Big) \Big\} \\ &\qquad + F^\epsilon(u^\epsilon, v_h, w).\end{aligned}$$

Thus, for any $w \in H_0^1(\Omega)$, there holds the following error equation:

$$\sum_{K\in\mathcal{T}_h} \big\{ -\epsilon(D^2 e^\epsilon, D^2 v_h)_K + (\det(D^2 u^\epsilon) - \det(D^2 u_h^\epsilon), v_h)_K \big\} \\ + G^\epsilon(u^\epsilon, v_h) + F^\epsilon(u^\epsilon, v_h, w) = 0 \qquad \forall v_h \in V_0^h.$$

Thus, using the mean value theorem, we have for all $v_h \in V_0^h$, $w \in H_0^1(\Omega)$

$$\sum_{K\in\mathcal{T}_h} \big\{ -\epsilon(D^2 e^\epsilon, D^2 v_h)_K + (\Upsilon_K^\epsilon : D^2 e^\epsilon, v_h)_K \big\} + G^\epsilon(u^\epsilon, v_h) + F^\epsilon(u^\epsilon, v_h, w) = 0, \quad (6.34)$$

where $\Upsilon_K^\epsilon = \mathrm{cof}(D^2 u^\epsilon - \tau_K e^\epsilon)$ $\tau_K \in [0, 1]$.

Step 2. A duality argument:

Next, denote $e_h^\epsilon := I_h e^\epsilon = I_h u^\epsilon - u_h^\epsilon \in V_0^h$ to be the interpolant of e^ϵ into V_0^h, and let $e_0^\epsilon \in H_0^1(\Omega)$ be the linear interpolant of e_h^ϵ as defined in Lemma 6.2.3.

Let $v \in H^3(\Omega) \cap H_0^1(\Omega)$ be the solution to the following problem:

$$\begin{aligned} L_{u^\epsilon}(v) &= -\Delta e_0^\epsilon \qquad && \text{in } \Omega, \\ v &= 0 && \text{on } \partial\Omega, \\ \Delta v &= 0 && \text{on } \partial\Omega. \end{aligned}$$

Assuming $\partial\Omega$ is smooth, it follows from standard elliptic regularity theory that there exists such a v, and furthermore (cf. Theorem 3.2.2)

$$\|v\|_{H^3} \le C\epsilon^{-2} \|\Delta e_0^\epsilon\|_{H^{-1}} \le C\epsilon^{-2} \|D e_0^\epsilon\|_{L^2}. \quad (6.35)$$

Recalling $\alpha^\epsilon = u^\epsilon - I_h u^\epsilon$, we use Lemma A.0.1 and integrate by parts to obtain

$$\begin{aligned}\|De_0^\epsilon\|_{L^2}^2 &= \epsilon(\Delta^2 v, e_0^\epsilon) - (\Phi^\epsilon : D^2 v, e_0^\epsilon) = -\epsilon(D(\Delta v), De_0^\epsilon) + (\Phi^\epsilon Dv, De_0^\epsilon)\\ &= \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2 v, D^2 e_h^\epsilon)_K + (\Phi^\epsilon Dv, De_h^\epsilon)\\ &\qquad + (\Phi^\epsilon Dv, D(e_0^\epsilon - e_h^\epsilon))_K - \epsilon(D(\Delta v), D(e_0^\epsilon - e_h^\epsilon))_K\Big\} - G^\epsilon(v, e_h^\epsilon)\end{aligned}$$

$$\begin{aligned}&= \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2 v, D^2 e_h^\epsilon)_K + (\Phi^\epsilon Dv, De^\epsilon)_K - (\Phi^\epsilon Dv, D\alpha^\epsilon)_K \\ &\qquad + (\Phi^\epsilon Dv, D(e_0^\epsilon - e_h^\epsilon))_K - \epsilon(D(\Delta v), D(e_0^\epsilon - e_h^\epsilon))_K\Big\} - G^\epsilon(v, e_h^\epsilon).\end{aligned} \tag{6.36}$$

Step 3: Bounding the last four terms in (6.36)*:*

By (6.3),(6.7), and a Sobolev inequality, we bound the third, fourth, and fifth term in (6.36) as follows:

$$\begin{aligned}&\Big|\sum_{K\in\mathcal{T}_h}\Big\{-(\Phi^\epsilon Dv, D\alpha^\epsilon)_K + (\Phi^\epsilon Dv, D(e_0^\epsilon - e_h^\epsilon))_K \\ &\qquad - \epsilon(D(\Delta v), D(e_0^\epsilon - e_h^\epsilon))_K\Big\}\Big| \\ &\le C\|\Phi^\epsilon\|_{L^2}\|Dv\|_{L^\infty}\big(\|\alpha^\epsilon\|_{1,h} + \|e_0^\epsilon - e_h^\epsilon\|_{1,h}\big) + C\epsilon\|e_0^\epsilon - e_h^\epsilon\|_{1,h}\|v\|_{H^3} \\ &\le C\epsilon^{-\frac{1}{2}}\big(h\|e^\epsilon\|_{2,h} + h^2\|u^\epsilon\|_{H^3}\big)\|v\|_{H^3}.\end{aligned} \tag{6.37}$$

Using (6.18), we also have

$$\big|G^\epsilon(v, e_h^\epsilon)\big| \le C\epsilon h\|e^\epsilon\|_{2,h}\|v\|_{H^3}. \tag{6.38}$$

Step 4: Bounding the first two terms in (6.36)

To bound the first two terms in the last line of (6.36), we write

$$\begin{aligned}&\sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2 v, D^2 e_h^\epsilon)_K + (\Phi^\epsilon De^\epsilon, Dv)_K\Big\} \\ &= \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2 v, D^2 e^\epsilon)_K + (\Phi^\epsilon De^\epsilon, Dv)_K - \epsilon(D^2\alpha^\epsilon, D^2 v)_K\Big\} \\ &= \tilde{a}_h^\epsilon(e^\epsilon, v) - \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2\alpha^\epsilon, D^2 v)_K - \sum_{F\in\mathcal{E}_h^i(K)}\Big\langle\Phi^\epsilon De^\epsilon\cdot\eta_F, v\Big\rangle_F\Big\} \\ &= \tilde{a}_h^\epsilon(e^\epsilon, I_h v) + \tilde{a}_h^\epsilon(e^\epsilon, v - I_h v) \\ &\qquad - \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2\alpha^\epsilon, D^2 v)_K - \sum_{F\in\mathcal{E}_h^i(K)}\Big\langle\Phi^\epsilon De^\epsilon\cdot\eta_F, v\Big\rangle_F\Big\},\end{aligned} \tag{6.39}$$

where

$$\tilde{a}_h^\epsilon(e^\epsilon, v) := \sum_{K\in\mathcal{T}_h} \big\{\epsilon(D^2 e^\epsilon, v)_K - (\Phi^\epsilon : D^2 e^\epsilon, v)_K\big\}.$$

To bound the fourth term in (6.39), we have

$$\begin{aligned}
&\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)} \big\langle \Phi^\epsilon De^\epsilon\cdot\eta_F, v\big\rangle_F \\
&= \sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)} \Big\{\langle(\Phi^\epsilon - P_F(\Phi^\epsilon))(De^\epsilon - P_F(De^\epsilon))\cdot\eta_F, v\big\rangle_F \\
&\qquad + \big\langle P_F(\Phi^\epsilon)(De^\epsilon - P_F(De^\epsilon))\cdot\eta_F, v - P_F(v)\big\rangle_F\Big\}.
\end{aligned}$$

Thus,

$$\Big|\sum_{K\in\mathcal{T}_h}\sum_{F\in\mathcal{E}_h^i(K)} \big\langle \Phi^\epsilon De^\epsilon\cdot\eta_F, v\big\rangle_F\Big| \le Ch\|\Phi^\epsilon\|_{H^1}\|e^\epsilon\|_{2,h}\|v\|_{1,h}. \tag{6.40}$$

To bound the third term in (6.39), we use the identity

$$-\epsilon\sum_{K\in\mathcal{T}_h}(D^2 v, D^2\alpha^\epsilon)_K = \epsilon\sum_{K\in\mathcal{T}_h}(D(\Delta v), D\alpha^\epsilon)_K - G^\epsilon(v,\alpha^\epsilon).$$

Thus,

$$\epsilon\Big|\sum_{K\in\mathcal{T}_h}(D^2 v, D^2\alpha^\epsilon)_K\Big| \le C\epsilon h^2\|u^\epsilon\|_{H^3}\|v\|_{H^3}. \tag{6.41}$$

Bounding the second term in (6.39), we have

$$\big|\tilde{a}_h^\epsilon(e^\epsilon, v - I_h v)\big| \le C\epsilon^{-\frac{1}{2}}h\|e^\epsilon\|_{2,h}\|v\|_{H^3}. \tag{6.42}$$

To bound $\tilde{a}_h^\epsilon(e^\epsilon, I_h v)$, we use (6.34) to conclude

$$\begin{aligned}
\tilde{a}_h^\epsilon(e^\epsilon, I_h v) &= \sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2 e^\epsilon, D^2(I_h v))_K - (\Phi^\epsilon : D^2 e^\epsilon, I_h v)_K\Big\} \\
&= \sum_{K\in\mathcal{T}_h}((\Upsilon_K^\epsilon - \Phi^\epsilon) : D^2 e^\epsilon, I_h v)_K + G^\epsilon(u^\epsilon, I_h v) + F^\epsilon(u^\epsilon, I_h v, v).
\end{aligned}$$

In the case $n = 2$, we have for each $K\in\mathcal{T}_h$,

$$\|\Upsilon_K^\epsilon - \Phi^\epsilon\|_{L^2(K)} = \tau_K\|D^2 e^\epsilon\|_{L^2(K)} \le \|D^2 e^\epsilon\|_{L^2(K)}.$$

Thus,

$$\Big|\sum_{K\in\mathcal{T}_h}((\Upsilon_K^\epsilon-\Phi^\epsilon):D^2e^\epsilon,I_hv)_K\Big|\le C\|e^\epsilon\|_{2,h}^2\|v\|_{H^3}. \tag{6.43}$$

Next, since $u^\epsilon\in H^3(\Omega),\ v\in H^2(\Omega)$, we have the following identity:

$$G^\epsilon(u^\epsilon,I_hv)=G^\epsilon(u^\epsilon,I_hv-v).$$

Therefore, by (6.22) and (6.18),

$$\begin{aligned}&\big|F^\epsilon(u^\epsilon,I_hv,v)+G^\epsilon(u^\epsilon,I_hv)\big| \\ &\le C\epsilon\Big(\big(\|v-I_hv\|_{1,h}+h\|v-I_hv\|_{2,h}\big)\|u^\epsilon\|_{H^3}+\|v-I_hv\|_{L^2}\|\Delta^2u^\epsilon\|_{L^2}\Big)\\ &\le C\epsilon\big(h^2\|u^\epsilon\|_{H^3}+h^3\|\Delta^2u^\epsilon\|_{L^2}\big)\|v\|_{H^3}.\end{aligned} \tag{6.44}$$

Thus, combining (6.39)–(6.44), we bound the first two terms in the last line of (6.36) as follows:

$$\begin{aligned}&\Big|\sum_{K\in\mathcal{T}_h}\Big\{\epsilon(D^2v,D^2e_h^\epsilon)_K+(\Phi^\epsilon De^\epsilon,Dv)_K\Big\}\Big| \\ &\le C\Big((\|\Phi^\epsilon\|_{H^1}+\epsilon^{-\frac12})h\|e^\epsilon\|_{2,h}+\|e^\epsilon\|_{2,h}^2+\epsilon\big(h^2\|u^\epsilon\|_{H^3}+h^3\|\Delta^2u^\epsilon\|_{L^2}\big)\Big)\|v\|_{H^3}.\end{aligned} \tag{6.45}$$

Step 5: Combining Steps 2–4:

Combining (6.36)–(6.38), (6.45), and (6.35) we obtain

$$\begin{aligned}\|De_0^\epsilon\|_{L^2}^2&\le C\Big\{\big((\epsilon^{-\frac12}+\|\Phi^\epsilon\|_{H^1})h+\|e^\epsilon\|_{2,h}\big)\|e^\epsilon\|_{2,h}\\&\qquad+\epsilon\big(h^2|u^\epsilon|_{H^3}+h^3\|\Delta^2u^\epsilon\|_{L^2}\big)\Big\}\|v\|_{H^3}\\&\le C\epsilon^{-2}\Big\{\big((\epsilon^{-\frac12}+\|\Phi^\epsilon\|_{H^1})h+\|e^\epsilon\|_{2,h}\big)\|e^\epsilon\|_{2,h}\\&\qquad+\epsilon\big(h^2|u^\epsilon|_{H^3}+h^3\|\Delta^2u^\epsilon\|_{L^2}\big)\Big\}\|De_0^\epsilon\|_{L^2}.\end{aligned}$$

Dividing by $\|De_0^\epsilon\|_{L^2}$, using Theorem 6.4.4, and applying Poincaré's inequality, we have

$$\begin{aligned}\|e_0^\epsilon\|_{H^1}\le C\epsilon^{-2}\Big\{&\big(\epsilon^{-\frac12}+\|\Phi^\epsilon\|_{H^1}\big)C_5\big(\epsilon^{\frac32(1-n)}h^2\|u^\epsilon\|_{H^3}+h^3\|\Delta^2u^\epsilon\|_{L^2}\big)\\&+C_5^2\big(\epsilon^{\frac32(1-n)}h\|u^\epsilon\|_{H^3}+h^2\|\Delta^2u^\epsilon\|_{L^2}\big)^2\Big\}.\end{aligned}$$

Thus, using the inequality

$$\|e^\epsilon\|_{1,h} \leq C\|e_h^\epsilon\|_{1,h} \leq C\|e_0^\epsilon\|_{H^1},$$

we obtain (6.33) with $C_6(\epsilon) = CC_5\epsilon^{-2}(\epsilon^{-\frac{1}{2}} + \|\Phi^\epsilon\|_{H^1}) = O(\epsilon^{-3})$ and $C_7(\epsilon) = CC_5^2\epsilon^{-2} = O(\epsilon^{-2})$.

□

Remark 6.4.6. *The reason we strict ourselves to the case $n = 2$ in Theorem 6.4.5 is that we are currently unable to estimate the term $\|\Upsilon_K^\epsilon - \Phi^\epsilon\|_{L^2(K)} = \|\mathrm{cof}(D^2u^\epsilon - \tau_K e^\epsilon) - \mathrm{cof}(D^2u^\epsilon)\|_{L^2(K)}$ in the three dimensional case. Doing so would require optimal error estimates of $\|D^2u^\epsilon - D^2u_h^\epsilon\|_{L^\infty(K)}$ or $\|D^2u^\epsilon - D^2u^\epsilon\|_{L^4(K)}$ $\forall K \in \mathcal{T}_h$.*

6.5 Numerical Experiments and Rates of Convergence

In this section, we provide several 2-D experiments to gauge the efficiency of the finite element method developed in the previous sections. We also compare the results with the tests in Chapters 3 and 5, where (2.8)–(2.10) was approximated by C^1 and quadratic mixed finite element methods, respectively. All of the tests given below are computed on the domain $\Omega = (0,1)^2$.

We emphasize the considerable advantage of using the Morley finite element to approximate (2.8)–(2.10), as the resulting algebraic system is much smaller than any C^1-conforming finite element method or mixed finite element method. Table 6.1 lists the resulting number of unknowns after discretizing (2.8)–(2.10) using the finite element methods presented in Chapters 3 and 5 and the finite element method developed in this chapter.

As we can see from the table, the use of mixed finite element methods to approximate (2.8)–(2.10) results in roughly four times more unknowns than that of using Morley elements. Also, by Table 6.1, we can expect to have at least 2.5 times more unknowns using any C^1-conforming finite element (e.g. Argyris) compared to the Morley element.

Table 6.1: Approximate number of DOF's on domain $\Omega = (0,1)^2$ using the Argyris element, quadratic mixed finite elements, and the Morley element.

h	Argyris	Mixed Method	Morley
0.25	351	425	129
0.1	1947	2978	801
0.05	7489	12351	3201
0.025	29372	50299	12801
0.01	181420	317741	80001
0.005	722834	1275479	320001

Test 6.1:

In this test, we calculate $\|u - u_h^\epsilon\|$ for fixed $h = 0.0277$, while varying ϵ in order to approximate the error $\|u - u^\epsilon\|$. We set to solve problem (6.4) with the following data:

$$\begin{aligned}
&\text{(a) } u = e^{\frac{x_1^2+x_2^2}{2}}, && f = (1 + x_1^2 + x_2^2)e^{x_1^2+x_2^2}, && g = e^{\frac{x_1^2+x_2^2}{2}}.\\
&\text{(b) } u = x_1^2 + x_2^2, && f = 4, && g = x_1^2 + x_2^2.
\end{aligned}$$

After computing the error, we plot to results in Figures 6.2 and 6.3 to estimate the rate of convergence in ϵ for each norm. Figure 6.2 clearly shows $\|u - u_h^\epsilon\|_{L^2}$ and $\|u - u_h^\epsilon\|_{L^\infty}$ converge linearly in ϵ, where as Figure 6.3 shows $|u-u_h^\epsilon|_{1,h} = O(\epsilon^{\frac{3}{4}})$ and $|u-u_h^\epsilon|_{2,h} = O(\epsilon^{\frac{1}{4}})$. Since we have fixed h small, we would expect $\|u-u^\epsilon\|_{L^\infty} \approx O(\epsilon)$, $\|u-u^\epsilon\|_{L^2} \approx O(\epsilon)$, $\|u-u^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$, and $\|u - u^\epsilon\|_{H^2} \approx O(\epsilon^{\frac{1}{4}})$. We note that these are the same rates of convergence found in both Chapters 3 and 5 (cf. Tests 3.1 and 5.1).

Test 6.2:

The purpose of this test is to calculate the rate of convergence of $\|u^\epsilon - u_h^\epsilon\|$ for fixed $\epsilon = 0.01$ in various norms. As in Test 6.1, we solve problem (6.4), but with the boundary condition $\Delta u^\epsilon\big|_{\partial\Omega} = \epsilon$ replaced by $\Delta u^\epsilon\big|_{\partial\Omega} = \phi^\epsilon$. We use the following test functions and data:

$$\begin{aligned}
&\text{(a) } u^\epsilon = e^{\frac{x_1^2+x_2^2}{2}}, && f^\epsilon = (1 + x_1^2 + x_2^2)e^{x_1^2+x_2^2}\\
&&& \quad - \epsilon(8 + 8(x_1^2 + x_2^2) + 2x_1^2x_2^2 + x_1^4 + x_2^4)e^{\frac{x_1^2+x_2^2}{2}},\\
&\quad g^\epsilon = e^{\frac{x_1^2+x_2^2}{2}}, && \phi^\epsilon = (2 + x_1^2 + x_2^2)e^{\frac{x_1^2+x_2^2}{2}}.\\
&\text{(b) } u^\epsilon = \frac{1}{12}\big(x_1^4 + x_2^4\big), && f^\epsilon = x_1^2x_2^2 - 4\epsilon,\\
&\quad g^\epsilon = \frac{1}{12}\big(x_1^4 + x_2^4\big), && \phi^\epsilon = x_1^2 + x_2^2.
\end{aligned}$$

After calculating the error, we divide each norm by the power of h expected to be the convergence rate by the analysis of the previous section. As seen by Table 6.2, we have $\|u^\epsilon - u_h^\epsilon\|_{2,h} = O(h)$ and $\|u^\epsilon - u_h^\epsilon\|_{1,h} = O(h^2)$ as expected. The tests also indicate that $\|u^\epsilon - u_h^\epsilon\|_{L^\infty} = O(h^2)$ although a theoretical proof has yet to be developed.

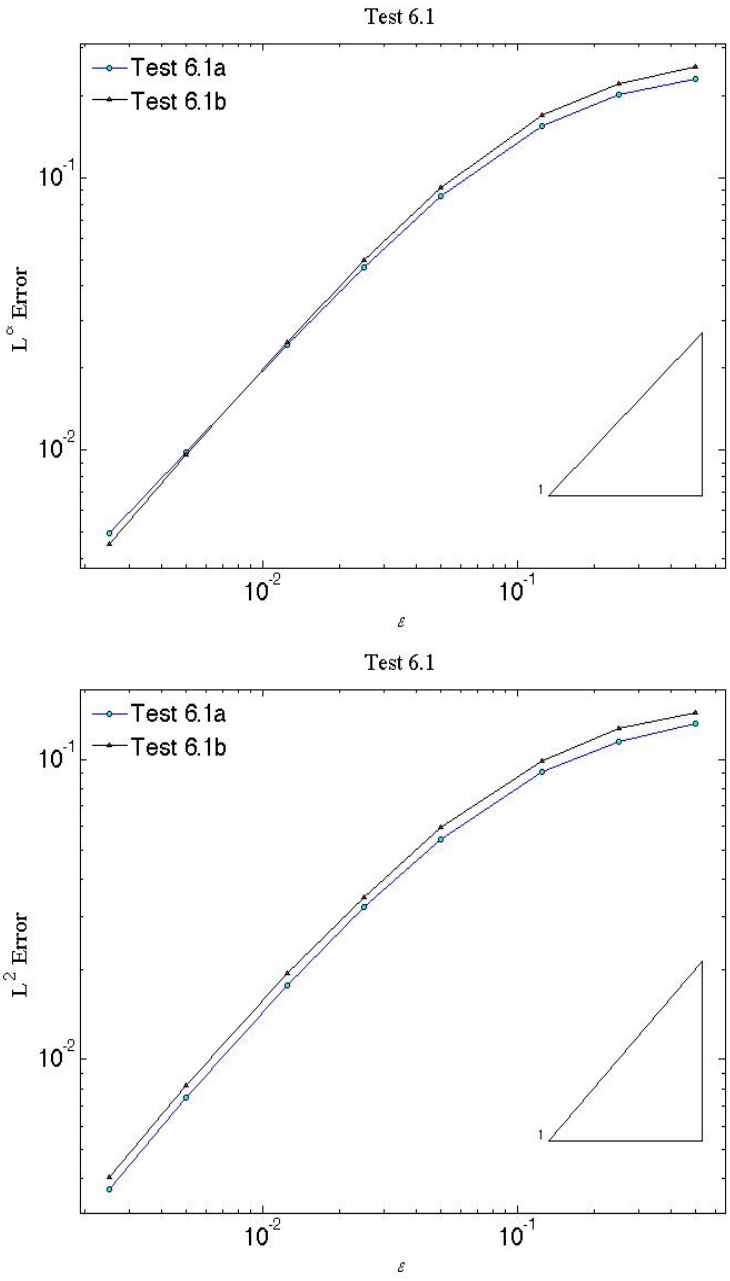


Figure 6.2: Test 6.1: L^∞ errors (top) and L^2 errors (bottom) w.r.t. ϵ ($h = 0.0277$).

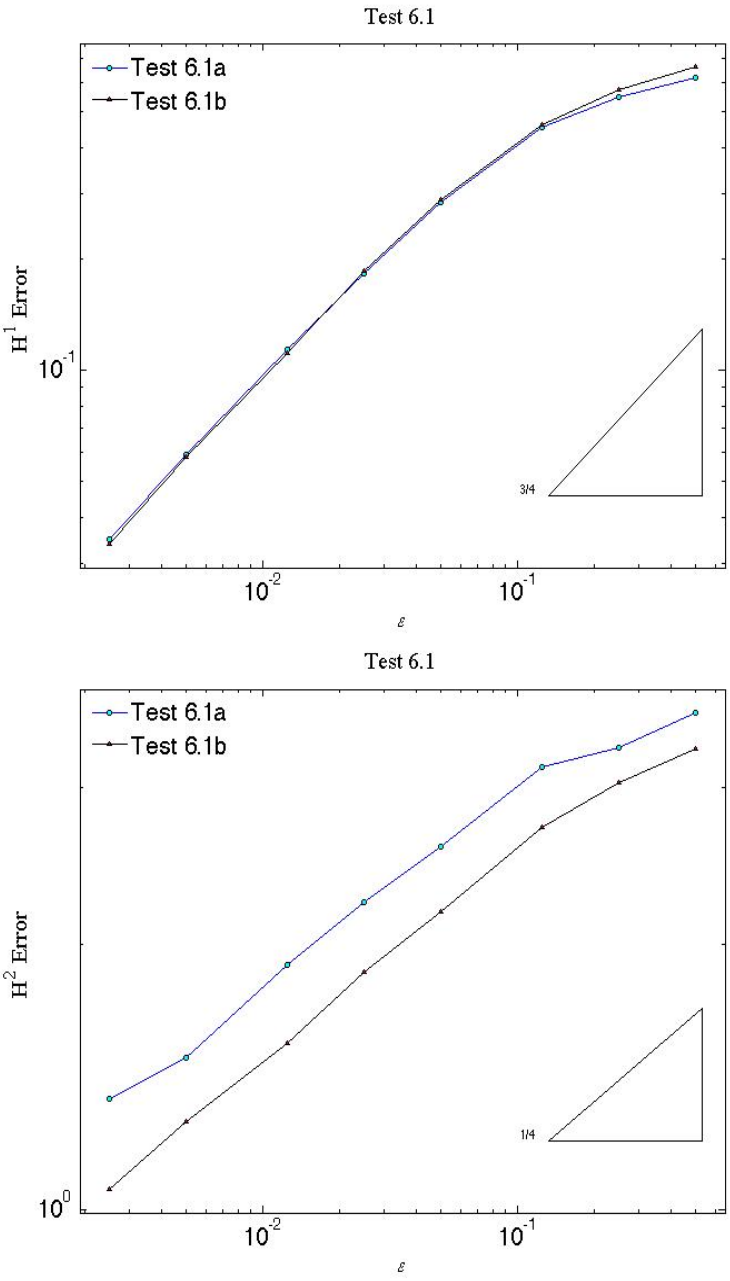


Figure 6.3: Test 6.1: H^1 errors (top) and H^2 errors (bottom) w.r.t. ϵ ($h = 0.0277$).

Table 6.2: Test 6.2: Change of $\|u^\epsilon - u_h^\epsilon\|$ w.r.t. h ($\epsilon = 0.01$).

	h	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^\infty}}{h^2}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^2}}{h^2}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{1,h}}{h^2}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{2,h}}{h}$
Test 6.2a	0.2357	0.109941429	0.046356469	0.481550276	1.339301103
	0.1286	0.121488987	0.057427881	0.565240816	1.497037325
	0.0884	0.125419729	0.062162169	0.608255615	1.602056335
	0.0673	0.124684001	0.066041256	0.639762523	1.696395691
	0.0544	0.125689338	0.066075097	0.634235105	1.721571324
	0.0456	0.12058518	0.067607341	0.645972799	1.664270614
	0.0393	0.124267558	0.068792935	0.662004934	1.765003053
	0.0345	0.12357908	0.067011132	0.646318	1.740322609
	0.0277	0.121948677	0.067393033	0.657561026	1.761516606
Test 6.2b	0.2357	0.034175275	0.015383995	0.186182611	0.473255325
	0.1286	0.031703427	0.014626955	0.199801547	0.515604588
	0.0884	0.028603018	0.013479966	0.201015796	0.526897059
	0.0673	0.028139068	0.01345244	0.205462666	0.526926152
	0.0544	0.025931282	0.013110943	0.198485483	0.535665257
	0.0456	0.025002886	0.012984765	0.198695753	0.542492982
	0.0393	0.024538845	0.012852139	0.202843657	0.537316285
	0.0345	0.025692081	0.012972065	0.200772947	0.544215362
	0.0277	0.023954437	0.013189277	0.204616247	0.543368231

Test 6.3:

This test is exactly the same as Test 6.1, but we now use the following data:

$$f = 1, \qquad g = 0.$$

We note that in this case, there exists a unique convex viscosity solution, but there does not exist a classical solution (cf. [39], [60]). Figure 6.4 displays the computed solution using $\epsilon = 0.005$, $h = 0.0393$, and it clearly shows that the vanishing moment method approximation correctly captures the convex viscosity solution.

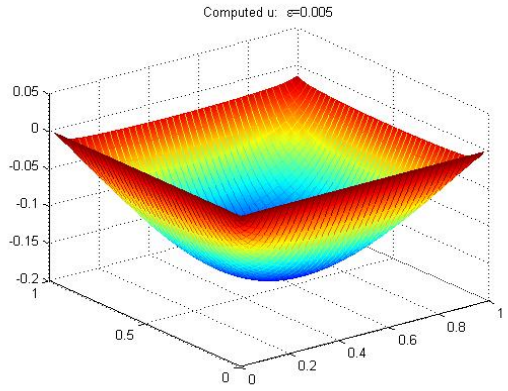


Figure 6.4: Test 6.3: Computed solution. $\epsilon = 0.005$, $h = 0.0393$.

Chapter 7

Finite Element Methods for the Nonlinear Balance Equation

The nonlinear balance equation is a Monge-Ampère type equation that models two dimensional geostrophic wind [91], which is the wind resulting from the exact balance between the Coriolis force and the pressure gradient force. Although the true wind almost always differs from the geostrophic wind due to friction and centrifugal forces, geostrophic flow can be a valuable first approximation. The goal of this chapter is to apply the methodology of the vanishing moment method to the nonlinear balance equation, and then analyze its finite element approximation.

The chapter is organized as follows. In Section 7.1 we derive the nonlinear balance equation, starting with the geostrophic balance and the f-plane momentum equations. In Section 7.2, we provide the theoretical background and PDE analysis of the nonlinear balance equation. We find that if an ellipticity condition is satisfied, then under a suitable change of variables, the nonlinear balance equation can be written as an elliptic Monge-Ampère equation. Making use of this observation, we directly apply the analysis of Chapters 3–6 to approximate the nonlinear balance equation in Section 7.3. Finally in Section 7.4, we provide numerical examples showing the effectiveness of the methods developed in the previous sections.

7.1 Derivation of the Nonlinear Balance Equation

To derive the nonlinear balance equation, we follow the presentation in [91]. Let $\Omega \subset \mathbf{R}^2$ be an open bounded, convex domain and set $\mathbf{u} := (u_1\ u_2)$ to be the horizontal wind, where $\mathbf{u}$ is composed of the u_1 velocity in the east-west (x_1) direction and the u_2 velocity in the north-south (x_2) direction. Let f be the Coriolis parameter (assumed to be constant) so that $(f\mathbf{u})$ is the Coriolis force, and let p be the pressure. To start, we state the *geostrophic*

balance, which describes the balance between the pressure gradient force and the Coriolis force in the horizontal directions:

$$Dp = f\mathbf{u}^{\perp}, \tag{7.1}$$

where $\mathbf{u}^{\perp} := (u_2, -u_1)$. Taking the divergence of (7.1), we obtain

$$\Delta p = f\mathrm{div}(\mathbf{u}^{\perp}) = f\left(\frac{\partial u_2}{\partial x_1} - \frac{\partial u_1}{\partial x_2}\right). \tag{7.2}$$

While equation (7.2) is useful for pointwise estimation, it does not include any type of centrifugal force and is therefore only used to approximate straight flows. Furthermore, the dynamics of the fluids are missing in the description. A more accurate representation is given by the f-plane momentum equations, which is a version of the Boussinesq equations:

$$\frac{D\mathbf{u}}{Dt} + Dp = f\mathbf{u}^{\perp} \qquad \Omega \times (0,T], \tag{7.3}$$

$$\mathrm{div}\,\mathbf{u} = 0 \qquad \Omega \times (0,T], \tag{7.4}$$

where

$$\frac{D}{Dt} := \frac{\partial}{\partial t} + \mathbf{u}\cdot D$$

denotes the material derivative.

Remark 7.1.1. *Equation* (7.3) *can be written as*

$$\begin{aligned}
&\frac{\partial u_1}{\partial t} + u_1\frac{\partial u_1}{\partial x_1} + u_2\frac{\partial u_1}{\partial x_2} + \frac{\partial p}{\partial x_1} = fu_2,\\
&\frac{\partial u_2}{\partial t} + u_1\frac{\partial u_2}{\partial x_1} + u_2\frac{\partial u_2}{\partial x_2} + \frac{\partial p}{\partial x_2} = -fu_1.
\end{aligned}$$

Taking the divergence of (7.3) and using (7.4), we obtain

$$\begin{aligned}
0 &= \frac{\partial}{\partial t}(\mathrm{div}\,\mathbf{u}) + \frac{\partial}{\partial x_1}\left(u_1\frac{\partial u_1}{\partial x_1} + u_2\frac{\partial u_1}{\partial x_2} - fu_2\right) + \frac{\partial}{\partial x_2}\left(u_1\frac{\partial u_2}{\partial x_1} + u_2\frac{\partial u_2}{\partial x_2} + fu_1\right) + \Delta p\\
&= \left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + 2\frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} + f\left(\frac{\partial u_1}{\partial x_2} - \frac{\partial u_2}{\partial x_1}\right)\\
&\quad + u_1\left(\frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_1 \partial x_2}\right) + u_2\left(\frac{\partial^2 u_1}{\partial x_1 \partial x_2} + \frac{\partial^2 u_2}{\partial x_2^2}\right) + \Delta p\\
&= \left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + 2\frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} + f\left(\frac{\partial u_1}{\partial x_2} - \frac{\partial u_2}{\partial x_1}\right)\\
&\quad + u_1\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2}\right) + u_2\frac{\partial}{\partial x_2}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2}\right) + \Delta p
\end{aligned}$$

$$= \left(\frac{\partial u_1}{\partial x_1}\right)^2 + \left(\frac{\partial u_2}{\partial x_2}\right)^2 + 2\frac{\partial u_1}{\partial x_2}\frac{\partial u_2}{\partial x_1} + f\left(\frac{\partial u_1}{\partial x_2} - \frac{\partial u_2}{\partial x_1}\right) + \Delta p. \tag{7.5}$$

Next, since the flow is two dimensional, we can write the velocity $\mathbf{u}$ in terms of its stream function, ψ

$$u_1 := -\frac{\partial \psi}{\partial x_2}, \quad u_2 := \frac{\partial \psi}{\partial x_1}. \tag{7.6}$$

Using identity (7.6) in (7.5) results in the *nonlinear balance equation*:

$$-2\left(\frac{\partial^2 \psi}{\partial x_1 \partial x_2}\right)^2 + 2\frac{\partial^2 \psi}{\partial x_2^2}\frac{\partial^2 \psi}{\partial x_1^2} + f\left(\frac{\partial^2 \psi}{\partial x_2^2} + \frac{\partial^2 \psi}{\partial x_1^2}\right) - \Delta p = 0,$$

that is,

$$\det(D^2\psi) + \frac{f}{2}\Delta\psi - \frac{1}{2}\Delta p = 0. \tag{7.7}$$

7.2 Theoretical Results

In this section, we show that many properties and results of the Monge-Ampère equation can be migrated to the nonlinear balance equation.

Before we proceed, we assume the following *ellipticity condition* holds:

$$\Delta p + \frac{1}{2}f^2 > 0. \tag{7.8}$$

The reasons to assume this condition are twofold; one is analytical, while the other has physical significance, as we now explain.

Let ψ solve (7.7), and set $\varphi = \psi + \frac{f}{4}(x_1^2 + x_2^2)$. We then have

$$\begin{aligned} \frac{1}{2}\Delta p + \frac{f^2}{4} &= \det(D^2\psi) + \frac{f}{2}\Delta\psi + \frac{f^2}{4} \\ &= \det(D^2(\varphi - \frac{f}{4}(x_1^2 + x_2^2))) + \frac{f}{2}\Delta(\varphi - \frac{f}{4}(x_1^2 + x_2^2)) + \frac{f^2}{4} \\ &= \left(\frac{\partial^2 \varphi}{\partial x_1^2} - \frac{f}{2}\right)\left(\frac{\partial^2 \varphi}{\partial x_2^2} - \frac{f}{2}\right) - \left(\frac{\partial^2 \varphi}{\partial x_1 \partial x_2}\right)^2 + \frac{f}{2}\left(\frac{\partial^2 \varphi}{\partial x_1^2} + \frac{\partial^2 \varphi}{\partial x_2^2}\right) - \frac{f^2}{4} \\ &= \det(D^2\varphi). \end{aligned} \tag{7.9}$$

Thus, we have converted the nonlinear balance equation for ψ into the Monge-Ampère equation for φ. Also, by the ellipticity condition (7.8), the left-hand side of (7.9) is positive, and therefore there exists a unique convex viscosity solution. We now show that a viscosity solution of (7.9) corresponds to a viscosity solution of (7.7).

For given $p \in H^1(\Omega)$ and f with $\Delta p + \frac{1}{2}f^2 > 0$, let $\varphi \in C^0(\overline{\Omega})$ be a convex viscosity

solution of (7.9), and set $\psi = \varphi - \frac{f}{4}(x_1^2+x_2^2) \in C^0(\overline{\Omega})$. Suppose $\psi - \phi$ has a local maximum at $x^0 \in \Omega$ for some $\phi \in C^2(\Omega)$. Set $\xi = \phi + \frac{f}{4}(x_1^2+x_2^2) \in C^2(\Omega)$ and note that $\varphi - \xi = \psi - \phi$, and thus $\varphi - \xi$ has a local maximum at x^0. From the definition of viscosity solutions (cf. Definition 1.3.1), we have

$$\frac{1}{2}\Delta p(x_0) + \frac{f^2}{2} \leq \det(D^2\xi(x_0)) = \det(D^2\phi(x_0)) + \frac{f}{2}\Delta\phi(x_0) + \frac{f^2}{4}.$$

Similarly, if $\psi - \phi$ has a local minimum at x^0, then

$$\frac{1}{2}\Delta p(x_0) \geq \det(D^2\phi(x^0)) + \frac{f}{2}\Delta\phi(x_0).$$

It follows that ψ is a viscosity solution of (7.7) such that $\psi + \frac{f}{4}(x_1^2+x_2^2)$ is convex. We note that ψ is not necessarily a convex function. Also, we recall that there exist exactly two solutions to the Monge-Ampère equation in two dimensions; one being convex, the other concave. It immediately follows that there exists a unique solution ψ to (7.7) such that $\psi + \frac{f}{4}(x_1^2+x_2^2)$ is concave.

The physical importance of the ellipticity condition (7.8) can be seen by noting that (7.7) can be rewritten as follows:

$$\left(\frac{f}{2} + \frac{\partial^2\psi}{\partial x_1^2}\right)\left(\frac{f}{2} + \frac{\partial^2\psi}{\partial x_2^2}\right) = \frac{1}{2}\Delta p + \left(\frac{\partial^2\psi}{\partial x_1 \partial x_2}\right)^2 + \frac{f^2}{4} \geq \frac{1}{2}\Delta p + \frac{f^2}{4}. \tag{7.10}$$

Thus, $\left(\frac{f}{2} + \frac{\partial^2\psi}{\partial x_1^2}\right)$ and $\left(\frac{f}{2} + \frac{\partial^2\psi}{\partial x_2^2}\right)$ have the same sign by (7.8). Hence, the *absolute vorticity*, $(f + \nabla \times \mathbf{u}) = (f + \Delta\psi)$, will either be positive or negative in the whole domain. The solution with positive absolute vorticity (i.e. the solution such that $\psi + \frac{f}{4}(x_1^2+x_2^2)$ is convex) corresponds to the solution in the Northern Hemisphere, where as the solution that has negative absolute vorticity (i.e. the solution such that $\psi + \frac{f}{4}(x_1^2+x_2^2)$ is concave) corresponds to the solution in the Southern Hemisphere [91].

7.2.1 Vanishing Moment Approximation

Let ψ be the viscosity solution of (7.7) with prescribed Dirichlet boundary condition $\psi\big|_{\partial\Omega} = g$ such that $\psi + \frac{f}{4}(x_1^2+x_2^2)$ is convex. That is, ψ satisfies (in the viscosity sense)

$$\det(D^2\psi) + \frac{1}{2}\Delta\psi = \frac{1}{2}\Delta p \qquad \text{in } \Omega, \tag{7.11}$$

$$\psi = g \qquad \text{on } \partial\Omega. \tag{7.12}$$

Let φ be the corresponding viscosity solution to (7.9) with prescribed Dirichlet boundary conditions (in the viscosity sense) $\varphi\big|_{\partial\Omega} = g + \frac{f}{4}(x_1^2 + x_2^2)$, that is

$$\det(D^2\varphi) = \frac{1}{2}\Delta p + \frac{f^2}{4} \qquad \text{in } \Omega, \tag{7.13}$$

$$\varphi = g + \frac{f}{4}(x_1^2 + x_2^2) \qquad \text{on } \partial\Omega. \tag{7.14}$$

Employing the vanishing moment methodology developed in Chapter 2, we approximate φ by φ^ϵ, where φ^ϵ solves

$$-\epsilon\Delta^2\varphi^\epsilon + \det(D^2\varphi^\epsilon) = \frac{1}{2}\Delta p + \frac{f^2}{4} \qquad \text{in } \Omega, \tag{7.15}$$

$$\varphi^\epsilon = g + \frac{f}{4}(x_1^2 + x_2^2) \qquad \text{on } \partial\Omega, \tag{7.16}$$

$$\Delta\varphi^\epsilon = \epsilon \qquad \text{on } \partial\Omega. \tag{7.17}$$

Applying the PDE results of Chapter 2 to (7.15)–(7.17), making the substitution

$$\psi^\epsilon = \varphi^\epsilon - \frac{f}{4}(x_1^2 + x_2^2), \tag{7.18}$$

and noting

$$\Delta^2\left(\frac{f}{4}(x_1^2 + x_2^2)\right) = 0, \qquad \Delta\psi^\epsilon = \Delta\varphi^\epsilon - f,$$

we have the following result.

Theorem 7.2.1. *Suppose $p \in H^1(\Omega)$, the ellipticity condition (7.8) holds, and ψ is the unique viscosity solution to (7.11)–(7.12) such that $\psi + \frac{f}{4}(x_1^2 + x_2^2)$ is convex. Then for every $\epsilon > 0$, there exists a unique solution to the following problem:*

$$-\epsilon\Delta^2\psi^\epsilon + \det(D^2\psi^\epsilon) + \frac{f}{2}\Delta\psi^\epsilon = \frac{1}{2}\Delta p \qquad \textit{in } \Omega, \tag{7.19}$$

$$\psi^\epsilon = g \qquad \textit{on } \partial\Omega, \tag{7.20}$$

$$\Delta\psi^\epsilon = \epsilon - f \qquad \textit{on } \partial\Omega. \tag{7.21}$$

Furthermore,

$$\psi^\epsilon + \frac{f}{4}(x_1^2 + x_2^2) \textit{ is convex for each } \epsilon > 0,$$
$$\psi^\epsilon \to \psi \textit{ unifomly as } \epsilon \to 0^+,$$
$$\psi^\epsilon \rightharpoonup \psi \textit{ in } H^1(\Omega) \textit{ as } \epsilon \to 0^+.$$

Remark 7.2.2. *Theorem 7.2.1 provides a direct way to approximate the viscosity solution of* (7.11)–(7.12) *via the vanishing moment approximation* (7.19)–(7.21). *We could then develop several finite element and spectral Galerkin methods based on this approximation to construct convergent schemes for the nonlinear balance equation. However, a simpler approach in both analysis and practice is to compute* (7.15)–(7.17) *using the numerical methods analyzed in Chapters 3–6 and then make the substitution* (7.18). *This is the path we take.*

7.3 Finite Element Formulations and Analysis

In this section we apply the results of Chapters 3–6 to construct convergent numerical methods for problem (7.15)–(7.17). By making the substitution (7.18), we obtain approximated solutions for the nonlinear balance equation.

In what follows, we let $\mathcal{T}_h$ be a quasiuniform triangular (or rectangular in the case of C^1 finite element or mixed finite element methods) mesh of Ω with mesh size $h \in (0,1)$.

7.3.1 C^1 Finite Element Methods

In this section, we construct and analyze C^1 finite element methods to approximate the solution to (7.15)–(7.17) which in turn approximates the viscosity solution to the nonlinear balance equation (7.11)–(7.12) via the substitution (7.18).

To derive the finite element formulation of (7.15)–(7.17), we let V^h and V_0^h be the C^1-conforming finite element spaces of degree r (≥ 5) defined in Chapter 3. Define

$$\tilde{V}_g^h = \{v_h \in V^h,\ v_h\big|_{\partial\Omega} = g + \frac{f}{4}(x_1^2 + x_2^2)\}, \qquad V_g^h = \{v_h \in V^h,\ v_h\big|_{\partial\Omega} = g\}.$$

We define the finite element method for (7.15)–(7.17) as seeking $\varphi_h^\epsilon \in \tilde{V}_g^h$ such that

$$\begin{aligned} -\epsilon(\Delta\varphi_h^\epsilon, \Delta v_h) + (\det(D^2\varphi_h^\epsilon), v_h) & \\ = \frac{1}{4}\big(f^2, v_h\big) - \frac{1}{2}(Dp, Dv_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta} \right\rangle_{\partial\Omega} & \qquad \forall v_h \in V_0^h. \end{aligned} \tag{7.22}$$

Setting $\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2) \in V_g^h$ and applying the analysis of Chapter 3 to (7.22) (cf. Theorems 3.3.4, 3.3.5, and 3.3.7) give us the following results.

Theorem 7.3.1. *Suppose $p \in H^1(\Omega)$, the ellipticity condition holds, and suppose $\varphi^\epsilon \in H^s(\Omega)$ ($s \geq 3$) is the unique solution to* (7.15)–(7.17) *and ψ^ϵ is the unique solution to* (7.19)–(7.21). *Furthermore, assume that the linearized problem,* (3.7)–(3.9) *is H^4 regular. Then there exists an $h_0 > 0$ such that for $h \leq h_0$, there exists a unique $\varphi_h^\epsilon \in \tilde{V}_g^h$ solving*

(7.22). *Furthermore, by setting* $\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2)$, *we obtain the following error estimates:*

$$\|\psi^\epsilon - \psi_h^\epsilon\|_{H^2} \le C\epsilon^{-\frac{3}{2}} h^{\ell-2} \|\varphi^\epsilon\|_{H^\ell}, \tag{7.23}$$

$$\|\psi^\epsilon - \psi_h^\epsilon\|_{H^1} \le C\epsilon^{-4} h^{\ell-1} \|\varphi^\epsilon\|_{H^\ell}, \tag{7.24}$$

$$\|\psi^\epsilon - \psi_h^\epsilon\|_{L^2} \le C\epsilon^{-5} \big(h^\ell \|\varphi^\epsilon\|_{H^\ell} + \epsilon^{-1} h^{2\ell-4} \|\varphi^\epsilon\|_{H^\ell}^2\big), \tag{7.25}$$

where $\ell = \min\{r+1, s\}$.

Remark 7.3.2. *A similar bound holds for spectral Galerkin methods by using the relation* $h = \frac{1}{N}$ *and setting* $\ell = \min\{N+1, s\}$, *where* N *denotes the polynomial degree in the spectral element space (cf. Chapter 4).*

7.3.2 Mixed Finite Element Methods

We treat the mixed finite element method analysis similarly. Let V_0^h, V_g^h, W^h, W_0^h be the Lagrange finite element spaces of degree k (≥ 2) defined in Chapter 5, and let

$$\tilde{V}_g^h := \{v \in V^h;\ v\big|_{\partial\Omega} = g + \frac{f}{4}(x_1^2 + x_2^2)\}.$$

We define the mixed finite element formulation for (7.15)–(7.17) as finding $(\kappa_h^\epsilon, \varphi_h^\epsilon) \in W_\epsilon^h \times \tilde{V}_g^h$ such that

$$(\kappa_h^\epsilon, \mu_h) + (\mathrm{div}(\mu_h), D\varphi_h^\epsilon) = \big\langle \frac{\partial}{\partial\tau}\big(g + \frac{f}{4}(x_1^2 + x_2^2)\big), \mu\eta\cdot\tau\big\rangle_{\partial\Omega}, \tag{7.26}$$

$$(\mathrm{div}(\kappa_h^\epsilon), Dv_h) + \epsilon^{-1}(\det(\kappa_h^\epsilon), v_h) = \frac{1}{4\epsilon}(f^2, v_h) - \frac{1}{2\epsilon}(Dp, Dv_h). \tag{7.27}$$

Applying the analysis of Chapter 5 to (7.26)–(7.27), and making the substitutions $\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2)$ and $\sigma_h^\epsilon = \kappa_h^\epsilon - I_{2\times 2}\frac{f}{2}$, we have the following result (cf. Theorems 5.3.4 and 5.3.6).

Theorem 7.3.3. *Suppose* $p \in H^1(\Omega)$ *and the ellipticity condition holds. Let* $\varphi^\epsilon \in H^{s+2}(\Omega)$ *be the unique solution to* (7.15)–(7.17) *and let* ψ^ϵ *be the unique solution to* (7.19)–(7.21). *Set* $\kappa^\epsilon = D^2\varphi^\epsilon$ *and* $\sigma^\epsilon = D^2\psi^\epsilon$. *Then there exists an* $h_1 > 0$ *such that for* $h \le h_1$, *there exists a unique solution,* $(\kappa_h^\epsilon, \psi_h^\epsilon) \in W_\epsilon^h \times \tilde{V}_g^h$, *solving* (7.26)–(7.27). *Furthermore, setting* $\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2)$, $\sigma_h^\epsilon = \kappa_h^\epsilon - I_{2\times 2}\frac{f}{2}$, *we have the following estimates:*

$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{L^2} \le C\epsilon^{-\frac{9}{2}} h^{\ell-2} \big(\|\kappa^\epsilon\|_{H^\ell} + \|\varphi^\epsilon\|_{H^\ell}\big), \tag{7.28}$$

$$\|\sigma^\epsilon - \sigma_h^\epsilon\|_{H^1} \le C\epsilon^{-\frac{9}{2}} h^{\ell-3} \big(\|\kappa^\epsilon\|_{H^\ell} + \|\varphi^\epsilon\|_{H^\ell}\big), \tag{7.29}$$

$$\|\psi^\epsilon - \psi_h^\epsilon\|_{H^1} \le C\epsilon^{-7} (h^{\ell-1}\big(\|\kappa^\epsilon\|_{H^\ell} + \|\varphi^\epsilon\|_{H^\ell}\big) + \epsilon^{-\frac{1}{2}} h^{2\ell-4}\big(\|\kappa^\epsilon\|_{H^\ell} + \|\varphi^\epsilon\|_{H^\ell}\big)^2), \tag{7.30}$$

where $\ell = \min\{k+1, s\}$.

7.3.3 A Nonconforming Morley Finite Element Method

We end this section with a nonconforming Morley finite element method for problem (7.15)–(7.17). Let V^h and V_0^h be the finite element spaces corresponding to the Morley element defined in Chapter 6, and let

$$\tilde{V}_g^h = \{v_h \in V^h,\ v_h\big|_{\partial\Omega} = g + \frac{f}{4}(x_1^2 + x_2^2)\}, \qquad V_g^h = \{v_h \in V^h,\ v_h\big|_{\partial\Omega} = g\}.$$

We define the nonconforming finite element method for (7.15)–(7.17) as seeking $\varphi_h^\epsilon \in \tilde{V}_g^h$ such that for all $v_h \in V_0^h$

$$\begin{aligned} &\sum_{K\in\mathcal{T}_h} \Big\{ -\epsilon(D^2\varphi_h^\epsilon, D^2 v_h)_K + (\det(D^2\varphi_h^\epsilon), v_h)_K \Big\} \qquad (7.31)\\ &= \frac{1}{4}\big(f^2, v_h\big) - \frac{1}{2}\sum_{K\in\mathcal{T}_h}(Dp, Dv_h)_K - \sum_{F\in\mathcal{E}_h^b}\left\langle \frac{\partial^2}{\partial^2\tau}\big(g + \frac{f}{4}(x_1^2+x_2^2)\big) + \epsilon, \frac{\partial v_h}{\partial \eta_F}\right\rangle_F. \end{aligned}$$

Making the substitution $\psi_h^\epsilon = \varphi_h^\epsilon + \frac{f}{4}(x_1^2 + x_2^2)$ and applying Theorems 6.4.4 and 6.4.5 gives us the following results.

Theorem 7.3.4. *Suppose* $p \in H^1(\Omega)$*, the ellipticity condition holds, and* $\varphi^\epsilon \in H^s(\Omega)$ *is the unique solution to* (7.15)–(7.17)*. Furthermore, assume that the linearized problem,* (3.7)-(3.9) *is* H^3 *regular. Then there exists an* $h_2 > 0$ *such that for* $h \le h_2$*, there exists a unique* $\varphi_h^\epsilon \in \tilde{V}_g^h$ *solving* (7.31)*. Furthermore, by setting* $\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2)$*, we have the following error estimates:*

$$\|\psi^\epsilon - \psi_h^\epsilon\|_{2,h} \le C\big(\epsilon^{-\frac{3}{2}} h\|\varphi^\epsilon\|_{H^3} + h^2\|\Delta^2\varphi^\epsilon\|_{L^2}\big) \qquad (7.32)$$

$$\begin{aligned} \|\psi^\epsilon - \psi_h^\epsilon\|_{1,h} \le{}& C\epsilon^{-3}\big(\epsilon^{-\frac{3}{2}}h^2\|\varphi^\epsilon\|_{H^3} + h^3\|\Delta^2\varphi^\epsilon\|_{L^2}\big) \qquad (7.33)\\ &+ \epsilon^{-2}\big(\epsilon^{-\frac{3}{2}}h\|\varphi^\epsilon\|_{H^3} + h^2\|\Delta^2\varphi^\epsilon\|_{L^2}\big)^2 \end{aligned}$$

7.4 Numerical Experiments and Rates of Convergence

In this section we show the efficiency and accuracy of the moment method to approximate the nonlinear balance equation using mixed finite element methods developed in the previous section.

Test 7.1

In this test we fix $h = 0.013$ and vary ϵ to approximate $\|\psi - \psi^\epsilon\|$ in various norms. We first compute $(\kappa^\epsilon, \varphi_h^\epsilon)$ using the mixed method formulation (7.26)–(7.27) with Lagrange quadratic elements and then make the substitution

$$\psi_h^\epsilon = \varphi_h^\epsilon - \frac{f}{4}(x_1^2 + x_2^2), \qquad \sigma_h^\epsilon = \kappa_h^\epsilon - I_{n\times n}\frac{f}{2}.$$

We set $\Omega = (0,1)\times(0,1)$ and for simplicity, set $f = 2$ and $F = \frac{1}{2}\Delta p$. We note that the ellipticity condition is now $F + 1 > 0$, and that ψ satisfies

$$\begin{aligned} \det(D^2\psi) + \Delta\psi &= F \qquad \text{in } \Omega, \\ \psi &= g \qquad \text{on } \partial\Omega. \end{aligned}$$

We use the following test functions and parameters.

$$\begin{aligned}
&\text{(a) } \psi = \frac{1}{4}(x_1^2 + x_2^2)^2, && F = 3(x_1^4 + x_2^4) + 6x_1^2x_2^2 + 4(x_1^2 + x_2^2), \\
&\text{(b) } \psi = \frac{1}{12}x_1^4 + \frac{1}{6}x_2^3 - \frac{1}{2}x_2^2, && F = x_2(x_1^2 + 1) - 1, \\
&\text{(c) } \psi = \frac{1}{12}x_1^4 + \frac{1}{6}x_2^3 - x_2^2, && F = (x_1^2 + 1)(x_2 - 1) - 1, \\
&\text{(d) } \psi = \frac{1}{12}x_1^4 + \frac{1}{6}x_2^3 - \frac{3}{4}x_2^2, && F = (x_1^2 + 1)(x_2 - \frac{1}{2}) - 1,
\end{aligned}$$

We note that the first two test functions satisfy the ellipticity condition. The third test function does not satisfy the ellipticity condition anywhere in Ω, where as the fourth test function satisfies the ellipticity condition for $x_2 > \frac{1}{2}$.

After finding the error, we divide by various powers of ϵ to estimate the rate at which each norm converges. As seen in Figure 7.1, when the ellipticity condition holds, $\|\psi - \psi_h^\epsilon\|_{L^2} \approx O(\epsilon)$, $\|\psi - \psi_h^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$, and $\|\sigma - \sigma_h^\epsilon\|_{L^2} = O\big(\epsilon^{\frac{1}{4}}\big)$ as expected (cf. Tests 3.1, 5.1, and 6.1). Since we have fixed h small, we can predict that $\|\psi - \psi^\epsilon\|$ behaves similarly.

However, when the ellipticity condition is violated, convergence is not guaranteed. This is seen in Figure 7.2. The error $\|\psi - \psi_h^\epsilon\|$ diverges as $\epsilon \to 0^+$.

Test 7.2

For this test, we first compute φ_h^ϵ using the finite element formulation (7.22). We then recover an approximation of the velocity field $\mathbf{u}$ by using the following identity:

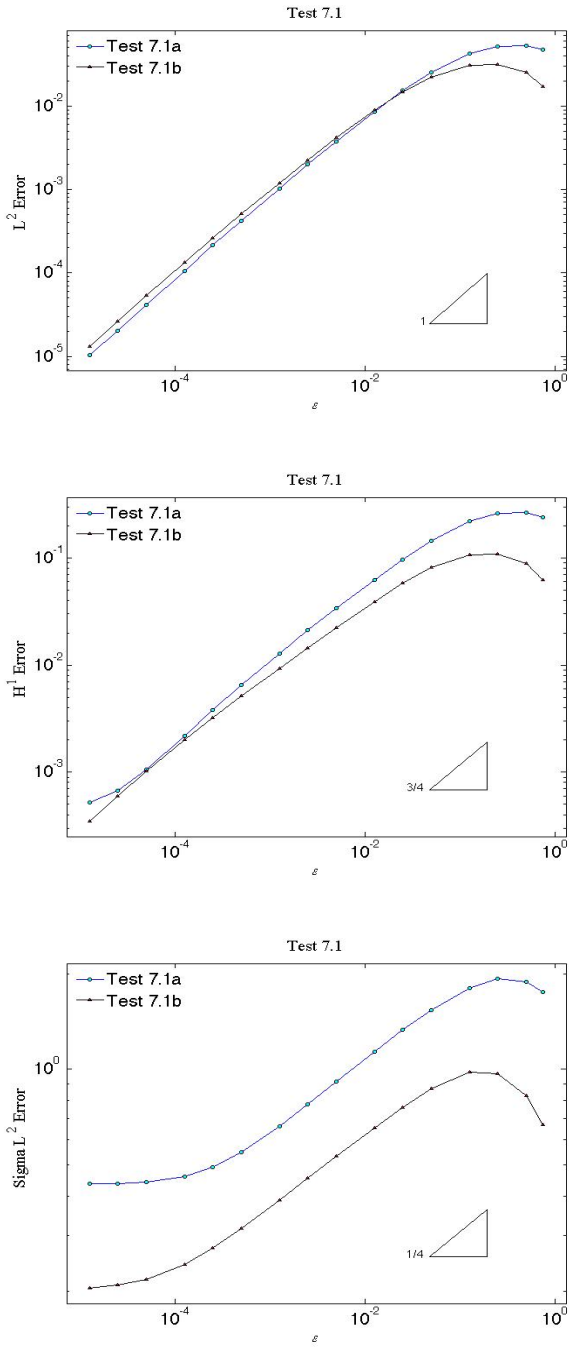


Figure 7.1: Tests 7.1a and 7.1b. Change of $\|\psi - \psi_h^\epsilon\|$ w.r.t. ϵ ($h = 0.017$)

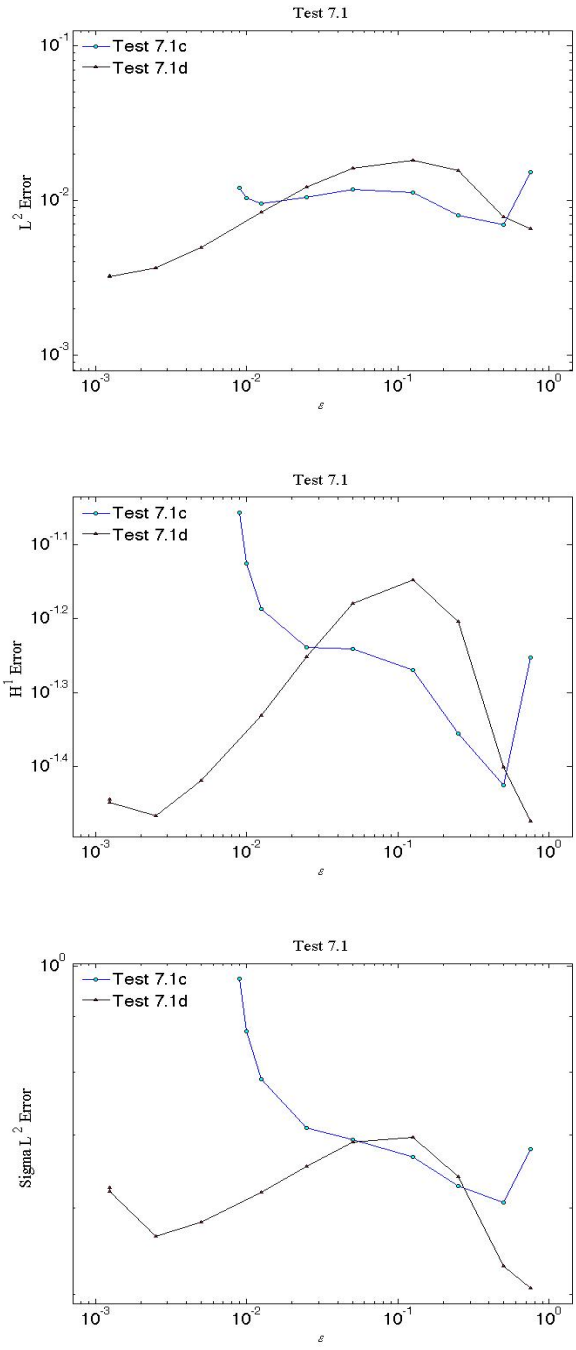


Figure 7.2: Tests 7.1c and 7.1d. Change of $\|\psi - \psi_h^\epsilon\|$ w.r.t. ϵ ($h = 0.017$)

$$\mathbf{u}_h^\epsilon := \left[-\frac{\partial \varphi_h^\epsilon}{\partial x_2} + \frac{f}{2}x_2,\ \frac{\partial \varphi_h^\epsilon}{\partial x_1} - \frac{f}{2}x_1\right] = \left[-\frac{\partial \psi_h^\epsilon}{\partial x_2},\ \frac{\partial \psi_h^\epsilon}{\partial x_1}\right]$$
$$\approx \left[-\frac{\partial \psi}{\partial x_2},\ \frac{\partial \psi}{\partial x_1}\right] = \mathbf{u}.$$

We let $\Omega = (0,1) \times (0,1)$, and use the following function parameters:

$$p = \sin(\pi x)\sin(\pi y), \qquad f = 2\pi, \qquad g = 0.$$

We then plot the computed velocity field in Figure 7.3 with parameters $\epsilon = 0.01$ and $h = 0.05$.

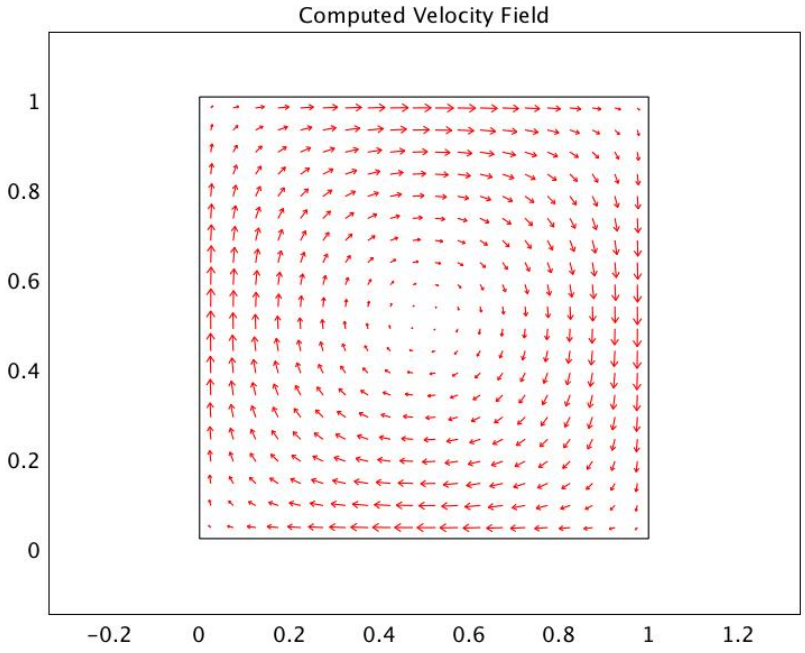


Figure 7.3: Tests 7.2. Computed velocity field with $\epsilon = 0.01$, $h = 0.05$

Chapter 8

Finite Element Methods for the Semigeostrophic Flow Equations

The semi-geostrophic equations, derived by B.J. Hoskins [63], are used in meteorology to model slowly varying flows constrained by rotation and stratification. They can be considered as an approximation of the Euler equations and are thought to be an efficient model to describe front formation ([69, 36]). Under certain assumptions and in some appropriately chosen curve coordinates, they can be formulated as a coupled system consisting of the Monge-Ampère equation and the transport equation. The derivation of the nonlinear system is presented in Section 8.1. The goal of this chapter is to formulate and analyze numerical methods of the nonlinear formulation.

To achieve this goal in Section 8.2, we apply the vanishing moment methodology presented in Chapter 2, and then state certain assumptions about this approximation. In Section 8.3, we formulate a modified characteristic finite element method based upon the vanishing moment approximation. In Section 8.4, we show optimal order error estimates of the proposed finite element method under certain time-stepping and mesh constraints. The main idea of the proof is to use an inductive argument that is based on the results of Section 3.3. Finally, in Section 8.5, we provide numerical tests to validate the analysis in the previous section and reinforce certain assumptions we have made.

8.1 Derivation of the Nonlinear Formulation

To introduce the three dimensional semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem, we suppose a fluid is moving inside a bounded open domain $\Omega \subset \mathbf{R}^3$ satisfying the following incompressible Boussinesq equations which are a

version of the incompressible Euler equations:

$$\frac{D\mathbf{u}}{Dt} + Dp = f\mathbf{u}^\perp - \frac{\rho}{\rho_0} g\mathbf{e}_3 \qquad \text{in } \Omega \times (0,T], \tag{8.1}$$

$$\frac{D\rho}{Dt} = 0 \qquad \text{in } \Omega \times (0,T], \tag{8.2}$$

$$\operatorname{div} \mathbf{u} = 0 \qquad \text{in } \Omega \times (0,T], \tag{8.3}$$

$$\mathbf{u} = \mathbf{0} \qquad \text{on } \partial\Omega \times [0,T], \tag{8.4}$$

where $\mathbf{e}_3 := (0,0,1)$, $\mathbf{u} = (u_1, u_2, u_3)$ is the velocity field, p is the pressure, ρ is the temperature of the atmosphere or the density of the ocean water, and ρ_0 is a reference value of ρ. Also,

$$\frac{D}{Dt} := \frac{\partial}{\partial t} + \mathbf{u} \cdot D$$

denotes the material derivative, and

$$\mathbf{u}^\perp := (u_2, -u_1, 0).$$

Remark 8.1.1. *Omitting the gravitational term in* (8.1)–(8.4)*, the flow becomes two dimensional, and we obtain the f-plane momentum equations* (7.3)–(7.4) *introduced in Chapter 7.*

Ignoring the material derivative in (8.1), we have

$$D_H p = f\mathbf{u}^\perp, \tag{8.5}$$

$$\frac{\partial p}{\partial x_3} = -\frac{\rho}{\rho_0} g, \tag{8.6}$$

where

$$D_H := \Big(\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, 0\Big).$$

Equation (8.5) is the three dimensional *geostrophic balance* (see equation (7.1) for the two dimensional version), and equation (8.6) is known as the *hydrostatic balance*, which describes the balance between the pressure gradient force and the gravitational force in the vertical direction.

Remark 8.1.2. *The geostrophic balance* (8.5) *is equivalent to*

$$\mathbf{u}_H = -f^{-1} D^\perp p,$$

where

$$\mathbf{u}_H := (u_1, u_2, 0), \qquad D^\perp p := (Dp)^\perp$$

The geostrophic and hydrostatic balances give very simple relations between the pressure field and the velocity field. However, the dynamics of the fluids are missing in the description. To overcome this limitation, J. B. Hoskins [63] proposed so-called semigeostrophic approximation which is based on replacing the material derivative of the full velocity $\frac{D\mathbf{u}}{Dt}$ by the material derivative of the *geostrophic wind* $\frac{D\mathbf{u}_g}{Dt}$ in (8.1), where

$$\mathbf{u}_g := -f^{-1}D^{\perp}p. \tag{8.7}$$

This then leads to the following semigeostrophic flow equations (in the primitive variables):

$$\frac{D\mathbf{u}_g}{Dt} + D_H p = f\mathbf{u}^{\perp} \qquad \text{in } \Omega \times (0,T], \tag{8.8}$$

$$\frac{\partial p}{\partial x_3} = -\frac{\rho}{\rho_0}g \qquad \text{in } \Omega \times (0,T], \tag{8.9}$$

$$\frac{D\rho}{Dt} = 0 \qquad \text{in } \Omega \times (0,T], \tag{8.10}$$

$$\operatorname{div}\mathbf{u} = 0 \qquad \text{in } \Omega \times (0,T], \tag{8.11}$$

$$\mathbf{u} = 0 \qquad \text{on } \partial\Omega \times (0,T]. \tag{8.12}$$

Remark 8.1.3. *Using* $(\mathbf{u}^{\perp})^{\perp} = -\mathbf{u}_H$ *and* $(D_H p)^{\perp} = D^{\perp}p$*, equation* (8.8) *can be written as*

$$\frac{D\mathbf{u}_g^{\perp}}{Dt} + D^{\perp}p = -f\mathbf{u}_H. \tag{8.13}$$

There are no explicit dynamic equations for $\mathbf{u}$ in the above semigeostrophic flow model, but rather (8.8) is now an evolution equation for Dp. However, we note that the full velocity $\mathbf{u}$ still appears in the material derivative,

$$\frac{D\mathbf{u}_g}{Dt} = \frac{\partial \mathbf{u}_g}{\partial t} + (\mathbf{u}\cdot D)\mathbf{u}_g.$$

Should $\mathbf{u}\cdot D$ be replaced by $\mathbf{u}_g \cdot D$ in the material derivative, the resulting model is known as *the quasi-geostrophic flow equations* (cf. [71]).

Due to the peculiar structure of the semigeostrophic flow equations, it is difficult to analyze and to numerically solve the equations. The first successful analytical approach is one based on a fully nonlinear reformulation, which was first proposed in [16] and was further developed in [10, 69] (see [34] for a different approach). The main idea of the reformulation is to use time-dependent curved coordinates so the resulting system becomes partially decoupled. The trade-off is the presence of stronger nonlinearity in the new formulation.

The derivation of the fully nonlinear reformulation starts with introducing the *geopotential*

$$\psi := f^{-2}p + \frac{1}{2}|x_H|^2, \tag{8.14}$$

and *geostrophic transformation*

$$\Phi := D\psi. \tag{8.15}$$

Here, $x_H := (x_1, x_2, 0)$. Using the definition of $\mathbf{u}_g$ and equation (8.9), we have

$$\Phi := f^{-2}Dp + x_H = f^{-2}D_H p - \frac{\rho}{\rho_0}g\mathbf{e}_3 + x_H = f^{-1}\mathbf{u}_g^{\perp} - \frac{\rho}{\rho_0}g\mathbf{e}_3 + x_H.$$

Calculating the material derivative of Φ and using (8.8), (8.10), (8.13) we have

$$\begin{aligned}\frac{D\Phi}{Dt} &= f^{-1}\frac{D\mathbf{u}_g^{\perp}}{Dt} + \frac{g\mathbf{e}_3}{\rho_0}\frac{D\rho}{Dt} + \frac{Dx_H}{Dt} \\ &= f^{-1}\frac{D\mathbf{u}_g^{\perp}}{Dt} + \mathbf{u}_H \\ &= -f^{-1}D^{\perp}p. \\ &= f(x-\Phi)^{\perp}.\end{aligned} \tag{8.16}$$

Next, for any $x \in \Omega$, let $X(x,t)$ denote the fluid particle trajectory originating from x, that is,

$$\begin{aligned}\frac{dX(x,t)}{dt} &= \mathbf{u}(X(x,t),t) \qquad \forall\, t > 0, \\ X(x,0) &= x.\end{aligned}$$

Define the composite function

$$\Psi(\cdot,t) := \Phi(X(\cdot,t),t) = D\psi(X(\cdot,t),t). \tag{8.17}$$

We then have from (8.16)

$$\frac{\partial\Psi(x,t)}{\partial t} = f(X(x,t) - \Psi(x,t))^{\perp}. \tag{8.18}$$

Next, the incompressibility assumption (8.10) implies X is volume preserving, and therefore

$$\det(DX) = 1,$$

which is equivalent to

$$\int_\Omega g(X(x,t))dx = \int_\Omega g(x)dx \qquad \forall\, g \in C(\overline{\Omega}). \tag{8.19}$$

To summarize, we have reduced (8.8)–(8.11) into (8.17)–(8.19). It is easy to see that $\Psi(x,t)$ is not unique because one has a freedom in choosing the geopotential ψ. However, Cullen, Norbury, and Purser [35] (also see [36, 10, 69]) discovered the so-called *Cullen-Norbury-Purser principle* which says that $\Psi(x,t)$ must minimize the geostrophic energy at each time t. A consequence of this minimum energy principle is that the geopotential ψ must be a convex function. Using the assumption that ψ is convex and Brenier's polar factorization theorem [16], Brenier and Benamou [10] proved existence of such a convex function ψ and a measure preserving mapping X which solves (8.17)–(8.19).

Continuing, we let $\alpha(y,t)dy$ be the image measure of the Lebesgue measure dx by $\Psi(x,t)$, that is

$$\int_\Omega g(\Psi(x,t))dx = \int_{\mathbf{R}^3} g(y)\alpha(y,t)dy \qquad \forall g \in C_c(\mathbf{R}^3).$$

We note that the image measure $\alpha(y,t)dy$ is the push-forward $\Psi_\# dx$ of dx by $\Psi(x,t)$, and $\alpha(y,t)$ is the density of $\Psi_\# dx$ with respect to the Lebesgue measure dy.

Assuming ψ is sufficiently regular, it follows from (8.17) and (8.19) that

$$\int_\Omega g(\Psi(x,t))dx = \int_\Omega g(D\psi(X(x,t),t))dx = \int_\Omega g(D\psi(x,t))dx \qquad \forall\, g \in C_c(\mathbf{R}^3). \tag{8.20}$$

Using a change of variable $y = D\psi(x,t)$ on the right and the definition of $\alpha(y,t)dy$ on the left we obtain

$$\int_{\mathbf{R}^3} g(y)\alpha(y,t)dy = \int_{\mathbf{R}^3} g(y)d\mu = \int_{\mathbf{R}^3} g(y)\det(D^2\psi^*(y,t))dy \qquad \forall\, g \in C_c(\mathbf{R}^3),$$

where ψ^* denotes the Legendre transform of ψ, that is,

$$\psi^*(y,t) = \sup_{x\in\Omega}\big(x\cdot y - \psi(x,t)\big). \tag{8.21}$$

Hence (α, ψ^*) satisfy the following Monge-Ampère equation:

$$\alpha(y,t) = \det(D^2\psi^*(y,t)).$$

For a convex function ψ and by a property of the Legendre transform we have $D\psi^*(y,t) =$

$x \in \Omega$ for all $y \in \mathbf{R}^n$. Hence

$$D\psi^* \subset \Omega.$$

Finally, for any $w \in C_c^\infty([-1,T];\mathbf{R}^3)$, it follows from integration by parts and (8.18) that

$$\begin{aligned}
-\int_\Omega w(\Psi(x,0),0)\,dx &= \int_0^T \int_\Omega \frac{dw(\Psi(x,t),t)}{dt}\,dxdt \\
&= \int_0^T \int_\Omega \Big\{ Dw(\Psi(x,t),t)\cdot \frac{\partial \Psi(x,t)}{\partial t} + \frac{\partial w(\Psi(x,t),t)}{\partial t} \Big\}\,dxdt \\
&= \int_0^T \int_\Omega \Big\{ Dw(\Psi(x,t),t)\cdot f(X(x,t)-\Psi(x,t))^\perp + \frac{\partial w(\Psi(x,t),t)}{\partial t} \Big\}\,dxdt.
\end{aligned}$$

Making the change of variable $y = D\psi(x,t)$ and using the definition of $\alpha(y,t)dy$ we obtain

$$\int_0^T \int_{\mathbf{R}^3} \Big\{ \frac{\partial w(y,t)}{\partial t} + f\mathbf{v}(y,t)\cdot Dw(y,t) \Big\} \alpha(y,t)\,dydt + \int_{\mathbf{R}^3} w(y,0)\alpha(y,0)\,dy = 0, \quad (8.22)$$

where

$$\mathbf{v} := \big(\frac{\partial \psi^*}{\partial x_2} - x_2, x_1 - \frac{\partial \psi^*}{\partial x_1}, 0 \big) = \big(D\psi^* - x \big)^\perp.$$

Hence,

$$\frac{\partial \alpha(y,t)}{\partial t} + \operatorname{div}(\mathbf{v}(y,t)\alpha(y,t)) = 0,$$

with the assumption $f = 1$.

In summary, (ψ^*, α) satisfy the following coupled system consisting of the Monge-Ampère equation and the transport equation:

$$\det(D^2\psi^*) = \alpha \qquad \text{in } \mathbf{R}^3 \times (0,T], \quad (8.23)$$

$$\frac{\partial \alpha}{\partial t} + \operatorname{div}(\mathbf{v}\alpha) = 0 \qquad \text{in } \mathbf{R}^3 \times (0,T], \quad (8.24)$$

$$\alpha(x,0) = \alpha_0 \qquad \text{in } \mathbf{R}^3 \times \{t=0\}, \quad (8.25)$$

$$D\psi^* \subset \Omega. \quad (8.26)$$

Remark 8.1.4. *As a comparison, the two-dimensional incompressible Euler equations (in*

the vorticity-stream function formulation) has the form

$$\begin{aligned} \Delta\phi &= \omega && \text{in } \Omega \times (0,T], \\ \frac{\partial \omega}{\partial t} + \operatorname{div}(\mathbf{u}\omega) &= 0 && \text{in } \Omega \times (0,T], \\ \mathbf{u} &= (D\phi)^{\perp}. \end{aligned}$$

Clearly, the main difference is that ϕ-equation above is a linear equation while ψ^* in (8.23) *is a fully nonlinear equation.*

We now cite the following existence and regularity results for (8.23)–(8.26). The proof can be found in [11]

Theorem 8.1.5. *Let $\Omega_0, \Omega \subset \mathbf{R}^3$ be two bounded Lipschitz domain. Suppose further that $\alpha_0 \in L^p(\mathbf{R}^3)$ with $\alpha_0 \geq 0$, $supp(\alpha_0) \subset \Omega_0$, and $\int_{\Omega_0} \alpha_0(x)dx = |\Omega|$. Then for any $T > 0$, $p > 1$,* (8.23)–(8.26) *has a weak solution (ψ^*, α) in the sense of* (8.20) *and* (8.22). *Furthermore, there exists an $R > 0$ such that $supp(\alpha(x,t)) \subset B_R(0)$ for all $t \in [0,T]$ and*

$$\begin{aligned} &\alpha \in L^\infty([0,T]; L^p(B_R(0))) && \textit{nonnegative}, \\ &\psi \in L^\infty([0,T]; W^{1,\infty}(\Omega)) && \textit{convex in physical space}, \\ &\psi^* \in L^\infty([0,T]; W^{1,\infty}(\mathbf{R}^3)) && \textit{convex in dual space}. \end{aligned}$$

The main task for the rest of this chapter is to formulate and analyze numerical methods for the system (8.23)–(8.26). We remark that since α and ψ^* are not physical variables, one needs to recover the physical variables $\mathbf{u}$ and p from α and ψ^*. This can be done by first constructing the geopotential ψ from its Legendre transform ψ^*. Numerically, this can be done by fast inverse Legendre transform algorithms [70]. Second, one recovers the pressure field p from the geopotential ψ using (8.14). Third, one obtains the geostrophic wind $\mathbf{u}_g$ and the full velocity field $\mathbf{u}$ from the pressure field p using (8.7).

We conclude this section by remarking that in the case that the gravity is omitted, the flow becomes two-dimensional. Repeating the derivation of this section and dropping the third component of all vectors, we then obtained a two dimensional semigeostrophic flow model which has exactly the same form as (8.23)–(8.26) except that the definition of $\mathbf{v}$ becomes

$$\mathbf{v} = (\frac{\partial \psi^*}{\partial x_2} - x_2, x_1 - \frac{\partial \psi^*}{\partial x_1}).$$

8.2 Vanishing Moment Approximation

By inspecting the above system, one easily observes that there are three clear difficulties for approximating the solution of (8.23)–(8.26). First, the equations are posed over an

unbounded domain, which makes numerically solving the system infeasible. The second difficulty is the full nonlinearity in equation (8.23). Third, equation (8.25) imposes a nonstandard constraint on the solution ψ^*, which often is called the second kind boundary condition for ψ^* in the PDE community (cf. [10, 36]).

As a first step to approximate the solution of the above system, we solve (8.23)–(8.26) over a finite and convex domain, $U \subset \mathbf{R}^3$. For the second difficulty, we employ the vanishing moment methodology introduced in Chapter 2 and approximate the fully nonlinear system (8.23) by the following quasilinear problem:

$$-\epsilon\Delta^2\psi^\epsilon + \det(D^2\psi^\epsilon) = \alpha^\epsilon \qquad \text{in } U \times (0,T], \tag{8.27}$$

$$\frac{\partial \alpha^\epsilon}{\partial t} + \operatorname{div}(\mathbf{v}^\epsilon \alpha^\epsilon) = 0 \qquad \text{in } U \times (0,T], \tag{8.28}$$

$$\alpha^\epsilon(x,0) = \alpha_0(x) \qquad \text{in } \mathbf{R}^3 \times \{t=0\}, \tag{8.29}$$

where $\epsilon > 0$ and

$$\mathbf{v}^\epsilon := \big(D\psi^\epsilon - x\big)^\perp.$$

System (8.27)–(8.29) is under-constrained, so extra constraints are required to ensure uniqueness. To this end, we impose the following conditions:

$$\frac{\partial \psi^\epsilon}{\partial \eta} = 0 \qquad \text{on } U \times (0,T], \tag{8.30}$$

$$\frac{\partial \Delta\psi^\epsilon}{\partial \eta} = 0 \qquad \text{on } U \times (0,T], \tag{8.31}$$

$$(\psi^\epsilon, 1) = 0 \qquad t \in (0,T]. \tag{8.32}$$

We remark that the choice of (8.30) intends to minimize the "reflection" due to the introduction of the finite computational domain U. It can be regarded as a simple radiation boundary condition. An additional consequence of (8.30) is that it also effectively overcomes the third difficulty, which is caused by the nonstandard constraint (8.26) for solving system (8.23)–(8.26). Finally, (8.32) is a mathematical technique for selecting a unique function from a class of functions differing from each other by an additive constant.

Since (8.27)–(8.32) is a quasilinear system, we can define weak solutions in the usual way using integration by parts.

Definition 8.2.1. *A pair of functions* $(\psi^\epsilon, \alpha^\epsilon) \in L^\infty((0,T); H^2(U)) \times L^2((0,T); H^1(U)) \cap H^1((0,T); L^2(U))$ *is called a weak solution to* (8.27)–(8.32) *if they satisfy the following integral identities for almost every* $t \in (0,T)$*:*

$$-\epsilon\big(\Delta\psi^\epsilon, \Delta v\big) + \big(\det(D^2\psi^\epsilon), v\big) = (\alpha^\epsilon, v) + \langle \epsilon^2, v\rangle_{\partial U} \qquad \forall v \in H^2(U), \tag{8.33}$$

$$\Big(\frac{\partial \alpha^\epsilon}{\partial t}, w\Big) + \big(\mathbf{v}^\epsilon \cdot D\alpha^\epsilon, w\big) = 0 \qquad \forall w \in H^1(U), \tag{8.34}$$

$$\big(\alpha^\epsilon(\cdot, 0), z\big) = \big(\alpha_0, z\big) \qquad \forall z \in L^2(U), \tag{8.35}$$

$$(\psi^\epsilon, 1) = 0, \tag{8.36}$$

where we have used the fact that $\operatorname{div} \mathbf{v}^\epsilon = 0$.

For the continuation of the paper, we assume that there exists a unique solution to (8.27)–(8.32) such that $\psi^\epsilon(x,t)$ is convex, $\alpha^\epsilon(x,t) \geq 0$, and supp $\alpha^\epsilon(x,t) \subset B_R(0) \subset U$ for all $t \in [0,T]$. We also assume $\psi^\epsilon \in L^2((0,T); H^s(U))$ $(s \geq 3)$, $\alpha^\epsilon \in L^2((0,T); H^p(U))$ $(p \geq 2)$, and that the following bounds hold (cf. (2.11)) for almost all $t \in [0,T]$

$$\|\psi^\epsilon(t)\|_{H^j} = O(\epsilon^{\frac{1-j}{2}}) \ (j = 1,2,3), \qquad \|\Phi^\epsilon(t)\|_{L^\infty} = O(\epsilon^{-1}), \tag{8.37}$$

$$\|\psi^\epsilon(t)\|_{W^{j,\infty}} = O(\epsilon^{1-j}) \ (j = 1,2), \qquad \|\alpha^\epsilon(t)\|_{W^{1,\infty}} = O(\epsilon^{-1}), \tag{8.38}$$

where $\Phi^\epsilon = \operatorname{cof}(D^2\psi^\epsilon)$ denotes the cofactor matrix of $D^2\psi^\epsilon$.

The following lemma provides a key assertion, that is, $\alpha^\epsilon(x,t) \geq 0$ in $U \times [0,T]$ provided that $\alpha_0(x) \geq 0$ in $\mathbf{R}^n (\, n = 2,3)$. The proof can be found in [53].

Lemma 8.2.2. *Suppose* $(\alpha^\epsilon, \psi^\epsilon)$ *is a regular solution of* (8.27)–(8.32). *Assume* $\alpha_0(x) \geq 0$ *in* $\mathbf{R}^n (\, n = 2,3)$. *Then* $\alpha^\epsilon(x,t) \geq 0$ *in* $U \times [0,T]$. *Furthermore, if* α_0 *is compactly supported, then* $\alpha^\epsilon(\cdot, t)$ *is also compactly supported for all* $t \in [0,T]$.

The remainder of this chapter is concerned with formulating and analyzing a modified characteristic finite element method for problem (8.27)–(8.32). The proposed method approximates the elliptic equation for ψ^ϵ by conforming finite element methods (cf. [27, 17] and Chapter 3) and discretizes the transport equation for α^ϵ by a modified characteristic method due to Douglas and Russell [42]. When deriving error estimates, we are particularly interested in the explicit dependence on ϵ for the proposed numerical method.

8.3 Formulation of a Modified Characteristic Finite Element Method

Let $\mathcal{T}_h$ be a quasiuniform triangulation or rectangular partition of U with mesh size $h \in (0,1)$ and $V^h \subset H^2(U)$ denote a conforming finite element space (such as Argyris, Bell, Bogner–Fox–Schmit, and Hsieh–Clough–Tocher finite element spaces [27] when

$n = 2$) consisting of piecewise polynomial functions of degree $r\,(\geq 5)$ such that for any $v \in H^s(U)$ $(s \geq 3)$

$$\inf_{v_h \in V^h} \|v - v_h\|_{H^j} \leq h^{\ell - j} \|v\|_{H^s}, \quad j = 0, 1, 2;\ \ell = \min\{r+1, s\}.$$

Let W^h be a finite dimensional subspace of $H^1(U))$ consisting of piecewise polynomials of degree $k\,(\geq 1)$ associated with the mesh $\mathcal{T}_h$.

Set

$$V_0^h := \Big\{ v_h \in V^h;\ \frac{\partial v_h}{\partial \eta}\Big|_{\partial U} = 0 \Big\}, \qquad V_1^h := \{ v_h \in V_0^h;\ (v_h, 1) = 0 \},$$
$$W_0^h := \{ w_h \in W^h;\ w_h\big|_{\partial U} = 0 \}, \qquad \tau := \frac{(1, \mathbf{v}^\epsilon)}{\sqrt{1 + |\mathbf{v}^\epsilon|^2}} \in \mathbf{R}^{n+1}.$$

We then have

$$\frac{\partial}{\partial \tau} := \tau \cdot \Big(\frac{\partial}{\partial t}, D \Big) = \frac{1}{\sqrt{1 + |\mathbf{v}^\epsilon|^2}} \Big(\frac{\partial}{\partial t} + \mathbf{v}^\epsilon \cdot D \Big).$$

Hence, we have

$$\frac{\partial \alpha^\epsilon}{\partial \tau} = \frac{1}{\sqrt{1 + |\mathbf{v}^\epsilon|^2}} \Big(\frac{\partial \alpha^\epsilon}{\partial t} + \mathbf{v}^\epsilon \cdot D\alpha^\epsilon \Big) = 0, \tag{8.39}$$

where we have used the fact that $\operatorname{div} \mathbf{v}^\epsilon = 0$.

Next, for a fixed positive integer M, let $\Delta t := \frac{T}{M}$ and $t_m := m\Delta t$ for $m = 0, 1, 2, \cdots, M$. For any $x \in U$, let $\overline{x} := x - \mathbf{v}^\epsilon(x, t)\Delta t$. It follows from Taylor's formula that (cf. [40, 42])

$$\frac{\partial \alpha^\epsilon(x, t_m)}{\partial \tau} = \frac{\alpha^\epsilon(x, t_m) - \alpha^\epsilon(\bar{x}, t_{m-1})}{\Delta t} + O(\Delta t) \qquad \text{for } m = 1, 2, \cdots, M. \tag{8.40}$$

Borrowing ideas from [40, 42], we propose the following modified characteristic finite element method for problem (8.27)–(8.32):

Step 1: Let α_h^0 be the finite element interpolation or the elliptic projection of α_0.

Step 2: For $m = 0, 1, 2, \ldots M$, find $(\psi_h^m, \alpha_h^{m+1}) \in V_1^h \times W_0^h$ such that

$$-\epsilon(\Delta \psi_h^m, \Delta v_h) + (\det(D^2 \psi_h^m), v_h) = (\alpha_h^m, v_h) + \big\langle \epsilon^2, v_h \big\rangle_{\partial U} \qquad \forall v_h \in V_0^h, \tag{8.41}$$
$$(\psi_h^m, 1) = 0, \tag{8.42}$$
$$\big(\alpha_h^{m+1} - \overline{\alpha}_h^m, w_h \big) = 0 \qquad \forall w_h \in W_0^h, \tag{8.43}$$

where

$$\overline{\alpha}_h^m := \alpha_h^m(\overline{x}_h), \qquad \overline{x}_h := x - \mathbf{v}_h^m \Delta t, \qquad \mathbf{v}_h^m := (D\psi_h^m - x)^\perp.$$

Let $(\psi^\epsilon, \alpha^\epsilon)$ be the solution of (8.27)–(8.32) and (ψ_h^m, α_h^m) be the solution of (8.41)–(8.43). In the subsequent sections we prove existence and uniqueness for (ψ_h^m, α_h^m) and provide optimal order error estimates for $\psi^\epsilon(t_m) - \psi_h^m$ and $\alpha^\epsilon(t_m) - \alpha_h^m$ under certain mesh and time stepping constraints. To this end, we first use the results of Section 3.4, where finite element approximations of the Monge-Ampère equation with small perturbations of the data was studied. The results of this section enables us to bound the error $\psi^\epsilon(t_m) - \psi_h^m$ in terms of of the error $\alpha^\epsilon(t_m) - \alpha_h^m$. With this result in hand, we use an inductive argument in Section 8.4 to get the desired error estimates for both $\psi^\epsilon(t_m) - \psi_h^m$ and $\alpha^\epsilon(t_m) - \alpha_h^m$.

8.4 Error Analysis for Finite Element Method (8.41)–(8.43)

In this section, we provide the main results of the chapter, where we obtain optimal error estimates of both $\|\psi^\epsilon - \psi_h^\epsilon\|$ and $\|\alpha^\epsilon - \alpha^\epsilon\|$ under certain time stepping and mesh constraints. First, we note Theorems 3.4.3 and 3.4.5 immediately give us the following result.

Theorem 8.4.1. *Assume that we have* $\|\alpha^\epsilon(t_m) - \alpha_h^m\|_{H^{-2}} = O(\epsilon^{n+1} h^{\frac{3n-6}{2}})$. *Then for* $h \leq \min\{h_1, h_2\}$, *there exists a unique solution,* ψ_h^m *to* (8.41), *where* h_1 *is defined in Theorem 3.3.4, and* h_2 *is defined in Theorem 3.4.5, that is*

$$h_1 = \begin{cases} O\Big(\frac{\epsilon^{\frac{5}{2}}}{\|\psi^\epsilon\|_{H^\ell}}\Big)^{\frac{1}{\ell-2}} & n = 2 \\ O\Big(\min\Big\{\Big(\frac{\epsilon^5}{\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}}\Big)^{\frac{1}{\ell-2}}, \Big(\frac{\epsilon^4}{\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}}\Big)^{\frac{2}{2\ell-7}}\Big\}\Big) & n = 3 \end{cases}$$

$$h_2 = O\Big(\frac{\epsilon^{\frac{1}{2}(5n-3)}}{\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}}\Big)^{\frac{2}{2\ell+2-3n}},$$

where $\ell = \min\{r+1, s\}$.

Moreover, there exists constants $C_1(\epsilon) = O(\epsilon^{\frac{3}{2}(1-n)})$, $C_2(\epsilon) = O(\epsilon^{-1})$, *such that*

$$\|\psi^\epsilon(t_m) - \psi_h^m\|_{H^2} \leq C_1(\epsilon) h^{\ell-2} \|\psi^\epsilon(t_m)\|_{H^\ell} + C_2(\epsilon) \|\alpha^\epsilon(t_m) - \alpha_h^m\|_{H^{-2}}, \tag{8.44}$$

$$\|\psi^\epsilon(t_m) - \psi_h^m\|_{H^1} \leq C\epsilon^{-2}\Big(C_1(\epsilon)\epsilon^{-\frac{1}{2}} h^{\ell-1} \|\psi^\epsilon(t_m)\|_{H^\ell} + (C_2(\epsilon)\epsilon^{-\frac{1}{2}} h + 1)\|\alpha^\epsilon(t_m) - \alpha_h^m\|_{H^{-2}}\Big). \tag{8.45}$$

Remark 8.4.2. *Let* $h_3 = O(\epsilon^{\frac{3}{2}})$. *Then under the same hypotheses of Theorem 8.4.1, we have for* $h \leq \min\{h_1, h_2, h_3\}$

$$\|\psi^\epsilon(t_m) - \psi_h^m\|_{H^1} \leq C\epsilon^{-2}\Big(C_1(\epsilon)\epsilon^{-\frac{1}{2}} h^{\ell-1} \|\psi^\epsilon(t_m)\|_{H^\ell} + \|\alpha^\epsilon(t_m) - \alpha_h^m\|_{H^{-2}}\Big). \tag{8.46}$$

Before proving our main result, we comment on the error estimates of the elliptic projection of α^ϵ, which we denote by $a_h \in W_0^h$. Letting $\omega = \alpha^\epsilon - a_h$, then it is well-known

that the following bounds hold [42, 41].

$$\begin{aligned}
&\|\omega\|_{L^2([0,T];L^2)} + \|\omega_t\|_{L^2([0,T];L^2)} \qquad (8.47)\\
&\qquad + h\big(\|\omega\|_{L^2([0,T];H^1)} + \|\omega_t\|_{L^2([0,T];H^1)}\big) \le Ch^j\Big\{\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\Big\},\\
&\|\omega(t)\|_{W^{1,\infty}} \le Ch^{j-1}\|\alpha(t)\|_{W^{j,\infty}} \qquad \text{for a.e. } t\in[0,T),
\end{aligned}$$

where $j = \min\{k+1, p\}$.

Using these results, we have the following lemma.

Lemma 8.4.3. *Suppose $k \ge 3$. Then there exists a constant $C > 0$, such that for any $m = 0, 1, ..., M$,*

$$\|\omega^m\|_{H^{-2}} \le Ch^{j+2}\|\alpha^\epsilon(t_m)\|_{H^j}. \qquad (8.48)$$

Proof. We use a standard duality argument to prove (8.48). For arbitrary $\phi \in H^2(U)$, let $w_\phi \in H^4(U)$ be the unique solution to

$$\begin{aligned}
-\Delta w_\phi &= \phi \qquad \text{in } U,\\
w_\phi &= 0 \qquad \text{on } \partial U.
\end{aligned}$$

Since ∂U is assumed to be smooth, we have $\|w_\phi\|_{H^4} \le C\|\phi\|_{H^2}$. Thus, for any $t_m \in [0,T]$ and any $w_h \in W_0^h$,

$$\begin{aligned}
(\omega^m, \phi) &= -(\omega^m, \Delta w_\phi) = (D\omega^m, Dw_\phi)\\
&= (D\omega^m, D(w_\phi - w_h)) \le \|\omega^m\|_{H^1}\|w_\phi - w_h\|_{H^1}\\
&\le Ch^{j-1}\|\alpha^\epsilon(t_m)\|_{H^j}\|w_\phi - w_h\|_{H^1}.
\end{aligned}$$

Thus, for appropriate choice of $w_h \in W_0^h$ (say $w_h = I_h w_\phi$, the finite element interpolant of w onto W_0^h), we have

$$(\omega^m, \phi) \le Ch^{j+2}\|\alpha^\epsilon(t_m)\|_{H^j}\|w_\phi\|_{H^4} \le Ch^{j+2}\|\alpha^\epsilon(t_m)\|_{H^j}\|\phi\|_{H^2}.$$

Dividing by $\|\phi\|_{H^2}$ and noting ϕ was arbitrary, we obtain (8.48). □

With this result in hand, we are able to present an inductive argument that will give us uniqueness and the sought after error estimates.

Theorem 8.4.4. *There exists h_4 such that for $h \le \min\{h_1, h_2, h_3, h_4\}$ there exists $\Delta t_1 > 0$ such that for $\Delta t \le \min\{\Delta t_1, h^2\}$ the following error estimates hold:*

$$\max_{0\le m\le M} \|\alpha^\epsilon(t_m) - \alpha_h^m\|_{L^2} \le C_3(\epsilon)\Big\{\Delta t \|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbb{R}^3)} + h^j\big[\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big] + C_4(\epsilon)h^\ell \|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}, \tag{8.49}$$

$$\max_{0\le m\le M} \|\psi^\epsilon(t_m) - \psi_h^m\|_{H^2} \le C_5(\epsilon)\Big\{\Delta t \|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbb{R}^3)} + h^j\big[\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big] + C_4(\epsilon)h^{\ell-2} \|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}, \tag{8.50}$$

$$\max_{0\le m\le M} \|\psi^\epsilon(t_m) - \psi_h^m\|_{H^1} \le C_6(\epsilon)\Big\{\Delta t \|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbb{R}^3)} + h^j\big[\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big] + C_4(\epsilon)h^{\ell-1} \|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}, \tag{8.51}$$

where

$$\alpha^\epsilon_{\tau\tau} := \frac{\partial^2 \alpha^\epsilon}{\partial \tau^2}, \qquad \alpha^\epsilon_t := \frac{\partial \alpha^\epsilon}{\partial t}, \qquad C_3(\epsilon) = O(\epsilon^{-1}),$$
$$C_4(\epsilon) = O(\epsilon^{-\frac{1}{2}(2+3n)}), \quad C_5(\epsilon) = O(\epsilon^{-2}), \quad C_6(\epsilon) = O(\epsilon^{-3}),$$
$$\ell = \min\{r+1, s\}, \qquad j = \min\{k+1, p\}.$$

Proof. We break the proof up into five steps.

Step 1: The proof is based on two induction hypotheses, where we assume for $m = 0, 1, ..., k$,

$$\|\alpha^\epsilon(t_m) - \alpha_h^m\|_{H^{-2}} = O\big(\epsilon^{n+1} h^{\frac{3n-6}{2}}\big), \tag{8.52}$$

$$\|D^2 \psi_h^m\|_{L^\infty} = O(\epsilon^{-1}) \tag{8.53}$$

We now show that the case $k = 0$ holds. Letting

$$h_5 = O\left(\frac{\epsilon^{n+1}}{\|\alpha_0\|_{H^j}}\right)^{\frac{2}{2j-3n-2}},$$

and using (8.47), we have for $h \le h_5$

$$\|\alpha_0 - \alpha_h^0\|_{H^{-2}} \le C h^{j+2} \|\alpha_0\|_{H^j} \le C\epsilon^{n+1} h^{\frac{3n-6}{2}}.$$

Next, by Theorem 8.4.1, there exists ψ_h^0 solving (8.41). Noting $h_1 \le C\left(\frac{\epsilon^{\frac{1}{2}(n-1)}}{\|\psi^\epsilon(0)\|_{H^\ell}}\right)^{\frac{2}{2\ell-7}}$, we have for $h \le \min\{h_1, h_2, h_5\}$,

$$\begin{aligned}
\|D^2\psi_h^0\|_{L^\infty} &\le \|D^2\psi^\epsilon(0)\|_{L^\infty} + h^{-\frac{3}{2}}\|D^2\psi^\epsilon(0) - D^2\psi_h^0\|_{L^2} \\
&\le C\Big(\epsilon^{-1} + h^{-\frac{3}{2}}C_1(\epsilon)h^{\ell-2}\|\psi^\epsilon(0)\|_{H^\ell} + C_2(\epsilon)\|\alpha_0^\epsilon - \alpha_h^0\|_{H^{-2}})\Big) \\
&\le C\Big(\epsilon^{-1} + C_1(\epsilon)h^{\frac{2\ell-7}{2}}\|\psi^\epsilon(0)\|_{H^\ell} + C_2(\epsilon)h^{j+2}\|\alpha_0^\epsilon\|_{H^j})\Big) \le C\epsilon^{-1}.
\end{aligned}$$

The remaining four steps are devoted to show that the estimates hold for general k at the end of the proof.

Step 2: Let $\xi^m = \alpha_h^m - a_h^m$. By (8.43) and (8.28), a straight-forward calculation shows,

$$\begin{aligned}
\big(\xi^{m+1} - \overline{\xi}^m, \xi^{m+1}\big) &= -\big(a_h^{m+1} - \overline{a}_h^m, \xi^{m+1}\big) \\
&= \big(\Delta t\, \alpha_\tau^\epsilon(t_{m+1}) - (a_h^{m+1} - \overline{a}_h^m), \xi^{m+1}\big) \\
&= \big(\Delta t\, \alpha_\tau^\epsilon(t_{m+1}) - \big(\alpha^\epsilon(t_{m+1}) - \overline{\alpha}^\epsilon(t_m)\big), \xi^{m+1}\big) \\
&\quad + \big(\omega^{m+1} - \overline{\omega}^m, \xi^{m+1}\big),
\end{aligned} \tag{8.54}$$

where $\overline{\xi}^m := \xi^m(\overline{x}_h)$, $\overline{\alpha}_h^\epsilon(t_m) := \alpha^\epsilon(\overline{x}_h, t_m)$, and $\overline{\omega}_h^m := \omega^m(\overline{x}_h)$.

We now bound the right hand side of (8.54). To bound the first term, we write

$$\begin{aligned}
&\Delta t\, \alpha_t^\epsilon(x, t_{m+1}) - \big(\alpha^\epsilon(x, t_{m+1}) - \alpha^\epsilon(\bar{x}, t_m)\big) \\
&\qquad = \Delta t\, \alpha_t^\epsilon(x, t_{m+1}) - \big(\alpha^\epsilon(x, t_{m+1}) - \alpha^\epsilon(\bar{x}, t_m)\big) + \big(\alpha^\epsilon(\bar{x}_h, t_m) - \alpha^\epsilon(\bar{x}, t_m)\big).
\end{aligned}$$

Using the identity

$$\Delta t\, \alpha_\tau^\epsilon(x, t_{m+1}) - \big(\alpha^\epsilon(x, t_{m+1}) - \alpha^\epsilon(\bar{x}, t_m)\big) = \int_{(\bar{x}, t_m)}^{(x, t_{m+1})} \sqrt{|x(\tau) - \bar{x}_h|^2 + (t(\tau) - t_m)^2}\; \alpha_{\tau\tau}^\epsilon d\tau.$$

and (8.37), we obtain

$$\begin{aligned}
&\|\Delta t\, \alpha_\tau^\epsilon(t_{m+1}) - \big(\alpha^\epsilon(t_{m+1}) - \overline{\alpha}^\epsilon(t_m)\big)\|_{L^2}^2 \\
&\qquad = \int_{\mathbf{R}^n} \Big| \int_{(\bar{x}, t_m)}^{(x, t_{m+1})} \sqrt{|x(\tau) - \bar{x}|^2 + (t(\tau) - t_m)^2}\; \alpha_{\tau\tau}^\epsilon d\tau \Big|^2 dx \\
&\qquad \le \Delta t \int_{\mathbf{R}^n} \sqrt{|\mathbf{v}^\epsilon(t_{m+1})|^2 + 1} \Big| \int_{(\bar{x}, t_m)}^{(x, t_{m+1})} \alpha_{\tau\tau}^\epsilon d\tau \Big|^2 dx \\
&\qquad \le C\Delta t^2 \|\mathbf{v}^\epsilon(t_{m+1})\|_{L^\infty} \int_{\mathbf{R}^3} \int_{(\bar{x}, t_m)}^{(x, t_{m+1})} \big|\alpha_{\tau\tau}^\epsilon\big|^2 d\tau dx
\end{aligned} \tag{8.55}$$

$$= C\Delta t^2 \|\mathbf{v}^\epsilon(t_{m+1})\|_{L^\infty} \|\alpha^\epsilon_{\tau\tau}\|^2_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}$$
$$\le C\Delta t^2 \|\alpha^\epsilon_{\tau\tau}\|^2_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)},$$

where $\overline{\alpha}^\epsilon(t_m) := \alpha^\epsilon(\bar{x}, t_m)$.

Next, since

$$\alpha^\epsilon(\bar{x}_h, t_m) - \alpha^\epsilon(\bar{x}, t_m) = \int_0^1 D\alpha^\epsilon(\bar{x}_h + s(\bar{x} - \bar{x}_h), t_m) \cdot (\bar{x} - \bar{x}_h) ds,$$

we have

$$\begin{aligned} &\|\overline{\alpha}^\epsilon(t_m) - \overline{\alpha}^\epsilon(t_m)\|^2_{L^2} \\ &\quad = \int_{\mathbf{R}^n} \Big| \int_0^1 D\alpha^\epsilon(\bar{x}_h + s(\bar{x} - \bar{x}_h), t_m) \cdot (\bar{x} - \bar{x}_h) ds \Big|^2 dx \\ &\quad = \Delta t^2 \int_{\mathbf{R}^n} \Big| \int_0^1 D\alpha^\epsilon(\bar{x}_h + s(\bar{x} - \bar{x}_h), t_m) \cdot (\mathbf{v}^m_h - \mathbf{v}^\epsilon(t_m)) ds \Big|^2 dx \\ &\quad \le \Delta t^2 \|\alpha^\epsilon(t_m)\|^2_{W^{1,\infty}} \|\mathbf{v}^m_h - \mathbf{v}^\epsilon(t_m)\|^2_{L^2} \\ &\quad \le C\epsilon^{-2}\Delta t^2 \|\mathbf{v}^m_h - \mathbf{v}^\epsilon(t_m)\|^2_{L^2}. \end{aligned} \tag{8.56}$$

Using (8.55)–(8.56), we can bound the the first term of the right hand side of (8.54) as follows:

$$\begin{aligned} &\big(\Delta t\, \alpha^\epsilon_\tau(t_{m+1}) - \big(\alpha^\epsilon(t_{m+1}) - \overline{\alpha}^\epsilon(t_m)\big), \xi^{m+1}\big) \\ &\quad \le C\Delta t^2 \big(\|\alpha_{\tau\tau}\|^2_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)} + \epsilon^{-2} \|\mathbf{v}^m_h - \mathbf{v}^\epsilon(t_m)\|^2_{L^2} \big) + \frac{1}{8} \|\xi^{m+1}\|^2_{L^2}. \end{aligned} \tag{8.57}$$

To bound the second term of the right hand side of (8.54), we write

$$\begin{aligned} &\omega^{m+1}(x) - \omega^m(\bar{x}_h) \\ &\quad = \big(\omega^{m+1}(x) - \omega^m(x)\big) + \big(\omega^m(x) - \omega^m(\bar{x})\big) + \big(\omega^m(\bar{x}) - \omega^m(\bar{x}_h)\big). \end{aligned}$$

We then have

$$\begin{aligned} \|\omega^{m+1} - \omega^m\|^2_{L^2} &\le \int_{\mathbf{R}^n} \Big| \int_{t_m}^{t_{m+1}} \omega_t(t) dt \Big|^2 dx \\ &\le \Delta t \int_{\mathbf{R}^n} \int_{t_m}^{t_{m+1}} |\omega_t(t)|^2 dt \\ &= \Delta t \|\omega_t\|^2_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}. \end{aligned} \tag{8.58}$$

Next, we bound $\omega^m(x) - \omega^m(\bar{x})$ by noting

$$\begin{aligned}\omega^m(x) - \omega^m(\bar{x}) &= \int_0^1 D\omega^m(x + s(\bar{x} - x)) \cdot (\bar{x} - x) ds \\ &= \Delta t \int_0^1 D\omega^m(x + s(\bar{x} - x)) \cdot \mathbf{v}^\epsilon(t_m) ds.\end{aligned}$$

It follows that

$$\|\omega^m(x) - \overline{\omega}^m\|_{L^2}^2 \leq C\Delta t^2 \|\mathbf{v}^\epsilon(t_m)\|_{L^\infty}^2 \|\omega^m\|_{H^1}^2 \leq C\Delta t^2 \|\omega^m\|_{H^1}^2, \tag{8.59}$$

with $\overline{\omega}^m := \omega^m(\bar{x})$.

Finally, we bound $\omega^m(\bar{x}) - \omega^m(\bar{x}_h)$ using the identity

$$\omega^m(\bar{x}) - \omega^m(\bar{x}_h) = \Delta t \int_0^1 D\omega^m(\bar{x} + s(\bar{x}_h - \bar{x})) \cdot (\mathbf{v}^\epsilon(t_m) - \mathbf{v}_h^m) ds,$$

yielding

$$\begin{aligned}\|\overline{\omega}^m - \overline{\omega}^m\|_{L^2}^2 &\leq C\Delta t^2 \|\omega^m\|_{W^{1,\infty}}^2 \|\mathbf{v}^\epsilon(t_m) - \mathbf{v}_h^m\|_{L^2}^2 \\ &\leq C\Delta t^2 \|\mathbf{v}^\epsilon(t_m) - \mathbf{v}_h^m\|_{L^2}^2.\end{aligned} \tag{8.60}$$

Combining (8.58)–(8.60), we bound the second term on the right hand side of (8.54) as follows:

$$\begin{aligned}\left(\omega^{m+1} - \overline{\omega}^m, \xi^{m+1}\right) \leq C\Big(&\Delta t \|\omega_t\|_{L^2([t_m, t_{m+1}] \times \mathbf{R}^n)}^2 + \Delta t^2 \|\omega^m\|_{H^1}^2 \\ &+ \Delta t^2 \|\mathbf{v}^\epsilon(t_m) - \mathbf{v}_h^m\|_{L^2}^2\Big) + \frac{1}{8}\|\xi^{m+1}\|_{L^2}^2.\end{aligned} \tag{8.61}$$

Step 3: To get a lower bound of $(\xi^{m+1} - \bar{\xi}^m, \xi^{m+1})$, let $F_m(x) := x - \Delta t \mathbf{v}_h^m(x)$. We then have

$$\det(J_{F_m}) = 1 + \Delta t^2 \left(1 + \frac{\partial^2 \psi_h^m}{\partial x_1^2}\frac{\partial^2 \psi_h^m}{\partial x_2^2} - \left(\frac{\partial^2 \psi_h^m}{\partial x_1 \partial x_2}\right)^2 - \left(\frac{\partial^2 \psi_h^m}{\partial x_1^2} + \frac{\partial^2 \psi_h^m}{\partial x_2^2}\right)\right),$$

where J_{F_m} denotes the Jacobian of F_m. Letting $\Delta t_2 = O(\epsilon)$, we can conclude from the induction hypotheses that for $\Delta t \leq \Delta t_2$, F_m is invertible and $\det(J_{F_m^{-1}}) = 1 + C\epsilon^{-2}\Delta t^2$. From this result, we get

$$\|\overline{\xi}^m\|_{L^2}^2 = (1 + C\epsilon^{-2}\Delta t^2)\|\xi^m\|_{L^2}^2. \tag{8.62}$$

Thus, using the inequality

$$\frac{1}{2}(x^2-y^2) \le \frac{1}{2}(x^2-y^2+(x-y)^2) = (x-y)x,$$

we have

$$\begin{aligned}(\xi^{m+1}-\overline{\xi}^m,\xi^{m+1}) &\ge \frac{1}{2}\big((\xi^{m+1},\xi^{m+1})-(\overline{\xi}^m,\overline{\xi}^m)\big) \\ &= \frac{1}{2}\big(\|\xi^{m+1}\|_{L^2}^2-(1+C\epsilon^{-2}\Delta t^2)\|\xi^m\|_{L^2}^2\big).\end{aligned} \tag{8.63}$$

Step 4: Combining (8.54), (8.57), (8.61), (8.63), and using the induction hypotheses with Theorem 8.4.1, we have

$$\begin{aligned}&\|\xi^{m+1}\|_{L^2}^2-\|\xi^m\|_{L^2}^2 \\ &\quad\le C\Delta t^2\Big(\epsilon^{-1}\|\alpha_{\tau\tau}^\epsilon\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+\epsilon^{-2}\|\omega^m\|_{H^1}^2+\epsilon^{-2}\|\mathbf{v}_h^m-\mathbf{v}^\epsilon(t_m)\|_{L^2}^2\Big) \\ &\qquad + C\Delta t\|\omega_t\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+C\epsilon^{-2}\Delta t^2\|\xi^m\|_{L^2}^2 \\ &\quad\le C\epsilon^{-2}\Big\{\Delta t^2\Big(\|\alpha_{\tau\tau}^\epsilon\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+\|\omega^m\|_{H^1}^2+\|\mathbf{v}_h^m-\mathbf{v}^\epsilon(t_m)\|_{L^2}^2\Big) \\ &\qquad + \Delta t\|\omega_t\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+\Delta t^2\|\xi^m\|_{L^2}^2\Big\} \\ &\quad\le C\epsilon^{-2}\Big\{\Delta t^2\Big(\|\alpha_{\tau\tau}^\epsilon\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+\|\omega^m\|_{H^1}^2 \\ &\qquad + \epsilon^{-4}\big(C_1^2(\epsilon)\epsilon^{-1}h^{2\ell-2}\|\psi^\epsilon(t_m)\|_{H^\ell}^2+\|\alpha^\epsilon(t_m)-\alpha_h^m\|_{H^{-2}}^2\big)\Big) \\ &\qquad + \Delta t\|\omega_t\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^3)}^2+\Delta t^2\|\xi^m\|_{L^2}^2\Big\}.\end{aligned}$$

Using the inequality $\|\alpha^\epsilon(t_m)-\alpha_h^m\|_{H^{-2}} \le \|\alpha^\epsilon(t_m)-\alpha_h^m\|_{L^2} \le \|\xi^m\|_{L^2}+\|\omega^m\|_{L^2}$ yields

$$\begin{aligned}\|\xi^{m+1}\|_{L^2}^2-\|\xi^m\|_{L^2}^2 \le C\epsilon^{-2}\Big\{&\Delta t^2\Big(\|\alpha_{\tau\tau}^\epsilon\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2+\epsilon^{-4}\|\omega^m\|_{H^1}^2 \\ &+\epsilon^{-5}C_1^2(\epsilon)h^{2\ell-2}\|\psi^\epsilon(t_m)\|_{H^\ell}^2\Big)+\Delta t\|\omega_t\|_{L^2([t_m,t_{m+1}]\times\mathbf{R}^n)}^2 \\ &+\epsilon^{-4}\Delta t^2\|\xi^m\|_{L^2}^2\Big\}.\end{aligned}$$

Applying the summation operator $\sum_{m=0}^k$ and noting $\xi^0=0$, we have

$$\begin{aligned}\|\xi^{k+1}\|_{L^2}^2 \le C\epsilon^{-2}\Big\{&\Delta t^2\|\alpha_{\tau\tau}^\epsilon\|_{L^2([0,T]\times\mathbf{R}^3)}^2+\Delta t\Big(\epsilon^{-4}\|\omega\|_{L^2([0,T];H^1)}^2 \\ &+\epsilon^{-5}C_1^2(\epsilon)h^{2\ell-2}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}^2+\|\omega_t\|_{L^2([0,T]\times\mathbf{R}^n)}^2\Big) \\ &+\epsilon^{-4}\Delta t^2\sum_{m=0}^k\|\xi^m\|_{L^2}^2\Big\}.\end{aligned}$$

Using the discrete Gronwall inequality, we get

$$\begin{aligned}\|\xi^{k+1}\|_{L^2} \le C\epsilon^{-1}\left(1+\epsilon^{-3}\Delta t\right)^{k+1}\Big\{&\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)} \\ &+\sqrt{\Delta t}\Big(\epsilon^{-2}\|\omega\|_{L^2([0,T];H^1)} \\ &\quad+\epsilon^{-\frac{5}{2}}C_1(\epsilon)h^{\ell-1}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}+\|\omega_t\|_{L^2([0,T]\times\mathbf{R}^n)}\Big)\Big\}.\end{aligned} \tag{8.64}$$

Let $\Delta t_3 = O(\epsilon^3)$. Then for $\Delta t \le \min\{\Delta t_3, h^2\}$, we have using (8.64), the triangle inequality, and (8.47),

$$\begin{aligned}&\|\alpha^\epsilon(t_{k+1}) - \alpha_h^{k+1}\|_{L^2} \\ &\quad\le C_3(\epsilon)\Big\{\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}+h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)}+\|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big) \\ &\qquad+C_4(\epsilon)h^\ell\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}.\end{aligned} \tag{8.65}$$

Thus, by Theorem 8.4.1, we have the following bounds:

$$\begin{aligned}&\|\psi^\epsilon(t_{k+1}) - \psi_h^{k+1}\|_{H^2} \\ &\quad\le C_2(\epsilon)C_3(\epsilon)\Big\{\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}+h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)}+\|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big) \\ &\qquad+C_4(\epsilon)h^\ell\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}+C_1(\epsilon)h^{\ell-2}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)} \\ &\quad\le C_5(\epsilon)\Big\{\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}+h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)}+\|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big) \\ &\qquad+C_4(\epsilon)h^{\ell-2}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\},\end{aligned} \tag{8.66}$$

$$\begin{aligned}&\|\psi^\epsilon(t_{k+1}) - \psi_h^{k+1}\|_{H^1} \\ &\quad\le C\epsilon^{-2}C_3(\epsilon)\Big\{\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}+h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)}+\|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big) \\ &\qquad+C_4(\epsilon)h^\ell\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}+C\epsilon^{-\frac{5}{2}}C_1(\epsilon)h^{\ell-1}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)} \\ &\quad\le C_6(\epsilon)\Big\{\Delta t\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}+h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)}+\|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big) \\ &\qquad+C_4(\epsilon)h^{\ell-1}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\}.\end{aligned} \tag{8.67}$$

Step 5: We now verify the induction hypotheses. Let $C_7(\epsilon) = C_3(\epsilon)\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha^\epsilon_t\|_{L^2([0,T];H^j)}\big)$, $C_8(\epsilon) = C_4(\epsilon)\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}$, and

$$\begin{aligned}\Delta t_4 &= O\left(\frac{\epsilon^{n+1}h^{\frac{3n-6}{2}}}{C_3(\epsilon)\|\alpha^\epsilon_{\tau\tau}\|_{L^2([0,T]\times\mathbf{R}^n)}}\right) \\ h_6 &= O\left(\min\Big\{\left(\frac{\epsilon^{n+1}}{C_7(\epsilon)}\right)^{\frac{2}{2j-3n-6}},\left(\frac{\epsilon^{n+1}}{C_8(\epsilon)}\right)^{\frac{2}{2\ell-3n-6}}\Big\}\right)\end{aligned}$$

Thus, for $h \le h_6$ and $\Delta t \le \Delta t_4$, we have by (8.65),(8.66), and (8.37),

$$\begin{aligned}
\|\alpha^\epsilon(t_{k+1}) - \alpha_h^{k+1}\|_{L^2} &\le C\epsilon^{n+1}h^{\frac{3n-6}{2}}, \\
\|D^2\psi_h^{k+1}\|_{L^\infty} &\le \|D^2\psi^\epsilon(t_{k+1})\|_{L^\infty} + Ch^{-\frac{3}{2}}\|D^2\psi^\epsilon(t_{k+1}) - D^2\psi_h^{k+1}\|_{L^2} \\
&\le C\epsilon^{-1} + h^{-\frac{3}{2}}C_5(\epsilon)\Big\{\Delta t\|\alpha_{\tau\tau}^\epsilon\|_{L^2([0,T]\times\mathbf{R}^3)} \\
&\qquad + h^j\big(\|\alpha^\epsilon\|_{L^2([0,T];H^j)} + \|\alpha_t^\epsilon\|_{L^2([0,T];H^j)}\big) + C_4(\epsilon)h^{\ell-2}\|\psi^\epsilon\|_{L^2([0,T];H^\ell)}\Big\} \\
&\le C\epsilon^{-1}.
\end{aligned}$$

Therefore, the induction hypotheses (8.52)–(8.53) hold, and the proof is complete by setting $h_4 = \min\{h_5, h_6\}$ and $\Delta t_1 = \min\{\Delta t_2, \Delta t_3, \Delta t_4\}$.

□

Remark 8.4.5. *Recalling the definitions of V^h and W^h, we require $k \ge r-2 \ge 3$ in order to obtain optimal error estimates.*

8.5 Numerical Experiments and Rates of Convergence

In this section, we shall present several 2-D numerical tests to gauge the effectiveness of the modified characteristic finite element method, and to verify the error estimates of the previous section. The first four tests are performed on the domain $U = (0,1)^2$, while the fifth test uses $U = (0,6)^2$. In all five tests, the fifth degree Argyris plate finite element is used for V^h, and the cubic Lagrange element is used for W^h.

Test 8.1

The purpose of this test is twofold. First, we compute α_h^m and ψ_h^m to view certain properties of these two functions. Specifically, we are interested if α_h^m, $\Delta\psi_h^m$, and $\det(D^2\psi_h^m)$ are strictly positive for $m = 0, 1, ..., M$. Second, we calculate $\|\psi^* - \psi_h^m\|$ and $\|\alpha - \alpha_h^\epsilon\|$ for fixed h and Δt in order to approximate $\|\psi^* - \psi^\epsilon\|$ and $\|\alpha - \alpha^\epsilon\|$. We set to solve (8.41)–(8.43), but with the right-hand side of (8.43) being replaced by (F, w_h), and V_1^h and W_0^h being replaced by $V_{g_N}^h$ and $W_{g_N}^h$, where

$$\begin{aligned}
V_{g_N}^h(t) &:= \big\{v_h \in V^h;\ \frac{\partial v_h}{\partial \eta}\big|_{\partial U} = g_N,\ (v_h, 1) = c(t)\big\}, \qquad c(t) = (\psi^*, 1), \\
W_{g_D}^h(t) &:= \big\{w_h \in W_0^h;\ w_h\big|_{\partial U} = g_D\big\}.
\end{aligned}$$

We use the following test functions and parameters:

$$
\begin{aligned}
&\text{(a) } \psi^* = t(x_1^2 + x_2^2), && \alpha = 4t^2 \\
&\quad g_N = 2t(x_1\nu_1 + x_2\nu_2), && g_D = 4t^2, \\
&\quad F = 8t. \\
&\text{(b) } \psi^* = e^{t(x_1^2+x_2^2)/2}, && \alpha = t^2(1 + t(x_1^2 + x_2^2))e^{t(x_1^2+x_2^2)}, \\
&\quad g_N = te^{t(x_1^2+x_2^2)/2}(x_1\nu_1 + \nu_2), && g_D = t^2(1 + t(x_1^2 + x_2^2))e^{t(x_1^2+x_2^2)}, \\
&\quad F = t\big(2 + 4t(x_1^2 + x_2^2) + t^2(x_1^2 + x_2^2)^2\big)e^{t(x_1^2+x_2^2)}.
\end{aligned}
$$

We plot α_h^m, $\Delta\psi_h^m$, and $\det(D^2u_h^m)$ for both $t_m = 0.5$ and $t_m = 1$ with $h = 0.05$, $\Delta t = 0.1$ in Figures 8.1–8.4. As seen in the figures, all three quantities are positive for both values of t_m. This observation supports the assumption that $\alpha^\epsilon(t) > 0$ and $\psi^\epsilon(t)$ is strictly convex for all $t \in [0, T]$.

Next, we plot the errors versus ϵ at $t_m = 0.25$ in Figures 8.5 and 8.6. The figures show $\|\psi^*(t_m) - \psi_h^m\|_{H^2} = O(\epsilon^{\frac{1}{4}})$, and since we have set both h and Δt very small, these results suggest that $\|\psi^*(t_m) - \psi^\epsilon(t_m)\|_{H^2} = O(\epsilon^{\frac{1}{4}})$. Similarly, we argue $\|\psi^*(t_m) - \psi^\epsilon(t_m)\|_{H^1} = O(\epsilon^{\frac{3}{4}})$ and $\|\psi^*(t_m) - \psi^\epsilon(t_m)\|_{L^2} = O(\epsilon)$ based on our results. We note that these are the same convergence results found in Chapters 3–6, where the single Monge-Ampère equation was considered. We also notice that this test suggests that $\|\alpha(t_m) - \alpha^\epsilon(t_m)\|_{L^2}$ may not converge, which suggests that the convergence can only be possible in a weaker norm such as $H^{-2}(\Omega)$.

Test 8.2

The goal of this test is to calculate the rate of convergence of $\|\psi^\epsilon(t_m) - \psi_h^m\|$ and $\|\alpha^\epsilon(t_m) - \alpha_h^m\|$ for fixed ϵ and h while varying Δt. We use the same domain and finite element spaces as in Test 8.1, and set to solve (8.41)–(8.43), but with the right-hand side of (8.41) being replaced by $(\alpha_h^m, v_h) + \epsilon\langle\phi^\epsilon, v_h\rangle_{\partial U}$, and the right-hand side of (8.43) being replaced by (F^ϵ, w_h). Also, V_1^h and W_0^h are replaced by $V_{g_N^\epsilon}^h$ and $W_{g_N^\epsilon}^h$, where

$$
\begin{aligned}
&V_{g_N^\epsilon}^h(t) := \big\{v_h \in V^h;\ \frac{\partial v_h}{\partial \eta}\big|_{\partial U} = g_N^\epsilon,\ (v_h, 1) = c^\epsilon(t)\big\}, \qquad c^\epsilon(t) = (\psi^\epsilon, 1), \\
&W_{g_D^\epsilon}^h(t) := \big\{w_h \in W_0^h;\ w_h\big|_{\partial U} = g_D^\epsilon\big\}.
\end{aligned}
$$

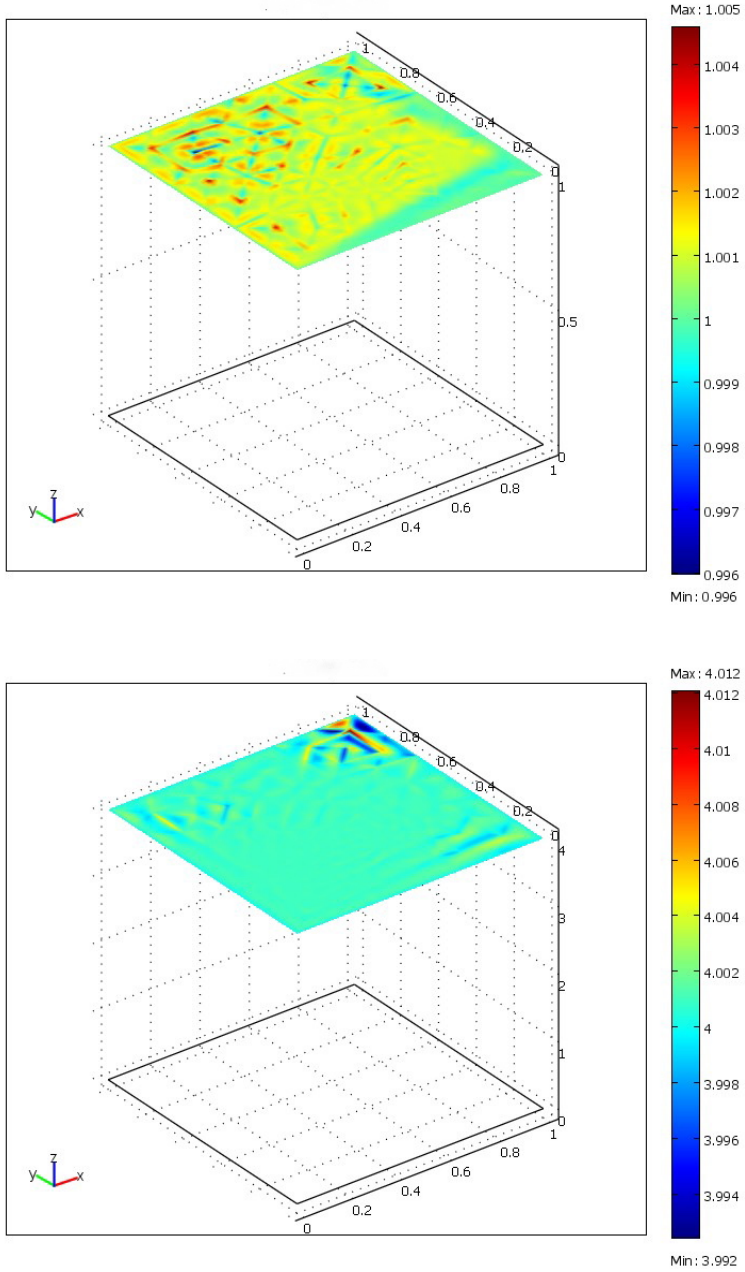


Figure 8.1: Test 8.1a: Computed α_h^M at $t_M = 0.5$ and $t_M = 1$. $\Delta t = 0.1$, $h = 0.05$.

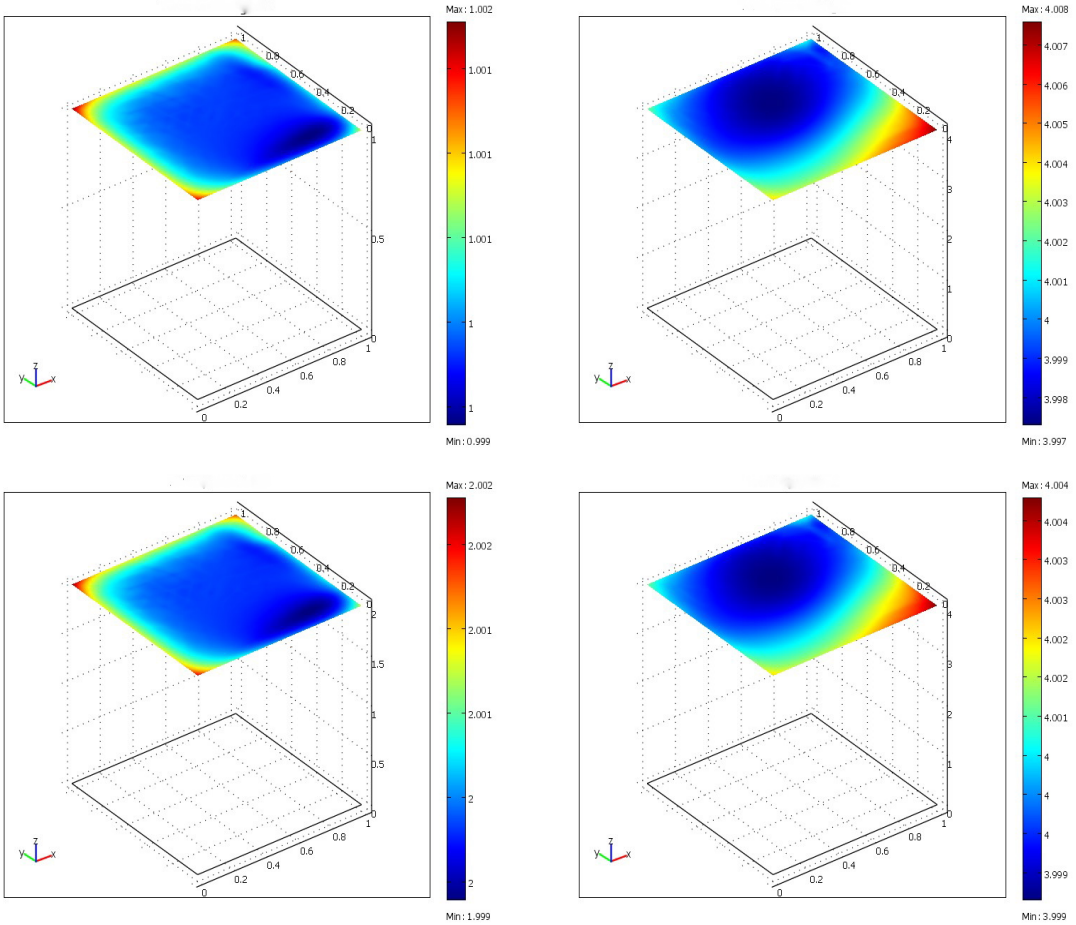


Figure 8.2: Test 8.1a: Computed determinant (top) and Laplacian (bottom) at $t_M = 0.5$ (left) and $t_M = 1$ (right). $\Delta t = 0.1$, $h = 0.05$.

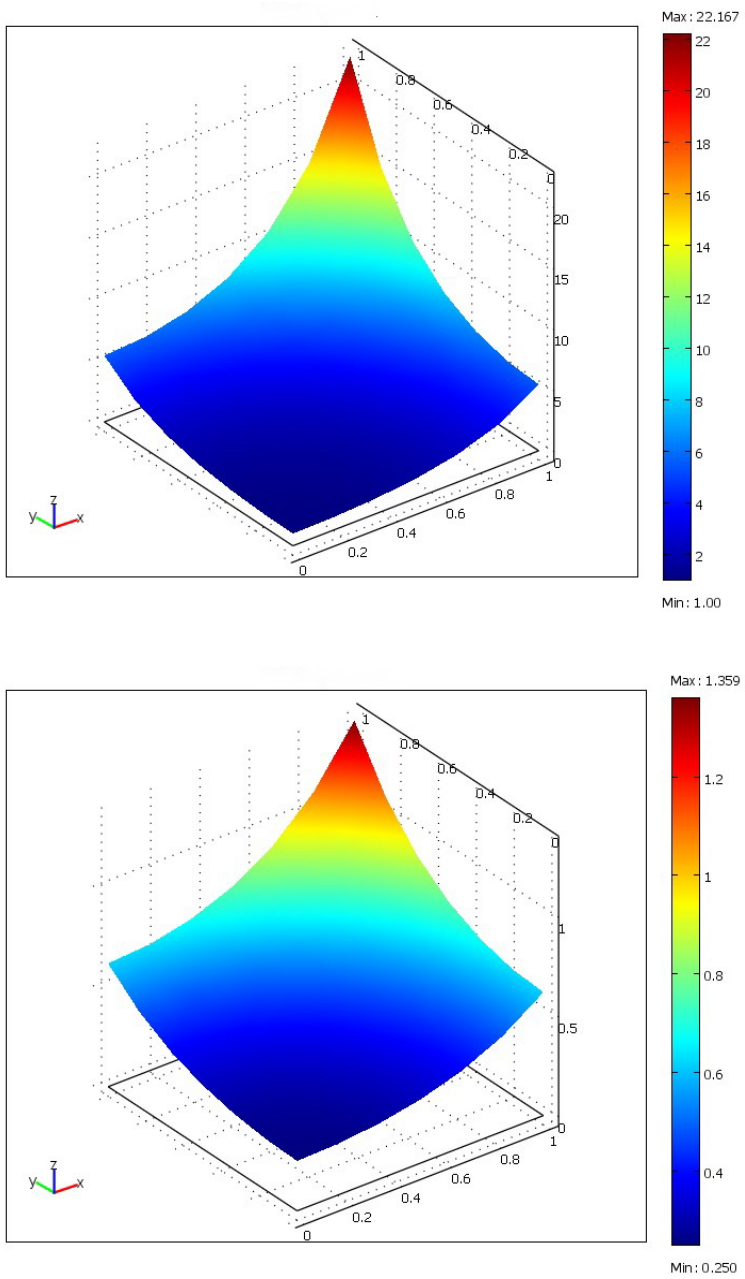


Figure 8.3: Test 8.1b: Computed α_h^ϵ at $t_M = 0.5$ (top) and $t_M = 1$. $\Delta t = 0.1$, $h = 0.05$.

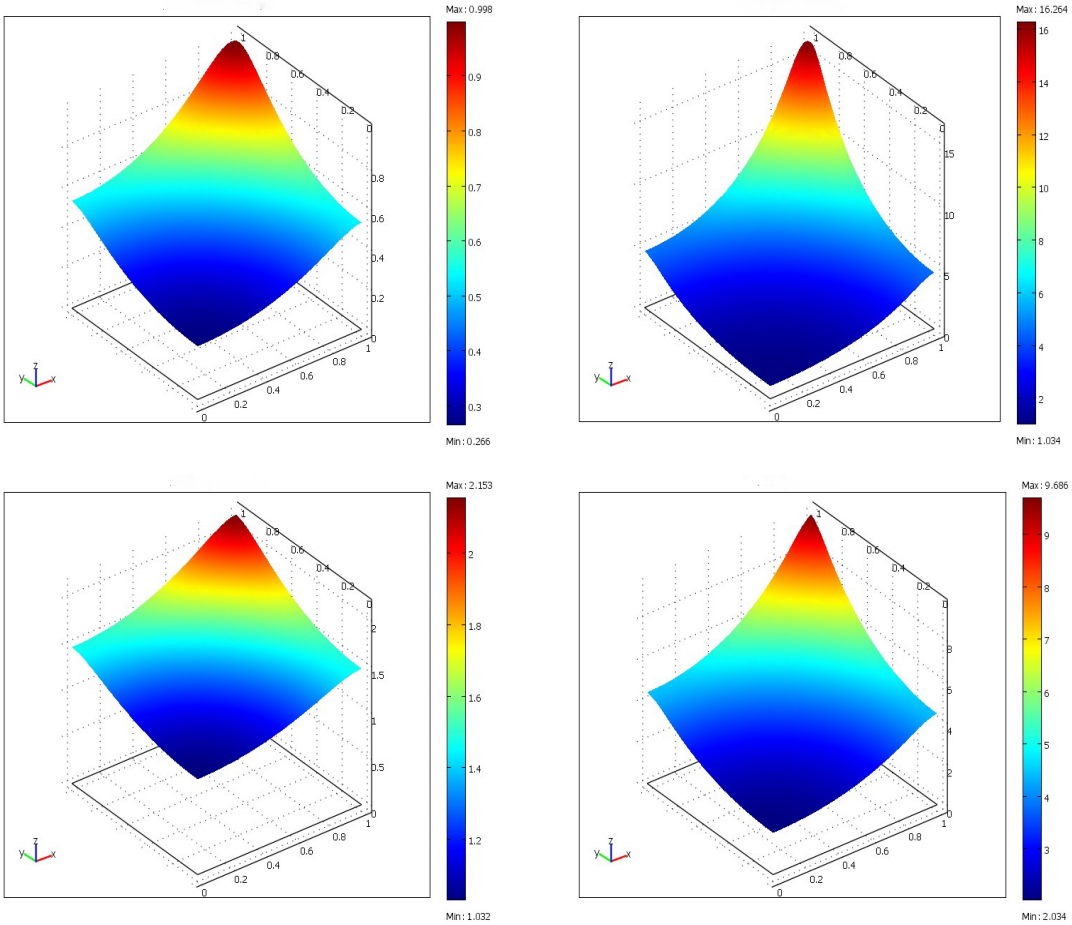


Figure 8.4: Test 8.1b: Computed determinant (top) and Laplacian (bottom) at $t_M = 0.5$ (left) and $t_M = 1$ (right). $\Delta t = 0.1$, $h = 0.05$.

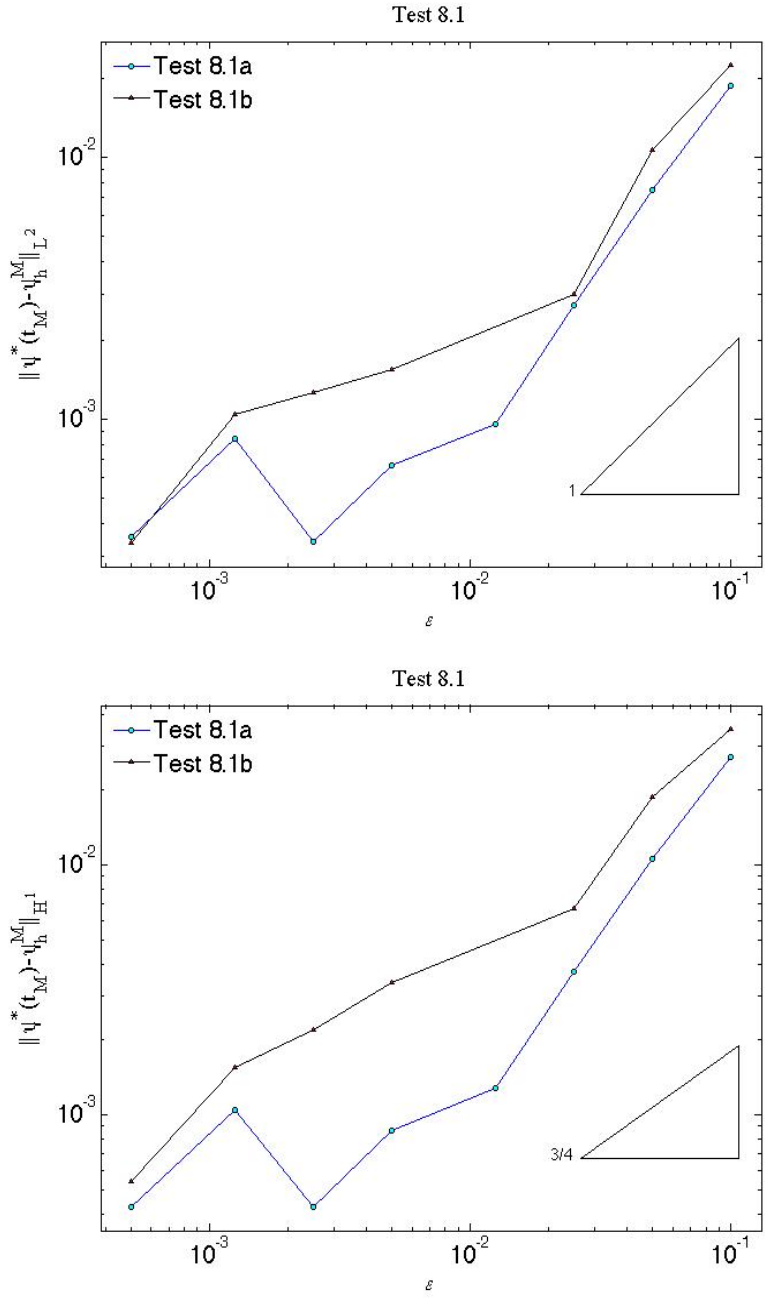


Figure 8.5: Test 8.1: Change of $\|\psi^*(t_M)-\psi_h^M\|$ w.r.t. ϵ. $h = 0.023$, $\Delta t = 0.0005$, $t_M = 0.25$.

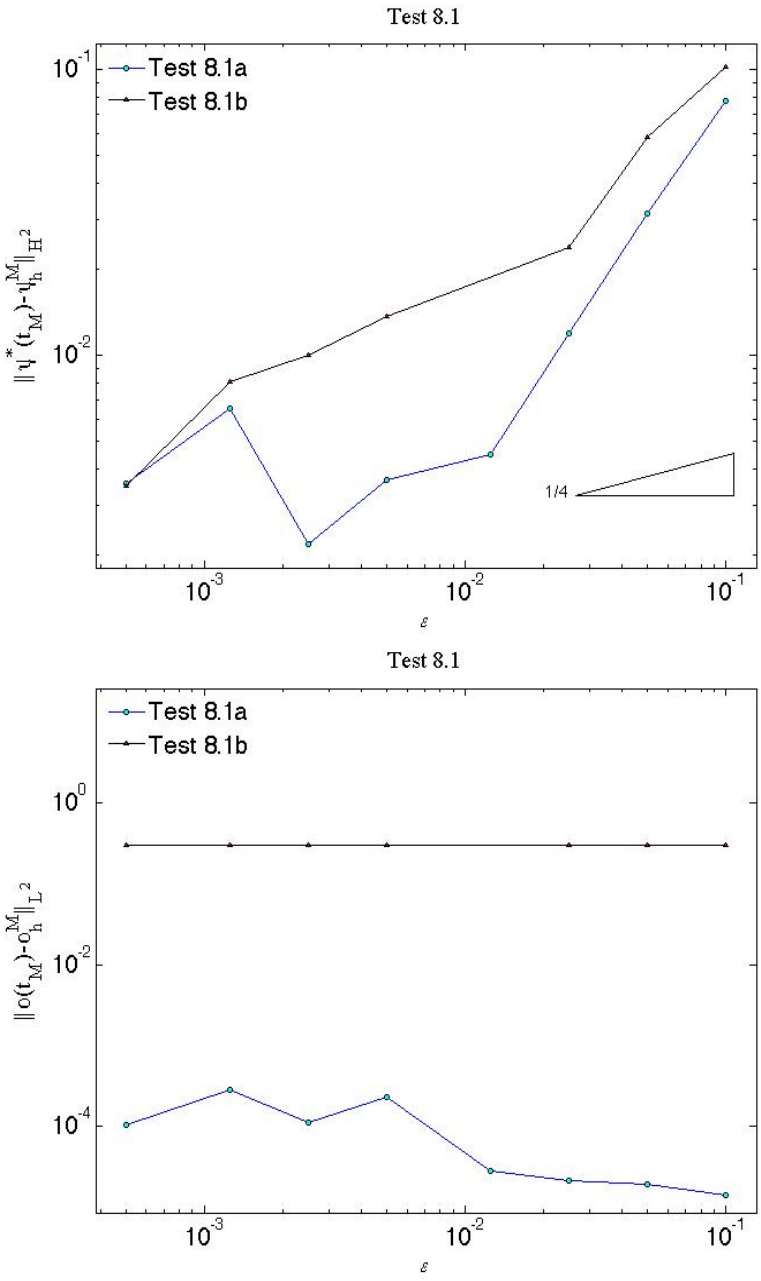


Figure 8.6: Test 8.1: Change of $\|\psi^*(t_M)-\psi_h^M\|$ w.r.t. ϵ. $h = 0.023$, $\Delta t = 0.0005$, $t_M = 0.25$.

We use the following test functions and parameters:

$$
\begin{aligned}
&\text{(a) } \psi^\epsilon = t(x_1^2 + x_2^2), && \alpha^\epsilon = 4t^2,\\
&\quad g_N^\epsilon = 2t(x_1\eta_1 + x_2\eta_2), && g_D^\epsilon = 4t^2,\\
&\quad F^\epsilon = 8t, && \phi^\epsilon = 0.\\
&\text{(b) } \psi^\epsilon = e^{t(x_1^2+x_2^2)/2}, && \alpha^\epsilon = t^2(1 + t(x_1^2 + x_2^2))e^{t(x_1^2+x_2^2)}\\
&&& \quad - \epsilon t^2 e^{t(x_1^2+x_2^2)/2}(8 + 8t(x_1^2 + x_2^2) + t^2(x_1^2 + x_2^2)^2),\\
&\quad g_N^\epsilon = te^{t(x_1^2+x_2^2)/2}(x_1\nu_1 + x_2\nu_2), && g_D^\epsilon = t^2(1 + t(x_1^2 + x_2^2))e^{t(x_1^2+x_2^2)},\\
&&& \quad - \epsilon t^2 e^{t(x_1^2+x_2^2)/2}(8 + 8t(x_1^2 + x_2^2) + t^2(x_1^2 + x_2^2)^2),
\end{aligned}
$$

$$
\begin{aligned}
F^\epsilon &= t\big(2 + 4t(x_1^2 + x_2^2) + t^2(x_1^2 + x_2^2)^2\big)e^{t(x_1^2+x_2^2)}\\
&\quad - \frac{\epsilon t}{2}e^{t(x_1^2+x_2^2)/2}\big(32 + 56(x_1^2 + x_2^2)t + 16t^2(x_1^2 + x_2^2)^2 + t^3(x_1^2 + x_2^2)^3\big),\\
\phi^\epsilon &= \Big(\big(4x_1t^2 + x_1t^3(x_1^2 + x_2^2)\big)\eta_1 + \big(4x_2t^2 + x_2t^3(x_1^2 + x_2^2)\big)\eta_2\Big)e^{t(x_1^2+x_2^2)/2}.
\end{aligned}
$$

We plot the data in Figures 8.7 and 8.8. As seen from the figures, the convergence of $\|\alpha^\epsilon(t_M) - \alpha_h^M\|_{L^2}$ and $\|\psi^\epsilon(t_M) - \psi_h^M\|$ is at least of order Δt in all norms.

Test 8.3

This test is exactly the same as in Test 8.2, but we now fix Δt and ϵ and vary h. We use the same test functions and parameters as in the previous test and plot the errors in Figures 8.9 and 8.10. We can conclude from Tests 8.2 and 8.3 that the convergence rate is dominated by Δt, as the figures show little change in the errors of $\|\alpha^\epsilon(t_M) - \alpha_h^M\|_{L^2}$ and $\|\psi^\epsilon(t_M) - \psi_h^M\|$ as h varies. These results coincide with the conclusions of Theorem 8.4.4.

Test 8.4

We again solve the same problem in Test 8.2 and use the same test functions, but we now fix ϵ and set $\Delta t = h^2$. The errors at time $t_M = 0.25$ are plotted verses Δt in Figures 8.11 and 8.12. We see that the figures are similar to Figures 8.7 and 8.8. This result is expected since we have concluded that h has very little contribution to the error estimate in relation to Δt.

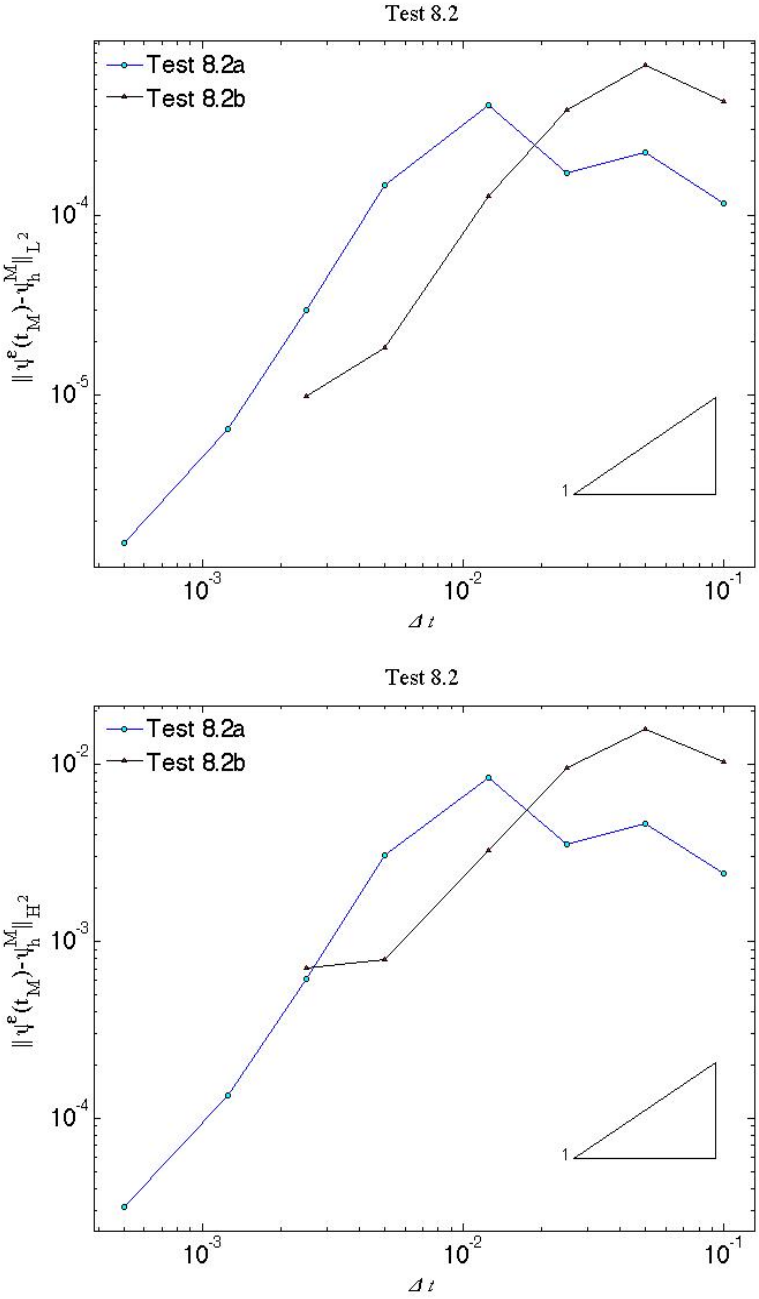


Figure 8.7: Test 8.2: Change of $\|\psi^{\epsilon}(t_M) - \psi_h^M\|$ w.r.t. Δt. $h = 0.05$, $\epsilon = 0.01$, $t_M = 0.25$.

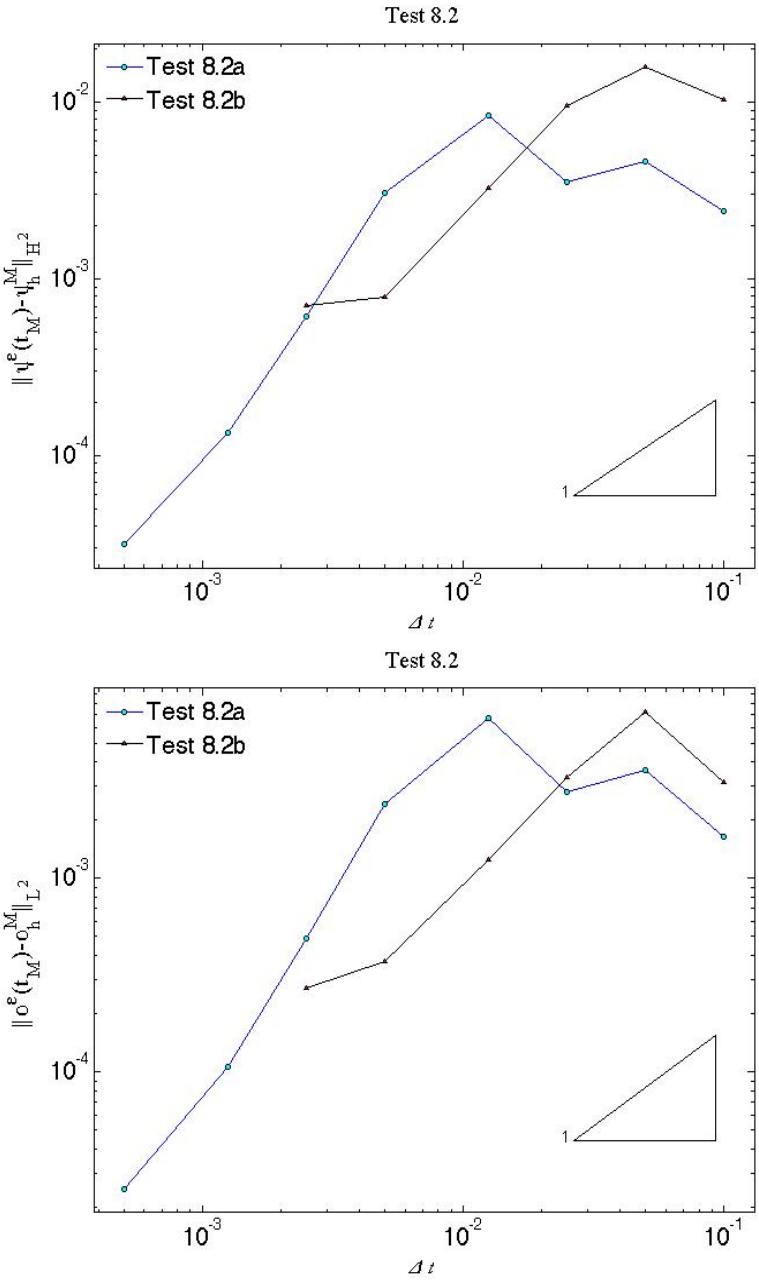


Figure 8.8: Test 8.2: Change of $\|\psi^{\epsilon}(t_M) - \psi_h^M\|$ w.r.t. Δt. $h = 0.05$, $\epsilon = 0.01$, $t_M = 0.25$.

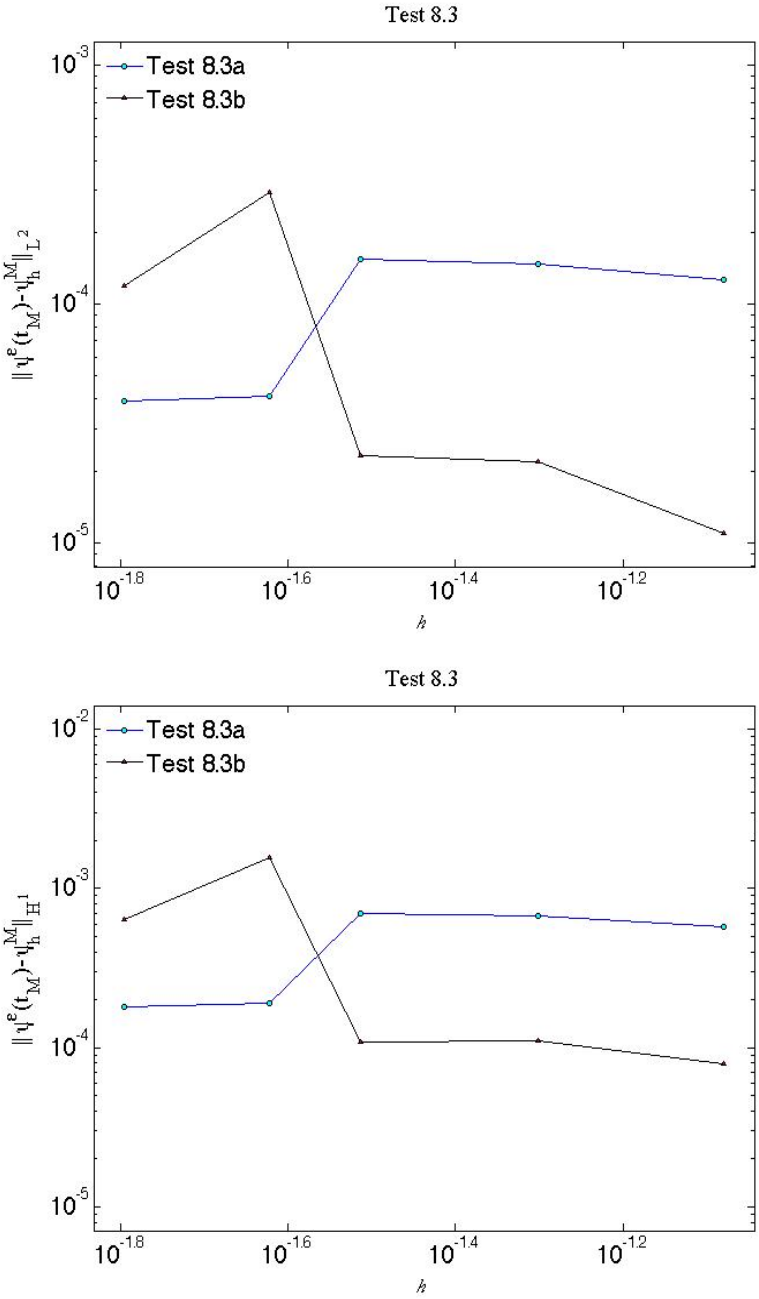


Figure 8.9: Test 8.3: Change of $\|\psi^{\epsilon}(t_M) - \psi_h^M\|$ w.r.t. h. $\epsilon = 0.01$, $\Delta t = 0.005$, $t_M = 0.25$.

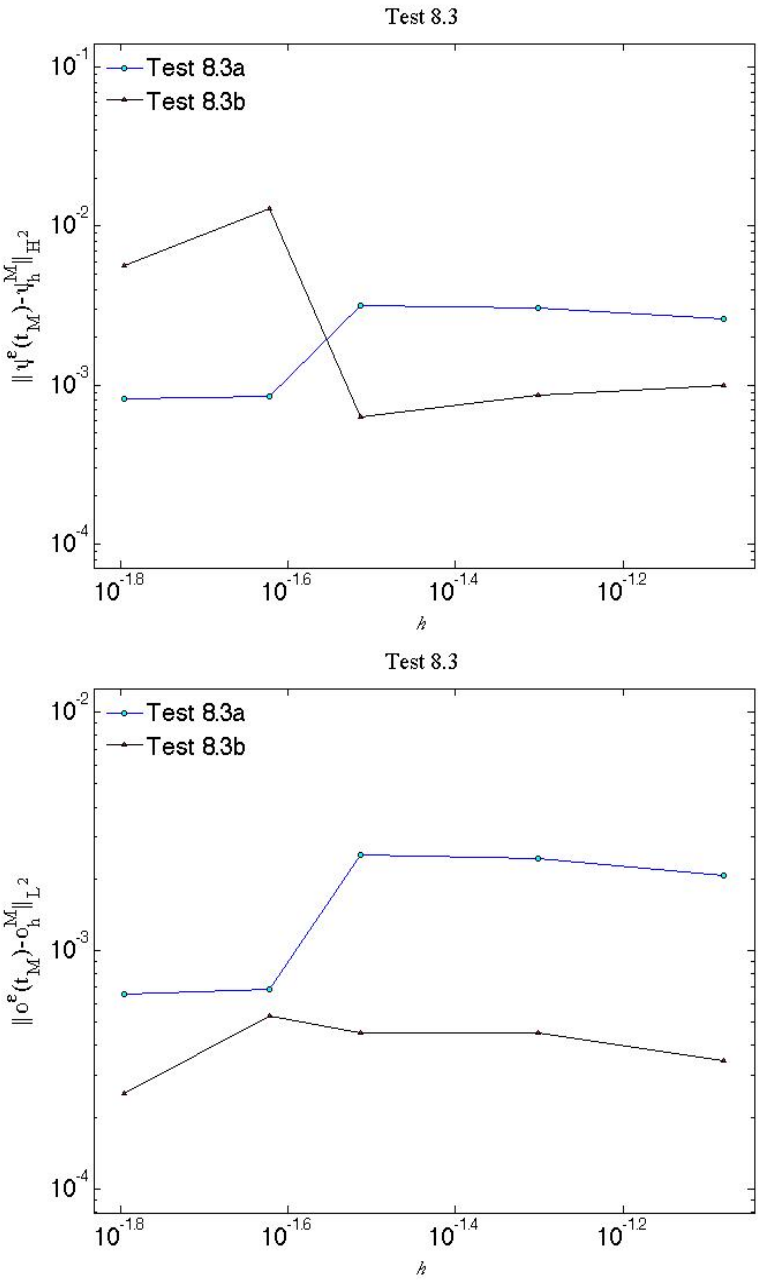


Figure 8.10: Test 8.3: Change of $\|\psi^\epsilon(t_M)-\psi_h^M\|$ w.r.t. h. $\epsilon = 0.01$, $\Delta t = 0.005$, $t_M = 0.25$.

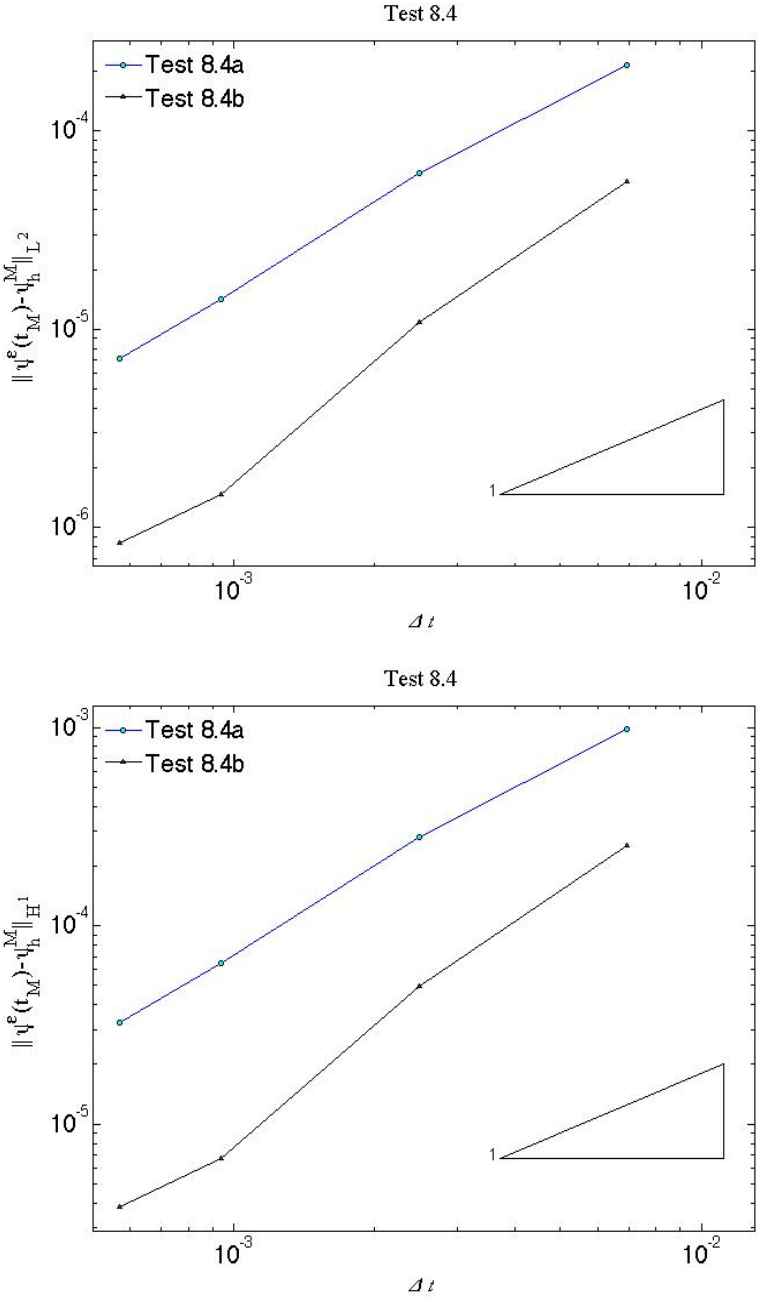


Figure 8.11: Test 8.4: Change of $\|\psi^{\epsilon}(t_M) - \psi_h^M\|$ w.r.t. $\Delta t = h^2$. $\epsilon = 0.01$, $t_M = 0.25$.

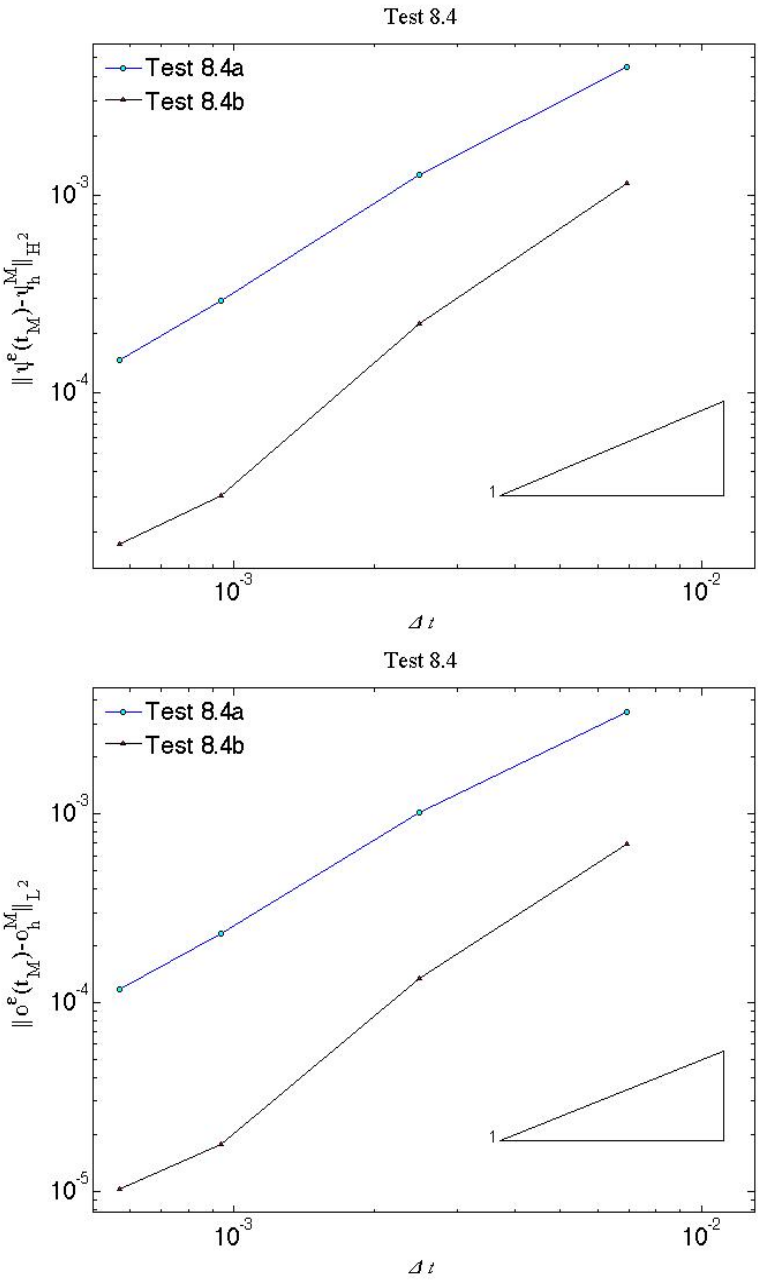


Figure 8.12: Test 8.4: Change of $\|\psi^{\epsilon}(t_M) - \psi_h^M\|$ w.r.t. $\Delta t = h^2$. $\epsilon = 0.01$, $t_M = 0.25$.

Test 8.5

For this test, we solve problem (8.41)–(8.43) with domain $U = (0, 6)^2$ and initial condition

$$\alpha_0(x) = \frac{1}{8}\chi_{[2,4]\times[2.25,3.75]}(4 - x_1)(x_1 - 2)(3.75 - x_2)(x_2 - 2.25),$$

where $\chi_{[2,4]\times[2.25,3.75]}$ denotes the characteristic function of the set $[2, 4] \times [2.25, 3.75]$. We comment that the exact solution of this problem is unknown. We plot the computed α_h^m and ψ_h^m at times $t_m = 0$, $t_m = 0.05$, and $t_m = 0.1$, and $t_m = 0.15$ in Figure 8.13 with parameters $\Delta t = 0.001$, $h = 0.05$, and $\epsilon = 0.01$. As expected, the figure shows that $\alpha_h^m > 0$ and ψ_h^m is convex for all m.

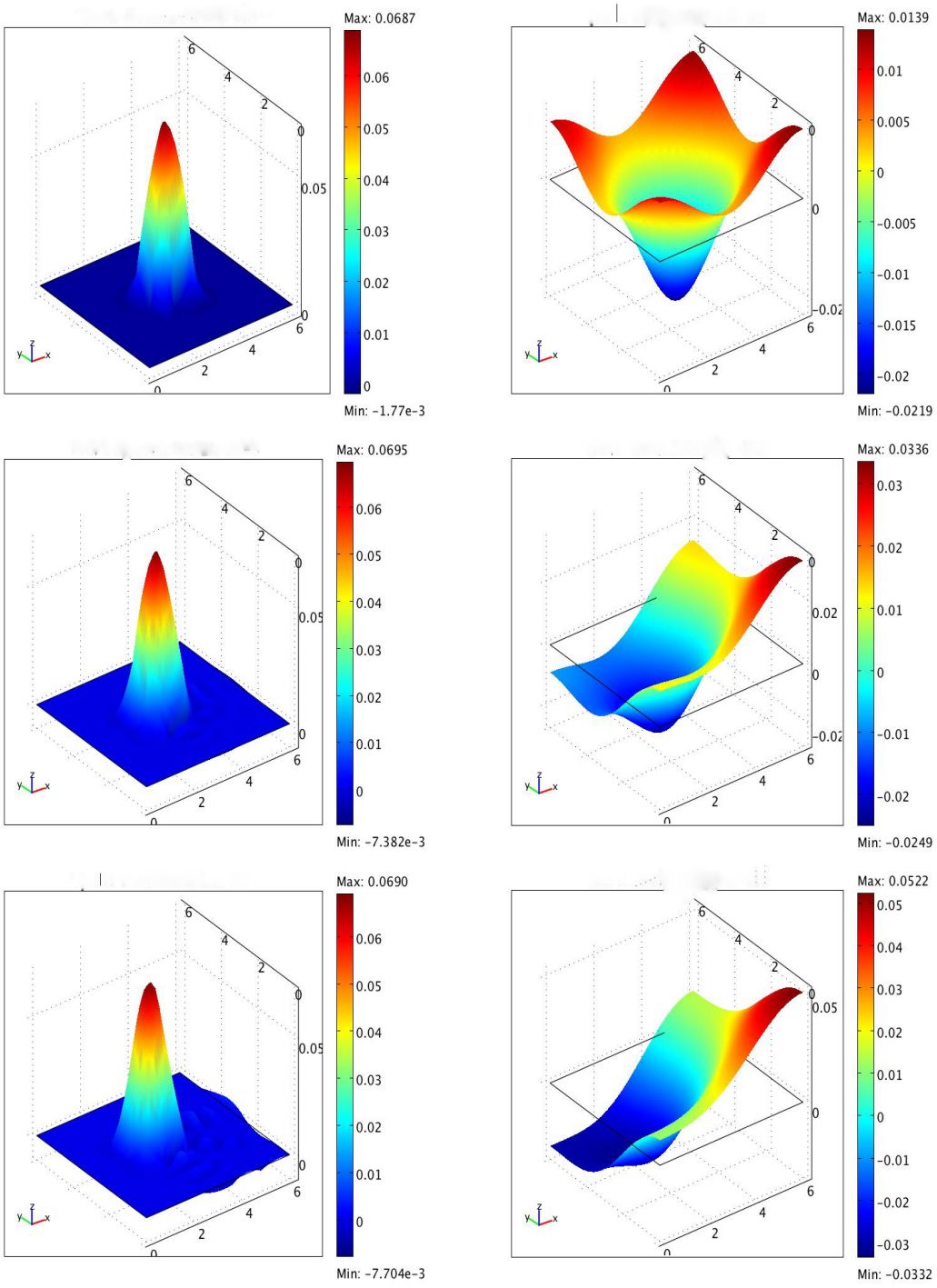


Figure 8.13: Test 8.5: Computed α_h^m (left) and ψ_h^m (right) at $t_m = 0$ (top), $t_m = 0.05$ (middle), and $t_m = 0.1$ (bottom). $\Delta t = 0.01,\ h = 0.05,\ \epsilon = 0.01$

Chapter 9

C^1 Finite Element Methods for General Fully Nonlinear Second Order PDEs

Motivated by the results obtained in Chapter 3 for the Monge-Ampère equation, we now analyze C^1 finite element approximations for general fully nonlinear second order PDEs (1.1). To do so, we employ the same vanishing moment methodology described in Chapter 2. That is, we approximate the fully nonlinear second order PDE

$$F(D^2u, Du, u, x) = 0 \qquad \text{in } \Omega, \tag{9.1}$$

$$u = g \qquad \text{on } \partial\Omega, \tag{9.2}$$

by the following fourth order quasi-linear PDE:

$$G_\epsilon(u^\epsilon) := \epsilon\Delta^2 u^\epsilon + F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x) = 0 \qquad \text{in } \Omega \ (\epsilon > 0), \tag{9.3}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{9.4}$$

$$\Delta u^\epsilon = \epsilon \qquad \text{on } \partial\Omega. \tag{9.5}$$

We assume there exists a unique solution to (9.3)–(9.5) and would like to construct and analyze finite element methods to approximate u^ϵ using a class of C^1 finite elements such as Argyris, Bogner-Fox-Schmit, and Hsieh-Clough-Tocher elements (cf. [27]).

To achieve this goal, we use the analysis in Chapter 3 as a guide. First in Section 9.1, we define the variational formulation of (9.3)–(9.5) and finite element method based upon the variational formulation. Next, we make certain assumptions about the properties of F which will play a crucial role in the analysis of the chapter. In Section 9.2, we show existence of solutions of the linearized PDE operator and prove stability and convergence

results of its finite element approximation. The main results of the chapter are found in Section 9.3, where we use a fixed point argument to simultaneously show existence, uniqueness, and convergence of the finite element approximation of (9.3)–(9.5). Finally, in Section 9.4, we apply the preceding analysis towards specific fully nonlinear second order PDEs.

9.1 Formulation of Finite Element Methods and Assumptions

Let $V := H^2(\Omega)$, and for notational convenience, we let $\|\cdot\|_V$ be the standard Sobolev norm, that is, $\|v\|_V := \|v\|_{H^2}\ \forall v \in V$. We also define the following subspace and subset of V:

$$V_0 := \{v \in V;\ v\big|_{\partial\Omega} = 0\}, \qquad V_g := \{v \in V;\ v\big|_{\partial\Omega} = g\}. \tag{9.6}$$

Multiplying (9.3) by $v \in V_0$, integrating over Ω, and integrating by parts yields

$$\epsilon(\Delta u^\epsilon, \Delta v) + (F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x), v) = \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega}. \tag{9.7}$$

Based on (9.7), we define the variational formulation of (9.3)–(9.5) as to find $u^\epsilon \in V_g$ such that

$$\epsilon(\Delta u^\epsilon, \Delta v) + (F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x), v) = \left\langle \epsilon^2, \frac{\partial v}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall v \in V_0. \tag{9.8}$$

Let $\mathcal{T}_h$ be a quasiuniform triangular or rectangular partition of Ω if $n = 2$ or a quasiuniform tetrahedral or $3D$-rectangular mesh if $n = 3$. Let $V^h \subset V$ be a conforming finite element space consisting of piecewise polynomials of degree r such that for any $v \in V \cap H^s(\Omega)$, we have

$$\inf_{v_h \in V^h} \|v - v_h\|_{H^j} \le Ch^{\ell - j}\|v\|_{H^\ell} \qquad j = 0, 1, 2, \qquad \ell = \min\{s, r+1\}. \tag{9.9}$$

Let

$$V_0^h := \{v_h \in V^h;\ v\big|_{\partial\Omega} = 0\}, \quad V_g^h := \{v_h \in V^h;\ v\big|_{\partial\Omega} = g\}.$$

Based on (9.8), we define the finite element formulation of (9.3)–(9.5) as to find $u_h^\epsilon \in V_g^h$

such that

$$\epsilon(\Delta u_h^\epsilon, \Delta v_h) + (F(D^2 u_h^\epsilon, Du_h^\epsilon, u_h^\epsilon, x), v_h) = \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall v_h \in V_0^h. \tag{9.10}$$

Let u^ϵ be the solution to (9.8) and let u_h^ϵ be the solution to (9.10). The main goal of this chapter is to prove existence and uniqueness for problem (9.10) and also derive error estimates of $u^\epsilon - u_h^\epsilon$ in the energy norm. To achieve these goals, we generalize the analysis of Chapter 3 for $F \in C^1(\mathbf{R}^{n\times n}, \mathbf{R}^n, \mathbf{R}, \Omega)$ satisfying certain structure conditions. First, we give the following additional notation:

$$F : (r, p, z, x) \in \mathbf{R}^{n\times n} \times \mathbf{R}^n \times \mathbf{R} \times \Omega \mapsto \mathbf{R}, \qquad F_r(r, p, z, x)(v) := \sum_{i,j=1}^n \frac{\partial F}{\partial r_{ij}} \frac{\partial^2 v}{\partial x_i \partial x_j},$$

$$F_p(r, p, z, x)(v) := \sum_{i=1}^n \frac{\partial F}{\partial p_i} \frac{\partial v}{\partial x_i}, \qquad F_z(r, p, z, x)(v) := \frac{\partial F}{\partial z} v,$$

$$F'(r, p, z, x)(v) := F_r(r, p, z, x)(v) + F_p(r, p, z, x)(v) + F_z(r, p, z, x)(v),$$

$$F(v) := F(D^2 v, Dv, v, x), \qquad F'[w](v) := F'(D^2 w, Dw, w, x)(v),$$

$$G'_\epsilon[w](v) := \epsilon \Delta^2 v + F'[w](v).$$

Next, it is essential in the analysis that we assume the following conditions:

(A1) There exists $\epsilon_0 > 0$ such that for all $\epsilon \in (0, \epsilon_0)$, there exists a unique solution to (9.3)–(9.5) with $u^\epsilon \in H^s(\Omega)$ $(s \geq 3)$.

(A2) The operator $(G'_\epsilon[u^\epsilon])^*$ (the adjoint of $G'_\epsilon[u^\epsilon]$) is an isomorphism from V_0 to V_0^*. That is for all $\varphi \in V_0^*$, there exists $v \in V_0$ such that

$$\langle (G'_\epsilon[u^\epsilon])^*(v), w \rangle = \langle \varphi, w \rangle \qquad \forall w \in V_0. \tag{9.11}$$

Furthermore, there exists positive constants $C_1(\epsilon), \gamma(\epsilon)$ such that the following Gårding inequality holds:

$$\langle G'_\epsilon[u^\epsilon](v), v \rangle \geq C_1(\epsilon) \|v\|_V^2 - \gamma(\epsilon) \|v\|_{L^2}^2 \qquad \forall v \in V_0, \tag{9.12}$$

and there exists $C_2(\epsilon) > 0$ such that

$$\|F'[u^\epsilon]\|_{VV^*} \leq C_2(\epsilon), \tag{9.13}$$

where

$$\|F'[u^\epsilon]\|_{VV^*} := \sup_{v\in V} \frac{\|F'[u^\epsilon](v)\|_{V^*}}{\|v\|_V} := \sup_{v\in V}\sup_{w\in V} \frac{\langle F'[u^\epsilon](v), w\rangle}{\|v\|_V\|w\|_V}.$$

Moreover, there exists $p > 2$ and $C_R(\epsilon) > 0$ such that if $\varphi \in L^2(\Omega)$ and $v \in V_0$ satisfies (9.11), then $v \in H^p(\Omega)$ and

$$\|v\|_{H^p} \le C_R(\epsilon)\|\varphi\|_{L^2}.$$

(A3) There exists a Banach space Y with $V^h \subset Y \subset V$ and a constant $C > 0$ such that

$$\sup_{y\in Y} \frac{\|F'[y]\|_{VV^*}}{\|y\|_Y} \le C, \tag{9.14}$$

where

$$\|F'[y]\|_{VV^*} := \sup_{v\in V} \frac{\|F'[y](v)\|_{V^*}}{\|v\|_V} := \sup_{v\in V}\sup_{w\in V} \frac{\langle F'[y](v), w\rangle}{\|v\|_V\|w\|_V}.$$

(A4) There exist $\tilde{u}_h^\epsilon \in V_g^h$ and constants $C_3(\epsilon),\ C_4(\epsilon) > 0$ independent of h such that

$$\|u^\epsilon - \tilde{u}_h^\epsilon\|_V \le C_3(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \qquad \ell = \min\{r+1, s\}, \tag{9.15}$$

$$\|\tilde{u}_h^\epsilon\|_Y \le C_4(\epsilon)\|u^\epsilon\|_Y. \tag{9.16}$$

(A5) There exists a constant $\delta > 0$, such that for any $v_h \in V_g^h$ with $\|u_h^\epsilon - w_h\|_V \le \delta$, there holds

$$\|F'[u^\epsilon] - F'[w_h]\|_{VV^*} \le L(\epsilon, h)\|u^\epsilon - w_h\|_V, \tag{9.17}$$

where $L(\epsilon, h) = L(\epsilon, h, n, \Omega, \delta, u^\epsilon)$.

9.2 Analysis of the Linearized Problem and its Finite Element Approximation

To construct the necessary tools to analyze finite element method (9.10), we first study finite element approximation of the linearization of (9.3).

9.2.1 Linearization

For given $\varphi \in V_0^*$ and $\psi \in H^{-\frac{1}{2}}(\partial\Omega)$, we consider the following linear problem:

$$G'_\epsilon[u^\epsilon](v) = \varphi \qquad \text{in } \Omega, \tag{9.18}$$

$$v = 0 \qquad \text{on } \partial\Omega, \tag{9.19}$$

$$\Delta v = \psi \qquad \text{on } \partial\Omega. \tag{9.20}$$

Multiplying the equation $G'_\epsilon[u^\epsilon]$ by $w \in V_0$, integrating over Ω, and integrating by parts, we obtain

$$\langle G'_\epsilon[u^\epsilon](v), w\rangle = \epsilon(\Delta v, \Delta w) + \langle F'[u^\epsilon](v), w\rangle - \epsilon \left\langle \Delta v, \frac{\partial w}{\partial \eta} \right\rangle_{\partial\Omega}.$$

Based on this calculation, we define the weak formulation of (9.18)–(9.20) as to find $v \in V_0$ such that

$$B_\epsilon[v, w] = \langle \varphi, w\rangle + \epsilon \left\langle \psi, \frac{\partial w}{\partial \eta} \right\rangle_{\partial\Omega} \qquad \forall w \in V_0, \tag{9.21}$$

where

$$B_\epsilon[v, w] := \epsilon(\Delta v, \Delta w) + \langle F'[u^\epsilon](v), w\rangle.$$

In view of assumptions (A1)–(A2), we immediately have the following theorem.

Theorem 9.2.1. *Assume assumptions* (A1)–(A2) *hold. Then there exists a unique solution* $v \in V_0$ *to* (9.21)*. Furthermore, there exists* $C_5(\epsilon) > 0$ *such that*

$$\|v\|_V^2 \le C_5(\epsilon)\big(\|\varphi\|_{V^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\big). \tag{9.22}$$

Proof. From the Gårding-type inequality (9.12) and the fact $(G'_\epsilon[u^\epsilon])^*$ is injective on V_0, it follows that $G'_\epsilon[u^\epsilon]$ is a an isomorphism from V_0 to V_0^* using a Fredholm alternative argument [1, Theorem 8.5].

We now claim that there exists $C(\epsilon)$ such that $\|v\|_{L^2} \le C(\epsilon)\|\varphi\|_{V^*}$. If not, there would exist sequences $\{\varphi_k\}_{k=1}^\infty \subset V_0^*$ and $\{v_k\}_{k=1}^\infty \subset V_0$ such that

$$\langle G'_\epsilon[u^\epsilon](v_k), w\rangle = \langle \varphi_k, w\rangle \qquad w \in V_0,$$

but

$$\|v_k\|_{L^2} > k\|\varphi_k\|_{V^*}.$$

Without loss of generality, we may as well suppose $\|v_k\|_{L^2} = 1$ (and therefore $\|\varphi_k\|_{V^*} \to 0$ as $k \to \infty$). In light of (9.12), $\{v_k\}_{k=1}^{\infty}$ is bounded in V_0. Therefore by compactness, there exists a subsequence $\{v_{k_j}\}_{j=1}^{\infty}$ and $v \in V_0$ such that

$$v_{k_j} \rightharpoonup v \quad \text{weakly in } V_0, \tag{9.23}$$

$$v_{k_j} \to v \quad \text{in } H_0^1(\Omega). \tag{9.24}$$

Therefore,

$$\langle G'_\epsilon[u^\epsilon](v), w\rangle = 0 \quad \forall w \in V_0,$$

Since $G'_\epsilon[u^\epsilon]$ is an isomorphism, $v \equiv 0$. However (9.24) implies that $\|v\|_{L^2} = 1$, a contradiction.

Hence there exists $C(\epsilon)$ such that $\|v\|_{L^2} \le C(\epsilon)\|\varphi\|_{V^*}$, and therefore by (9.12) and a trace inequality, we have

$$\begin{aligned} C_1(\epsilon)\|v\|_V^2 &\le \epsilon\|\Delta v\|_{L^2}^2 + \langle F'[u^\epsilon](v), v\rangle + \gamma(\epsilon)\|v\|_{L^2}^2 \\ &= B_\epsilon[v, v] + \gamma(\epsilon)\|v\|_{L^2}^2 \\ &= \langle \varphi, v\rangle + \epsilon\left\langle \psi, \frac{\partial v}{\partial \eta}\right\rangle_{\partial\Omega} + \gamma(\epsilon)\|v\|_{L^2}^2 \\ &\le \left(\|\varphi\|_{V^*} + C\epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)} + \gamma(\epsilon)C(\epsilon)\|\varphi\|_{V^*}\right)\|v\|_V. \end{aligned}$$

Dividing by $C_1(\epsilon)\|v\|_V$, we obtain (9.22) with $C_5(\epsilon) = C(\epsilon)C_1^{-1}(\epsilon)\gamma(\epsilon)$. □

9.2.2 Finite Element Approximation

Let V_0^h be one of the finite dimensional subspaces of V_0 as defined in Section 9.1. Based on the variational formulation (9.21), we define the finite element method for (9.18)–(9.20) as to find $v_h \in V_0^h$ such that

$$B_\epsilon[v_h, w_h] = \langle \varphi, w_h\rangle + \epsilon\left\langle \psi, \frac{\partial w_h}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall w_h \in V_0^h. \tag{9.25}$$

Using a modification of the well-known Schatz's argument (cf. [17, Theorem 5.7.6]), we obtain the following result.

Theorem 9.2.2. *Let assumptions* (A1)–(A2) *hold and suppose that* $v \in H^s(\Omega)$ $(s \ge 3)$ *is the unique solution to* (9.21). *Then for* $h \le h_0$, *there exists a unique solution* $v_h \in V_0^h$ *to* (9.25), *where*

$$h_0 = C\left(\frac{C_1(\epsilon)}{C_2^2(\epsilon)C_R^2(\epsilon)\gamma(\epsilon)}\right)^{\frac{1}{2p-4}}. \tag{9.26}$$

Furthermore, there holds the following inequalities:

$$\|v_h\|_V \le C_6(\epsilon)\big(\|\varphi\|_{V^*} + \epsilon\|\psi\|_{H^{-\frac{1}{2}}(\partial\Omega)}\big), \tag{9.27}$$

$$\|v - v_h\|_V \le C_7(\epsilon)h^{\ell-2}\|v\|_{H^\ell}, \tag{9.28}$$

$$\|v - v_h\|_{L^2} \le C_8(\epsilon)h^{\ell+p-4}\|v\|_{H^\ell}, \tag{9.29}$$

where

$$C_7(\epsilon) = CC_1^{-1}(\epsilon)C_2(\epsilon), \qquad C_8(\epsilon) = CC_1^{-1}(\epsilon)C_2^2(\epsilon)C_R(\epsilon), \qquad \ell = \min\{s, r+1\}.$$

Proof. To show existence, we begin by deriving estimates for a solution v_h to (9.25) that may exist. We start with the error equation:

$$B_\epsilon[v - v_h, w_h] = 0 \qquad \forall w_h \in V_0^h.$$

Then using (9.12) and (9.13), we have for any $w_h \in V_0^h$

$$\begin{aligned}
&C_1(\epsilon)\|v - v_h\|_V^2 \\
&\quad\le \epsilon\|\Delta(v - v_h)\|_{L^2}^2 + \langle F'[u^\epsilon](v - v_h), v - v_h\rangle + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \\
&\quad= B_\epsilon[v - v_h, v - v_h] + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \\
&\quad= B_\epsilon[v - v_h, v - w_h] + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \\
&\quad\le \epsilon\|\Delta(v - v_h)\|_{L^2}\|(\Delta(v - w_h)\|_{L^2} \\
&\qquad\quad + \|F'[u^\epsilon]\|_{VV^*}\|v - v_h\|_V\|v - w_h\|_V + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \\
&\quad\le CC_2(\epsilon)\|v - v_h\|_V\|v - w_h\|_V + \gamma(\epsilon)\|v - v_h\|_{L^2}^2.
\end{aligned}$$

Thus, by (9.9)

$$C_1(\epsilon)\|v - v_h\|_V^2 \le CC_1^{-1}C_2^2(\epsilon)h^{2\ell-4}\|v\|_{H^\ell}^2 + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \tag{9.30}$$

Next, we let $w \in V_0 \cap H^p(\Omega)$ be the solution to the following problem:

$$\langle \big(G_\epsilon'[u^\epsilon]\big)^*(w), z\rangle = (v - v_h, z) \qquad \forall z \in V_0,$$

with

$$\|w\|_{H^p} \le C_R(\epsilon)\|v - v_h\|_{L^2} \tag{9.31}$$

We then have for any $w_h \in V_0^h$

$$\begin{aligned}\|v - v_h\|_{L^2}^2 &= \langle (G'_\epsilon[u^\epsilon])^*(w), (v - v_h)\rangle \\ &= \langle G'_\epsilon[u^\epsilon](v - v_h), w\rangle \\ &= B_\epsilon[v - v_h, w] \\ &= B_\epsilon[v - v_h, w - w_h] \\ &\le CC_2(\epsilon)\|v - v_h\|_V \|w - w_h\|_V.\end{aligned}$$

Consequently from (9.9) and (9.31)

$$\begin{aligned}\|v - v_h\|_{L^2}^2 &\le CC_2(\epsilon)h^{p-2}\|v - v_h\|_V \|w\|_{H^p} \\ &\le CC_2(\epsilon)C_R(\epsilon)h^{p-2}\|v - v_h\|_V \|v - v_h\|_{L^2},\end{aligned}$$

and thus,

$$\|v - v_h\|_{L^2} \le CC_2(\epsilon)C_R(\epsilon)h^{p-2}\|v - v_h\|_V. \tag{9.32}$$

Applying the inequality (9.32) into (9.30) gives us

$$\begin{aligned}C_1(\epsilon)\|v - v_h\|_V^2 &\le CC_1^{-1}(\epsilon)C_2^2(\epsilon)h^{2\ell-4}\|v\|_{H^\ell}^2 + \gamma(\epsilon)\|v - v_h\|_{L^2}^2 \\ &\le CC_1^{-1}(\epsilon)C_2^2(\epsilon)h^{2\ell-4}\|v\|_{H^\ell}^2 + CC_2^2(\epsilon)C_R^2(\epsilon)\gamma(\epsilon)h^{2p-4}\|v - v_h\|_V^2.\end{aligned}$$

Thus, for $h \le h_0$

$$C_1(\epsilon)\|v - v_h\|_V^2 \le CC_1^{-1}(\epsilon)C_2^2(\epsilon)h^{2\ell-4}\|v\|_{H^\ell}^2,$$

and therefore (cf. (9.32))

$$\begin{aligned}\|v - v_h\|_V &\le CC_1^{-1}(\epsilon)C_2(\epsilon)h^{\ell-2}\|v\|_{H^\ell}, \\ \|v - v_h\|_{L^2} &\le CC_1^{-1}(\epsilon)C_2^2(\epsilon)C_R(\epsilon)h^{\ell+p-4}\|v\|_{H^\ell}.\end{aligned}$$

So far, we have been under the assumption that there exists a solution v_h. We now consider the question of existence and uniqueness. First since the problem is linear and in a finite dimensional setting, existence and uniqueness are equivalent. Now suppose $\varphi \equiv 0$, $\psi \equiv 0$. In light of (9.22), we have $v \equiv 0$, and therefore, (9.28) implies $v_h \equiv 0$ as well provided h is sufficiently small. In particular, this means that (9.25) has unique solutions for $h \le h_0$. Finally, (9.27) follows from (9.22) and (9.28). □

9.3 Finite Element Approximation of (9.21)

In this section, we give the main results of the chapter. First, we define an operator $T : V_g^h \mapsto V_g^h$ such that for a given $v_h \in V_g^h$, $T(v_h)$ is the solution to the following linear problem:

$$B_\epsilon[v_h - T(v_h), w_h] = \epsilon(\Delta v_h, \Delta w_h) + \langle F(v_h), w_h\rangle - \left\langle \epsilon^2, \frac{\partial w_h}{\partial \eta} \right\rangle_{\partial\Omega} \quad \forall w_h \in V_0^h. \tag{9.33}$$

In view of Theorem 9.2.2, T is well-defined provided assumptions (A1)–(A2) hold and $h \le h_0$. The goal now is to show that $T(\cdot)$ has a unique fixed point in a neighborhood of u^ϵ, which will be a solution to (9.10). To this end, we set

$$\mathbb{B}_h(\rho) := \{v_h \in V_g^h;\ \|v_h - \tilde{u}_h^\epsilon\|_V \le \rho\},$$

where $\tilde{u}_h^\epsilon$ is defined by (A4).

Let $\ell = \min\{s, r+1\}$, where r is the polynomial degree of the finite element space V^h and s is defined by (A1). The following lemma bounds the distance between the center of $\mathbb{B}_h$ and $T(\tilde{u}_h^\epsilon)$.

Lemma 9.3.1. *Let assumptions* (A1)–(A4) *hold. Then there exists an* $h_1 > 0$ *such that for* $h \le \min\{h_0, h_1\}$,

$$\|\tilde{u}_h^\epsilon - T(\tilde{u}_h^\epsilon)\|_V \le C_9(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell}, \tag{9.34}$$

where $C_9(\epsilon) = CC_3(\epsilon)C_4(\epsilon)C_6(\epsilon)\|u^\epsilon\|_Y$.

Proof. Let $\alpha^\epsilon = \tilde{u}_h^\epsilon - u^\epsilon$. Using the definition of T and the mean value theorem, we have for any $z_h \in V_0^h$

$$\begin{aligned} B_\epsilon[\tilde{u}_h^\epsilon - T(\tilde{u}_h^\epsilon), z_h] &= \epsilon(\Delta \tilde{u}_h^\epsilon, \Delta z_h) + (F(\tilde{u}_h^\epsilon), z_h) - \left\langle \epsilon^2, \frac{\partial z_h}{\partial \eta} \right\rangle_{\partial\Omega} \\ &= \epsilon(\Delta \alpha^\epsilon, \Delta z_h) + (F(\tilde{u}_h^\epsilon) - F(u^\epsilon), z_h) \\ &= \epsilon(\Delta \alpha^\epsilon, \Delta z_h) + \langle F'[\xi](\alpha^\epsilon), z_h\rangle, \end{aligned} \tag{9.35}$$

where $\xi = \tilde{u}_h^\epsilon - \tau\alpha^\epsilon$ for some $\tau \in [0,1]$.

In light of (9.27), we have

$$\|\tilde{u}_h^\epsilon - T(\tilde{u}_h^\epsilon)\|_V \le C_6(\epsilon)\|\varphi\|_{V^*} \tag{9.36}$$

where $\varphi = \epsilon\Delta_h^2\alpha^\epsilon + F'[\xi](\alpha^\epsilon)$ and Δ_h^2 is the discrete biharmonic operator, that is,

$$\langle \Delta_h^2 w_h, z_h\rangle = (\Delta w_h, \Delta z_h) - \left\langle \Delta w_h, \frac{\partial z_h}{\partial \eta}\right\rangle_{\partial\Omega} \qquad \forall w_h, z_h \in V_0^h.$$

Using (A3) and (A4), we have

$$\begin{aligned}\langle \varphi, z_h\rangle &= \epsilon(\Delta\alpha^\epsilon, \Delta z_h) + \langle F'[\xi](\alpha^\epsilon), z_h\rangle \\ &\le \epsilon\|\Delta\alpha^\epsilon\|_{L^2}\|\Delta z_h\|_{L^2} + \|F'[\xi](\alpha^\epsilon)\|_{V^*}\|z_h\|_V \\ &\le (\epsilon + C)\|\xi\|_Y\|\alpha^\epsilon\|_V\|z_h\|_V \\ &\le CC_4(\epsilon)\|u^\epsilon\|_Y\|\alpha^\epsilon\|_V\|z_h\|_V.\end{aligned} \tag{9.37}$$

Next using a density argument, we can choose h_1 such that for $h \le h_1$,

$$\|\varphi\|_{V^*} = \sup_{z\in V_0}\frac{\langle\varphi, z\rangle}{\|z\|_V} \le 2\sup_{z_h\in V_0^h}\frac{\langle\varphi, z_h\rangle}{\|z_h\|_V}.$$

Therefore by (9.36) and (9.37),

$$\begin{aligned}\|\tilde{u}_h^\epsilon - T(\tilde{u}_h^\epsilon)\|_V &\le C_6(\epsilon)\|\varphi\|_{V^*} \\ &\le CC_4(\epsilon)C_6(\epsilon)\|u^\epsilon\|_Y\|\alpha^\epsilon\|_V \\ &\le CC_3(\epsilon)C_4(\epsilon)C_6(\epsilon)h^{\ell-2}\|u^\epsilon\|_Y\|u^\epsilon\|_{H^\ell}.\end{aligned}$$

□

Lemma 9.3.2. *Suppose assumptions* (A1)–(A5) *hold. Suppose further that* $L(\epsilon, h) = o(h^{2-\ell})$. *Then there exists an* $h_2 > 0$ *such that for* $h \le \min\{h_0, h_1, h_2\}$, *the operator* T *is a contracting mapping in the ball* $\mathbb{B}_h(\rho_0)$, *where* $\rho_0 = O\left(\min\{\delta, (C_6(\epsilon)L(\epsilon, h))^{-1}\}\right)$.

Proof. Using the definition of T, we have for any $v_h, w_h \in \mathbb{B}_h(\rho_0),\ z_h \in V_0^h$,

$$\begin{aligned}B_\epsilon[T(v_h) - T(w_h), z_h] &= B_\epsilon[v_h, z_h] - B_\epsilon[w_h, z_h] + \epsilon(\Delta(w_h - v_h), \Delta z_h) \\ &\quad + \langle F(w_h) - F(v_h), z_h\rangle \\ &= \langle F'[u^\epsilon](v_h - w_h), z_h\rangle + \langle F(w_h) - F(v_h), z_h\rangle.\end{aligned}$$

Using the mean value theorem, we obtain

$$\begin{aligned}B_\epsilon[T(v_h) - T(w_h), z_h] &= \langle F'[u^\epsilon](v_h - w_h), z_h\rangle + \langle F(w_h) - F(v_h), z_h\rangle \\ &= \langle (F'[u^\epsilon] - F'[\xi])(v_h - w_h), z_h\rangle,\end{aligned}$$

where $\xi = w_h + \tau(v_h - w_h)$ for some $\tau \in [0,1]$. Here, we have abused the notation of ξ, defining it differently in two different proofs.

By (9.27), there holds

$$\|T(v_h) - T(w_h)\|_V \le C_6(\epsilon)\|\varphi\|_{V^*}, \tag{9.38}$$

where $\varphi = (F'[u^\epsilon] - F'[\xi])(v_h - w_h)$.

Noting $\rho_0 \le \delta$, it follows from (A5) that for any $z_h \in V_0^h$

$$\begin{aligned}\langle \varphi, z_h\rangle &= \langle F'[u^\epsilon] - F'[\xi](v_h - w_h), z_h\rangle \\ &\le \|F'[u^\epsilon] - F'[\xi]\|_{VV^*}\|v_h - w_h\|_V\|z_h\|_V \\ &\le L(\epsilon,h)\|u^\epsilon - \xi\|_V\|v_h - w_h\|_V\|z_h\|_V.\end{aligned}$$

Next using the triangle inequality, we have

$$\begin{aligned}\|u^\epsilon - \xi\|_V &\le \|u^\epsilon - w_h\|_V + \|v_h - w_h\|_V \\ &\le \|u^\epsilon - \tilde{u}_h^\epsilon\|_V + 2\|\tilde{u}_h^\epsilon - w_h\|_V + \|v_h - \tilde{u}_h^\epsilon\|_V \\ &\le Ch^{\ell-2}\|u^\epsilon\|_{H^\ell} + 3\rho_0,\end{aligned}$$

and therefore,

$$\langle \varphi, z_h\rangle \le CL(\epsilon,h)\big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0\big)\|v_h - w_h\|_V\|z_h\|_V. \tag{9.39}$$

It follows from (9.38) and (9.39) that for $h \le \min\{h_0, h_1\}$

$$\begin{aligned}\|T(v_h) - T(w_h)\|_V &\le C_6(\epsilon)\|\varphi\|_{V^*} \\ &\le 2C_6(\epsilon)\sup_{z_h\in V_0^h}\frac{\langle\varphi, z_h\rangle}{\|z_h\|_V} \\ &\le CC_6(\epsilon)L(\epsilon,h)\big(h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_0\big)\|v_h - w_h\|_V\end{aligned}$$

Choosing h_2 such that

$$h_2 = O\left(C_6(\epsilon)L(\epsilon,h_2)\|u^\epsilon\|_{H^\ell}\right)^{\frac{1}{2-\ell}}$$

we have

$$\|T(v_h) - T(w_h)\|_V \le \frac{1}{2}\|v_h - w_h\|_V.$$

□

With these two lemmas in hand, we can now derive the main results of the chapter.

Theorem 9.3.3. *Under the same hypotheses of Lemma 9.3.2, there exists $h_3 > 0$ such that for $h \leq \min\{h_0, h_1, h_2, h_3\}$, there exists a unique solution to* (9.10)*. Furthermore, there holds the following error estimate:*

$$\|u^\epsilon - u_h^\epsilon\|_V \leq C_{10}(\epsilon) h^{\ell-2} \|u^\epsilon\|_{H^\ell}, \tag{9.40}$$

with $C_{10}(\epsilon) = CC_9(\epsilon) = C_3(\epsilon)C_4(\epsilon)C_6(\epsilon)\|u^\epsilon\|_Y$. Moreover, there exists $h_4 > 0$ such that for $h \leq \min\{h_0, h_1, h_2, h_3, h_4\}$

$$\|u^\epsilon - u_h^\epsilon\|_{L^2} \leq C_{11}(\epsilon)\Big(C_2(\epsilon)h^{\ell+p-4}\|u^\epsilon\|_{H^\ell} + L(\epsilon, h)C_{10}(\epsilon)h^{2\ell-4}\|u^\epsilon\|_{H^\ell}^2\Big), \tag{9.41}$$

where $C_{11}(\epsilon) = CC_{10}(\epsilon)C_R(\epsilon)$.

Proof. Let $\rho_1 := 2C_9(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell}$, and choose $h_3 > 0$ such that

$$h_3 = O\left(\frac{\min\{\delta, (C_6(\epsilon)L(\epsilon, h_3))^{-1}\}}{C_9(\epsilon)\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}.$$

Then for $h \leq h_3$, we have $\rho_1 \leq \rho_0$.

Thus for $h \leq \min\{h_0, h_1, h_2, h_3\}$, we use Lemmas 9.3.1 and 9.3.2 to conclude that for any $v_h \in \mathbb{B}_h(\rho_1)$,

$$\begin{aligned}
\|\tilde{u}_h^\epsilon - T(v_h)\|_V &\leq \|\tilde{u}_h^\epsilon - T(\tilde{u}_h^\epsilon)\|_V + \|T(\tilde{u}_h^\epsilon) - T(v_h)\|_V \\
&\leq C_9(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} + \frac{1}{2}\|\tilde{u}_h^\epsilon - v_h\|_V \\
&\leq \frac{\rho_1}{2} + \frac{\rho_1}{2} = \rho_1.
\end{aligned}$$

Hence, T maps $\mathbb{B}_h(\rho_1)$ into $\mathbb{B}_h(\rho_1)$. Since T is a contraction mapping in $\mathbb{B}_h(\rho_1)$, T has a unique fixed point in $\mathbb{B}_h(\rho_1)$, which is the unique solution to (9.10). To derive the error estimate (9.40), we use the triangle inequality to get

$$\begin{aligned}
\|u^\epsilon - u_h^\epsilon\|_V &\leq \|u^\epsilon - \tilde{u}_h^\epsilon\|_V + \|\tilde{u}_h^\epsilon - u_h^\epsilon\|_V \\
&\leq Ch^{\ell-2}\|u^\epsilon\|_{H^\ell} + \rho_1 \\
&\leq C_{10}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell}.
\end{aligned}$$

To obtain the L^2 error estimate (9.41), we start with the error equation:

$$(\Delta e_h^\epsilon, \Delta v_h) + \langle F(u^\epsilon) - F(u_h^\epsilon), v_h\rangle = 0 \quad \forall v_h \in V_0^h,$$

where $e_h^\epsilon := u^\epsilon - u_h^\epsilon$. Using the mean value theorem, we obtain

$$(\Delta e_h^\epsilon, \Delta v_h) + \langle F'[\xi](e_h^\epsilon), v_h\rangle = 0 \quad \forall v_h \in V_0^h, \tag{9.42}$$

where $\xi = u^\epsilon - \tau e_h^\epsilon$ for some $\tau \in [0,1]$. Again, we have abused the notation of ξ, defining it differently in different proofs.

Next, let $w \in H^p(\Omega) \cap V_0$ be the solution to the following auxiliary problem:

$$\langle (G'_\epsilon[u^\epsilon])^*(w), z\rangle = (e_h^\epsilon, z) \qquad \forall z \in V_0,$$

with

$$\|w\|_{H^p} \le C_R(\epsilon)\|e_h^\epsilon\|_{L^2}. \tag{9.43}$$

Using (9.42), we then have for any $w_h \in V_0^h$

$$\begin{aligned}
\|e_h^\epsilon\|_{L^2}^2 &= \langle (G'_\epsilon[u^\epsilon])^*(w), e_h^\epsilon\rangle \\
&= \langle G'_\epsilon[u^\epsilon](e_h^\epsilon), w\rangle \\
&= B_\epsilon[e_h^\epsilon, w] \\
&= B_\epsilon[e_h^\epsilon, w - w_h] + \epsilon(\Delta e_h^\epsilon, \Delta w_h) + \langle F'[u^\epsilon](e_h^\epsilon), w_h\rangle \\
&= B_\epsilon[e_h^\epsilon, w - w_h] + \langle \big(F'[u^\epsilon] - F'[\xi]\big)(e_h^\epsilon), w_h\rangle \\
&\le CC_2(\epsilon)\|e_h^\epsilon\|_V\|w - w_h\|_V + \|F'[u^\epsilon] - F'[\xi]\|_{VV^*}\|e_h^\epsilon\|_V\|w_h\|_V.
\end{aligned} \tag{9.44}$$

Let

$$h_4 = O\left(\frac{\delta}{C_{10}(\epsilon)\|u^\epsilon\|_{H^\ell}}\right)^{\frac{1}{\ell-2}}.$$

Then by (9.40) for $h \le h_4$

$$\|u^\epsilon - \xi\|_V = \tau\|e_h^\epsilon\|_V \le \delta.$$

Therefore for appropriate choice of w_h in (9.44), we have for $h \le \min\{h_0, h_1, h_2, h_3, h_4\}$,

$$\begin{aligned}
\|e_h^\epsilon\|_{L^2}^2 &\le C\Big(C_2(\epsilon)h^{p-2}\|e_h\|_V + L(\epsilon, h)\|e_h^\epsilon\|_V^2\Big)\|w\|_{H^p} \\
&\le CC_R(\epsilon)\Big(C_2(\epsilon)h^{p-2}\|e_h\|_V + L(\epsilon, h)\|e_h^\epsilon\|_V^2\Big)\|e_h^\epsilon\|_{L^2}.
\end{aligned}$$

Thus,

$$\begin{aligned}\|e_h^\epsilon\|_{L^2} &\le CC_R(\epsilon)\Big(C_2(\epsilon)h^{p-2}\|e_h\|_V + L(\epsilon,h)\|e_h^\epsilon\|_V^2\Big)\\ &\le CC_{10}(\epsilon)C_R(\epsilon)\Big(C_2(\epsilon)h^{\ell+p-4}\|u^\epsilon\|_{H^\ell} + L(\epsilon,h)C_{10}(\epsilon)h^{2\ell-4}\|u^\epsilon\|_{H^\ell}^2\Big).\end{aligned}$$

□

Remark 9.3.4. *H^1-norm error estimates can be obtained from their L^2 and H^2-norm errors by using norm interpolation techniques [17, Theorem 14.3.12].*

9.4 Examples

9.4.1 Monge-Ampère Equation

A detailed analysis of the Monge-Ampère case was carried out in Chapter 3, where we proved optimal error estimates in the energy norm. We now apply the results of the previous section to verify that we reach the same conclusions.

Recall in the case of the Monge-Ampère equation, we have

$$\begin{aligned}F(u) = F(D^2u, Du, u, u) &= f - \det(D^2u),\\ F'[v](w) &= -\mathrm{cof}(D^2v) : D^2w.\end{aligned}$$

Therefore, (9.3)–(9.5) become

$$-\Delta^2 u^\epsilon + \det(D^2u^\epsilon) = f \qquad \text{in } \Omega, \tag{9.45}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{9.46}$$

$$\Delta u^\epsilon = \epsilon \qquad \text{on } \partial\Omega, \tag{9.47}$$

and the finite element method for (9.45)–(9.47) (cf. (9.10)) is to find $u_h^\epsilon \in V_g^h$ such that

$$-\epsilon(\Delta u_h^\epsilon, \Delta v_h) + (\det(D^2u_h^\epsilon), v_h) = (f, v_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta}\right\rangle_{\partial\Omega}. \tag{9.48}$$

The linearization of G_ϵ at the solution u^ϵ is

$$G_\epsilon'[u^\epsilon](v) = \epsilon\Delta^2 v - \Phi^\epsilon : D^2v = \epsilon\Delta^2 v - \mathrm{div}(\Phi^\epsilon Dv),$$

where Φ^ϵ denotes the cofactor matrix of D^2u^ϵ, and we have used Lemma A.0.1 to get the

last equality. Finally, the bilinear form $B_\epsilon[\cdot,\cdot]$ becomes

$$B_\epsilon[v,w] = \epsilon(\Delta v, \Delta w) + (\Phi^\epsilon Dv, Dw).$$

By Theorems 2.2.2, 3.2.1, and 3.2.2, assumptions (A1)–(A2) are true with

$$\begin{aligned} &C_1(\epsilon) = O(\epsilon), \quad C_2(\epsilon) = O(\epsilon^{-\frac{1}{2}}), \quad \gamma(\epsilon) \equiv 0, \\ &p = 4, \qquad\qquad C_R(\epsilon) = O(\epsilon^{-1}), \quad C_5(\epsilon) = O(\epsilon^{-1}), \quad C_6(\epsilon) = O(\epsilon^{-1}). \end{aligned}$$

The goal of this section is to prove conditions (A3)–(A5) hold and derive the explicit dependence of constants $C_i(\epsilon)$, δ, and $L(\epsilon, h)$ on ϵ. For the reader's convenience, we recall the bounds stated in Chapter 2:

$$\begin{aligned} &\|u^\epsilon\|_{H^j} = O(\epsilon^{\frac{1-j}{2}}) \ (j=1,2,3), \quad \|u^\epsilon\|_{W^{j,\infty}} = O(\epsilon^{1-j}) \ (j=1,2), \\ &\|\Phi^\epsilon\|_{L^\infty} = O(\epsilon^{-1}), \qquad\qquad \|\Phi^\epsilon\|_{L^2} = O(\epsilon^{-\frac{1}{2}}), \end{aligned} \tag{9.49}$$

where $\Phi^\epsilon = \mathrm{cof}(D^2 u^\epsilon)$, denotes the cofactor matrix of $D^2 u^\epsilon$.

To confirm (A3), we take $Y = H^2(\Omega)$ with $\|\cdot\|_Y = \|\cdot\|_{H^2}$ for the case $n = 2$, and $Y = W^{2,\infty}(\Omega)$ with $\|\cdot\|_Y = \|\cdot\|^2_{W^{2,\infty}}$ for the case $n = 3$. Applying Lemma A.0.1, a Sobolev inequality, and an argument involving mollifiers (similar to Lemma 3.3.2), yields

$$\begin{aligned} \sup_{y\in Y} \frac{\|F'[y]\|_{VV^*}}{\|y\|_Y} &= \sup_{y\in Y}\sup_{v\in V} \frac{\|F'[y](v)\|_{V^*}}{\|y\|_Y\|v\|_V} \\ &= \sup_{y\in Y}\sup_{v\in V}\sup_{w\in V} \frac{|\langle \mathrm{cof}(D^2 y) : D^2 v, w\rangle|}{\|y\|_Y\|v\|_V\|w\|_V} \\ &= \sup_{y\in Y}\sup_{v\in V}\sup_{w\in V} \frac{|(\mathrm{cof}(D^2 y) Dv, Dw)|}{\|y\|_Y\|v\|_V\|w\|_V} \\ &\le C\Big(\sup_{y\in Y} \frac{\|\mathrm{cof}(D^2 y)\|_{L^2}}{\|y\|_Y}\Big) \le C. \end{aligned}$$

Thus, (A3) holds. We also note by (9.49)

$$\|u^\epsilon\|_Y \le C\epsilon^{\frac{1}{2}(5-3n)}. \tag{9.50}$$

Next, (A4) holds by setting $\tilde{u}^\epsilon_h = I_h u^\epsilon$, where $I_h u^\epsilon$ is the standard finite element interpolant of u^ϵ. It follows from standard interpolation theory [27, 17] that $C_3(\epsilon) = O(1)$, $C_4(\epsilon) = O(1)$. Thus,

$$C_9(\epsilon) = CC_3(\epsilon)C_4(\epsilon)C_6(\epsilon)\|u^\epsilon\|_Y = O(\epsilon^{\frac{3}{2}(1-n)}).$$

To verify (A5), we derive the following identity for any $v_h \in V_g^h$:

$$\begin{aligned}\|F'[u^\epsilon] - F'[v_h]\|_{VV^*} &= \sup_{w\in V} \frac{\|\big(\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)\big) : D^2w\|_{V^*}}{\|w\|_V} \\ &= \sup_{w\in V}\sup_{z\in V} \frac{|((\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h))Dw, Dz)|}{\|w\|_V\|z\|_V} \\ &\le C\|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)\|_{L^2}.\end{aligned}$$

It follows for $n = 2$,

$$\|F'[u^\epsilon] - F'[v_h]\|_{VV^*} \le C\|u^\epsilon - v_h\|_V.$$

Thus, $L(\epsilon, h) = O(1)$ in the case $n = 2$. For the $n = 3$ case, we use the same notation as Lemma 3.3.2 to conclude by the mean value theorem that for any $i, j = 1, 2, 3$

$$\begin{aligned}\|\mathrm{cof}(D^2u^\epsilon)_{ij} - \mathrm{cof}(D^2v_h)_{ij}\|_{L^2} &= \|\det(D^2u^\epsilon\big|_{ij}) - \det(D^2v_h\big|_{ij})\|_{L^2} \\ &= \|\Lambda^{ij} : (D^2u^\epsilon\big|_{ij} - D^2v_h\big|_{ij})\|_{L^2} \\ &= \|\Lambda^{ij}\|_{L^\infty}\|D^2u^\epsilon\big|_{ij} - D^2v_h\big|_{ij}\|_{L^2} \\ &\le \|\Lambda^{ij}\|_{L^\infty}\|u^\epsilon - v_h\|_V,\end{aligned}$$

where $\Lambda^{ij} = \mathrm{cof}(D^2u^\epsilon\big|_{ij} + \tau(D^2v_h\big|_{ij} - D^2u^\epsilon|_{ij}))$ for some $\tau \in [0, 1]$. Noting $\Lambda^{ij} \in \mathbf{R}^{2\times 2}$, we have $\|\Lambda^{ij}\|_{L^\infty} \le C\|u^\epsilon + v_h\|_{W^{2,\infty}}$. Thus, for any $\delta > 0$ and $v_h \in V_g^h$ with $\|\tilde{u}_h^\epsilon - v_h\|_V \le \delta$, we have using the inverse inequality and (9.49)

$$\begin{aligned}\|F'[u^\epsilon] - F'[v_h]\|_{VV^*} &\le C\|u^\epsilon + v_h\|_{W^{2,\infty}}\|u^\epsilon - v_h\|_V \\ &\le C(\epsilon^{-1} + h^{-\frac{3}{2}}\delta)\|u^\epsilon - v_h\|_V.\end{aligned}$$

Thus $L(\epsilon, h) = O(\epsilon^{-1} + h^{-\frac{3}{2}})$ in the three dimensional case. We note $L(\epsilon, h) = o(h^{2-\ell})$ provided $\ell > \frac{7}{2}$.

Next, using (9.50) yields

$$\|u^\epsilon - u_h^\epsilon\|_V \le C_{10}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \le CC_9(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell} \le C\epsilon^{\frac{3}{2}(1-n)}h^{\ell-2}\|u^\epsilon\|_{H^\ell}. \tag{9.51}$$

Finally, by (9.41)

$$\begin{aligned}\|u^\epsilon - u_h^\epsilon\|_{L^2} &\le C_{11}(\epsilon)\Big(C_2(\epsilon)h^{\ell+p-4}\|u^\epsilon\|_{H^\ell} + L(\epsilon,h)C_{10}(\epsilon)h^{2\ell-4}\|u^\epsilon\|_{H^\ell}^2\Big) \\ &\le CC_{10}C_R(\epsilon)\Big(\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + L(\epsilon,h)C_{10}(\epsilon)h^{2\ell-4}\|u^\epsilon\|_{H^\ell}^2\Big) \\ &\le C\epsilon^{-3}C_{10}(\epsilon)\Big(\epsilon^{-\frac{1}{2}}h^\ell\|u^\epsilon\|_{H^\ell} + \epsilon^{2-n}h^{2\ell-1-\frac{3}{2}n}C_{10}(\epsilon)\|u^\epsilon\|_{H^\ell}^2\Big).\end{aligned} \tag{9.52}$$

We note that the error estimates (9.51)–(9.52) are exactly the conclusions of Theorems 3.3.4 and 3.3.5.

9.4.2 The Equation of Prescribed Gauss Curvature

For given $K > 0$, the equation of prescribed Gauss Curvature is as follows:

$$\det(D^2u) = K(1+|Du|^2)^{\frac{n+2}{2}} \qquad \text{in } \Omega, \tag{9.53}$$

$$u = g \qquad \text{on } \partial\Omega. \tag{9.54}$$

Equation (9.53) is a fully nonlinear Monge-Ampère-type equation which arises in differential geometry. Indeed, given a manifold which is the graph of a function u embedded in $\mathbf{R}^{n+1}$, the Gauss curvature of the manifold (the product of the principal curvatures) is given by

$$K = \frac{\det(D^2u)}{(1+|Du|^2)^{\frac{n+2}{2}}}.$$

It is known [59] that there exists a constant $K^* > 0$ such that for each $K \in [0, K^*)$, problem (9.53)–(9.54) has a unique convex viscosity solution. Theoretically, it is very difficult to give an accurate estimate for the upper bound K^*. This then calls for help from accurate numerical methods. Indeed, the methodology and analysis of the vanishing moment method works very well for solving this problem and for estimating K^*.

In the case of the equation of prescribed Gauss curvature, we have

$$\begin{aligned}F(D^2u, Du, u, x) &= K(1+|Du|^2)^{\frac{n+2}{2}} - \det(D^2u) \\ F'[v](w) &= K(n+2)(1+|Dv^\epsilon|^2)^{\frac{n}{2}}Dv^\epsilon \cdot Dw - \operatorname{cof}(D^2v) : D^2w.\end{aligned}$$

Therefore, (9.3)–(9.5) becomes

$$-\Delta^2 u^\epsilon + \det(D^2u^\epsilon) - K(1+|Du^\epsilon|^2)^{\frac{n+2}{2}} = 0 \qquad \text{in } \Omega, \tag{9.55}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{9.56}$$

$$\Delta u^\epsilon = \epsilon \qquad \text{on } \partial\Omega, \tag{9.57}$$

and the finite element method for (9.55)–(9.57) (cf. (9.10)) is to find $u_h^\epsilon \in V_g^h$ such that

$$
\begin{aligned}
&-\epsilon(\Delta u_h^\epsilon, \Delta v_h) + (\det(D^2 u_h^\epsilon), v_h) \\
&\qquad - K\big((1+|Du_h^\epsilon|^2)^{\frac{n+2}{2}}, v_h\big) = -\left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta} \right\rangle_{\partial\Omega}.
\end{aligned} \tag{9.58}
$$

We also note that the linearization of G_ϵ at the solution u^ϵ is

$$
G'_\epsilon[u^\epsilon](v) = \epsilon\Delta^2 v - \operatorname{div}(\Phi^\epsilon Dv) + K(n+2)(1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv,
$$

where Φ^ϵ denotes the cofactor matrix of D^2u^ϵ. The associated bilinear form is

$$
B_\epsilon[v,w] = \epsilon(\Delta v, \Delta w) + (\Phi^\epsilon Dv, Dw) + K(n+2)\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, w\big).
$$

For the continuation of this chapter, we assume there exists a unique convex solution to (9.55)–(9.57). Furthermore, we assume that bounds (9.49) hold for the solution of (9.55)–(9.57). We consider if assumptions (A2)–(A3) hold.

We first prove the following identity.

Lemma 9.4.1. *For any $w \in V_0$, there holds*

$$
\begin{aligned}
&\big(\big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} Du^\epsilon \cdot Dw, w\big) \\
&\qquad = -\frac{1}{2}\big(w^2, [1+|Du^\epsilon|^2]^{\frac{n}{2}} \Delta u^\epsilon + n[1+|Du^\epsilon|^2]^{\frac{n-2}{2}} \Delta_\infty u^\epsilon\big),
\end{aligned} \tag{9.59}
$$

where Δ_∞ is the infinite Laplace operator, that is,

$$
\Delta_\infty u^\epsilon := D^2 u^\epsilon Du^\epsilon \cdot Du^\epsilon.
$$

Proof. Integrating by parts, we obtain

$$
\begin{aligned}
&\big(\big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} Du^\epsilon \cdot Dw, w\big) \\
&\quad = \left(\big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} Du^\epsilon, D\left(\frac{w^2}{2}\right)\right) \\
&\quad = -\frac{1}{2}\Big(w^2, \big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} \Delta u^\epsilon\Big) - \frac{1}{2}\Big(w^2, \sum_{i=1}^n \frac{\partial}{\partial x_i}(1+|Du^\epsilon|^2)^{\frac{n}{2}} \frac{\partial u^\epsilon}{\partial x_i}\Big).
\end{aligned}
$$

Expanding the last term yields

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_i}(1+|Du^\epsilon|^2)^{\frac{n}{2}} \frac{\partial u^\epsilon}{\partial x_i} = n(1+|Du^\epsilon|^2)^{\frac{n-2}{2}} \sum_{i,j=1}^{n} \frac{\partial^2 u^\epsilon}{\partial x_i \partial x_j} \frac{\partial u^\epsilon}{\partial x_i} \frac{\partial u^\epsilon}{\partial x_j}$$
$$= n(1+|Du^\epsilon|^2)^{\frac{n-2}{2}} \Delta_\infty u^\epsilon.$$

Thus,

$$\big(\big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} Du^\epsilon \cdot Dw, w\big) = -\frac{1}{2}\big(w^2, [1+|Du^\epsilon|^2]^{\frac{n}{2}} \Delta u^\epsilon + n[1+|Du^\epsilon|^2]^{\frac{n-2}{2}} \Delta_\infty u^\epsilon\big).$$

□

Since u^ϵ is convex, both Δu^ϵ and $\Delta_\infty u^\epsilon$ are positive, leading to the following corollary.

Corollary 9.4.2. *For any $w \in V_0$ there holds*

$$\big(\big[1+|Du^\epsilon|^2\big]^{\frac{n}{2}} Du^\epsilon \cdot Dw, w\big) \leq 0$$

with equality only holding for $w \equiv 0$.

Now we are ready to show that a Gårding-type inequality (9.12) holds. Since u^ϵ is convex, there exists a constant $\theta > 0$ such that

$$\begin{aligned} &\epsilon\|\Delta v\|_{L^2}^2 + \langle F'[u^\epsilon](v), v\rangle \\ &\quad = \epsilon\|\Delta v\|_{L^2}^2 + (\Phi^\epsilon Dv, Dv) + K(n+2)\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, v\big) \\ &\quad \geq \epsilon\|\Delta v\|_{L^2}^2 + \theta\|Dv\|_{L^2}^2 - K(n+2)\big|\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, v\big)\big| \\ &\quad \geq C_1(\epsilon)\|v\|_{H^2}^2 - K(n+2)\big|\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, v\big)\big|. \end{aligned} \tag{9.60}$$

Bounding the last term in the above expression, we use (9.49) and Lemma 9.4.1 to obtain

$$\begin{aligned} &\big|K(n+2)\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, v\big)\big| \\ &\quad = \frac{K(n+2)}{2}\big(v^2, [1+|Du^\epsilon|^2]^{\frac{n}{2}} \Delta u^\epsilon + n[1+|Du^\epsilon|^2]^{\frac{n-2}{2}} \Delta_\infty u^\epsilon\big) \\ &\quad \leq \frac{K(n+2)}{2}\|v\|_{L^2}^2\big(\|Du^\epsilon\|_{L^\infty}^n \|\Delta u^\epsilon\|_{L^\infty} + n\|Du^\epsilon\|_{L^\infty}^{n-2}\|\Delta_\infty u^\epsilon\|_{L^\infty}\big) \\ &\quad \leq C\|v\|_{L^2}^2\big(\|\Delta u^\epsilon\|_{L^\infty} + \|\Delta_\infty u^\epsilon\|_{L^\infty}\big) \\ &\quad = \gamma(\epsilon)\|v\|_{L^2}^2. \end{aligned} \tag{9.61}$$

Using bound (9.61) in (9.60), we immediately obtain

$$C_1(\epsilon)\|v\|_V^2 \le \epsilon\|\Delta v\|_{L^2}^2 + \langle F'[u^\epsilon](v), v\rangle + \gamma(\epsilon)\|v\|_{L^2}^2.$$

Next, for any $v, w \in V_0$, we have using (9.49)

$$\begin{aligned}\langle F'[u^\epsilon](v), w\rangle &= (\Phi^\epsilon Dv, Dw) + K(n+2)\big((1+|Du^\epsilon|^2)^{\frac{n}{2}} Du^\epsilon \cdot Dv, w\big)\\ &\le C\epsilon^{-\frac{1}{2}}\|v\|_V\|w\|_V + K(n+2)(1+\|u^\epsilon\|_{W^{1,\infty}}^2)^{\frac{n}{2}}\|u^\epsilon\|_{W^{1,\infty}}\|v\|_{H^1}\|w\|_{L^2}\\ &\le C\epsilon^{-\frac{1}{2}}\|v\|_V\|w\|_V\\ &= C_2(\epsilon)\|v\|_V\|w\|_V.\end{aligned}$$

Therefore, we reach the following conclusion.

Proposition 9.4.3. *Suppose $\big(G'_\epsilon[u^\epsilon]\big)^*$ is an isomorphism from V_0 to V_0^*, that is for all $\varphi \in V_0^*$ there exists $v \in V_0$ such that*

$$\langle (G'_\epsilon[u^\epsilon])^*(v), w\rangle = \langle \varphi, w\rangle \qquad \forall w \in V_0. \tag{9.62}$$

Furthermore, suppose that there exists $p > 2$ and $C_R(\epsilon) > 0$ such that if $\varphi \in L^2(\Omega)$ and $v \in V_0$ satisfies (9.62) then $v \in H^p(\Omega)$ and

$$\|v\|_{H^p} \le C_R(\epsilon)\|\varphi\|_{L^2}. \tag{9.63}$$

Then assumption (A2) *holds.*

To confirm (A3), we take $Y = H^2(\Omega)$ with norm $\|\cdot\|_Y = \|\cdot\|_{H^2} + \|\cdot\|_{W^{1,4}}^2$ in the $n = 2$ case, and let $Y = W^{2,\infty}(\Omega)$ with norm $\|\cdot\|_Y = \|\cdot\|_{W^{2,\infty}}^2 + \|\cdot\|_{W^{1,6}}^3$ in the $n = 3$ case. Here, we used the Sobolev embeddings

$$\begin{aligned} H^2(\Omega) &\hookrightarrow W^{1,4}(\Omega) \qquad n = 2,\\ W^{2,\infty}(\Omega) &\hookrightarrow W^{1,6}(\Omega) \qquad n = 3.\end{aligned}$$

Applying Lemma A.0.1 yields

$$\begin{aligned}&\sup_{y\in Y}\frac{\|F'[y]\|_{VV^*}}{\|y\|_Y}\\ &= \sup_{y\in Y}\sup_{v\in V}\sup_{z\in V}\frac{|\langle \mathrm{cof}(D^2y) : D^2v, z\rangle - K(n+2)\langle (1+|Dy|^2)^{\frac{n}{2}} Dy\cdot Dv, z\rangle|}{\|y\|_Y\|v\|_V\|z\|_V}\\ &= \sup_{y\in Y}\sup_{v\in V}\sup_{z\in V}\frac{|(\mathrm{cof}(D^2y)Dv, Dz) - K(n+2)((1+|Dy|^2)^{\frac{n}{2}} Dy\cdot Dv, z)|}{\|y\|_Y\|v\|_V\|z\|_V}\end{aligned}$$

$$
\begin{aligned}
&\le C\Big(\sup_{y\in Y}\sup_{v\in V}\sup_{z\in V}\frac{\|\mathrm{cof}(D^2y)\|_{L^2}\|v\|_V\|z\|_V + K(n+2)\|(1+|Dy|^2)^{\frac{n}{2}}Dy\cdot Dv\|_{L^1}\|z\|_{L^\infty}}{\|y\|_Y\|v\|_V\|z\|_V}\Big)\\
&\le C\Big(\sup_{y\in Y}\sup_{v\in V}\sup_{z\in V}\frac{\|\mathrm{cof}(D^2y)\|_{L^2}\|v\|_V\|z\|_V + K(n+2)\||Dy|^n\|_{L^2}\|y\|_V\|v\|_V\|z\|_{L^\infty}}{\|y\|_Y\|v\|_V\|z\|_V}\Big)\\
&\le C\Big(\sup_{y\in Y}\sup_{v\in V}\sup_{z\in V}\frac{\|\mathrm{cof}(D^2y)\|_{L^2}\|v\|_V\|z\|_V + K(n+2)\|Dy\|^n_{L^{2n}}\|y\|_V\|v\|_V\|z\|_{L^\infty}}{\|y\|_Y\|v\|_V\|z\|_V}\Big)\\
&\le C\Big(\sup_{y\in Y}\frac{\|\mathrm{cof}(D^2y)\|_{L^2} + K(n+2)\|Dy\|^n_{L^{2n}}}{\|y\|_Y}\Big)\\
&\le C.
\end{aligned}
$$

Thus, (A3) holds. We also note $\|u^\epsilon\|_Y \le C\epsilon^{\frac{1}{2}(5-3n)}$.

Next (A4) holds by setting $\tilde{u}^\epsilon_h = I_h u^\epsilon$ and using standard interpolation theory. We also trivially have $C_3(\epsilon),\ C_4(\epsilon) = O(1)$ and

$$
C_9(\epsilon) = CC_3(\epsilon)C_4(\epsilon)C_6(\epsilon)\|u^\epsilon\|_Y = O(C_6(\epsilon)\epsilon^{\frac{1}{2}(5-3n)}).
$$

To verify condition (A5), we first make the following calculation.

$$
\begin{aligned}
&\|F'[u^\epsilon] - F'[v_h]\|_{VV^*}\\
&\quad= \Big\{\sup_{w\in V}\|(\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)) : D^2w\\
&\qquad\quad - K(n+2)\big[(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\big]\cdot Dw\|_{V^*}\Big\}\Big/\|w\|_V\\
&\quad= \sup_{w\in V}\sup_{z\in V}\Big\{\big|\big((\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h))Dw, Dz\big)\\
&\qquad\quad - K(n+2)\big(\big[(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\big]\cdot Dw, z\big)\big|\Big\}\Big/\|w\|_V\|z\|_V\\
&\quad\le C\|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)\|_{L^2}\\
&\qquad\quad + K(n+2)\sup_{w\in V}\sup_{z\in V}\frac{\big|\big(\big[(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\big]\cdot Dw, z\big)\big|}{\|w\|_V\|z\|_V}\\
&\quad\le C\Big\{\|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)\|_{L^2}\\
&\qquad\quad + \sup_{w\in V}\frac{\|\big[(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\big]\cdot Dw\|_{L^1}}{\|w\|_V}\Big\}.
\end{aligned}
$$

For the case $n = 2$, we have

$$
\begin{aligned}
&(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\\
&\qquad= (1+|Du^\epsilon|^2)(Du^\epsilon - Dv_h) + (Du^\epsilon - Dv_h)\cdot(Du^\epsilon + Dv_h)Dv_h.
\end{aligned}
$$

Thus using a Sobolev inequality and (9.49), we have for any $\delta > 0$ and $v_h \in V^h_g$ with

$\|u^\epsilon - v_h\|_V \le \delta$

$$
\begin{aligned}
&\|\big[(1+|Du^\epsilon|^2)^{\frac{n}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{n}{2}}Dv_h\big]\cdot Dw\|_{L^1}\\
&\qquad = \|\big[(1+|Du^\epsilon|^2)(Du^\epsilon - Dv_h) + (Du^\epsilon - Dv_h)\cdot(Du^\epsilon + Dv_h)Dv_h\big]\cdot Dw\|_{L^1}\\
&\qquad \le C\big(\|1+|Du^\epsilon|^2\|_{L^2} + \|(Du^\epsilon + Dv_h)\cdot Dv_h\|_{L^2}\big)\|Du^\epsilon - Dv\|_{L^4}\|Dw\|_{L^4}\\
&\qquad \le C\big(1+\|u^\epsilon + v_h\|_{H^1}^2\big)\|u^\epsilon - v\|_V\|w\|_V\\
&\qquad \le C(1+\delta^2)\|u^\epsilon - v\|_V\|w\|_V.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|F'[u^\epsilon] - F'[v]\|_{VV^*} &\le C\Big\{\|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v)\|_{L^2} + (1+\delta^2)\|u^\epsilon - v\|_V\Big\}\\
&\le C(1+\delta^2)\|u^\epsilon - v\|_V.
\end{aligned}
$$

Therefore, (A5) holds with $L(\epsilon,h) = O(1)$ in the two dimensional case.

For the case $n=3$, we use the mean value theorem to get

$$
\begin{aligned}
&(1+|Du^\epsilon|)^{\frac{3}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{3}{2}}Dv_h\\
&\qquad = (1+|Du^\epsilon|^2)^{\frac{3}{2}}(Du^\epsilon - Dv_h) + \big[(1+|Du^\epsilon|^2)^{\frac{3}{2}} - (1+|Dv_h|^2)^{\frac{3}{2}}\big]Dv_h\\
&\qquad = (1+|Du^\epsilon|^2)^{\frac{3}{2}}(Du^\epsilon - Dv_h) + \big[3(1+|D\xi|^2)^{\frac{1}{2}}D\xi\cdot(Du^\epsilon - Dv_h)\big]Dv_h,
\end{aligned}
$$

where $\xi = u^\epsilon + \tau(v-u^\epsilon)$ for some $\tau\in[0,1]$.

Using this identity, (9.49), and the fact $L^6 \hookrightarrow H^1$ for $n=3$, we have assuming $\delta<1$

$$
\begin{aligned}
&\|\big[(1+|Du^\epsilon|^2)^{\frac{3}{2}}Du^\epsilon - (1+|Dv_h|^2)^{\frac{3}{2}}Dv_h\big]\cdot Dw\|_{L^1}\\
&\qquad = \|\big[(1+|Du^\epsilon|^2)^{\frac{3}{2}}(Du^\epsilon - Dv_h) + \big(3(1+|D\xi|^2)^{\frac{1}{2}}D\xi\cdot(Du^\epsilon - Dv_h)\big)Dv_h\big]\cdot Dw\|_{L^1}\\
&\qquad \le \|(1+|Du^\epsilon|^2)^{\frac{3}{2}}\|_{L^2}\|Du^\epsilon - Dv_h\|_{L^4}\|Dw_h\|_{L^4}\\
&\qquad\qquad + 3\|Dw_h\|_{L^6}\|Du^\epsilon - Dv_h\|_{L^6}\|(1+|D\xi|^2)^{\frac{1}{2}}|D\xi||Dv_h|\|_{L^{\frac{3}{2}}}\\
&\qquad \le C\big(\|(1+|Du^\epsilon|^2)^{\frac{3}{2}}\|_{L^2} + \|(1+|D\xi|^2)^{\frac{1}{2}}|D\xi||Dv_h|\|_{L^{\frac{3}{2}}}\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(1+\|Du^\epsilon\|_{L^6}^3 + \||D\xi||Dv_h|\|_{L^{\frac{3}{2}}} + \||D\xi|^2|Dv_h|\|_{L^{\frac{3}{2}}}\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(1+(\|D\xi\|_{L^3} + \|D\xi\|_{L^6}^3)\|Dv_h\|_{L^3}\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(\epsilon^{-\frac{3}{2}} + (\|\xi\|_V + \|\xi\|_V^3)\|v_h\|_V\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(1+(\|u^\epsilon\|_V + \delta + \|u^\epsilon\|_V^3 + \delta^3)(\|u^\epsilon\|_V + \delta)\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(1+(\epsilon^{-\frac{3}{2}} + \delta)(\epsilon^{-\frac{1}{2}} + \delta)\big)\|u^\epsilon - v_h\|_V\|w\|_V\\
&\qquad \le C\big(\epsilon^{-2} + \epsilon^{-\frac{3}{2}}\delta\big)\|u^\epsilon - v_h\|_V\|w\|_V.
\end{aligned}
$$

Thus, using the same argument as used in the Monge-Ampère analysis above, we have

$$\begin{aligned}\|F'[u^\epsilon] - F'[v_h]\|_{VV^*} &\le C\Big\{\|\mathrm{cof}(D^2u^\epsilon) - \mathrm{cof}(D^2v_h)\|_{L^2} + \big(\epsilon^{-2} + \epsilon^{-\frac{3}{2}}\delta\big)\|u^\epsilon - v_h\|_V\Big\}\\ &\le C\Big\{\|u^\epsilon + v_h\|_{W^{2,\infty}} + \big(\epsilon^{-2} + \epsilon^{-\frac{3}{2}}\delta\big)\Big\}\|u^\epsilon - v_h\|_V\\ &\le C\big(\epsilon^{-2} + h^{-\frac{3}{2}}\delta + \epsilon^{-\frac{3}{2}}\delta\big)\|u^\epsilon - v_h\|_V.\end{aligned}$$

Thus, (A5) holds in the three dimensional case with $L(\epsilon, h) = C\big(\epsilon^{-2} + h^{-\frac{3}{2}}\big)$. We note $L(\epsilon, h) = o(h^{2-\ell})$ provided $\ell > \frac{7}{2}$.

Gathering up our results, and applying Theorem 9.3.3, we make the following conclusion.

Theorem 9.4.4. *Suppose $(G'_\epsilon[u^\epsilon])^*$ is an isomorphism from V_0 to V_0^*, and that there exists $p > 2$ and $C_R(\epsilon) > 0$ such that if $v \in V_0$ satisfies* (9.62) *then $v \in H^p(\Omega)$ and the bound* (9.63) *holds. Then for h sufficiently small, there exists a unique solution to* (9.58). *Furthermore, there exists positive constants $C_{12}(\epsilon)$, $C_{13}(\epsilon)$ such that*

$$\begin{aligned}&\|u^\epsilon - u_h^\epsilon\|_V \le C_{12}(\epsilon)h^{\ell-2}\|u^\epsilon\|_{H^\ell},\\ &\|u^\epsilon - u_h^\epsilon\|_{L^2} \le C_{13}(\epsilon)\Big(\epsilon^{-\frac{1}{2}}h^{\ell+p-4}\|u^\epsilon\|_{H^\ell} + \epsilon^{4-2n}h^{2\ell-1-\frac{3}{2}n}C_{12}(\epsilon)\|u^\epsilon\|_{H^\ell}^2\Big).\end{aligned}$$

9.5 Numerical Experiments and Rates of Convergence

In this section, we provide several 2-D numerical experiments to gauge the efficiency of the finite element methods developed in the previous sections. Since we have already conducted numerical experiments for the Monge-Ampère equation in Chapter 3 (cf. Section 3.6), we are specifically interested in approximating the equation of prescribed Gauss curvature.

Test 9.1

In this test, we fix $h = 0.024$ in order to study the behavior of u^ϵ. Notably, we are interested if there exists a solution to (9.55)–(9.57) and whether $\|u - u^\epsilon\| \to 0$ as $\epsilon \to 0^+$. To this end, we solve the following problem: find $u_h^\epsilon \in V_g^h$ such that

$$\begin{aligned}-\epsilon(\Delta u_h^\epsilon, \Delta v_h) + (\det(D^2u_h^\epsilon), v_h)&\\ - K\big((1 + |Du_h^\epsilon|^2)^{\frac{n+2}{2}}, v_h\big) = (f, v_h) - \left\langle \epsilon^2, \frac{\partial v_h}{\partial \eta}\right\rangle_{\partial\Omega}&.\end{aligned}$$

Here, we take V^h to be the Argyris finite element space and $\Omega = (0,1)^2$. We use the following test functions and parameters:

(a) $u = e^{\frac{x_1^2+x_2^2}{2}}$,

$f = (1 + x - 1^2 + x_2^2)e^{x_1^2+x_2^2} - 0.1(1 + (x_1^2+x_2^2)e^{x_1^2+x_2^2})^2$,

$g = e^{\frac{x_1^2+x_2^2}{2}}$, $\qquad K = 0.1$.

(b) $u = \cos(\sqrt{x_1}\pi) + \cos(\sqrt{x_2}\pi)$,

$f = \frac{\pi^2}{16}\big(x_1^{-\frac{3}{2}}\sin(\sqrt{x_1}\pi) - x_1^{-1}\pi\cos(\sqrt{x_1}\pi)\big)\,\big(x_2^{-\frac{3}{2}}\sin(\sqrt{x_2}\pi) - x_2^{-1}\pi\cos(\sqrt{x_2}\pi)\big)$

$\quad - 0.025\big(1 + \frac{\pi^2}{4}\big(x_1^{-1}\sin^2(\sqrt{x_1}\pi) + x_2^{-1}\sin^2(\sqrt{x_2}\pi)\big)\big)^2$,

$g = \cos(\sqrt{x_1}\pi) + \cos(\sqrt{x_2}\pi)$, $\qquad K = 0.025$.

The computed solution is compared to the exact solution in Figure 9.1. As seen from Figure 9.1, the behavior of $\|u-u_h^\epsilon\|$ behaves similarly to that of the Monge-Ampère equation (cf. Tests 3.1 and 5.1) in that $\|u-u_h^\epsilon\|_{L^2} \approx O(\epsilon)$, $\|u-u_h^\epsilon\|_{H^1} \approx O(\epsilon^{\frac{3}{4}})$, and $\|u-u_h^\epsilon\|_{H^2} \approx O(\epsilon^{\frac{1}{4}})$. Since we have fixed h very small, we expect that $\|u-u^\epsilon\|$ behaves similarly.

Test 9.2

The purpose of this test is to calculate the rate of convergence of $\|u^\epsilon - u_h^\epsilon\|$ for fixed ϵ in various norms. As in Test 9.1, we use Argyris elements and solve problem (9.58) with boundary condition $\Delta u^\epsilon = \epsilon$ on $\partial\Omega$ being replaced by $\Delta u^\epsilon = \phi^\epsilon$ on $\partial\Omega$. We use the following test functions and data:

(a) $u^\epsilon = e^{\frac{x_1^2+x_2^2}{2}}$, $\qquad f^\epsilon = (1 + x_1^2 + x_2^2)e^{x_1^2+x_2^2} - 0.1(1 + (x_1^2 + x_2^2)e^{x_1^2+x_2^2})^2$

$\qquad\qquad - \epsilon(4(1 + x_1^2 + x_2^2) + (2 + x_1^2 + x_2^2)^2)e^{\frac{x_1^2+x-2^2}{2}}$,

$g^\epsilon = e^{\frac{x_1^2+x_2^2}{2}}$, $\qquad \phi^\epsilon = (2 + x_1^2 + x_2^2)e^{\frac{x_1^2+x_2^2}{2}}$,

$K = 0.1$.

(b) $u^\epsilon = \frac{1}{8}(x_1^2 + x_2^2)^4$, $\qquad f^\epsilon = 7(6x_1^2x_2^2(x_1^8 + x_2^8) + 15x_1^4x_2^4(x_1^4 + x_2^4) + 20x_1^6x_2^6 + x_1^{12} + x_2^{12})$

$\qquad\qquad - 0.1(1 + x_1^2(x_1^2 + x_2^2)^6 + x_2^2(x_2^2 + x_1^2)^6)^2 - 288\epsilon(x_1^2 + x_2^2)^2$,

$g^\epsilon = \frac{1}{8}(x_1^2 + x_2^2)^4$, $\qquad \phi^\epsilon = 8(x_1^2 + x_2^2)^3$,

$K = 0.1$.

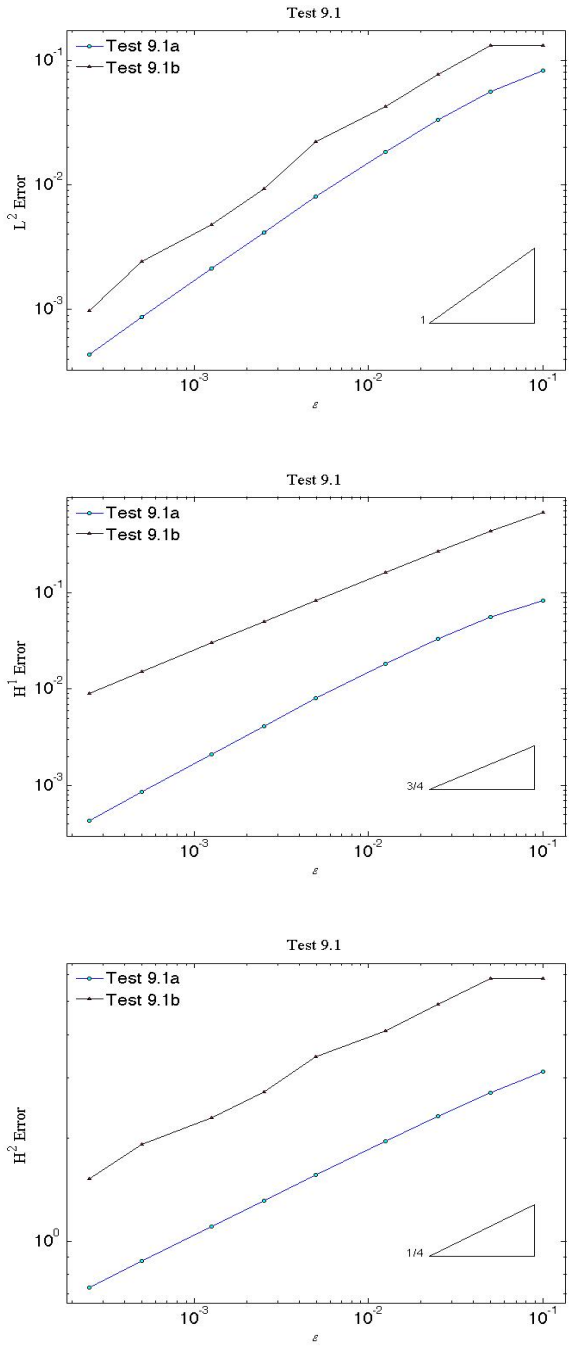


Figure 9.1: Test 9.1. Change of $\|u - u_h^\epsilon\|$ w.r.t. ϵ ($h = 0.024$)

Table 9.1: Test 9.2: Change of $\|u^\epsilon - u_h^\epsilon\|$ w.r.t. h ($\epsilon = 0.01$)

	h	$\frac{\|u^\epsilon - u_h^\epsilon\|_{L^2}}{h^6}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^1}}{h^5}$	$\frac{\|u^\epsilon - u_h^\epsilon\|_{H^2}}{h^4}$
Test 9.2a	0.083	0.8550556	2.1834919	15.655336
	0.05	0.6712376	1.6932808	12.28983
	0.031	0.5649840	1.3817137	9.998401
	0.024	0.5717870	1.3629	9.344284
	0.016	0.4016084	0.9571877	6.6707473
Test 9.2b	0.083	19.351388	48.506286	334.35795
	0.05	14.167003	35.054243	247.87093
	0.031	11.363703	27.634275	196.08477
	0.024	11.305043	26.313195	180.61956
	0.016	8.0321681	18.411873	127.12244

We record the results in Table 9.1. The data clearly indicates that $\|u^\epsilon - u_h^\epsilon\|_{H^2} = O(h^4)$ as expected from the previous section. We also see that $\|u^\epsilon - u_h^\epsilon\|_{L^2} = O(h^6)$ and $\|u^\epsilon - u_h^\epsilon\|_{H^1} = O(h^5)$, although a theoretical proof has not been shown for these rates of convergence. We note that these rates of convergence are comparable to the numerical experiments for the Monge-Ampère equation (cf. Test 3.2).

Test 9.3

For this test, we use our numerical method to approximate K^* and compare our results with those found in [6], where the method of continuity (which was used to prove existence of the equation of prescribed Gauss curvature) was implemented at the discrete level. We compute (9.55)–(9.57) with the following Dirichlet boundary conditions and domains as used in [6]:

(a) $g = \sqrt{1 - x_1^2 - y^2}$, $\Omega = (-0.57, 0.57)^2$.

(b) $g = 1 - x_1^2 - x_2^2$, $\Omega = (-0.57, 0.57)^2$.

(c) $g = 1 - (x_1 - 0.075)^2 - (x_2 - 0.015)^2$, $\Omega = (-0.57, 0.57)^2$.

(d) $g = \sqrt{1 - x_1^2 - y^2}$, $\Omega = (-0.72, 0.72) \times (-0.36, 0.36)$.

(e) $g = 1 - x_1^2 - x_2^2$, $\Omega = (-0.72, 0.72) \times (-0.36, 0.36)$.

(f) $g = 1 - (x_1 - 0.075)^2 - (x_2 - 0.015)^2$, $\Omega = (-0.72, 0.72) \times (-0.36, 0.36)$.

We remark that for the above choice of data, the solution of the prescribed Gauss curvature equation is concave, and so we set $\epsilon < 0$ in order to approximate the solution (cf. Test 3.2). Table 9.2 compares our results and those of [6]. We also plot the computed

Table 9.2: Test 9.3. Computed K^* with $\epsilon = -0.001$, $h = 0.031$

	Computed K^*	K^* in [6]
Test 9.3a	2.07	2.10
Test 9.3b	2.2	2.24
Test 9.3c	1.95	1.85
Test 9.3d	2.68	2.61
Test 9.3e	2.71	2.73
Test 9.3f	2.2	2.27

,

solution of Test 9.3a for K-values $0.5, 1, 1.5$ and 2.07 in Figure 9.2. Tables 9.2 shows that our numerical method gives comparable values to those computed in [6].

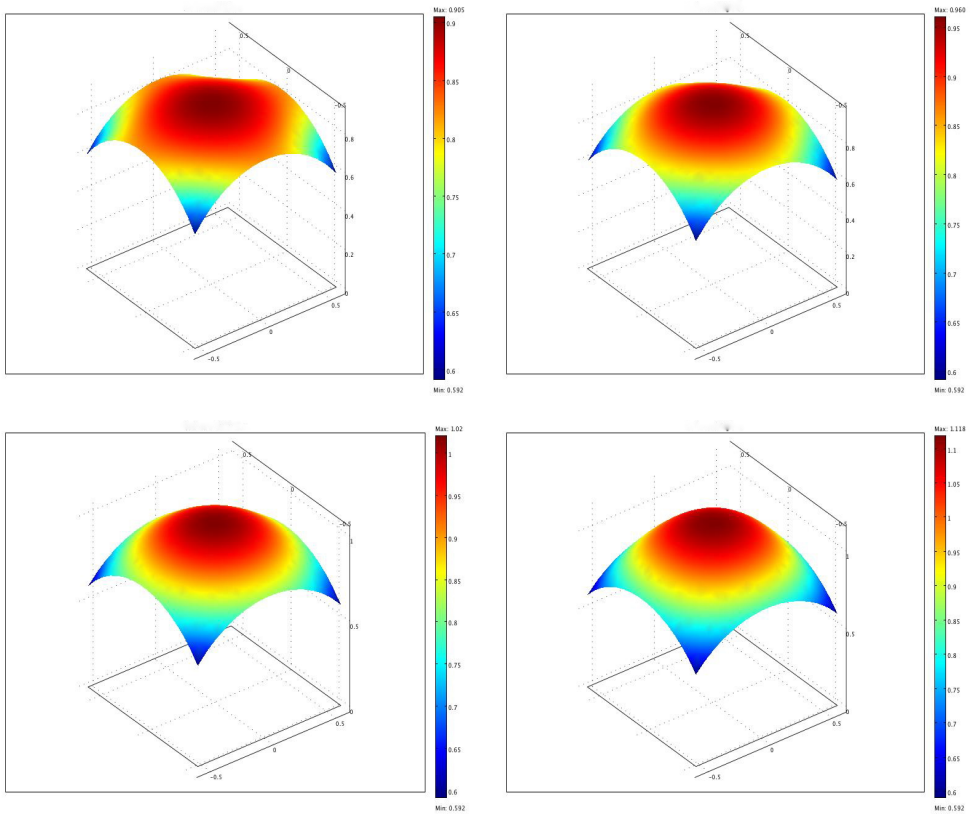


Figure 9.2: Test 9.3a. Compute solution for K-values 0.5 (top left), 1 (top right), 1.5 (bottom left), and 2.07 (bottom right). $\epsilon = -0.001$ ($h = 0.024$)

Chapter 10

Summary and Future Directions

The research presented in this dissertation developed and analyzed various numerical methods to approximate the viscosity solutions of fully nonlinear second order PDEs. We introduced a new notion of weak solutions for these types of PDEs called moment solutions which is based on the vanishing moment method, which unlike viscosity solutions are constructive by definition. The notion of moment solutions and the vanishing moment method are exactly in the same spirit as the original notion of viscosity solutions and the vanishing viscosity method proposed by Crandall and Lions in [30] for the Hamilton-Jacobi equations.

We concentrated on the Dirichlet problem for a prototypical fully nonlinear second order PDE, namely the Monge-Ampère equation. We presented four classes of numerical methods towards this equation, showing existence, uniqueness, and optimal error estimates in each case. We then extended this work to other problems including the nonlinear balance equation and a nonlinear formulation of the semigeostrophic flow equations. Because fully nonlinear second order PDEs appear in many application areas, and the numerical approximation of these types of problems is a relatively untouched area, the results presented in this work are expected to have a significant impact in the scientific community.

As expected, the research effort has created more work to be done than could reasonably be achieved in the time frame to complete the dissertation. There are many possible extensions and unanswered questions of this research which we now touch upon.

10.1 A General Moment Solution Theory

Recall (cf. Chapter 2) that the principle of the vanishing moment method is to approximate second order fully nonlinear PDEs

$$F(D^2u, Du, u, x) = 0 \qquad \text{in } \Omega, \tag{10.1}$$

$$u = g \qquad \text{on } \partial\Omega, \tag{10.2}$$

by the following higher order quasilinear PDEs:

$$G_\epsilon(D^r u^\epsilon) + F(D^2 u^\epsilon, Du^\epsilon, Du^\epsilon, u^\epsilon, x) = 0 \qquad \text{in } \Omega, \tag{10.3}$$

$$u^\epsilon = g \qquad \text{on } \partial\Omega, \tag{10.4}$$

$$\Delta u^\epsilon = c_\epsilon, \quad \text{or} \quad \frac{\partial \Delta u^\epsilon}{\partial \eta} = c_\epsilon, \quad \text{or} \quad D^2 u^\epsilon \eta \cdot \eta = c_\epsilon \qquad \text{on } \partial\Omega, \tag{10.5}$$

where $\{G_\epsilon\}$ is a family of suitably chosen linear or quasilinear differential operators of order $r \geq 3$. We call $\lim_{\epsilon \to 0^+} u^\epsilon$ (if it exists) a moment solution to problem (10.1)–(10.2).

To establish a complete theory of moment solutions and the vanishing moment method for fully nonlinear second order PDEs, there are many issues that must be addressed including

- How to choose the operator G_ϵ?
- Is there a unique solution to (10.3)–(10.5)?
- Does the limit $\lim_{\epsilon \to 0^+} u^\epsilon$ always exist, and if it does, what is the rate of convergence?
- How do moment solutions relate to viscosity solutions?

In the case of the Monge-Ampère equation

$$F(D^2 u, Du, u, x) = f(x) - \det(D^2 u(x)),$$

it has been shown [49] that there exists a unique moment solution to (10.3)–(10.5) with $G_\epsilon(D^4 u^\epsilon) = \epsilon \Delta^2 u^\epsilon$. Furthermore, the moment solution and the convex viscosity solution of the Monge-Ampère equation coincide. However, the proof of this result relies heavily on the structure of the Monge-Ampère equation, specifically, writing the Monge-Ampère equation in terms of eigenvalues of the Hessian of $D^2 u^\epsilon$ and using the identities

$$\det(D^2 u^\epsilon) = \prod_{i=1}^n \lambda_i^\epsilon, \qquad \Delta u^\epsilon = \sum_{i=1}^n \lambda_i^\epsilon.$$

For these reasons, it is not clear how to generalize this result to general nonlinear operators, $F(D^2 u, Du, u, x)$.

10.2 Discontinuous Galerkin Methods for Fully Nonlinear Second Order Equations

Discontinuous Galerkin (DG) methods are promising numerical methods for computing fully nonlinear second order PDEs via the vanishing moment method. The advantage of

such a method lies in the fact that there are no continuity requirements across the interfaces of the elements, and hence, arbitrarily high order polynomials and unstructured meshes can be used [7, 5]. Also, the test and trial spaces are easy to construct, and DG methods can easily handle inhomogeneous boundary conditions and curved boundaries. In this section, we formulate a symmetric interior penalty discontinuous Galerkin type formulation for problem (10.3)–(10.5) with $G_\epsilon(D^r u^\epsilon) = \epsilon\Delta^2 u^\epsilon$ and $c_\epsilon = \epsilon$.

First, we introduce the following notation. Let $\mathcal{T}_h$ be a quasiuniform triangular or rectangular partition of Ω if $n = 2$ or a quasiuniform tetrahedral or 3-D rectangular mesh if $n = 3$ with mesh size $h \in (0, 1)$. Let K be an $n-$dimensional simplex with $n+1$ vertices, and let F_j $(1 \le j \le n+1)$ denote the $(n-1)$-dimensional subsimplex of K. Next, we let $\mathcal{E}_h$ denote the set of all $(n-1)$-dimensional subsimplexes in the mesh $\mathcal{T}_h$, and define the set of interior and boundary $(n-1)$-dimensional subsimplexes as follows:

$$\mathcal{E}_h^I := \{F \in \mathcal{E}_h;\ F \cap \partial\Omega = \emptyset\},$$
$$\mathcal{E}_h^B := \{F \in \mathcal{E}_h;\ F \cap \partial\Omega \neq \emptyset\}.$$

Define the energy space

$$E_h = \prod_{K\in\mathcal{T}_h} H^4(K),$$

and its finite element subspace

$$V_r^h = \prod_{K\in\mathcal{T}_h} \mathbb{P}_r(K),$$

where $\mathbb{P}_r(K)$ denotes the space of polynomials of degree r on K.

For any $F \in \mathcal{E}_h^I$ there are two triangles K^+ and K^- such that $F = \partial K^+ \cap \partial K^-$. For any $v \in H^1(K^+) \cap H^1(K^-)$, define the jumps and averages of v across e as follows:

$$[v]\big|_F = v^+\big|_F - v^-\big|_F, \qquad \{v\}\big|_F = \frac{1}{2}\left(v^+\big|_F + v^-\big|_F\right),$$

where $v^+ = v\big|_{K^+}$, $v^- = v\big|_{K^-}$. Let η_F denote the unit normal of F pointing from K^- to K^+ and let $\{\tau_F^{(i)}\}_{i=1}^{n-1}$ denote an orthogonal frame of the tangent space of ∂F. For $v \in H^2(K^+) \cap H^2(K^-)$ define the jumps and averages of $\frac{\partial v}{\partial \eta}$ and $\frac{\partial v}{\partial \tau}$ across e as follows:

$$\left[\frac{\partial v}{\partial \eta}\right]\Big|_F = \frac{\partial v^+}{\partial \eta_F}\Big|_F - \frac{\partial v^-}{\partial \eta_F}\Big|_F, \qquad \left\{\frac{\partial v}{\partial \eta}\right\}\Big|_F = \frac{1}{2}\left(\frac{\partial v^+}{\partial \eta_F}\Big|_F + \frac{\partial v^-}{\partial \eta_F}\Big|_F\right),$$
$$\left[\frac{\partial v}{\partial \tau}\right]\Big|_F = \sum_{i=1}^{n-1}\left(\frac{\partial v^+}{\partial \tau_F^{(i)}}\Big|_F - \frac{\partial v^-}{\partial \tau_F^{(i)}}\Big|_F\right), \qquad \left\{\frac{\partial v}{\partial \tau}\right\}\Big|_F = \frac{1}{2}\sum_{i=1}^{n-1}\left(\frac{\partial v^+}{\partial \tau_F^{(i)}}\Big|_F + \frac{\partial v^-}{\partial \tau_F^{(i)}}\Big|_F\right).$$

Likewise, for $v \in H^3(K^+) \cap H^3(K^-)$, we define the jumps and averages of $\frac{\partial^2 v}{\partial \tau \partial \eta}$ and $\frac{\partial^2 v}{\partial \tau^2}$ as

$$\begin{aligned}
\left[\frac{\partial^2 v}{\partial \tau \partial \eta}\right]\Big|_F &= \sum_{i=1}^{n-1}\left(\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \eta_F}\Big|_F - \frac{\partial^2 v^-}{\partial \tau^{(i)} \partial \eta_F}\Big|_F\right), \\
\left\{\frac{\partial v}{\partial \tau \partial \eta}\right\}\Big|_F &= \frac{1}{2}\sum_{i=1}^{n-1}\left(\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \eta_F}\Big|_F + \frac{\partial^2 v^-}{\partial \tau_F^{(i)} \partial \eta_F}\Big|_F\right), \\
\left[\frac{\partial^2 v}{\partial \tau^2}\right]\Big|_F &= \sum_{i=1}^{n-1}\left(\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \tau_F^{(i)}}\Big|_F - \frac{\partial v^-}{\partial \tau_F^{(i)}}\Big|_F\right), \\
\left\{\frac{\partial^2 v}{\partial \tau^2}\right\}\Big|_F &= \frac{1}{2}\sum_{i=1}^{n-1}\left(\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \tau_F^{(i)}}\Big|_F + \frac{\partial^2 v^-}{\partial \tau_F^{(i)} \partial \tau_F^{(i)}}\Big|_F\right).
\end{aligned}$$

If $e \in \mathcal{E}_h^B$, then there exists K^+ such that $e = \partial K^+ \cap \partial\Omega$. Denote η_F to be the unit normal of e that points outside of K^+, and for any $v \in H^1(K^+)$ set

$$[v]\big|_F = v^+\big|_F, \qquad \{v\}\big|_F = v^+\big|_F.$$

For $v \in H^2(K^+)$ set

$$\begin{aligned}
\left[\frac{\partial v}{\partial \eta}\right]\Big|_F &= \frac{\partial v^+}{\partial \eta_F}\Big|_F, & \left\{\frac{\partial v}{\partial \eta_F}\right\}\Big|_F &= \frac{\partial v^+}{\partial \eta_F}\Big|_F, \\
\left[\frac{\partial v}{\partial \tau}\right]\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial v^+}{\partial \tau_F^{(i)}}\Big|_F, & \left\{\frac{\partial v}{\partial \tau_F}\right\}\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial v^+}{\partial \tau_F^{(i)}}\Big|_F,
\end{aligned}$$

and for $v \in H^3(K^+)$

$$\begin{aligned}
\left[\frac{\partial^2 v}{\partial \tau \partial \eta}\right]\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \eta_F}\Big|_F, & \left\{\frac{\partial v}{\partial \tau \partial \eta}\right\}\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \eta_F}\Big|_F, \\
\left[\frac{\partial^2 v}{\partial \tau^2}\right]\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \tau_F^{(i)}}\Big|_F, & \left\{\frac{\partial^2 v}{\partial \tau^2}\right\}\Big|_F &= \sum_{i=1}^{n-1}\frac{\partial^2 v^+}{\partial \tau_F^{(i)} \partial \tau_F^{(i)}}\Big|_F.
\end{aligned}$$

Multiplying (2.4) by $v_h \in V_r^h$ and integrating by parts, we obtain (see Lemma 6.1.1)

$$\begin{aligned}
0 &= \epsilon(\Delta^2 u^\epsilon, v_h) + (F(D^2 u^\epsilon, Du^\epsilon, u^\epsilon, x), v_h) \qquad (10.6)\\
&= \sum_{K \in \mathcal{T}_h}\Big(\epsilon(D^2 u^\epsilon, D^2 v)_K + (F(D^2 u^\epsilon, Du^\epsilon, u^\epsilon, x), v_h)_K + \epsilon\left\langle \frac{\partial \Delta u^\epsilon}{\partial \eta_K}, v_h\right\rangle_{\partial K} \\
&\quad - \epsilon\left\langle \Delta u^\epsilon, \frac{\partial v_h}{\partial \eta_K}\right\rangle_{\partial K} + \epsilon\sum_{i=1}^{n-1}\left\langle \frac{\partial^2 u^\epsilon}{\partial \tau_K^{(i)} \eta_K}, \frac{\partial v_h}{\partial \tau_K^{(i)}}\right\rangle_{\partial K} - \epsilon\sum_{i=1}^{n-1}\left\langle \frac{\partial^2 u^\epsilon}{\partial \tau_K^{(i)} \partial \tau_K^{(i)}}, \frac{\partial v_h}{\partial \eta_K}\right\rangle_{\partial K}\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{K\in\mathcal{T}_h} \Big(\epsilon(D^2u^\epsilon, D^2v_h)_K + (F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x), v_h)_K\Big) \\
&\quad + \epsilon \sum_{F\in\mathcal{E}_h} \Big(\Big\langle \Big\{ \frac{\partial \Delta u^\epsilon}{\partial \eta} \Big\}, [v_h] \Big\rangle_F - \Big\langle \Big\{ \Delta u^\epsilon + \frac{\partial^2 u^\epsilon}{\partial \tau^2} \Big\}, \Big[\frac{\partial v_h}{\partial \eta} \Big] \Big\rangle_F + \Big\langle \Big\{ \frac{\partial^2 u^\epsilon}{\partial \tau \partial \eta} \Big\}, \Big[\frac{\partial v_h}{\partial \tau} \Big] \Big\rangle_F \Big) \\
&= \sum_{K\in\mathcal{T}_h} \Big(\epsilon(D^2u^\epsilon, D^2v_h)_K + (F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x), v_h)_K\Big) \\
&\quad + \epsilon \sum_{F\in\mathcal{E}_h} \Big(\Big\langle \Big\{ \frac{\partial \Delta u^\epsilon}{\partial \eta} \Big\}, [v_h] \Big\rangle_F + \Big\langle \Big\{ \frac{\partial^2 u^\epsilon}{\partial \tau \partial \eta} \Big\}, \Big[\frac{\partial v_h}{\partial \tau} \Big] \Big\rangle_F \Big) \\
&\quad - \epsilon \sum_{F\in\mathcal{E}_h^I} \Big\langle \Big\{ \Delta u^\epsilon + \frac{\partial^2 u^\epsilon}{\partial \tau^2} \Big\}, \Big[\frac{\partial v_h}{\partial \eta} \Big] \Big\rangle_F - \epsilon \sum_{F\in\mathcal{E}_h^B} \Big\langle \epsilon + \frac{\partial^2 g}{\partial \tau^2}, \frac{\partial v_h}{\partial \eta} \Big\rangle_F .
\end{aligned}
$$

Here, $\{\tau_K^{(i)}\}_{i=1}^{n-1}$ and η_K denote the an orthogonal tangent frame and unit normal vector of ∂K, respectively. We have also assumed u^ϵ is sufficiently smooth in the above calculation, so that

$$
\Big[\frac{\partial \Delta u^\epsilon}{\partial \eta} \Big] \Big|_F = [\Delta u^\epsilon] \Big|_F = \Big[\frac{\partial^2 u^\epsilon}{\partial \tau^2} \Big] \Big|_F = \Big[\frac{\partial^2 u^\epsilon}{\partial \tau \partial \eta} \Big] \Big|_F = 0 \qquad \forall F \in \mathcal{E}_h^I.
$$

Defining

$$
\begin{aligned}
A_{h,\epsilon}(v, w) :&= \sum_{K\in\mathcal{T}_h} \Big(\epsilon(D^2v, D^2w)_K + (F(D^2v, Dv, v, x), w)_K\Big), \\
J_{h,\epsilon}(v, w) :&= \epsilon \sum_{F\in\mathcal{E}_h} \Big(\Big\langle \Big\{ \frac{\partial \Delta v}{\partial \eta} \Big\}, [w] \Big\rangle_F + \Big\langle \Big\{ \frac{\partial^2 v}{\partial \tau \partial \eta} \Big\}, \Big[\frac{\partial w}{\partial \tau} \Big] \Big\rangle_F \Big) \\
&\quad - \epsilon \sum_{F\in\mathcal{E}_h^I} \Big\langle \Big\{ \Delta v + \frac{\partial^2 v}{\partial \tau^2} \Big\}, \Big[\frac{\partial w}{\partial \eta} \Big] \Big\rangle_F ,
\end{aligned}
$$

we may write (10.6) as follows:

$$
A_{h,\epsilon}(u^\epsilon, v_h) + J_{h,\epsilon}(u^\epsilon, v_h) = \epsilon \sum_{F\in\mathcal{E}_h^B} \Big\langle \epsilon + \frac{\partial^2 g}{\partial \tau^2}, \frac{\partial v_h}{\partial \eta} \Big\rangle_F .
$$

We now make some modifications in order to provide the bilinear form with certain desirable properties, namely symmetry and coercivity in the higher order linear terms. First, we note

$$
J_{h,\epsilon}(v_h, u^\epsilon) = \epsilon \sum_{F\in\mathcal{E}_h^B} \Big(\Big\langle g, \frac{\partial \Delta v_h}{\partial \eta} \Big\rangle_F + \Big\langle \frac{\partial g}{\partial \tau}, \frac{\partial^2 v_h}{\partial \tau \partial \eta} \Big\rangle_F \Big),
$$

and

$$A_{h,\epsilon}(u^\epsilon, v_h) + J_{h,\epsilon}(u^\epsilon, v_h) + J_{h,\epsilon}(v_h, u^\epsilon)$$
$$= \epsilon \sum_{F\in\mathcal{E}_h^B} \Big(\Big\langle \epsilon + \frac{\partial^2 g}{\partial \tau^2}, \frac{\partial v_h}{\partial \eta} \Big\rangle_F + \Big\langle g, \frac{\partial \Delta v_h}{\partial \eta} \Big\rangle_F + \Big\langle \frac{\partial g}{\partial \tau}, \frac{\partial^2 v_h}{\partial \tau \partial \eta} \Big\rangle_F \Big).$$

Next, we penalize the jump terms by introducing the following bilinear form:

$$J_h^\gamma(v,w) = \sum_{F\in\mathcal{E}_h} \Big(\gamma_1 h_F^{-3} \langle [v], [w] \rangle_F + \gamma_2 h_F^{-1} \Big\langle \Big[\frac{\partial v}{\partial \tau}\Big], \Big[\frac{\partial w}{\partial \tau}\Big] \Big\rangle_F \Big)$$
$$+ \sum_{F\in\mathcal{E}_h^I} \gamma_3 h_F^{-1} \Big\langle \Big[\frac{\partial v}{\partial \eta}\Big], \Big[\frac{\partial w}{\partial \eta}\Big] \Big\rangle_F,$$

where γ_i $(i = 1, 2, 3)$ are positive constants independent of h and $h_F = \mathrm{diam}(F)$.

We then have the following identity:

$$A_{h,\epsilon}(u^\epsilon, v_h) + J_{h,\epsilon}(u^\epsilon, v_h) + J_{h,\epsilon}(v_h, u^\epsilon) + J_h^\gamma(u^\epsilon, v_h) = F(v_h),$$

where

$$F(v_h) = \sum_{F\in\mathcal{E}_h^B} \Big(\epsilon \Big\langle \epsilon + \frac{\partial^2 g}{\partial \tau^2}, \frac{\partial v_h}{\partial \eta} \Big\rangle_F + \Big\langle g, \epsilon \frac{\partial \Delta v_h}{\partial \eta} + \gamma_1 h_F^{-3} v_h \Big\rangle_F$$
$$+ \Big\langle \frac{\partial g}{\partial \tau}, \epsilon \frac{\partial^2 v_h}{\partial \tau \partial \eta} + \gamma_2 h_F^{-1} \frac{\partial v_h}{\partial \tau} \Big\rangle_F \Big).$$

These calculations motivate the following discontinuous Galerkin formulation for problem (10.3)–(10.5): Find $u_h^\epsilon \in V_r^h$ such that

$$a_{h,\epsilon}^\gamma(u_h^\epsilon, v_h) = F_h^\epsilon(v_h) \qquad \forall v_h \in V_r^h, \tag{10.7}$$

where

$$a_{h,\epsilon}^\gamma(u_h^\epsilon, v_h) := \sum_{K\in\mathcal{T}_h} \Big(\epsilon (D^2 u_h^\epsilon, D^2 v_h)_K + (F(D^2 u_h^\epsilon, D u_h^\epsilon, u_h^\epsilon, x), v_h)_K \Big)$$
$$+ \sum_{F\in\mathcal{E}_h} \Big(\epsilon \Big\langle \Big\{ \frac{\partial \Delta u_h^\epsilon}{\partial \eta} \Big\}, [v_h] \Big\rangle_F + \epsilon \Big\langle [u_h^\epsilon], \Big\{ \frac{\partial \Delta v_h}{\partial \eta} \Big\} \Big\rangle_F + \epsilon \Big\langle \Big\{ \frac{\partial^2 u_h^\epsilon}{\partial \tau \partial \eta} \Big\}, \Big[\frac{\partial v_h}{\partial \tau} \Big] \Big\rangle_F$$
$$+ \epsilon \Big\langle \Big[\frac{\partial u_h^\epsilon}{\partial \tau} \Big], \Big\{ \frac{\partial^2 v_h}{\partial \tau \partial \eta} \Big\} \Big\rangle_F + \gamma_1 h_F^{-3} \langle [u_h^\epsilon], [v_h] \rangle_F + \gamma_2 h_F^{-1} \Big\langle \Big[\frac{\partial u_h^\epsilon}{\partial \tau} \Big], \Big[\frac{\partial v_h}{\partial \tau} \Big] \Big\rangle_F \Big)$$

$$
\begin{aligned}
&- \sum_{F\in\mathcal{E}_h^I} \Big(\epsilon \left\langle \left\{ \Delta u_h^\epsilon + \frac{\partial^2 u_h^\epsilon}{\partial \tau^2} \right\}, \left[\frac{\partial v_h}{\partial \eta} \right] \right\rangle_F + \epsilon \left\langle \left[\frac{\partial u_h^\epsilon}{\partial \eta} \right], \left\{ \Delta v_h + \frac{\partial^2 v_h}{\partial \tau^2} \right\} \right\rangle_F \\
&- \gamma_3 h_F^{-1} \left\langle \left[\frac{\partial u_h^\epsilon}{\partial \eta} \right], \left[\frac{\partial v_h}{\partial \eta} \right] \right\rangle_F \Big).
\end{aligned}
$$

Remark 10.2.1. *Assuming $\gamma_2 = \gamma_3$, it is easy to check that the bilinear form $a_{h,\epsilon}^\gamma(\cdot,\cdot)$ can be written as follows:*

$$
\begin{aligned}
a_{h,\epsilon}^\gamma(u_h^\epsilon, v_h) = &\sum_{K\in\mathcal{T}_h} \Big(\epsilon (D^2 u_h^\epsilon, D^2 v_h)_K + (F(D^2 u_h^\epsilon, Du_h^\epsilon, u_h^\epsilon, x), v_h)_K \Big) \\
&+ \sum_{F\in\mathcal{E}_h^I} \Big(\epsilon \left\langle \left\{ \frac{\partial \Delta u_h^\epsilon}{\partial \eta} \right\}, [v_h] \right\rangle_F + \epsilon \left\langle [u_h^\epsilon], \left\{ \frac{\partial \Delta v_h}{\partial \eta} \right\} \right\rangle_F + \epsilon \left\langle \left\{ \frac{\partial D u_h^\epsilon}{\partial \eta} \right\}, [Dv_h] \right\rangle_F \\
&+ \epsilon \left\langle [Du_h^\epsilon], \left\{ \frac{\partial D v_h}{\partial \eta} \right\} \right\rangle_F - 2\epsilon \left\langle \{\Delta u_h^\epsilon\}, \left[\frac{\partial v_h}{\partial \eta} \right] \right\rangle_F - 2\epsilon \left\langle \left[\frac{\partial u_h^\epsilon}{\partial \eta} \right], \{\Delta v_h\} \right\rangle_F \\
&+ \gamma_1 h_e^{-3} \langle [u_h^\epsilon], [v_h] \rangle_F + \gamma_2 h_e^{-1} \langle [Du_h^\epsilon], [Dv_h] \rangle_F \Big) + \sum_{F\in\mathcal{E}_h^B} \Big(\epsilon \left\langle \frac{\partial \Delta u_h^\epsilon}{\partial \eta}, v \right\rangle \\
&+ \epsilon \left\langle u_h^\epsilon, \frac{\partial \Delta v_h}{\partial \eta} \right\rangle_F + \epsilon \left\langle \frac{\partial^2 u_h^\epsilon}{\partial \tau \partial \eta}, \frac{\partial v_h}{\partial \tau} \right\rangle_F + \epsilon \left\langle \frac{\partial u_h^\epsilon}{\partial \tau}, \frac{\partial^2 v_h}{\partial \tau \partial \eta} \right\rangle_F \\
&+ \gamma_1 h_F^{-3} \langle u_h^\epsilon, v_h \rangle_F + \gamma_2 h_F^{-1} \left\langle \frac{\partial u_h^\epsilon}{\partial \tau}, \frac{\partial v_h}{\partial \tau} \right\rangle_F \Big)
\end{aligned}
$$

We expect that the analysis of the DG scheme (10.7) to be nontrivial due to the nonlinearity in the bilinear form $a_{h,\epsilon}(\cdot,\cdot)$. Also of importance is determining the relationship between γ and ϵ in order for the associated linearized problem to be coercive, and what role they play in a priori error estimates.

10.3 Numerical Methods for the Optimal Mass Transport Problem

The original mass transport problem, proposed by Gaspard Monge in the 18th century, questions the optimal way to move soil to an excavation with minimal transportation cost, i.e., the total distance that the soil is moved, at an infinitesimal level, should be minimal [75]. A modern version of this problem was later formulated and studied by Kantorovich in 1942 leading to the famous Monge-Kantorovich (MK) problem [66], which has received considerable attention in recent years [3, 4, 45, 46, 85]. In general, the MK problem asks the following: Given two sets of equal volume, find the optimal mass-preserving mapping between them, where optimality is measured by a given positive cost density.

In a modern setting, the mass transport problem is described as follows. Let f_1 and f_2

be density functions satisfying the mass balance condition:

$$\int_X f_1(x)dx = \int_Y f_2(y)dy,$$

where $X = \text{supp}(f_1)$, $Y = \text{supp}(f_2)$. Let

$$\mathcal{S} := \left\{ \mathbf{s} : \mathbf{R}^n \to \mathbf{R}^n;\ \int_{\mathbf{s}^{-1}(E)} f_1(x)dx = \int_E f_2(y)dy\ \forall E \text{ Borel} \right\}$$

be the admissible set of mappings, and $c : \mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^+$ be the cost density. For $\mathbf{s} \in \mathcal{S}$, define

$$I[s] = \int_{\mathbf{R}^n} c(x, \mathbf{s}(x)) f_1(x)dx$$

to be the total work of the mapping $\mathbf{s}$. The optimal mass transport problem is then defined as seeking $\mathbf{s}^* \in \mathcal{S}$ that minimizes the total work

$$I[\mathbf{s}^*] \leq I[\mathbf{s}] \qquad \forall \mathbf{s} \in \mathcal{S}. \tag{10.8}$$

We note by definition that $\mathbf{s} \in \mathcal{S}$ implies

$$\int_X \phi(\mathbf{s}(x)) f_1(x)dx = \int_Y \phi(y) f_2(y)(y) \quad \forall \phi \in C^0(Y).$$

It can then be shown [11] that this mass-preserving condition implies $\mathbf{s}$ satisfies

$$f_2(\mathbf{s}(x)) \det(D\mathbf{s}(x)) = f_1(x) \qquad \forall x \in X, \tag{10.9}$$

$$\mathbf{s}(X) \subset Y. \tag{10.10}$$

For the case $c(x, y) = |x - y|^2$, it can be shown under reasonable conditions that there exists an optimal mapping satisfying (10.8) with $\mathbf{s}^* = D\varphi^*$ for some convex function $\varphi^* : X \to \mathbf{R}$ [56, 45]. Substituting this identity into (10.9)–(10.10), we obtain the following Monge-Ampère-type equation:

$$f_2(D\varphi^*(x)) \det(D^2\varphi^*(x)) = f_1(x) \qquad \forall x \in X, \tag{10.11}$$

$$D\varphi^*(X) \subset Y. \tag{10.12}$$

It is then possible to apply the vanishing moment methodology developed in Chapter 2 and approximate the nonlinear PDE (10.11) by the following regularized PDE:

$$-\epsilon\Delta^2\varphi^\epsilon + \det(D^2\varphi^\epsilon(x)) = \frac{f_1(x)}{f_2(D\varphi^\epsilon(x))} =: f(\varphi^\epsilon, x). \tag{10.13}$$

Generalizations of the numerical methods developed in Chapters 3–6 and 9 can then be used to approximate (10.13) to compute the optimal mapping.

10.4 Numerical Methods for Parabolic Fully Nonlinear Second Order Equations

There are several different versions of legitimate parabolic generalizations to elliptic PDE (10.1) (cf. [68, 95]). In this section, we shall only consider the following widely studied class of fully nonlinear second order parabolic PDEs:

$$F(D^2u, Du, u, x, t) + \frac{\partial u}{\partial t} = 0, \tag{10.14}$$

Clearly, this is the most natural parabolic generalization to equation (10.1). For example, the corresponding parabolic Monge-Amperè type equation reads as

$$\det(D^2u) - \frac{\partial u}{\partial t} = f \geq 0. \tag{10.15}$$

In the past two decades the viscosity solution theory has been well developed for equations (10.14) and (10.15) (cf. [68, 95, 61]). On the other hand, numerical approximation to these fully nonlinear parabolic PDEs is a completely untouched area. To the best of our knowledge, no numerical result (in fact, no attempt) is known in the literature.

As in Chapter 2, we can define the notion of moment solutions using the vanishing moment methodology for initial and initial-boundary value problems for (10.14). Mimicking the derivation of Chapter 2, we propose the following vanishing moment approximations to (10.14) and (10.15), respectively,

$$\epsilon\Delta^2 u^\epsilon + F(D^2u^\epsilon, Du^\epsilon, u^\epsilon, x, t) + \frac{\partial u^\epsilon}{\partial t} = 0, \tag{10.16}$$

$$-\epsilon\Delta^2 u^\epsilon + \det(D^2u^\epsilon) - \frac{\partial u^\epsilon}{\partial t} = f. \tag{10.17}$$

We note that each of the above equations is now a semi-ilinear fourth order parabolic PDEs.

By adopting the method of lines approach, generalizations of the numerical methods discussed in previous Chapters to the corresponding parabolic equations (10.16) and (10.17) are standard (cf. [43, 48] and the references therein). Assuming that an implicit time stepping method such as the backward Euler and the Crank-Nicolson schemes are used for time discretization, then at each time step we only need to solve a fully nonlinear elliptic equation of the form (10.3). As a result, all numerical methods discussed in Chapters 3–6 immediately apply. However, it should be pointed out that the convergence and error

analysis of all fully discrete schemes are expected to be harder, in particular, establishing error estimates which depend on ϵ^{-1} *polynomially* instead of *exponentially* will be very challenging.

10.5 Fast Solvers for Fully Nonlinear Second Order Equations

As expected, the discretization of (10.3) using any of the methods presented in this dissertation results in a large sparse nonlinear system. Thus, an efficient nonlinear solver must be employed to exploit this sparsity in its design and implementation. The most attractive solver for the discretization of (10.3) is Newton's method due to its ease of use and relatively fast rate of convergence. However, Newton's method requires a good initial guess for the algorithm to converge. To obtain a good initial guess and to speed up convergence, we propose a multi-resolution strategy, where we use solutions with larger values of ϵ as its initial guess for the case of smaller ϵ. In fact, we have used this method in many of the numerical tests in this dissertation, but we have not rigorously studied its convergence properties nor obtained optimal values of ϵ to make this method efficient.

Within each Newton iteration, a linear solver must be invoked. Direct solver techniques such as Gaussian elimination are too costly and destroy sparsity due to fill-in, and therefore linear iterative solvers become attractive. Unfortunately, the resulting algebraic system in the discretization of (10.3) is expected to be very ill-conditioned. In fact, using any standard discretization method for the biharmonic equation, the condition number of the resulting system is of order $O(h^{-4})$, where h denotes the mesh parameter in the numerical method. Worse yet, we expect that the parameter ϵ would cause the condition number to blow up further as $\epsilon \to 0^+$. As a result, standard iterative methods such as Gauss-Seidel and the Jacobi method will be very inefficient. However, linear systems of elliptic finite element methods can be solved in optimal computational order by multigrid methods, where by optimal, the computational cost is linear with respect to the unknowns. Furthermore, the convergence rate does not deteriorate when the discretization is refined. Another option to solve the linear system within each Newton iteration is the use of domain decomposition methods. The principle of domain decomposition methods is to divide the domain into overlapping or nonoverlapping subdomains in order to construct a preconditioner. In most cases, the resulting matrix has a condition number independent of the mesh parameter, and thus, one can apply iterative methods to achieve computational efficiency. We note there are many multigrid and domain decomposition methods for different discretizations for the biharmonic equation [14, 19, 20, 21, 62]. It then seems plausible to adapt these established solvers to the nonlinear equation (10.3), and as a result, we would obtain computationally

efficient solvers to compute fully nonlinear second order PDEs.

Bibliography

[1] S. Agmon, *Lectures on Elliptic Boundary Value Problems*, Van Nostrand Mathemetical Studies, Princeton, NJ, 1965.

[2] A. D. Aleksandrov, *Certain estimates for the Dirichlet problem*, Soviet Math. Dokl., 1:1151–1154, 1961.

[3] L. Ambrosio, *Lecture Notes on Optimal Transport problems*, in Mathematical Aspects of Evolving Interfaces, Lect. Notes in Math., 1812:1-52, 2003.

[4] L. Ambrosio, *Optimal Transport Maps in Monge-Kantorovich Problem*, Proceedings of he ICM, Higher Ed. Press, Beijing, 131-140, 2002.

[5] D. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Num. Anall, 19:742–760, 1982.

[6] F. E. Baginski and N. Whitaker, *Numerical solutions of boundary value problems for $\mathcal{K}$-surfaces in* $\mathbf{R}^3$, Numer. Methods for PDEs, 12(4):525–546, 1996.

[7] G. Baker, *Finite element methods for elliptic equations using nonconforming elements,* Math. Comp., 31:45–59, 1977.

[8] G. Barles and P. E. Souganidis, *Convergence of approximation schemes for fully nonlinear second order equations*, Asymptotic Anal., 4(3):271–283, 1991.

[9] T. Barth and J. Sethian, *Numerical schemes for the Hamilton-Jacobi and level set equations on triangulated domains*, J. Comput. Phys. 145(1):1–40, 1998.

[10] J. Benamou and Y. Brenier, *Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem*, SIAM. J. Appl. Math., 58(5):1450-1461, 1998.

[11] J.-D. Benamou and Y. Brenier, *A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, Numer. Math., 84(3):375–393, 2000.

[12] C. Bernardi and Y. Maday, *Spectral methods*, in *Handbook of numerical analysis, Vol. V*, pages 209–485. North-Holland, Amsterdam, 1997.

[13] K. Böhmer, *On finite element methods for fully nonlinear elliptic equations of second order*, SIAM J. Numer. Anal., 46(3):1212–1249, 2008.

[14] J.H. Bramble and X. Zhang, *Multigrid methods for the biharmonic problem discretized by conforming C^1 finite elements on nonnested meshes*, Numer. Funct. Anal. Optim. 16:835–846, 1995.

[15] Y. Brenier and G. Loeper, *A geometric approximation to the Euler equations: The Vlasov-Monge-Ampère System*, http://arxiv.org/abs/math/0504135v1.

[16] Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions.* Comm. Pure Appl. Math., 44:375–417, 1991.

[17] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, second edition, Springer, 2002.

[18] S. C. Brenner and L. Y. Sung, *C^0 interior penalty methods for fourth order elliptic boundary value problems on polygonal domains*, J. Sci. Comp., 22:83–118, 2005.

[19] S. C. Brenner and L. Y. Sung, *Multigrid algorithms for C^0 interior penalty methods*, SIAM J. Numer. Anal., 44(1):199–223, 2006.

[20] S. C. Brenner, *An optimal-order nonconforming multigrid method for the the biharmonic equation*, SIAM J. Numer. Anal., 26:1124–1138, 1989.

[21] S. C. Brenner, *Convergence of nonconforming multigrid methods without full elliptic regularity*, Math. Comp., 68(225):25–53, 1999.

[22] S. Bryson and D. Levy, *High-order central WENO schemes for multidimension Hamilton-Jacobi equations,* SIAM J. Numer. Anal. 41(4):1339–1369, 2003.

[23] L. A. Caffarelli and X. Cabré, *Fully nonlinear elliptic equations*, volume 43 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1995.

[24] L. A. Caffarelli and M. Milman, *properties of the solutions of the linearized Monge-Ampère equation*, Amer. J. Math., 119(2):423–465, 1997.

[25] L. A. Caffarelli and M. Milman, *Monge Ampère Equation: Applications to Geometry and Optimization*, *Contemporary Mathematics*, American Mathematical Society, Providence, RI, 1999.

[26] S. Y. Cheng and S. T. Yau, *On the regularity of the Monge-Ampère equation* $\det(\partial^2 u/\partial x_i \partial x_j) = F(x,u)$, Comm. Pure Appl. Math., 30(1):41–68, 1977.

[27] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems.* North-Holland, Amsterdam, 1978.

[28] B. Cockburn, *Continuous dependence and error estimation for viscosity methods*, Acta Numer., 12:127–180, 2003.

[29] COMSOL Multiphysics v3.5a, Comsol Group, Stockholm, Sweeden, 2008.

[30] M. G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc., 277(1):1–42, 1983.

[31] M. G. Crandall, L. Evans, and P.-L. Lions, *Some properties of viscosity solutions of the Hamilton-Jacobi equations*, Trans. Am. Math. Soc., 282(2):487–502, 1984.

[32] M. G. Crandall, H. Ishii, and P.-L. Lions, *User's guide to viscosity solutions of second order partial differential equations*, Bull. Amer. Math. Soc. (N.S.), 27(1):1–67, 1992.

[33] M. G. Crandall and P.-L. Lions, *Convergent difference schemes for nonlinear parabolic equations and mean curvature motion*, Numer. Math., 75(1):17–41, 1996.

[34] M. Cullen and M. Feldman, *Lagrangian solutions of semigeostrophic equations in physical space*, SIAM J. Math. Anal., 37:1371–1395, 2006.

[35] M. Cullen, J. Norbury, and R. J. Purser, *Generalized Lagrangian solutions for atmospheric and oceanic flows*, SIAM J. Appl. Math., 51:20–31, 1991.

[36] M. Cullen, R. Douglas, *Applications of the Monge-Ampère equation and Monge transport problem to meteorology and oceanography,* Contemporary Mathematics, 206:33–53, 1999.

[37] C. Dawson and M. Martinez-Canales, A characteristic-Galerkin approximation to a system of shallow water equations, Numer. Math., 86:239–256, 2000.

[38] T. A. Davis and I. S. Duff, *An unsymmetric-pattern multifrontal method for sparse LU factorization*, SIAM J. Matrix Anal. Appl., 18(1):140–158, 1997.

[39] E. J. Dean and R. Glowinski, *Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type*, Comput. Methods Appl. Mech. Engrg., 195(13-16):1344–1386, 2006.

[40] J. Douglas, Jr., *Numerical methods for the flow of miscible fluids in porous media*, in Numerical Methods in Coupled Systems (R. W. Lewis, P. Bettess, and E. Hinton eds.), John Wiley & Songs, New York.

[41] J. Douglas, Jr., R. Ewing, and M. Wheeler, *A time-discretization procedure for a mixed finite element approximation of miscible displacement in porous media*, R.A.I.R.O. Numer. Anal., 17(3):249-265, 1983.

[42] J. Douglas, Jr. and T. Russell, *Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures*, SIAM. J. Numer. Anal., 19(5):871-885, 1982.

[43] C. M. Elliott, D. A. French, and F. A. Milner, *A second order splitting method for the Cahn-Hillard equation*, Numer. Math., 54:575–590, 1989.

[44] L. C. Evans, *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI, 1998.

[45] L. Evans, *Partial differential equations and Monge-Kantorovich mass transfer*, in Current Developments in Mathematics, Int. Press, 65-126, 1999.

[46] L. Evans and W. Gangbo, *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc., 137, no. 653, 1999.

[47] R. S. Falk, J. E. Osborn, *Error estimates for mixed methods*, R.A.I.R.O. Anal. Numér., 14(3):249–277, 1980.

[48] X, Feng and O. A. Karakashian, *Fully discrete dynamic mesh discontinuous Galerkin methods for the Cahn-Hilliard equation of phase transition*, Math. Comp. 76:1093–1117, 2007.

[49] X. Feng, *Convergence of the vanishing moment method for the Monge-Ampère equations in two spatial dimensions*, submitted to Trans. AMS.

[50] X. Feng and M. Neilan, *Vanishing moment method and moment solutions for second order fully nonlinear partial differential equations*, J. Scient. Comp., 38(1):74–98, 2009.

[51] X. Feng and M. Neilan, *Mixed finite element methods for the fully nonlinear Monge-Ampère equation based on the vanishing moment method*, SIAM J. Numer. Anal., 47(2):1226–1250, 2009.

[52] X. Feng and M. Neilan, *Analysis of Galerkin methods for the fully nonlinear Monge-Ampère equation*, submitted to Math. Comp.

[53] X. Feng and M. Neilan, *A modified characteristic finite element method for a fully nonlinear formulation of the semigeostrophic ow equations*, accepted in SIAM J. Numer. Anal.

[54] X. Feng, M. Neilan, and A. Prohl, *Error analysis of finite element approximations of the inverse mean curvature flow arising from the general relativity*, Numer. Math. 108:93–119, 2007.

[55] W. H. Fleming and H. M. Soner, *Controlled Markov Processes and Viscosity Solutions*, volume 25 of *Stochastic Modelling and Applied Probability*. Springer, New York, second edition, 2006.

[56] W. Gangbo and R. J. McCann, *The geometry of optimal transport*, Acta Math., 177:113-161, 1996.

[57] D. Gilbarg and N. S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, *Classics in Mathematics*, Springer-Verlang, Berlin, 2001. Reprint of the 1998 edition.

[58] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman (Advanced Publishing Program), Boston, MA, 1985.

[59] B. Guan, *On the existence and regularity of hypersurfaces of prescribed Gauss curvature with boundary,* Indiana Univ. Math. J. 44(1):21–241, 1995.

[60] C. E. Gutierrez, *The Monge-Ampère Equation*, volume 44 of *Progress in Nonlinear Differential Equations and Their Applications*, Birkhauser, Boston, MA, 2001.

[61] C. E. Gutierrez and Q. Huang, $W^{2,p}$ *estimates for the parabolic Monge-Ampère equation*, Arch. Ration. Mech. Anal., 159:137-177, 2001.

[62] M. R. Hanisch, *Multigrid preconditioning for the biharmonic Dirichlet problem*, SIAM J. Numer. Anal. 30:184–214, 1993.

[63] B. J. Hoskins, *The geostrophic momentum approximation and the semigeostrophic equations*, J. Atmospheric Sci., 32:233–242, 1975.

[64] H. Ishii, *On uniqueness and existence of viscosity solutions of fully nonlinear second order PDE's*, Comm. Pure Appl. Math., 42:14–45, 1989.

[65] R. Jensen, *The maximum principle for viscosity solutions of fully nonlinear second order partial differential equations*, 101:1–27, 1988.

[66] L. V. Kantorovich, *On the transfer of masses*, Dokl. Akad. Nauk. SSSR 37:227-229, 1942.

[67] O. A. Ladyzhenskaya and N. N. Ural'tseva, *Linear and Quasilinear Elliptic Equations*, Academic Press, New York, 1968.

[68] G. M. Lieberman, *Second order parabolic differential equations*, World Scientific, Singapore, 1996.

[69] G. Loeper, *A fully non-linear version of the incompressible Euler equations: The semigeostrophic system*, http://arxiv.org/abs/math/0504138v1.

[70] Y. Lucet, *Faster than the fast Legendre transform, the linear-time Legendre transform,* Numer. Algorithms 16(2):171–185, 1998.

[71] A. Majda, *Introduction to PDEs and Waves for Atmosphere and Ocean*, American Mathematical Society, 2003.

[72] MATLAB v.7.8, The MathWorks, Natick, MA, 2009.

[73] R. J. McCann and A. M. Oberman. *Exact semi-geostrophic flows in an elliptical ocean basin*, Nonlinearity, 17(5):1891–1922, 2004.

[74] W. Ming and J. Xu, *The Morley element for fourth order elliptic equations in any dimension*, Numer. Math., 103:155–169, 2006.

[75] G. Monge, *Memoire sur la thèorie des dèblais at des remblais*, Historire de l'Académie Royale des Sciences de Paris, p.666-704, 1781.

[76] L.S.D. Morley, *The triangular equilibrium element in the solution of plate bending problems,* Aero. Quart., 19:149–169, 1968.

[77] I. Mozolevski and E. Süli, *A priori error analysis for the hp-version of the discontinuous Galerkin finite element method for the biharmonic equation*, Comput. Meth. Appl. Math., 3:596–607, 2003.

[78] M. Neilan, *A nonconforming Morley finite element method for the fully nonlinear Monge-Ampère equation*, submitted to Numer. Math.

[79] A. M. Oberman, *Wide stencil finite difference schemes for the elliptic Monge-Ampére equation and functions of the eigenvalues of the Hessian*, Discrete Contin. Dyn. Syst, 1:221–238, 2008.

[80] V. I. Oliker and L. D. Prussner, *On the numerical solution of the equation* $(\partial^2 z/\partial x^2)(\partial^2 z/\partial y^2) - ((\partial^2 z/\partial x \partial y))^2 = f$ *and its discretizations. I.*, Numer. Math., 54(3):271–293, 1988.

[81] S. Osher and J. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys. 79(1):12–49, 1988.

[82] S. Osher and S. W. Shu, *High-order essentially nonoscillatory schemes for Hamilton-Jacobi equations*, SIAM J. Numer. Anal., 28(4):907–922, 1991.

[83] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2003.

[84] Eun-Jae Park, *Mixed finite element methods for nonlinear second-order elliptic problems*, SIAM J. Numer. Anal., 32(3):865–885, 1995.

[85] S. Rachev, *The Monge-Kantorovich mass transference problem and its stochastic applications*, Theory of Prob. and Appl., 29:647-676, 1984.

[86] M. Saum, *Adaptive Discontinuous Galerkin Finite Element Methods for Second and Fourth Order Elliptic Partial Differential Equations*, PhD Thesis. University of Tennessee, 2006.

[87] J. A. Sethian, *Level set methods and fast marching methods. Evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science.* Second edition. Cambridge University Press, Cambridge, 1999.

[88] J. Shen, *Efficient spectral-Galerkin method. I. Direct solvers of second- and fourth-order equations using Legendre polynomials*, SIAM J. Sci. Comput., 15(6):1489–1505, 1994.

[89] Z.-C. Shi, *On the error estimate of Morley element.* Numerica Mathematica Sinica 12(2):113–118, 1990.

[90] T. Nilssen, X. -C. Tai, and R. Wagner, *A robust nonconfirming H^2 element*, Math. Comp., 70:489–505, 2000.

[91] R. Thelwell, *The Nonlinear Balance Equation: a Survey of Numerical Methods.* Masters Thesis. Colorado State University, 2005.

[92] D. W. Thoe and E.C Zachmanoglou, *Introduction to Partial Differential Equations with Applications*, Dover Publications, New York, NY, 1986.

[93] M. Wang, Z. Shi, and J. Xu, *A new class of Zienkiewicz-type non-conforming elements in any dimension*, Numer. Math., 106(2):335–347, 2007.

[94] M. Wang and J. Xu, *Some tetrahedron nonconforming elements for fourth order elliptic equations*, Math. Comp., 76:1–18, 2007.

[95] L. Wang, *On the regularity theory of fully nonlinear parabolic equations*, I. Commun. Pure Appl. Math., 45:27–76, 1992.

Appendices

Appendix A: Useful Results

In this section, we state many well-known results that are used throughout the dissertation. The first concerns the divergence-free row property for the cofactor matrices. The proof of the lemma can be found in [44, p. 440].

Lemma A.0.1. *Let* $\mathbf{v} = (v_1, v_2, \cdots, v_n) : \Omega \to \mathbf{R}^n$ *be given a vector-valued function, and assume* $\mathbf{v} \in [C^2(\Omega)]^n$. *Then the cofactor matrix* $\mathrm{cof}(D\mathbf{v})$ *of the gradient matrix* $D\mathbf{v}$ *of* $\mathbf{v}$ *satisfies the following row divergence-free property:*

$$\mathrm{div}(\mathrm{cof}(D\mathbf{v}))_i = \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(\mathrm{cof}(D\mathbf{v}))_{ij} = 0 \qquad \text{for } i = 1, 2, \cdots, n, \tag{A.18}$$

where $(\mathrm{cof}(D\mathbf{v}))_i$ *and* $(\mathrm{cof}(D\mathbf{v}))_{ij}$ *denote respectively the* i*th row and the* (i,j)*-entry of* $\mathrm{cof}(D\mathbf{v})$.

The next theorem bounds the interpolation error for affine families of finite elements.

Theorem A.0.2 ([27], Theorem 3.1.6)**.** *Let there be a regular affine family of finite elements* (K, P_K, Σ_K) *whose reference finite element* $(\hat{K}, \hat{P}, \hat{\Sigma})$ *satisfies the following:*

- $W^{k+1,p}(\hat{K}) \hookrightarrow C^s(\hat{K})$,
- $W^{k+1,p}(\hat{K}) \hookrightarrow W^{m,q}(\hat{K})$,
- $\mathbb{P}_k(\hat{K}) \subset \hat{P} \subset W^{m,q}(\hat{K})$,

where $\mathbb{P}_k$ *denotes the space of polynomials of degree less than or equal to* k. *Then there exists a constant* C, *such that for all finite elements* K *in the family, and all functions* $v \in W^{k+1,p}(K)$,

$$\|v - I_K v\|_{W^{m,q}(K)} \leq C h^{n(\frac{1}{q}-\frac{1}{p})} h^{k+1-m} |v|_{W^{k+1,p}(K)}, \tag{A.19}$$

where $I_K v$ *denotes the standard interpolation operator of* v.

We note that the Argyris finite element used in Chapter 3 is not affine-equivalent. Thus, we need the following result.

Theorem A.0.3 ([27], Theorem 6.1.1)**.** *A regular family of Argyris triangles is almost-affine. That is, for all* $p \in [1, \infty]$ *and all pairs* (m, q) *with* $m \geq 0$ *and* $q \in [1, \infty]$ *compatible with the inclusion* $W^{6,p}(K) \hookrightarrow W^{m,q}(K)$, *there exists a constant* C *independent of* K *such that for all* $v \in W^{6,p}(K)$,

$$\|v - I_K v\|_{W^{m,q}(K)} \leq C h^{n(\frac{1}{q}-\frac{1}{p})} h^{6-m} |v|_{W^{6,p}(K)}. \tag{A.20}$$

Theorem A.0.4 (The Inverse Inequality – [27], Theorem 3.2.6)**.** *Let $\mathcal{T}_h$ denote a regular family of triangulations, and suppose all the finite elements (K, P_K, Σ_K) are affine equivalent to a single reference finite element $(\hat{K}, \hat{P}, \hat{\Sigma})$. Suppose that for two pairs (ℓ, r) and (m, q) with $\ell, m \geq 0$ and $(r, q) \in [1, \infty]$ such that $\ell \leq m$ and $\hat{P} \subset W^{\ell,r}(\hat{K}) \cap W^{m,q}(\hat{K})$. Then there exists a constant $C > 0$ such that for all v_h in the finite elements space,*

$$\left(\sum_{K \in \mathcal{T}_h} |v_h|^q_{W^{m,q}(K)} \right)^{\frac{1}{q}} \leq C h^{n(\frac{1}{q} - \frac{1}{r})} h^{m-\ell} \left(\sum_{K \in \mathcal{T}_h} |v_h|^r_{W^{\ell,r}(K)} \right)^{\frac{1}{r}}. \tag{A.21}$$

Theorem A.0.5 (Brouwer's Fixed Point Theorem - [44], p.441)**.** *Assume*

$$u : B(0,1) \to B(0,1)$$

is continuous, where $B(0,1)$ denotes the closed unit ball in $\mathbf{R}^n$. Then u has a fixed point, that is, there exists a point $x \in B(0,1)$ with $u(x) = x$.

Remark A.0.6. *We note that Brouwer's Theorem can be extended to continuous mappings of closed convex bodies in an $n-$dimensional topological vector space.*

Theorem A.0.7 (Trace Theorem I - [44] p.258)**.** *Assume Ω is bounded and $\partial\Omega$ is C^1. Then for $p \in [1, \infty]$, there exists a bounded linear operator*

$$T : W^{1,p}(\Omega) \to L^p(\partial\Omega)$$

such that $Tu = u\big|_{\partial\Omega}$ if $u \in W^{1,p}(\Omega) \cap C^0(\bar{\Omega})$ and $\|Tu\|_{L^p(\partial\Omega)} \leq C\|u\|_{W^{1,p}}$.

Theorem A.0.8 (Trace Theorem II - [17], Theorem 1.6.6)**.** *Suppose that Ω has a Lipschitz boundary, and that p is a real number in the range $1 \leq p \leq \infty$. Then there is a constant, C, such that*

$$\|Tv\|_{L^p(\partial\Omega)} \leq C\|v\|^{1-\frac{1}{p}}_{L^p} \|v\|^{\frac{1}{p}}_{W^{1,p}} \qquad \forall v \in W^{1,p}(\Omega).$$

Theorem A.0.9 (Poincare's Inequality)**.** *If Ω is a bounded domain that can be written as a finite union of domains that are star-shaped with respect to a ball there, then there is a constant $C < \infty$ such that*

$$\|v\|_{W^{1,p}} \leq C|v|_{W^{1,p}} \qquad \forall W^{1,p}_0(\Omega).$$

Appendix B: Numerical Test Data

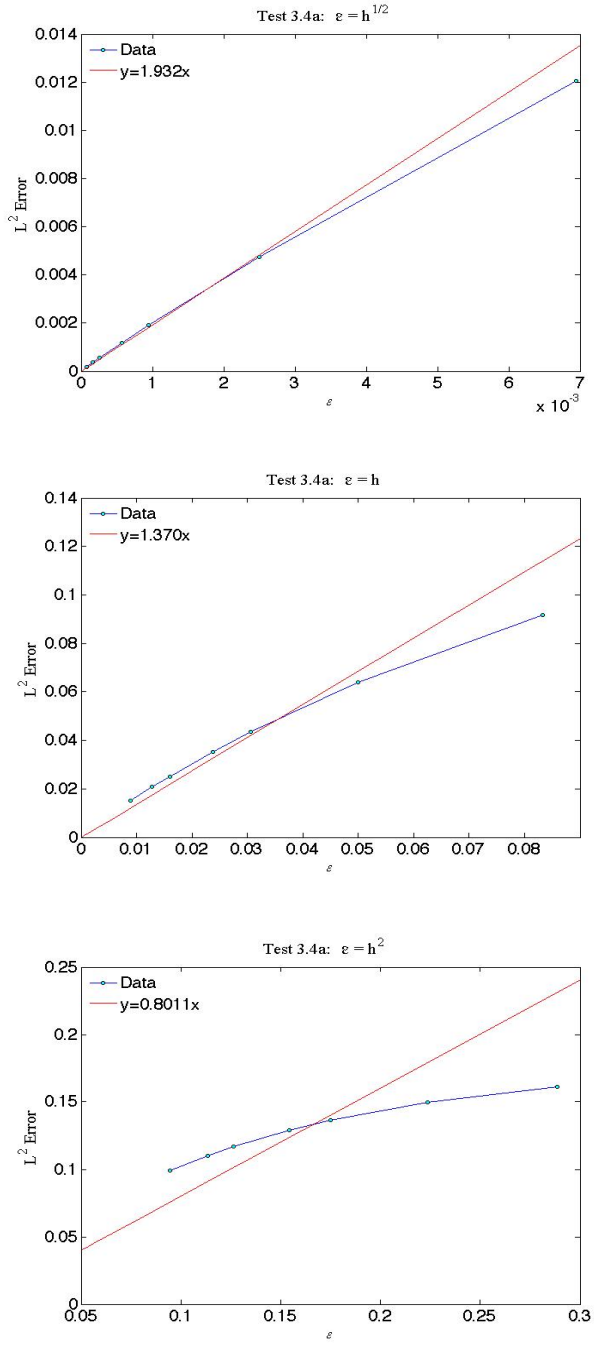


Figure B.1: Test 3.4a. L^2 Error of u_h^ϵ

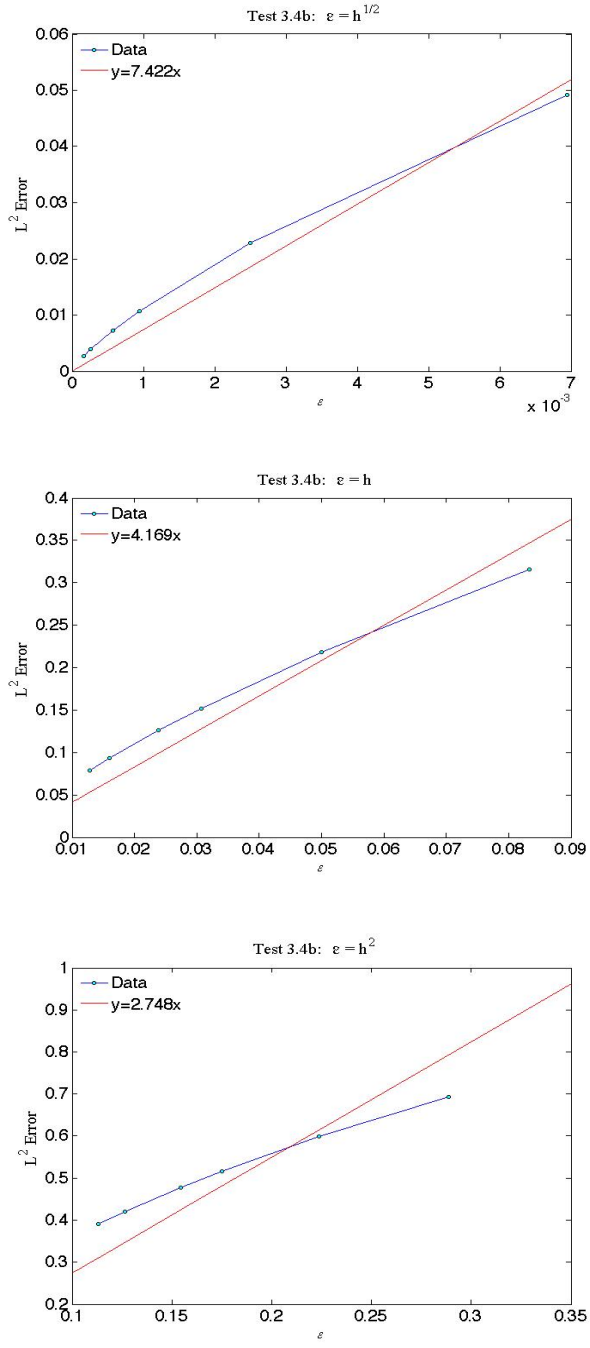


Figure B.2: Test 3.4b. L^2 Error of u_h^ϵ

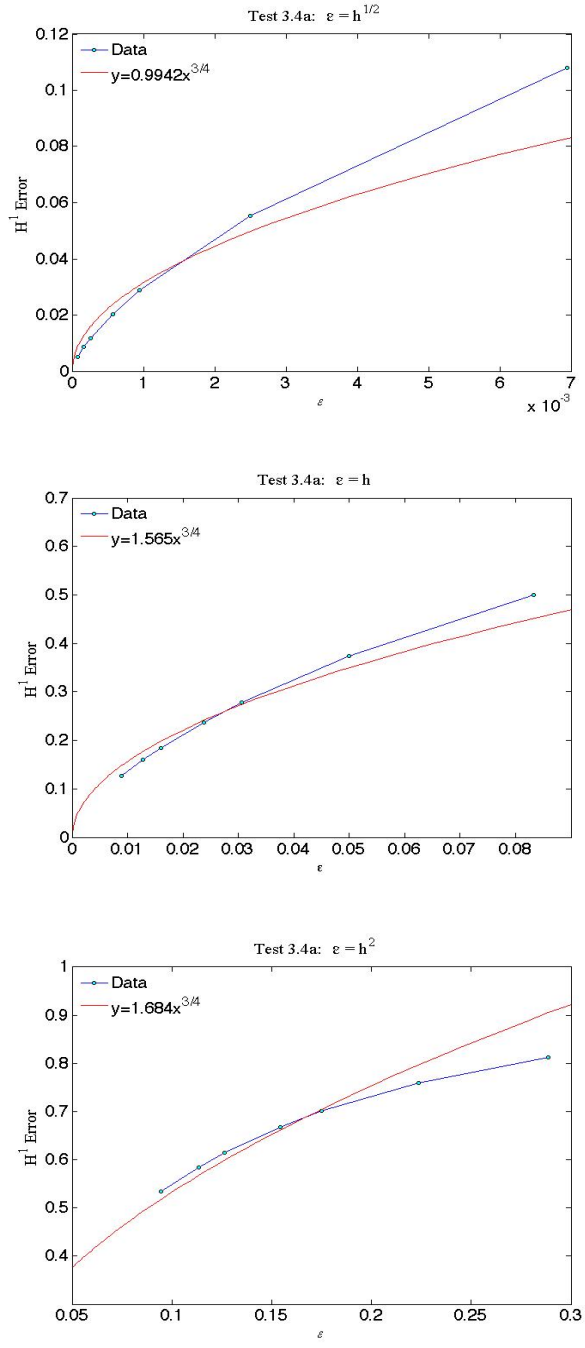


Figure B.3: Test 3.4a. H^1 Error of u_h^ϵ

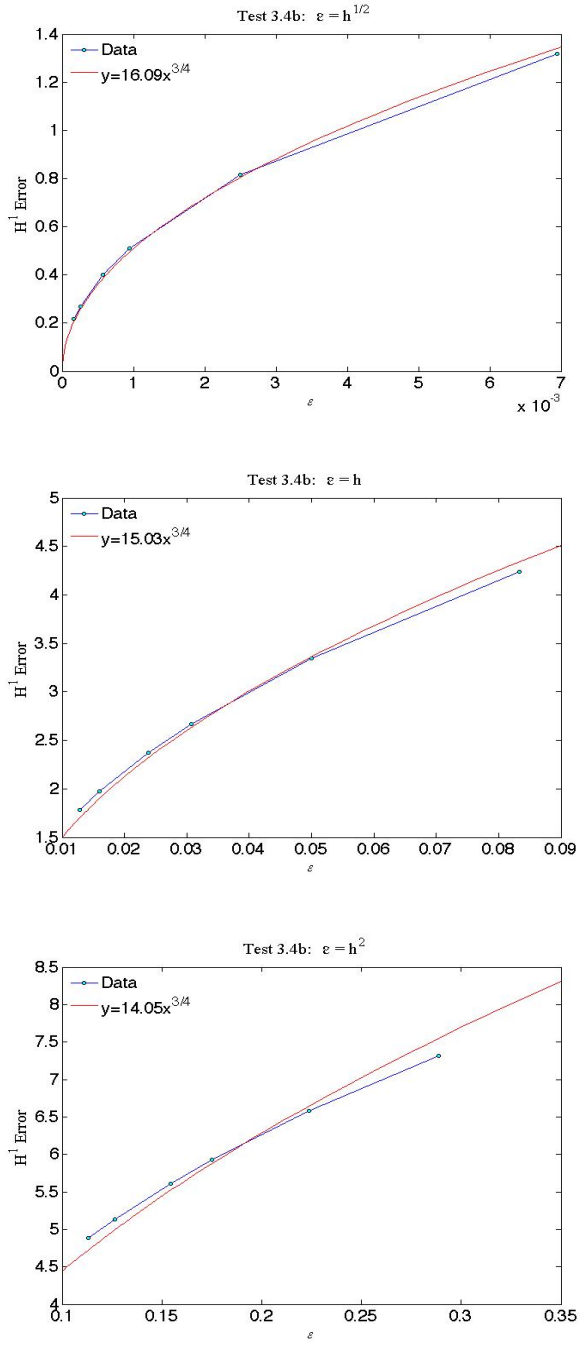


Figure B.4: Test 3.4b. H^1 Error of u_h^ϵ

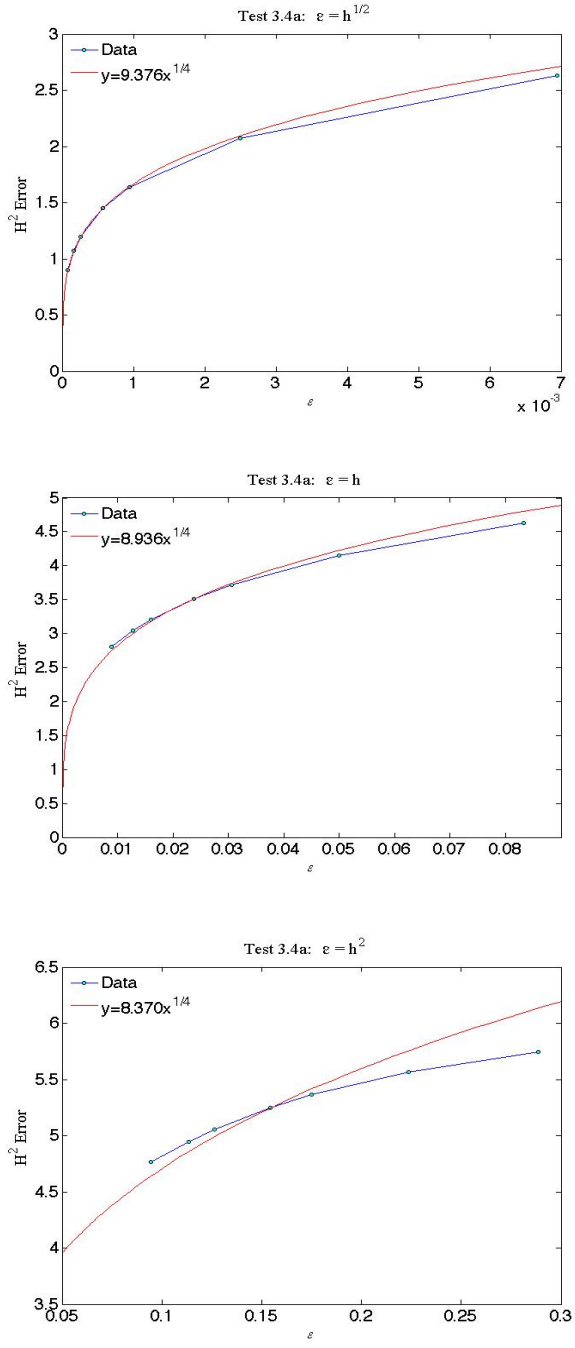


Figure B.5: Test 3.4a. H^2 Error of u_h^ϵ

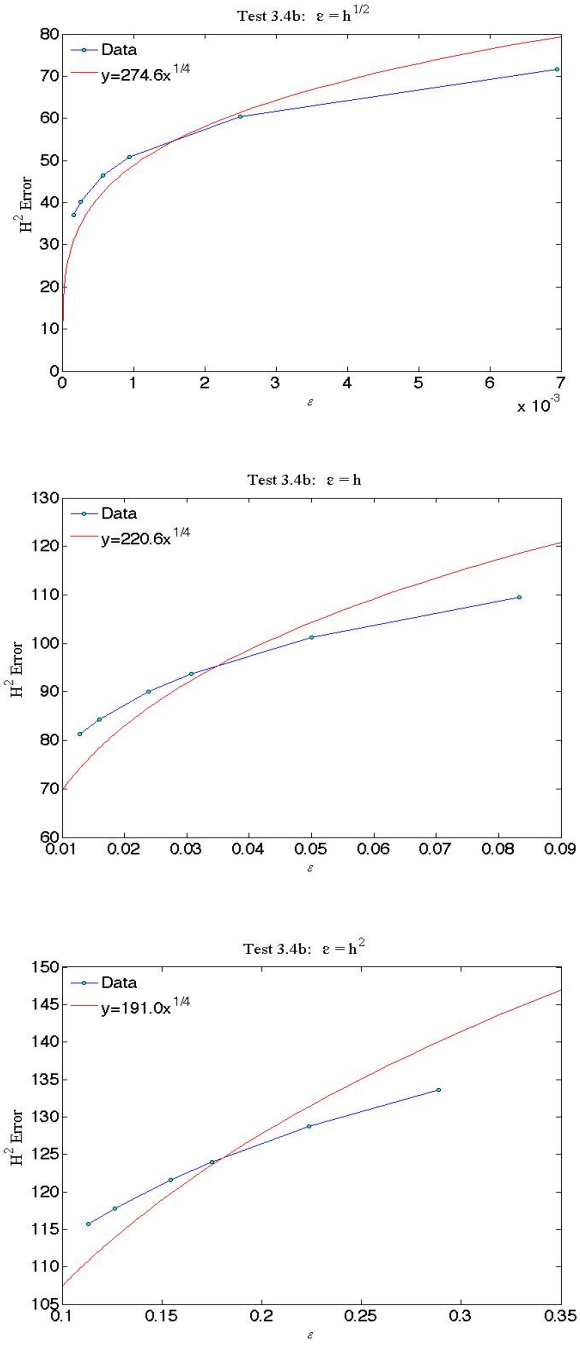
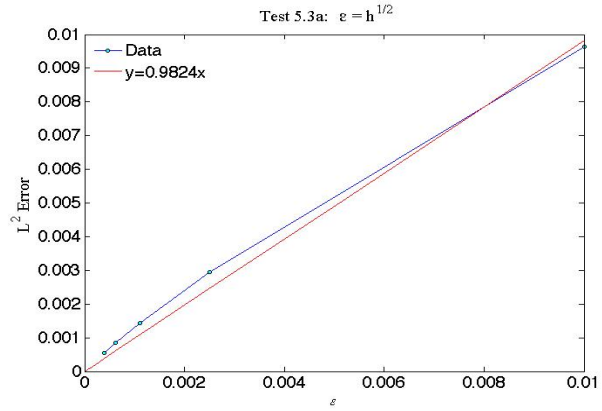
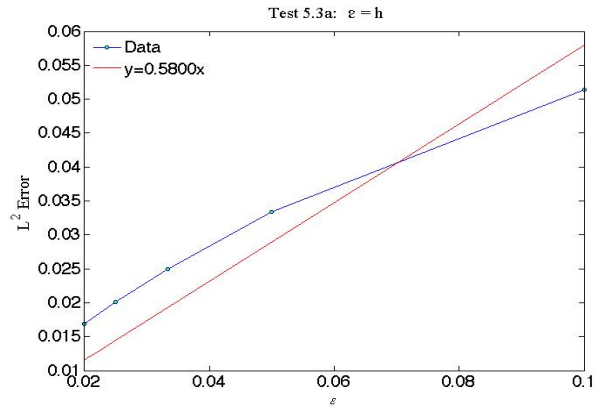


Figure B.6: Test 3.4b. H^2 Error of u_h^ϵ





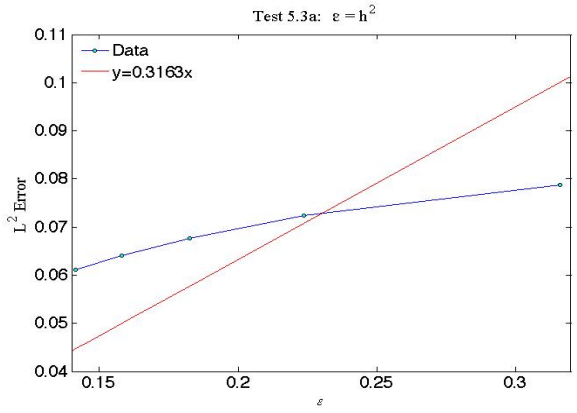
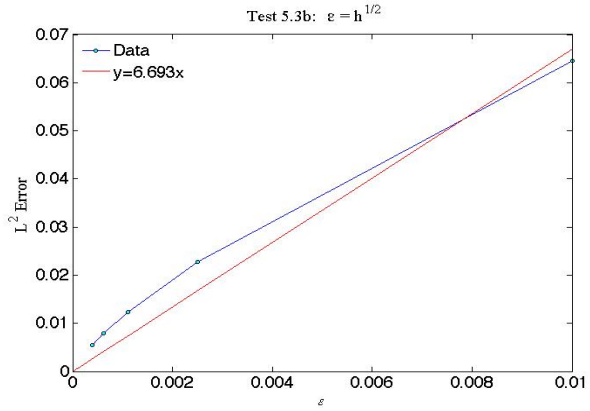
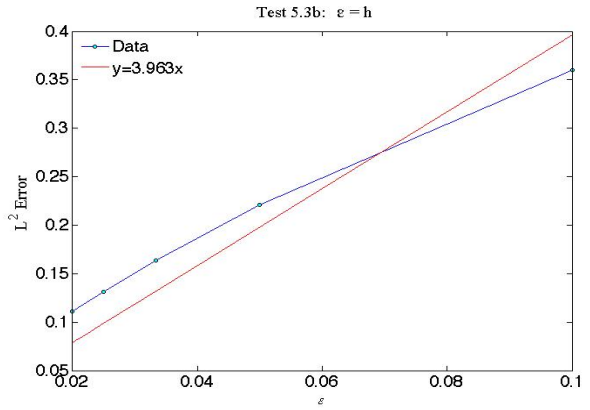


Figure B.7: Test 5.3a. L^2 Error of u_h^ϵ





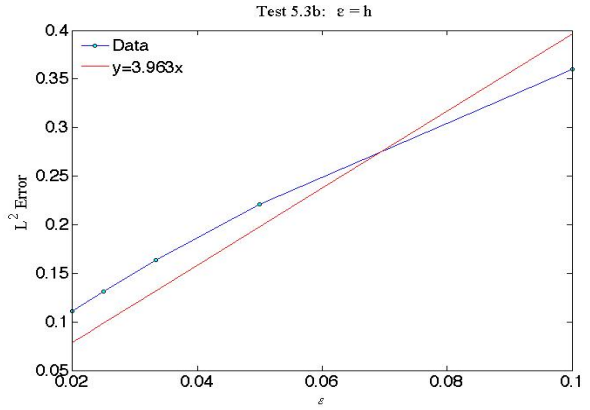
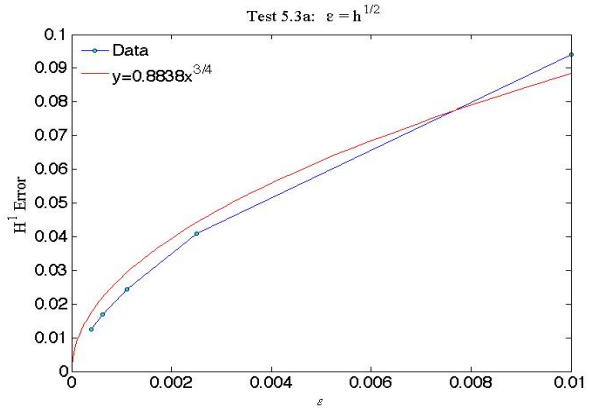
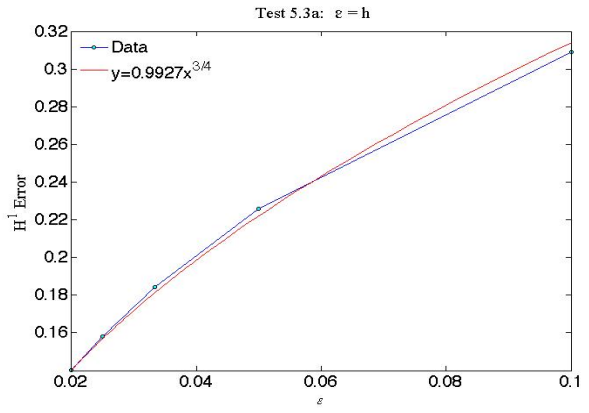


Figure B.8: Test 5.3b. L^2 Error of u_h^ϵ





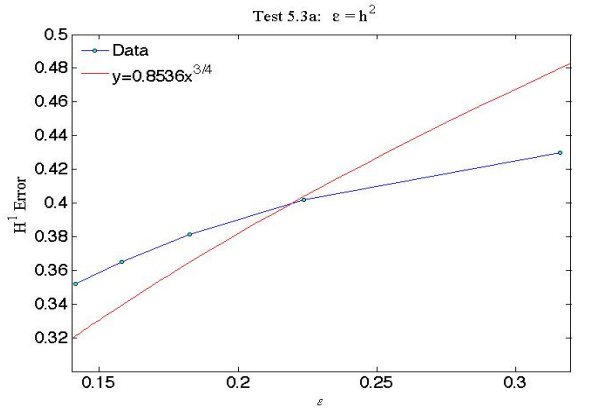


Figure B.9: Test 5.3a. H^1 Error of u_h^ϵ

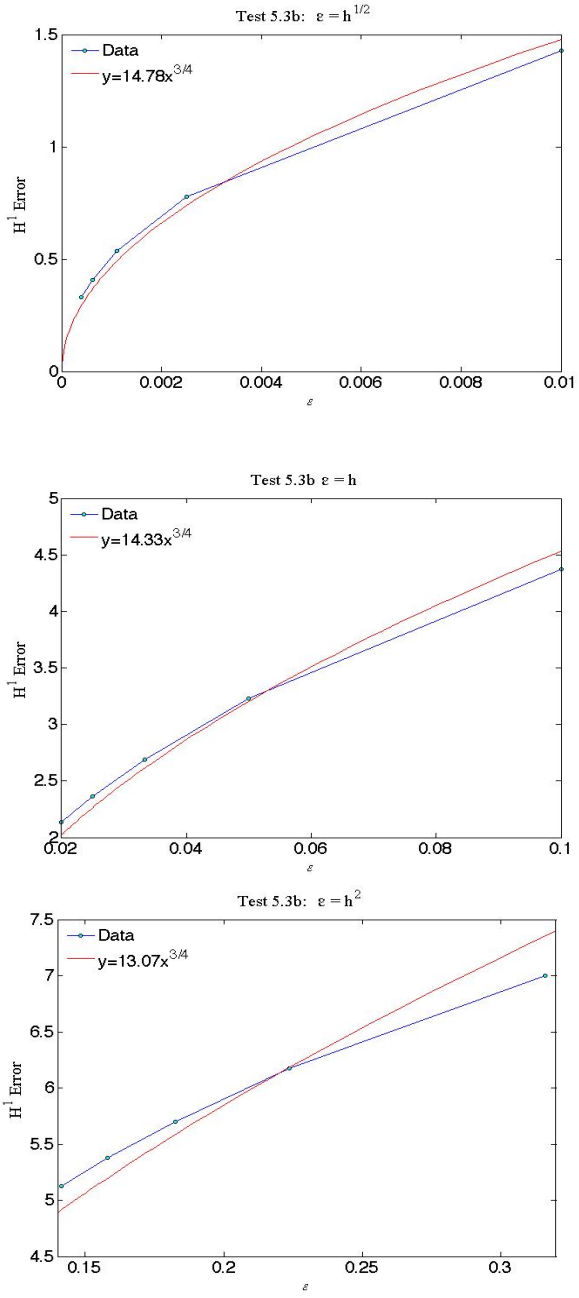


Figure B.10: Test 5.3b. H^1 Error of u_h^ϵ

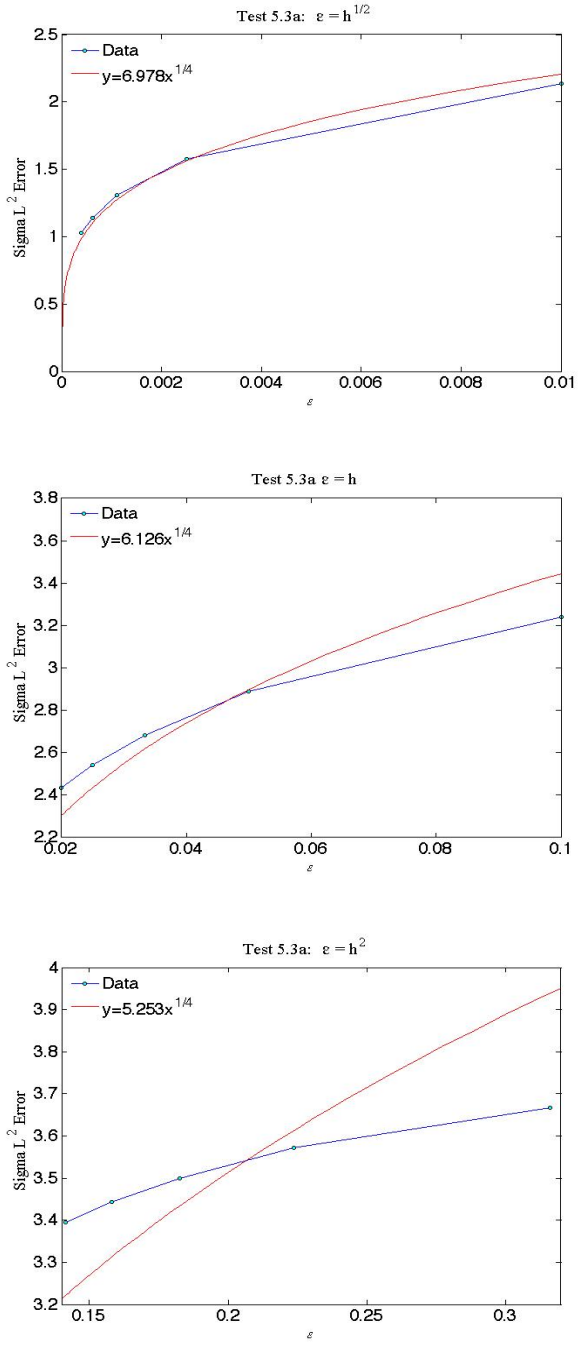
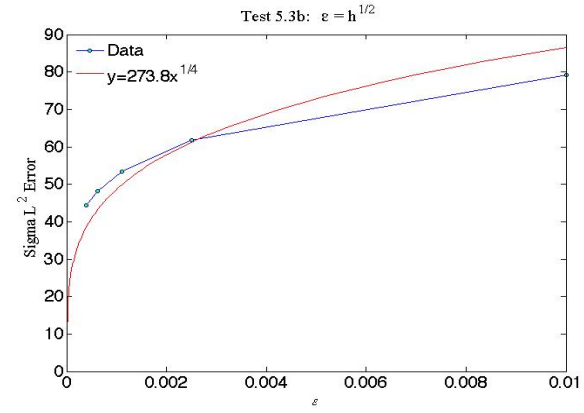
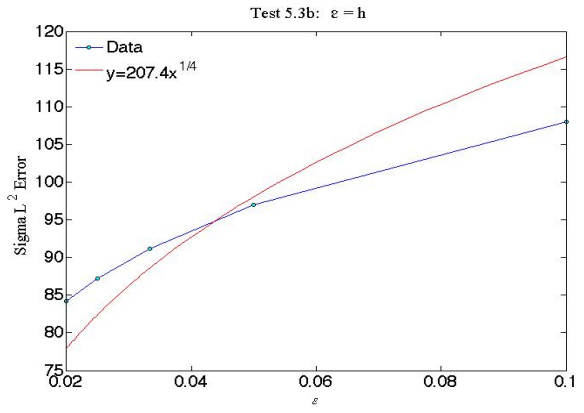


Figure B.11: Test 5.3a. L^2 Error of σ_h^ϵ





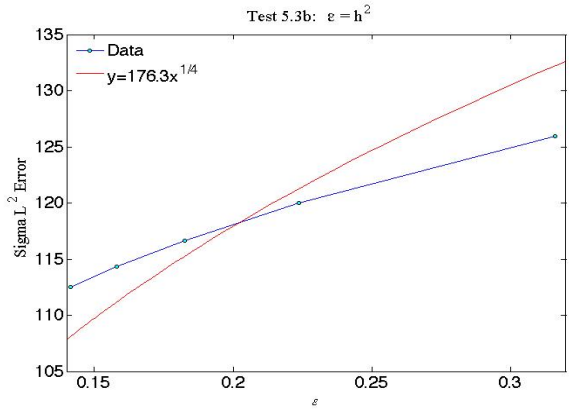


Figure B.12: Test 5.3b. L^2 Error of σ_h^ϵ

Vita

Michael Joseph Neilan was born in Wilmington, Delaware on February 28, 1982 to James Patrick Neilan and Irene Margaret Neilan. In 1994, he moved to Brentwood, Tennessee where he remained until he completed his work at Brentwood High School in 2000. In the same year, he enrolled in the University of Tennessee - Knoxville, where in 2004 he received his Bachelor degrees in Mathematics and Computer Science. The following year, he entered the Mathematics PhD program at the University of Tennessee.