8-2017

Numerical Methods for Non-divergence Form Second Order Linear Elliptic Partial Differential Equations and Discontinuous Ritz Methods for Problems from the Calculus of Variations

Stefan Raymond Schnake
University of Tennessee, Knoxville, sschnak1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Stefan Raymond Schnake entitled "Numerical Methods for Non-divergence Form Second Order Linear Elliptic Partial Differential Equations and Discontinuous Ritz Methods for Problems from the Calculus of Variations." 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, Tuoc Phan, Stanimire Tomov

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

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

Numerical Methods for Non-divergence Form Second Order Linear Elliptic Partial Differential Equations and Discontinuous Ritz Methods for Problems from the Calculus of Variations

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Stefan Raymond Schnake

August 2017

This dissertation is dedicated to my wife Kacy and my parents Betty and David who have given me overwhelming love and support during my time in Knoxville.

Acknowledgements

I would like to thank all of those who have supported me during my Ph.D. study here at the University of Tennessee. I foremost would like to thank my advisor, Professor Xiaobing Feng, for his countless hours of contribution to my Ph.D. research as well as the innumerable words of wisdom. Every week Dr. Feng and I would spend two hours in his office, and after every occasion I would walk out of his office learning something novel - mathematically related or not. His guidance and willingness to share his previous experiences has shaped me into the mathematician writing this prose today, and I will always remember his advice during my future academic endeavors.

I would also like to thank my doctoral committee: Professors Ohannes Karakashian, Tuoc Phan, and Stanimire Tomov. The critiques and points raised by Dr. Karakashian during my presentations were extremely valuable; in addition, I have gained much from his insight during the courses I have taken under him. Dr. Phan's advanced PDE course presented me with the beauty of elliptic existence and regularity theory and was influential in my passion for elliptic PDEs. Finally, Dr. Tomov's lectures on high performance computing reminded me of the big step one has from the mathematical construction of the methods presented in this dissertation to the efficient and practical implementations of said methods.

In addition, I would like to thank a few other mathematical faculty members - Professors Steven Wise, Abner Salgado and Vasilios Alexiades - whose unique interpretations and specialties have given me a greater breadth of knowledge in the field of numerical analysis. In this list as well is Professor Ted Porter, my academic advisor at Murray State University, who kindled my growing interest in Mathematics and was influential in my decision to pursue a Ph.D.

I would like to thank my family and friends for their emotional support and encouragement. Specifically, I would like to thank my parents-in-law, Dennis and Jane-Ann Aslinger and my parents Betty and David Schnake. Finally, I would like to thank my wife Kacy, whom I met in Knoxville, for the abundance of love and support she has provided.

Finally, I would like to acknowledge the support of the NSF grants DMS-1620168 and DMS-1318486 for my research from June of 2015 to July of 2017.

Greater love hath no man than to lay down his life for a friend. -John 15:13

Abstract

This dissertation consists of three integral parts. Part one studies discontinuous Galerkin approximations of a class of non-divergence form second order linear elliptic PDEs whose coefficients are only continuous. An interior penalty discontinuous Galerkin (IP-DG) method is developed for this class of PDEs. A complete analysis of the proposed IP-DG method is carried out, which includes proving the stability and error estimate in a discrete $W^{2,p}$-norm [Wˆ2,p-norm]. Part one also studies the convergence of the vanishing moment method for this class of PDEs. The vanishing moment method refers to a PDE technique for approximating these PDEs by a family of fourth order PDEs. Detailed proofs of uniform H^1 [Hˆ1] and H^2 [Hˆ2]-stability estimates for the approximate solutions and their convergence are presented.

Part two studies finite element approximations of a class of calculus of variations problems which exhibit so-called Lavrentiev gap phenomenon (LGP), whose solutions often contain singularities. The LGP incapacitates all standard numerical methods, especially the finite element method, as they fail to produce a correct approximate solution. To overcome the difficulty, an enhanced finite element method based on a truncation technique is developed in this part of the dissertation. The proposed enhanced finite element method is shown to numerically converge on several benchmark problems with the LGP.

Part three of the dissertation develops a discontinuous Galerkin numerical framework for general calculus of variations problems, which is called the discontinuous Ritz (DR) methodology and can be regarded as the counterpart of the discontinuous

Galerkin (DG) methodology for PDEs. Conceptually, it approximates the admissible space by the DG spaces which consist of totally discontinuous piecewise polynomials and approximates the underlying energy functional by discrete energy functionals defined on the DG spaces. The main idea here is to construct the desired discrete energy functional by using the newly developed DG finite element calculus theory, which only requires replacing the gradient operator in the energy functional by the corresponding DG finite element discrete gradient and adding the standard interior penalty terms. It is shown that for a certain class of functionals the proposed DR method does indeed converge to the true solution.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Prelude

Differential Equations describe relations between a function and its derivatives. A Partial Differential Equation (PDE) is a differential equation which involves a multivariate function and its partial derivatives. PDEs are ubiquitous and appear as mathematical models for many application problems from physical and biological sciences and engineering. This dissertation focuses on second order PDEs which have a general form

$$F(D^2u(x), \nabla u(x), u(x), x) = 0, \tag{1.1.1}$$

where $F : \mathbb{R}^{d\times d} \times \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}$, $d \geq 1$ is the dimension, Ω is an open, bounded domain in $\mathbb{R}^d$, and $u : \Omega \to \mathbb{R}$ is the unknown solution. Here ∇u denotes the gradient of u, i.e.,

$$\nabla u = \left[u_{x_1}, u_{x_2}, \cdots, u_{x_d}\right],$$

and D^2u denotes the Hessian of u, i.e.,

$$D^2u = \begin{bmatrix} u_{x_1x_1} & u_{x_1x_2} & \cdots & u_{x_1x_d} \\ u_{x_2x_1} & u_{x_2x_2} & \cdots & u_{x_2x_d} \\ \vdots & \vdots & \ddots & \vdots \\ u_{x_dx_1} & u_{x_dx_2} & \cdots & u_{x_dx_d} \end{bmatrix}.$$

There are two primary sources which produce PDEs that are of importance and interest to study analytically and approximate numerically. First, a PDE arises as a mathematical description of a natural, physical or biological law or process. For example, the well-known diffusion equation

$$u_t - \operatorname{div}(D\nabla u) = S \tag{1.1.2}$$

describes the conversation law of mass:

$$u_t + \operatorname{div}(F(u)) = S,$$

combined with Fick's law of diffusion:

$$F(u) = -D\nabla u.$$

Another example is the following celebrated Navier-Stokes equations for incompressible inviscid fluids (c.f. [21]):

$$\begin{aligned} \boldsymbol{u}_t - \nu\Delta\boldsymbol{u} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} + \nabla p = \boldsymbol{f}, \\ \operatorname{div}(\boldsymbol{u}) = 0, \end{aligned} \tag{1.1.3}$$

where $\boldsymbol{u}$ is the velocity field of the fluid, p denotes the pressure, $\nu > 0$ is the viscosity coefficient, and $\boldsymbol{f}$ is the body force acting on the fluid. Here in the system, the first equation, called the momentum equation, describes the conservation of momentum and is the mathematical description of Newton's second law of motion. The second equation, called the continuity equation, is the mathematical statement of the conservation of mass (or the incompessibility).

Second, a PDE arises as the so-called Euler-Lagrange equation of a calculus of variations problem. To illustrate this point, we must introduce some notation. Let V be some function space (called the admissible set) and let $\mathcal{J}$ be a functional on V (called the energy), that has the following form:

$$\mathcal{J}(v) = \int_\Omega f(\nabla u(x), u(x), x)\,\mathrm{d}x, \tag{1.1.4}$$

where Ω is an open, bounded domain in $\mathbb{R}^d$ and $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}$ is called the density function.

The calculus of variations seeks a function $u \in V$ such that

$$\mathcal{J}(u) \le \mathcal{J}(v) \quad \forall v \in V. \tag{1.1.5}$$

Such a u, if it exists, is called a *minimzer* of $\mathcal{J}$ over V and is written as

$$u \in \operatorname*{arg\,min}_{v \in V} \mathcal{J}(v). \tag{1.1.6}$$

For example, if $d = 1$, then the shortest path for a particle to move under the force of gravity from the point (a, α) to the point (b, β) in the plane is given by the minimization problem (1.1.5) with

$$\Omega = (a, b), \quad f(\xi, v, x) = \frac{\sqrt{1+\xi^2}}{\sqrt{2gv}},$$

and the energy space

$$V = \{v \in C^1([a, b]) : v(a) = \alpha, v(b) = \beta\},$$

where g is the gravitational constant. This is the famous Brachistochrome problem (see [19]).

Another example is the minimal surface problem, an important problem in the field of differential geometry, which seeks a function $u : \Omega \to \mathbb{R}$ to solve (1.1.5) with the density function

$$f(\xi, v, x) = \sqrt{1 + \xi^2}$$

and the energy space

$$V = \{v \in W^{1,1}(\Omega) : v\big|_{\partial\Omega} = \varphi\}.$$

where φ is some given function on the boundary. Here the graph of the minimizer u is a minimal surface because the mean curvature is zero at every point on the graph.

Lastly, it can be shown that the solution u to the Laplace equation

$$\Delta u = 0 \qquad \text{in } \Omega, \tag{1.1.7a}$$

$$u = \varphi \qquad \text{on } \partial\Omega, \tag{1.1.7b}$$

also solves problem (1.1.5) with the energy functional

$$\mathcal{J}(v) = \int_\Omega \frac{1}{2} |\nabla v|^2 \,\mathrm{d}x \tag{1.1.8}$$

and energy space

$$V = \{v \in H^1(\Omega) : v\big|_{\partial\Omega} = \varphi\}.$$

Here (1.1.8) is called the Dirichlet integral (c.f. [19]).

In general, if u is a minimizer of (1.1.5), then it must satisfy the Euler-Lagrange equation:

$$\sum_{i=1}^{d} \frac{\partial}{\partial x_i} \left(f_{\xi_i}(\nabla u, u, x) \right) = f_u(\nabla u, u, x) \quad \forall x \in \Omega, \tag{1.1.9}$$

which is a second order PDE. For example, the Euler-Lagrange equation of the Dirichlet integral (1.1.8) is $\Delta u = 0$. A popular strategy for finding u is to solve (1.1.9) rather than using the minimization formulation. However, it must be stressed that the Euler-Lagrange equation provides only a necessary condition for all minimizers of (1.1.5) but not a sufficient condition. The importance of this will be discussed later in the introduction. Also, while a calculus of variations problem always gives a PDE problem, the converse may not be true, that is, not every PDE has a minimization counterpart. One example is the 1-D advection equation:

$$u_t + u_x = S.$$

1.2 Scope and Objective of Dissertation

The class of PDEs to be considered in this dissertation is the following linear, elliptic, non-divergence form PDE:

$$\mathcal{L}u := -A(x) : D^2 u = f \text{ in } \Omega, \tag{1.2.1a}$$

$$u = 0 \text{ on } \partial\Omega. \tag{1.2.1b}$$

Here $\Omega \subset \mathbb{R}^n$ is an open, bounded domain with boundary $\partial\Omega$, $f \in L^p(\Omega)$ with $1 < p < \infty$, and $A \in [C^0(\overline{\Omega})]^{n\times n}$ is positive definite in $\overline{\Omega}$. Here the matrix inner product $A : B$ is defined by

$$A : B = \sum_{i,j=1}^{d} a_{i,j} b_{i,j} = \operatorname{tr}(AB),$$

where the last equality holds if either A or B is symmetric. Non-divergence form elliptic PDEs appear inside of a class of second order fully nonlinear PDEs, known as Hamilton-Jacobi-Bellman (HJB) equations:

$$F(u) := \inf_{\alpha \in \Lambda} \left(-A^\alpha : D^2 u + b^\alpha \cdot \nabla u + c^\alpha u - f^\alpha \right) = 0, \tag{1.2.2}$$

where Λ is a parameter set and $\{A^\alpha\}, \{b^\alpha\}, \{c^\alpha\}, \{f^\alpha\}$ are families of functions indexed by $\alpha \in \Lambda$. The HJB equations arise from many applications such as stochastic optimal control and game theory [30]. Non-divergence form PDEs are also encountered in the linearization of fully nonlinear PDEs such as Monge-Ampère-type equations [11]:

$$F(u) := \det(D^2 u) = f, \tag{1.2.3}$$

for $f \geq 0$, in one of two ways. First, the linearization of (1.2.3) about a point u is

$$\operatorname{tr}(\operatorname{cof}(D^2 u) D^2 \varphi) = \operatorname{cof}(D^2 u) : D^2 \varphi, \tag{1.2.4}$$

where $\operatorname{cof}(D^2 u)$ is the cofactor of $D^2 u$ (see [11]). Thus we recover a non-divergence form operator for the linearization of the Monge-Ampère equation. Second, from [23] we can write the Monge-Ampère equation as a special case of the HJB equations (1.2.2):

$$\sup_{B \in \mathcal{S}_1} \left(-B : D^2 u + d \sqrt[d]{f \det B} \right) = 0, \tag{1.2.5}$$

where

$$\mathcal{S}_1 = \{ B \in \mathbb{R}^{d \times d}, B \text{ is symmetric positive semi-definite}, \operatorname{tr}(B) = 1 \}.$$

Non-divergence form PDEs in some sense are the best linear approximations of these fully non-linear PDEs. The techniques used to study and solve these PDEs will be helpful for solving their non-linear counterparts.

We also consider a special class of variational problems which exhibit the *Lavrentiev Gap Phenomonon* (LGP). This phenomenon arises when the minimizer u has certain types of singularities, and makes it impossible to approximate u by Lipschitz functions using only the energy $\mathcal{J}$.

While the existence and uniqueness of solutions to PDEs and variational problems are well understood, very few of these results give a constructive glimpse of the form of the solution u. Indeed, closed form solutions do not exist for a majority of PDEs and variational problems, even very simple and "nice" ones. The situation makes seeking approximate numerical solutions the only practical approach to solve these PDEs and variational problems. This in turn calls for developing accurate and efficient numerical methods for these problems on computers.

The goal of this dissertation is to construct, implement, and analyze accurate and efficient numerical methods for solving the non-divergence form PDEs and calculus of variations problems, especially those exhibiting the Lavrentiev Gap Phenomenon, using both the continuous and discontinuous Galerkin finite element framework.

1.3 Facts about PDEs to be Studied in this Dissertation

In this section, we collect some basic facts about the PDEs and calculus of variation problems which will be studied in this dissertation. These include the existence and uniqueness as well as the regularity results under different conditions on the data.

1.3.1 Linear Elliptic Non-divergence Form PDEs

The general theory for linear, elliptic, non-divergence PDEs is rich, culminating in three separate solution theories each depending on the specific regularity of A, f, and $\partial\Omega$. Let A be a uniformly positive definite matrix on Ω, that is, there exist constants

$\lambda, \Lambda > 0$ such that

$$\lambda \|\xi\|^2 \leq A(x)\xi \cdot \xi \leq \Lambda \|\xi\|^2 \quad \forall x \in \overline{\Omega}, \forall \xi \in \mathbb{R}^d. \tag{1.3.1}$$

The first theory is the classical solution (or Schauder's) theory (see [35, Chapter 6]). For $0 < \alpha < 1$, let $A \in [C^\alpha(\overline{\Omega})]^{n \times n}$, $f \in C^\alpha(\overline{\Omega})$, and $\partial\Omega \in C^{2,a}$, where $C^{k,\alpha}(\overline{\Omega})$ denotes the space of classically differentiable functions u of order k and $D^k u$ is Hölder continuous with modulus of continuity α. Under these conditions, there exists a unique solution $u \in C^{2,\alpha}(\overline{\Omega})$ to (1.2.1).

The second theory is the $W^{2,p}$ strong solution theory (see [35, Chapter 9]) which seeks solutions in $W^{2,p}(\Omega) \cap W^{1,p}_0(\Omega)$ that satisfy the PDE almost everywhere in Ω. For $1 < p < \infty$, let $A \in [C^0(\overline{\Omega})]^{n \times n}$, $f \in L^p(\Omega)$, and $\partial\Omega \in C^{1,1}$. Under these conditions (1.2.1) has a unique strong solution u in the Sobolev space $W^{2,p}(\Omega) \cap W^{1,p}_0(\Omega)$. There are also cases where we can relax the regularity of the boundary and still maintain well-posedness. If $d, p = 2$ and Ω is convex, then the regularity of $\partial\Omega$ may be dropped (c.f. [3]). We can also relax to coefficient matrix to $A \in [L^\infty(\Omega)]^{n \times n}$ under certain assumptions. First we still maintain a unique strong solution if we assume $\partial\Omega \in C^{1,1}$ and $A \in [\mathrm{VMO}(\Omega)]^{d \times d}$, that is $A \in [\mathrm{BMO}(\Omega)]^{d \times d}$ with

$$\lim_{r \to 0} \fint_{B_r \cap \Omega} |A - \bar{A}| \, \mathrm{d}x = 0,$$

where B_r is a ball of radius r and $A = \fint_{B_r \cap \Omega} A \, \mathrm{d}x$ (see [15]), or if Ω is convex and A satisfies the Cordès condition, that is, there exists $\varepsilon \in (0, 1)$ such that

$$\frac{\sum_{i,j=1}^{d} (a_{ij})^2}{\left(\sum_{i=1}^{d} a_{i,i}\right)^2} \leq \frac{1}{d - 1 + \varepsilon}, \tag{1.3.2}$$

where $a_{i,j}$ denotes the components of A (see [55, 44]).

Both of these theories were established using the freezing coefficient technique which we now describe. Since A is continuous, then, in a small enough ball, A is essentially constant. For a constant coefficient matrix A_0, we have $-A_0 : D^2u = -\operatorname{div}(A_0\nabla u)$. Since the operator $-\operatorname{div}(A_0\nabla u)$ is a change of basis away from the Laplacian $-\Delta u$, we can apply estimates from the the Poisson problem

$$\begin{aligned} -\Delta u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned}$$

to the operator $-A_0 : D^2u$ in the small ball. We then use a partition of unity and covering argument to derive a global Gärding-type stability estimate for our non-divergence form operator:

$$\|D^2u\| \leq C(\|\mathcal{L}u\| + \|u\|), \tag{1.3.4}$$

where $\|\cdot\|$ stands for the Hölder norm in the Schauder theory and for the L^2 norm in the strong solution theory. From here, each theory uses a different technique to arrive at the existence and uniqueness of the solution. The freezing coefficient technique will be used heavily in Chapter 3 and on the discrete level in Chapter 2.

The final theory is the viscosity solution theory which seeks solutions in $C^0(\overline{\Omega})$ that satisfy the PDE in the viscosity sense found in [36]. If we assume $A \in L^\infty(\Omega)$ and $f \in C^0(\overline{\Omega})$, then there exists a viscosity solution $u \in C^0(\overline{\Omega})$, moreover, we have the interior estimate $u \in C^\alpha(\Omega)$ (c.f. [36]).

1.3.2 The Calculus of Variations

Unlike partial differential equations, whose existence and uniqueness heavily rely on the structure of the PDE operator, the existence and uniqueness of solutions to problems from the calculus of variations is quite general.

To set up the theory and result from [19], we consider $\mathcal{J}(v)$ from (1.1.4) and the admissible set

$$V = W_g^{1,p}(\Omega) := \{v + g : v \in W_0^{1,p}(\Omega)\},$$

where $g \in W^{1,p}(\Omega)$ for some $1 < p < \infty$ and Ω is an open, bounded domain. We also assume $\mathcal{J}$ is proper, meaning there exists $v \in V$ such that $\mathcal{J}(v) < \infty$.

The existence of a minimizer to $\mathcal{J}$ over V comes from sufficient conditions placed on the density function $f(\xi, v, x)$ in (1.1.4). We assume the following on f:

(H1) f is a Carathédory function, that is,

$$\begin{aligned} x \to f(\xi, v, x) &\quad \text{is measurable for every } (\xi, v) \in \mathbb{R}^n \times \mathbb{R}, \\ (\xi, v) \to f(\xi, v, x) &\quad \text{is continuous for almost every } x \in \Omega; \end{aligned}$$

(H2) the function $\xi \to f(\xi, v, x)$ is convex for every $(v, x) \in \mathbb{R} \times \overline{\Omega}$;

(H3) there exists $q \in [1, p)$ and constants $\alpha_1 > 0$, $\alpha_2, \alpha_3 \in \mathbb{R}$ such that

$$f(\xi, v, x) \geq \alpha_1 |\xi|^p + \alpha_2 |u|^q + \alpha_3.$$

Under these assumptions, there exists $u \in W_g^{1,p}(\Omega)$ such that

$$u \in \underset{v \in V}{\arg\min}\, \mathcal{J}(v).$$

To show how exactly each assumption is used we give rough sketch of the proof. Let

$$m = \inf_{v \in V} \mathcal{J}(v),$$

which is finite since $\mathcal{J}$ is proper. We can extract a minimizing sequence $\{u_j\} \subset V$ such that $\mathcal{J}(u_j) \searrow m$. By assumption (H3), $\mathcal{J}$ is coercive on V so that $\{u_j\}$ is a bounded set in V. Since $V = W_g^{1,p} \subset W^{1,p}(\Omega)$ and $W^{1,p}(\Omega)$ is a reflexive Banach space for $1 < p < \infty$, we can use its compactness to extract a convergent subsequence $\{u_j\}$ (not relabeled) and $\hat{u} \in W^{1,p}(\Omega)$ such that $u_j \rightharpoonup \hat{u}$ weakly in $W^{1,p}(\Omega)$. Since V is affine, $\hat{u} \in V$. Since f is convex in ξ by (H2) and satisfies (H1), $\mathcal{J}$ is weakly lower semi-continuous in $W^{1,p}(\Omega)$, that is,

$$\mathcal{J}(\hat{u}) \le \liminf_{j\to\infty} \mathcal{J}(u_j).$$

Thus we have

$$m \le \mathcal{J}(\hat{u}) \le \liminf_{j\to\infty} \mathcal{J}(u_j) = m.$$

which implies $\mathcal{J}(\hat{u}) = m$ and $\hat{u}$ is the minimizer of $\mathcal{J}$ over V. The conditions on f are refined such that weakening (H1)-(H3) will lead to a counterexample which violates the existence of a minimizer u (see [19]).

While this framework covers existence, an additional assumption is required for uniqueness. For example, if $(\xi, v) \to f(\xi, v, x)$ is strictly convex for every $x \in \overline{\Omega}$, then the minimizer u is unique.

1.4 A Literature Survey of Previous Numerical Methods

We give a brief review of existing numerical methods to the problems of interest in this dissertation, namely second order, linear, elliptic, non-divergence form PDEs and problems from the calculus of variations - especially those exhibiting the Lavrentiev gap phenomenon.

1.4.1 Linear Elliptic Non-divergence Form PDEs

In contrast to the wealth of results for the PDE analysis, very little progress has been made in the field of numerical methods for second order, non-divergence form, elliptic PDEs with a non-differentiable coefficient matrix A (1.2.1a-1.2.1b). The difficulty is two-fold. First, the non-divergence structure of the PDE prevents the use of integration by parts on it to define weak solutions, which is a pre-requisite for formulating finite element methods for a given PDE problem. The second lies in the lack of differentiability of the coefficient matrix A. If A is differentiable, then the non-divergence operator $-A : D^2u$ may be written as the sum of a diffusion operator and a lower order advection operator, that is,

$$-A : D^2u = -\nabla \cdot (A\nabla u) + (\nabla \cdot A) \cdot \nabla u. \tag{1.4.1}$$

The diffusion operator fits well with the Galerkin framework, but without differentiability of A, it is not possible to rewrite the non-divergence operator in such a way.

However, while progress has been slow, a few numerical methods for these non-divergence form PDEs with continuous A or weaker have been reported very recently in the literature.

The first work, by Smears and Süli in 2013 (see [55]) provides an hp discontinuous Galerkin method for $A \in [L^\infty(\Omega)]^{n\times n}$ satisfying the Cordès condition (1.3.2) and $f \in L^2(\Omega)$ that approximates the strong solution $u \in H^2(\Omega) \cap H_0^1(\Omega)$. This method is constructed by adding an artificial discrete Laplacian to the bilinear form. The stability of their method relies on the fact that for a weight $\gamma \in L^\infty(\Omega)$ dependent on the Cordès condition, the quantity $\|\gamma\mathcal{L}v - \Delta v\|$ is controllable. With this, they were able to achieve a convergent method which is optimal in h and sub-optimal in p by a half order.

The second method, by Nochetto and Wang in 2014 (see [47]), develops a continuous finite element method for continuous A and f. The construction relies on the identity

$$-A : D^2u = -\frac{\lambda}{2}\Delta u - \left(A - \frac{\lambda}{2}I\right) : D^2u, \tag{1.4.2}$$

where λ is from the ellipticity condition (1.3.1). The first term of (1.4.2) is treated as expected while the second term is converted into a non-local integral operator. The method then uses linear finite elements and a weakly acute mesh to recover a Discrete

Maximum Principle. The method is also proved to converge to the viscosity solution $u \in C(\overline{\Omega})$.

The next method, by Wang and Wang in 2014 (see [56]), uses a weak Galerkin approach, that is, to decompose a DG function v_h into two functions (w_h, z_h) where w_h lives in the interior of each element and z_h lives on the skeleton of the mesh. The paper creates a discrete weak version of the Hessian using these decoupled functions and proposes a primal-dual method to define the solution u_h. The method only requires A to be piecewise continuous and $f \in L^2(\Omega)$ in order for the existence and unique of u_h. In addition the converges to the strong solution $u \in H^2(\Omega) \cap H^1_0(\Omega)$ provided such a strong solutions exists.

The final method, by Feng, Hennings, and Neilan in 2015 (see [22]), discretizes the PDE using a nonstandard continuous finite element method with quadratic or higher order elements. The weak form is built first by using (1.4.1) and rewriting the non-divergence PDE as the following diffusion-advection equation:

$$-\nabla \cdot (A \nabla u) + (\nabla \cdot A) \cdot \nabla u = f. \tag{1.4.3}$$

From here they use standard finite element techniques to create a weak form for (1.4.3), then they integrate by parts, on each element, to recover back the non-divergence operator. The stability of the method is proved by use of the freezing coefficient technique (see Subsection 1.3.1). This method converges optimally in the discrete $W^{2,p}$ norm provided that $A \in [C^0(\overline{\Omega})]^{n \times n}$ and $f \in L^p(\Omega)$ for some $1 < p < \infty$, and is the natural extension of the finite element method to non-divergence form PDEs as it recovers the standard finite element method when A is a constant matrix.

1.4.2 The Calculus of Variations

There are two common approaches to numerically approximating the minimizers satisfying (1.1.5): indirect and direct methods. Indirect methods use the Euler-Lagrange equation (1.1.9) to convert the minimization problem into a PDE problem, which then can be discretized using a variety of methods such as Finite Difference, Finite Element, or Discontinuous Galerkin. This is often the preferred approach because of the vast wealth of material available for numerical approximations of PDEs, but it does have one drawback: the Euler-Lagrange equation is only a necessary condition for a minimum and not a sufficient one. More information must be known of $\mathcal{J}$ in order to determine if the solution of the Euler-Lagrange equation does indeed globally minimize $\mathcal{J}$. In addition, such a discretization may lose some important properties of the original energy, such as conservation or dissipation.

Another, less common, approach is the direct approach, which seeks to directly approximate $\mathcal{J}$ by a discrete functional $\mathcal{J}_h$. We then seek u_h such that

$$u_h \in \operatorname*{arg\,min}_{v_h \in X_h} \mathcal{J}_h(v_h), \tag{1.4.4}$$

where X_h is a discrete approximation space. Since problem (1.4.4) is now an algebraic problem, a variety of methods may be employed to recover u_h. For example, we may minimize $\mathcal{J}_h$ by using a quasi-Newton minimization solver or by applying the discrete Euler-Lagrange equation to $\mathcal{J}_h$ and then solve for u_h. The key to this approach is how to construct a "good" discrete energy $\mathcal{J}_h$ since we are not dealing directly with a PDE. While the literature on this approach is not very extensive, we list a few examples of numerical methods based on this direct approach.

First, we have the discrete variational derivative method by Furihata and Matsuo (see [33]). This method uses a finite difference method to discretize $\mathcal{J}$ for energies arising from the KdV equations, the nonlinear Schrödinger equations, and the Cahn-Hilliard equations. The key to the method is to construct a discrete energy to ensure important properties of the continuous energy such as conservation or dissipation in time are preserved. The method is comprehensive in that it defines methods for higher order temporal and spacial schemes as well as robust discrete solvers.

Second, we have the (see [10]) which provides an interior penalty discontinuous Galerkin finite element discretization of $\mathcal{J}$ based on f satisfying conditions similar to (H1)-(H3) and $V = W_g^{1,p}(\Omega)$. The key property of their discretization is to use a lifting operator to approximate the distributional gradient of a piecewise polynomial function rather than just the piecewise gradient. Standard penalty parameters are added to weakly enforce continuity and Dirichlet boundary data.

The convergence of the Variational DGFEM and many direct methods are proven using a special convergence theory: Γ-Convergence. In order to prove convergence of the method, it is necessary to show that u_h from (1.4.4) converges to u from (1.1.6), that is, to show that the minimizers of $\mathcal{J}_h$ to converge to the minimizer of $\mathcal{J}$. Pointwise convergence of $\mathcal{J}_h$ to $\mathcal{J}$ is not enough to ensure the convergence of minimizers to minimizers and uniform convergence is too strong for practical applications. The convergence that preserves the convergence of minimizers is Γ-Convergence. We recall the definition of Γ-Convergence from [6]:

Definition 1.1. *Let X be a topological vector space and let $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$. Let $F : X \to \overline{\mathbb{R}}$ and $\{F_n\}$ be a sequence of functions from X to $\overline{\mathbb{R}}$. We say F_n Gamma-converges to F, written $F_n \xrightarrow{\Gamma} F$, provided the following two conditions hold for every $x \in X$.*

1. *For every sequence $\{x_n\}$ such that $x_n \to x$ in X as $n \to \infty$ we have*

$$F(x) \leq \liminf_{n\to\infty} F_n(x_n).$$

2. *There exists a sequence $\{x_n\}$ such that $x_n \to x$ in X as $n \to \infty$ and*

$$F(x) \geq \limsup_{n\to\infty} F_n(x_n).$$

The first criterion of Definition (1.1) satisfies a general lower semi-continuity condition needed for the existence of minimizers. The second criterion, however, requires the existence of a *recovery sequence*, that is, a sequence $\{x_n\}_{n\in\mathbb{N}}$ such that $F_n(x_n) \to F(x)$ if $x_n \to x$. Here we see that $F_n(x_n)$ “recovers” $F(x)$.

An important result of Γ-Convergence is that if $\mathcal{J}_h \xrightarrow{\Gamma} \mathcal{J}$ and if u_h from (1.4.4) converges to $\hat{u}$, then $\hat{u}$ minimizes $\mathcal{J}$, which is exactly the convergence result we want. To show this, let $x_n \in X$ such that

$$x_n \in \arg\min_{y \in X} F_n(y),$$

and suppose $x_n \to x$ in X for some $x \in X$. Let $x' \in X$ minimize F over X. By Criterion 2 of (1.1), there exists a sequence $\{y_n\}_{n\in\mathbb{N}}$ of X such that

$$F(x') \geq \limsup_{n\to\infty} F_n(y_n). \tag{1.4.5}$$

Since Criterion 1 of (1.1) holds for any sequence converging to x, we have

$$F(x') \leq F(x) \leq \liminf_{n\to\infty} F_n(x_n) \leq \liminf_{n\to\infty} F_n(y_n) \leq \limsup_{n\to\infty} F_n(y_n) \leq F(x').$$

Since $F(x) = F(x')$, then x must be a minimizer of F over X.

It is important to note that Γ-Convergence does not imply that the discrete minimizers x_n will converge, only that, if they converge, they will converge to the minimizer of F. The initial convergence must be shown separately, usually by a compactness argument from the coerciveness of F.

The Lavrentiev Gap Phenomenon

A specific class of functionals that exhibit the Lavrentiev Gap Phenomenon (LGP) is analyzed in this dissertation. To define the phenomenon, let $\mathcal{A} = W_g^{1,1}(\Omega)$ and let $\mathcal{A}_\infty := \mathcal{A} \cap W^{1,\infty}(\Omega)$. Since Ω is bounded, then $\mathcal{A} \subset \mathcal{A}_\infty$ and consequently there holds

$$\inf_{v\in\mathcal{A}_1} \mathcal{J}(v) \leq \inf_{v\in\mathcal{A}_\infty} \mathcal{J}(v). \tag{1.4.6}$$

$\mathcal{J}$ is said to exhibit the Lavrentiev gap phenomenon whenever

$$\inf_{v\in\mathcal{A}} \mathcal{J}(v) < \inf_{v\in\mathcal{A}_\infty} \mathcal{J}(v), \tag{1.4.7}$$

in other words, when the strict inequality holds in (1.4.6).

The LGP presents itself in a variety of problems from applications including materials sciences, nonlinear elasticity, and image processing (see [32, 57, 12]).

The gap between the minimum values on both sides of (1.4.7) suggests that the the minimizer of the left-hand side must have some singularity which causes the gap. It has been known in the literature [32, 57, 12] that the gap phenomenon could happen not only for non-convex energy functionals but also for strictly convex and coercive energy functionals. As a result, it is a very complicated phenomenon to characterize, analyze, approximate, because the gap phenomenon can be triggered by quite different mechanisms, and the definition of the LGP is a very broad concept which covers many different types of singularities. In addition, there are no known general sufficient conditions which guarantee the existence of the gap phenomenon.

The simplest and best known example of the gap phenomenon is Maniá's 1-D problem [42], where one minimizes the functional

$$\mathcal{J}(v) = \int_0^1 v'(x)^6 \big(v(x)^3 - x\big)^2 \,\mathrm{d}x \tag{1.4.8}$$

over all functions $v \in W^{1,1}(0,1)$ satisfying $v(0) = 0$ and $v(1) = 1$. By inspection it is easy to see that $u(x) = x^{\frac{1}{3}}$ minimizes (1.4.8) with a minimum value zero. However, it can be shown that the minimum over space $W^{1,\infty}(0,1)$, that is, the space of all Lipschitz functions, is strictly larger than zero. As a result, Maniá's problem does exhibit the LGP. Notice that $u'(x) = \frac{1}{3}x^{-\frac{2}{3}}$ which blows up rapidly as $x \to 0^+$. Moreover, a more striking property, which was stated by Ball and Knowles (cf. [5]), is that if u_j is a sequence of functions in $W^{1,q}(0,1)$ for $q \geq \frac{3}{2}$ with $u_j(0) = 0$ and $u_j(1) = 1$ such that $u_j \to u$ a.e. as $j \to \infty$, then $\mathcal{J}(u_j) \to \infty$ as $j \to \infty$. Since conforming finite element spaces are a subspace of $W^{1,\infty}$, the above properties of the functional $\mathcal{J}$ imply that the standard finite element approximations to Maniá's problem must fail to approximate both the minimizer and the minimum value of the functional.

To achieve convergent numerical discretizations for energies exhibiting the LGP, we must use non-standard discretizations of $\mathcal{J}$, and several numerical techniques have been shown to curb the phenomenon. We mention that one may resolve the LGP by not necessarily changing $\mathcal{J}$, but rather minimizing over a different space. Indeed, Ortner used the non-conforming Crouzeix-Raviart element instead of a conforming finite element and achieves convergence on a specific class of energies exhibiting the LGP (see [48]). However, a focus of this dissertation is on conforming discretizations, where we only change the discrete functional $\mathcal{J}_h$ and not the discrete space S_h.

Over the past thirty years, there have been several conforming discrete discretizations. Below we only briefly discuss these methods; a deeper explanation of these methods and why they overcome the Lavrentiev gap phenomenon will be given in Chapter 4.

In 1987, Ball and Knowles (see [5]) introduced a penalty type method. In this method they decouple the finite element function v_h and its derivative w_h, and then minimize $\mathcal{J}$ over both functions while weakly enforcing $w_h' = w_h$. They prove convergence of the method for a variety of 1-D problems, and the penalty method has been extended to higher-dimension problems [45, 12].

Another technique, the truncation/removal method, was developed by Li and Bai (see [41, 4]). This truncation method modifies $\mathcal{J}(v_h)$ on the elements where $\mathcal{J}(v_h)$ is larger than a constant times the Sobolev norm of v_h to tame the LGP. This leads to a robust scheme that converges for a wide variety of gap phenomenon problems.

1.5 Summary of the Dissertation Contributions

This dissertation is the accumulation of several research projects which can be divided into three parts.

In part one, we study numerical and PDE approximations to non-divergence form second order linear elliptic PDEs. We develop several interior-penalty discontinuous Galerkin (IP-DG) methods for second order, linear, elliptic, non-divergence form PDEs following the technique of [22]. We show the stability of these methods using a freezing coefficient argument on the discrete level with a non-standard duality argument involving the discrete adjoint. Included as well is a $W^{1,p}$ stability result for IP-DG methods for the constant coefficient case - a result that has independent interest and is used in the stability argument. We show optimal error estimates in broken $W^{2,p}$ norm and give several numerical tests for cases inside and outside of the theory. We also develop a vanishing moment method for second order linear elliptic non-divergence form PDEs. This PDE technique approximates the second order PDE by a sequence of fourth order PDEs by the addition of a vanishing biharmonic term. Uniform H^1 and H^2 stability estimates are obtained which to the convergence of the method. In addition, we derive L^2 and H^1 error estimates for the vanishing moment approximations. We present a C^0 DG finite element method for the fourth order method and give numerical results supporting the convergence of the method. We also give numerical test results of a method combining the IP-DG schemes and the vanishing moment method and apply it to the several examples of the Hamilton-Jacobi-Bellman equations.

In part two, we introduce an enhanced finite element method for variational problems exhibiting the Lavrentiev Gap Phenomenon. We show the advantages of the method, include heuristics on how to tune the method to achieve convergence, and prove the Γ-Convergence of the method on the continuous space $W^{1,\infty}(\Omega)$. In addition, we give a few numerical results showing the convergence of the method for a selection of 1-D and 2-D problems, some of which include the phenomenon while others do not.

Finally, in part three, we focus on discontinuous Ritz methods for a class of variational problems that are coercive and convex. We use the discontinuous Galerkin, finite element, (DG-FE) numerical calculus developed in [25] to construct a discontinuous Ritz framework for variational problems. Noting the similarities of this method and Variational DGFEM method of Buffa and Ortner (see [10]), we obtain the convergence of the method as well as a compactness result. We also develop a MATLAB toolbox to implement the DG-FE calculus and discontinuous Ritz methods which has several numerical examples shown and a complete documentation manual included.

1.6 Notation

Standard function and space notation will be used in this dissertation, and to improve its readability we write $a \lesssim b$ and $a \gtrsim b$ for $a \leq Cb$ and $a \geq Cb$ respectively for some constant $C > 0$ which does not depend on any discretization or approximation parameters.

Let Ω be an open and bounded domain in $\mathbb{R}^d$. For a subdomain D of Ω with boundary ∂D, let $L^p(D)$ and $W^{s,p}(D)$ for $s \geq 0$ and $1 \leq p \leq \infty$ denote the standard Lebesgue and Sobolev spaces respectively with norms

$$\|v\|_{L^p(D)} = \begin{cases} \left(\int_D |v(x)|^p \, \mathrm{d}x \right)^{1/p} & \text{if } p < \infty, \\ \operatorname*{ess\,sup}_{x \in D} |v(x)| & \text{if } p = \infty, \end{cases}$$

and

$$\|v\|_{W^{s,p}(D)} = \sum_{|\alpha| \leq s} \|D^\alpha v\|_{L^p(D)},$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ is a multi-index and $|\alpha| = \alpha_1 + \cdots + \alpha_n$. With convention, we denote $H^k(\Omega) = W^{k,2}(\Omega)$. Let $W_0^{1,p}(D)$ be the closure of $C_c^\infty(D)$ in $W^{1,p}(D)$. Let

$$(f, g)_D = \int_D f(x)g(x)\,\mathrm{d}x$$

denote the L^2 inner product on D and $(\cdot,\cdot) := (\cdot,\cdot)_\Omega$. We also define the H^{-1} norm as follows:

$$\|v\|_{H^{-1}(\Omega)} = \sup_{w\in H_0^1(\Omega)} \frac{|(v,w)_\Omega|}{\|\nabla w\|_{L^2(\Omega)}}.$$

Let $\mathcal{T}_h$ be a shape-regular, conforming, and quasi-uniform triangulation of Ω with $h \approx \operatorname{diam}(T)$ for all $T \in \mathcal{T}_h$. Let $\mathcal{E}_h^I$ and $\mathcal{E}_h^B$ denote respectively the sets of all interior and boundary edges/faces of $\mathcal{T}_h$, and set $\mathcal{E}_h := \mathcal{E}_h^I \cup \mathcal{E}_h^B$. We introduce the broken Sobolev spaces

$$\begin{aligned} W^{s,p}(\mathcal{T}_h) &:= \prod_{T\in\mathcal{T}_h} W^{s,p}(T), \qquad & L^p(\mathcal{T}_h) &:= W^{0,p}(\mathcal{T}_h), \\ W_h^{s,p}(D) &:= W^{s,p}(\mathcal{T}_h)\big|_D, \qquad & L_h^p(D) &:= L^p(\mathcal{T}_h)\big|_D. \end{aligned}$$

For any interior edge/face $e = \partial T^+ \cap \partial T^- \in \mathcal{E}_h^I$, we define the jump and average of a scalar or vector valued function v as

$$[v]\big|_e := v^+ - v^-, \qquad \{v\}\big|_e := \frac{1}{2}\left(v^+ + v^-\right),$$

where $v^\pm = v|_{T^\pm}$. On a boundary edge/face $e \in \mathcal{E}_h^B$ with $e = \partial T^+ \cap \partial\Omega$, we set $[v]\big|_e = \{v\}\big|_e = v^+$. For any $e \in \mathcal{E}_h^I$ we use ν_e to denote the unit outward normal vector pointing in the direction of the element with the smaller global index. For $e \in \mathcal{E}_h^B$ we set ν_e to be the outward normal to $\partial\Omega$ restricted to e. The standard Continuous Galerkin (CG) and Discontinuous Galerkin (DG) finite element spaces are defined as

$$
\begin{aligned}
S_h = S_h^k &:= \Big\{ v_h \in W^{2,p}(\mathcal{T}_h) \cap W_0^{1,p}(\Omega);\; v_h\big|_T \in \mathbb{P}_k(T) \quad \forall T \in \mathcal{T}_h \Big\},\\
V_h = V_h^k &:= \Big\{ v_h \in W^{2,p}(\mathcal{T}_h);\; v_h\big|_T \in \mathbb{P}_k(T) \quad \forall T \in \mathcal{T}_h \Big\},
\end{aligned}
$$

where $\mathbb{P}_k(T)$ denotes the set of all polynomials of degree less than or equal to k on T. We also introduce for any $D \subset \Omega$

$$
\begin{aligned}
V_h(D) &:= \Big\{ v \in V_h;\; v\big|_{\Omega\setminus\overline{D}} \equiv 0 \Big\},\\
S_h(D) &:= \Big\{ v \in S_h;\; v\big|_{\Omega\setminus\overline{D}} \equiv 0 \Big\}.
\end{aligned}
$$

Note that $V_h(D)$ and $S_h(D)$ are nontrivial provided that there exists an inscribed ball B with radius $r \geq 2h$ such that $B \subset D$. Also note that $S_h(D)$ is not a subspace of $S_h(\Omega)$. In addition, we define the vector valued discrete space $[V_h]^d$ as

$$
[V_h]^d = \{\varphi_h = (\varphi_{h,1}, \varphi_{h,2}, \ldots, \varphi_{h,d}) : \varphi_{h,i} \in V_h \quad \forall i = 1, 2, \ldots, d\}.
$$

For each $e \in \mathcal{E}_h$, let $\gamma_e > 0$ be constant on e. We define the following mesh-dependent norms on $W_h^{1,p}(D)$ and $W_h^{2,p}(D)$:

$$
\begin{aligned}
\|v\|_{W_h^{2,p}(D)} := \|D_h^2 v\|_{L^p(D)} &+ \Big(\sum_{e \in \mathcal{E}_h^I} h_e^{1-p} \| | [\nabla v] | \|_{L^p(e\cap\bar{D})}^p \Big)^{\frac{1}{p}} \qquad (1.6.1)\\
&+ \Big(\sum_{e \in \mathcal{E}_h} \gamma_e^p h_e^{1-2p} \|[v]\|_{L^p(e\cap\bar{D})}^p \Big)^{\frac{1}{p}},\\
\|v\|_{W_h^{1,p}(D)} := \|\nabla_h v\|_{L^p(D)} &+ \Big(\sum_{e \in \mathcal{E}_h} \gamma_e^p h_e^{1-p} \|[v]\|_{L^p(e\cap\bar{D})}^p \Big)^{\frac{1}{p}} \qquad (1.6.2)\\
&+ \Big(\sum_{e \in \mathcal{E}_h} h_e \|\{\nabla v\}\|_{L^p(e\cap\bar{D})}^p \Big)^{\frac{1}{p}},
\end{aligned}
$$

(1.6.3)

where $\nabla_h v$ and $D_h^2 v$ denote the piecewise gradient and Hessian of v. In addition, we define the discrete $W_h^{-2,p}$-norm and $W_h^{-1,p}$-norm as follows:

$$\|q\|_{W_h^{-2,p}(D)} := \sup_{0\neq v_h\in V_h(D)} \frac{(q, v_h)_D}{\|v_h\|_{W_h^{2,p'}(D)}}, \tag{1.6.4}$$

$$\|q\|_{W_h^{-1,p}(D)} := \sup_{0\neq v\in W_h^{1,p'}(D)} \frac{(q, v)_D}{\|v\|_{W_h^{1,p'}(D)}}, \tag{1.6.5}$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. Finally, for any domain $D \subseteq \Omega$ and any $w \in L_h^p(D)$, we introduce the following mesh-dependent semi-norm

$$\|w\|_{L_h^p(D)} := \sup_{0\neq v_h\in V_h(D)} \frac{\big(w, v_h\big)_D}{\|v_h\|_{L^{p'}(D)}}. \tag{1.6.6}$$

It can be proved that (cf. [22])

$$\|w_h\|_{L^p(\Omega)} \lesssim \|w_h\|_{L_h^p(\Omega)} \qquad \forall w_h \in V_h. \tag{1.6.7}$$

1.7 Mathematical Software and Implementation

A majority of the numerical results in this dissertation, namely those given in Chapters 2, 4, 5, 6, and the Hamilton Jacobi Bellman results in 3, are obtained with the programming language MATLAB (see [Mathworks]). Results for the C^0 interior penalty finite element method in Chapter 3 are obtained using the FEniCS Project software collection (see [1]).

Chapter 2

Interior Penalty Discontinuous Galerkin Methods for Second Order Linear Non-Divergence Form Elliptic Partial Differential Equations

2.1 Introduction

In this chapter, we develop interior penalty discontinuous (IP-DG) Galerkin methods for approximating the $W^{2,p}$ strong solution to the following second order linear elliptic PDE in non-divergence form:

$$\mathcal{L}u := -A : D^2 u = f \quad \text{in } \Omega, \tag{2.1.1a}$$

$$u = 0 \quad \text{on } \partial\Omega. \tag{2.1.1b}$$

where A is merely continuous. Non-divergence form PDEs are related to the fully nonlinear Hamilton-Jacobi-Bellman equations (1.2.2) which have applications in stochastic optimal control and financial mathematics, and the Monge-Ampère equation (1.2.3), which has applications to differential geometry and optimal mass transport.

Convergent finite element methods for (2.1.1) are non-trivial in construction; the reason for this is two-fold. First, the non-divergence structure does not allow for integration by parts which is essential for the establishment of a weak formulation. Merely testing (2.1.1a) by an H^1 conforming finite element function does not produce a convergent discretization. However, if A is differentiable, then it is easy to check that equation (2.1.1a) can be rewritten as a diffusion-advection equation

$$-A : D^2u = -\operatorname{div}(A\nabla u) + \operatorname{div}(A) \cdot \nabla u \tag{2.1.2}$$

with A as the diffusion coefficient and $\operatorname{div}(A)$ as the advection coefficient. This rewritten equation, now with a second order divergence form operator, is well suited for classical finite element methods. However, if $A \in \left[C^0(\overline{\Omega})\right]^{d\times d}$, then this formulation is not possible since $(\nabla \cdot A)$ does not exist as a function, but rather only as a measure. This is the second challenge of developing convergent finite element methods.

Because of these challenges, the discretization of the non-divergence term $-A : D^2u$ is not trivial, and only a few numerical schemes have been developed that are convergent for continuous A (see [55, 47, 56, 22]). Each of these schemes discretize the non-divergence term in a different fashion, and, because of this, each scheme has advantages and disadvantages, such as extensions to discontinuous A and ease of computation. Subsection 1.4.1 gives a more thorough explanation of each method. Of these methods, we focus on the finite element method designed to approximate the $W^{2,p}$ strong solution of (2.1.1) in the case of continuous A, which was developed by Feng, Hennings, and Neilan in [22]. To formulate their finite element method, they first assume $A \in C^1(\overline{\Omega})$ allowing the equality (2.1.2) to hold. Because the second order term on left hand side of (2.1.2) is in divergence form, it is easy to formulate the standard bilinear form for such an equation, namely,

$$a^{FE}(w_h, v_h) = \int_\Omega \nabla w_h \cdot \nabla v_h \,\mathrm{d}x + \int_\Omega \mathrm{div}(A) \cdot \nabla w_h v_h \,\mathrm{d}x \qquad \forall w_h, v_h \in S_h. \tag{2.1.3}$$

Since a finite element function $v_h \in S_h$ may not be globally in $H^2(\Omega)$, they apply the following well-known DG integration by parts formula:

$$\int_\Omega \tau \cdot \nabla_h v \,\mathrm{d}x = -\int_\Omega (\nabla_h \cdot \tau) v \,\mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_e [\tau \cdot \nu_e]\{v\} \,\mathrm{d}S + \sum_{e \in \mathcal{E}_h} \int_e \{\tau \cdot \nu_e\}[v] \,\mathrm{d}S, \tag{2.1.4}$$

where v and τ are any scalar and vector valued functions, respectively, defined on each $T \in \mathcal{T}_h$, to the second term of (2.1.3) which gives them the following bilinear form:

$$a_h^{FE}(w_h, v_h) = -\sum_{T \in \mathcal{T}_h} \int_T (A : D^2 w_h) v_h \,\mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_e [A \nabla w_h \cdot \nu_e] v_h \,\mathrm{d}S \qquad \forall w_h, v_h \in S_h. \tag{2.1.5}$$

Since (2.1.5) does not contain $\mathrm{div}(A)$, their finite element method for continuous A is then defined as seeking $u_h \in S_h$ such that

$$a_h^{FE}(u_h, v_h) = (f, v_h) \qquad \forall v_h \in S_h. \tag{2.1.6}$$

With the freezing coefficient technique explained in Subsection 1.3.1 and a non-standard duality argument, they prove stability of the bilinear form, and prove convergence of the method with an optimal discrete $W^{2,p}$-norm error estimate. We note that this finite element method has several advantages over the other methods listed. First, the method is quite simple in construction and is implementable on many

H^1-conforming finite element software packages - such as FEniCS [1]. Secondly, we consider this discretization of the non-divergence term to be the most natural, since this method is equivalent to the standard finite element method when A is a constant coefficient matrix. One drawback of this method is that the convergence analysis is proven using the freezing coefficient technique, which is insufficient for showing convergence for $A \in [L^\infty(\Omega)]^{d\times d}$. However, the method shows convergent results when it is tested on an numerical example with discontinuous A.

The goal of this chapter is to extend the formulation of the finite element method in [22], whose approximate solutions belong to the H^1 conforming space S_h, to V_h - the space of discontinuous polynomials. The basis for this extension is threefold. First, since the jumps of the normal derivatives are used in (2.1.3), it is natural to extend this bilinear form to a completely discontinuous space. Our extension will be the IP-DG formulation of problem (2.1.1). Secondly, the IP-DG framework brings with it several computational advantages over traditional finite element methods, for example, simplicity and ease of computation, flexibility of mesh generation, and ease of adaptivity. Lastly, since non-divergence form PDEs are used as the building blocks for the Hamilton-Jacobi-Bellman equations (1.2.2) whose viscosity solutions have low regularity, approximations from the discontinuous Galerkin space V_h should be better at resolving these solutions.

This chapter is organized as follows. In Section 2.2, we establish some preliminary results related to discontinuous discrete functions. In Section 2.3 we present several IP-DG schemes related to the case of constant coefficient A, derive a $W_h^{1,p}$ stability result for the symmetric IP-DG scheme, then establish a $W_h^{2,p}$ stability result for all of the these methods. In Section 2.4 we formulate our IP-DG schemes for the case of continuous A, prove the $W_h^{2,p}$ stability of the methods using the stability of the constant coefficient case, and derive optimal order error estimates in the $W_h^{2,p}$ norm. Finally, in Section 2.5, we present several numerical experiments showing the validity of methods for test problems both inside and outside of the $W^{2,p}$ strong solution theory. This chapter is based on a joint research project which was reported in [28].

2.2 Properties of the Discrete Functions of V_h

In this section we collect some technical lemmas that cover the basic properties of functions V_h which is defined in Section 1.6. These facts will be used many times throughout the whole chapter.

We first state the standard trace inequalities for broken Sobolev functions, a proof of this lemma can be found in [8].

Lemma 2.1. *For any $T \in \mathcal{T}_h$ and $1 < p < \infty$, there holds*

$$\|v\|^p_{L^p(\partial T)} \lesssim \left(h_T^{p-1} \|\nabla v\|^p_{L^p(T)} + h_T^{-1} \|v\|^p_{L^p(T)} \right) \qquad \forall v \in W^{1,p}(T). \tag{2.2.1}$$

Therefore by scaling we have

$$\sum_{e \in \mathcal{E}_h^I} h_e \|v\|^p_{L^p(e \cap \overline{D})} \lesssim \begin{cases} \|v\|^p_{L^p(D)} & \forall v \in V_h(D), \\ \|v\|^p_{L^p(D)} + h^p \|\nabla v\|^p_{L^p(D)} & \forall v \in W_h^{2,p}(D). \end{cases} \tag{2.2.2}$$

Next, we prove an inverse inequality between the $W_h^{2,p}$-norm and the $W_h^{1,p}$-norm.

Lemma 2.2. *For any $v_h \in V_h$, $D \subseteq \Omega$, there holds for $1 < p < \infty$*

$$\|v_h\|_{W_h^{2,p}(D)} \lesssim h^{-1} \|v_h\|_{W_h^{1,p}(D_h)}, \tag{2.2.3}$$

where

$$D_h = \{x \in \Omega;\, \mathrm{dist}(x, D) \leq h\}. \tag{2.2.4}$$

Proof. To show (2.2.3), we use (1.6.1), (2.2.1), and standard inverse estimates [8] to obtain

$$\|v_h\|_{W_h^{2,p}(D)} \lesssim \|D_h^2 v_h\|_{L^p(D)} + \Big(\sum_{e \in \mathcal{E}_h} \gamma_e^p h_e^{1-2p} \|[v]\|^p_{L^p(e \cap \bar{D})} \Big)^{\frac{1}{p}}$$

$$
\begin{aligned}
&\quad + \sum_{\substack{T\in\mathcal{T}_h\\ T\subset D_h}} \Big(h_T^{1-p}\Big(h_T^{p-1}\|D^2 v_h\|_{L^p(T)}^p + h_T^{-1}\|\nabla v_h\|_{L^p(T)}^p\Big)\Big)^{\frac{1}{p}}\\
&\lesssim \|D_h^2 v_h\|_{L^p(D)} + \Big(\sum_{e\in\mathcal{E}_h}\gamma_e^p h_e^{-p} h_e^{1-p}\|[v]\|_{L^p(e\cap\bar{D})}^p\Big)^{\frac{1}{p}}\\
&\quad + \sum_{\substack{T\in\mathcal{T}_h\\ T\subset D_h}} \Big(\|D^2 v_h\|_{L^p(T)}^p + h_T^{-p}\|\nabla v_h\|_{L^p(T)}^p\Big)^{\frac{1}{p}}\\
&\lesssim h^{-1}\|v_h\|_{W_h^{1,p}(D_h)} + h^{-1}\Big(\sum_{e\in\mathcal{E}_h^I}\gamma_e^p h_e^{1-p}\|[v]\|_{L^p(e\cap\bar{D})}^p\Big)^{\frac{1}{p}}\\
&\lesssim h^{-1}\|v_h\|_{W_h^{1,p}(D_h)}.
\end{aligned}
$$

□

We also prove an inverse inequality between the L^p-norm and the $W_h^{-1,p}$-norm.

Lemma 2.3. *Let $v \in V_h(D)$. For any $1 < p < \infty$ and subdomain $D \subset \Omega$ we have*

$$
\|v_h\|_{L^p(D)} \lesssim h^{-1}\|v_h\|_{W_h^{-1,p}(D)}. \tag{2.2.5}
$$

Proof. Using the relation (1.6.7) and the definition of $\|\cdot\|_{L_h^p(\Omega)}$, we find that

$$
\|v_h\|_{L^p(D)} \le \|v_h\|_{L^p(\Omega)} \lesssim \|v_h\|_{L_h^p(\Omega)} = \sup_{0\neq w_h\in V_h}\frac{(v_h, w_h)_D}{\|w_h\|_{L^{p'}(\Omega)}} \quad \forall v_h \in V_h(D).
$$

Therefore, by the standard inverse estimate $h\|w_h\|_{W_h^{1,p'}(D)} \le h\|w_h\|_{W_h^{1,p'}(\Omega)} \lesssim \|w_h\|_{L^{p'}(\Omega)}$ and noting that $V_h(D) \subset W_h^{1,p'}(D)$, we obtain

$$
\|v_h\|_{L^p(D)} \lesssim h^{-1}\sup_{0\neq w_h\in V_h}\frac{(v_h, w_h)_D}{\|w_h\|_{W_h^{1,p'}(D)}} \le h^{-1}\sup_{0\neq w\in W_h^{1,p'}(D)}\frac{(v_h, w)_D}{\|w\|_{W_h^{1,p'}(D)}} = \|v_h\|_{W_h^{-1,p}(D)}.
$$

The proof is complete. □

The following lemma shows that the broken Sobolev norms are controlled by their corresponding Sobolev norms.

Lemma 2.4. *For any $1 < p < \infty$ there holds the following inequality:*

$$\|\varphi\|_{W_h^{2,p}(\Omega)} \le \|\varphi\|_{W^{2,p}(\Omega)} \quad \forall \varphi \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega).$$

Proof. Since the inequality holds for all $\varphi \in C^\infty(\Omega) \cap W_0^{1,p}(\Omega)$, it can be extended to all $\varphi \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ by a density argument. □

The next lemma establishes a Poincaré-Friedrichs' inequality for DG functions.

Lemma 2.5. *Let $D \subset \Omega$ such that $V_h(D) \neq \{0\}$ and $\operatorname{diam}(D) \ge h$. Then for any $v_h \in V_h(D)$ there holds the following inequalities:*

$$\|v_h\|_{L^p(D)} \lesssim \operatorname{diam}(D)\|v_h\|_{W_h^{1,p}(D)}, \tag{2.2.6}$$

$$\|v_h\|_{W_h^{1,p}(D)} \lesssim \operatorname{diam}(D)\|v_h\|_{W_h^{2,p}(D)}. \tag{2.2.7}$$

Proof. Let $\tilde{V}_h$ denote the generalized Hsiegh–Clough–Tochner space [20], and let $E_h : V_h \to \tilde{V}_h$ be the reconstruction operator constructed in [34]. The arguments given in [34] show that, for $v_h \in V_h(D)$,

$$\begin{aligned} &E_h v_h \in H_0^2(D_h), \qquad\qquad (2.2.8)\\ &\|v_h - E_h v_h\|_{L^p(\Omega)} \lesssim h\|v_h\|_{W_h^{1,p}(D)},\\ &\|v_h - E_h v_h\|_{W_h^{m,p}(\Omega)} \lesssim h^{s-m}\|v_h\|_{W_h^{s,p}(D)}, \qquad 1 \le m \le s \le 2, \end{aligned}$$

where D_h is the same as in Lemma 2.2. Therefore, by the triangle inequality, the Poincarè-Friedrichs inequality, and the assumption $\operatorname{diam}(D) \ge h$,

$$\begin{aligned} \|v_h\|_{L^p(D)} &\le \|E_h v_h\|_{L^p(D)} + \|v_h - E_h v_h\|_{L^p(\Omega)}\\ &\lesssim \operatorname{diam}(D)\|E_h v_h\|_{W^{1,p}(D_h)} + h\|v_h\|_{W_h^{1,p}(D)} \lesssim \operatorname{diam}(D)\|v_h\|_{W_h^{1,p}(D)}. \end{aligned}$$

Likewise, we find

$$\|v_h\|_{W_h^{1,p}(D)} \le \|E_h v_h\|_{W^{1,p}(D_h)} + \|v_h - E_h v_h\|_{W_h^{1,p}(\Omega)},$$
$$\lesssim \operatorname{diam}(D)\|E_h v_h\|_{W^{2,p}(D_h)} + h\|v_h\|_{W_h^{2,p}(\Omega)} \lesssim \operatorname{diam}(D)\|v_h\|_{W_h^{2,p}(D)}.$$

The proof is complete. □

Next we establish a discrete Sobolev interpolation estimate for DG functions.

Lemma 2.6. *Let $1 < p < \infty$. For all $v_h \in V_h$ we have*

$$\|v_h\|^2_{W_h^{1,p}(\Omega)} \lesssim \|v_h\|_{L^p(\Omega)}\|v_h\|_{W_h^{2,p}(\Omega)}. \tag{2.2.9}$$

Proof. Let $E_h : V_h \to \tilde{V}_h$ be the enriching operator in the proof of Lemma 2.5. By the triangle inequality and scaling we find

$$\|v_h\|^2_{W_h^{1,p}(\Omega)} \lesssim \|v_h - E_h v_h\|^2_{W_h^{1,p}(\Omega)} + \|E_h v_h\|^2_{W^{1,p}(\Omega)}. \tag{2.2.10}$$

Since $E_h v_h \in W^{2,p}(\Omega)$ we can apply the Gagliardo–Nirenberg estimate [9] to get

$$\|E_h v_h\|^2_{W^{1,p}(\Omega)} \lesssim \|E_h v_h\|_{W^{2,p}(\Omega)}\|E v_h\|_{L^p(\Omega)}.$$

Applying estimates (2.2.8), we conclude that

$$\|E_h v_h\|^2_{W^{1,p}(\Omega)} \lesssim \|v_h\|_{W_h^{2,p}(\Omega)}\|v_h\|_{L^p(\Omega)}. \tag{2.2.11}$$

Likewise, by (2.2.8) and an inverse estimate,

$$\|v_h - E_h v_h\|^2_{W_h^{1,p}(\Omega)} \lesssim h^2\|v_h\|^2_{W_h^{2,p}(\Omega)} \lesssim \|v_h\|_{L^p(\Omega)}\|v_h\|_{W_h^{2,p}(\Omega)}. \tag{2.2.12}$$

Combining (2.2.10)–(2.2.12) completes the proof. □

Next we prove some local super approximation estimates for the DG nodal interpolation in various discrete norms. The derivation of the lemma is standard (cf. [46]);

Lemma 2.7. *Let $I_h : C^0(\mathcal{T}_h) := \Pi_{T\in\mathcal{T}_h} C^0(\overline{T}) \to V_h$ denote the nodal interpolation operator, and $\eta \in C^\infty(\Omega)$ with $|\eta|_{W^{j,\infty}(\Omega)} \lesssim d^{-j}$ for $0 \le j \le k$. Then for any $v_h \in V_h$ and $D \subseteq \Omega$ we have*

$$\|\eta v_h - I_h(\eta v_h)\|_{L^p(D)} \lesssim \frac{h}{d}\|v_h\|_{L^p(D_h)}, \tag{2.2.13}$$

$$h\|\nabla_h(\eta v_h - I_h(\eta v_h))\|_{L^p(D)} \lesssim \frac{h}{d}\|v_h\|_{L^p(D_h)}, \tag{2.2.14}$$

$$h^2\|D_h^2(\eta v_h - I_h(\eta v_h))\|_{L^p(D)} \lesssim \frac{h}{d}\|v_h\|_{L^p(D_h)}, \tag{2.2.15}$$

$$\|\eta v_h - I_h(\eta v_h)\|_{W_h^{2,p}(D)} \lesssim \frac{1}{d^2}\left(\|v_h\|_{L^p(D_h)} + \|\nabla_h v_h\|_{L^p(D_h)}\right), \tag{2.2.16}$$

where D_h is the same as in Lemma 2.2. Moreover, there holds

$$\|\eta v_h - I_h(\eta v_h)\|_{W_h^{2,p}(D)} \lesssim \frac{h}{d^3}\|v_h\|_{W_h^{2,p}(D_h)} \tag{2.2.17}$$

if the polynomial degree $k \ge 2$.

Proof. From [2, Lemma 3] we have the following estimates for I_h

$$h^{mp}|\eta v_h - I_h(\eta v_h)|^p_{W^{m,p}(T)} \lesssim h^{p(k+1)}|\eta v_h|^p_{W^{k+1,p}(T)}, \quad 0 \le m \le k+1. \tag{2.2.18}$$

By the assumptions on η, the fact that $|v_h|_{W^{k+1,p}(T)} = 0$, and a standard inverse inequality we get

$$\begin{aligned} |\eta v_h|_{W^{k+1,p}(T)} &\lesssim \sum_{|\alpha|+|\beta|=k+1} \int_T |D^\alpha \eta|^p |D^\beta v_h|^p \,\mathrm{d}x \\ &\lesssim \sum_{j=0}^{k} \frac{1}{d^{p(k+1-j}}|v_h|^p_{W^{j,p}(T)} \lesssim \sum_{j=0}^{k} \frac{h^{-jp}}{d^{p(k+1-j)}}\|v_h\|^p_{L^p(T)}. \end{aligned} \tag{2.2.19}$$

It follows from (2.2.18) and (2.2.19) with $h \leq d$ that

$$h^{mp}|\eta v_h - I_h(\eta v_h)|^p_{W^{m,p}(T)} \lesssim \sum_{j=0}^{k} \frac{h^{p(k+1-j)}}{d^{p(k+1-j)}} \|v_h\|^p_{L^p(T)} \lesssim \frac{h^p}{d^p}\|v_h\|_{L^p(T)}.$$

Thus we have

$$\begin{aligned}
\|\eta v_h - I_h(\eta v_h)\|_{L^p(D)} &\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} |\eta v_h - I_h(\eta v_h)|^p_{L^p(T)} \\
&\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} \frac{h^p}{d^p}\|v_h\|_{L^p(T)} \lesssim \frac{h^p}{d^p}\|v_h\|_{L^p(D_h)},
\end{aligned}$$

$$\begin{aligned}
h^p\|\nabla_h(\eta v_h - I_h(\eta v_h))\|_{L^p(D)} &\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} h^p|\eta v_h - I_h(\eta v_h)|^p_{W^{1,p}(T)} \\
&\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} \frac{h^p}{d^p}\|v_h\|_{L^p(T)} \lesssim \frac{h^p}{d^p}\|v_h\|_{L^p(D_h)}, \\
h^{2p}\|\nabla_h(\eta v_h - I_h(\eta v_h))\|_{L^p(D)} &\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} h^{2p}|\eta v_h - I_h(\eta v_h)|^p_{W^{2,p}(T)} \\
&\lesssim \sum_{\substack{T\in\mathcal{T}_h \\ T\cap D\neq\emptyset}} \frac{h^p}{d^p}\|v_h\|_{L^p(T)} \lesssim \frac{h^p}{d^p}\|v_h\|_{L^p(D_h)}.
\end{aligned}$$

Hence (2.2.13), (2.2.14), and (2.2.15) hold.

To show (2.2.16), using (2.2.19) and an inverse estimate we have

$$\begin{aligned}
h^{p(k-1)}|\eta v_h|^p_{W^{k+1,p}(T)} &\lesssim \sum_{j=0}^{k} \frac{h^{p(k-1)}}{d^{p(k+1-j)}}|v_h|^p_{W^{j,p}(T)} \\
&\lesssim \frac{1}{d^{2p}}\|v_h\|^p_{L^p(T)} + \sum_{j=1}^{k} \frac{h^{p(k-j)}}{d^{p(k+1-j)}}|v_h|^p_{W^{1,p}(T)} \\
&\lesssim \frac{1}{d^{2p}}\Big(\|v_h\|^p_{L^p(T)} + |v_h|^p_{W^{1,p}(T)}\Big).
\end{aligned} \tag{2.2.20}$$

It follows from (2.2.18) and (2.2.20) that

$$\begin{aligned}
\|D^2(\eta v_h - I_h(\eta v_h))\|_{L^p(T)}^p &\lesssim h^{p(k-1)}|\eta v_h|_{W^{k+1,p}(T)}^p \lesssim \frac{1}{d^{2p}}\Big(\|v_h\|_{L^p(T)}^p + |v_h|_{W^{1,p}(T)}^p\Big),\\
h^{-p}\|\nabla(\eta v_h - I_h(\eta v_h))\|_{L^p(T)}^p &\lesssim h^{p(k-1)}|\eta v_h|_{W^{k+1,p}(T)}^p \lesssim \frac{1}{d^{2p}}\Big(\|v_h\|_{L^p(T)}^p + |v_h|_{W^{1,p}(T)}^p\Big),\\
h^{-2p}\|\eta v_h - I_h(\eta v_h)\|_{L^p(T)}^p &\lesssim h^{p(k-1)}|\eta v_h|_{W^{k+1,p}(T)}^p \lesssim \frac{1}{d^{2p}}\Big(\|v_h\|_{L^p(T)}^p + |v_h|_{W^{1,p}(T)}^p\Big).
\end{aligned}$$

Using the previous three estimates and Lemma 2.1 we get

$$\begin{aligned}
\|\eta v_h - I_h(\eta v_h)\|_{W_h^{2,p}(D)}^p &\lesssim \sum_{T\in\mathcal{T}_h} \|D^2(\eta v_h - I_h(\eta v_h))\|_{L^p(T)}^p\\
&\quad + \sum_{e\in\mathcal{E}_h^I} h_e^{1-p}\|[\nabla(\eta v_h - I_h(\eta v_h))]\|_{L^p(e)}^p + \sum_{e\in\mathcal{E}_h^I} h_e^{1-2p}\|[\eta v_h - I_h(\eta v_h)]\|_{L^p(e)}^p\\
&\quad + \sum_{e\in\mathcal{E}_h^B} h_e^{1-2p}\|\eta v_h - I_h(\eta v_h)\|_{L^p(e)}^p\\
&\lesssim \sum_{\substack{T\in\mathcal{T}_h\\ T\cap D\neq\emptyset}} \|D^2(\eta v_h - I_h(\eta v_h))\|_{L^p(T)}^p + \sum_{\substack{T\in\mathcal{T}_h\\ T\cap D\neq\emptyset}} h^{-p}\|\nabla(\eta v_h - I_h(\eta v_h))\|_{L^p(T)}^p\\
&\quad + \sum_{\substack{T\in\mathcal{T}_h\\ T\cap D\neq\emptyset}} h^{-2p}\|\eta v_h - I_h(\eta v_h)\|_{L^p(T)}^p\\
&\lesssim \sum_{\substack{T\in\mathcal{T}_h\\ T\cap D\neq\emptyset}} \frac{1}{d^{2p}}\Big(\|v_h\|_{L^p(T)}^p + |v_h|_{W^{1,p}(T)}^p\Big) \lesssim \frac{1}{d^{2p}}\Big(\|v_h\|_{L^p(T)}^p + \|\nabla_h v_h\|_{L^p(T)}^p\Big).
\end{aligned}$$

Thus, (2.2.16) holds.

Finally, the proof of (2.2.17) is similar to that of (2.2.16) except one minor detail. Since $k \geq 2$, by (2.2.19) and an inverse inequality we get

$$\begin{aligned}
h^{p(k-1)}|\eta v_h|_{W^{k+1,p}(T)}^p &\lesssim \sum_{j=0}^{k} \frac{h^{p(k-1)}}{d^{p(k+1-j)}}|v_h|_{W^{j,p}(T)}^p \qquad (2.2.21)\\
&= h^p\Big(\sum_{j=0}^{k} \frac{h^{p(k-2)}}{d^{p(k+1-j)}}|v_h|_{W^{j,p}(T)}^p\Big)
\end{aligned}$$

$$
\begin{aligned}
&= h^p\Big(\frac{1}{d^{3p}}\|v_h\|^p_{L^p(T)} + \frac{1}{d^{2p}}|v_h|^p_{W^{1,p}(T)} + \sum_{j=2}^{k}\frac{h^{p(k-j)}}{d^{p(k+1-j)}}|v|^p_{W^{2,p}(T)}\Big)\\
&\lesssim \frac{h^p}{d^{3p}}\|v_h\|^p_{W^{2,p}(T)}.
\end{aligned}
$$

Thus, we can obtain (as in the derivation on (2.2.16)) using Lemma 2.5 that

$$
\begin{aligned}
\|\eta v_h - I_h(\eta v_h)\|_{W_h^{2,p}(D)} &\lesssim \frac{h}{d^3}\sum_{\substack{T\in\mathcal{T}_h\\ T\cap D\neq\emptyset}}\|v_h\|_{W^{2,p}(T)}\\
&= \frac{h}{d^3}\Big(\|v_h\|_{L^p(D_h)} + \|\nabla_h v_h\|_{L^p(D_h)} + \|D_h^2 v_h\|_{L^p(D_h)}\Big)\\
&\lesssim \frac{h}{d^3}\|v_h\|_{W_h^{2,p}(D_h)}.
\end{aligned}
$$

The proof is complete. □

2.3 DG discrete $W^{1,p}$ and Calderon-Zygmund estimates for PDEs with constant coefficients

In this section we consider the constant coefficient case, that is, $A(x) \equiv A_0 \in \mathbb{R}^{n\times n}$ on Ω. We define three interior-penalty discontinuous Galerkin discretizations $\mathcal{L}^{\varepsilon}_{0,h}$ to the PDE operator $\mathcal{L}$ and extend their domains to the broken Sobolev space $W^{2,p}(\mathcal{T}_h)$. Our goal in this subsection is to prove global stability estimates for $\mathcal{L}^{\varepsilon}_{0,h}$ which will be crucial in the next section. The final global stability estimate given in Theorem 2.2 can be regarded as a DG discrete Calderon-Zygmund estimate for $\mathcal{L}^{\varepsilon}_{0,h}$.

Let A_0 be a constant, positive-definite matrix in $\mathbb{R}^{n\times n}$ and define

$$
\mathcal{L}_0 w := -A_0 : D^2 w = -\operatorname{div}(A_0\nabla w). \tag{2.3.1}
$$

From this we gather the standard PDE weak form:

$$a_0(w,v) := \int_\Omega A_0 \nabla w \cdot \nabla v \,\mathrm{d}x \qquad \forall w, v \in H^1_0(\Omega). \tag{2.3.2}$$

The Lax-Milgram theorem [8] yields the existence and boundedness of $\mathcal{L}_0^{-1} : H^{-1}(\Omega) \to H^1_0(\Omega)$. Moreover, if $\partial\Omega \in C^{1,1}$ we have from Calderon-Zygmund theory [35] that $\mathcal{L}_0^{-1} : L^p(\Omega) \to W^{2,p}(\Omega) \cap W^{1,p}_0(\Omega)$ exists and

$$\|\mathcal{L}_0^{-1}\varphi\|_{W^{2,p}(\Omega)} \lesssim \|\varphi\|_{L^p(\Omega)} \qquad \forall \varphi \in L^p(\Omega),$$

and therefore

$$\|w\|_{W^{2,p}(\Omega)} \lesssim \|\mathcal{L}_0 w\|_{L^p(\Omega)} \qquad \forall w \in W^{2,p}(\Omega) \cap W^{1,p}_0(\Omega).$$

Define $\mathcal{L}^\varepsilon_{0,h} : V_h \to V_h$ by

$$\big(\mathcal{L}^\varepsilon_{0,h} w_h, v_h\big) := a^\varepsilon_{0,h}(w_h, v_h) \qquad \forall v_h, w_h \in V_h, \tag{2.3.3}$$

where the IP-DG bilinear form is defined by

$$\begin{aligned} a_{0,h}(w_h,v_h) := \int_\Omega A_0 \nabla_h w_h \cdot \nabla_h v_h \,\mathrm{d}x - \sum_{e\in\mathcal{E}_h} \int_e \{A_0 \nabla w_h \cdot \nu_e\}[v_h] \,\mathrm{d}S \\ - \varepsilon \sum_{e\in\mathcal{E}_h} \int_e \{A_0 \nabla v_h \cdot \nu_e\}[w_h] \,\mathrm{d}S + \sum_{e\in\mathcal{E}_h} \int_e \frac{\gamma_e}{h_e}[w_h][v_h] \,\mathrm{d}S, \end{aligned} \tag{2.3.4}$$

and $\gamma_e > 0$ is a penalization parameter. The parameter choices $\varepsilon \in \{1, 0, -1\}$ give respectively the SIP-DG, IIP- DG, and NIP-DG formulations. For the sake of clarity and readability we shall assume for the rest of this chapter that ε may be either 1, 0, or -1 unless otherwise stated. Applying the DG-integration by parts formula (2.1.4) to the first term on the right-hand side of (2.3.4) yields

$$\begin{aligned} a_{0,h}^{\varepsilon}(w_h, v_h) = &- \int_{\Omega} (A_0 : D_h^2 w_h) v_h \, \mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_e [A_0 \nabla w_h \cdot \nu_e]\{v_h\} \, \mathrm{d}S \\ &- \varepsilon \sum_{e \in \mathcal{E}_h} \int_e \{A_0 \nabla v_h \cdot \nu_e\}[w_h] \, \mathrm{d}S + \sum_{e \in \mathcal{E}_h} \int_e \frac{\gamma_e}{h_e} [w_h][v_h] \, \mathrm{d}S \end{aligned} \tag{2.3.5}$$

for any $w_h, v_h \in V_h$. By Hölder's inequality, it is easy to check that the above new form of $a_{0,h}^{\varepsilon}(\cdot,\cdot)$ is also well-defined on $W^{2,p}(\mathcal{T}_h) \times W^{2,p'}(\mathcal{T}_h)$ with $\frac{1}{p} + \frac{1}{p'} = 1$. As a result, this new form enables us to extend the domain of $a_{0,h}^{\varepsilon}(\cdot,\cdot)$ to $W^{2,p}(\mathcal{T}_h) \times W^{2,p'}(\mathcal{T}_h)$ and $\mathcal{L}_{0,h}^{\varepsilon} : W^{2,p}(\mathcal{T}_h) \to (W^{2,p}(\mathcal{T}_h))^*$.

2.3.1 DG discrete $W^{1,p}$ error estimates

From the standard IP-DG theory [49], there exists $\gamma^* = \gamma^*(\|A_0\|_{L^\infty(\Omega)}, \mathcal{T}_h) > 0$ depending only on the shape regularity of the mesh and on $\|A_0\|_{L^\infty(\Omega)}$ such that $\mathcal{L}_{0,h}^{\varepsilon}$ is invertible on V_h provided $\gamma_e \geq \gamma^*$; in the non–symmetric case $\varepsilon = -1$, γ^* can be any positive number. Moreover, if $w \in W^{2,2}(\mathcal{T}_h) \cap H_0^1(\Omega)$ and $w_h \in V_h$ satisfy

$$a_{0,h}^{\varepsilon}(w - w_h, v_h) = 0 \qquad \forall v_h \in V_h, \tag{2.3.6}$$

then the quasi-optimal error estimate

$$\|w - w_h\|_{W_h^{1,2}(\Omega)} \lesssim \inf_{v_h \in V_h} \|w - v_h\|_{W_h^{1,2}(\Omega)} \tag{2.3.7}$$

is satisfied. The goal of this subsection is to generalize this result to general exponent $p \in (1, \infty)$ for the SIP-DG method. In particular, we have

Theorem 2.1. *Suppose $w \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ $(1 < p < \infty)$ and $w_h \in V_h$ satisfy (2.3.6) with $\varepsilon = 1$. Then there holds*

$$\|w - w_h\|_{W_h^{1,p}(\Omega)} \lesssim h|\log h|^t \|w\|_{W^{2,p}(\Omega)}, \tag{2.3.8}$$

where $t = (p+1)/p$ *if* $k = 1$ *and* $t = 0$ *if* $k \geq 2$.

This result, while of independent interest, is quite technical and borrows the techniques of L^∞ error estimates from [14] to prove Theorem 2.1 for $p > 2$. Then we perform a duality argument using the symmetry of $a^1_{0,h}(\cdot,\cdot)$ to prove the case $1 < p < 2$. We include the proof for completeness which can also be found in [28].

To prove Theorem 2.1 we introduce some notation given in [14] (also see [52]). For given $z \in \overline{\Omega}$, we define the weight function σ_z as

$$\sigma_z(x) = \frac{h}{|x-z|+h}. \tag{2.3.9}$$

For $1 \leq p < \infty$ and $s \in \mathbb{R}$, we define the following weighted norms

$$\begin{aligned}
\|v\|_{L^p(D),z,s} &= \left(\int_D \left|\sigma_z^s(x)v(x)\right|^p \,\mathrm{d}x\right)^{1/p}, \\
\|v\|_{W^{1,p}(D),z,s} &= \|v\|_{L^p(D),z,s} + \|\nabla_h v\|_{L^p(D),z,s}, \\
\|v\|_{W_h^{1,p}(D),z,s} &= \|v\|_{W^{1,p}(D),z,s} + \Big(\sum_{e\in\mathcal{E}_h} h_e^{1-p}\big\|\sigma_z^s[v]\big\|^p_{L^p(e\cap\overline{D})}\Big)^{1/p} \\
&\quad + \Big(\sum_{e\in\mathcal{E}_h} h_e\big\|\sigma_z^s\{\nabla_h v\}\big\|^p_{L^p(e\cap\overline{D})}\Big)^{1/p}.
\end{aligned} \tag{2.3.10}$$

The weighted norms in the case $p = \infty$ are defined analogously.

The derivation of $W^{1,p}$ error estimates of DG approximations is based on the work [14], where localized pointwise estimates of DG approximations are obtained. There it was shown that if $w \in W^{2,\infty}(\Omega)$ and $w_h \in V_h$ satisfy (2.3.6) with $\varepsilon = 1$, then

$$|\nabla(w-w_h)(z)| \lesssim \inf_{v_h\in V_h} \|w-v_h\|_{W_h^{1,\infty}(\Omega),z,s} \quad 0 \leq s < k \tag{2.3.11}$$

for all $z \in \overline{\Omega}$. Similar to pointwise estimates of finite element approximations (e.g., [52, 8]), the ingredients to prove (2.3.11) include duality arguments and DG approximation estimates of regularized Green functions in a weighted (discrete) $W^{1,1}$-norm. These results are rather technical and involve dyadic decompositions of Ω, local DG error estimates, and Green function estimates.

Here, we follow a similar argument to derive $W^{1,p}$ estimates; the main difference being that we derive DG approximation estimates of regularized Green functions in a weighted (discrete) $W^{1,p'}$-norm with $1/p + 1/p' = 1$ (cf. Lemma 2.3). Using these estimates and applying similar arguments in [52, 14] then yield the estimate

$$|\nabla(w - w_h)(z)|^p \lesssim h^{-n} \inf_{v_h \in V_h} \|w - v_h\|^p_{W_h^{1,p}(\Omega),z,s}$$

for certain values of s. Integrating this expression with respect to z and applying Fubini's theorem (cf. Lemma 2.1) then yields L^p estimates of the piecewise gradient error.

Unfortunately, the strategy just described does not immediately give us estimates for the terms $h_e^{1-p}\|[w - w_h]\|^p_{L^p(e)}$ appearing in the $W_h^{1,p}$-norm. To bypass this difficulty, we first use the trace inequality

$$\sum_{e \in \mathcal{E}_h} h_e^{1-p}\|[w - w_h]\|^p_{L^p(e)} \lesssim \|\nabla_h(w - w_h)\|^p_{L^p(\Omega)} + h^{-p}\|w - w_h\|^p_{L^p(\Omega)},$$

and then derive estimates for $h^{-p}\|w - w_h\|^p_{L^p(\Omega)}$. We note that the standard duality argument to derive L^p estimates yields

$$\|w - w_h\|_{L^p(\Omega)} \lesssim h\|w - w_h\|_{W_h^{1,p}(\Omega)},$$

which is of little benefit. Rather, our strategy is to modify the arguments given in [14, Theorem 5.1] and estimate $|(w - w_h)(z)|$ in terms of $\inf_{v_h \in V_h} \|w - v_h\|_{W_h^{1,p}(\Omega),z,s}$ (cf. Lemma 2.2) and then apply Fubini's theorem. We note that it is due to this term that the $|\log h|^t$ factor appears in Theorem 2.1.

Lemma 2.1. *Let $p \in [2, \infty)$ and $v \in L^p(\Omega)$. Let $z \in \Omega$ and $T_z \in \mathcal{T}_h$ such that $z \in T_z$. Then there holds*

$$\int_\Omega \int_{T_z} |v(x)|^p \,\mathrm{d}x\,\mathrm{d}z \lesssim h^n \|v\|^p_{L^p(\Omega)}. \tag{2.3.12}$$

Moreover for any $s > n/p$ and $w \in W^{2,p}(\mathcal{T}_h)$, there holds

$$\int_\Omega \|v\|^p_{W^{1,p}_h(\Omega),z,s} \,\mathrm{d}z \lesssim \frac{h^n}{ps-n}\big(\|\nabla_h v\|^p_{L^p(\Omega)} + h^{-p}\|v\|^p_{L^p(\Omega)} + h^p\|D^2_h v\|^p_{L^p(\Omega)}\big). \tag{2.3.13}$$

If $s = n/p$, then we have

$$\int_\Omega \|v\|^p_{W^{1,p}_h(\Omega),z,n/p} \,\mathrm{d}z \lesssim |\log h| h^n\big(\|\nabla_h v\|^p_{L^p(\Omega)} + h^{-p}\|v\|^p_{L^p(\Omega)} + h^p\|D^2_h v\|^p_{L^p(\Omega)}\big). \tag{2.3.14}$$

Proof. (i) Let $v \in L^p(\Omega)$ and extend v to $\mathbb{R}^n$ by zero. Denote by $B_h(z)$ the ball of radius h and center z. Then by a change of variables and interchanging integrals, we find

$$\begin{aligned}
\int_\Omega \int_{T_z} |v(x)|^p \,\mathrm{d}x\,\mathrm{d}z &\le \int_\Omega \int_{B_h(z)} |v(x)|^p \,\mathrm{d}x\,\mathrm{d}z \\
&= h^n \int_\Omega \int_{B_1(0)} |v(z+hy)|^p \,\mathrm{d}y\,\mathrm{d}z \\
&= h^n \int_{B_1(0)} \int_\Omega |v(z+hy)|^p \,\mathrm{d}z\,\mathrm{d}y \\
&\lesssim h^n \int_{B_1(0)} \|v\|^p_{L^p(\Omega)} \,\mathrm{d}y \lesssim h^n \|v\|^p_{L^p(\Omega)}.
\end{aligned}$$

This proves (2.3.12).

(ii) To prove (2.3.13) we again extend v to $\mathbb{R}^n$ by zero and make a change of variables to obtain

$$\int_\Omega \|\sigma^s_z v\|^p_{L^p(\Omega)} \,\mathrm{d}z = \int_\Omega \int_\Omega \Big(\frac{h}{|x-z|+h}\Big)^{sp} |v(x)|^p \,\mathrm{d}x\,\mathrm{d}z$$

$$\begin{aligned}&\leq h^n \int_\Omega \int_{\hat\Omega} \frac{h^{sp}}{(|hy|+h)^{sp}} |v(z+hy)|^p \,\mathrm{d}y \,\mathrm{d}z\\ &= h^n \int_{\hat\Omega} \Big(\int_\Omega |v(z+hy)|^p \,\mathrm{d}z \Big) \frac{1}{(|y|+1)^{sp}} \,\mathrm{d}y,\end{aligned}$$

where $\hat\Omega = \{2h^{-1}x : \ x \in \Omega\}$ is a dilation of Ω. Therefore,

$$\int_\Omega \|\sigma_z^s v\|^p_{L^p(\Omega)} \,\mathrm{d}z \lesssim h^n \|v\|^p_{L^p(\Omega)} \int_{\hat\Omega} \frac{1}{(|y|+1)^{sp}} \,\mathrm{d}y. \tag{2.3.15}$$

For $sp > n$, there holds

$$\int_{\hat\Omega} \frac{1}{(|y|+1)^{sp}} \,\mathrm{d}y \lesssim \int_0^\infty \frac{r^{n-1}}{(r+1)^{sp}} \,\mathrm{d}r = (n-1)! \prod_{j=1}^n (sp-j)^{-1} \leq \frac{(n-1)!}{sp-n}.$$

Combining this identity with (2.3.15) yields the inequality

$$\int_\Omega \|\sigma_z^s v\|^p_{L^p(\Omega)} \,\mathrm{d}z \lesssim \frac{h^n}{sp-n} \|v\|^p_{L^p(\Omega)}. \tag{2.3.16}$$

If $sp = n$, then we find by a direct calculation that

$$\int_{\hat\Omega} \frac{1}{(|y|+1)^n} \,\mathrm{d}y \lesssim \int_0^{h^{-1}} \frac{r^{n-1}}{(r+1)^n} \,\mathrm{d}r = -\sum_{j=1}^{n-1} \frac{1}{(h+1)^{n-j}} + \log(1+h^{-1}) \lesssim |\log h|,$$

and therefore by (2.3.15),

$$\int_\Omega \|\sigma_z^{p/n} v\|^p_{L^p(\Omega)} \,\mathrm{d}z \lesssim |\log h| h^n \|v\|_{L^p(\Omega)}. \tag{2.3.17}$$

Next, by trace inequalities given in Lemma 2.1, we have

$$\begin{aligned}\sum_{e\in\mathcal{E}_h} h_e^{1-p} \|[\sigma_z^s v]\|^p_{L^p(e)} &\lesssim h^{-p} \|\sigma_z^s v\|^p_{L^p(\Omega)} + \|\nabla_h(\sigma_z^s v)\|^p_{L^p(\Omega)}\\ &\lesssim h^{-p} \|\sigma_z^s v\|^p_{L^p(\Omega)} + \|v\nabla(\sigma_z^s)\|^p_{L^p(\Omega)} + \|\sigma_z^s \nabla_h v\|^p_{L^p(\Omega)}.\end{aligned}$$

Noting that

$$|\nabla(\sigma_z^s)| \lesssim \frac{h^s}{(|x-z|+h)^{s+1}} = \frac{\sigma_z^s}{|x-z|+h} \lesssim h^{-1}\sigma_z^s,$$

we obtain

$$\sum_{e\in\mathcal{E}_h} h_e^{1-p}\|[\sigma_z^s v]\|_{L^p(e)}^p \lesssim h^{-p}\|\sigma_z^s v\|_{L^p(\Omega)}^p + \|\sigma_z^s \nabla_h v\|_{L^p(\Omega)}^p. \tag{2.3.18}$$

Likewise we have

$$\sum_{e\in\mathcal{E}_h} h_e\|\sigma_z^s\{\nabla v\}\|_{L^p(e)}^p \lesssim \|\sigma_z^s \nabla_h v\|_{L^p(\Omega)}^p + h^p\|\sigma_z^s D_h^2 v\|_{L^p(\Omega)}. \tag{2.3.19}$$

Combining (2.3.18)–(2.3.19) yields

$$\|v\|_{W_h^{1,p}(\Omega),z,s} \lesssim \|\sigma_z^s \nabla_h v\|_{L^p(\Omega)}^p + h^{-p}\|\sigma_z^s v\|_{L^p(\Omega)}^p + h^p\|\sigma_z^s D_h^2 v\|_{L^p(\Omega)}^p. \tag{2.3.20}$$

Finally applying the identities (2.3.16)–(2.3.17) to (2.3.20) yields the desired result (2.3.13)–(2.3.14). The proof is complete. □

Lemma 2.2. *Let $w \in W^{2,p}(\mathcal{T}_h)$ $(2 \le p \le \infty)$ and $w_h \in V_h$ satisfy (2.3.6) with $\varepsilon = 1$. Then for any $0 \le s \le k-1+n/p$ and $z \in \overline{\Omega}$,*

$$|(w-w_h)(z)| \lesssim h^{1-n/p}|\log h|^{\bar{s}(p)} \inf_{v_h\in V_h} \|w-v_h\|_{W_h^{1,p}(\Omega),z,s},$$

where $\bar{s}(p) = 1$ if $k = s+1-n/p$ and $\bar{s}(p) = 0$ for $k > s+1-n/p$.

Proof. Step 1: Set-up. By the triangle inequality, an inverse estimate, and Hölder's inequality we obtain

$$\begin{aligned}|(w-w_h)(z)| &\le |(w-v_h)(z)| + \|v_h - w_h\|_{L^\infty(T_z)}\\ &\le |(w-v_h)(z)| + h^{-n/2}\|v_h - w_h\|_{L^2(T_z)}\end{aligned}$$

$$
\begin{aligned}
&\leq |(w - v_h)(z)| + h^{-n/2}\big(\|w - w_h\|_{L^2(T_z)} + \|w - v_h\|_{L^2(T_z)}\big) \\
&\leq \|w - v_h\|_{L^\infty(T_z)} + h^{-n/2}\|w - w_h\|_{L^2(T_z)}.
\end{aligned}
$$

Therefore by standard approximation theory, and since $\sigma_z \approx 1$ on T_z, we have

$$
\begin{aligned}
|(w - w_h)(z)| &\leq h^{1-n/p}\|w\|_{W_h^{1,p}(T_z)} + h^{-n/2}\|w - w_h\|_{L^2(T_z)} \\
&\leq h^{1-n/p}\|w\|_{W_h^{1,p}(\Omega),z,s} + h^{-n/2}\|w - w_h\|_{L^2(T_z)}.
\end{aligned}
$$

Replacing w and w_h by $w - v_h$ and $w_h - v_h$, respectively, yields

$$
|(w - w_h)(z)| \lesssim h^{1-n/p}\|w - v_h\|_{W_h^{1,p}(\Omega),z,s} + h^{-n/2}\|w - w_h\|_{L^2(T_z)}. \tag{2.3.21}
$$

Next, define $\rho \in L^2(\Omega)$ by

$$
\rho(x) = \begin{cases} \frac{h^{-n/2}(w-w_h)(x)}{\|w-w_h\|_{L^2(T_z)}} & \text{if } x \in T_z \\ 0 & \text{otherwise,} \end{cases}
$$

and let $g_z \in H_0^1(\Omega)$ be the regularized Green's function satisfying

$$
\mathcal{L}_0 g_z = \rho. \tag{2.3.22}
$$

Setting $g_{z,h}$ to be the DG approximation of g_z, i.e., $a_{0,h}(v_h, g_z - g_{z,h}) = 0,\ \forall v_h \in V_h$, and $e_z := g_z - g_{z,h}$, we have by Galerkin orthogonality and the continuity of the bilinear form,

$$
\begin{aligned}
h^{-n/2}\|w - w_h\|_{L^2(T_z)} &= (\rho, w - w_h) = a_{0,h}(w - w_h, g_z) \\
&= a_{0,h}(w - v_h, e_z) \lesssim \|w - v_h\|_{W_h^{1,p}(\Omega),s,z}\|e_z\|_{W_h^{1,p'}(\Omega),-s,z}.
\end{aligned}
$$

Consequently, by (2.3.21), we have

$$|(w - w_h)(z)| \lesssim \|w - v_h\|_{W_h^{1,p}(\Omega),z,s}\big(h^{1-n/p} + \|e_z\|_{W_h^{1,p'}(\Omega),-s,z}\big) \quad \forall v_h \in V_h. \quad (2.3.23)$$

Thus, the proof will be completed once it is shown that $\|e_z\|_{W_h^{1,p'}(\Omega),-s,z} \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p}$. This result is derived in the following steps.

Step 2: Dyadic decomposition of Ω*.* To estimate $\|e_z\|_{W_h^{1,p'}(\Omega),-s,z}$ we require some more notation. Without loss of generality, assume that $\operatorname{diam}(\Omega) = 1$. Let $d_j = 2^{-j}$ and set

$$\begin{aligned}
\Omega_j &= \{x \in \Omega : \ d_{j+1} < |z - x| < d_j\},\\
\Omega_j' &= \{x \in \Omega : \ d_{j+2} < |z - x| < d_{j-1}\},\\
\Omega_j'' &= \{x \in \Omega : \ d_{j+3} < |z - x| < d_{j-2}\}.
\end{aligned}$$

Let $M > 1$ be a real number to be determined later, and let $J \approx |\log h|$ be an integer such that $Mh = 2^{-J}$. We then write

$$\|e_z\|_{W_h^{1,p'}(\Omega),z,-s} \lesssim \|e_z\|_{W_h^{1,p'}(B_{Mh}(z)),z,-s} + \sum_{j=0}^{J} \|e_z\|_{W_h^{1,p'}(\Omega_j),z,-s}. \quad (2.3.24)$$

Note that, by the definition of Ω_j, the weighted norms, and Hölder's inequality that

$$\begin{aligned}
\|e_z\|_{W_h^{1,p'}(\Omega_j),z,-s} &\lesssim d_j^{n/q+s} h^{-s} \|e_z\|_{W_h^{1,2}(\Omega_j)},\\
\|e_z\|_{W_h^{1,p'}(B_{Mh}),z,-s} &\lesssim h^{n/q} \|e_z\|_{W_h^{1,2}(B_{Mh}(z))} \le h^{n/q} \|e_z\|_{W_h^{1,2}(\Omega)},
\end{aligned}$$

where

$$q \in [2, \infty] \text{ satisfies } 1/q + 1/p = 1/2.$$

Applying these estimates to (2.3.24) yields

$$\begin{aligned}\|e_z\|_{W_h^{1,p'}(\Omega),z,-s} &\lesssim h^{n/q}\|e_z\|_{W_h^{1,2}(\Omega)} + \sum_{j=0}^{J} d_j^{n/q+s} h^{-s}\|e_z\|_{W_h^{1,2}(\Omega_j)} \\ &= h^{n/q}\|e_z\|_{W_h^{1,2}(\Omega)} + Q_h,\end{aligned} \tag{2.3.25}$$

where

$$Q_h := h^{-s}\sum_{j=0}^{J} d_j^{n/q+s}\|e_z\|_{W_h^{1,2}(\Omega_j)}. \tag{2.3.26}$$

To estimate the first term in the right–hand side of (2.3.25), we apply elliptic regularity and the identity $\|\rho\|_{L^2(\Omega)} = h^{-n/2}$ to obtain

$$\|e_z\|_{W_h^{1,2}(\Omega)} \lesssim h\|g_z\|_{W^{2,2}(\Omega)} \lesssim h\|\rho\|_{L^2(\Omega)} = h^{1-n/2}.$$

Applying this estimate in (2.3.25) and using the identity $1 - n/2 + n/q = 1 - n/p$ yields

$$\|e_z\|_{W_h^{1,p'}(\Omega),z,-s} \lesssim h^{1-n/p} + Q_h. \tag{2.3.27}$$

It remains to find an appropriate upper bound of Q_h to complete the proof.

Step 3: Estimate of Q_h –Local error estimates. Lemma 4.4 in [14] states that

$$\|e_z\|_{W_h^{1,2}(\Omega_j)} \lesssim h^k d_j^{1-k-n/2} + d_j^{-1}\|e_z\|_{L^2(\Omega_j')}.$$

Applying this estimate to the definition of Q_h (2.3.26) yields

$$\begin{aligned}Q_h &\lesssim h^{k-s}\sum_{j=0}^{J} d_j^{n/q+s+1-k-n/2} + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|e_z\|_{L^2(\Omega_j')} \\ &= h^{k-s}\sum_{j=0}^{J} d_j^{-(k-s+n/p-1)} + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|e_z\|_{L^2(\Omega_j')}\end{aligned}$$

$$= h^{1-n/p}\Theta(k-s+n/p-1) + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|e_z\|_{L^2(\Omega'_j)},$$

where

$$\Theta(\tau) := \sum_{j=0}^{J}\left(\frac{h}{d_j}\right)^{\tau}.$$

Therefore, since (cf. [14, (5.19)])

$$\Theta(\tau) \lesssim \begin{cases} |\log h| & \text{if } \tau = 0, \\ \frac{1}{M^{\tau}(1-2^{-\tau})} & \text{if } \tau > 0, \end{cases}$$

we find that

$$Q_h \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|e_z\|_{L^2(\Omega'_j)}. \tag{2.3.28}$$

Step 4: Estimate of Q_h – Duality Arguments. Applying [14, (5.24)] yields

$$\|e_z\|_{L^2(\Omega'_j)} \lesssim h^k d_j^{1-k-n/2}\|e_z\|_{W_h^{1,1}(\Omega)} + h\|e_z\|_{W_h^{1,2}(\Omega''_j)}. \tag{2.3.29}$$

Using estimates (2.3.29) and (2.3.28), and noting that $\max_{0\le j\le J} d_j^{-1} = 2^J = 1/(hM)$, we find

$$\begin{aligned} Q_h &\lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{k-s}\sum_{j=0}^{J} d_j^{s-k-n/p}\|e_z\|_{W_h^{1,1}(\Omega)} + h^{1-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|e_z\|_{W_h^{1,2}(\Omega''_j)} \\ &\lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{k-s}\sum_{j=0}^{J} d_j^{s-k-n/p}\|e_z\|_{W_h^{1,1}(\Omega)} + \frac{h^{-s}}{M}\sum_{j=0}^{J} d_j^{n/q+s}\|e_z\|_{W_h^{1,2}(\Omega''_j)} \\ &\lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{-n/p}\Theta(k-s+n/p)\|e_z\|_{W_h^{1,1}(\Omega)} + \frac{1}{M}Q_h. \end{aligned}$$

Taking M sufficiently large yields

$$Q_h \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{-n/p}\Theta(k-s+n/p)\|e_z\|_{W_h^{1,1}(\Omega)}.$$

Applying this estimate to (2.3.27) then yields

$$\|e_z\|_{W_h^{1,p}(\Omega),z,-s} \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p} + h^{-n/p}\Theta(k-s+n/p)\|e_z\|_{W_h^{1,1}(\Omega)}. \tag{2.3.30}$$

In particular, the case $s=0$, $p=\infty$, $p'=1$ gives

$$\|e_z\|_{W_h^{1,1}(\Omega)} \lesssim |\log h|^{\bar{s}(\infty)} h + \Theta(k)\|e_z\|_{W_h^{1,1}(\Omega)}.$$

Since

$$\Theta(k) \lesssim \frac{1}{M^k(1-2^{-k})},$$

we can take M sufficiently large to conclude that

$$\|e_z\|_{W_h^{1,1}(\Omega)} \lesssim |\log h|^{\bar{s}(\infty)} h.$$

Finally, applying this last estimate to (2.3.30) yields

$$\|e_z\|_{W_h^{1,p}(\Omega),z,-s} \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p}\big(1+\Theta(k-s+n/p)\big) \lesssim |\log h|^{\bar{s}(p)} h^{1-n/p}.$$

Applying this last estimate to (2.3.23) completes the proof. □

Lemma 2.3. *Let z and T_z be as in Lemma 2.1. For arbitrary $\varphi \in C_0^\infty(T_z)$, with $\|\varphi\|_{W^{1,2}(T_z)} = 1$, we extend φ to Ω by zero, and let $\hat{g}_z$ be the solution to*

$$\mathcal{L}_0^* \hat{g}_z = h^{-n/2-1}\partial\varphi/\partial x_i \quad \textit{in } \Omega, \qquad \hat{g}_z = 0 \quad \textit{on } \partial\Omega.$$

Let $\hat{g}_{z,h} \in V_h$ satisfy the discrete adjoint problem

$$a_{0,h}(v_h, \hat{g}_{z,h}) = h^{-n/2-1} \int_\Omega (\partial\varphi/\partial x_i) v_h \,\mathrm{d}x \qquad \forall v_h \in V_h,$$

where we have dropped the superscript of the bilinear form for notational simplicity. Let $p \in [2,\infty]$, $p' \in [1,2]$ such that $1/p + 1/p' = 1$. Then for any $0 \le s \le k + n/p$ there holds

$$\|\hat{g}_z - \hat{g}_{z,h}\|_{W_h^{1,p'}(\Omega),z,-s} \lesssim |\log h|^{\bar{\bar{s}}(p)} h^{-n/p},$$

where $\bar{\bar{s}}(p) = 1$ if $s = k + n/p$ and $\bar{\bar{s}}(p) = 0$ otherwise.

Proof. Set $\hat{e}_z = \hat{g}_z - \hat{g}_{z,h}$, and for $M > 0$, let J satisfy $Mh = 2^{-J}$. Then by applying similar arguments as the proof of Lemma 2.2, we obtain

$$\begin{aligned}
\|\hat{e}_z\|_{W_h^{1,p'}(\Omega),z,-s} &\le \|\hat{e}_z\|_{W_h^{1,p'}(B_{Mh}(z)),z,-s} + \sum_{j=0}^{J} \|\hat{e}_z\|_{W_h^{1,p'}(\Omega_j)} \\
&\lesssim h^{n/q} \|\hat{e}_z\|_{W_h^{1,2}(\Omega)} + h^{-s} \sum_{j=0}^{J} d_j^{n/q+s} \|\hat{e}_z\|_{W_h^{1,2}(\Omega_j)} \\
&\lesssim h^{n/q+1} \|\hat{g}_z\|_{W_h^{2,2}(\Omega)} + h^{-s} \sum_{j=0}^{J} d_j^{n/q+s} \|\hat{e}_z\|_{W_h^{1,2}(\Omega_j)} \\
&\lesssim h^{-n/p} + \hat{F}_h,
\end{aligned}$$

with

$$\hat{F}_h := h^{-s} \sum_{j=0}^{J} d_j^{n/q+s} \|\hat{e}_z\|_{W_h^{1,2}(\Omega_j)}. \tag{2.3.32}$$

By the local error estimate given in [14, Lemma 4.2] we have

$$\|\hat{e}_z\|_{W_h^{1,2}(\Omega_j)} \lesssim h^k \|\hat{g}_z\|_{W^{k+1,2}(\Omega'_j)} + d_j^{-1} \|\hat{e}_z\|_{L^2(\Omega'_j)},$$

and Green function estimates show that $\|\hat{g}_z\|_{W^{k+1,2}(\Omega_j')} \lesssim d_j^{-n/2-k}$. Applying these estimates into (2.3.32) yield

$$\begin{aligned}\hat{F}_h &\lesssim h^{k-s}\sum_{j=0}^{J} d_j^{n(1/q-1/2)+s-k} + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|\hat{e}_z\|_{L^2(\Omega_j')}\\ &= h^{-n/p}\Theta(k-s+n/p) + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|\hat{e}_z\|_{L^2(\Omega_j')}\\ &\lesssim |\log h|^{\bar{\bar{s}}(p)}h^{-n/p} + h^{-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|\hat{e}_z\|_{L^2(\Omega_j')}.\end{aligned}$$

Applying [14, (5.39)], we have

$$\|\hat{e}_z\|_{L^2(\Omega_j')} \lesssim h^k d_j^{1-k-n/2}\|\hat{e}_z\|_{W_h^{1,1}(\Omega)} + h\|\hat{e}_z\|_{W_h^{1,2}(\Omega_j'')},$$

and therefore

$$\begin{aligned}\hat{F}_h &\lesssim |\log h|^{\bar{\bar{s}}(p)}h^{-n/p} + h^{-n/p}\Theta(k-s+n/p)\|\hat{e}_z\|_{W_h^{1,1}(\Omega)} + h^{1-s}\sum_{j=0}^{J} d_j^{n/q+s-1}\|\hat{e}_z\|_{W_h^{1,2}(\Omega_j'')}\\ &\lesssim |\log h|^{\bar{\bar{s}}(p)}h^{-n/p} + h^{-n/p}\Theta(k-s+n/p)\|\hat{e}_z\|_{W_h^{1,1}(\Omega)} + \frac{\hat{F}_h}{M}.\end{aligned}$$

By taking M sufficiently large, we obtain

$$\hat{F}_h \lesssim |\log h|^{\bar{\bar{s}}(p)}h^{-n/p} + h^{-n/p}\Theta(k-s+n/p)\|\hat{e}_z\|_{W_h^{1,1}(\Omega)},$$

and therefore

$$\|\hat{e}_z\|_{W_h^{1,p'}(\Omega),z,-s} \lesssim |\log h|^{\bar{\bar{s}}(p)}h^{-n/p} + h^{-n/p}\Theta(k-s+n/p)\|\hat{e}_z\|_{W_h^{1,1}(\Omega)}.$$

The case $s=0$, $p'=1$, $p=\infty$ yields

$$\|\hat{e}_z\|_{W_h^{1,1}(\Omega)} \lesssim 1 + \Theta(k)\|\hat{e}_z\|_{W_h^{1,1}(\Omega)},$$

and therefore, we conclude by taking $M > 0$ sufficiently large that

$$\|\hat{e}_z\|_{W_h^{1,1}(\Omega)} \lesssim 1$$

We then conclude that

$$\|\hat{e}_z\|_{W_h^{1,p'}(\Omega),z,-s} \lesssim |\log h|^{\bar{\bar{s}}(p)} h^{-n/p}.$$

The proof is complete. □

Proof of Theorem 2.1 for $p \geq 2$

We now prove Theorem 2.1 in the case $p \in [2, \infty)$. To this end, let $z \in \Omega$ and $T_z \in \mathcal{T}_h$ such that $z \in T_z$. Using an inverse estimate, (2.2.5), and the triangle inequality we obtain

$$\begin{aligned}
|\partial w_h(z)/\partial x_i| &\lesssim h^{-n/2}\|\partial w_h/\partial x_i\|_{L^2(T_z)} \qquad (2.3.33)\\
&\lesssim h^{-n/2-1}\|\partial w_h/\partial x_i\|_{W^{-1,2}(T_z)}\\
&\lesssim h^{-n/2-1}\Big(\|\partial(w-w_h)/\partial x_i\|_{W^{-1,2}(T_z)} + \|\partial w/\partial x_i\|_{W^{-1,2}(T_z)}\Big).
\end{aligned}$$

Note that, by the Poincaré-Friedrichs and Hölder inequalities,

$$\begin{aligned}
\|\partial w/\partial x_i\|_{W^{-1,2}(T_z)} &= \sup_{\substack{\varphi \in C_0^\infty(T_z)\\ \|\varphi\|_{W^{1,2}(T_z)}=1}} (\partial w/\partial x_i, \varphi)_{T_z}\\
&\lesssim \sup_{\substack{\varphi \in C_0^\infty(T_z)\\ \|\varphi\|_{W^{1,2}(T_z)}=1}} |T_z|^{\frac{p-2}{2p}}\|\partial w/\partial x_i\|_{L^p(T_z)}\|\varphi\|_{L^2(T_z)}\\
&\lesssim |T_z|^{\frac{p-2}{2p}}\operatorname{diam}(T_z)\|\partial w/\partial x_i\|_{L^p(T_z)} \lesssim h^{1+n/2-n/p}\|\partial w/\partial x_i\|_{L^p(T_z)}.
\end{aligned}$$

Inserting this estimate into (2.3.33) yields

$$|\partial w_h(z)/\partial x_i| \lesssim h^{-n/p}\|\partial w/\partial x_i\|_{L^p(T_z)} + h^{-n/2-1}\|\partial(w-w_h)/\partial x_i\|_{W^{-1,2}(T_z)}.$$

Replacing w by $w - v_h$ and w_h by $w_h - v_h$ for some $v_h \in V_h$ in the argument above, we conclude

$$\begin{aligned} |\partial(w_h - v_h)(z)/\partial x_i| \lesssim h^{-n/p}&\|\partial(w-v_h)/\partial x_i\|_{L^p(T_z)} \\ &+ h^{-n/2-1}\|\partial(w-w_h)/\partial x_i\|_{W^{-1,2}(T_z)}. \end{aligned} \tag{2.3.34}$$

Let φ, $\hat{g}_z$ and $\hat{g}_{z,h}$ be as in Lemma 2.3. Setting $\hat{e}_z = \hat{g}_z - \hat{g}_{z,h}$, we have for arbitrary $v_h \in V_h$

$$\begin{aligned} h^{-n/2-1}\int_{T_z}(w-w_h)\partial\varphi/\partial x_i\,dx &= a_{0,h}(w-v_h,\hat{e}_z) \\ &\lesssim \|w-v_h\|_{W_h^{1,p}(\Omega),z,s}\|\hat{e}_z\|_{W_h^{1,p'}(\Omega),z,-s} \\ &\lesssim \|w-v_h\|_{W_h^{1,p}(\Omega),z,s}|\log h|^{\bar{\bar{s}}(p)}h^{-n/p}, \end{aligned}$$

where $\bar{\bar{s}}(p)$ is defined in Lemma 2.3.

Applying this last estimate into (2.3.34) yields

$$\begin{aligned} |\nabla(w_h - v_h)(z)| \lesssim h^{-n/p}&\|\nabla(w-v_h)\|_{L^p(T_z)} \\ &+ h^{-n/p}|\log h|^{\bar{\bar{s}}(p)}\|w-v_h\|_{W_h^{1,p}(\Omega),z,s}. \end{aligned} \tag{2.3.35}$$

Raising (2.3.35) by the power p and integrating over Ω with respect to z, we conclude

$$\begin{aligned} \|\nabla_h(w_h - v_h)\|_{L^p(\Omega)} \lesssim &\left(h^{-n}\int_\Omega \|\nabla_h(w-v_h)\|_{L^p(T_z)}^p\,\mathrm{d}z\right)^{1/p} \\ &+ \left(h^{-n}|\log h|^{\bar{\bar{s}}(p)p}\int_\Omega \|w-v_h\|_{W_h^{1,p}(\Omega),z,s}^p\,\mathrm{d}z\right)^{1/p}. \end{aligned}$$

Next, we choose s such that $n/p < s < k + n/p$. Then $\bar{\bar{s}}(p) = 0$, and by (2.3.12)–(2.3.13)

$$\|\nabla_h(w_h - v_h)\|^p_{L^p(\Omega)} \lesssim \|\nabla_h(w - v_h)\|^p_{L^p(\Omega)} + h^{-p}\|w - v_h\|^p_{L^p(\Omega)} + h^p\|D^2_h(w - v_h)\|^p_{L^p(\Omega)},$$

and therefore by the triangle inequality, and by taking $v_h = I_h w$, the nodal interpolant of w,

$$\|\nabla_h(w - w_h)\|_{L^p(\Omega)} \lesssim h\|w\|_{W^{2,p}(\Omega)}. \tag{2.3.36}$$

Next we bound the jumps $\|[w - w_h]\|_{L^p(e)}$. First, by the trace inequalities stated in Lemma 2.1 we have

$$\sum_{e\in\mathcal{E}_h} h_e^{1-p}\|[w - w_h]\|^p_{L^p(e)} \lesssim C\Big(\|\nabla_h(w - w_h)\|^p_{L^p(\Omega)} + h^{-p}\|w - w_h\|^p_{L^p(\Omega)}\Big). \tag{2.3.37}$$

By Lemma 2.2 we have for any $z \in \overline{\Omega}$ and $v_h \in V_h$,

$$|(w - w_h)(z)|^p \lesssim h^{p-n}|\log h|^{p\bar{s}(p)}\|w - v_h\|^p_{W^{1,p}_h(\Omega),z,s},$$

where $\bar{s}(p) = 1$ if $k = s + 1 - n/p$ and $\bar{s}(p) = 0$ for $k > s + 1 - n/p$. Integrating this expression with respect to z yields

$$\|w - w_h\|^p_{L^p(\Omega)} \lesssim h^{p-n}|\log h|^{p\bar{s}(p)} \int_\Omega \|w - v_h\|^p_{W^{1,p}_h(\Omega),z,s}\,\mathrm{d}z. \tag{2.3.38}$$

If $k = 1$, then we set $s = n/p$, so that $\bar{s}(p) = 1$, and by (2.3.14) with $v_h = I_h w$,

$$\begin{aligned}\|w - w_h\|^p_{L^p(\Omega)} &\lesssim h^{p-n}|\log h|^p \int_\Omega \|w - v_h\|^p_{W^{1,p}_h(\Omega),z,n/p}\,\mathrm{d}z \\ &\lesssim h^{2p}|\log h|^{p+1}\|w\|^p_{W^{2,p}(\Omega)}.\end{aligned} \tag{2.3.39}$$

On the other hand, if $k \geq 2$, then we choose s such that $n/p < s < k-1+n/p$. Then $\bar{s}(p) = 0$, and by (2.3.38) and (2.3.13),

$$\|w - w_h\|^p_{L^p(\Omega)} \lesssim h^{2p} \|w\|^p_{W^{2,p}(\Omega)}. \tag{2.3.40}$$

Combining (2.3.37) with (2.3.36), (2.3.39) and (2.3.40) then yields,

$$\sum_{e \in \mathcal{E}_h} h_e^{1-p} \|[w - w_h]\|^p_{L^p(e)} \lesssim |\log h|^{p+1} h^p \|w\|^p_{W^{2,p}(\Omega)}. \tag{2.3.41}$$

Finally combining (2.3.36), (2.3.41) and applying standard scaling arguments yields (2.3.8). This completes the proof of Theorem 2.1 in the case $p \geq 2$.

Proof of Theorem 2.1 for $1 < p < 2$

The proof of $W^{1,p}$ error estimates in the range $p \in (1,2)$ is based on the following result.

Lemma 2.4. *There holds, for $p' \in [2, \infty)$,*

$$\|v_h\|_{W^{1,p'}_h(\Omega)} \lesssim |\log h|^{t'} \sup_{0 \neq z_h \in V_h} \frac{a_{0,h}(v_h, z_h)}{\|z_h\|_{W^{1,p}_h(\Omega)}} \qquad \forall v_h \in V_h,$$

where $p \in (1,2]$ satisfies $1/p + 1/p' = 1$ and $t' = (p'+1)/p'$ if $k = 1$ and $t' = 0$ for $k \geq 2$.

Proof. For a fixed $v_h \in V_h$, let $v \in H^1_0(\Omega)$ satisfy $\mathcal{L}_0 v = \mathcal{L}_{0,h} v_h$ in Ω. Then $v \in W^{2,p'}(\Omega)$ with

$$\|v\|_{W^{2,p'}(\Omega)} \lesssim \|\mathcal{L}_{0,h} v_h\|_{L^{p'}(\Omega)}. \tag{2.3.42}$$

Moreover, due to the definition of $\mathcal{L}_{0,h}$ and the consistency of $a_{0,h}(\cdot,\cdot)$, we find that

$$a_{0,h}(v, z_h) = a_{0,h}(v_h, z_h) \qquad \forall z_h \in V_h.$$

Since $p' \geq 2$, we can apply the results of the previous section to conclude that

$$\|v_h\|_{W_h^{1,p'}(\Omega)} \lesssim |\log h|^{t'}\big(\|v\|_{W^{1,p'}(\Omega)} + h\|v\|_{W^{2,p'}(\Omega)}\big). \tag{2.3.43}$$

Denote by $\mathcal{P}_h : L^2(\Omega) \to V_h$ the L^2 projection onto V_h. We then write

$$\|v\|_{W^{1,p'}(\Omega)} \lesssim \sup_{z\in W_0^{1,p}(\Omega)} \frac{(A_0\nabla v, \nabla z)}{\|z\|_{W^{1,p}(\Omega)}} = \sup_{z\in W_0^{1,p}(\Omega)} \frac{(\mathcal{L}_0 v, z)}{\|z\|_{W^{1,p}(\Omega)}} = \sup_{z\in W_0^{1,p}(\Omega)} \frac{(\mathcal{L}_{0,h} v_h, \mathcal{P}_h z)}{\|z\|_{W^{1,p}(\Omega)}}.$$

Standard arguments show that $\|\mathcal{P}_h z\|_{W_h^{1,p}(\Omega)} \lesssim \|z\|_{W^{1,p}(\Omega)}$ for all $z \in W^{1,p}(\Omega)$; thus,

$$\|v\|_{W^{1,p'}(\Omega)} \lesssim \sup_{0\neq z_h\in V_h} \frac{(\mathcal{L}_{0,h} v_h, z_h)}{\|z_h\|_{W_h^{1,p}(\Omega)}} = \sup_{0\neq z_h\in V_h} \frac{a_{0,h}(v_h, z_h)}{\|z_h\|_{W_h^{1,p}(\Omega)}}. \tag{2.3.44}$$

Likewise, using (2.3.42), (1.6.7) and an inverse estimate yields

$$\|v\|_{W^{2,p'}(\Omega)} \lesssim \|\mathcal{L}_{0,h} v_h\|_{L_h^{p'}(\Omega)} = \sup_{0\neq z_h\in V_h} \frac{a_{0,h}(v_h, z_h)}{\|z_h\|_{L^p(\Omega)}} \lesssim h^{-1} \sup_{0\neq z_h\in V_h} \frac{a_{0,h}(v_h, z_h)}{\|z_h\|_{W^{1,p}(\Omega)}}. \tag{2.3.45}$$

Applying the estimates (2.3.44)–(2.3.45) to (2.3.43) then gives the desired result. □

We now prove Theorem 2.1 for $1 < p < 2$. To this end, for $w_h \in V_h$ and $w \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ satisfying (2.3.6), let $v_h \in V_h$ be the unique solution to

$$a_{0,h}(v_h, z_h) = \int_\Omega |\nabla_h w_h|^{p-2} \nabla_h w_h \cdot \nabla_h z_h \,\mathrm{d}x + \sum_{e\in\mathcal{E}_h} h_e^{1-p} \int_e |[w_h]|^{p-2}[w_h][z_h]\,\mathrm{d}S$$

for all $z_h \in V_h$. Setting $z_h = w_h$ and using a scaling argument yields

$$\|w_h\|^p_{W_h^{1,p}(\Omega)} \lesssim a_{0,h}(v_h, w_h).$$

Moreover, Lemma 2.4 and Hölder's inequality gets

$$\|v_h\|_{W_h^{1,p'}(\Omega)} \lesssim |\log h|^{t'} \|w_h\|^{p-1}_{W_h^{1,p}(\Omega)}.$$

Consequently,

$$\begin{aligned}\|w_h\|_{W_h^{1,p}(\Omega)} &= \frac{\|w_h\|^p_{W_h^{1,p}(\Omega)}}{\|w_h\|^{p-1}_{W_h^{1,p}(\Omega)}} \lesssim |\log h|^{t'} \frac{a_{0,h}(v_h, w_h)}{\|v_h\|_{W_h^{1,p'}(\Omega)}} \\ &= |\log h|^{t'} \frac{a_{0,h}(w, v_h)}{\|v_h\|_{W_h^{1,p'}(\Omega)}} \lesssim |\log h|^{t'} \|w\|_{W_h^{1,p}(\Omega)}.\end{aligned}$$

Standard arguments then show that this estimate implies

$$\|w - w_h\|_{W_h^{1,p}(\Omega)} \lesssim |\log h|^{t'} h \|w\|_{W^{2,p}(\Omega)} \quad 1 < p < 2. \tag{2.3.46}$$

This completes the proof of Theorem 2.1 upon noting that $t' = (p'+1)/p' = (2p-1)/p \leq (p+1)/p = t$ for $p \in (1, 2]$.

2.3.2 DG discrete Calderon-Zygmund estimates for PDEs with constant coefficients

The goal of this subsection is to establish a stability result for the operator $\mathcal{L}^\varepsilon_{0,h}$ in the $W_h^{2,p}$-norm, which is a discrete counterpart of (2.3.3). Such an estimate can be regarded as a DG discrete Calderon-Zygmund estimate for $\mathcal{L}^\varepsilon_{0,h}$.

Theorem 2.2. *(i) For $\varepsilon = 1$ and $1 < p < \infty$ we have*

$$\|w_h\|_{W_h^{2,p}(\Omega)} \lesssim |\log h|^t \|\mathcal{L}^\varepsilon_{0,h} w_h\|_{L^p(\Omega)} \qquad \forall w_h \in V_h, \tag{2.3.47}$$

where $t = (p+1)/p$ if $k = 1$ and $t = 0$ if $k \geq 2$.

(ii) (2.3.47) also holds with $t = 0$ for $\varepsilon \in \{1, 0, -1\}$ and $p = 2$.

Proof. (i) We observe that (2.3.47) is equivalent to showing

$$\|(\mathcal{L}^\varepsilon_{0,h})^{-1} \varphi_h\|_{W_h^{2,p}(\Omega)} \lesssim |\log h|^t \|\varphi_h\|_{L^p(\Omega)} \qquad \forall \varphi_h \in V_h. \tag{2.3.48}$$

For any $\varphi_h \in V_h$, let $w := \mathcal{L}_0^{-1}\varphi_h \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ and $w_h := (\mathcal{L}_{0,h}^{\varepsilon})^{-1}\varphi_h \in V_h$. Since $w \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ we have

$$a_{0,h}^{\varepsilon}(w, v_h) = (\varphi_h, v_h) = a_{0,h}^{\varepsilon}(w_h, v_h) \qquad \forall v_h \in V_h.$$

Thus w_h is the IP-DG approximate solution to w. Applying Theorem 2.1 and the elliptic regularity estimate, we obtain

$$\|w - w_h\|_{W_h^{1,p}(\Omega)} \lesssim |\log h|^t h \|w\|_{W^{2,p}(\Omega)} \lesssim |\log h|^t h \|\varphi_h\|_{L^p(\Omega)}. \tag{2.3.49}$$

Moreover, by Lemma 2.4 and the Calderon-Zygmund estimate for $\mathcal{L}_0$ we have

$$\|w\|_{W_h^{2,p}(\Omega)} \le \|w\|_{W^{2,p}(\Omega)} \lesssim \|\varphi_h\|_{L^p(\Omega)}. \tag{2.3.50}$$

Denote by $I_h : C^0(\Omega) \to V_h$ the nodal interpolation operator onto V_h. By finite element interpolation theory [16] we have

$$h^{-1}\|w - I_h w\|_{W_h^{1,p}(\Omega)} + \|w - I_h w\|_{W_h^{2,p}(\Omega)} \lesssim \|w\|_{W^{2,p}(\Omega)}. \tag{2.3.51}$$

Therefore by the triangle inequality, an inverse estimate, Lemma 2.4, (2.3.49), and (2.3.50), we obtain

$$\begin{aligned}
\|w_h\|_{W_h^{2,p}(\Omega)} &\le \|w - I_h w\|_{W_h^{2,p}(\Omega)} + \|I_h w - w_h\|_{W_h^{2,p}(\Omega)} + \|w\|_{W_h^{2,p}(\Omega)} \\
&\lesssim h^{-1}\|I_h w - w_h\|_{W_h^{1,p}(\Omega)} + \|\varphi_h\|_{L^p(\Omega)} \\
&\le h^{-1}\big(\|w - w_h\|_{W_h^{1,p}(\Omega)} + \|w - I_h w\|_{W_h^{1,p}(\Omega)}\big) + \|\varphi_h\|_{L^p(\Omega)} \\
&\lesssim |\log h|^t \|\varphi_h\|_{L^p(\Omega)} = |\log h|^t \|\mathcal{L}_{0,h} w_h\|_{L^p(\Omega)}.
\end{aligned}$$

(ii) The proof of this part is exactly same as that of Part (i), the only difference is that now (2.3.7), instead of (2.3.8), should be called in the proof. □

2.4 IP-DG methods and their convergence analysis

Our primary goal is to develop stable and convergent IP-DG schemes to approximate the $W^{2,p}$ strong solution of (1.2.1). We assume in (1.2.1) that $A \in [C^0(\overline{\Omega})]^{n\times n}$ satisfies (1.3.1) and $f \in L^p(\Omega)$. Given sufficient smoothness of the boundary, we have the existence and uniqueness of a strong solution $u \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ (see Chapter 1, Subsection 1.3.1 for details). Moreover, we have the following Calderon-Zygmund stability estimate for $\mathcal{L}$:

$$\|u\|_{W^{2,p}(\Omega)} \lesssim \|f\|_{L^p(\Omega)}. \tag{2.4.1}$$

We now turn our attention to the development and analysis of IP-DG methods for continuous A.

2.4.1 Formulation of IP-DG methods

We follow the same recipe as in the constant coefficient case to build our IP-DG methods. To this end, we momentarily assume $A \in [C^1(\Omega)]^{n\times n}$, so that we can rewrite the PDE (2.1.1a) in divergence form as follows:

$$-\nabla \cdot (A\nabla u) + \mathrm{div}(A) \cdot \nabla u = f, \tag{2.4.2}$$

where $\mathrm{div}(A)$ is defined row-wise. We then define the following (standard) IP-DG methods for problem (2.4.2) by seeking $u_h \in V_h$ such that

$$\begin{aligned}\int_\Omega (A\nabla_h u_h)\cdot\nabla_h v_h \,\mathrm{d}x + \int_\Omega ((\nabla\cdot A)\cdot\nabla_h u_h)v_h\,\mathrm{d}x \qquad\qquad (2.4.3)\\ - \sum_{e\in\mathcal{E}_h}\int_e \{A\nabla u_h\cdot\nu_e\}[v_h]\,\mathrm{d}S - \varepsilon\sum_{e\in\mathcal{E}_h}\int_e\{A\nabla v_h\cdot\nu_e\}[u_h]\,\mathrm{d}S\end{aligned}$$

$$+\sum_{e\in\mathcal{E}_h}\int_e \frac{\gamma_e}{h_e}[u_h][v_h]\,\mathrm{d}S = \int_\Omega f v_h\,\mathrm{d}x,$$

where $\gamma_e \geq \gamma_*(\|A\|_{L^\infty(\Omega)}, \mathcal{T}_h) > 0$. We emphasize that γ^* is independent of the derivatives of A.

Now we come back to the case in hand with $A \in [C^0(\Omega)]^{n\times n}$. Clearly, the term $\mathrm{div}(A)$ does not exist as a function (it is in fact a Radon measure), so the above formulation is not defined for the case we are considering. To overcome this difficulty, our idea is to apply the DG integration by parts formula (2.1.4) to the first term on the left-hand side of (2.4.3), yielding

$$\begin{aligned} a_h(w_h, v_h) :=& -\int_\Omega (A : D_h^2 w_h) v_h\,\mathrm{d}x + \sum_{e\in\mathcal{E}_h^I}\int_e [A\nabla w_h\cdot\nu_e]\{v_h\}\,\mathrm{d}S \\ &- \varepsilon\sum_{e\in\mathcal{E}_h}\int_e \{A\nabla v_h\cdot\nu_e\}[w_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h}\int_e \frac{\gamma_e}{h_e}[w_h][v_h]\,\mathrm{d}S. \end{aligned} \tag{2.4.4}$$

No derivative of A appears in the above new form of $a_h^\varepsilon(\cdot,\cdot)$; thus, it is well-defined on $V_h \times V_h$. This leads to the following definition.

Definition 2.1. *Our IP-DG methods are defined by seeking $u_h \in V_h$ such that*

$$a_h^\varepsilon(u_h, v_h) = (f, v_h) \qquad \forall v_h \in V_h, \quad \varepsilon \in \{1, 0, -1\}. \tag{2.4.5}$$

When $\varepsilon = 1$ we refer to the method as “symmetrically induced” even though the bilinear form is not symmetric. Likewise, $\varepsilon = 0$ and $\varepsilon = -1$ yield an “incompletely induced” and “non-symmetrically induced” methods, respectively.

2.4.2 Stability analysis

As in Section 2.3 we define the IP-DG approximation $\mathcal{L}_h^\varepsilon$ of $\mathcal{L}$ on V_h using the bilinear form $a_h^\varepsilon(\cdot,\cdot)$; precisely, we define $\mathcal{L}_h^\varepsilon : V_h \to V_h$ by

$$\big(\mathcal{L}_h^\varepsilon w_h, v_h\big) := a_h^\varepsilon(w_h, v_h) \qquad \forall w_h, v_h \in V_h. \tag{2.4.6}$$

Since we can extend the domain of $a_h^\varepsilon(\cdot,\cdot)$ to $W^{2,p}(\mathcal{T}_h) \times W^{2,p'}(\mathcal{T}_h)$, then the domain and co-domain of $\mathcal{L}_h$ can be extended to the broken Sobolev spaces $W^{2,p}(\mathcal{T}_h)$ and $(W^{2,p'}(\mathcal{T}_h))^*$ respectively.

The goal of this subsection is to establish a DG discrete Calderon-Zygmund estimate similar to (2.3.47) for the operator $\mathcal{L}_h^\varepsilon$. To achieve this, we appeal to the freezing coefficient technique found in Subsection 1.3.1 along with a covering argument to derive a Gärding-type estimate similar to (1.3.4) for $(\mathcal{L}_h^\varepsilon)^*$, the discrete adjoint of $(\mathcal{L}_h^\varepsilon)^*$. While (1.3.4) can be proven for $\mathcal{L}_h^\varepsilon$, more is needed to obtain the stability of $\mathcal{L}_h^\varepsilon$. Traditionally, to derive the stability of $\mathcal{L}_h^\varepsilon$ from this Gärding-type inequality, one must either have the existence/uniqueness of u_h satisfying (2.4.5) in hand, or use a duality argument. We wish to prove existence and uniqueness using the stability of $\mathcal{L}_h^\varepsilon$, thus leaving us the duality argument as the only choice. However, since the formal adjoint of $\mathcal{L}$ does not have a stability result mirroring (2.4.1) for non-differentiable A, a standard duality argument cannot be used. To remedy this, we seek to prove the stability of the discrete adjoint. Proving the stability of $(\mathcal{L}_h^\varepsilon)^*$ immediately gives us the invertibility of the stiffness matrix generated by $(\mathcal{L}_h^\varepsilon)^*$, which is equivalent to the invertibility of the stiffness matrix generated by $\mathcal{L}_h^\varepsilon$, thus providing us the existence and uniqueness of our IP-DG scheme (2.4.5). We are able to carry out a duality argument to obtain the stability of $(\mathcal{L}_h^\varepsilon)^*$ since the continuous dual problem is exactly (1.2.1) whose Calderon-Zygmund estimate (2.4.1) gives us stability of $\mathcal{L}$.

We now proceed to establish a few auxiliary lemmas which will be needed to show the desired estimate.

Lemma 2.1. *For all $\delta > 0$, there exists $R_\delta > 0$ and $h_\delta > 0$ such that for all $x_0 \in \Omega$ and $A_0 \equiv A(x_0)$*

$$\|(\mathcal{L}_h^\varepsilon - \mathcal{L}_{0,h}^\varepsilon)w\|_{L_h^p(B_{R_\delta}(x_0))} \lesssim \delta \|w\|_{W_h^{2,p}(B_{R_\delta}(x_0))} \quad \forall w \in W^{2,p}(\mathcal{T}_h),\ \forall h \le h_\delta. \tag{2.4.7}$$

Here, $B_{R_\delta}(x_0) := \{x \in \Omega : \ |x - x_0| < R_\delta\}$ denotes the ball with center x_0 and radius R_δ.

Proof. Since A is continuous on $\overline{\Omega}$, then it is uniformly continuous. Therefore, for every $\delta > 0$ there exists $R_\delta > 0$ such that if $x, y \in \Omega$ satisfies $|x - y| < R_\delta$, we have $|A(x) - A(y)| < \delta$. Consequently, for any $x_0 \in \Omega$

$$\|A - A_0\|_{L^\infty(B_{R_\delta})} \leq \delta, \tag{2.4.8}$$

where we have used the shorthand notation $B_{R_\delta} := B_{R_\delta}(x_0)$.

Set $h_\delta = \min\{h_0, \frac{R_\delta}{4}\}$ and let $0 < h < h_\delta$, $w \in W^{2,p}(\mathcal{T}_h)$, and $v_h \in V_h(B_{R_\delta})$. Since $(\mathcal{L}^\varepsilon_{0,h} - \mathcal{L}^\varepsilon_h)w \in W^{2,p}(\mathcal{T}_h)$, it follows from (2.3.5) and (2.4.4) that for every $v_h \in V_h(B_{R_\delta})$ we have

$$\begin{aligned}
\big((\mathcal{L}^\varepsilon_{0,h} - \mathcal{L}^\varepsilon_h)w, v_h\big) &= -\int_{\Omega\cap B_{R_\delta}} ((A_0 - A) : D_h^2 w) v_h \,\mathrm{d}x \\
&\quad + \sum_{e\in\mathcal{E}_h^I} \int_{e\cap\overline{B}_{R_\delta}} [(A_0 - A)\nabla w \cdot \nu_e]\{v_h\} \,\mathrm{d}S - \varepsilon \sum_{e\in\mathcal{E}_h} \int_{e\cap\overline{B}_{R_\delta}} \{(A - A_0)\nabla v_h \cdot \nu_e\}[w_h] \,\mathrm{d}S \\
&\leq \|A - A_0\|_{L^\infty(B_{R_\delta})} \Big(\|D_h^2 w\|_{L^p(\Omega\cap B_{R_\delta})} \|v_h\|_{L^{p'}(\Omega\cap B_{R_\delta})} \\
&\quad + \Big(\sum_{e\in\mathcal{E}_h} h_e^{1-2p} \|[w]\|^p_{L^p(e\cap\overline{B}_{R_\delta}))}\Big)^{\frac{1}{p}} \Big(\sum_{e\in\mathcal{E}_h} h_e h_e^{p'} \|\{\nabla v_h \cdot \nu_e\}\|^{p'}_{L^{p'}(e\cap\overline{B}_{R_\delta}))}\Big)^{\frac{1}{p'}} \\
&\quad + \Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p} \|[\nabla w]\|^p_{L^p(e\cap\overline{B}_{R_\delta})}\Big)^{\frac{1}{p}} \Big(\sum_{e\in\mathcal{E}_h^I} h_e \|\{v_h\}\|^{p'}_{L^{p'}(e\cap\overline{B}_{R_\delta})}\Big)^{\frac{1}{p'}} \Big) \\
&\lesssim \|A - A_0\|_{L^\infty(B_{R_\delta})} \|w\|_{W_h^{2,p}(B_{R_\delta})} \Big(\|v_h\|_{L^{p'}(B_{R_\delta})} + h\|\nabla_h v_h\|_{L^{p'}(B_{R_\delta})} \Big) \\
&\lesssim \delta \|w\|_{W_h^{2,p}(B_{R_\delta})} \|v_h\|_{L^{p'}(B_{R_\delta})} = \delta \|w\|_{W_h^{2,p}(B_{R_\delta})} \|v_h\|_{L^{p'}(B_{R_\delta})}.
\end{aligned}$$

Dividing both sides by $\|v_h\|_{L^{p'}(B_{R_\delta})}$ yields the desired estimate. The proof is complete. □

The next lemma shows that $\mathcal{L}^\varepsilon_h$ is locally a bounded operator on $W^{2,p}(\mathcal{T}_h)$.

Lemma 2.2. *For any $x_0 \in \Omega$ and $R \geq h$, there holds*

$$\|\mathcal{L}^\varepsilon_h w\|_{L_h^p(B_R(x_0))} \lesssim \|w\|_{W_h^{2,p}(B_R(x_0))} \qquad \forall w \in W^{2,p}(\mathcal{T}_h). \tag{2.4.9}$$

Proof. Set $B_R := B_R(x_0)$ and let $v_h \in V_h(B_R)$. For $e \in \mathcal{E}_h$, set $e_R := e \cap \overline{B}_R$. By the trace estimate (2.2.2)), the definition of $\mathcal{L}_h^\varepsilon$, and an inverse inequality we have

$$\begin{aligned}
\big(\mathcal{L}_h^\varepsilon w, v_h\big) &= -\int_{\Omega \cap B_R} (A : D_h^2 w) v_h \,\mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_{e_R} [A \nabla w \cdot \nu_e]\{v_h\} \,\mathrm{d}S \\
&\quad - \varepsilon \sum_{e \in \mathcal{E}_h} \int_{e_R} \{A \nabla v_h \cdot \nu_e\}[w] \,\mathrm{d}S + \sum_{e \in \mathcal{E}_h} \int_{e_R} \frac{\gamma_e}{h_e} [w][v_h] \,\mathrm{d}S \\
&\lesssim \|D^2 w\|_{L^p(\Omega \cap B_R)} \|v_h\|_{L^{p'}(\Omega \cap B_R)} \\
&\quad + \Big(\sum_{e \in \mathcal{E}_h^I} h_e^{1-p} \|[\nabla w]\|_{L^p(e_R)}^p\Big)^{\frac{1}{p}} \Big(\sum_{e \in \mathcal{E}_h^I} h_e \|\{v_h\}\|_{L^p(e_R)}^{p'}\Big)^{\frac{1}{p'}} \\
&\quad + \Big(\sum_{e \in \mathcal{E}_h} \gamma_e^p h_e^{1-2p} \|[w]\|_{L^p(e_R)}^p\Big)^{\frac{1}{p}} \Big(\sum_{e \in \mathcal{E}_h} h_e^{p'} h_e \|\{\nabla_h v_h \cdot \nu_e\}\|_{L^{p'}(e_R)}^{p'}\Big)^{\frac{1}{p'}} \\
&\quad + \Big(\sum_{e \in \mathcal{E}_h} \gamma_e^p h_e^{1-2p} \|[w]\|_{L^p(e_R)}^p\Big)^{\frac{1}{p}} \Big(\sum_{e \in \mathcal{E}_h} h_e \|[v_h]\|_{L^{p'}(e_R)}^{p'}\Big)^{\frac{1}{p'}} \\
&\lesssim \|w\|_{W_h^{2,p}(B_R)} \|v_h\|_{L^{p'}(B_R)}.
\end{aligned}$$

Dividing both sides by $\|v_h\|_{L^{p'}(B_R)}$ yields the desired estimate. □

Our last lemma establishes a left-side inf-sup condition for $\mathcal{L}_h^\varepsilon$. This estimate relies on the formal adjoint operator $\mathcal{L}_h^* := (\mathcal{L}_h^\varepsilon)^*$ and some techniques from [53].

Lemma 2.3. *There exists an $h_0 > 0$ such that for all $h \leq h_0$ and $k \geq 2$ we have*

$$\|v_h\|_{L^{p'}(\Omega)} \lesssim \sup_{0 \neq w_h \in V_h} \frac{(\mathcal{L}_h^\varepsilon w_h, v_h)}{\|w_h\|_{W_h^{2,p}(\Omega)}} \qquad \forall v_h \in V_h, \tag{2.4.10}$$

where $1 < p < \infty$ if $\varepsilon = 1$ and $p = 2$ if $\varepsilon \in \{0, -1\}$.

Proof. Note that (2.4.10) is equivalent to

$$\|v_h\|_{L^{p'}(\Omega)} \lesssim \sup_{0 \neq w_h \in V_h} \frac{(\mathcal{L}_h^\varepsilon w_h, v_h)}{\|w_h\|_{W_h^{2,p}(\Omega)}} = \sup_{0 \neq w_h \in V_h} \frac{(\mathcal{L}_h^* v_h, w_h)}{\|w_h\|_{W_h^{2,p}(\Omega)}} = \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} \tag{2.4.11}$$

for all $v_h \in V_h$. We divide the remaining proof into three steps.

Step 1: Local estimates. Let $x_0 \in \Omega$, $A_0 \equiv A(x_0)$, δ_0, h_{δ_0}, R_{δ_0}, $R_1 := (1/3)R_{\delta_0}$, and $B_1 := B_{R_1}(x_0)$ be as in Lemma 2.1 with $\delta_0 > 0$ to be determined, and set $h \leq h_{\delta_0}$.

By the elliptic regularity of $\mathcal{L}$, for any $v_h \in V_h(B_1)$, there exists $\varphi \in W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ such that $\mathcal{L}\varphi = v_h|v_h|^{p-2}$ in Ω and satisfies the estimate

$$\|\varphi\|_{W^{2,p}(\Omega)} \lesssim \|v_h\|_{L^{p'}(\Omega)}^{p'-1} = \|v_h\|_{L^{p'}(B_1)}^{p'-1}. \tag{2.4.12}$$

Since $\mathcal{L}_h^\varepsilon$ is consistent with $\mathcal{L}$ for any $\varphi_h \in V_h$ we have

$$\begin{aligned}\|v_h\|_{L^{p'}(B_1)} &= \|v_h\|_{L^{p'}(\Omega)} = (\mathcal{L}\varphi, v_h) = (\mathcal{L}_h^\varepsilon\varphi, v_h) \\ &= (\mathcal{L}_h^\varepsilon\varphi_h, v_h) + \big(\mathcal{L}_h^\varepsilon(\varphi - \varphi_h), v_h\big) \\ &= (\mathcal{L}_h^* v_h, \varphi_h) + \big(\mathcal{L}_{0,h}^\varepsilon(\varphi - \varphi_h), v_h\big) + \big((\mathcal{L}_h^\varepsilon - \mathcal{L}_{0,h}^\varepsilon)(\varphi - \varphi_h), v_h\big).\end{aligned} \tag{2.4.13}$$

From the existence-uniqueness of the IP-DG scheme (2.3.4), there exists $\varphi_h \in V_h$ such that

$$\big(\mathcal{L}_{0,h}^\varepsilon(\varphi - \varphi_h), w_h\big) = 0 \quad \forall w_h \in V_h.$$

Combining Galerkin orthogonality, Theorem 2.2, and (2.4.12) gives us the solution estimate

$$\|\varphi_h\|_{W_h^{2,p}(\Omega)} \lesssim \|\mathcal{L}_{0,h}^\varepsilon\varphi_h\|_{L_h^p(\Omega)} = \|\mathcal{L}_{0,h}^\varepsilon\varphi\|_{L_h^p(\Omega)} \lesssim \|\varphi\|_{W^{2,p}(\Omega)} \lesssim \|v_h\|_{L^{p'}(B_1)}^{p'-1}. \tag{2.4.14}$$

Using Lemma 2.1 and (2.4.12)–(2.4.14) we have

$$\begin{aligned}\|v_h\|_{L^{p'}(B_1)}^{p'} &= (\mathcal{L}_h^* v_h, \varphi_h) + ((\mathcal{L}_h^\varepsilon - \mathcal{L}_{0,h}^\varepsilon)(\varphi - \varphi_h), v_h) \\ &\leq \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)}\|\varphi_h\|_{W_h^{2,p}(\Omega)} + \|(\mathcal{L}_h^\varepsilon - \mathcal{L}_{0,h}^\varepsilon)(\varphi - \varphi_h)\|_{L_h^p(B_1)}\|v_h\|_{L^{p'}(B_1)} \\ &\lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)}\|v_h\|_{L^{p'}(B_1)}^{p'-1} + \delta_0\|\varphi - \varphi_h\|_{W_h^{2,p}(B_1)}\|v_h\|_{L^{p'}(B_1)} \\ &\lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)}\|v_h\|_{L^{p'}(B_1)}^{p'-1} + \delta_0\|v_h\|_{L^{p'}(B_1)}^{p'}.\end{aligned}$$

Taking δ_0 sufficiently small to move the right hand term to the left side and dividing by $\|v_h\|_{L^{p'}(B_1)}^{p'-1}$ gives us the local estimate

$$\|v_h\|_{L^{p'}(B_1)} \lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p}(B_1)} \qquad \forall v_h \in V_h(B_1). \tag{2.4.15}$$

Step 2: A Gärding type inequality by a covering argument. Given R_1 from Step 1, let $R_2 = 2R_1$ and $R_3 = 3R_1$. Let $\eta \in C^3(\Omega)$ be a cutoff function satisfying

$$0 \le \eta \le 1, \quad \eta\big|_{B_1} = 1, \quad \eta\big|_{\Omega\setminus B_2} = 0, \quad |\eta|_{W^{m,\infty}(\Omega)} = O(R_1^{-m}). \tag{2.4.16}$$

For any $v_h \in V_h$, we have by (2.4.15),

$$\begin{aligned}\|v_h\|_{L^{p'}(B_1)} &= \|\eta v_h\|_{L^{p'}(B_1)} \le \|\eta v_h - I_h(\eta v_h)\|_{L^{p'}(B_1)} + \|I_h(\eta v_h)\|_{L^{p'}(B_1)} \\ &\lesssim \|\eta v_h - I_h(\eta v_h)\|_{L^{p'}(B_1)} + \|\mathcal{L}_h^*(I_h(\eta v_h))\|_{W_h^{-2,p'}(B_1)} \\ &\lesssim \|\eta v_h - I_h(\eta v_h)\|_{L^{p'}(B_1)} + \|\mathcal{L}_h^*(I_h(\eta v_h) - \eta v_h)\|_{W_h^{-2,p'}(B_1)} + \|\mathcal{L}_h^*(\eta v_h)\|_{W_h^{-2,p'}(B_1)}.\end{aligned} \tag{2.4.17}$$

We now bound the second term on the right hand side of (2.4.17). By the definition of $\|\cdot\|_{W_h^{-2,p}}$, Lemma 2.2 and (1.6.7), for any $w_h \in V_h$ we have

$$\begin{aligned}\|\mathcal{L}_h^*(I_h(\eta v_h) - \eta v_h)\|_{W_h^{-2,p'}(B_1)} &= \sup_{0\ne w_h\in V_h} \frac{(\mathcal{L}_h^*(I_h(\eta v_h) - \eta v_h), w_h)}{\|w_h\|_{W_h^{2,p}(B_1)}} \\ \le \sup_{w_h\in V_h} \frac{(\mathcal{L}_h^\varepsilon w_h, I_h(\eta v_h) - \eta v_h)}{\|w_h\|_{W_h^{2,p}(B_1)}} &\lesssim \sup_{w_h\in V_h} \frac{\|\mathcal{L}_h^\varepsilon w_h\|_{L_h^p(B_1)} \|I_h(\eta v_h) - \eta v_h)\|_{L^{p'}(B_1)}}{\|w_h\|_{W_h^{2,p}(B_1)}} \\ \lesssim \sup_{w_h\in V_h} \frac{\|w_h\|_{W_h^{2,p}(B_1)} \|I_h(\eta v_h) - \eta v_h)\|_{L^{p'}(B_1)}}{\|w_h\|_{W_h^{2,p}(B_1)}} &= \|I_h(\eta v_h) - \eta v_h\|_{L^{p'}(B_1)}.\end{aligned}$$

Thus (2.4.17) becomes

$$\|v_h\|_{L^{p'}(B_1)} \lesssim \|\eta v_h - I_h(\eta v_h)\|_{L^{p'}(B_1)} + \|\mathcal{L}_h^*(\eta v_h)\|_{W_h^{-2,p'}(B_1)}. \tag{2.4.18}$$

Using Lemmas 2.3, 2.7, and 2.2 with (2.4.18) yields

$$\begin{aligned} \|v_h\|_{L^{p'}(B_1)} &\lesssim \frac{h}{R_1}\|v_h\|_{L^{p'}(B_3)} + \|\mathcal{L}_h^*(\eta v_h)\|_{W_h^{-2,p'}(B_3)} \\ &\lesssim \frac{1}{R_1}\|v_h\|_{W^{-1,p'}(B_3)} + \|\mathcal{L}_h^*(\eta v_h)\|_{W_h^{-2,p'}(B_3)}. \end{aligned} \tag{2.4.19}$$

We now want to remove the cutoff function η from the adjoint operator appearing in the right-hand side of (2.4.19). For $w_h \in V_h(B_3)$, we break up $\mathcal{L}_h^*(\eta v_h)$ as follows:

$$\begin{aligned} (\mathcal{L}_h^*(\eta v_h), w_h) = (\mathcal{L}_h^\varepsilon w_h, \eta v_h) = (\mathcal{L}_h^\varepsilon w_h \eta, v_h) + \Big[(\mathcal{L}_h^\varepsilon w_h, \eta v_h) - (\mathcal{L}_h^\varepsilon w_h \eta, v_h)\Big] \\ = (\mathcal{L}_h^\varepsilon(I_h(w_h\eta)), v_h) + (\mathcal{L}_h^\varepsilon(w_h\eta - I_h(w_h\eta)), v_h) \\ + \Big[(\mathcal{L}_h^\varepsilon w_h, \eta v_h) - (\mathcal{L}_h^\varepsilon w_h \eta, v_h)\Big] \\ =: I_1 + I_2 + I_3. \end{aligned} \tag{2.4.20}$$

We then seek to bound each I in order. To bound I_1, we will use the definition of $\|\cdot\|_{W_h^{-2,p}}$, the stability of I_h, and Lemma 2.5 to obtain

$$\begin{aligned} I_1 = (\mathcal{L}_h^* v_h, I_h(w_h\eta)) &\lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)}\|I_h(\eta w_h)\|_{W_h^{2,p}(B_3)} \\ &\lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)}\|\eta w_h\|_{W_h^{2,p}(B_3)} \lesssim \frac{1}{R_1^2}\|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)}\|w_h\|_{W_h^{2,p}(B_3)}. \end{aligned} \tag{2.4.21}$$

For I_2 we use Lemmas 2.3, 2.7, 2.2 to get

$$\begin{aligned} I_2 = (\mathcal{L}_h^\varepsilon(w_h\eta - I_h(w_h\eta)), v_h) &\lesssim \|w_h\eta - I_h(w_h\eta)\|_{W_h^{2,p}(B_3)}\|v_h\|_{L^{p'}(B_3)} \\ &\lesssim \frac{h}{R_1^3}\|w_h\|_{W_h^{2,p}(B_3)}\|v_h\|_{L^{p'}(B_3)} \lesssim \frac{1}{R_1^3}\|w_h\|_{W_h^{2,p}(B_3)}\|v_h\|_{W^{-1,p'}(B_3)}. \end{aligned} \tag{2.4.22}$$

To bound I_3 we introduce the operator $\mathcal{L}_{0,h}^\varepsilon$. For $e \in \mathcal{E}_h$ let $e_3 := e \cap B_3$, and define $\tilde{A} := A - A_0$. We then write

$$I_3 = (\mathcal{L}_h^\varepsilon w_h, \eta v_h) - (\mathcal{L}_h^\varepsilon w_h \eta, v_h) \tag{2.4.23}$$

$$
\begin{aligned}
&= (\mathcal{L}_{0,h}^{\varepsilon} w_h, \eta v_h) - (\mathcal{L}_{0,h}^{\varepsilon} w_h \eta, v_h) + \\
&\qquad + \big[(\mathcal{L}_{h}^{\varepsilon} w_h, \eta v_h) - (\mathcal{L}_{h}^{\varepsilon} w_h \eta, v_h) - (\mathcal{L}_{0,h}^{\varepsilon} w_h, \eta v_h) + (\mathcal{L}_{0,h}^{\varepsilon} w_h \eta, v_h)\big] \\
&= -\int_{B_3} \left(w_h A_0 : D^2\eta + (A_0 + A_0^T)\nabla\eta \cdot \nabla_h w_h\right) v_h \,\mathrm{d}x \\
&\quad - \varepsilon \sum_{e\in\mathcal{E}_h} \int_{e_3} (A_0 \nabla\eta \cdot \nu_e)\{v_h\}[w_h] \,\mathrm{d}S \\
&\quad - \int_{B_3} \left(w_h(\tilde{A} : D^2\eta + \left(\tilde{A} + \tilde{A}^T\right)\nabla\eta \cdot \nabla_h w_h\right) v_h \,\mathrm{d}x \\
&\quad - \varepsilon \sum_{e\in\mathcal{E}_h} \int_{e_3} (\tilde{A} \nabla\eta \cdot \nu_e)\{v_h\}[w_h] \,\mathrm{d}S =: K_1 + K_2 + K_3 + K_4.
\end{aligned}
$$

We now must bound each K_i. To bound K_1 we use the definition of $\|\cdot\|_{W_h^{-1,p}(B_3)}$ and Lemma 2.3 to get

$$
\begin{aligned}
K_1 &\lesssim \left(\|w_h A_0 : D^2\eta\|_{W_h^{1,p}(B_3)} + \|(A_0 + A_0^T)\nabla\eta \cdot \nabla_h w_h\|_{W_h^{1,p}(B_3)}\right) \|v_h\|_{W_h^{-1,p'}(B_3)} \qquad (2.4.24)\\
&\lesssim \frac{1}{R_1^3} \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{W_h^{-1,p'}(B_3)}.
\end{aligned}
$$

The bound of K_2 uses Lemmas 2.1, 2.5 to obtain

$$
\begin{aligned}
K_2 &\lesssim \frac{1}{R_1} \left(\sum_{e\in\mathcal{E}_h} h_e^{1-2p} \|[w_h]\|_{L^p(e_3)}^p\right)^{\frac{1}{p}} \left(\sum_{e\in\mathcal{E}_h} h_e h_e^{p'} \|\{v_h\}\|_{L^{p'}(e_3)}^{p'}\right)^{\frac{1}{p'}} \qquad (2.4.25)\\
&\lesssim \frac{1}{R_1} \|w_h\|_{W_h^{2,p}(B_3)} \left(h\|v_h\|_{L^{p'}(B_3)}\right) \\
&\lesssim \frac{1}{R_1} \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{W_h^{-1,p'}(B_3)}.
\end{aligned}
$$

We use similar techniques as (2.4.24), (2.4.25) and the fact that $\|\tilde{A}\|_{L^\infty(B_3)} \le \delta_0$ to get

$$
\begin{aligned}
K_3 &\lesssim \left(\|w_h \tilde{A} : D^2\eta\|_{L^p(B_3)} + \|(\tilde{A} + \tilde{A}^T)\nabla\eta \cdot \nabla_h w_h\|_{L^p(B_3)}\right) \|v_h\|_{L^{p'}(B_3)} \qquad (2.4.26)\\
&\lesssim \delta_0 \left(\frac{1}{R_1^2} \|w_h\|_{L^p(B_3)} + \frac{1}{R_1} \|w_h\|_{W_h^{1,p}(B_3)}\right) \|v_h\|_{L^{p'}(B_3)}
\end{aligned}
$$

$$\lesssim \delta_0 \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{L^{p'}(B_3)},$$

where we have used Lemma 2.5 to derive the last inequality. Likewise, we find

$$\begin{aligned} K_4 &\lesssim \frac{1}{R_1} \left(\sum_{e \in \mathcal{E}_h} h_e^{1-2p} \|[w_h]\|_{L^p(e_3)}^p \right)^{\frac{1}{p}} \left(\sum_{e \in \mathcal{E}_h} h_e h_e^{p'} \|\{v_h\}\|_{L^{p'}(e_3)}^{p'} \right)^{\frac{1}{p'}} \tag{2.4.27} \\ &\lesssim \frac{\delta_0}{R_1} \|w_h\|_{W_h^{2,p}(B_3)} \left(h \|v_h\|_{L^{p'}(B_3)} \right) \\ &\lesssim \delta_0 \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{L^{p'}(B_3)}, \end{aligned}$$

where we have used the inequality $h \leq R_1$. Combining (2.4.23)-(2.4.27) we get

$$I_3 \lesssim \frac{1}{R_1^3} \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{W_h^{-1,p'}(B_3)} + \delta_0 \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{L^{p'}(B_3)}, \tag{2.4.28}$$

and bringing together (2.4.20)-(2.4.22), and (2.4.28) gives us

$$\begin{aligned} (\mathcal{L}_h^*(\eta v_h), w_h) \lesssim \frac{1}{R_1^3} &\left(\|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)} + \|v_h\|_{W_h^{-1,p'}(B_3)} \right) \|w_h\|_{W_h^{2,p}(B_3)} \tag{2.4.29} \\ &+ \delta_0 \|w_h\|_{W_h^{2,p}(B_3)} \|v_h\|_{L^{p'}(B_3)}. \tag{2.4.30} \end{aligned}$$

By the definition of $\| \cdot \|_{W_h^{-2,p'}(B_3)}$ and (2.4.29) we get

$$\|\mathcal{L}_h^*(\eta w_h)\|_{W_h^{-2,p'}(B_3)} \lesssim \frac{1}{R_1^3} \left(\|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)} + \|v_h\|_{W_h^{-1,p'}(B_3)} \right) + \delta_0 \|v_h\|_{L^{p'}(B_3)}. \tag{2.4.31}$$

Using (2.4.19) and (2.4.31) gives us

$$\|v_h\|_{L^{p'}(B_1)} \lesssim \frac{1}{R_1^3} \left(\|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(B_3)} + \|v_h\|_{W_h^{-1,p'}(B_3)} \right) + \delta_0 \|v_h\|_{L^{p'}(B_3)}.$$

Since $\overline{\Omega}$ is compact, employing a covering argument (cf. [22, 35]) then yields

$$\|v_h\|_{L^{p'}(\Omega)} \lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} + \|v_h\|_{W_h^{-1,p'}(\Omega)} + \delta_0 \|v_h\|_{L^{p'}(\Omega)}.$$

Because δ_0 is small, we can absorb the last term on the right-hand side to the left-hand side to arrive at the global estimate

$$\|v_h\|_{L^{p'}(\Omega)} \lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} + \|v_h\|_{W_h^{-1,p'}(\Omega)}, \tag{2.4.32}$$

which is a Gärding-type inequality.

Step 3: Duality argument on the adjoint operator. To control the last term in (2.4.32) we now use a duality argument for $\mathcal{L}_h^*$. This argument uses the regularity estimate of the original problem $\mathcal{L}$.

Define the set

$$X = \{g \in W_h^{1,p}(\Omega);\ \|g\|_{W_h^{1,p}(\Omega)} = 1\}.$$

By the discrete Poincaré inequality, with constant $C = C(p, \Omega)$, we have for all $g \in X$

$$\|g\|_{L^p(\Omega)} \leq C\|g\|_{W_h^{1,p}(\Omega)} < \infty,$$

since X is bounded in $W_h^{1,p}(\Omega)$. Thus, X is precompact in $L^p(\Omega)$ by Sobolev embedding. Next we define the set

$$W = \{\varphi := \mathcal{L}^{-1} g;\ g \in X\}.$$

Note that $\mathcal{L}^{-1} : L^p(\Omega) \to W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega) \subset W^{2,p}(\mathcal{T}_h)$ is well defined by well-posedness of the PDE. Also since $\mathcal{L}^{-1}$ is linear and satisfies the estimate

$$\|\mathcal{L}^{-1} g\|_{W_h^{2,p}(\Omega)} = \|\varphi\|_{W_h^{2,p}(\Omega)} \leq \|\varphi\|_{W^{2,p}(\Omega)} \lesssim \|g\|_{L^p(\Omega)},$$

it is bounded in $W^{2,p}(\mathcal{T}_h)$. Thus W is precompact in $W^{2,p}(\mathcal{T}_h)$. From [53, Lemma 5], for every $\tau > 0$ there exists $h_* > 0$ that only depends on τ and $\overline{W}$ such that for each

$\varphi \in W$ and $0 < h \leq h_*$ there is a $\varphi_h \in V_h$ such that if $k \geq 2$ we have

$$\|\varphi - \varphi_h\|_{W_h^{2,p}(\Omega)} \leq \tau. \tag{2.4.33}$$

Note by the reverse triangle inequality and (2.4.33) we have

$$\|\varphi_h\|_{W_h^{2,p}(\Omega)} \leq \|\varphi\|_{W_h^{2,p}(\Omega)} \lesssim \|g\|_{L^p(\Omega)} \leq C$$

and hence

$$\{\varphi_h \in V_h;\, |\varphi_h - \varphi| \leq \tau\}$$

is uniformly bounded in φ and h. Let $g \in X$ and choose $\varphi_g = \mathcal{L}^{-1} g \in W$ which tells us that $\mathcal{L}\varphi_g = g$. Let $v_h \in V_h$ and $\varphi_h \in V_h$. By Lemma 2.2 and the definition of $\|\cdot\|_{W_h^{-2,p'}(\Omega)}$ we have

$$\begin{aligned}\int_\Omega v_h g \,\mathrm{d}x &= (\mathcal{L}_h^\varepsilon \varphi_g, v_h) = (\mathcal{L}_h^\varepsilon \varphi_h, v_h) + (\mathcal{L}_h^\varepsilon(\varphi_g - \varphi_h), v_h) \\ &= (\mathcal{L}_h^* v_h, \varphi_h) + (\mathcal{L}_h^\varepsilon(\varphi_g - \varphi_h), v_h) \\ &\lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} \|\varphi_h\|_{W_h^{2,p}(\Omega)} + \|\varphi_g - \varphi_h\|_{W_h^{2,p}(\Omega)} \|v_h\|_{L^{p'}(\Omega)}.\end{aligned}$$

Selecting φ_h to satisfy (2.4.33) and taking the supremum on g gives us

$$\|v_h\|_{W^{-1,p}(\Omega)} \lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} \|\varphi_h\|_{W_h^{2,p}(\Omega)} + \tau \|v_h\|_{L^{p'}(\Omega)}. \tag{2.4.34}$$

Combining (2.4.32) and (2.4.34) yields

$$\|v_h\|_{L^{p'}(\Omega)} \lesssim \|\mathcal{L}_h^* v_h\|_{W_h^{-2,p'}(\Omega)} + \tau \|v_h\|_{L^{p'}(\Omega)}. \tag{2.4.35}$$

By choosing τ sufficiently small to kick back the right-most term we have (2.4.11). This completes the proof upon taking $h_0 = \min\{h_{\delta_0}, h_*\}$. □

We are now ready to prove the global stability of the operator $\mathcal{L}_h^\varepsilon$.

Theorem 2.3. *Suppose that $h \leq h_0$ and $k \geq 2$. Then there holds the following stability estimate:*

$$\|w_h\|_{W_h^{2,p}(\Omega)} \lesssim \|\mathcal{L}_h^\varepsilon w_h\|_{L_h^p(\Omega)} \qquad \forall w_h \in V_h, \tag{2.4.36}$$

where $1 < p < \infty$ if $\varepsilon = 1$, and $p = 2$ if $\varepsilon \in \{0, -1\}$.

Proof. Let $w_h \in V_h$ be fixed, and consider the auxiliary problem of finding $q_h \in V_h$ such that

$$\begin{aligned}\big(v_h, \mathcal{L}_h^* q_h\big) = \big(\mathcal{L}_h^\varepsilon v_h, q_h\big) = \int_\Omega |D_h^2 w_h|^{p-2} D_h^2 w_h : D^2 v_h \,\mathrm{d}x \\ + \sum_{e\in\mathcal{E}_h^I} h_e^{1-p} \int_e |[\nabla w_h]|^{p-2}[\nabla w_h]\cdot[\nabla v_h]\,\mathrm{d}S \\ + \sum_{e\in\mathcal{E}_h} h_e^{1-2p} \int_e |[w_h]|^{p-2}[w_h][v_h]\,\mathrm{d}S \qquad \forall v_h \in V_h.\end{aligned} \tag{2.4.37}$$

Since V_h is finite dimensional and the operator is linear, the existence is equivalent to the uniqueness. To show the uniqueness, let $q_h^{(1)}$ and $q_h^{(2)}$ both solve (2.4.37). Then by Lemma 2.3 we get

$$\|q_h^{(1)} - q_h^{(2)}\|_{L^{p'}(\Omega)} \lesssim \sup_{0\neq v_h\in V_h} \frac{(\mathcal{L}_h^\varepsilon v_h, q_h^{(1)} - q_h^{(2)})}{\|v_h\|_{W_h^{2,p}(\Omega)}} = 0.$$

Hence (2.4.37) has a unique solution $q_h \in V_h$. Also by Lemma 2.3 and Hölder's inequality,

$$\|q_h\|_{L^{p'}(\Omega)} \lesssim \sup_{0\neq v_h\in V_h} \frac{(\mathcal{L}_h^\varepsilon v_h, q_h)}{\|v_h\|_{W_h^{2,p}(\Omega)}} \lesssim \|w_h\|_{W_h^{2,p}(\Omega)}^{p-1}.$$

Consequently, we find

$$\|w_h\|_{W_h^{2,p}(\Omega)}^p \lesssim (w_h, \mathcal{L}_h^* q_h) = (\mathcal{L}_h w_h, q_h) \lesssim \|\mathcal{L}_h w_h\|_{L_h^p(\Omega)} \|q_h\|_{L^{p'}(\Omega)}$$

$$\lesssim \|\mathcal{L}_h w_h\|_{L_h^p(\Omega)} \|w_h\|_{W_h^{2,p}(\Omega)}^{p-1}.$$

Dividing by $\|w_h\|_{W_h^{2,p}(\Omega)}^{p-1}$ now yields the desired result. □

2.4.3 Well-posedness and error estimates

The goals of this subsection are to establish the well-posedness for the IP-DG scheme (2.4.5) and to derive the optimal order error estimates in $W_h^{2,p}$-norm for the IP-DG solutions.

Theorem 2.4. *Under the assumptions of Lemma 2.3, the IP-DG scheme (2.4.5) has a unique solution $u_h \in V_h$ such that*

$$\|u_h\|_{W_h^{2,p}(\Omega)} \lesssim \|f\|_{L^p(\Omega)}, \tag{2.4.38}$$

where the hidden constant depends on the dimension n, exponent p, the maximum penalty parameter $\max_{e\in\mathcal{E}_h} \gamma_e$, and the modulus of continuity of A.

Proof. Since (2.4.5) is equivalent to a linear system, hence it suffices to prove the uniqueness. To show the uniqueness, we first prove (2.4.38).

Let $u_h \in V_h$ be a solution of (2.4.5), then from (2.4.36) and the definition of $\|\cdot\|_{L_h^p(\Omega)}$ we have

$$\|u_h\|_{W_h^{2,p}(\Omega)} \lesssim \|\mathcal{L}_h^\varepsilon u_h\|_{L_h^p(\Omega)} = \sup_{v_h\in V_h} \frac{(\mathcal{L}_h^\varepsilon u_h, v_h)}{\|v_h\|_{L^{p'}(\Omega)}} = \sup_{v_h\in V_h} \frac{(f, v_h)}{\|v_h\|_{L^{p'}(\Omega)}} \le \|f\|_{L^p(\Omega)}.$$

Hence, (2.4.38) holds.

Suppose that $u_h^1, u_h^2 \in V_h$ solve (2.4.5). Let $\tilde{u}_h = u_h^1 - u_h^2$. Then by (2.4.38) we have

$$\|\tilde{u}_h\|_{W_h^{2,p}(\Omega)} \le \|0\|_{L^p(\Omega)} = 0.$$

Since $\tilde{u}_h \in V_h$ with $\|\tilde{u}_h\|_{W_h^{2,p}(\Omega)} = 0$ we conclude that $\tilde{u}_h \in C^1(\Omega)$, $\tilde{u}_h\big|_{\partial\Omega} = 0$, and $D_h^2\tilde{u}_h = 0$ in Ω. The only way this can happen is if $\tilde{u}_h = 0$. Thus, the IP-DG solution must be unique. The proof is complete. □

Next we show a Céa-type lemma for the IP-DG scheme, which immediately deduces the optimal order error estimates in the $W_h^{2,p}$-norm.

Theorem 2.5. *Suppose that $h \leq h_0$ and $k \geq 2$. Let $u \in W^{2,p} \cap W_0^{1,p}(\Omega)$ be the solution of problem (1.2.1) and $u_h \in V_h$ solve (2.4.5). Then*

$$\|u - u_h\|_{W_h^{2,p}(\Omega)} \lesssim \inf_{w_h \in V_h} \|u - w_h\|_{W_h^{2,p}(\Omega)}, \tag{2.4.39}$$

where the hidden constant depends on the same parameters as those given in Theorem 2.4. Moreover, if $u \in W^{s,p}(\Omega)$ for some $s \geq 2$, we have

$$\|u - u_h\|_{W_h^{2,p}(\Omega)} \lesssim h^{r-2}\|u\|_{W^{r,p}(\Omega)}, \qquad r = \min\{s, k+1\}. \tag{2.4.40}$$

Proof. By the consistency of $\mathcal{L}_h^\varepsilon$ we have the following Galerkin orthogonality:

$$\big(\mathcal{L}_h^\varepsilon(u - u_h), v_h\big) = 0 \qquad \forall v_h \in V_h. \tag{2.4.41}$$

Let $w_h \in V_h$, by Theorem 2.3, Lemma 2.2, (2.4.41), and the definition of $\|\cdot\|_{L_h^p(\Omega)}$ we have

$$\begin{aligned} \|u_h - w_h\|_{W_h^{2,p}(\Omega)} &\lesssim \|\mathcal{L}_h^\varepsilon(u_h - w_h)\|_{L_h^p(\Omega)} = \sup_{0 \neq v_h \in V_h} \frac{(\mathcal{L}_h^\varepsilon(u_h - w_h), v_h)}{\|v_h\|_{L^{p'}(\Omega)}} \\ &= \sup_{0 \neq v_h \in V_h} \frac{(\mathcal{L}_h^\varepsilon(u - w_h), v_h)}{\|v_h\|_{L^{p'}(\Omega)}} = \|\mathcal{L}_h^\varepsilon(u - w_h)\|_{L_h^p(\Omega)} \\ &\lesssim \|u - w_h\|_{W_h^{2,p}(\Omega)}. \end{aligned} \tag{2.4.42}$$

Thus by (2.4.42) and the triangle inequality we get

$$\|u - u_h\|_{W_h^{2,p}(\Omega)} \leq \|u - w_h\|_{W_h^{2,p}(\Omega)} + \|u_h - w_h\|_{W_h^{2,p}(\Omega)} \lesssim \|u - w_h\|_{W_h^{2,p}(\Omega)}. \tag{2.4.43}$$

Taking the infimum on both sides over all $w_h \in V_h$ yields (2.4.39). Finally, (2.4.40) follows from taking $w_h = I_h u$ and using the finite element interpolation theory [8]. The proof is complete. □

2.5 Numerical Experiments

In this section we present a number of 2D numerical tests to verify our error estimate and to gauge the performance of our IP-DG methods. In particular, we shall compare our IP-DG methods to the related conforming finite element counterpart developed in [22]. Moreover, we shall also perform numerical tests which are not covered by our convergence theory; this includes the cases when the coefficient matrix is either discontinuous or degenerate.

2.5.1 Hölder continuous coefficient

For this test we take A as the following Hölder continuous matrix-valued function:

$$A(x) = \begin{bmatrix} |x|^{1/2} + 1 & -|x|^{1/2} \\ -|x|^{1/2} & 5|x|^{1/2} + 1 \end{bmatrix}, \qquad x \in \mathbb{R}^2.$$

Let $\Omega = (-1/2, 1/2)^2$ and choose f such that the exact solution is given by

$$u(x_1, x_2) = \sin(2\pi x_1)\sin(2\pi x_2)\exp(x_1\cos(x_2)),$$

which has zero trace on the boundary.

Figure 2.1 shows the errors in the $L^2(\Omega), W_h^{1,2}(\Omega)$, and $W_h^{2,2}(\Omega)$ norms of both the symmetrically and incompletely induced methods. The convergence rates observed for the symmetrically induced method are

$$\|u - u_h\|_{L^2(\Omega)} = \mathcal{O}(h^{k+1}) \text{ for all } k,$$

$$\|\nabla_h(u - u_h)\|_{L^2(\Omega)} = \mathcal{O}(h^{k}) \text{ for all } k,$$

$$\|D_h^2(u-u_h)\|_{L^2(\Omega)} = \mathcal{O}(h^{k-1}) \text{ for } k = 2, 3.$$

As expected, these convergence rates are optimal. However, for the incompletely induced method we find that the rate of convergence in the L^2-norm is sub-optimal for even degree polynomials and optimal with all other norms and degrees. This should be expected since the incomplete scheme is sub-optimal even for smooth A [49].

2.5.2 Uniformly continuous coefficients

In this test we take $\Omega = (0, 1/2)^2$ and let

$$A(x) = \begin{bmatrix} -\dfrac{5}{\log(|x|)} + 15 & 1 \\ 1 & -\dfrac{1}{\log(|x|)} + 3 \end{bmatrix}.$$

f is chosen such that $u(x) = |x|^{7/4}$ is the exact solution. From [22] we see that the expected convergence rates are

$$\begin{aligned} \|\nabla_h(u-u_h)\|_{L^2(\Omega)} &= \mathcal{O}(h^{\min\{k,7/4-\delta\}}) \text{ for all } k, \\ \|D_h^2(u-u_h)\|_{L^2(\Omega)} &= \mathcal{O}(h^{\min\{k,7/4-\delta\}-1}) \text{ for } k = 2, 3 \end{aligned}$$

for any $\delta > 0$.

Figure 2.2 gives the computed results for both the symmetrically and incompletely induced schemes which match exactly the expected rates of convergence.

2.5.3 Degenerate coefficients

In this test we take $\Omega = (0, 1)^2$ and the matrix

$$A(x) = \frac{16}{9}\begin{bmatrix} x_1^{2/3} & -x_1^{1/3}x_2^{1/3} \\ -x_1^{1/3}x_2^{1/3} & x_2^{2/3} \end{bmatrix}.$$

$f = 0$ and the exact solution $u(x) = x_1^{4/3} - x_2^{4/3}$. For an explanation for this example we refer to [22]. Note that $\det(A) = 0$ for every $x \in \Omega$ so this PDE is degenerate everywhere and is outside of the strong solution theory. We also observe that $u \in W^{m,p}(\Omega)$ provided $(4 - 3m)p > -1$.

Figure 2.3 shows the L^2 and piecewise H^1 errors for both the symmetrically and incompletely induced methods. The numerical results suggest the following rates of convergence:

$$\begin{aligned} \|u - u_h\|_{L^2(\Omega)} &= \mathcal{O}(h^{4/3}), \\ \|\nabla_h(u - u_h)\|_{L^2(\Omega)} &= \mathcal{O}(h^{5/6}) \end{aligned}$$

for $k = 1, 2, 3$. These rates are consistent with the results of the related conforming finite element method given in [22].

2.5.4 L^∞ Cordès coefficients

Our next test is taken from [55, 56] where a different DG method and a weak Galerkin method were used to solve this problem. Let $\Omega = [-1, 1]^2$ and

$$A(x) = \frac{16}{9}\begin{bmatrix} 2 & x_1x_2/|x_1x_2| \\ x_1x_2/|x_1x_2| & 2 \end{bmatrix}.$$

f is chosen so that the exact solution is $u(x) = x_1x_2\big(1 - e^{1-|x_1|}\big)\big(1 - e^{1-|x_2|}\big)$. Notice that the matrix A is discontinuous across the x_1-axis and x_2-axis, and it satisfies the Cordès condition. While our convergence theory does not apply to this example, we still compute the numerical solution on a uniform triangulation that has edges on all discontinuities of A. Due to its inconsistent behavior we list the L^2 error and convergence rates in Table 2.1. The following H^1 semi-norm rates are observed:

$$\|\nabla_h(u - u_h)\|_{L^2(\Omega)} = \mathcal{O}(h^k)$$

for $k = 1, 2, 3$ as shown in Figure 2.4.

Table 2.1: The L^2 errors and rates for the symmetrically induced method. The rates for the incompletely induced method are similar. $\gamma_e \equiv 10000$ is used as the penalty parameter.

	$k = 1$		$k = 2$		$k = 3$	
h	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|u - u_h\|_{L^2(\Omega)}$	rate
1	1.3e-1	-	7.7e-2	-	2.6e-2	-
1/2	8.9e-2	0.58	1.8e-2	2.09	1.5e-3	4.12
1/4	4.6e-2	0.95	2.9e-3	2.62	7.6e-4	4.27
1/8	1.9e-2	1.22	4.8e-2	2.62	4.2e-6	4.19
1/16	7.6e-3	1.35	8.0e-5	2.57	3.3e-7	3.65
1/32	2.9e-3	1.41	1.4e-5	2.54	3.2e-8	3.36

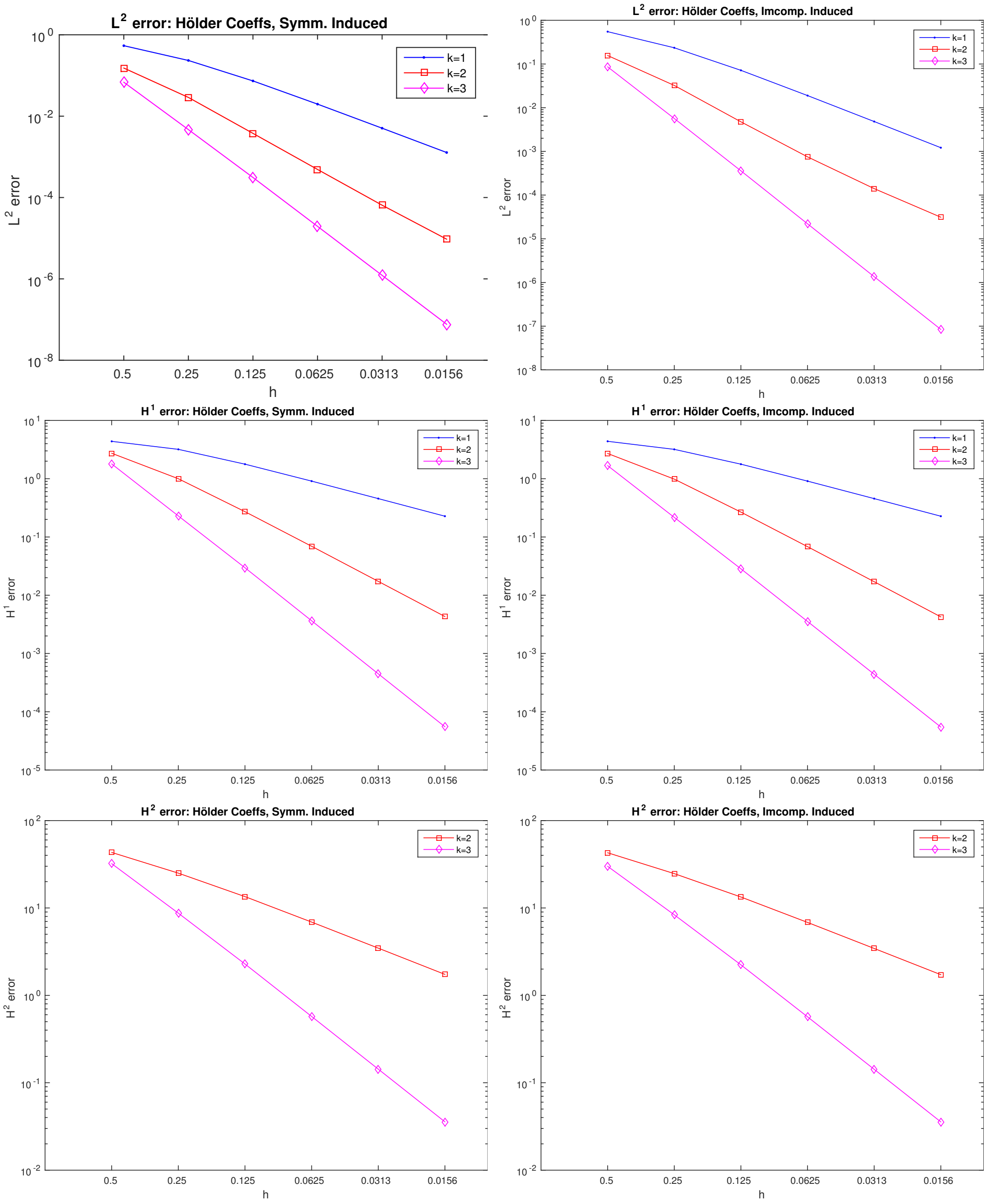


Figure 2.1: The L^2 (top), piecewise H^1 (middle), and piecewise H^2 (bottom) errors for both the symmetrically (left) and incompletely (right) induced schemes with polynomial degree $k = 1, 2, 3$. $\gamma_e \equiv 100$ is used as the penalty parameter.

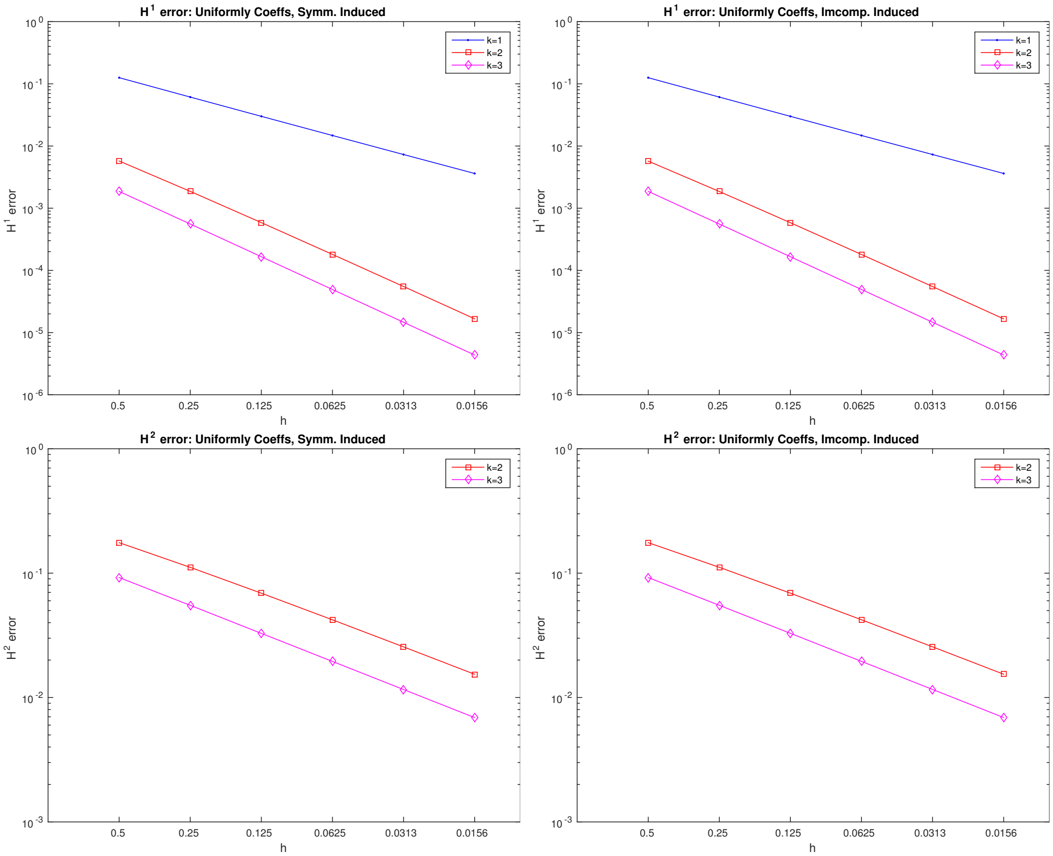


Figure 2.2: The piecewise H^1 (top) and piecewise H^2 (bottom) errors for both the symmetrically (left) and incompletely (right) induced schemes with polynomial degree $k = 1, 2, 3$. $\gamma_e \equiv 1000$ is used as the penalty parameter.

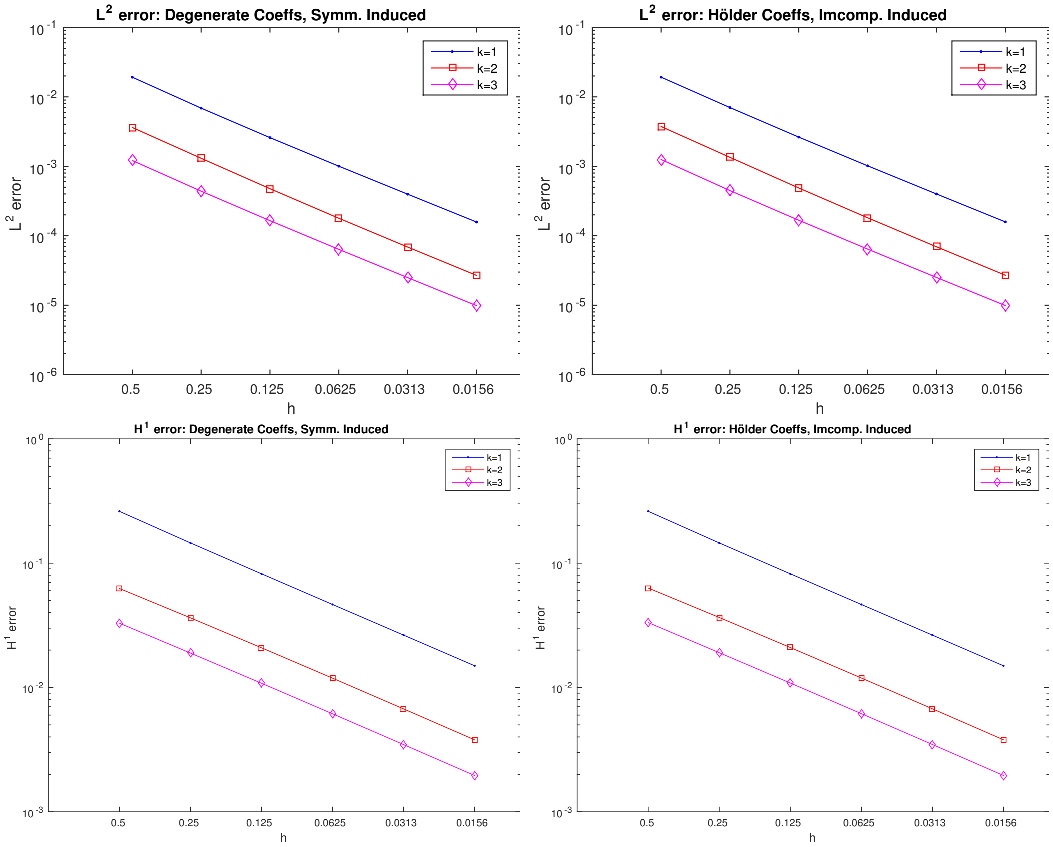


Figure 2.3: The L^2 (top) and piecewise H^1 (bottom) errors for both the symmetrically (left) and incompletely (right) induced schemes with polynomial degree $k = 1, 2, 3$. $\gamma_e \equiv 100$ is used as the penalty parameter.

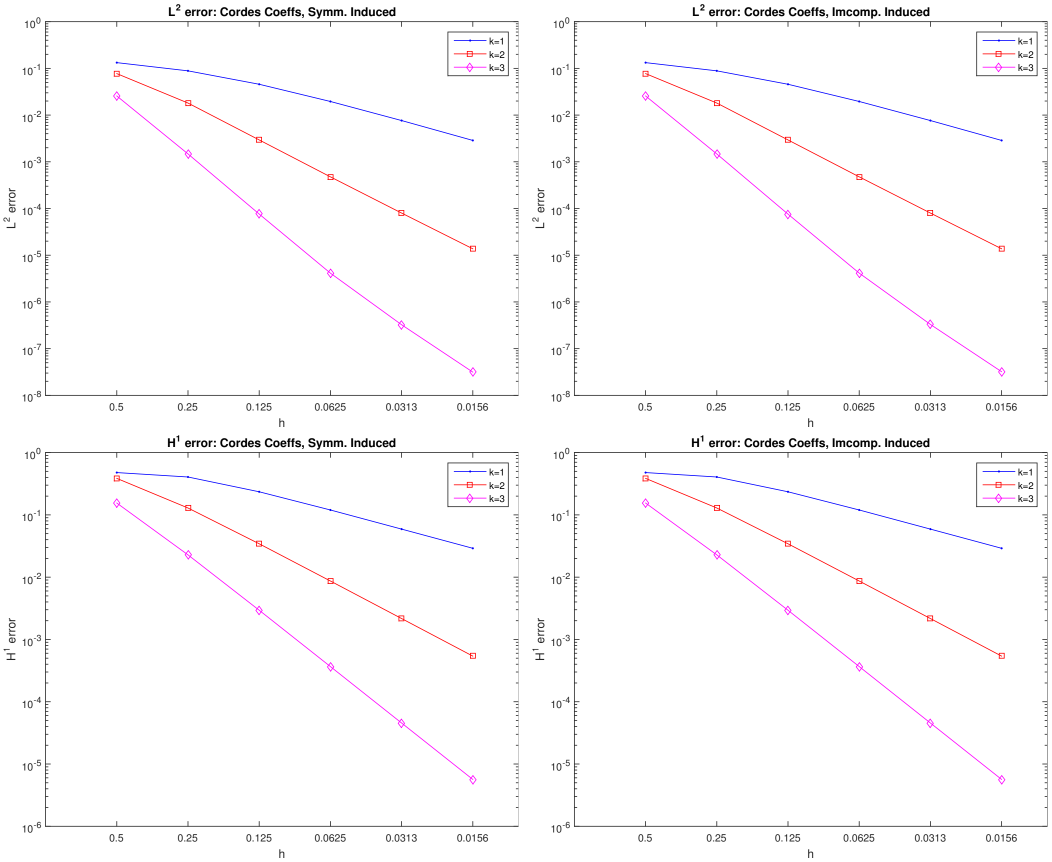


Figure 2.4: The L^2 (top) and piecewise H^1 (bottom) errors for both the symmetrically (left) and incompletely (right) induced schemes with polynomial degree $k = 1, 2, 3$. $\gamma_e \equiv 10000$ is used as the penalty parameter.

Chapter 3

The Vanishing Moment Method for Second Order Linear Elliptic Non-divergence Form PDEs

3.1 Introduction

In Chapter 2, we introduced an interior-penalty discontinuous Galerkin method for the following non-divergence form second order linear elliptic PDE:

$$\begin{aligned} \mathcal{L}u := -A : D^2 u = f \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{aligned} \tag{P}$$

where $A \in \left[C(\overline{\Omega})\right]^{d\times d}$ is uniformly positive definite and $f \in L^p(\Omega)$. While this IP-DG method is accurate, gives optimal error estimates in the discrete $W^{2,p}$ norm, and may even converge when A is a coefficient matrix outside of the strong solution theory, the treatment of the non-divergence form is quite delicate. Indeed, every numerical method for these non-divergence problems reviewed in Subsection 1.4.1 requires special handling of the non-divergence term.

In an attempt to bypass these delicate discretizations to the non-divergence term, we approximate the strong solution u to (P) on the PDE level via the solution u^ε to an approximate problem where u^ε solves the following fourth order problem:

$$\begin{aligned} \mathcal{L}^\varepsilon u^\varepsilon := \varepsilon\Delta^2 u^\varepsilon - A : D^2 u^\varepsilon = f \text{ in } \Omega, \\ u^\varepsilon = 0 \text{ on } \partial\Omega, \\ \Delta u^\varepsilon = 0 \text{ on } \partial\Omega. \end{aligned} \tag{P_ε}$$

Here $\varepsilon > 0$ is small. We call this process the *vanishing moment method* (VMM). Since the non-divergence operator is not the highest order term in (P_ε), numerical methods to approximate the solution u^ε to (P_ε) will trivially discretize the non-divergence operator while giving special attention to the discretization of the biharmonic operator, which has been extensively studied in the literature for finite element, discontinuous Galerkin, and finite difference methods. We note that the boundary condition for simply supported plates $\Delta u^\varepsilon = 0$ is purely artificial and is only needed for the well-posedness of (P_ε). Other suitable choices for the essential boundary condition are $\nabla\Delta u^\varepsilon \cdot \nu = 0$ or $A : D^2 u^\varepsilon = 0$ where ν is the unit outward normal vector of Ω.

To give a motivation of the formulation of this method, we recall the *vanishing viscosity method* - the first order analog of the vanishing moment method. Consider the following first order fully nonlinear stationary Hamilton-Jacobi Equation:

$$\begin{aligned} H(\nabla u, u, x) = 0 \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \end{aligned} \tag{H}$$

where H is the Hamiltonian. Due to the nonlinearity of the highest order derivative, well-posedness of (H) is not trivial; moreover, in which sense one defines a solution to (H) is not obvious. In 1983, Crandall and Lions (see [18]) proved the existence of solutions to (H) by considering solutions to the following second order approximate problem:

$$\begin{aligned} -\varepsilon \Delta u + H^\varepsilon(\nabla u, u, x) &= 0 \text{ in } \Omega, \\ u &= z^\varepsilon \text{ on } \partial\Omega \end{aligned} \tag{H_ε}$$

where $H^\varepsilon \to H$ and $z^\varepsilon \to 0$ uniformly. Since (H_ε) is now a quasi-linear problem, that is, the PDE operator is linear in the highest order derivative, a notion of weak solutions u^ε to (H_ε) can be defined. Moreover, if $u^\varepsilon \in W^{2,p}_{\text{loc}}(\Omega)$, $p > d$ converges to $u^* \in C(\overline{\Omega})$ uniformly, then u^* is a viscosity solution to (H). This method is called the vanishing viscosity method since ε is the viscosity coefficient if u^ε represents the velocity of a fluid in a fluid dynamics problem.

The vanishing moment method was first proposed by Feng and Neilan [27], where they used the technique to approximate fully nonlinear second order equations such as the Monge-Ampère equation (1.2.3). Various numerical methods have been developed subsequently (see [27, 38, 26]); however, convergence of the VMM is only proved in special cases (see [24]). The goal of this chapter is to present the convergence of the vanishing moment method for non-divergence form second order linear elliptic PDEs. Specifically, we show the solutions u^ε of (P_ε) converge to $u \in H^2(\Omega) \cap H^1_0(\Omega)$ where u is the strong solution to (P). In addition, we derive error estimates for $\|u^\varepsilon - u\|$ in powers of ε in various norms. To motivate the need for error estimates, if one were to discretize (P_ε) and produce an approximate solution u^ε_h, then a straight forward approach to show $u^\varepsilon_h \to u$ would be to use the triangle inequality to show

$$\|u^\varepsilon_h - u\| \leq \|u^\varepsilon_h - u^\varepsilon\| + \|u^\varepsilon - u\|. \tag{3.1.3}$$

Note that the error $\|u^\varepsilon_h - u^\varepsilon\|$ is purely the discretization error from the numerical method. However, the error $\|u^\varepsilon - u\|$ is the PDE approximation error which is independent of the numerical method used. Thus in order to obtain sharp rates of convergence for $\|u - u^\varepsilon_h\|$, we must have sharp error estimates for $\|u^\varepsilon - u\|$ in terms of powers of ε.

Since the highest order derivative in (P_ε) is in divergence form, we can easily define a concept of weak solutions for (P_ε).

Definition 3.1. *Let $\varepsilon > 0$. We say that $u^\varepsilon \in H^2(\Omega) \cap H_0^1(\Omega)$ is a weak solution to (P_ε) provided*

$$\varepsilon(\Delta u^\varepsilon, \Delta v) - (A : D^2 u^\varepsilon, v) = (f, v) \tag{3.1.4}$$

for all $v \in H^2(\Omega) \cap H_0^1(\Omega)$.

We assume that for any $f \in L^2(\Omega)$, and ε sufficiently small, there exists a unique weak solution $u^\varepsilon \in H^2(\Omega) \cap H_0^1(\Omega)$ to (P_ε). Moreover, we assume that we can increase the regularity of the weak solution such that to $u^\varepsilon \in H$, where

$$H := \{v \in H^2(\Omega) \cap H_0^1(\Omega) : \Delta v \in H_0^1(\Omega)\}.$$

However, we do not assume any stability estimate for $\mathcal{L}^\varepsilon$.

To prove the convergence of u^ε, we obtain an H^1 and H^2 uniform in ε stability estimates which will give us weak compactness in $H^2(\Omega) \cap H_0^1(\Omega)$. To derive estimates we will utilize the freezing coefficient technique mentioned in Subsection 1.3.2. The H^1 and H^2 uniform stability estimates for $\mathcal{L}^\varepsilon$ when A is a constant coefficient matrix is proven in Section 3.2. Then in Section 3.3, we extend theses estimates to $\mathcal{L}^\varepsilon$ for continuous A. In Section 3.4, we give the proof of convergence for $u^\varepsilon \to u$ where u is the strong solution of (P) as well as error estimates for $\|u^\varepsilon - u\|$ in the H^1 and L^2 norm. To test the effectiveness of the vanishing moment method, we formulate a C^0 interior penalty method for (P_ε) in Section 3.5 and apply the method to several test examples from Section 2.5. In Section 3.6, we test a hybrid method combinting the vanishing moment method with the IP-DG method for non-divergence form PDEs from Chapter 2 on several Hamilton-Jacobi-Bellman examples.

3.2 Stability Estimates for Constant Coefficient Operators

In this section, we consider the case when $A = A_0$ is a constant matrix that is uniformly positive definite. This leads to the following problem:

$$\begin{aligned} \mathcal{L}_0^\varepsilon u^\varepsilon := \varepsilon \Delta^2 u^\varepsilon - A_0 : D^2 u^\varepsilon = f \text{ in } \Omega, \\ u = 0 \text{ on } \partial\Omega, \\ \Delta u = 0 \text{ on } \partial\Omega. \end{aligned} \tag{P_ε^0}$$

Since $A_0 : D^2 u = \operatorname{div}(A_0 \nabla u)$, we can define a standard weak solution $u^\varepsilon \in H^2(\Omega) \cap H_0^1(\Omega)$ to (P_ε^0). Our first result of this section is the following local H^1 estimate for $\mathcal{L}_0^\varepsilon$.

Lemma 3.1. *Let $B \subset \Omega$ be open and let $v \in H^2(B) \cap H_0^1(B)$. Then we have the following estimate:*

$$\sqrt{\varepsilon}\|\Delta v\|_{L^2(B)} + \sqrt{\lambda}\|\nabla v\|_{L^2(B)} \lesssim \|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B)}. \tag{3.2.2}$$

Note that while we have control on Δv for fixed ε, we lose this control as $\varepsilon \to 0$. Thus, we refer to (3.2.2) only as an H^1 estimate. Also the analysis of this method is reliant on having control of the H^{-1} norm of $\mathcal{L}_0^\varepsilon v$.

Proof. Testing (P_ε^0) by v, integrating by parts, using the ellipticity condition (1.3.1), and using Poincaré's inequality gives us

$$\begin{aligned} \varepsilon\|\Delta v\|_{L^2(B)}^2 + \lambda\|\nabla v\|_{L^2(B)}^2 &\leq \varepsilon(\Delta v, \Delta v)_B + (A_0 \nabla v, \nabla v)_B \\ &= (\mathcal{L}_0^\varepsilon v, v)_B \\ &\leq \|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B)} \|v\|_{H^1(B)} \end{aligned}$$

$$
\begin{aligned}
&\lesssim \|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B)}\|\nabla v\|_{L^2(B)}\\
&\leq \frac{1}{\delta}\|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B)} + \delta\|\nabla v\|_{L^2(B)}.
\end{aligned}
$$

Choosing δ, only dependent on λ and the constant from the Poincaré inequality, sufficiently small allows us to move $\|\nabla u^\varepsilon\|_{L^2(B)}$ on the right side to the left and obtain (3.2.2). The proof is complete.

□

While we have a uniform H^1 estimate, uniform control on the Hessian D^2u^ε is required in order to show convergence to the strong solution u. To this end, we next establish a uniform H^2 stability for our constant coefficient operator $\mathcal{L}_0^\varepsilon$.

Lemma 3.2. *Let $B \subset \Omega$ and let $v \in H$ with* $\mathrm{supp}(v) \subset B$*, then the following estimate holds:*

$$
\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(B)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B)} \lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B)}. \tag{3.2.3}
$$

Proof. Testing $\mathcal{L}_0^\varepsilon v$ by $-\Delta v$ and integrating by parts we get

$$
(\mathcal{L}_0^\varepsilon v, -\Delta v) = (\varepsilon\Delta^2 v - A_0 : D^2 v, -\Delta v) = \varepsilon\|\nabla\Delta v\|_{L^2(B)}^2 + (A_0 : D^2 v, \Delta v). \tag{3.2.4}
$$

Since A_0 is symmetric and positive definite. There exists an orthogonal matrix $Q \in \mathbb{R}^{n\times n}$ such that $Q^TAQ = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n) =: \Lambda$ where $\lambda \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Let $y = Q^Tx$ and $\hat{v}(y) = v(Qy) = v(x)$. Since the Laplacian is preserved under orthogonal change of basis we have the following:

$$
\Delta_x v(x) = \Delta_y \hat{v}(y), \qquad \Delta_x^2 v(x) = \Delta_y^2 \hat{v}(y),
$$

$$
A_0 : D_x^2 v(x) = \Lambda : D_y^2 \hat{v}(y) = \sum_{j=1}^n \lambda_j \hat{v}_{y_j y_j}(y).
$$

WLOG we may assume that $A_0 = \Lambda$ in (3.2.4). Hence

$$\begin{aligned}
(A_0 : D^2 v, \Delta v) &= (\operatorname{div}(A_0 \nabla v), \Delta v) \\
&= -(A_0 \nabla v, \nabla \Delta v) \\
&= -(A_0 \nabla v, \operatorname{div}(D^2 v)) \\
&= (\nabla(A_0 \nabla v), D^2 v) \\
&= \sum_{j=1}^{n} \lambda_j \|\nabla v_{x_i}\|^2_{L^2(B)} \geq \lambda \|D^2 v\|^2_{L^2(B)}.
\end{aligned} \tag{3.2.5}$$

Combining (3.2.4) and (3.2.5) gives us

$$\begin{aligned}
\varepsilon \|\nabla \Delta v\|^2_{L^2(B)} + \lambda \|D^2 v\|^2_{L^2(B)} &\leq \|\mathcal{L}_0^\varepsilon v\|_{L^2(B)} \|\Delta v\|_{L^2(B)} \\
&\leq \frac{\delta}{2} \|\Delta v\|^2_{L^2(B)} + \frac{1}{2\delta} \|\mathcal{L}_0^\varepsilon v\|^2_{L^2(B)} \\
&\leq \frac{\delta}{2} \|D^2 v\|^2_{L^2(B)} + \frac{1}{2\delta} \|\mathcal{L}_0^\varepsilon v\|^2_{L^2(B)}.
\end{aligned}$$

Choosing δ sufficiently small, dependent only on the ellipticity condition, to move $\|\mathcal{L}_0^\varepsilon v\|^2_{L^2(B)}$ to the right hand side gives the desired result. The proof is complete. □

Next, we derive similar boundary estimates. Let $B^+ = B \cap \mathbb{R}^d_+ := \{x = (x', x_d) \in \mathbb{R}^d : x_d > 0\}$ and $(\partial B^+)^+ = \partial(B^+) \cap \mathbb{R}^d_+$ where B is a small ball with it's center on the x_d-axis.

Lemma 3.3. *Let $v \in H^2(B+)$ with $\Delta v \in H^1(B+)$ and $v, \nabla v, \Delta v = 0$ near ∂B^+ and $v = \Delta v = 0$ on $\partial B^+ \setminus (\partial B^+)^+$. Then we have the following estimates:*

$$\sqrt{\varepsilon} \|\nabla \Delta v\|_{L^2(B^+)} + \lambda \|D^2 v\|_{L^2(B^+)} \lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B^+)}, \tag{3.2.6}$$

$$\sqrt{\varepsilon} \|\Delta v\|_{L^2(B^+)} + \lambda \|\nabla v\|_{L^2(B^+)} \lesssim \|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B^+)}. \tag{3.2.7}$$

Proof. We will extend v from B^+ to B by an odd reflection, that is $v(x', x_n) = -v(x', -x_n)$ for all $x \in B \setminus B^+$. Since v and Δv are both zero on $\partial B \cap \{x_n = 0\}$, we have $v \in H_0^2(B)$ and $\Delta v \in H_0^1(B)$ by the odd extension. Again we test the PDE by Δv and use a similar argument as to Lemma 3.2. This gives us

$$\sqrt{\varepsilon}\|\nabla \Delta v\|_{L^2(B)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B)} \lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B)}. \tag{3.2.8}$$

Since the odd reflection is a bounded linear operator on L^2, independent of the diameter of B, and $B^+ \subset B$, we have

$$\begin{aligned}\sqrt{\varepsilon}\|\nabla \Delta v\|_{L^2(B^+)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B^+)} &\leq \sqrt{\varepsilon}\|\nabla \Delta v\|_{L^2(B)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B)} \\ &\lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B)} \\ &\lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B^+)}.\end{aligned}$$

which is precisely (3.2.6). We can repeat a similar argument to achieve (3.2.7). The proof is complete.

□

3.3 Uniform Stability Estimates for Variable Coefficient Operators

Let $A \in \left[C(\overline{\Omega})\right]^{d\times d}$ be uniformly positive definite. In this section we seek uniform H^1 and H^2 stability estimates for $\mathcal{L}^\varepsilon$. Following the freezing coefficients technique, we first need to drive local H^1 and H^2 stability estimates, which in turn require the following lemma controlling the bound of the H^{-1} norm of the Hessian.

Lemma 3.4. *Let B be an open ball and $v \in H^2(B)$, then we have the following estimate:*

$$\|D^2 v\|_{H^{-1}(B)} \leq d^2\|\nabla u\|_{L^2(B)}, \tag{3.3.1}$$

where d is the dimension of the domain Ω, and

$$\|D^2 v\|_{H^{-1}(B)} = \left(\sum_{i,j=1}^{d} \|v_{x_i,x_j}\|^2_{H^{-1}(B)} \right)^{\frac{1}{2}}. \tag{3.3.2}$$

Proof. Let $i, j = 1, \ldots, d$ and $w \in H^1_0(B)$ with $w \not\equiv 0$. Integrating by parts gives us

$$(v_{x_i x_j}, w) = (v_{x_i}, w_{x_j}) \leq \|\nabla v\|_{L^2(B)} \|\nabla w\|_{L^2(B)}.$$

Thus by the definition of $\|v\|_{H^{-1}(B)}$ we have

$$\|v_{x_i x_j}\|_{H^{-1}(B)} = \sup_{\substack{w \in H^1_0(B) \\ w \not\equiv 0}} \frac{|(v_{x_i x_j}, w)_B|}{\|\nabla w\|_{L^2(B)}} \leq \|\nabla v\|_{L^2(B)}.$$

Summing over i and j gives us (3.3.1). The proof is complete. □

We are now ready to prove the local H^1 and H^2 stability of $\mathcal{L}^\varepsilon$.

Lemma 3.5. *Let $x_0 \in \Omega$ and $B_R(x_0) \subset \Omega$ to be the ball of radius R centered at x_0. There exists $R_0 > 0$, independent of ε, such that for all $v \in H$ with* $\operatorname{supp}(v) \subset B :=$ *$B_{R_0}(x_0)$, the following estimates hold:*

$$\sqrt{\varepsilon}\|\nabla \Delta v\|_{L^2(B)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B)} \lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(B)}, \tag{3.3.3}$$

$$\sqrt{\varepsilon}\|\Delta v\|_{L^2(B)} + \sqrt{\lambda}\|\nabla v\|_{L^2(B)} \lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(B)}. \tag{3.3.4}$$

Proof. Let $\delta > 0$ and define $A_0 := A(x_0)$. Since A is continuous, there exists $R_0 > 0$ such that

$$\|A - A_0\|_{L^\infty(B_{R_0}(x_0))} \leq \delta.$$

By Lemma 3.2 we have

$$\begin{aligned}\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(B)} + \sqrt{\lambda}\|D^2 v\|_{L^2(B)} &\lesssim \|\mathcal{L}_0^\varepsilon v\|_{L^2(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(B)} + \|(\mathcal{L}_0^\varepsilon - \mathcal{L}^\varepsilon)v\|_{L^2(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(B)} + \|(A - A_0) : D^2 v\|_{L^2(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(B)} + \|A - A_0\|_{L^\infty(B)}\|D^2 v\|_{L^2(B)} \\ &\leq \|\mathcal{L}^\varepsilon v\|_{L^2(B)} + \delta\|D^2 v\|_{L^2(B)}.\end{aligned}$$

Since δ only depends on λ, we can make it sufficiently small to move $\|D^2 v\|_{L^2(B)}$ to the left hand side and arrive at (3.3.3).

To show (3.3.4), we follow a similar technique. Using Lemma 3.1 and Lemma 3.4, we have

$$\begin{aligned}\sqrt{\varepsilon}\|\Delta v\|_{L^2(B)} + \sqrt{\lambda}\|\nabla v\|_{L^2(B)} &\lesssim \|\mathcal{L}_0^\varepsilon v\|_{H^{-1}(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(B)} + \|(\mathcal{L}_0^\varepsilon - \mathcal{L}^\varepsilon)v\|_{H^{-1}(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(B)} + \|(A - A_0) : D^2 v\|_{H^{-1}(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(B)} + \|A - A_0\|_{L^\infty(B)}\|D^2 v\|_{H^{-1}(B)} \\ &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(B)} + \delta\|\nabla v\|_{L^2(B)}.\end{aligned}$$

Thus we may make δ sufficiently small to move the $\|\nabla v\|_{L^2(B)}$ from the left and side and obtain (3.3.4). The proof is complete. □

Finally, using cutoff functions and a covering argument, we obtain some interior Gärding-type inequalities.

Lemma 3.6. *For any $\Omega' \subset\subset \Omega$ and $v \in H$, the following estimates hold:*

$$\begin{aligned}\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega')} + \sqrt{\lambda}\|D^2 v\|_{L^2(\Omega')} &\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)} \\ &\quad + \varepsilon\big(\|\nabla v\|_{L^2(\Omega)} + \|\Delta v\|_{L^2(\Omega)} + \|\nabla\Delta v\|_{L^2(\Omega)}\big), \qquad (3.3.5)\\ \sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega')} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega')} &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(\Omega)} + \|v\|_{L^2(\Omega)} \qquad (3.3.6)\end{aligned}$$

$$+\,\varepsilon\big(\|\nabla v\|_{L^2(\Omega)} + \|\Delta v\|_{L^2(\Omega)}\big).$$

Proof. For a ball B_R with radius R let $\sigma = 1/2$ and choose the cutoff function $\eta \in C_0^\infty(B_R)$ with $0 \le \eta \le 1$, $\eta \equiv 1$ in $B_{\sigma R}$, $\eta \equiv 0$ on $B_R \setminus B_{\sigma' R}$ where $\sigma' = 3/4$. Moreover, $\|D^k\eta\|_{L^\infty(B_R)} \lesssim (1-\sigma)^{-k}R^{-k}$ for $k = 0,1,2,3,4$. Applying (3.3.3) to ηv on the ball $B_{\sigma' R}$ gives us

$$\begin{aligned}\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(B_{\sigma R})}+&\sqrt{\lambda}\|D^2 v\|_{L^2(B_{\sigma R})}\\ &= \sqrt{\varepsilon}\|\nabla\Delta(\eta v)\|_{L^2(B_{\sigma R})} + \sqrt{\lambda}\|D^2(\eta v)\|_{L^2(B_{\sigma R})}\\ &\le \sqrt{\varepsilon}\|\nabla\Delta(\eta v)\|_{L^2(B_{\sigma' R})} + \sqrt{\lambda}\|D^2(\eta v)\|_{L^2(B_{\sigma' R})}\\ &\lesssim \|\mathcal{L}^\varepsilon(\eta v)\|_{L^2(B_{\sigma' R})} = \|\varepsilon\Delta^2(\eta v) - A : D^2(\eta v)\|_{L^2(B_{\sigma' R})}.\end{aligned} \tag{3.3.7}$$

Expanding $\Delta^2(\eta v)$ and $A : D^2(\eta v)$ yields

$$\Delta^2(\eta v) = \eta\Delta^2 v + 4\nabla\Delta v\cdot\nabla\eta + 6\Delta v\Delta\eta + 4\nabla v\cdot\nabla\Delta\eta + v\Delta^2\eta, \tag{3.3.8}$$

$$A : D^2(\eta v) = \eta A : D^2 v + 2A\nabla v\cdot\nabla\eta + vA : D^2\eta. \tag{3.3.9}$$

Using (3.3.8) and (3.3.9) we obtain (below the L^2 norm is taken on $B_{\sigma' R}$),

$$\begin{aligned}\|\varepsilon\Delta^2(\eta v) - A : D^2(\eta v)\|_{L^2} \lesssim \|\mathcal{L}^\varepsilon v\|_{L^2} &+ \frac{1}{(1-\sigma)R}\|\nabla v\|_{L^2} + \frac{1}{(1-\sigma)^2R^2}\|v\|_{L^2}\\ &+ \frac{\varepsilon}{(1-\sigma)^4R^4}\|v\|_{L^2} + \frac{\varepsilon}{(1-\sigma)^3R^3}\|\nabla v\|_{L^2}\\ &+ \frac{\varepsilon}{(1-\sigma)^2R^2}\|\Delta v\|_{L^2} + \frac{\varepsilon}{(1-\sigma)R}\|\nabla\Delta v\|_{L^2}.\end{aligned} \tag{3.3.10}$$

The treatment of the first three terms on the right follows from [35, p.236]. Keeping the rest of the terms on the right and using a covering argument we arrive at (3.3.5). Since $\overline{\Omega'}$ is compact it only takes a finite number of balls to cover Ω', thus the estimate does not depend on R.

To show (3.3.6), using the same η as prescribed above and estimate (3.3.4), we recover a similar estimate to (3.3.7):

$$\begin{aligned}\sqrt{\varepsilon}\|\Delta v\|_{L^2(B_{\sigma R})}+&\sqrt{\lambda}\|\nabla v\|_{L^2(B_{\sigma R})}\\ &\lesssim \|\mathcal{L}^\varepsilon(\eta v)\|_{H^{-1}(B_{\sigma' R})}\\ &= \|\varepsilon\Delta^2(\eta v)-A:D^2(\eta v)\|_{H^{-1}(B_{\sigma' R})}\\ &= \sup_{w\in H_0^1(B_{\sigma' R}))}\frac{(\varepsilon\Delta^2(\eta v)-A:D^2(\eta v),w)}{\|\nabla w\|_{L^2(B_{\sigma' R})}}.\end{aligned}\tag{3.3.11}$$

Let $w\in H_0^1(B_{\sigma' R}))$, by integration by parts we have

$$\begin{aligned}(\varepsilon\Delta^2(\eta v)-A:D^2(\eta v),w) &= -\varepsilon(\nabla\Delta(\eta v),\nabla w)-(A:D^2(\eta v),w)\\ &:= \varepsilon I_1+I_2.\end{aligned}\tag{3.3.12}$$

We first focus on I_2, expanding $\nabla\Delta(\eta v)$ similar to (3.3.8) and integrating by parts yields

$$\begin{aligned}I_1 &= -(\nabla\Delta(\eta v),\nabla w)\\ &= -(\eta\nabla\Delta v,\nabla w)-(3\Delta v\nabla\eta+3\Delta\eta\nabla v+v\nabla\Delta\eta,\nabla w)\\ &= (\operatorname{div}(\eta\nabla\Delta v),w)-(3\Delta v\nabla\eta+3\Delta\eta\nabla v+v\nabla\Delta\eta,\nabla w)\\ &= (\Delta^2 v,\eta w)+(\nabla\Delta v,\nabla\eta w)-(3\Delta v\nabla\eta+3\Delta\eta\nabla v+v\nabla\Delta\eta,\nabla w)\\ &= (\Delta^2 v,\eta w)-(\Delta v,\operatorname{div}(\nabla\eta w))-(3\Delta v\nabla\eta+3\Delta\eta\nabla v+v\nabla\Delta\eta,\nabla w)\\ &= (\Delta^2 v,\eta w)-(\Delta v,w\Delta\eta+\nabla\eta\cdot\nabla w)-(3\Delta v\nabla\eta+3\Delta\eta\nabla v+v\nabla\Delta\eta,\nabla w)\\ &\lesssim (\Delta^2 v,\eta w)+\Bigg(\frac{1}{(1-\sigma)^2R^2}\|\Delta v\|_{L^2}+\frac{1}{(1-\sigma)^2R^2}\|\nabla v\|_{L^2}\\ &\qquad+\frac{1}{(1-\sigma)^3R^3}\|v\|_{L^2}\Bigg)\|\nabla w\|_{L^2}.\end{aligned}\tag{3.3.13}$$

Using (3.3.9) on I_2 we get

$$I_2 = -(A:D^2(\eta v),w)\tag{3.3.14}$$

$$
\begin{aligned}
&= -(\eta A : D^2 v + 2A\nabla v \cdot \nabla \eta + vA : D^2\eta, w) \\
&\lesssim -(A : D^2 v, \eta w) + \left(\frac{1}{(1-\sigma)R}\|\nabla v\|_{L^2} + \frac{1}{(1-\sigma)^2R^2}\|v\|_{L^2} \right) \|\nabla w\|_{L^2}
\end{aligned}
$$

Combining (3.3.13) and (3.3.14) with (3.3.12) gives us

$$
\begin{aligned}
\sqrt{\varepsilon}\|\Delta v\|_{L^2(B_{\sigma R})} &+ \sqrt{\lambda}\|\nabla v\|_{L^2(B_{\sigma R})} \\
&\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2} + \frac{1}{(1-\sigma)R}\|\nabla v\|_{L^2} + \frac{1}{(1-\sigma)^2R^2}\|v\|_{L^2} \\
&\quad + \frac{\varepsilon}{(1-\sigma)^3R^3}\|v\|_{L^2} + \frac{\varepsilon}{(1-\sigma)^2R^2}\|\nabla v\|_{L^2} \\
&\quad + \frac{\varepsilon}{(1-\sigma)^2R^2}\|\Delta v\|_{L^2}.
\end{aligned}
\tag{3.3.15}
$$

Following similar treatment as (3.3.10) we arrive at (3.3.6). This completes the proof. □

We now extend this interior estimates to the boundary to obtain a global estimate.

Lemma 3.7. *Let $\partial\Omega \in C^{2,1}$. For any $v \in H$, the following estimates hold:*

$$
\begin{aligned}
\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|D^2 v\|_{L^2(\Omega)} &\lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)} \\
&\quad + \varepsilon\big(\|\nabla v\|_{L^2(\Omega)} + \|\Delta v\|_{L^2(\Omega)} + \|\nabla\Delta v\|_{L^2(\Omega)}\big),
\end{aligned}
\tag{3.3.16}
$$

$$
\begin{aligned}
\sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)} &\lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(\Omega)} + \|v\|_{L^2(\Omega)} \\
&\quad + \varepsilon(\|\nabla v\|_{L^2(\Omega)} + \|\Delta v\|_{L^2(\Omega)}).
\end{aligned}
\tag{3.3.17}
$$

Proof. Since $\partial\Omega \in C^{2,1}$ for any $x_0 \in \partial\Omega$, we may flatten $\partial\Omega$ near x_0, use Lemma 3.3 and the proof of Lemma 3.5 to create a local boundary estimate mimicking (3.3.3) and (3.3.4). We then follow the same argument as in Lemma 3.6 achieving estimates similar (3.3.5) and (3.3.6) near the boundary. These estimates combined with (3.3.5) and (3.3.6) give us (3.3.16) and (3.3.17). The proof is complete. □

Now we must deal with the terms involving ε on the right hand side of (3.3.16) and (3.3.18). However, since $\sqrt{\varepsilon}$ vanishes slower than ε as $\varepsilon \to 0$, we can easily hide these terms for ε sufficiently small.

Lemma 3.8. *There exists $\varepsilon_0 > 0$ such that for any $\varepsilon < \varepsilon_0$ the following estimates hold for any $v \in H$:*

$$\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|D^2 v\|_{L^2(\Omega)}\sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)} \lesssim \|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)}, \tag{3.3.18}$$

$$\sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)} \lesssim \|\mathcal{L}^\varepsilon v\|_{H^{-1}(\Omega)} + \|v\|_{L^2(\Omega)}. \tag{3.3.19}$$

Proof. Adding (3.3.16) and (3.3.17) and noting $\|\mathcal{L}^\varepsilon v\|_{H^{-1}(\Omega)} \leq \|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)}$ we obtain

$$\begin{aligned}&\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega)} + \sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|D^2 v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)}\\ &\quad\leq C\big(\|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)} + \varepsilon(\|\nabla v\|_{L^2(\Omega)} + \|\Delta v\|_{L^2(\Omega)} + \|\nabla\Delta v\|_{L^2(\Omega)})\big),\end{aligned} \tag{3.3.20}$$

where C is a positive constant independent of ε. Choosing $\varepsilon_0 = \min\{4/C^2, \sqrt{\lambda}/(2C)\}$ gives us $C\varepsilon < \sqrt{\varepsilon}/2$ and $C\varepsilon < \sqrt{\lambda}/2$ for all $\varepsilon < \varepsilon_0$. Let $\varepsilon < \varepsilon_0$, then we subtract the first three terms on the right side of (3.3.20) from both sides and obtain

$$\begin{aligned}\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega)} + \sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)}+&\sqrt{\lambda}\|D^2 v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)}\\ &\leq C\left(\|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)} + \|v\|_{L^2(\Omega)}\right).\end{aligned} \tag{3.3.21}$$

Dropping $\sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)}$ gives us (3.3.18). (3.3.19) can be shown similarly. The proof is complete. □

Using the existence and uniqueness of weak solutions to (P_ε), we now prove a full stability estimate using Lemma 3.8.

Lemma 3.9. *For all $\varepsilon < \varepsilon_0$ and $v \in H$, we have the following stability estimates:*

$$\sqrt{\varepsilon}\|\nabla\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|D^2 v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)} \leq C\|\mathcal{L}^\varepsilon v\|_{L^2(\Omega)}, \tag{3.3.22}$$

$$\sqrt{\varepsilon}\|\Delta v\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla v\|_{L^2(\Omega)} \leq C\|\mathcal{L}^\varepsilon v\|_{H^{-1}(\Omega)}. \tag{3.3.23}$$

where C is a positive independent of ε and u^ε.

Proof. Fix $\varepsilon < \varepsilon_0$ and for the sake of contradiction suppose for $k \in \mathbb{N}$ we have $u^\varepsilon_k \in H^3(\Omega)$ with $u^\varepsilon_k = \Delta u^\varepsilon_k = 0$ on $\partial\Omega$, $\|u^\varepsilon_k\|_{H^3(\Omega)} = 1$, and $\|\mathcal{L}^\varepsilon u^\varepsilon_k\|_{L^2(\Omega)} \to 0$ as $k \to \infty$. Since $\|u^\varepsilon_k\|_{H^3(\Omega)}$ is bounded way may extract a convergent subsequence (not relabeled) such that $u^\varepsilon_k \rightharpoonup u^\varepsilon_*$ weakly in $H^3(\Omega)$ for some $u^\varepsilon_* \in H^3(\Omega)$. In addition, $u^\varepsilon_* = \Delta u^\varepsilon_* = 0$ on $\partial\Omega$, and $\|u^\varepsilon_*\|_{H^3(\Omega)} = 1$. Moreover, since our problem is linear we have

$$\|\mathcal{L}^\varepsilon u^\varepsilon_*\|_{L^2(\Omega)} = \lim_{k\to\infty} \|\mathcal{L}^\varepsilon u^\varepsilon_k\|_{L^2(\Omega)} = 0.$$

By the existence and uniqueness of (P_ε), $u^\varepsilon_* = 0$ which is a contradiction to $\|u^\varepsilon_*\|_{H^3(\Omega)} = 1$. The proof is complete. □

3.4 Convergence of u^ε in $H^2(\Omega)$ and Error Estimates

With the help of Lemma 3.9 we are ready to establish our main results on this chapter. First we show that the sequence $\{u^\varepsilon\}_{\varepsilon<\varepsilon_0}$ indeed converges by compactness, and that the limit function is the strong solution to (P). Moreover, we derive a stability estimate for $\mathcal{L} : H^1 \to H^{-1}$. To the best of our knowledge, this estimate has never been shown in the literature.

Theorem 3.1. *Let $\partial\Omega \in C^{2,1}$, $\varepsilon < \varepsilon_0$ and $u^\varepsilon \in H$ be the weak solution to (P_ε). Then u^ε converges to $u \in H^2(\Omega) \cap H^1_0(\Omega)$ weakly in $H^2(\Omega)$ where u is the strong solution to (P). Moreover, we have the following H^1 stability result for $\mathcal{L}$:*

$$\|\nabla u\|_{L^2(\Omega)} \lesssim \|\mathcal{L}u\|_{H^{-1}(\Omega)}. \tag{3.4.1}$$

Proof. Since $\mathcal{L}^\varepsilon u^\varepsilon = f$ weakly and $u^\varepsilon = 0$ on $\partial\Omega$, we have the boundedness of $\|u^\varepsilon\|_{H^2(\Omega)}$ from the Poincaré inequality and Lemma 3.9. By compactness there exists a subsequence $\{u^\varepsilon\}$ (not relabeled) and a function $u^* \in H^2(\Omega) \cap H_0^1(\Omega)$ such that $u^\varepsilon \rightharpoonup u^*$ weakly in $H^2(\Omega)$. Moreover since $\mathcal{L}^\varepsilon u^\varepsilon = f$ weakly, we have for any $\varphi \in C_0^\infty(\Omega)$

$$\varepsilon(\Delta u^\varepsilon, \Delta\varphi) - (A : D^2 u^\varepsilon, \varphi) = (f, \varphi). \tag{3.4.2}$$

By (3.3.23) we have

$$|\varepsilon(\Delta u^\varepsilon, \Delta\varphi)| \leq \big(\sqrt{\varepsilon}\|\Delta u^\varepsilon\|_{L^2(\Omega)}\big)\big(\sqrt{\varepsilon}\|\Delta\varphi\|_{L^2(\Omega)}\big) \lesssim \sqrt{\varepsilon}\|f\|_{H^{-1}(\Omega)}\|\Delta\varphi\|_{L^2(\Omega)},$$

which vanishes as $\varepsilon \to 0$. Using weak convergence, we can pass the limit as $\varepsilon \to 0$ in (3.4.2) to obtain

$$-(A : D^2 u^*, \varphi) = (f, \varphi) \tag{3.4.3}$$

for any $\varphi \in C_0^\infty(\Omega)$. Thus u^* is a strong solution to (P). By uniqueness of $\mathcal{L}$ we have $u^* = u$, and the whole sequence u^ε weakly converges to u.

We now derive (3.4.1). Since $\|\mathcal{L}^\varepsilon u^\varepsilon\|_{H^{-1}(\Omega)} = \|f\|_{H^{-1}(\Omega)}$ is constant with respect to ε and the L^2 norm is weakly lower semi-continuous, we take the $\liminf$ of (3.3.23) and use $u^\varepsilon \rightharpoonup u$ to get

$$\begin{aligned}
\|f\|_{H^{-1}(\Omega)} &= \liminf_{\varepsilon\to 0} \|\mathcal{L}^\varepsilon u^\varepsilon\|_{H^{-1}(\Omega)} \\
&\gtrsim \liminf_{\varepsilon\to 0} \left(\sqrt{\varepsilon}\|\Delta u^\varepsilon\|_{L^2(\Omega)} + \sqrt{\lambda}\|\nabla u^\varepsilon\|_{L^2(\Omega)}\right) \\
&\gtrsim \liminf_{\varepsilon\to 0} \sqrt{\varepsilon}\|\Delta u^\varepsilon\|_{L^2(\Omega)} + \liminf_{\varepsilon\to 0} \sqrt{\lambda}\|\nabla u^\varepsilon\|_{L^2(\Omega)} \\
&\gtrsim \|\nabla u\|_{L^2(\Omega)}.
\end{aligned}$$

The proof is complete. □

Next we derive some error estimates for $u - u^\varepsilon$ in powers of ε. To this end, we use the H^1 stability estimate for $\mathcal{L}$ along with the stability for $\mathcal{L}^\varepsilon$ to achieve the desired estimates.

Theorem 3.2. *For $\varepsilon < \varepsilon_0$, let $u \in H^2(\Omega) \cap H^1_0(\Omega)$ and $u^\varepsilon \in H$ be the solutions to (P) and (P_ε) respectively. Then we have the following error estimates:*

$$\|\nabla(u^\varepsilon - u)\|_{L^2(\Omega)} \lesssim \sqrt{\varepsilon}\|f\|_{L^2(\Omega)}, \tag{3.4.4}$$

$$\|u^\varepsilon - u\|_{L^2(\Omega)} \lesssim \sqrt{\varepsilon}\|f\|_{L^2(\Omega)}. \tag{3.4.5}$$

Proof. Let $e^\varepsilon = u^\varepsilon - u$. By linearity of $\mathcal{L}$ we get $\mathcal{L}e^\varepsilon = \varepsilon\Delta^2 u^\varepsilon \in H^{-1}(\Omega)$. Using (3.4.1) and (3.3.23) we have

$$\begin{aligned}
\|\nabla e^\varepsilon\|_{L^2(\Omega)} &\lesssim \varepsilon\|\Delta^2 u^\varepsilon\|_{H^{-1}(\Omega)} \\
&= \sup_{v \in H^1_0(\Omega)} \frac{-\varepsilon(\nabla\Delta u^\varepsilon, \nabla v)}{\|\nabla v\|_{L^2(\Omega)}} \\
&\leq \sup_{v \in H^1_0(\Omega)} \frac{\varepsilon\|\nabla\Delta u^\varepsilon\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)}}{\|\nabla v\|_{L^2(\Omega)}} \\
&\leq \varepsilon\|\nabla\Delta u^\varepsilon\|_{L^2(\Omega)} \\
&\leq \sqrt{\varepsilon}\left(\sqrt{\varepsilon}\|\nabla\Delta u^\varepsilon\|_{L^2(\Omega)}\right) \\
&\lesssim \sqrt{\varepsilon}\|f\|_{L^2(\Omega)}.
\end{aligned}$$

which is exactly (3.4.4). By using the Poincaré inequality on $\|\nabla(u^\varepsilon - u)\|_{L^2(\Omega)}$ we obtain (3.4.5). The proof is complete. □

3.5 A C^0 Interior Penalty Method for (P_ε)

As mentioned previously, the goal of applying the vanishing moment method to (P) is to numerically approximate the fourth order equation (P_ε). To this end, we define the following C^0 interior penalty method for (P_ε).

Definition 3.2. *The C^0 interior penalty method for (P_ε) is to seek $u_h \in S_h$ such that*

$$\varepsilon a_h(u_h, v_h) + b_h(u_h, v_h) = (f, v_h) \qquad \forall v_h \in S_h, \tag{3.5.1}$$

where

$$\begin{aligned} a_h(w_h, v_h) := &\sum_{T \in \mathcal{T}_h} \int_T \Delta w_h \Delta v_h \, \mathrm{d}x \\ &- \sum_{e \in \mathcal{E}_h^I} \int_e \{\Delta w_h\}[\nabla v_h \cdot \nu_e] \, \mathrm{d}S - \sum_{e \in \mathcal{E}_h^I} \int_e \{\Delta v_h\}[\nabla w_h \cdot \nu_e] \, \mathrm{d}S \\ &+ \sum_{e \in \mathcal{E}_h^I} \frac{\gamma_e}{h_e} \int_e [\nabla w_h \cdot \nu_e][\nabla v_h \cdot \nu_e] \, \mathrm{d}S, \end{aligned} \tag{3.5.2}$$

and

$$b_h(w_h, v_h) = -\sum_{T \in \mathcal{T}_h} \int_T \left(A : D^2 w_h\right) v_h \, \mathrm{d}x. \tag{3.5.3}$$

Here $a_h(\cdot, \cdot)$ reflects the discrete biharmonic operator while $b_h(\cdot, \cdot)$ represents the discrete non-divergence operator. The C^0 interior penalty discretization of $a_h(\cdot, \cdot)$ is motivated by Brenner in [7]; however, note the only discretization used to create $b_h(\cdot, \cdot)$ is the piecewise discrete Hessian. If $\varepsilon = 0$, then the method may not converge as $h \to 0$ in general.

To test our C^0 interior penalty method, we use a selection of 2-D examples from Section 2.5. However, our solution u to (P) will always be smooth so that the convergence rate will not deteriorate due to the lack of regularity of u. For all tests, we will choose the source f such that the exact solution u is given by

$$u(x_1, x_2) = \sin(2\pi x_1) \sin(2\pi x_2) \exp(x_1 \cos(x_2)).$$

We will test the C^0 interior penalty method using two metrics. First, we set $\varepsilon = h^k$ where k is the polynomial degree of S_h, and compute the errors and rates of convergence of $\|u_h - u\|$ in the L^2- and H^1-norm for $k = 2$, 3, and 4. Second, we fix $h = 1/64$ and $k = 3$, and then we compute the errors and rates of convergence of $\|u_h - u\|$ in the L^2- and H^1-norm with varying ε. With this test as an approximation of $\|u^\varepsilon - u\|$ we hope to corroborate the error estimates (3.4.4-3.4.5); however, since u_h is not exactly u, we will see divergence as ε becomes too small.

3.5.1 Identity Coefficient Matrix A

Let $\Omega = (-1/2, 1/2)^2$ and take $A = I_{d\times d}$. This gives the standard Poisson problem:

$$\begin{aligned} -A : D^2 u = -\Delta u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega. \end{aligned}$$

The L^2 errors and rates of convergence for varying h are shown in Table 3.1 while the H^1 errors and rates of convergence for varying h are shown in Table 3.2. As we see, the method is convergent for all k listed. Also for large ε, we see the method produces poor solutions with lower rates of convergence. This further supports the need for $\varepsilon < \varepsilon_0$ in the theory.

The L^2 errors and rates of convergence for varying ε are shown in Table 3.3 while the H^1 errors and rates of convergence for varying ε are shown in Table 3.4. We see that both the L^2 and H^1 error estimate are of order ε. This is a half order better than our error estimates (3.4.4-3.4.5). In addition, we see that the convergence of the method is dependent on the relationship between h and ε; choosing ε too small will give an inaccurate numerical solution, which is expected.

Table 3.1: The L^2 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with $A = I_{d\times d}$. Here $\varepsilon = h^k$.

	$k=2$		$k=3$		$k=4$	
$1/h$	error	rate	error	rate	error	rate
2	5.21e-01	-	5.11e-01	-	4.50e-01	-
4	5.13e-01	0.02	3.36e-01	0.61	1.28e-01	1.81
8	4.66e-01	0.14	7.57e-02	2.15	1.03e-02	3.64
16	2.96e-01	0.65	1.04e-02	2.87	6.57e-04	3.97
32	8.19e-02	1.85	1.32e-03	2.98	4.10e-05	4.00
64	1.39e-02	2.56	1.67e-04	2.98	2.53e-06	4.02
128	2.81e-03	2.30	2.15e-05	2.96	1.52e-07	4.05

Table 3.2: The H^1 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with $A = I_{d\times d}$. Here $\varepsilon = h^k$.

	$k=2$		$k=3$		$k=4$	
$1/h$	error	rate	error	rate	error	rate
2	5.15e+00	-	5.08e+00	-	4.10e+00	-
4	4.69e+00	0.14	3.07e+00	0.73	1.17e+00	1.81
8	4.10e+00	0.19	7.03e-01	2.13	9.65e-02	3.60
16	2.71e+00	0.60	9.84e-02	2.84	6.58e-03	3.88
32	8.01e-01	1.76	1.29e-02	2.93	4.70e-04	3.81
64	1.37e-01	2.54	1.81e-03	2.84	3.43e-05	3.78
128	2.74e-02	2.33	2.77e-04	2.71	2.73e-06	3.66

Table 3.3: The L^2 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) with $A = I_{d\times d}$. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{L^2(\Omega)}$	rate	$-\log_2 \varepsilon$	$\|u - u_h\|_{L^2(\Omega)}$	rate
1	5.20e-01	-	21	3.17e-05	0.57
2	5.07e-01	0.04	22	3.24e-05	-0.03
3	4.84e-01	0.07	23	4.97e-05	-0.62
4	4.43e-01	0.13	24	9.28e-05	-0.90
5	3.80e-01	0.22	25	1.87e-04	-1.01
6	2.95e-01	0.36	26	3.93e-04	-1.08
7	2.05e-01	0.53	27	8.76e-04	-1.15
8	1.27e-01	0.69	28	2.11e-03	-1.27
9	7.22e-02	0.81	29	5.69e-03	-1.43
10	3.88e-02	0.90	30	1.73e-02	-1.60
11	2.02e-02	0.94	31	5.86e-02	-1.76
12	1.03e-02	0.97	32	2.15e-01	-1.88
13	5.20e-03	0.98	33	8.77e-01	-2.03
14	2.62e-03	0.99	34	1.54e+01	-4.13
15	1.31e-03	1.00	35	1.52e+02	-3.30
16	6.58e-04	1.00	36	9.11e+01	0.74
17	3.30e-04	0.99	37	1.00e+02	-0.14
18	1.67e-04	0.99	38	1.79e+02	-0.84
19	8.57e-05	0.96	39	2.89e+02	-0.69
20	4.71e-05	0.86	40	9.99e+02	-1.79

Table 3.4: The H^1 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) for $A = I_{d\times d}$. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate	$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate
1	4.65e+00	-	21	3.40e-04	0.64
2	4.54e+00	0.03	22	3.12e-04	0.12
3	4.33e+00	0.07	23	4.45e-04	-0.51
4	3.97e+00	0.13	24	8.12e-04	-0.87
5	3.41e+00	0.22	25	1.63e-03	-1.00
6	2.66e+00	0.36	26	3.42e-03	-1.07
7	1.85e+00	0.52	27	7.64e-03	-1.16
8	1.15e+00	0.68	28	1.85e-02	-1.28
9	6.60e-01	0.81	29	4.99e-02	-1.43
10	3.57e-01	0.89	30	1.52e-01	-1.61
11	1.87e-01	0.93	31	7.94e-01	-2.38
12	9.59e-02	0.96	32	2.16e+00	-1.44
13	4.90e-02	0.97	33	8.79e+00	-2.02
14	2.50e-02	0.97	34	1.78e+02	-4.34
15	1.28e-02	0.97	35	8.12e+02	-2.19
16	6.58e-03	0.96	36	2.15e+03	-1.40
17	3.43e-03	0.94	37	8.14e+03	-1.92
18	1.81e-03	0.93	38	6.24e+03	0.39
19	9.62e-04	0.91	39	6.95e+03	-0.16
20	5.32e-04	0.85	40	1.78e+04	-1.36

Table 3.5: The L^2 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with Hölder continuous A. Here $\varepsilon = h^k$.

	$k=2$		$k=3$		$k=4$	
$1/h$	error	rate	error	rate	error	rate
2	5.20e-01	-	4.84e-01	-	3.52e-01	-
4	4.98e-01	0.06	2.15e-01	1.17	5.63e-02	2.65
8	4.34e-01	0.20	3.18e-02	2.76	3.90e-03	3.85
16	2.23e-01	0.96	3.94e-03	3.01	2.46e-04	3.98
32	4.26e-02	2.39	5.00e-04	2.98	1.55e-05	3.99
64	5.80e-03	2.88	6.64e-05	2.91	9.86e-07	3.97
128	1.07e-03	2.43	1.00e-05	2.73	7.01e-08	3.81

3.5.2 Hölder Continuous Coefficient Matrix A

Let $\Omega = (-1/2, 1/2)^2$ and take A as the following Hölder continuous matrix-valued function:

$$A(x) = \begin{bmatrix} |x|^{1/2}+1 & -|x|^{1/2} \\ -|x|^{1/2} & 5|x|^{1/2}+1 \end{bmatrix}, \qquad x \in \mathbb{R}^2.$$

The L^2 errors and rates of convergence are shown in Table 3.5 while the H^1 errors and rates of convergence are shown in Table 3.6. As we see, the method is convergent though the orders of convergence are suboptimal when compared to the IP-DG methods in Chapter 2.

The L^2 errors and rates of convergence for varying ε are shown in Table 3.3 while the H^1 errors and rates of convergence for varying ε are shown in Table 3.4. We see that both the L^2 and H^1 error estimate are of order ε. This tells us the error estimates (3.4.4-3.4.5) are not sharp - even for non-smooth A.

3.5.3 Uniformly Continuous Coefficient A

Let $\Omega = (0, 1/2)^2$ and take A as the following uniformly continuous matrix-valued function:

$$A(x) = \begin{bmatrix} -\dfrac{5}{\log(|x|)}+15 & 1 \\ 1 & -\dfrac{1}{\log(|x|)}+3 \end{bmatrix}.$$

Table 3.6: The H^1 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with Hölder continuous A. Here $\varepsilon = h^k$.

	$k = 2$		$k = 3$		$k = 4$	
$1/h$	error	rate	error	rate	error	rate
2	5.15e+00	-	4.84e+00	-	3.26e+00	-
4	4.45e+00	0.21	2.02e+00	1.26	5.32e-01	2.61
8	3.63e+00	0.29	3.20e-01	2.66	3.87e-02	3.78
16	1.99e+00	0.87	4.13e-02	2.95	2.81e-03	3.79
32	4.24e-01	2.23	5.56e-03	2.89	2.14e-04	3.71
64	6.14e-02	2.79	8.64e-04	2.69	1.54e-05	3.80
128	1.16e-02	2.41	1.50e-04	2.53	2.26e-06	2.76

Table 3.7: The L^2 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) with Hölder continuous A. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{L^2(\Omega)}$	rate
1	4.97e-01	-
2	4.66e-01	0.09
3	4.15e-01	0.17
4	3.41e-01	0.28
5	2.52e-01	0.44
6	1.66e-01	0.60
7	9.95e-02	0.74
8	5.53e-02	0.85
9	2.93e-02	0.91
10	1.52e-02	0.95
11	7.70e-03	0.98
12	3.89e-03	0.99
13	1.95e-03	0.99
14	9.80e-04	1.00
15	4.91e-04	1.00
16	2.47e-04	0.99
17	1.26e-04	0.98
18	6.64e-05	0.92
19	3.97e-05	0.74
20	3.23e-05	0.30
21	4.06e-05	-0.33
22	6.80e-05	-0.74
23	1.26e-04	-0.89
24	2.36e-04	-0.90
25	4.20e-04	-0.83
26	6.45e-04	-0.62
27	7.03e-04	-0.12
28	3.06e-03	-2.12
29	2.13e-02	-2.80
30	1.36e-01	-2.67
31	1.00e+00	-2.89
32	9.16e+00	-3.19
33	8.39e+00	0.13
34	2.42e+01	-1.53
35	1.21e+03	-5.65
36	9.00e+00	7.07
37	5.18e+01	-2.52
38	2.11e+01	1.29
39	3.46e+01	-0.71
40	2.09e+02	-2.59

Table 3.8: The H^1 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) with Hölder continuous A. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate	$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate
1	4.97e-01	-	21	4.06e-05	-0.33
2	4.66e-01	0.09	22	6.80e-05	-0.74
3	4.15e-01	0.17	23	1.26e-04	-0.89
4	3.41e-01	0.28	24	2.36e-04	-0.90
5	2.52e-01	0.44	25	4.20e-04	-0.83
6	1.66e-01	0.60	26	6.45e-04	-0.62
7	9.95e-02	0.74	27	7.03e-04	-0.12
8	5.53e-02	0.85	28	3.06e-03	-2.12
9	2.93e-02	0.91	29	2.13e-02	-2.80
10	1.52e-02	0.95	30	1.36e-01	-2.67
11	7.70e-03	0.98	31	1.00e+00	-2.89
12	3.89e-03	0.99	32	9.16e+00	-3.19
13	1.95e-03	0.99	33	8.39e+00	0.13
14	9.80e-04	1.00	34	2.42e+01	-1.53
15	4.91e-04	1.00	35	1.21e+03	-5.65
16	2.47e-04	0.99	36	9.00e+00	7.07
17	1.26e-04	0.98	37	5.18e+01	-2.52
18	6.64e-05	0.92	38	2.11e+01	1.29
19	3.97e-05	0.74	39	3.46e+01	-0.71
20	3.23e-05	0.30	40	2.09e+02	-2.59

Table 3.9: The L^2 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with uniformly continuous A. Here $\varepsilon = h^k$.

	$k = 2$		$k = 3$		$k = 4$	
$1/h$	error	rate	error	rate	error	rate
2	3.18e-01	-	1.70e-01	-	9.28e-02	-
4	2.38e-01	0.42	3.07e-02	2.47	7.90e-03	3.55
8	9.87e-02	1.27	4.02e-03	2.93	5.05e-04	3.97
16	1.70e-02	2.54	5.10e-04	2.98	3.15e-05	4.00
32	2.58e-03	2.72	6.53e-05	2.96	1.94e-06	4.02
64	5.17e-04	2.32	8.81e-06	2.89	1.18e-07	4.05
128	1.22e-04	2.09	1.40e-06	2.66	7.80e-09	3.92

Table 3.10: The H^1 error and rates of convergence in h for the C^0 interior penalty method (3.5.1) applied to (P_ε) with uniformly continuous A. Here $\varepsilon = h^k$.

	$k = 2$		$k = 3$		$k = 4$	
$1/h$	error	rate	error	rate	error	rate
2	2.93e+00	-	1.57e+00	-	8.37e-01	-
4	2.20e+00	0.41	2.89e-01	2.44	7.25e-02	3.53
8	9.89e-01	1.15	3.86e-02	2.90	4.86e-03	3.90
16	1.91e-01	2.37	5.01e-03	2.95	3.33e-04	3.87
32	2.94e-02	2.70	6.86e-04	2.87	2.36e-05	3.82
64	5.43e-03	2.44	1.03e-04	2.73	2.08e-06	3.51
128	1.28e-03	2.09	1.75e-05	2.56	5.18e-07	2.00

The L^2 errors and rates of convergence are shown in Table 3.9 while the H^1 errors and rates of convergence are shown in Table 3.10. As we see, the method is convergent though the orders of convergence are suboptimal when compared to the IP-DG methods in Chapter 2.

The L^2 errors and rates of convergence for varying ε are shown in Table 3.3 while the H^1 errors and rates of convergence for varying ε are shown in Table 3.4. Again, we see order ε convergence as we descrease ε, until ε is too small for h and the numerical solution does not well approximate the true solution u.

Table 3.11: The L^2 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) with uniformly continuous A. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{L^2(\Omega)}$	rate	$-\log_2 \varepsilon$	$\|u - u_h\|_{L^2(\Omega)}$	rate
1	2.45e-01	-	21	6.81e-06	-0.44
2	1.98e-01	0.31	22	1.15e-05	-0.76
3	1.43e-01	0.47	23	2.03e-05	-0.82
4	9.22e-02	0.64	24	3.30e-05	-0.70
5	5.38e-02	0.78	25	3.70e-05	-0.17
6	2.94e-02	0.87	26	4.52e-05	-0.29
7	1.54e-02	0.93	27	5.55e-04	-3.62
8	7.89e-03	0.96	28	3.01e-03	-2.44
9	4.00e-03	0.98	29	1.40e-02	-2.21
10	2.01e-03	0.99	30	6.19e-02	-2.15
11	1.01e-03	1.00	31	2.80e-01	-2.18
12	5.05e-04	1.00	32	1.38e+00	-2.31
13	2.53e-04	1.00	33	4.69e+00	-1.76
14	1.27e-04	1.00	34	7.14e+00	-0.61
15	6.35e-05	1.00	35	9.52e+00	-0.41
16	3.20e-05	0.99	36	1.06e+02	-3.48
17	1.64e-05	0.97	37	7.01e+03	-6.05
18	8.81e-06	0.89	38	5.69e+02	3.62
19	5.57e-06	0.66	39	3.83e+02	0.57
20	5.02e-06	0.15	40	7.60e+02	-0.99

Table 3.12: The H^1 error and rates of convergence in ε for the C^0 interior penalty method (3.5.1) applied to (P_ε) with uniformly continuous A. Here $h = 1/64$ and $k = 3$.

$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate	$-\log_2 \varepsilon$	$\|u - u_h\|_{H^1(\Omega)}$	rate
1	2.19e+00	-	21	6.47e-05	-0.37
2	1.77e+00	0.30	22	1.09e-04	-0.76
3	1.29e+00	0.46	23	1.93e-04	-0.83
4	8.30e-01	0.63	24	3.16e-04	-0.71
5	4.86e-01	0.77	25	3.71e-04	-0.23
6	2.66e-01	0.87	26	4.29e-04	-0.21
7	1.40e-01	0.93	27	5.08e-03	-3.57
8	7.22e-02	0.96	28	2.77e-02	-2.45
9	3.68e-02	0.97	29	3.76e-01	-3.77
10	1.87e-02	0.98	30	5.78e-01	-0.62
11	9.51e-03	0.98	31	5.01e+00	-3.12
12	4.85e-03	0.97	32	1.28e+01	-1.35
13	2.49e-03	0.96	33	4.68e+01	-1.87
14	1.29e-03	0.94	34	9.50e+01	-1.02
15	6.80e-04	0.93	35	2.09e+02	-1.14
16	3.59e-04	0.92	36	9.14e+03	-5.45
17	1.91e-04	0.92	37	4.51e+05	-5.62
18	1.03e-04	0.89	38	3.29e+04	3.78
19	6.15e-05	0.75	39	7.34e+04	-1.16
20	4.99e-05	0.30	40	1.49e+05	-1.03

3.6 Numerical Tests for the Hamilton Jacobi Bellman Equations

In this section, we introduce a hybrid IP-DG method based on the vanishing moment approach to solve the Hamilton Jacobi Bellman equation:

$$\inf_{\alpha\in\Lambda}\left(-A^{\alpha} : D^2 u - f^{\alpha}\right) = 0 \quad \text{in } \Omega, \tag{3.6.1a}$$

$$u = g \quad \text{on } \partial\Omega. \tag{3.6.1b}$$

where Λ is a parameter set and $\{A^{\alpha}\}$ and $\{f^{\alpha}\}$ are families of functions indexed by $\alpha \in \Lambda$.

To construct our numerical scheme, we define our symmetrically induced bilinear form $a_h(\cdot,\cdot)$ from (2.4.4):

$$\begin{aligned} a_h^{\alpha}(w_h, v_h) := &-\sum_{T\in\mathcal{T}_h}\int_T (A : D^2 w_h) v_h \,\mathrm{d}x + \sum_{e\in\mathcal{E}_h^I}\int_e [A\nabla w_h\cdot\nu_e]\{v_h\}\,\mathrm{d}S \\ &-\sum_{e\in\mathcal{E}_h}\int_e \{A\nabla v_h\cdot\nu_e\}[w_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h}\int_e \frac{\gamma_e}{h_e}[w_h][v_h]\,\mathrm{d}S, \end{aligned} \tag{3.6.2}$$

which induces a stiffness matrix A_h^{α} for each $\alpha \in \Lambda$. We also recall the standard symmetric IP-DG formulation for the biharmonic equation:

$$-\Delta^2 u = 0 \quad \text{in } \Omega, \tag{3.6.3a}$$

$$u = g \quad \text{on } \partial\Omega, \tag{3.6.3b}$$

$$\Delta u = 0 \quad \text{on } \partial\Omega. \tag{3.6.3c}$$

which has the form

$$b_h(w_h, v_h) := -\sum_{T\in\mathcal{T}_h}\int_T \Delta u_h \Delta v_h \,\mathrm{d}x \tag{3.6.4}$$

$$
\begin{aligned}
&+ \sum_{e\in\mathcal{E}_h} \int_e \{\nabla\Delta w_h \cdot \nu_e\}[v_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h} \int_e \{\nabla\Delta v_h \cdot \nu_e\}[w_h]\,\mathrm{d}S \\
&- \sum_{e\in\mathcal{E}_h^I} \int_e \{\Delta w_h\}[\nabla v_h \cdot \nu_e]\,\mathrm{d}S - \sum_{e\in\mathcal{E}_h^I} \int_e \{\Delta v_h\}[\nabla w_h \cdot \nu_e]\,\mathrm{d}S \\
&+ \sum_{e\in\mathcal{E}_h^I} \int_e \frac{\gamma_{1,e}}{h_e}[w_h][v_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h} \int_e \frac{\gamma_{0,e}}{h_e^3}[w_h][v_h]\,\mathrm{d}S.
\end{aligned}
$$

The bilinear form $b_h(\cdot,\cdot)$ induces a stiffness matrix, which we denote as B_h. Lastly we define the source terms with added contributions from the symmetrization terms and the Dirichlet data:

$$
\begin{aligned}
F_1^\alpha(v_h) &:= \int_\Omega f^\alpha v_h\,\mathrm{d}x - \sum_{e\in\mathcal{E}_h^B} \int_e A^\alpha \nabla v_h \cdot \nu_e g\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h^B} \int_e \frac{\gamma_e}{h_e} g v_h\,\mathrm{d}S \\
F_2(v_h) &:= \sum_{e\in\mathcal{E}_h^B} \int_e \nabla\Delta v_h \cdot \nu_e g\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h^B} \int_e \frac{\gamma_{e,0}}{h_e^3} g v_h\,\mathrm{d}S
\end{aligned}
$$

which induce vectors $\boldsymbol{F}_1^\alpha$ and $\boldsymbol{F}_2$. Given $\varepsilon = \varepsilon(h) > 0$, we define our hybrid interior penalty discontinuous Galerkin vanishing moment method (IP-DG-VMM) to approximate the viscosity solution of (3.6.1) as finding $u_h \in V_h$ such that

$$
\varepsilon b(u_h, v_h) + \inf_{\alpha\in\Lambda} \big(a_h^\alpha(u_h, v_h) - F_1^\alpha(v_h)\big) - \varepsilon F_2(v_h) = 0 \qquad \forall v_h \in V_h, \tag{3.6.5}
$$

or, as the nonlinear system

$$
\varepsilon B_h \boldsymbol{u}_h + \inf_{\alpha\in\Lambda} \big(A_h^\alpha \boldsymbol{u}_h - \boldsymbol{F}_1^\alpha\big) - \varepsilon \boldsymbol{F}_2 = 0, \tag{3.6.6}
$$

where $\boldsymbol{u}_h$ is the coefficient vector in the basis expansion of $u_h \in V_h$. We see that (3.6.5) is the symmetric IP-DG discretization of the vanishing moment method applied to the HJB equations - namely:

$$\varepsilon\Delta^2 u + \inf_{\alpha\in\Lambda}(-A^\alpha : D^2 u - f^\alpha) = 0 \quad \text{in } \Omega, \tag{3.6.7a}$$

$$u = g \quad \text{on } \partial\Omega, \tag{3.6.7b}$$

$$\Delta u = 0 \quad \text{on } \partial\Omega. \tag{3.6.7c}$$

We present two numerical examples of this method applied to the Hamilton-Jacobi-Bellman equations. With these examples we wish to answer two questions. First, if the methods do indeed globally converge. Second, if they do converge, what is the dependence on ε. For simpler problems, it is conjectured that setting $\varepsilon = 0$, that is, no moment is added, can provide a convergent method, but for more complicated problems the moment should be required. The system (3.6.6) was solved using Matlab's nonlinear solver `fsolve` with a zero initial guess unless otherwise specified. Since `fsolve` uses a quasi-Newton algorithm, we also list the number of quasi-Newton iterations needed to terminate the solver. The penalty parameters used are $\gamma_e, \gamma_{0,e}, \gamma_{1,e} = 10000$.

3.6.1 Test 1

The first example from [38] provides a simple example of the Hamilton-Jacobi-Bellman equation which we list below. Let $\Omega = (0,\pi)\times(-\pi/2,\pi/2)$.

$$\min\{-\Delta u, -\Delta u/2\} = 0 \text{ in } \Omega, \tag{3.6.8a}$$

$$u = g \text{ on } \partial\Omega. \tag{3.6.8b}$$

Here f is defined by

$$f(x,y) = \begin{cases} 2\cos(x)\sin(x) & \text{if } (x,y)\in S, \\ \cos(x)\sin(x) & \text{otherwise,} \end{cases}$$

Table 3.13: The error and rates of convergence of the L^2 and H^1-norms of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = 0$.

h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
1.57e+00	05	1.71e-01	0.00	5.18e-01	0.00
7.85e-01	04	3.58e-02	2.26	1.40e-01	1.89
3.93e-01	06	8.42e-03	2.09	3.63e-02	1.95
1.96e-01	11	2.07e-03	2.02	9.16e-03	1.99
9.82e-02	21	5.16e-04	2.01	2.29e-03	2.00
4.91e-02	18	1.29e-04	2.00	5.71e-04	2.00
2.45e-02	20	3.22e-05	2.00	1.43e-04	2.00

where $S = (0, \pi/2] \times (-\pi/2, 0] \cup (\pi/2, \pi] \times (0, \pi/2)$ and g is chosen such that the solution is $u(x, y) = \cos(x)\sin(y)$. We approximate the solution u_h using (3.6.5) for $k = 2$ and $\varepsilon = 0$. As we can see if Table 3.13, even without the moment term, the method is convergent with an order of convergence of two for both the L^2- and H^1-norm.

3.6.2 Test 2

For the second test we let $\Omega = (0, 1)^2$ and A^α be chosen from the finite control set

$$A^\alpha \in \left\{ \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & -2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}, \right.$$
$$\left. \begin{bmatrix} 2 & -2 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 2 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix} \right\}.$$

Choose f^α such that the exact solution is $u(x, y) = \sin(2\pi(1.2x - y))$. First we let $\varepsilon = 0$ and $k = 2$, and then compute u_h to see if the method is convergent. Table 3.14 shows the error and convergence rates in the L^2- and H^1-norm. The table indicates that the method is not converging. In addition, the `fsolve` algorithm was taking many more iterations to find a solution, in the order of thousands rather than the 10-20 used for Test 1. However, this could be a result of the initial guess being too

Table 3.14: The error and rates of convergence of the L^2 and H^1-norm of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = 0$ and $k = 2$ starting at an initial guess of $u_0 \equiv 0$.

h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
2.50e-01	16	2.59e-01	0.00	2.66e+00	0.00
1.25e-01	132	2.08e-01	0.32	1.76e+00	0.59
6.25e-02	339	5.10e-01	-1.29	2.80e+00	-0.67
3.12e-02	637	8.54e-02	2.58	1.08e+00	1.37
1.56e-02	846	1.42e-01	-0.73	1.44e+00	-0.40
7.81e-03	1433	3.28e-01	-1.21	2.30e+00	-0.68

Table 3.15: The error and rates of convergence of the L^2 and H^1-norm of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = 0$ and $k = 2$ starting at an initial guess of the L^2 of u onto the space V_h.

h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
2.50e-01	05	2.95e-01	0.00	2.79e+00	0.00
1.25e-01	73	1.92e-02	3.95	7.54e-01	1.89
6.25e-02	27	2.27e-02	-0.24	2.82e-01	1.42
3.12e-02	19	7.02e-03	1.69	8.47e-02	1.74
1.56e-02	04	1.71e-03	2.04	2.24e-02	1.92

far away from the true solution. Thus we repeat the test with $\varepsilon = 0$ and $k = 2$, but setting the initial guess $u_0 = \mathcal{P}_h u$ - the L^2 projection into the space V_h. Again, Table 3.15 shows the error and convergence rates in the L^2- and H^1-norm. We see that the method does converge with a close enough guess. Thus setting $\varepsilon = 0$ may give a locally convergent method, but does not produce a globally convergent one.

Next we set $k = 2$ and $\varepsilon = h^\delta$ where $\delta = 2, 3, 4$, and compute u_h to see if the method is convergent. Tables 3.16, 3.17, 3.18 show the the error and convergence rates in the L^2- and H^1-norm for $\delta = 2, 3, 4$ respectively. As we can see, the method is convergent for all three values of δ. In addition, the number of iterations needed is under 40. We also see that ε must be sufficiently small in order to achieve convergence, similar to the VMM for non-divergence form PDEs. To see how the error is changing as we decrease h, we include Figure 3.1 which plots the error $|u - u_h|$ for various h.

Table 3.16: The error and rates of convergence of the L^2 and H^1-norm of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = h^2$ and $k = 2$ starting at an initial guess of the $u_0 \equiv 0$.

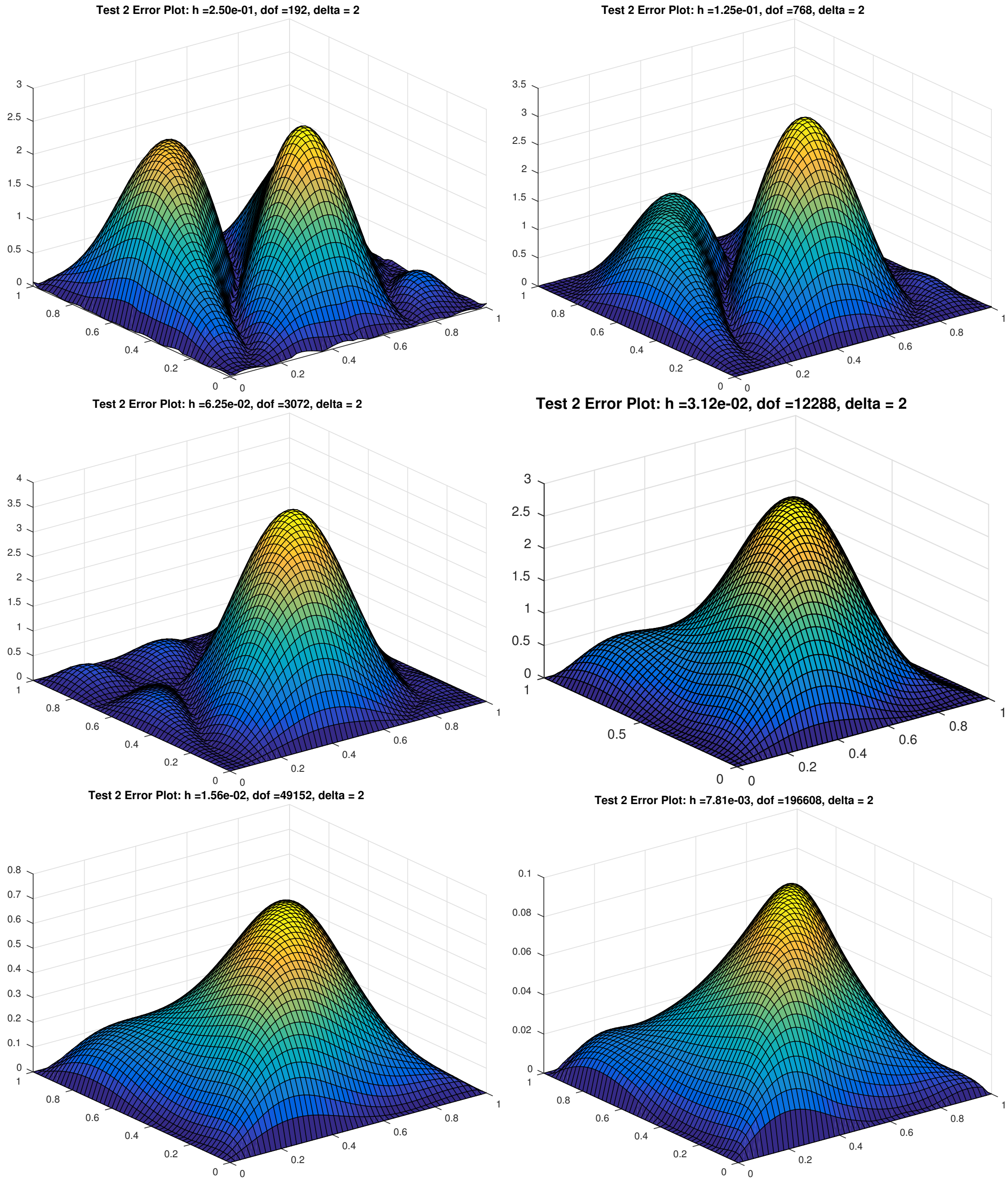
h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
2.50e-01	05	1.16e+00	0.00	9.25e+00	0.00
1.25e-01	05	1.21e+00	-0.07	9.00e+00	0.04
6.25e-02	06	1.43e+00	-0.24	8.65e+00	0.06
3.12e-02	07	1.29e+00	0.15	6.58e+00	0.40
1.56e-02	08	3.39e-01	1.93	1.73e+00	1.93
7.81e-03	09	4.63e-02	2.87	2.53e-01	2.77
3.91e-03	12	7.24e-03	2.68	4.73e-02	2.42

Table 3.17: The error and rates of convergence of the L^2 and H^1-norm of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = h^3$ and $k = 2$ starting at an initial guess of the $u_0 \equiv 0$.

h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
2.50e-01	05	1.21e+00	0.00	9.34e+00	0.00
1.25e-01	05	1.29e+00	-0.10	8.72e+00	0.10
6.25e-02	06	1.32e+00	-0.03	6.77e+00	0.36
3.12e-02	08	2.13e-01	2.63	1.10e+00	2.62
1.56e-02	10	1.08e-02	4.30	6.68e-02	4.04
7.81e-03	13	4.24e-04	4.67	5.73e-03	3.54
3.91e-03	19	6.63e-05	2.68	1.16e-03	2.31

Table 3.18: The error and rates of convergence of the L^2 and H^1-norm of Test 2 of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = h^4$ and $k = 2$ starting at an initial guess of the $u_0 \equiv 0$.

h	iterations	$\|u - u_h\|_{L^2(\Omega)}$	rate	$\|\nabla(u - u_h)\|_{L^2(\Omega)}$	rate
2.50e-01	06	1.27e+00	0.00	9.39e+00	0.00
1.25e-01	06	1.48e+00	-0.22	8.34e+00	0.17
6.25e-02	07	4.22e-01	1.81	2.17e+00	1.94
3.12e-02	10	8.75e-03	5.59	8.26e-02	4.71
1.56e-02	15	1.74e-03	2.33	2.15e-02	1.94
7.81e-03	36	5.45e-04	1.67	7.08e-03	1.60



Test 2 Error Plot: h =2.50e-01, dof =192, delta = 2

Test 2 Error Plot: h =1.25e-01, dof =768, delta = 2

Test 2 Error Plot: h =6.25e-02, dof =3072, delta = 2

Test 2 Error Plot: h =3.12e-02, dof =12288, delta = 2

Test 2 Error Plot: h =1.56e-02, dof =49152, delta = 2

Test 2 Error Plot: h =7.81e-03, dof =196608, delta = 2

Figure 3.1: The plot of $|u - u_h|$ for Test 2 of of the hybrid IP-DG-VMM applied Hamilton-Jacobi-Bellman equations (3.6.1) with $\varepsilon = h^2$ and $k = 2$ starting at an initial guess of the $u_0 \equiv 0$.

Chapter 4

An Enhanced Finite Element Method for a Class of Variational Problems Exhibiting the Lavrentiev Gap Phenomenon

4.1 Introduction

In this chapter, we focus on finite element methods for calculus of variations problems exhibiting the Lavrentiev Gap Phenomenon (LGP). To define the LGP, recall $\mathcal{J}$ from (1.1.4), and let $\mathcal{A} = \{v \in W^{1,1}(\Omega) : v = g \text{ on } \partial\Omega\}$ for an open, bounded domain $\Omega \subset \mathbb{R}^d$ and given g. Let $\mathcal{A} = \mathcal{A} \cap W^{1,\infty}(\Omega)$, that is, all admissible Lipschitz functions. We say that $\mathcal{J}$ exhibits the Lavrentiev Gap Phenomenon if the strict inequality

$$\inf_{v \in \mathcal{A}_1} \mathcal{J}(v) < \inf_{v \in \mathcal{A}_\infty} \mathcal{J}(v) \tag{4.1.1}$$

holds.

A simple 1-D example of the LGP was introduced by Maniá in 1934 (see [42]). Let $d = 1$, $\Omega = (0,1)$, $\mathcal{A} = \{v \in W^{1,1}(\Omega) : v(0) = 0, v(1) = 1\}$. The Maniá example seeks the minimizer u of

$$\mathcal{J}(v) := \int_0^1 (v')^6 (v^3 - x)^2 \,\mathrm{d}x \tag{4.1.2}$$

over $\mathcal{A}$. Since then, many other variational problems have been shown to exhibit the LGP (see [32, 57, 12]).

We emphasize that the Laverntiev gap phenomenon does not hinder the existence of minimizers, and that indeed there are examples of energies that are both coercive and convex that exhibit the LGP (see [32, 57, 12]). In the proof of existence in the direct method of the calculus of varations, given in Subsection 1.4.2, the minimizing sequence obtained from the infimum is not required to be Lipschitz continuous, but only to be a subset of $\mathcal{A}$. Thus problems with the Lavrentiev gap phenomenon can be quite well-posed.

However, we seek to approximate the unique minimizer u of (1.1.5) by minimizing a discrete functional $\mathcal{J}_h$ over the set of finite element functions S_h as seen in (1.4.4). In doing so is where the Lavrentiev gap phenomenon causes problems. It is easy to see since $S_h \subset \mathcal{A}_\infty \subset \mathcal{A}$ for all $h > 0$, an obvious attempt to formulate a numerical method for the variational problem (1.1.5) is the following *standard finite element method* which seeks $u_h \in S_h$ such that

$$u_h = \operatorname*{arg\,min}_{v_h \in S_h} \mathcal{J}(v_h). \tag{4.1.3}$$

Unfortunately, this finite element method fails to give a convergent method because the method cannot give the true minimum value if (4.1.1) holds and will not converge to the correct minimizer as the numerical test shows in Figure 4.1.

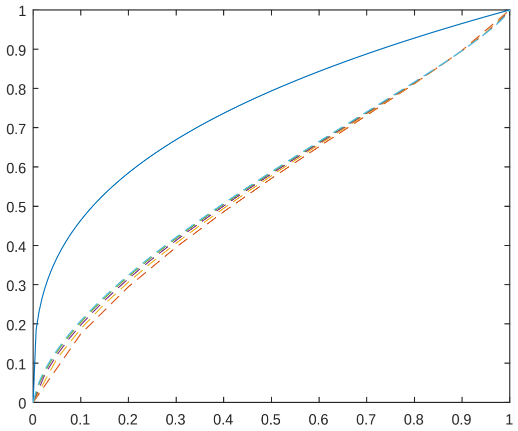


Figure 4.1: The standard finite element method applied to Maniá's problem (4.1.2). The solid line is the true solution $u(x) = x^{\frac{1}{3}}$ and the dashed lines are the finite element minimizers u_h for $h = \frac{1}{N}$ where $N = 10, 20, 40, 80, 160$. All minimizations were implemented by using the MATLAB minimization routine `fminunc` with initial function $u_0(x) = x$.

To see the deeper reason, we note that for any $v \in \mathcal{A}$ the existence of a recovery sequence, defined in Definition 1.1, of functions $\hat{v}_h \in S_h$ with $\hat{v}_h \to v$ in $\mathcal{A}$ such that

$$\lim_{h\to 0} \mathcal{J}_h(\hat{v}_h) = \mathcal{J}(v) \tag{4.1.4}$$

is a key step to show convergence of the discrete minimizers. It is clear that (4.1.1) implies that

$$\mathcal{J}(\hat{u}_h) \geq \inf_{v\in\mathcal{A}_\infty} \mathcal{J}(v) > \mathcal{J}(u)$$

for any $\hat{u}_h \in S_h$ with $\hat{u}_h \to u$ in $\mathcal{A}$, which contradicts with (4.1.4) for the minimizer u. In fact, it was proved by C. Ortner in [48] that for a class of convex energies the convergence (to the exact solution) of the standard finite element method is equivalent to (4.1.1) not holding (i.e., the gap phenomenon does not occur). Thus our primary goal is to construct an effective and robust finite element method to approximate u.

As expected, there have been a few successful attempts to design convergent numerical methods for variational problems with the gap phenomenon. Below we only focus on discussing the methods which use conforming finite element methods to approximate variational problems with the Maniá-type gap phenomenon, by which we mean that the minimizers of the variational problems blow up in the $W^{1,\infty}$-norm, but it is important to note that some gap phenomenon problems have been solved with the use of nonconforming finite element methods [12, 13, 48].

The first numerical method was proposed by Ball and Knowles in [5]. To handle the difficulty caused by the rapid blow-up in $W^{1,\infty}$-norm of the minimizer u, they proposed to approximate u and its derivative u' simultaneously, an idea which is often seen in mixed finite element methods. Specifically, the authors proposed to minimize the discrete energy functional

$$\mathcal{J}_h^{BK}(v_h, w_h) = \int_\Omega f(w_h, v_h, x)\,\mathrm{d}x \tag{4.1.5}$$

under the constraint

$$\|\Phi(v_h' - w_h)\|_{L^1(\Omega)} \leq \varepsilon_h$$

for some super-linear function Φ over all functions $(v_h, w_h) \in S_h^1 \times V_h^0$, where $\{\varepsilon_h\}$ is a sequence such that $\varepsilon_h \to 0$ as $h \to 0$. Notice that $\mathcal{J}_h^{BK}$ essentially has the same form as the original functional $\mathcal{J}$ after setting $w_h = v_h'$. While this method works and is well-posed on the discrete level, the decoupling of v_h and v_h' adds an additional layer of unknowns which increases the complexity of the discrete minimization problem.

The other major numerical developments were carried out by Z. Li *et al.* in [41, 39, 4]. Their work has brought two similar methods: an element removal method and a truncation method. Here we only detail the truncation method and briefly mention the element removal method because the latter is similar to the former and the truncation method is more closely related to our method to be introduced in this paper. Let $s \geq 1$ and $M_h > 0$. Define the discrete energy functional

$$\mathcal{J}_h^{Li}(v_h) = \sum_{T \in \mathcal{T}_h} \mathcal{J}_h^{Li}(v_h; T), \tag{4.1.6}$$

where

$$\mathcal{J}_h^{Li}(v_h;T) = \min\left\{\mathcal{J}_h(v_h;T),\ M_h\left(1+\|\nabla v_h\|_{L^s(T)}\right)\right\},$$
$$\mathcal{J}_h(v_h;T) = \int_T f(\nabla v_h, v_h, x)\,\mathrm{d}x.$$

Here the truncation substitutes the contribution of $\mathcal{J}_h(v_h,T)$ by another constant if v_h behaves "poorly" on T. The element removal method simply discards (i.e., sets $\mathcal{J}_h^{Li}(v_h,T)=0$ on) those "bad" elements. Both methods are robust and calculate the minimum value of $\mathcal{J}$ over $\mathcal{A}_\infty$ (assuming the minimizer u uniquely exists). However, the determination of M_h and s (or "bad" elements) requires a litany of *a priori* assumptions, some of which depend on the sought-after exact minimizer u.

The goal of this chapter is to introduce an effective and robust numerical method which remedis the standard finite element method by a novel and simple cut-off procedure. Our approach is motivated by the rationale that the standard finite element method fails to work because the magnitude of the gradient ∇u_h becomes too large (independent of the magnitude of u_h, where u_h stands for the standard finite element solution) near the singularity points. So the idea of our cut-off procedure is simply to limit the growth of $|\nabla u_h|$ to $\mathcal{O}(h^{-\alpha})$ order in the whole domain Ω, the resulting discrete energy functional is then given by

$$\mathcal{J}_h^\alpha(w_h) = \int_\Omega f\big(\chi_h^\alpha(\nabla w_h), w_h, x\big)\,dx, \tag{4.1.7}$$

where $\chi_h^\alpha(\cdot)$ denotes the cut-off function (see Section 4.2 for its definition). It is important to note that, unlike the truncation method of [4], the choice of the crucial parameter α does not depend on any *a priori* knowledge about the exact minimizer u, instead, it only depends on the structure of the energy density function f and the space $\mathcal{A}$. Moreover, we shall provide a sufficient condition, which is easy to use, for determining an upper bound for α to ensure the convergence.

The organization of this chapter is as follows. In Section 4.2 we state the variational problems we aim to solve and the assumptions under which we develop our numerical method. We then define our finite element method with a help of the above cut-off procedure. In Section 4.3 we show a Γ-convergence result for $\mathcal{J}_h^{\alpha}$ when minimizing the Maniá example under Lipschitz functions. In addition, we also present the alluded sufficient condition for determining an upper bound for α and demonstrate its utility using Maniá's problem. In Section 4.4 we provide some extensive numerical experiment results for two specific application problems to gauge the performance of the proposed enhanced finite element method. In addition, we test the proposed method on the well known minimal surface problem to show the effectiveness of the enhanced finite element method for non-gap phenomenon problems. A portion of this chapter is based on a joint research project which was reported in [29].

4.2 Formulation of the Enhanced Finite Element Method

From the analysis given in the previous section, we conclude that, in order to construct a convergent numerical method which uses S_h as an approximation space, we must design a discrete energy functional $\mathcal{J}_h$ which should not coincide with $\mathcal{J}$ on the finite element space S_h. In this section we shall construct a discrete energy functional $\mathcal{J}_h$ which meets this criterion and provides a convergent (nonstandard) finite element method for problem (1.1.5).

Before introducing our method, let us give a heuristic discussion about why the gap phenomenon is appearing and how the existing methods assuage its effect. Consider Maniá's problem (4.1.2). For any $v_h \in S_h$ (or in $\mathcal{A}_\infty$) sufficiently approximating $u(x) = x^{\frac{1}{3}}$, the quantity $(v_h^3 - x)^2$ will be small but always nonzero. However, at the same time $|v_h'|$ will be very large near the origin. If $|v_h'|$ is raised to a high enough power - six in this case - then the error of $(v_h^3 - x)^2$ will be magnified to be so large that the quantity

$$\int_0^h (v_h')^6 (v_h^3 - x)^2 \,\mathrm{d}x$$

will not vanish as $h \to 0$. For this reason, all of the existing methods were designed to dampen the effect of the derivative in the integral. The method of Ball and Knowles [5] weakly enforces $v_h' = w_h$ which allows the method to soften the effect of v_h', where v_h' has a singularity and achieves convergence. The methods of Li *et al.* [4] leave the function f unchanged, but remove or replace the functional value on the elements where something has gone wrong.

With this in mind we now introduce a discrete energy functional which is much simpler and has a majority of the characteristics of the methods in [41, 39, 4]. Our approach is motivated by the belief that the standard finite element method fails to work because the magnitude of the gradient ∇u_h becomes too large (independent of the magnitude of u_h, where u_h denotes the solution to (4.1.3)) near the singularity points. So our idea is simply to use a cut-off procedure to limit the growth of $|\nabla u_h|$ to $\mathcal{O}(h^{-\alpha})$ on the whole domain Ω in our discrete energy functional $\mathcal{J}_h$. To this end, let $\alpha > 0$, define the cut-off function $\chi_h^\alpha : \mathbb{R}^d \to \mathbb{R}^d$ in the ith component by

$$[\chi_h^\alpha(s)]_i = \begin{cases} s_i & \text{if } |s_i| \leq h^{-\alpha} \\ \operatorname{sgn}(s_i) h^{-\alpha} & \text{if } |s_i| > h^{-\alpha} \end{cases}, \qquad i = 1, 2, \ldots n. \tag{4.2.1}$$

It is clear that this function merely cuts the value of s_i to a constant $\operatorname{sgn}(s_i)h^{-\alpha}$ if $|s_i|$ is too large. Then our *cutoff functional* is simply defined as

$$\mathcal{J}_h^\alpha(v_h) = \int_\Omega f\big(\chi_h^\alpha(\nabla v_h), v_h, x\big) \,\mathrm{d}x, \tag{4.2.2}$$

and our enhanced finite element method is defined by seeking $u_h \in S_h$ such that

$$u_h \in \operatorname*{arg\,min}_{v_h \in S_h} \mathcal{J}_h^\alpha(v_h). \tag{4.2.3}$$

Since our discrete energy functional $\mathcal{J}_h^\alpha$ curbs the gap phenomenon by capping the derivative of its input on a scale of $\mathcal{O}(h^{-\alpha})$, spiritually it is similar to the truncation method of Li *et al.* [4], but, unlike the truncation method, it keeps the dynamics of f with respect to v and x much like Ball and Knowles' approach in [5]. Implementing the cut-off procedure is very simple and can be done by adding a few lines of code. Moreover, unlike the truncation method, our enhanced finite element method does not require *a priori* knowledge about the exact minimizer u of (1.1.5). Further adding to the simplicity is the existence of only one parameter α in the method. Here α controls the rate at which the cut-off grows and is the key for the convergence of the method. In general, α needs to be chosen in order to obtain equation (4.1.4) for all $v \in \mathcal{A}$ where $I_h v \in S_h$ is the finite element interpolant of v. Indeed, (4.1.4) is the only restriction we impose upon α. A permissible range for α, which guarantees convergence, can be determined from the density function f. In Section 4.3, we give a process on how to choose such an α.

4.3 Analysis of the Cutoff Functional $\mathcal{J}_h^\alpha$

In this section we show several results about the cutoff functional $\mathcal{J}_h^\alpha$, beginning with a general lower semi-continuity theorem, and ending with the Γ-convergence of $\mathcal{J}_h^\alpha$ when minimizing over Lipschitz functions to $\mathcal{J}$ for the Maniá example 4.1.2 as $h \to 0$. In addition, we show the process on how to choose α for the Maniá example.

We first state a few definitions and cite a general lower semi-continuity theorem from [40].

Definition 4.1. *A function $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R} \cup \{+\infty\}$ is called $L \otimes B$-measurable, if it is measurable with respect to the σ-algebra generated by the product of Borel subsets of $\mathbb{R}^d \times \mathbb{R}$ and measurable subsets of Ω.*

Definition 4.2. *A function $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R} \cup \{+\infty\}$ is a Carathédory function if* *(a)* $x \to f(\xi, v, x)$ *is measurable for every* $(\xi, v) \in \mathbb{R}^d \times \mathbb{R}$, *and*

(b) $(\xi, v) \to f(\xi, v, x)$ *is continuous for every* $x \in \Omega$.

Definition 4.3. *A sequence of functions* $f_M : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R} \cup \{+\infty\}$ *is said to converge to* $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}_{+\infty}$ *locally uniformly in* $\mathbb{R}^d \times \mathbb{R} \times \Omega$ *if there exists a sequence of measurable sets* $\Omega_l \subseteq \Omega$ *with* $|\Omega \setminus \Omega_l| \to 0$ *as* $l \to \infty$ *such that, for each* l *and any compact subset* $G \subset \mathbb{R} \times \mathbb{R}$, *we have*

$$f_M(\xi, v, x) \to f(\xi, v, x)$$

uniformly on $G \times \Omega_l$ *as* $M \to \infty$.

Lemma 4.1. *Let* $p, q \in [1, \infty]$. *Let* $f : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ *satisfy*

(i) $f(\cdot, \cdot, \cdot)$ *is a Carathédory function,*

(ii) $f(\cdot, x, u)$ *is convex for all* $(u, x) \in \mathbb{R} \times \Omega$, *and*

(iii) $f(\xi, u, x) \leq a(x)$, *for all* $(\xi, u) \in \mathbb{R} \times \mathbb{R}$ *where* $a \in L^1(\Omega)$.

Let $f_M : \mathbb{R} \times \mathbb{R} \times \Omega \to \mathbb{R}$ *be a sequence of functions satisfying*

(a) $f_M(\cdot, \cdot, \cdot)$ *are* $L \otimes B$ *measurable,*

(b) $f_M \to f$ *locally uniformly in* $\mathbb{R} \times \mathbb{R} \times \Omega$, *and*

(c) $f(\xi, u, x) \leq b(x)$, *for all* $(\xi, u) \in \mathbb{R} \times \mathbb{R}$ *where* $b \in L^1(\Omega)$.

Let $\{u_M\}, u \in L^p(\Omega)$ *and* $\{\xi_M\}, \xi \in L^q(\Omega)$, *be such that* $u_M \to u$ *strongly in* $L^p(\Omega)$ *and* $\xi_M \rightharpoonup \xi$ *weakly in* $L^q(\Omega)$, *then*

$$\int_\Omega f(\xi, u, x)\,\mathrm{d}x \leq \liminf_{M\to\infty} \int_\Omega f_M(\xi_M, u_M, x)\,\mathrm{d}x.$$

We now state our lower semi-continiuty result for $\mathcal{J}_h^\alpha$. Note that the assumptions placed on f are weaker than the assumptions used in Bai and Li's truncation method [4].

Theorem 4.1. *Let $\alpha > 0$ and let $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}$ satisfy the following properties.*

(i) $f(\xi, v, x)$ is a Carathédory function.

(ii) $f(\xi, v, x)$ is convex in ξ,

(iii) $f(\xi, v, x) \geq a(x)$ for all $(\xi, v) \in \mathbb{R}^d \times \mathbb{R}$ with $a \in L^1(\Omega)$.

Then for any sequence $v^h \in \mathcal{A}$ with $v_h \rightharpoonup v$ weakly in $W^{1,1}(\Omega)$, we have

$$\mathcal{J}(v) \leq \liminf_{h\to 0} \mathcal{J}_h^\alpha(v^h). \tag{4.3.1}$$

Proof. Since the embedding $W^{1,1}(\Omega) \hookrightarrow L^1(\Omega)$ is compact, we have that $v^h \to v$ strongly in $L^1(\Omega)$.

Define $f_h(\xi, u, x) := f(\chi_h^\alpha(\xi), u, x)$. Since f is continuous in ξ and u, and χ_h^α is continuous, we know f and f_h are Carathédory functions for every $h > 0$. Thus f_h is $L \otimes B$ measurable. Also $f, f_h \geq a$ on Ω for every $(\xi, v) \in \mathbb{R}^d \times \mathbb{R}$. Finally we want to show f_h converges locally uniformly to f in $\mathbb{R}^d \times \mathbb{R} \times \Omega$. Choose $\Omega_l = \Omega$ for all $l > 0$ and let $G \subset \mathbb{R}^d \times \mathbb{R}$ be compact. Since G is a compact subset of $\mathbb{R}^{n+1}$ it must be bounded by the Heine-Borel Theorem. Thus there exists $N > 0$ such that $|\xi| \leq |\xi| + |u| \leq N$ for all $(\xi, u) \in G$. Since $\alpha > 0$, there exists some $h_0 > 0$ such that $h^{-\alpha} > N$ for all $h < h_0$. Thus $\chi_h^\alpha(\xi) = \xi$ for all $h < h_0$, $(\xi, u) \in G$. Because of our definition of f_h we have $f_h \equiv f$ on $G \times \Omega$ for all $h < h_0$, and consequently f_h converges locally uniformly to f on $\mathbb{R}^d \times \mathbb{R} \times \Omega$ as $h \to 0$. Hence by Lemma 4.1, we have exactly (4.3.1). The proof is complete. □

Next we show an upper semi-continuous result for the Maniá example 4.1.2 showing the cutoff functional allows us to approximate $\mathcal{J}(v)$ with $\mathcal{J}_h^\alpha(v^h)$ for any $\alpha > 0$ and v^h Lipschitz continuous.

Lemma 4.1. *Let $d = 1$, $\Omega = (0, 1)$, $\mathcal{A} := \{v \in W^{1,1}(\Omega) : v(0) = 0, v(1) = 1\}$, and $f(\xi, v, x) = \xi^6(v^3 - x)^2$. For any $\alpha > 0$ and $v \in \mathcal{A}$, there exists a sequence $\{v^h\}_{h\geq 0}$*

with $v^h \in \mathcal{A}_\infty$ and $v^h \to v$ in $W^{1,1}(\Omega)$ such that

$$\mathcal{J}(v) \geq \limsup_{h\to 0} \mathcal{J}_h^\alpha(v^h). \tag{4.3.2}$$

Proof. First note if $\mathcal{J}(v) = \infty$, then (4.3.2) holds trivially. Thus we assume $\mathcal{J}(v) < \infty$. For $M > 0$ define

$$q_M(x) = \begin{cases} v'(x) & \text{if } |v'(x)| < M, \\ M & \text{if } |v'(x)| \geq M. \end{cases}$$

Clearly $q_M \in L^\infty(\Omega)$ for all M. Define

$$Q_M = \int_0^1 q_M(y)\,\mathrm{d}y.$$

Since $q_M \to v'$ in $L^1(\Omega)$, then

$$Q_M \to \int_0^1 v'\,\mathrm{d}x = 1$$

as $M \to \infty$. Finally define

$$w_M(x) = \frac{1}{Q_M}\int_0^x q_M(y)\,\mathrm{d}y.$$

By construction $w_M \in \mathcal{A} \cap W^{1,\infty}(\Omega)$. We will show $w_M \to v$ in $W^{1,1}(\Omega)$ as $M \to \infty$. Since $d = 1$, we have by Hölder's inequality and Sobolev embedding that

$$\|w_M - v\|_{L^p(\Omega)} \leq \|w_M - v\|_{L^\infty(\Omega)} \leq C\|w_M' - v'\|_{L^1(\Omega)}$$

for all $1 \leq p \leq \infty$ and for some pure constant $C > 0$. Thus it is sufficient to show $w_m' \to v'$ in $L^1(\Omega)$. By definition of w_M and q_M we have

$$\int_0^1 |w_M' - v'|\,\mathrm{d}x = \int_0^1 \left|\frac{1}{Q_M} q_M - v'\right| \mathrm{d}x$$

$$
\begin{aligned}
&= \int_{\{|v'|\geq M\}} \left|\frac{1}{Q_M} q_M - v'\right| \mathrm{d}x + \int_{\{|v'|<M\}} \left|\frac{1}{Q_M} q_M - v'\right| \mathrm{d}x \\
&= \int_{\{|v'|\geq M\}} \left|\frac{1}{Q_M} M - v'\right| \mathrm{d}x + \int_{\{|v'|<M\}} \left|\frac{1}{Q_M} v' - v'\right| \mathrm{d}x \\
&\leq \int_{\{|v'|\geq M\}} \frac{1}{Q_M} M + |v'| \,\mathrm{d}x + \left|\frac{1-Q_M}{Q_M}\right| \int_{\{|v'|<M\}} |v'| \,\mathrm{d}x \\
&\leq \int_{\{|v'|\geq M\}} \frac{1}{Q_M} |v'| + |v'| \,\mathrm{d}x + \left|\frac{1-Q_M}{Q_M}\right| \|v'\|_{L^1(\Omega)} \\
&\leq \left(\frac{1}{|Q_M|} + 1\right) \int_{\{|v'|\geq M\}} |v'| \,\mathrm{d}x + \left|\frac{1-Q_M}{Q_M}\right| \|v'\|_{L^1(\Omega)} \\
&\leq \left(\frac{1}{|Q_M|} + 1\right) \int_0^1 \mathbb{1}_{\{|v'|\geq M\}} |v'| \,\mathrm{d}x + \left|\frac{1-Q_M}{Q_M}\right| \|v'\|_{L^1(\Omega)},
\end{aligned}
$$

where $\mathbb{1}_E$ is the indicator function of the set E. Since $\mathbb{1}_{\{|v'|\geq M\}}|v'| \to 0$ pointwise,

$$
\|\mathbb{1}_{\{|v'|\geq M\}}|v'|\|_{L^1(\Omega)} \leq \|v'\|_{L^1(\Omega)} < \infty
$$

for all $M > 0$, and $Q_M \to 1$ as $M \to \infty$, we have $\|w_M' - v'\|_{L^1(\Omega)} \to 0$ as $M \to \infty$ by the Dominated Convergence Theorem. Thus $w_m \to v$ in $W^{1,1}(\Omega)$.

Furthermore since $w_M \to v$ in $L^2(\Omega)$ we can choose $M_h > 0$ such that if $q_h := q_{M_h}$, $Q_h := Q_{M_h}$, $v^h := w_{M_h}$, then $\|v - v^h\|_{L^2(\Omega)}^2 = \mathcal{O}(h^{6\alpha+1})$. Let $\delta > 0$. By Young's inequality with weight h^δ, we get for $0 < \delta < 1$,

$$
\begin{aligned}
\mathcal{J}_h(v^h) &= \int_0^1 (\chi_h^\alpha((v^h)'))^6((v^h)^3 - x)^2 \\
&= \int_0^1 (\chi_h^\alpha((v^h)'))^6((v^h)^3 - v^3 + v^3 - x)^2 \,\mathrm{d}x \\
&\leq \int_0^1 (1 + h^{-\delta})(\chi_h^\alpha((v^h)'))^6((v^h)^3 - v^3)^2 \,\mathrm{d}x \\
&\qquad + \int_0^1 (1 + h^{\delta})(\chi_h^\alpha((v^h)'))^6(v^3 - x)^2 \,\mathrm{d}x \\
&\leq \int_0^1 (1 + h^{-\delta})h^{-6\alpha}((v^h)^3 - v^3)^2 \,\mathrm{d}x \\
&\qquad + (1 + h^{\delta}) \int_0^1 ((v^h)')^6(v^3 - x)^2 \,\mathrm{d}x
\end{aligned}
$$

$$=: B_1^h + B_2^h.$$

We now show B_1^h vanishes. By Hölders inequality we have

$$\begin{aligned} B_1^h &= (1+h^{-\delta})h^{-6\alpha}\int_0^1 ((v^h)^3 - v^3)^2 \,\mathrm{d}x \\ &= (1+h^{-\delta})h^{-6\alpha}\int_0^1 (v^h - v)^2((v^h)^2 + v^h v + v^2)^2 \,\mathrm{d}x \\ &\le (1+h^{-\delta})h^{-6\alpha}||v^h - v||^2_{L^2(\Omega)}||((v^h)^2 + v^h v + v^2)^2||_{L^\infty(\Omega)}. \end{aligned}$$

Since $\|v - v^h\|^2_{L^2(\Omega)} = \mathcal{O}(h^{6\alpha+1})$, and v^h is uniformly bounded in h, then B_1^h clearly vanishes. For B_2^h, we use the definition of v^h and q_h to get

$$\begin{aligned} B_2^h/(1+h^\delta) &= \int_0^1 ((v^h)')^6 (v^3 - x)^2 \,\mathrm{d}x \\ &= \int_0^1 \frac{(q_h)^6}{Q_h^6}(v^3 - x)^2 \,\mathrm{d}x \\ &= \int_{\{|v'| \ge M_h\}} \frac{(q_h)^6}{Q_h^6}(v^3 - x)^2 \,\mathrm{d}x + \int_{\{|v'| < M_h\}} \frac{(q_h)^6}{Q_h^6}(v^3 - x)^2 \,\mathrm{d}x \\ &\le \frac{1}{Q_h^6}\left(\int_{\{|v'| \ge M_h\}} (M_h)^6 (v^3 - x)^2 \,\mathrm{d}x + \int_{\{|v'| < M_h\}} (v')^6 (v^3 - x)^2 \,\mathrm{d}x\right) \\ &\le \frac{1}{Q_h^6}\left(\int_{\{|v'| \ge M_h\}} (v')^6 (v^3 - x)^2 \,\mathrm{d}x + \int_{\{|v'| < M_h\}} (v')^6 (v^3 - x)^2 \,\mathrm{d}x\right) \\ &\le \frac{1}{Q_h^6}\mathcal{J}(v). \end{aligned}$$

Since $(1+h^\delta)/Q_h \to 1$ as $h \to 0$, we have (4.3.2). The proof is complete. □

With these two results in hand, we are ready to state our main result about our cutoff functional $\mathcal{J}_h^\alpha$ for the Maniá example.

Theorem 4.2. *Let $d = 1$, $\Omega = (0,1)$, $\mathcal{A} := \{v \in W^{1,1}(\Omega) : v(0) = 0, v(1) = 1\}$, and $f(\xi, v, x) = \xi^6(v^3 - x)^2$. For $h > 0$ let $\mathcal{J}^\alpha_{h,\mathcal{A}_\infty}(v) = \mathcal{J}^\alpha_h(v) + \Lambda_{\mathcal{A}_\infty}(v)$ where*

$$\Lambda_{\mathcal{A}_\infty}(v) = \begin{cases} 0 & \text{if } v \in \mathcal{A}_\infty, \\ \infty & \text{if } v \notin \mathcal{A}_\infty. \end{cases}$$

Then $\mathcal{J}^\alpha_{h,\mathcal{A}_\infty}(v) \xrightarrow{\Gamma} \mathcal{J}$ in the weak $W^{1,1}$ topology.

Proof. Let $v \in \mathcal{A}$. Since $\Lambda_{\mathcal{A}_\infty} \geq 0$ we have by properties of lim inf and Theorem 4.1,

$$\begin{aligned} \liminf_{h\to 0} \mathcal{J}^\alpha_{h,\mathcal{A}_\infty}(v) &\geq \liminf_{h\to 0} \mathcal{J}^\alpha_h(v) + \liminf_{h\to 0} \Lambda_{\mathcal{A}_\infty}(v) && (4.3.3) \\ &\geq \liminf_{h\to 0} \mathcal{J}^\alpha_h(v) && (4.3.4) \\ &\geq \mathcal{J}(v) && (4.3.5) \end{aligned}$$

for every sequence $\{v^h\}_{h\geq 0} \subset \mathcal{A}$ with $v^h \to v$ weakly in $W^{1,1}(\Omega)$. Thus (4.3.3) satisfies (1) of Definition 1.1. Moreover, Lemma 4.1 gives us (2) of Definition 1.1 since $\Lambda_{\mathcal{A}_\infty} \equiv 0$ on $\mathcal{A}_\infty$, and f is continuous and non-negative on $\mathbb{R} \times \mathbb{R} \times \Omega$. Thus $\mathcal{J}^\alpha_{h,\mathcal{A}_\infty}(v) \xrightarrow{\Gamma} \mathcal{J}$ in the weak $W^{1,1}$ topology. The proof is complete. □

It must be stressed that the Γ-convergence in Theorem 4.2 is only valid when minimizing over all Lipschitz functions, not over finite element functions. While we have not proven Lemma 4.1 with a sequence of finite element functions rather than Lipschtiz functions, we can with an unproven assumption. We will include the proof with this assumption in place to show how the parameter α may be tuned to achieve (4.1.4) without *a-priori* knowledge of the minimizer u.

Lemma 4.1. *Let $d = 1$, $\Omega = (0,1)$, $\mathcal{A} := \{v \in W^{1,1}(\Omega) : v(0) = 0, v(1) = 1\}$, and $f(\xi, v, x) = \xi^6(v^3 - x)^2$. Let $\alpha < 1/6$. For any $v \in \mathcal{A}$, let $v_h = I_h v$ be the nodal interpolant of v. We assume the following convergence holds:*

$$\int_0^1 (v_h')^6(v^3 - x)^2 \,\mathrm{d}x \to \mathcal{J}(v) \tag{4.3.6}$$

as $h \to 0$. Then we have

$$\mathcal{J}(v) \geq \limsup_{h \to 0} \mathcal{J}_h^\alpha(v_h).$$

Proof. Let $\delta > 0$. Adding and subtracting v, using Young's inequality with weight h^δ, and using the definition of χ_h^α we get

$$\begin{aligned}
\mathcal{J}_h^\alpha(v_h) &= \int_0^1 (\chi_h^\alpha(v_h'))^6 (v_h^3 - x)^2 \,\mathrm{d}x \\
&= \int_0^1 (\chi_h^\alpha(v_h'))^6 (v_h^3 - v^3 + v^3 - x)^2 \,\mathrm{d}x \\
&\leq \int_0^1 (1 + h^{-\delta})(\chi_h^\alpha(v_h'))^6 (v_h^3 - v^3)^2 \,\mathrm{d}x + \int_0^1 (1 + h^\delta)(\chi_h^\alpha(v_h'))^6 (v^3 - x)^2 \,\mathrm{d}x \\
&\leq \int_0^1 (1 + h^{-\delta}) h^{-6\alpha} (v_h^3 - v^3)^2 \,\mathrm{d}x + \int_0^1 (1 + h^\delta)(v_h')^6 (v^3 - x)^2 \,\mathrm{d}x \\
&=: A_1^h + A_2^h.
\end{aligned}$$

By (4.3.6), we have $A_2^h \to \mathcal{J}(v)$ as $h \to 0$. We claim that A_1^h vanishes as $h \to 0$ for $0 < \alpha < \frac{1}{6}$. The proof of the assertion goes as follows. By Hölder's inequality we have

$$\begin{aligned}
A_1^h &= (1 + h^{-\delta}) h^{-6\alpha} \int_0^1 (v_h^3 - v^3)^2 \,\mathrm{d}x \\
&= (1 + h^{-\delta}) h^{-6\alpha} \int_0^1 (v_h - v)^2 (v_h^2 + v_h v + v^2)^2 \,\mathrm{d}x \\
&\leq (1 + h^{-\delta}) h^{-6\alpha} \|v_h - v\|_{L^2(\Omega)}^2 \|(v_h^2 + v_h v + v^2)^2\|_{L^\infty(\Omega)}.
\end{aligned}$$

Since $v_h = I_h v$ we have that v_h is uniformly bounded in h and $\|v_h - v\|_{L^2(\Omega)}^2 = \mathcal{O}(h)$. Thus

$$0 \leq A_1^h \leq \|(v_h^2 + v_h v + v^2)^2\|_{L^\infty(\Omega)} (1 + h^{-\delta}) h^{1 - 6\alpha}.$$

Since $\alpha < \frac{1}{6}$ we may choose $\delta < 1 - 6\alpha$ such that $A_1^h \to 0$ as $h \to 0$ and we have (4.1.4). The proof is complete. □

Remark 4.1.

1. *Note that $v^3 - x$ factor in A_2^h would now have zero error since v is the actual input into $\mathcal{J}$ and not a Lipschtiz approximation. Thus multiplying by $(v_h')^6$ does not have a magnification effect which is the source of the gap phenomenon. With this in mind, we believe that (4.3.6) is true.*

2. *Clearly, the range of α does not depend on the solution u but only on the form of f and the regularity of the space $\mathcal{A}$. We regard this property as one crucial advantage of our method.*

4.4 Numerical Experiments

In this section we first present some numerical experiment results for two variational problems which are known to exhibit the gap phenomenon. The first problem is Maniá's 1-D problem which has been seen in the previous sections; the second problem, which was proposed by Foss in [31], is a 2-D variational problem from nonlinear elasticity. For each of the two test problems we solve it by using our enhanced finite element method with linear element (i.e., $k = 1$), and we solve the minimization problem (4.2.3) by using the MATLAB minimization function `fminunc`. We first demonstrate the convergence of the numerical method, we then numerically evaluate the effect and sharpness of the parameter α, and compare with the standard finite element method (which is known to be divergent). We also numerically compute the rate of convergence for $u - u_h$ although no theoretical rate convergence has yet been proved for the numerical method. To show that the proposed method also works for non-gap phenomenon problems, we present a numerical test for the minimal surface problem [21].

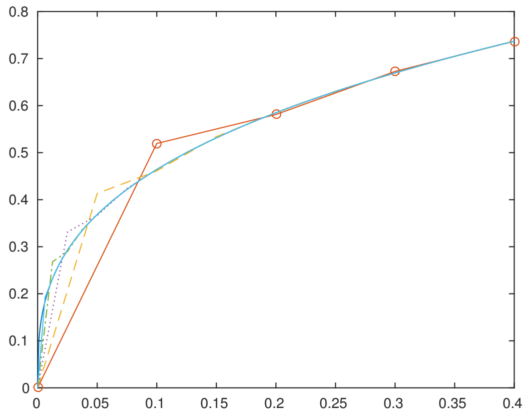


Figure 4.2: The graphs of the computed minimizers/solution u_h of the enhanced FEM applied to Maniá's problem (4.1.2) with parameter $\alpha = \frac{1}{4}$ from $x = 0$ to $x = 0.4$. The solid line is the exact solution $u(x) = x^{\frac{1}{3}}$ and the dashed and circled lines are the minimizers u_h for $h = \frac{1}{N}$ where $N = 10, 20, 40, 80, 160$. All minimizations were implemented by using the MATLAB minimization function `fminunc` with initial function $u_0(x) = x$.

4.4.1 Maniá's 1-D Problem

Once again, the energy functional of Maniá's 1-D problem is given by (4.1.2). A uniform mesh $\mathcal{T}_h$ with mesh size h and the linear finite element are used in the test. As mentioned above, we solve the resulting minimization problem (4.2.3) by using the MATLAB minimization function `fminunc` with initial function $u_0(x) = x$.

Figure 4.2 displays the computed solutions (minimizers) u_h with various mesh size h along with the exact solution $u(x) = x^{\frac{1}{3}}$. The parameter $\alpha = \frac{1}{4}$ is used for the tests. It is clear that the solutions u_h are correctly approximating u. Figure 4.3 shows the behavior of the absolute value of the error function $u - u_h$. As expected, we see that the location where the biggest error occurs moves closer to the singularity point $x = 0$ of u as the mesh size h gets smaller.

For a more detailed look, we also record the L^∞-norms of the error $u - u_h$ and compute the rate of convergence in Table 4.1. The table clearly shows the convergence of the computed solutions u_h. As a comparison and to see that these approximations would not be found using the standard finite element method, a comparison of the

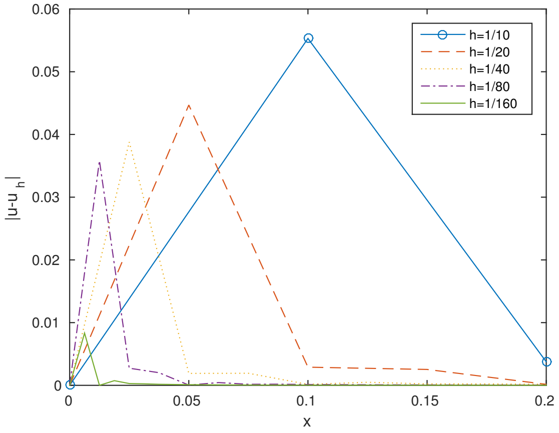


Figure 4.3: The graphs of the errors $|u_h - u|$ of the enhanced FEM applied to Maniá's problem (4.1.2) with parameter $\alpha = \frac{1}{4}$ from $x = 0$ to $x = 0.2$ for $h = \frac{1}{N}$ where $N = 10, 20, 40, 80, 160$. All minimizations were implemented by using the MATLAB minimization function `fminunc` with initial function $u_0(x) = x$.

values of $\mathcal{J}$ and $\mathcal{J}_h^\alpha$ at u_h, $I_h^1 u$, and $I_h^2(u)$ is given in Table 4.2, where I_h^1 and I_h^2 are the piecewise linear and quadratic interpolants respectively. We see here that $\mathcal{J}_h^\alpha$ correctly captures the dynamics needed to obtain a convergent sequence of solutions u_h while the sequences $\{\mathcal{J}(u_h)\}$ and $\{\mathcal{J}(I_h u)\}$ do not. In addition $\{\mathcal{J}_h^\alpha(I_h^1 u)\}$ and $\{\mathcal{J}_h^\alpha(I_h^2 u)\}$ converge with the same rate, $\mathcal{O}(h^{1.5})$. Thus employing higher order elements on this problem will not result in a larger convergence rate. To make this clear we plot the convergence rate of the numerical minimizers u_h of $\mathcal{J}_h^\alpha$ for linear and quadratic elements in Figure 4.4. Note both elements observe the same convergence rate of 1.5.

h	1/10	1/20	1/40	1/80	1/160
$\|u - u_h\|_{L^\infty}$	5.53e-2	4.50e-2	3.88e-2	3.59e-2	8.32e-3
rate	-	0.30	0.20	0.11	2.10

Table 4.1: The L^∞ errors between u and u_h where u_h are the solutions of the enhanced FEM applied to Maniá's problem (4.1.2) with parameter $\alpha = \frac{1}{4}$.

Finally, we examine the role of the parameter α. In section 4.2 we show that $\alpha < \frac{1}{2}$ is sufficient to ensure (4.1.1) for all $v \in \mathcal{A}$ with finite energy. Our numerical tests show that for any $\alpha < 1/2$ the enhanced finite element method converges for

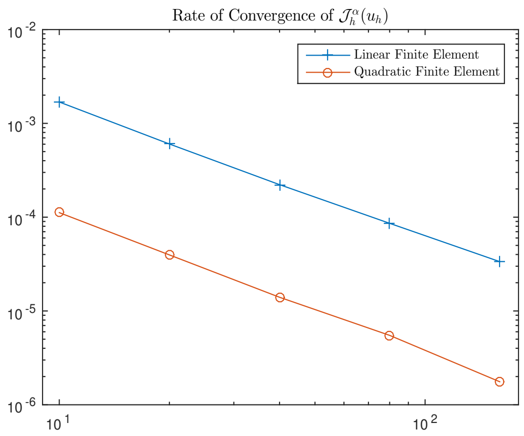


Figure 4.4: The rate of convergence of $\mathcal{J}_h^\alpha(u_h)$ where u_h is the solution to enhanced FEM applied to Maniá's problem (4.1.2) with parameter $\alpha = \frac{1}{4}$ for $h = \frac{1}{N}$ where $N = 10, 20, 40, 80, 160$. Plotted are the rates for the linear and quadratic finite element spaces. All minimizations were implemented by using the MATLAB minimization function `fminunc` with initial function $u_0(x) = x^{1/2}$.

Maniá's problem, and the convergence of $|\mathcal{J}_h^\alpha(u_h) - \mathcal{J}(u)| \to 0$ diminishes as $\alpha \to \frac{1}{2}$. So $\alpha^* := \frac{1}{2}$ seems a critical point for the choice of α for linear, quadratic, and higher order nodal finite elements. It must be noted that taking α close to α^* is not a good idea. Notice that the Euler-Lagrange equation of (4.1.2) is a nonlinear equation. To solve the nonlinear equation, a mesh restriction $h < h'$ is expected, and it takes up most of the total CPU time for solving the nonlinear equation. This mesh restriction is expected to depend on α. To see this, let

$$\tilde{u}_h = \operatorname*{arg\,min}_{v_h \in S_h} \mathcal{J}(u_h)$$

be the solution to the standard finite element method. Suppose that α is close to $\frac{1}{2}$, we observe that $\mathcal{J}_h^\alpha(\tilde{u}_h) \approx \mathcal{J}(\tilde{u}_h)$. While $\mathcal{J}_h^\alpha(I_h u)$ indeed converges to $\mathcal{J}(u)$ the convergence is very slow. Since the upper bound h' must be chosen such that for all $h < h'$ we have

$$\mathcal{J}_h^\alpha(I_h u) < \mathcal{J}_h^\alpha(\tilde{u}_h),$$

Table 4.2: The functional values $\mathcal{J}$ and $\mathcal{J}_h^\alpha$ for u_h, $I_h^1 u$, and $I_h^2 u$, where u_h are the solutions of the enhanced FEM applied to Maniá's problem (4.1.2) with parameter $\alpha = \frac{1}{4}$, and $I_h^1 u$ and $I_h^2 u$ are the piecewise linear and quadratic nodal interpolant of the exact solution/minimizer u.

h	1/10	1/20	1/40	1/80	1/160
$\mathcal{J}(u_h)$	8.23e-1	1.64	3.28	6.56	13.1
$\mathcal{J}_h^\alpha(u_h)$	1.68e-3	6.02e-4	2.22e-5	8.59e-6	3.31e-5
$\mathcal{J}(I_h^1 u)$	7.19e-1	1.52	3.04	6.09	12.9
$\mathcal{J}_h^\alpha(I_h^1 u)$	2.41e-3	8.63e-4	3.09e-4	1.10e-4	3.91e-5
$\mathcal{J}(I_h^2 u)$	1.16	2.31	4.62	9.24	18.5
$\mathcal{J}_h^\alpha(I_h^2 u)$	1.71e-4	6.03e-5	3.09e-4	7.54e-6	2.67e-6

so h' must be extremely small and approaches 0 as $\alpha \to \frac{1}{2}$. On noting the fact that for all $h \geq h'$ a small perturbation of $\tilde{u}_h$ will be a minimizer of $\mathcal{J}_h^\alpha$ over S_h, we see that α must be chosen carefully in order to guarantee that we can obtain good numerical solutions with any mesh sizes $h < h'$. To show this important detail graphically, Figure 4.5 displays the computed solutions/minimizers u_h to $\mathcal{J}_h^\alpha$ with $\alpha = \frac{2}{7}$. We observe that for $h = \frac{1}{10}$ and $h = \frac{1}{20}$, u_h do not approximate u well, but for $h = \frac{1}{40}$, $h = \frac{1}{180}$ and $h = \frac{1}{160}$, u_h gives much more accurate approximations.

4.4.2 Foss' 2-D Problem

We now consider a 2-D variational problem which exhibits the Lavrentiev gap phenomenon. It arises from nonlinear elasticity and was first studied by M. Foss in [31], and its numerical approximation was investigated by Li *et al.* in [4].

Let $\Omega = (0,1) \times (\frac{3}{2}, \frac{5}{2})$, the energy functional of Foss' problem is given by

$$\mathcal{J}(v) = 66\Big(\frac{13}{14}\Big)^{14} \int_\Omega \Big(\frac{y}{y-1}\Big)^{14} |u|^{\frac{14-3y}{y-1}} \big(|u|^{\frac{y}{y-1}} - x\big)^2 (u_x)^{14} \,\mathrm{d}x\,\mathrm{d}y, \tag{4.4.1}$$

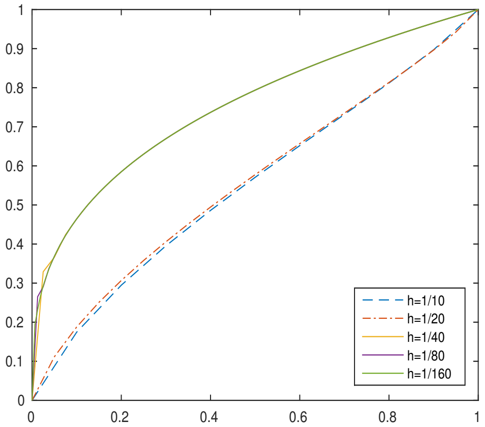


Figure 4.5: The graphs of the computed solutions/minimizers u_h of the enhanced FEM applied to the Maniá's problem (4.1.2) with parameter $\alpha = \frac{2}{7}$ for $h = \frac{1}{N}$ where $N = 10, 20, 40, 80, 160$. The dotted lines are for $N = 10$ and 20 while the solid lines are for $N = 40, 80$, and 160. All minimizations were implemented by using the MATLAB minimization function `fminunc` with initial function $u_0(x) = x$.

and the admissible set is $\mathcal{A} = \{u \in W^{1,1}(\Omega) : u(0,\cdot) = 0 \text{ and } u(1,\cdot) = 1\}$. It was shown by Foss [31] that

$$0 = \inf_{v\in\mathcal{A}} \mathcal{J}(v) < \inf_{v\in\mathcal{A}_\infty} \mathcal{J}(v) = 1,$$

which proves that $\mathcal{J}$ does exhibit the gap phenomenon. Moreover, the minimizer of $\mathcal{J}$ over $\mathcal{A}$ is given by $u(x,y) = x^{\frac{y-1}{y}}$, but the problem does not attain its minimum value in $\mathcal{A}_\infty$.

We apply our enhanced finite element method with $\alpha = \frac{1}{6}$ to solve Foss' problem. In order to generate a reasonably good initial guess for using the MATLAB minimization function `fminunc`, we first compute

$$\tilde{u}_h = \operatorname*{arg\,min}_{v_h\in S_h} \mathcal{J}(v_h) \tag{4.4.2}$$

using the MATLAB routine `fminunc` with initial guess $u(x,y) = x$ and then use $\tilde{u}_h$ as an initial condition for solving

$$u_h = \arg\min_{v_h \in S_h} \mathcal{J}_h^{\alpha}(v_h) \tag{4.4.3}$$

using the same MATLAB routine `fminunc`.

Figure 4.6 presents the error plots of both $|\hat{u}_h - u|$ and $|u_h - u|$ over the domain Ω. We observe that $|\tilde{u}_h - u|$ does not converge to zero while $|u_h - u|$ does. In addition, Table 4.3 shows that the cut-off procedure is sufficient in order to guarantee convergence for (4.4.1). Using the same reasoning as to show (4.1.4) for Maniá's problem, a value of $\alpha < \frac{3}{14}$ is sufficient for the proposed enhanced finite element method to work. However, computing the functional values $\mathcal{J}_h^{\alpha}(I_h u)$ with different values of α shows that $\alpha = \frac{1}{2}$ and $\alpha = \frac{10}{17}$ also result in convergent methods.

Table 4.3: The functional values $\mathcal{J}$ and $\mathcal{J}_h^{\alpha}$ at $\tilde{u}_h$ and u_h, where $\tilde{u}_h$ and u_h satisfy (4.4.2) and (4.4.3) respectively for problem (4.4.1). Here $\alpha = \frac{1}{6}$.

h	1/6	1/12	1/24
$\mathcal{J}(\tilde{u}_h)$	14.84	5.71	3.21
$\mathcal{J}_h^{\alpha}(\tilde{u}_h)$	11.46	4.74	2.68
$\mathcal{J}(u_h)$	3330	3914	4047
$\mathcal{J}_h^{\alpha}(u_h)$	1.28e-1	5.45e-3	5.28e-4

4.4.3 Minimal Surface Problem

Our last example will be a 2-D minimal surface problem. Let $\Omega = (0,1)^2$ and define the energy function $\mathcal{J}$ by

$$\mathcal{J}(v) = \int_{\Omega} \left(1 + |\nabla v|^2\right)^{1/2} \mathrm{d}x \tag{4.4.4}$$

and $\mathcal{A} := \{v \in W^{1,1}(\Omega) : v(x,0) = v(x,1) = x^2, v(0,\cdot) = 0, v(1,\cdot) = 1\}$. This functional arises from differential geometry, and the minimizer u of (4.4.4) should

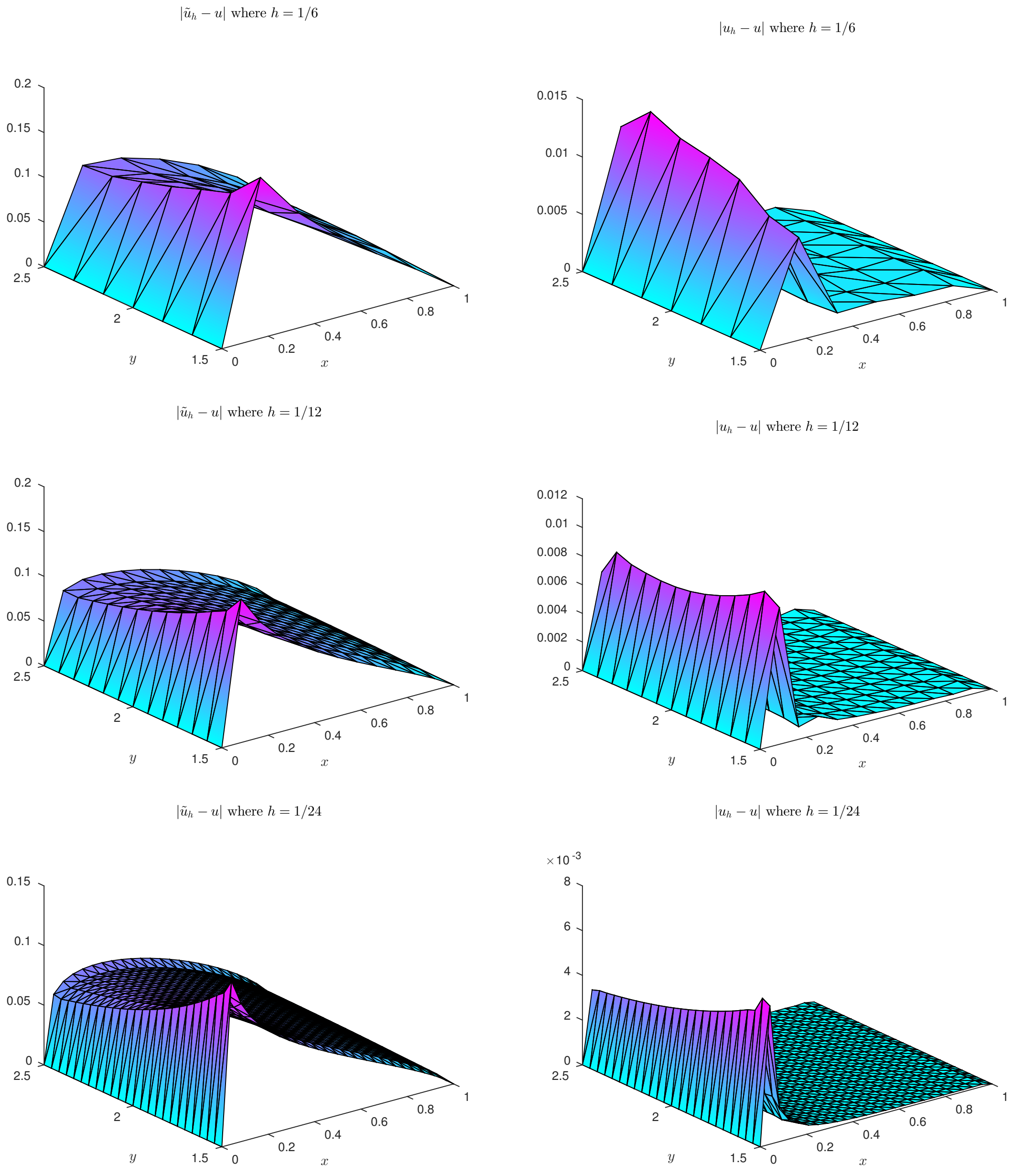


Figure 4.6: The graphs of the error function $|u - \tilde{u}_h|$ (left column) and the error function $|u - u_h|$ (right column) with $\alpha = \frac{1}{6}$ for $h = \frac{1}{6}, \frac{1}{12}$, and $\frac{1}{24}$. All minimizations were done by using the MATLAB minimization function `fminunc` with an intial guess $u_0(x, y) = x$.

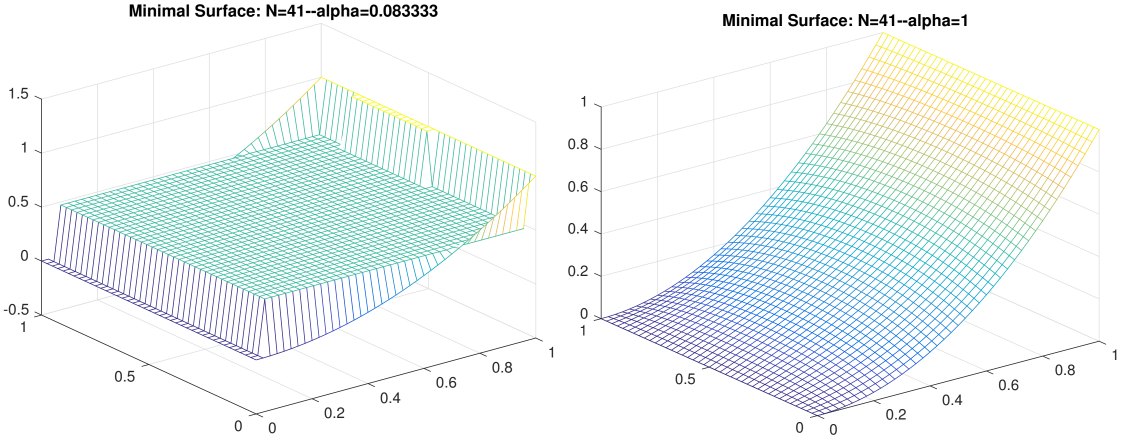


Figure 4.7: The graphs of the minimizer to the enhanced finite element method u_h with $h = 1/40$. Here $\alpha = \frac{1}{12}$ (left column) and $\alpha = 1$ (right column). All minimizations were done by using the MATLAB minimization function `fminunc` with an initial guess $u_0(x, y) = x$.

have zero mean curvature in Ω. Note that $\mathcal{J}$ does not exhibit the Lavrentiev gap phenomenon, but we can still test our enhanced finite element method to see the results. It can be shown that $\alpha \geq 1$ is sufficient to guarantee equi-coercivity of $\mathcal{J}_h^\alpha$ which is vital for the convergence of the method. Figure 4.7 shows the results of the enhanced finite element method for $\alpha = 1/12$ and $\alpha = 1$. As we can see, if we make α too small, then our enhanced finite element method will give us a minimizer that is constant on all interior nodes of $\mathcal{T}_h$. This is because the cutoff does not penalize large contributions of the gradient near the boundary, and, consequently, $\|\nabla u_h\|_{L^1(\Omega)} \to \infty$ as $h \to 0$. For larger α, equi-coercivity is maintained and the enhanced minimizer u_h agrees with the standard finite element minimizer $\tilde{u}_h$ (not pictured).

Chapter 5

A Discontinuous Ritz Framework for a Class of Convex and Coercive Problems from the Calculus of Variations

5.1 Introduction

In this chapter, we develop a discontinuous Ritz framework for numerically approximating solutions to problems from the Calculus of Variations:

$$u \in \operatorname*{arg\,min}_{v \in W_g^{1,p}(\Omega)} \mathcal{J}(v), \tag{5.1.1}$$

where

$$\mathcal{J}(v) = \int_\Omega f(\nabla v, v, x) \, \mathrm{d}x \tag{5.1.2}$$

is the energy and $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}$ is the density function, $W_g^{1,p}(\Omega) := \{v \in W^{1,p}(\Omega) : u = g \text{ on } \partial\Omega\}$, and $\Omega \subset \mathbb{R}^d$ is an open bounded domain. We seek to approximate the minimizer u of $\mathcal{J}$ over $W^{1,p}_g(\Omega)$. To do this, we follow the direct approach to construct the approximate solution u_h, that is, we construct a discrete energy $\mathcal{J}_h$ and let

$$u_h = \operatorname*{arg\,min}_{v_h \in X_h} \mathcal{J}_h(v_h), \tag{5.1.3}$$

where X_h is a discrete space which approximates $W^{1,p}_g(\Omega)$.

First, we let $X_h = V_h$ - the space of discontinuous, piecewise polynomial functions on a mesh $\mathcal{T}_h$ of Ω. The construction of $\mathcal{J}_h$ is crucial to the convergence of the method. In particular, since our discrete functions v_h are discontinuous across interior edges, the gradient ∇v_h is only piecewisely defined. Thus we must construct our discrete gradient $\nabla_h v_h$ judicially. A naive approach is to define the piecewise gradient as the discrete gradient. This gives us the following discrete energy functional:

$$\begin{aligned} \mathcal{J}_h^{pw}(v_h) = &\sum_{T \in \mathcal{T}_h} \int_T f(\nabla v_h, v_h, x)\,\mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_e \gamma_e h_e^{1-p} |[v_h]|^p \,\mathrm{d}S \\ &+ \sum_{e \in \mathcal{E}_h^B} \int_e \gamma_e h_e^{1-p} |v_h - g|^p \,\mathrm{d}S, \end{aligned} \tag{5.1.4}$$

where the last two terms are penalty terms to weakly enforce continuity and the Dirichlet boundary conditions. However, this approach does not always give an accurate scheme - even for nice f [10]. To show this, let $p = 2$, $g = 0$, and let $f(\xi, v, x) = \frac{1}{2}|\xi|^2 - F(x)v$ be the energy density to the Poisson problem:

$$-\Delta u = F \quad \text{in } \Omega, \tag{5.1.5a}$$

$$u = 0 \quad \text{on } \partial\Omega. \tag{5.1.5b}$$

Requiring the variational derivative of (5.1.4) to be zero for every $v_h \in V_h$, that is,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{J}_h^{pw}(u_h + t v_h)\bigg|_{t=0} = 0 \quad \forall v_h \in V_h,$$

we get the following problem: find $u_h \in V_h$ such that

$$a_h^{pw}(u_h, v_h) = (F, v_h) \qquad \forall v_h \in V_h,$$

where

$$a_h^{pw}(u_h, v_h) := \sum_{T \in \mathcal{T}_h} \int_T \nabla u_h \cdot \nabla v_h \, \mathrm{d}x + \sum_{e \in \mathcal{E}_h^I} \int_e \frac{\gamma_e}{h_e}[u_h][v_h] \, \mathrm{d}S + \sum_{e \in \mathcal{E}_h^B} \int_e \frac{\gamma_e}{h_e} u_h v_h \, \mathrm{d}S. \tag{5.1.6}$$

The bilinear form a_h^{pw} is coercive and continuous on V_h for any $\gamma_e > 0$, which immediately implies the existence and uniqueness of a discrete minimizer u_h; however, it is not *consistent*, that is, if u is the weak solution to (5.1.5), then there is a $v_h \in V_h$ such that

$$a_h^{pw}(u, v_h) \neq (F, v_h).$$

Instead we have

$$a_h^{pw}(u, v_h) = (F, v_h) + \sum_{e \in \mathcal{E}_h^I} \int_e \{\nabla u \cdot \nu_e\}[v_h] \, \mathrm{d}S$$

for every $v_h \in V_h$. Note the penalty terms are not the cause for the inconsistency, since the regularity and boundary data of u forces them to vanish. However, it is the discretization of gradient that causes the inconsistency. The inconsistency in this example, being $\mathcal{O}(1)$, leads to an non-convergent method.

In [10], Buffa and Ortner introduced a *variational DGFEM*. This method provided a consistent discretization of the gradient that produces a convergent method for a class of convex and coercive energies. Their discrete gradient is defined using the piecewise gradient and a lifting operator $R : W_h^{1,p}(\mathcal{T}_h) \to [V_h]^d$,

$$\int_\Omega R(v) \cdot \varphi_h = -\sum_{e \in \mathcal{E}_h^I} \int_e [v]\{\varphi_h \cdot \nu_e\} \, \mathrm{d}S \quad \forall \varphi_h \in [V_h]^d. \tag{5.1.7}$$

The motivation of this lifting operator arises from the contributions of the jumps of a discontinuous function in its distributional derivative. They then defined the following discrete energy:

$$
\begin{aligned}
\mathcal{J}_h^{BO}(v_h) = &\sum_{T\in\mathcal{T}_h} \int_T f(\nabla v_h + R(v_h), v_h, x)\,\mathrm{d}x \\
&+ \sum_{e\in\mathcal{E}_h^I} \int_e \gamma_e h_e^{1-p} |[v_h]|^p \,\mathrm{d}S + \sum_{e\in\mathcal{E}_h^B} \int_e \gamma_e h_e^{1-p} |v_h - g|^p \,\mathrm{d}S.
\end{aligned}
\tag{5.1.8}
$$

The bilinear form induced from this energy for the the Poisson problem is

$$
\begin{aligned}
a_h^{BO}(u_h, v_h) = &\sum_{T\in\mathcal{T}_h} \int_T \nabla u_h \cdot \nabla v_h \,\mathrm{d}x + \int_\Omega R(u_h)\cdot R(u_h)\,\mathrm{d}x \\
&- \sum_{e\in\mathcal{E}_h^I} \int_e [u_h]\{\nabla v_h \cdot \nu_e\}\,\mathrm{d}S - \sum_{e\in\mathcal{E}_h^I} \int_e [v_h]\{\nabla u_h \cdot \nu_e\}\,\mathrm{d}S \\
&+ \sum_{e\in\mathcal{E}_h^I} \int_e \frac{2\gamma_e}{h_e}[u_h][v_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h^B} \int_e \frac{2\gamma_e}{h_e} u_h v_h \,\mathrm{d}S,
\end{aligned}
$$

which is coercive and continuous on V_h for sufficiently large $\gamma_e > 0$. Moreover, $a_h^{BO}(\cdot,\cdot)$ is consistent since

$$
\int_\Omega R(u)\cdot R(v_h)\,\mathrm{d}x = \sum_{e\in\mathcal{E}_h^I} \int_e [u]\{\nabla R(v_h)\cdot \nu_e\}\,\mathrm{d}S = 0
$$

for all $v_h \in V_h$, which contributes to the convergence of the method for the Poisson problem.

It must be stressed that the piecewise gradient discretization has the ability to produce a consistent scheme if we include additional terms to the discrete energy. For example, for the Poisson problem, the standard symmetric interior penalty DG bilinear form is

$$
a_h^{SIPDG}(u_h, v_h) = \sum_{T\in\mathcal{T}_h} \int_T \nabla_h u_h \cdot \nabla_h v_h \,\mathrm{d}x
$$

$$
\begin{aligned}
&- \sum_{e \in \mathcal{E}_h^I} \int_e [u_h]\{\nabla v_h \cdot \nu_e\} \, \mathrm{d}S - \sum_{e \in \mathcal{E}_h^I} \int_e [v_h]\{\nabla u_h \cdot \nu_e\} \, \mathrm{d}S \\
&+ \sum_{e \in \mathcal{E}_h^I} \int_e \frac{\gamma_e}{h_e} [u_h][v_h] \, \mathrm{d}S + \sum_{e \in \mathcal{E}_h^B} \int_e \frac{\gamma_e}{h_e} u_h v_h \, \mathrm{d}S.
\end{aligned}
$$

It can be shown that $a_h^{SIPDG}(u_h, v_h)$, being symmetric, is induced by the following discrete energy:

$$
\begin{aligned}
\mathcal{J}_h^{SIPDG}(v_h) = &\sum_{T \in \mathcal{T}_h} \frac{1}{2} \int_T |\nabla v_h|^2 \, \mathrm{d}x - \sum_{e \in \mathcal{E}_h^I} \int_e [v_h]\{\nabla v_h \cdot \nu_e\} \, \mathrm{d}S \\
&+ \sum_{e \in \mathcal{E}_h^I} \frac{1}{2} \int_e \frac{\gamma_e}{h_e} |[v_h]|^2 \, \mathrm{d}S + \sum_{e \in \mathcal{E}_h^B} \frac{1}{2} \int_e \frac{\gamma_e}{h_e} |v_h - g|^2 \, \mathrm{d}S.
\end{aligned}
$$

Moreover, it was proved in [10] that the lifting operator ensures compactness of the discrete minimizers u_h. Since the minimizer of $\mathcal{J}_h^{BO}$ is sought in V_h, which is not a subset of $W^{1,p}(\Omega)$, the reflexive property of $W^{1,p}(\Omega)$ cannot be used to obtain a weakly convergent subsquence. However, V_h is a subset of $\mathrm{BV}(\Omega)$, the space of functions with bounded variation, which does have a compactness property. This compactness alone only shows that a subsequence u_{h_j} converges to a $u \in BV(\Omega)$, but Buffa and Ortner were able to prove a stronger result: if the sequence of discrete minimizers u_h is bounded in $W^{1,p}(\mathcal{T}_h)$, then a subsequence converges to $u \in W^{1,p}(\Omega)$. Moreover, there holds the weak convergence

$$
\nabla u_{h_j} + R(u_{h_j}) \rightharpoonup \nabla u \text{ in } L^p(\Omega),
$$

where ∇u_{h_j} is the piecewise gradient of u_{h_j}. This compactness requires the lifting operator to be present in the discretization in order to pass the week limit and prove convergence of the method.

The goal of this chapter is to develop a discontinuous Ritz (DR) framework for the minimization problem (5.1.1). As we have seen, the discretization of the gradient operator is critical to ensure the convergence of the method. Our main idea is to use discrete derivatives introduced in the discontinuous Galerkin finite elment (DG-FE) numerical calculus by Feng, Lewis, and Neilan (see [25]). In Section 5.2, we give the definition of the DG-FE derivatives, the motivation for using it, and then define our discontinuous Ritz method. Section 5.3 is devoted to the analysis of the DR method. We show that while both the Variational DGFEM and the DR methods are defined from different motivations, they are actually equivalent schemes, which leads to the convergence of the DR method for a specific class of f. In addition, we present a compactness result using our DG-FE numerical gradient. In Section 5.4, we show a few numerical tests using the discontinuous Ritz method on the p-Laplace problem.

5.2 The DG-FE Numerical Derivative and the Discontinuous Ritz Formulation

5.2.1 The DG-FE Numerical Derivatives

To define the DG-FE numerical derivatives, we first introduce some notation. Let $i = 1, \ldots, d$. We define the following trace operators $\mathcal{Q}_i^+, \mathcal{Q}_i^-, \mathcal{Q}_i$ as

$$\begin{aligned} \mathcal{Q}_i^{\pm}(v) &= \{v\} \pm \frac{1}{2}\operatorname{sgn}(\nu_e^i)[v], \\ \mathcal{Q}_i(v) &= \frac{1}{2}\left(\mathcal{Q}_i^+(v) + \mathcal{Q}_i^-(v)\right), \end{aligned} \tag{5.2.1}$$

where ν_e^i denotes the i^{th} component of ν_e, the normal vector to $e \in \mathcal{E}_h$, and

$$\operatorname{sgn}(\xi) = \begin{cases} 1 & \text{if } \xi \geq 0, \\ -1 & \text{if } \xi < 0. \end{cases}$$

With these trace operators in hand we can define three numerical partial derivative operators corresponding the the left, right, and central traces of v.

Definition 5.1. *Let $v \in W^{1,p}(\mathcal{T}_h)$ and $i = 1, \dots, d$. Define the numerical partial derivative operators in the x_i coordinate $\partial^+_{h,x_i}, \partial^-_{h,x_i}, \partial_{h,x_i} : W^{1,p}(\mathcal{T}_h) \to V_h$ by*

$$\int_\Omega \partial^\pm_{h,x_i}(v)\varphi_h \,\mathrm{d}x = \sum_{e\in\mathcal{E}_h} \int_e \mathcal{Q}^\pm_i(v)\nu^i_e[\varphi_h] \,\mathrm{d}S - \sum_{T\in\mathcal{T}_h} \int_T v\partial_{x_i}\varphi_h \,\mathrm{d}x \quad \forall\varphi_h \in V_h, \tag{5.2.2}$$

$$\partial_{h,x_i}(v) = \frac{1}{2}\left(\partial^+_{h,x_i}(v) + \partial^-_{h,x_i}(v)\right). \tag{5.2.3}$$

We call $\partial_{h,x_i}(v)$ the central partial derivative in the x_i coordinate. The motivation for these numerical derivatives is to require the standard integration by parts formula to hold when tested against any discrete function $\varphi_h \in V_h$. This allows many of the properties of the classical derivative to hold for the numerical derivatives; among them are the product rule, chain rule, and integration by parts. Because of this, a discrete energy built using the DG-FE derivative should be consistent. In addition, we can also define the discrete gradients $\nabla^+_h, \nabla^-_h, \nabla_h : W^{1,p}(\mathcal{T}_h) \to [V_h]^d$ by

$$\nabla^\pm_h v = [\partial^\pm_{h,x_1}(v), \partial^\pm_{h,x_2}(v), \dots, \partial^\pm_{h,x_d}(v)], \tag{5.2.4}$$

$$\nabla_h v = [\partial_{h,x_1}(v), \partial_{h,x_2}(v), \dots, \partial_{h,x_d}(v)]. \tag{5.2.5}$$

5.2.2 Formulation of the Discontinuous Ritz Method

With the DG-FE gradients in hand, we can define our discontinuous Ritz method.

Definition 5.2. *Our discontinuous Ritz method is defined as seeking $u_h \in V_h$ such that*

$$u_h \in \operatorname*{arg\,min}_{v_h\in V_h} \mathcal{J}_h(v_h), \tag{5.2.6}$$

where

$$\mathcal{J}_h(v) = \int_\Omega f(\nabla_h v, v, x) \,\mathrm{d}x + \sum_{e\in\mathcal{E}^I_h} \int_e \gamma_e h_e^{1-p} |[v_h]|^p \,\mathrm{d}S \tag{5.2.7}$$

$$+\sum_{e\in\mathcal{E}_h^B}\int_e \gamma_e h_e^{1-p}|v_h - g|^p \,\mathrm{d}S.$$

To compute the numerical derivative $\partial_{h,x_i} v$, we note that the mass matrix induced by the left-hand side of (5.2.2) is actually a block diagonal matrix which means the computation of the derivative can be done locally and in parallel. Moreover, when determining the DG-FE partial derivative of a discrete function, the linearity of $\partial^{\pm}_{h,x_i}$ and ∂_{h,x_i} allows the DG-FE partial derivatives to be written as a matrix which can be computed off-line.

5.3 Analysis of the Discontinuous Ritz Method

In this section we show the convergence of the discontinuous Ritz method defined in Definition 5.2 as well as several properties of the DG-FE derivative which will be useful in future implementations of the DR framework.

We first show that the DR method is actually equivalent to the Variational DGFEM developed by Buffa and Ortner in [10]. Specifically, we shall prove $\mathcal{J}_h \equiv \mathcal{J}_h^{BO}$ on V_h, thus giving equivalence of these two methods when minimizing over V_h.

Lemma 5.1. *Let $\mathcal{J}_h^{BO}$ and $\mathcal{J}_h$ be defined by* (5.1.8) *and* (5.2.7) *respectively, then for any $v_h \in V_h$ we have $\mathcal{J}_h(v_h) = \mathcal{J}_h^{BO}(v_h)$.*

Proof. Let $v_h \in V_h$, if we can show that $\nabla_h v_h = \nabla v_h + R(v_h)$, where ∇v_h is the piecewise gradient, then the equivalence of the two methods follows. Luckily, this property was already proved in Proposition 4.2 of [25], but below we include the whole proof for completeness.

By definition of $\nabla_h v_h$ and the DG integration by parts formula (2.1.4), we have

$$\begin{aligned}\int_\Omega \nabla_h v_h\cdot\varphi_h &= \sum_{e\in\mathcal{E}_h}\int_e \{v_h\}[\varphi_h\cdot\nu_e]\,\mathrm{d}S - \sum_{T\in\mathcal{T}_h}\int_T v_h \operatorname{div}\varphi_h \,\mathrm{d}x \qquad (5.3.1)\\ &= -\sum_{e\in\mathcal{E}_h^I}\int_e [v_h]\{\varphi_h\cdot\nu_e\}\,\mathrm{d}S + \sum_{T\in\mathcal{T}_h}\int_T \nabla v\cdot\varphi_h\,\mathrm{d}x\end{aligned}$$

$$
\begin{aligned}
&= \sum_{T\in\mathcal{T}_h} \int_T (\nabla v_h + R(v_h)) \cdot \varphi_h \,\mathrm{d}x \\
&= \int_\Omega (\nabla v_h + R(v_h)) \cdot \varphi_h \,\mathrm{d}x. \qquad \forall \varphi_h \in [V_h]^d,
\end{aligned}
$$

Since $\nabla_h v_h, \nabla v_h, R(v_h) \in [V_h]^d$ and

$$
\int_\Omega \big(\nabla_h v_h - (\nabla v_h + R(v_h))\big) \cdot \varphi_h \,\mathrm{d}x = 0 \qquad \forall \varphi_h \in [V_h]^d,
$$

Setting $\varphi_h = \nabla_h v_h - (\nabla v_h + R(v_h))$ we obtain $\nabla_h v_h = \nabla v_h + R(v_h)$ in Ω. Thus $\mathcal{J}_h(v_h) = \mathcal{J}_h^{BO}(v_h)$. The proof is complete. □

With the equivalence we can borrow and take advantage of the convergence result from Theorem 6.1 of [10] for a specific class of density functions f. To do this, we first need to introduce some notation. Consider the broken Sobolev space $W^{1,p}(\mathcal{T}_h)$ equipped with the following semi-norm and norm:

$$
|v|_{W_h^{1,p}(\Omega)} := \|\nabla v\|_{L^p(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p} \|[v]\|_{L^p(e)}^p\Big)^{\frac{1}{p}}, \tag{5.3.2}
$$

$$
\|v\|_{W_h^{1,p}(\Omega)} := |v|_{W_h^{1,p}(\Omega)} + \Big(\sum_{e\in\mathcal{E}_h^B} h_e^{1-p} \|[v]\|_{L^p(e)}^p\Big)^{\frac{1}{p}}. \tag{5.3.3}
$$

Note that the newly defined norm $\|\cdot\|_{W_h^{1,p}(\Omega)}$ and the norm defined in (1.6.2) are equivalent on V_h. Now we are ready to state the convergence result.

Theorem 5.1. *For $1 < p < \infty$, let p^* be the Sobolev conjugate of p, that is,*

$$
p^* = \begin{cases} \frac{dp}{d-p} & \textit{if } p < d, \\ \infty & \textit{if } p \geq d. \end{cases} \tag{5.3.4}
$$

Let $f : \mathbb{R}^d \times \mathbb{R} \times \Omega \to \mathbb{R}$ be a Carathédory function (see Definition 4.2) and let $\xi \to f(\xi, v, x)$ be convex for every $(v, x) \in \mathbb{R} \times \Omega$. In addition, suppose there exists constants $c_0, c_1 > 0$, functions $a_0, a_1 \in L^1(\Omega)$, and numbers r and q satisfying $r < p$

and $r \leq q < p^*$, such that the following growth estimate holds:

$$c_0(|\xi|^p - |v|^r + a_0(x)) \leq f(\xi, v, x) \leq c_1(|\xi|^p + |v|^q + a_1(x)).$$

For $h > 0$, let $u_h \in V_h$ satisfy (5.2.6). Then there exists a sequence $h_j \searrow 0$ and a function $u \in W^{1,p}_g(\Omega)$ such that the following hold:

$$\begin{aligned} u_{h_j} &\to u && \textit{in } L^q(\Omega) \quad \forall q < p^*, \\ \nabla_{h_j} u_{h_j} &\rightharpoonup \nabla u && \textit{in } [L^p(\Omega)]^d, \\ \mathcal{J}_{h_j}(u_{h_j}) &\to \mathcal{J}(u), \\ \sum_{e \in \mathcal{E}_h^B} \int_e h_e^{1-p} |u_{h_j} - g|^p \, \mathrm{d}S + \sum_{e \in \mathcal{E}_h^I} \int_e h_e^{1-p} |[u_{h_j}]|^p \, \mathrm{d}S &\to 0 \end{aligned}$$

as $j \to \infty$. Moreover, any accumulation point of the set $\{u_h\}_{h>0}$ is a minimizer of $\mathcal{J}$ from (5.1.2) over $W^{1,p}_g(\Omega)$. If $\xi \to f(\xi, v, x)$ is strictly convex for all $(v, x) \in \mathbb{R} \times \Omega$, then we have

$$\|u - u_{h_j}\|_{W^{1,p}_h(\Omega)} \to 0$$

as $j \to \infty$. If the minimizer u is unique, then the whole set $\{u_h\}_{h>0}$ converges.

The following results will be quite useful in later use of the DF-FE derivative and the discontinuous Ritz method. First, we give conditions to guarantee equivalence of the semi-norms $\|\nabla_h \cdot\|$ and $|\cdot|_{W^{1,p}_h(\Omega)}$ on V_h. To this end, we first quote a discrete inf-sup condition from Buffa and Ortner [10].

Lemma 5.1 (Lemma A.2 from [10])**.** *Let $1 \leq p < \infty$ and q be its Hölder conjugate. Then there exists a constant $C > 0$ independent of h such that*

$$\inf_{v_h \in V_h} \sup_{\varphi_h \in V_h} \frac{\int_\Omega v_h \varphi_h}{\|v_h\|_{L^p(\Omega)} \|\varphi_h\|_{L^q(\Omega)}} \geq C. \tag{5.3.6}$$

We first show the boundedness of $\|\nabla_h \cdot\|_{L^p(\Omega)}$ from $|\cdot|_{W^{1,p}_h(\Omega)}$ on $W^{1,p}(\mathcal{T}_h)$.

Lemma 5.2. *Let $1 < p < \infty$. Then there exists a constant $C > 0$ independent of h such that*

$$\|\nabla_h v\|_{L^p(\Omega)} \lesssim |v|_{W_h^{1,p}(\Omega)} \qquad \forall v \in W^{1,p}(\mathcal{T}_h), \tag{5.3.7}$$

Proof. Choose q to be the Hölder conjugate of p and let $v \in W^{1,p}(\mathcal{T}_h)$ and $\varphi_h \in [V_h]^d$. From (5.3.1) and the standard trace and inverse inequalities we have

$$\begin{aligned}
\int_\Omega \nabla_h v \cdot \varphi_h \, \mathrm{d}x &= -\sum_{e\in\mathcal{E}_h^I} \int_e [v]\{\varphi_h \cdot \nu_e\} \, \mathrm{d}S + \sum_{T\in\mathcal{T}_h} \int_T \nabla v \cdot \varphi_h \, \mathrm{d}x \\
&\leq \sum_{e\in\mathcal{E}_h^I} \int_e h_e^{\frac{1-p}{p}} |[v]| \cdot h_e^{\frac{1}{q}} |\{\varphi_h \cdot \nu_e\}| \, \mathrm{d}S + \sum_{T\in\mathcal{T}_h} \|\nabla v\|_{L^p(T)} \|\varphi_h\|_{L^q(T)} \\
&\leq \sum_{e\in\mathcal{E}_h^I} \int_e \left(h_e^{1-p}|[v]|^p\right)^{\frac{1}{p}} \left(h_e|\{\varphi_h \cdot \nu_e\}|^q\right)^{\frac{1}{q}} \mathrm{d}S + \|\nabla v\|_{L^p(\Omega)} \|\varphi_h\|_{L^q(\Omega)} \\
&\leq \left(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p} \|[v]\|_{L^p(e)}^p\right)^{\frac{1}{p}} \left(\sum_{e\in\mathcal{E}_h^I} h_e \|\{\varphi_h \cdot \nu_e\}\|_{L^q(e)}^q\right)^{\frac{1}{q}} \\
&\quad + \|\nabla v\|_{L^p(\Omega)} \|\varphi_h\|_{L^q(\Omega)} \\
&\lesssim \left(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p} \|[v]\|_{L^p(e)}^p\right)^{\frac{1}{p}} \|\varphi_h\|_{L^q(\Omega)} + \|\nabla v\|_{L^p(\Omega)} \|\varphi_h\|_{L^q(\Omega)} \\
&\lesssim |v|_{W_h^{1,p}(\Omega)} \|\varphi_h\|_{L^q(\Omega)}.
\end{aligned}$$

Since $\nabla_h v \in V_h$, it follows from Lemma 5.1 that

$$\|\nabla_h v\|_{L^p(\Omega)} \lesssim \sup_{\varphi_h \in V_h} \frac{\int_\Omega \nabla_h v \cdot \varphi_h}{\|\varphi_h\|_{L^q(\Omega)}} \lesssim |v|_{W_h^{1,p}(\Omega)}.$$

which is exactly (5.3.7). □

We next show the boundedness of $|\cdot|_{W_h^{1,p}(\Omega)}$ from $\|\nabla_h \cdot\|_{L^p(\Omega)}$ on V_h with the help of interior penalty terms. This is because $\{v_h\} = 0$ on $e \in \mathcal{E}_h^I$ in (5.3.1) does not imply that $v_h\big|_{T^+} = 0$ or $v_h\big|_{T^-} = 0$ on e.

Lemma 5.3. *Let $1 < p < \infty$. Then there exists a constant $C, \gamma^* > 0$ independent of h such that for all $\gamma_e > \gamma^*$ the following holds for every $v_h \in V_h$:*

$$|v_h|_{W_h^{1,p}(\Omega)} \leq C\|\nabla_h v_h\|_{L^p(\Omega)} + C\bigg(\sum_{e\in\mathcal{E}_h^I} \gamma_e h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\bigg)^{1/p}. \tag{5.3.8}$$

Proof. Choose q to be the Hölder conjugate of p and $v_h \in V_h$. From (5.3.1) we have

$$\int_\Omega \nabla_h v_h \cdot \varphi_h \,\mathrm{d}x = -\sum_{e\in\mathcal{E}_h^I}\int_e [v_h]\{\varphi_h \cdot \nu_e\}\,\mathrm{d}S + \int_\Omega \nabla v_h \cdot \varphi_h \,\mathrm{d}x \tag{5.3.9}$$

for every $\varphi_h \in [V_h]^d$. Let $\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})$ where $\mathcal{P}_h$ is the local L^2 projection onto $\mathcal{T}_h$ defined by

$$\int_T \mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\cdot\varphi_h\,\mathrm{d}x = \int_T \nabla v_h|\nabla v_h|^{p-2}\cdot\varphi_h\,\mathrm{d}x$$

for all $\varphi_h \in V_h$ and $T \in \mathcal{T}_h$. Choosing $\varphi_h = \mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})$ in (5.3.9) gives us

$$\begin{aligned}\int_\Omega \nabla_h v_h \cdot \mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\,\mathrm{d}x = &-\sum_{e\in\mathcal{E}_h^I}\int_e [v_h]\{\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\cdot\nu_e\}\,\mathrm{d}S \\ &+ \int_\Omega \nabla v_h \cdot \mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\,\mathrm{d}x.\end{aligned} \tag{5.3.10}$$

By the stability of $\mathcal{P}_h$ we obtain

$$\begin{aligned}\int_\Omega \nabla_h v_h \cdot \mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\,\mathrm{d}x &\leq \|\nabla_h v_h\|_{L^p(\Omega)}\|\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\|_{L^q(\Omega)} \\ &\leq \|\nabla_h v_h\|_{L^p(\Omega)}\|\nabla v_h|\nabla v_h|^{p-2}\|_{L^q(\Omega)} \\ &\leq \|\nabla_h v_h\|_{L^p(\Omega)}\|\nabla v_h\|_{L^p(\Omega)}^{p-1}.\end{aligned} \tag{5.3.11}$$

By the standard trace and inverse inequalities for DG functions, there exists $C_1 > 0$ independent of h such that

$$\begin{aligned}
&\sum_{e\in\mathcal{E}_h^I}\int_e [v_h]\{\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\cdot\nu_e\}\,\mathrm{d}S \qquad (5.3.12)\\
&\qquad\le \Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\Big)^{\frac{1}{p}}\Big(\sum_{e\in\mathcal{E}_h^I} h_e\|\{\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\cdot\nu_e\}\|_{L^q(e)}^q\,\mathrm{d}S\Big)^{\frac{1}{q}}\\
&\qquad\le C_1\Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\Big)^{\frac{1}{p}}\|\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\|_{L^q(\Omega)}\\
&\qquad\le C_1\Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\Big)^{\frac{1}{p}}\|\nabla v_h\|_{L^p(\Omega)}^{p-1}.
\end{aligned}$$

By the properties of P_h we have

$$\int_\Omega \nabla v_h\cdot\mathcal{P}_h(\nabla v_h|\nabla v_h|^{p-2})\,\mathrm{d}x = \int_\Omega \nabla v_h\cdot\nabla v_h|\nabla v_h|^{p-2}\,\mathrm{d}x = \|\nabla v_h\|_{L^p(\Omega)}^p. \qquad (5.3.13)$$

Thus by (5.3.10)-(5.3.13) and dividing by $\|\nabla v_h\|_{L^p(\Omega)}^{p-1}$ we have

$$\|\nabla_h v_h\|_{L^p(\Omega)} \ge -C_1\Big(\sum_{e\in\mathcal{E}_h^I} h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\Big)^{\frac{1}{p}} + \|\nabla v_h\|_{L^p(\Omega)}.$$

Choosing $\gamma^* = C_1^p + 1$ and $C = 1/2$ gives us the result. The proof is complete. □

We can also prove a compactness result using the DG-FE numerical derivatives. For this, we cite a discrete compactness result from Buffa and Ortner [10].

Lemma 5.4 (Theorem 5.2 and Lemma 8 from [10]). *For $1 < p < \infty$ and $0 < h < 1$, let $v^h \in W^{1,p}(\mathcal{T}_h)$ such that*

$$\sup_{0<h<1}\big(\|v^h\|_{L^1(\Omega)} + |v^h|_{W_h^{1,p}(\Omega)}\big) < \infty. \qquad (5.3.14)$$

Then there exists a sequence $h_j \searrow 0$ and a function $v \in W^{1,p}(\Omega)$ such that

$$v^{h_j} \to v \quad \text{in } L^q(\Omega) \qquad \forall\, 1 \le q < p^*, \tag{5.3.15a}$$

$$v^{h_j} \to v \quad \text{in } L^q(\partial\Omega) \qquad \forall\, 1 < q < q^*, \tag{5.3.15b}$$

$$\nabla v^{h_j} + R(v^{h_j}) \rightharpoonup \nabla v \ \text{ in } [L^p(\Omega)]^d, \tag{5.3.15c}$$

where p^* is the Sobolev conjugate of p defined in (5.3.4) and q^* is defined by

$$q^* = \begin{cases} \frac{(d-1)p}{d-p} & \text{if } p < d, \\ \infty & \text{if } p \ge d. \end{cases} \tag{5.3.16}$$

We are now ready to state our compactness result, which differs from Lemma 5.4 by controlling DG functions using the DG-FE numerical gradient as well as showing their DG-FE numerical gradients weakly converge.

Theorem 5.2. *Let $1 < p < \infty$. There exists $\gamma^* > 0$ such that for any $\gamma_e > \gamma^*$ suppose there is a family $\{v_h\}_{h\in(0,1)}$ with $v_h \in V_h$ and*

$$\sup_{0<h<1} \left(\|v_h\|_{L^1(\partial\Omega)} + \|\nabla_h v_h\|_{L^p(\Omega)} + \left(\sum_{e\in\mathcal{E}_h^I} \gamma_e h_e^{1-p} \|[v_h]\|^p_{L^p(e)} \right)^{1/p} \right) < \infty. \tag{5.3.17}$$

Then there exists a sequence $h_j \searrow 0$ and a function $v \in W^{1,p}(\Omega)$ such that

$$v_{h_j} \to v \quad \text{in } L^q(\Omega) \qquad \forall\, 1 \le q < p^*, \tag{5.3.18a}$$

$$v_{h_j} \to v \quad \text{in } L^q(\partial\Omega) \qquad \forall\, 1 < q < q^*, \tag{5.3.18b}$$

$$\nabla_{h_j} v_{h_j} \rightharpoonup \nabla v \ \text{ in } [L^p(\Omega)]^d, \tag{5.3.18c}$$

where p^* is the Sobolev conjugate of p defined in (5.3.4) and q^* is defined in (5.3.16).

Proof. From Lemma 5.3, we have

$$|v_h|_{W_h^{1,p}(\Omega)} \lesssim \|\nabla_h v_h\|_{L^p(\Omega)} + \left(\sum_{e\in\mathcal{E}_h^I} \gamma_e h_e^{1-p}\|[v_h]\|_{L^p(e)}^p\right)^{1/p}.$$

for sufficiently large γ^* which shows v_h is uniformly bounded in $W^{1,p}(\mathcal{T}_h)$. By the Poincarè-Fredrichs inequality, Theorem 10.6.12 of [8], we have

$$\|v_h\|_{L^1(\Omega)} \lesssim \|v_h\|_{L^p(\Omega)} \lesssim \|v_h\|_{L^p(\partial\Omega)} + |v_h|_{W_h^{1,p}(\Omega)}.$$

Therefore the family $\{v_h\}_{h\in(0,1)}$ satisfies the hypothesis of Lemma 5.4, which gives us everything in the theorem except for (5.3.18c). However, by Lemma 5.1, we have that $\nabla_h v_h = \nabla v_h + R(v_h)$ and consequently (5.3.18c). The proof is complete.

□

5.4 Numerical Experiments

In the section we give some numerical tests to show the effectiveness of the proposed discontinuous Ritz method. Our prototypical example is the following $p-$Laplace energy:

$$\mathcal{J}^p(v) = \int_\Omega \frac{1}{p}|\nabla v|^p - Fv\,\mathrm{d}x, \tag{5.4.1}$$

minimized over the space $W_g^{1,p}(\Omega)$. The Euler Lagrange equation 1.1.9 of $\mathcal{J}^p$ yields the following $p-$Laplace problem:

$$-\operatorname{div}(|\nabla u|^{p-2}\nabla u) = F \text{ in } \Omega, \tag{5.4.2a}$$

$$u = g \;\text{ on } \partial\Omega. \tag{5.4.2b}$$

Table 5.1: The L^p and $W_h^{1,p}$ errors and rates of convergence in h for the Discontinuous Ritz method 5.2 applied to $\mathcal{J}^p(\cdot)$ from (5.4.1) where $p = 2.5$.

$1/h$	$\|u - u_h\|_{L^p(\Omega)}$	rate	$\|\nabla u - \nabla_h u_h\|_{L^p(\Omega)}$	rate
10	5.12e-03	0.00	1.10e-01	0.00
20	3.06e-03	0.74	5.51e-02	0.99
40	1.67e-03	0.88	2.76e-02	1.00
80	8.74e-04	0.93	1.38e-02	1.00
160	4.49e-04	0.96	6.92e-03	1.00
320	2.28e-04	0.98	3.46e-03	1.00

Note that $p = 2$ gives the standard Poisson problem; however, here p can be any number such that $1 < p < \infty$. We will test two cases: one for $p > 2$ and another for $p < 2$.

5.4.1 Test 1: $p > 2$

Let $p = 2.5$, $d = 1$, $\Omega = (0, 1)$ and $g = x$. Choose $F(x) = -\sqrt{3}x^2$ so that the exact solution is $u(x) = x^3$. Table 5.1 shows the errors and rates in L^p and $W^{1,p}$-norm for $u - u_h$ where u_h is the discrete minimizer $u_h \in V_h$ of (5.2.6) with polynomial degree $k = 1$. The numerical results clearly indicate that the method is converging to the correct solution and we have optimal order convergence in the $W^{1,p}$ semi-norm, but we have sub-optimal convergence rate in the L^p norm.

5.4.2 Test 2: $p < 2$

Let $p = 1.5$, $d = 1$, $\Omega = (0, 1)$ and $g = 0$. Choose $F(x)$ such that the exact solution is $u(x) = \sin(\pi x)$. Note $w := |\nabla u|^{p-2}\nabla u = \frac{\sqrt{\pi}\cos(\pi x)}{\sqrt{|\cos(\pi x)|}}$ is not classically differentiable since $\cos(\pi x)$ is both positive and negative on $(0, 1)$, but $w \in W^{1,q}(\Omega)$ for all $1 < q < 2$ with ∇w having a discontinuity at $x = 0.5$. Table 5.2 shows the L^p and $W^{1,p}$ errors and rates of convergence for the method. We see that the rates of convergence are suboptimal for both the L^p and $W^{1,p}$ errors. This is most likely due to the degeneracy

Table 5.2: The L^p and $W_h^{1,p}$ errors and rates of convergence in h for the Discontinuous Ritz method 5.2 applied to $\mathcal{J}^p(\cdot)$ from (5.4.1) where $p = 1.5$.

$1/h$	$\|u - u_h\|_{L^p(\Omega)}$	rate	$\|\nabla u - \nabla_h u_h\|_{L^p(\Omega)}$	rate
10	8.50e-02	0.00	3.19e-01	0.00
20	5.77e-02	0.56	2.06e-01	0.63
40	4.03e-02	0.52	1.38e-01	0.57
80	2.85e-02	0.50	9.56e-02	0.53
160	2.02e-02	0.50	6.69e-02	0.51
320	1.43e-02	0.50	4.72e-02	0.51

of the PDE since largest error occurs at $x = 0.5$ where w is 0. This claim is supported by Figure 5.1.

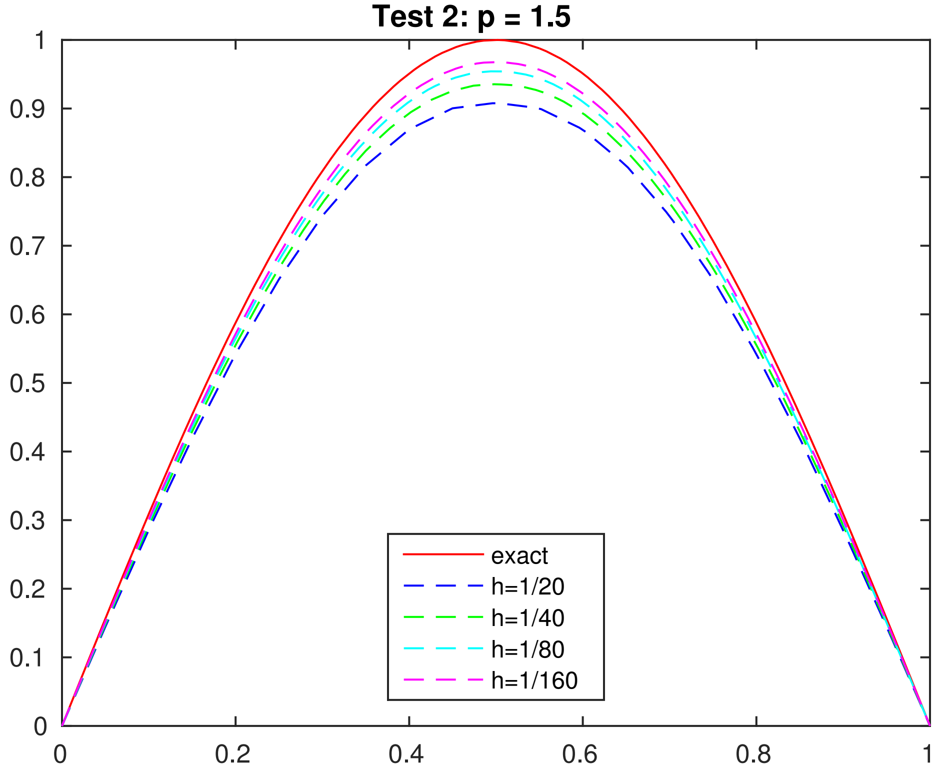


Figure 5.1: The plots of u and u_h where u is the exact minimizer for $\mathcal{J}^p(\cdot)$ from (5.4.1) with $p = 1.5$ and u_h is the discrete minimizer from (5.2.6). Here $h = 1/20, 1/40, 1/80, 1/160$.

Chapter 6

A MATLAB Toolbox for the Discontinuous Galerkin Finite Element Numerical Calculus

6.1 Introduction

The goal of this chapter is to showcase a MATLAB toolbox created to implement the discontinuous Galerkin finite element (DG-FE) numerical calculus introduced by Feng, Lewis, and Neilan (see [25]). The DG-FE derivatives were already defined in Defintion 5.1, but we include the definition here for completeness. Let $i = 1, \ldots, d$. we define three numerical partial derivative operators in the x_i coordinate $\partial^+_{h,x_i}, \partial^-_{h,x_i}, \partial_{h,x_i} : W^{1,p}(\mathcal{T}_h) \to V_h$, by

$$\int_\Omega \partial^\pm_{h,x_i}(v)\varphi_h \,\mathrm{d}x = \sum_{e\in\mathcal{E}_h} \int_e \mathcal{Q}^\pm_i(v)\nu^i_e[\varphi_h]\,\mathrm{d}S - \sum_{T\in\mathcal{T}_h} \int_T v\partial_{x_i}\varphi_h \,\mathrm{d}x \quad \forall\varphi_h \in V_h, \tag{6.1.1}$$

$$\partial_{h,x_i}(v) = \frac{1}{2}\left(\partial^+_{h,x_i}(v) + \partial^-_{h,x_i}(v)\right), \tag{6.1.2}$$

where $\mathcal{Q}^{\pm}$ and $\mathcal{Q}$ are from (5.2.1). Because V_h is a discontinuous piecewise polynomial space, we can also write $\partial^{+}_{h,x_i}, \partial^{-}_{h,x_i}, \partial_{h,x_i}$ locally on every element $T \in \mathcal{T}_h$ as

$$\int_T \partial^{\pm}_{h,x_i}(v)\big|_T \varphi_h \,\mathrm{d}x = \int_{\partial T} \mathcal{Q}^{\pm}_i(v)\nu^i_e \varphi_h \,\mathrm{d}S - \int_T v\partial_{x_i}\varphi_h \,\mathrm{d}x \quad \forall \varphi_h \in \mathbb{P}_r(T), \tag{6.1.3}$$

$$\partial_{h,x_i}(v)\big|_T = \frac{1}{2}\left(\partial^{+}_{h,x_i}(v)\big|_T + \partial^{-}_{h,x_i}(v)\big|_T\right). \tag{6.1.4}$$

The gradients $\nabla^{\pm}_h$ and ∇_h are defined similarly from (5.2.4) and (5.2.5). From the definition, we see that the DG-FE derivatives enforce that the integration by parts formula holds for every $\varphi_h \in V_h$. This DG-FE derivatives allow the authors to build a numerical calculus giving analogs of the classical calculus results such as the product rule, chain rule, and integration by parts.

This chapter presents a MATLAB toolbox that implements the DG-FE derivative on the variety of 1-D and 2-D domains. The code is written to accept a variety of algebraic or coordinate defined functions and has selectable options such as higher order quadrature approximation or degree of polynomial interpolation. The motivation for such a toolbox is two-fold. First, such a toolbox will make methods that use the DG-FE calculus, such as the Discontinuous Ritz method in Chapter 5 and the Dual Wind DG method [25, 37], easier to implement. Second, this toolbox can be used as an educational tool to provide students with a different notion of discrete derivatives. Currently, the main discrete derivative used in any undergraduate numerical analysis course is the finite difference numerical eriative. The DG-FE toolbox allows students to compute numerical derivatives for a variety of functions including those whose function values are only given at grid points.

The remainder of this chapter is organized as follows. In Section 6.2 we list a few properties of the DG-FE derivative, including the chain rule, product rule, and integration by parts formulas and give two examples of convergent numerical methods for the Poisson problem using the DG-FE derivative. In Section 6.3, we provide the documentation of the Matlab toolbox, including step by step examples on how to use the software.

6.2 Properties and Applications of the DG-FE Derivative

6.2.1 Properties of the DG-FE Derivative

In order to show that the DG-FE derivative indeed is a calculus - that is, it has its own versions of basic calculus rules such as the chain rule, product rule, and integration by parts - we cite such results as well as some characterizations of the DG-FE derivative from [25].

First, we note that the DG-FE derivative of any finite element function is equivalent to its weak derivative.

Theorem 6.1 (Corollary 4.1 of [25]). *Let $v_h \in S_h$. Then $\partial^+_{h,x_i} v_h, \partial^-_{h,x_i} v_h, \partial_{h,x_i} v_h = \partial_{x_i} v_h$ for every $i = 1, \dots, d$.*

Next, we have a version of the product rule and the chain rule where the DG-FE derivative is the L^2 projection of the product rule and chain rule onto the space V_h.

Theorem 6.2 (Theorem 4.2 of [25]). *Let $1 \leq p < \infty$ and $\mathcal{P}_h : L^p(\Omega) \to V_h$ be the local L^2 projection onto V_h, that is, for every $T \in \mathcal{T}_h$ we have*

$$\int_T \mathcal{P}_h(v) \varphi_h \, \mathrm{d}x = \int_T v \varphi_h \, \mathrm{d}x \qquad \forall \varphi_h \in V_h.$$

Let $F \in C^1(\mathbb{R})$ with $F' \in L^\infty(\mathbb{R})$. For $u, v \in W^{1,p}(\mathcal{T}_h) \cap C^0(\Omega)$, there holds, for $i = 1, \dots, d$,

$$\begin{aligned} \partial_{h,x_i}(uv) &= \mathcal{P}_h(u \partial_{x_i} v + v \partial_{x_i} u), \\ \partial_{h,x_i} F(u) &= \mathcal{P}_h(F'(u) \partial_{x_i} u). \end{aligned}$$

We also have an integration by parts formula when integrating against discrete functions.

Theorem 6.3 (Theorem 4.4 of [25]). *Let $v_h, \varphi_h \in V_h$, then the following integration by parts formulas hold:*

$$\int_\Omega \partial^{\pm}_{h,x_i} v_h \varphi_h \,\mathrm{d}x = -\int_\Omega v_h \partial^{\mp}_{h,x_i} \varphi_h \,\mathrm{d}x + \sum_{e\in\mathcal{E}_h^B} \int_e v_h \varphi_h \nu_e^i \,\mathrm{d}S,$$
$$\int_\Omega \partial_{h,x_i} v_h \varphi_h \,\mathrm{d}x = -\int_\Omega v_h \partial_{h,x_i} \varphi_h \,\mathrm{d}x + \sum_{e\in\mathcal{E}_h^B} \int_e v_h \varphi_h \nu_e^i \,\mathrm{d}S,$$

where ν_e^i is the i^{th} component of ν_e.

While in general broken Sobolev functions $v \in W^{1,p}(\mathcal{T}_h)$ do not have weak derivatives, the DG-FE derivative represents the L^2 projection of their distributional derivative on to V_h.

Theorem 6.4 (Proposition 4.2 of [25]). *Let $v \in W^{1,p}(\mathcal{T}_h)$. Then $\partial_{h,x_i} v$ coincides with the L^2 projection of $\mathcal{D}_{x_i} v$ onto V_h where $\mathcal{D}_{x_i} v$ is the distributional derivative of v, that is,*

$$\int_\Omega \partial_{h,x_i} v \varphi_h \,\mathrm{d}x = \langle \mathcal{D}_{x_i} v, \varphi_h \rangle \qquad \forall \varphi_h \in V_h.$$

6.2.2 Applications of the DG-FE Derivative

In this subsection we describe two convergent methods which were developed in [25], with the help of the DG-FE derivative. Both methods were formulated for the following model problem:

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned}$$

To introduce these methods, we first define a jump operator $j_h : W^{1,p}(\mathcal{T}_h) \to V_h$ as follows:

$$\sum_{T\in\mathcal{T}_h} \int_T j_h(v)\varphi_h \,\mathrm{d}x = \sum_{e\in\mathcal{E}_h^I} \int_e \frac{\gamma_e}{h_e}[v][\varphi_h]\,\mathrm{d}S + \sum_{e\in\mathcal{E}_h^B} \int_e \frac{\gamma_e}{h_e} v\varphi_h \,\mathrm{d}S \qquad \forall \varphi_h \in V_h.$$

The first method seeks a function $u_h \in V_h$ such that

$$\int_\Omega \nabla_h u_h \cdot \nabla_h \varphi_h \,\mathrm{d}x - \sum_{e\in\mathcal{E}_h^B} \int_e \nabla_h u_h \cdot \nu_e \varphi_h \,\mathrm{d}S + \int_\Omega j_h(u_h)\varphi_h \,\mathrm{d}x = \int_\Omega f\varphi_h \,\mathrm{d}x \quad (6.2.1)$$

for all $\varphi_h \in V_h$. This method is equivalent to the well-known LDG method for the model problem [17] and converges provided $\gamma_e > 0$.

The next method, the symmetric dual-wind discontinuous Galerkin (DWDG) method [37], is constructed from the ground up using the DG-FE gradients. The DWDG method seeks $u_h \in V_h$ such that

$$\frac{1}{2}\int_\Omega \left(\nabla_h^+ u_h \cdot \nabla_h^+ \varphi_h + \nabla_h^- u_h \cdot \nabla_h^- \varphi_h\right) \mathrm{d}x + \int_\Omega j_h(u_h)\varphi_h \,\mathrm{d}x = \int_\Omega f\varphi_h \,\mathrm{d}x \qquad (6.2.2)$$

for all $\varphi_h \in V_h$. Note the sided gradients ∇_h^+ and ∇_h^-, instead of the central gradient, are used in the formulation. If $\gamma_e > 0$, then the method is well-posed and convergent. Moreover, if $\mathcal{T}_h$ is quasi-uniform and if each element $T \in \mathcal{T}_h$ has at most one boundary edge, then the method is well-posed and converges provided $\gamma_e > -C_*$ for some constant $C_* > 0$ independent of h. Thus one could set $\gamma_e \equiv 0$, that is, ignoring the penalty terms, and still achieve convergence.

6.3 Documentation of the Matlab Toolbox

In this section we give the full documentation of the DG-FE Matlab toolbox. We first describe the algorithm used to implement the DG-FE numerical derivative on given meshes in one and two dimensions. We include a function call list with all of the available options included. In addition, we provide a few examples of how to compute the DG-FE numerical derivative. For 2-D polynomial evaluation, the Matlab function `polyval2`, written by Salmon Rodgers [50], is used in the toolbox. The toolbox is freely available and can be downloaded from [54]. Lastly, we refer to the central DG-FE partial derivative ∂_{h,x_i} as the "full" derivative in this documentation.

6.3.1 Algorithm

Since $\partial^{\pm}_{h,x_i} v$ and φ_h from (6.1.3) both lie in a finite dimenion space $\mathbb{P}_r(K)$, we can numerically compute $\partial^{\pm}_{h,x_i} v$ by converting (6.1.3) into a matrix problem. Let $\{\varphi_i\}_{i=1}^{NE}$ be a basis for $\mathbb{P}_r(K)$ with $\dim(\mathbb{P}_r(K)) = NE$. Then there exists constants $\{\alpha_i\}_{i=1}^{NE}$ such that $\partial^{\pm}_{h,x_i} v = \sum_{i=1}^{NE} \alpha_i \varphi_i$. This basis expression and linearity of the integral turns (6.1.3) into

$$\alpha_i \int_K \varphi_i \varphi_j \,\mathrm{d}x = \int_{\partial K} Q_i^{\pm}(v) \cdot \eta_K^{(i)} \varphi_j \,\mathrm{d}S - \int_K v \cdot \frac{\partial \varphi_j}{\partial x_i} \,\mathrm{d}x + \int_{\partial K \setminus \partial \Omega} \gamma_{i,e}[v] \cdot \varphi_j \,\mathrm{d}x, \tag{6.3.1}$$

for all $j = 1, 2, \ldots, NE$. Letting $A_K = \left[\int_K \varphi_i \varphi_j \,\mathrm{d}x \right]_{i,j=1}^{NE}$ and β_j be equal to the right hand side of (6.3.1), we get the matrix problem

$$A_K \alpha_K = \beta_K,$$

where $\alpha_K = \big[\alpha_i\big]_{i=1}^{NE}$ and $\beta_K = \big[\beta_j\big]_{j=1}^{NE}$. For efficiency, we map all interior integration to the simplex domain $K' = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ for all } i = 1, 2, \ldots, n, \ ||x||_1 < 1\}$ via an affine transformation.

For this computational implementation, the mass matrix A is dependent on the local basis chosen for V_h. In the 1-D case, we have used the Legendre polynomials since they give an orthogonal mass matrix while we have chosen the monomial basis $\{1, x, y, x^2, xy, y^2, x^3, x^2y, y^2x, y^3, \ldots\}$ in the 2-D case for its simplicity.

6.3.2 One Dimension Case

Syntax

The mian components of this package is broken up into two functions. `meshNumericalDerivative` is the function called to compute the numerical derivative data (plus, minus, or full) and `postProcessing` interpolates this data and can be used to evaluate the derivative a specified points.

meshNumericalDerivative

The proper syntax for calling this function is:

```
polydata = meshNumericalDerivative(v, degree, mesh, ...),
```

where `polydata` is the numerical derivative outputted as a matrix with each column being the coefficient vector $a_{K'}$ corresponding to the polynomial basis on K'. The three required arguments are

- `v`: The input function. Note that the function must be continuous and have a weak derivative on each element in order to have a numerical derivative. `v` can be several classes including:

  A discrete function represented as an array. This input should correspond to $\{v(x_j)\}_{j=0}^{J}$ where $\{x_j\}_{j=0}^{J}$ is the mesh. Note that `meshNumericalDerivative` will only accept a vector the same size as the mesh inputted and computes a cubic interpolation of the data before calculating the numerical derivative. If the user has any more information about the discrete function, midpoints for example, then the user should create a function that interpolates the data before input. See `interp1` in the MATLAB documentation for more information.

  A matrix having the same size as `polydata` which contains the representation in polynomial basis on K'. This is the preferred and fastest method when computing second, third, and higher ordered derivatives.

A `function_handle` loaded into the workspace.

A string of the `function_handle` that is not loaded into the workspace, e.g., `'exp'`.

Note: Any non-discrete function inputed must be a function of solely one argument. If the inputted function has more than one input - `func1(x,a)`, then the user should create an anonymous function to handle the extra argument - `testfunc = @(x) func1(x,3)`.

Since the input `v` can be discontinuous across edges, the user must be careful about how the outputs of `v` are displayed. Take for example the function

$$g(x) = \begin{cases} 0 & \text{if } x \geq 0, \\ 1 & \text{if } x < 0. \end{cases}$$

In order for the numerical derivative to work properly, $g(0)$ must output both 0 and 1. To clarify which value is which, the output will be a 2×1 column vector where the top element is the left-hand limit and the bottom value is the right-hand limit. For this example, `g(0) = [1;0]`. Because of this the code has been adapted to take a variety of inputs. `v` can either always output 2×1 column vector that is the same value on the interior of an element and (possibly) different on the edge, or an overloaded function that returns a single value on the interior and a 2×1 column vector on the edge. The tool does not make use of inputting vectors into `v`, so this is okay. Also scalar-valued, piecewise continuous functions work as expected.

- `degree`: A nonnegative integer that specifies the degree of polynomial space. Increase this number to increase the accuracy of the numerical derivative or if the user wants to take multiple derivatives of the inputed function. The `degree` should be non-negative integer not exceeding 5.

- mesh: An array that specifies the mesh of the domain - where the elements in the vector are increasing. In this case, the whole domain is an interval and each element is a subinterval. Note that the mesh does not have to be uniform - only increasing.

There are also optional arguments which must come in pairs. The first argument is the argument identifier and the second is the argument value. For example,

```
poly_derv = meshNumericalDerivative('exp', 3,
               [-1:.1:1],'accuracy','high').
```

Here 'accuracy' is the argument identifier and 'high' is the argument value.

- 'derivative': (Default value: 'full'). Specifies whether the output is $\partial_h^+ v$, $\partial_h^- v$, or $\partial_h v$. The possible argument values are 'plus', 'minus', and 'full' which represent $\partial_h^+ v$, $\partial_h^- v$, or $\partial_h v$ respectively.

- 'accuracy': (Default value: 'medium'). Specifies the accuracy of the numerical quadrature used when computing the numerical derivative. This program takes advantage of a Gaussian quadrature scheme.

  'low': A low order, 2 point, Gaussian quadrature scheme. While this is the fastest of the three options, only use it with a degree of 0, 1, or 2.

  'medium': A medium order, 3 point, Gaussian quadrature scheme. Use it with a degree of 2 or 3.

  'high': A high order, 5 point, Gaussian quadrature scheme. Use it with a degree of 3 or 4.

  'vhigh': A very high order, 7 point, Gaussian quadrature scheme. Use it with a degree of 5.

postProcessing

The proper syntax for calling this function is:

`pointvalues = postProcessing(polydata,mesh,format,x)`,

where `pointvalues` is a matrix of function values of the derivative for input `x`. The 4 required arguments are

- `polydata`: The derivative polynomial data outputted from `meshNumericalDerivative`.
- `mesh`: The same mesh as used in `polydata`.
- `format`: This should be either a 0 or 1 depending on what the user wants outputted when one of the values in `x` is on the boundary of an element.

  0 will give only scalar outputs with values on the edge being the average of the left and right hand value, that is,

  $$\frac{1}{2}\left(\lim_{x\to z^-}\partial_h v(x)+\lim_{x\to z^+}\partial_h v(x)\right).$$

  Note that values on the interior of an element will only have one value regardless. When 0 is specified, `x` can be any size matrix and `size(pointvalues) == size(x)`.

  1 will give a 3×1 column vector/matrix output. If the inputed value is on an edge of an element then the vector will be

  $$\begin{bmatrix} \lim_{x\to z^-}\partial_h v(x) \\ \lim_{x\to z^+}\partial_h v(x) \\ \frac{1}{2}\left(\lim_{x\to z^-}\partial_h v(x)+\lim_{x\to z^+}\partial_h v(x)\right) \end{bmatrix},$$

  giving the left, right, and average derivatives values. If the inputed value is on the interior of an element then the vector will be the same value repeated thrice. Note that if 1 is specified `x` may only be inputted as a row vector and not a matrix with more than one row.

- `x`: The specified points where the derivative will be evaluated. Look at the `format` paragraph above for how `x` should be entered.

Examples

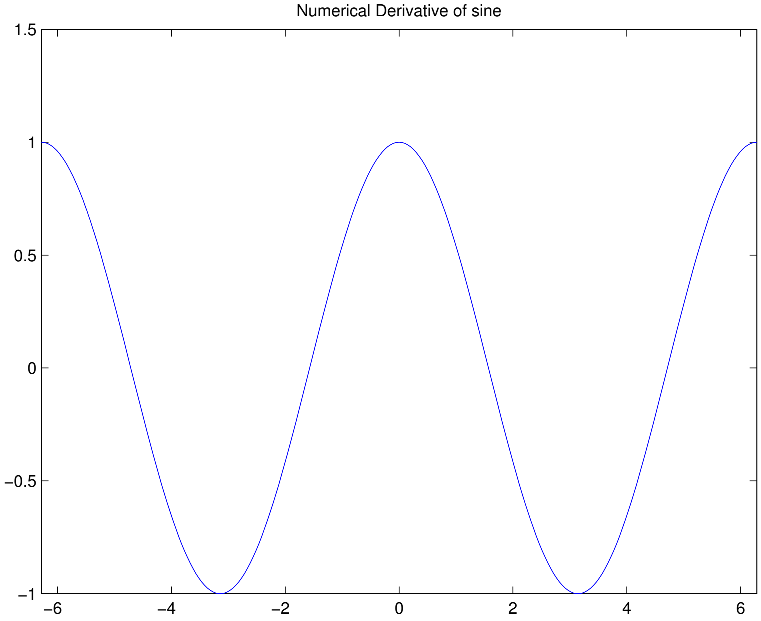
Below are a few examples of the code in action.

Derivative of $\sin(x)$ over $[-2\pi, 2\pi]$

Since the sine function is already a built-in MATLAB function, all we need to do is create a mesh and run the code. For this example we will use a quadratic approximation and will plot the derivative after.

```
>> mesh = [-2*pi:pi/8:2*pi];
>> poly_data = meshNumericalDerivative('sin',2,mesh);
>> poly = @(x) postProcessing(poly_data,mesh,0,x);
>> fplot(poly,[-2*pi,2*pi]);title('Numerical Derivative of sine');
```

Here is the graph outputted which does map cosine, the derivative of sine.

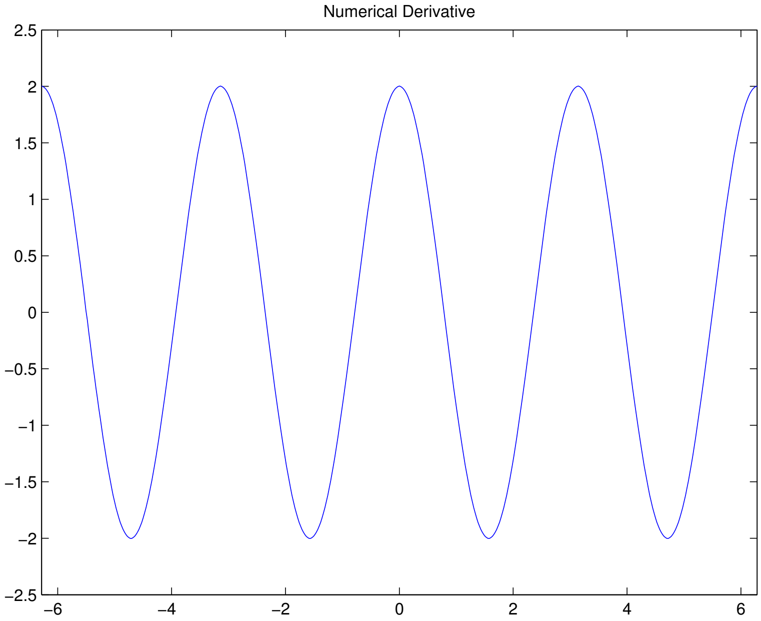


Derivative of $\sin(2x)$ over $[-2\pi, 2\pi]$

For this example we will use a `function_handle` that is loaded into the workspace. Again we will use quadratic approximation. We will also plot the computed derivative.

```
>> mesh = [-2*pi:pi/8:pi];
>> test_sine = @(x) sin(2*x);
>> poly_data = meshNumericalDerivative(test_sine,2,mesh);
>> poly = @(x) postProcessing(poly_data,mesh,0,x);
>> fplot(poly,[-2*pi,2*pi]);title('Numerical Derivative');
```

Here is the graph outputted which does map the derivative: $2\cos(2x)$.

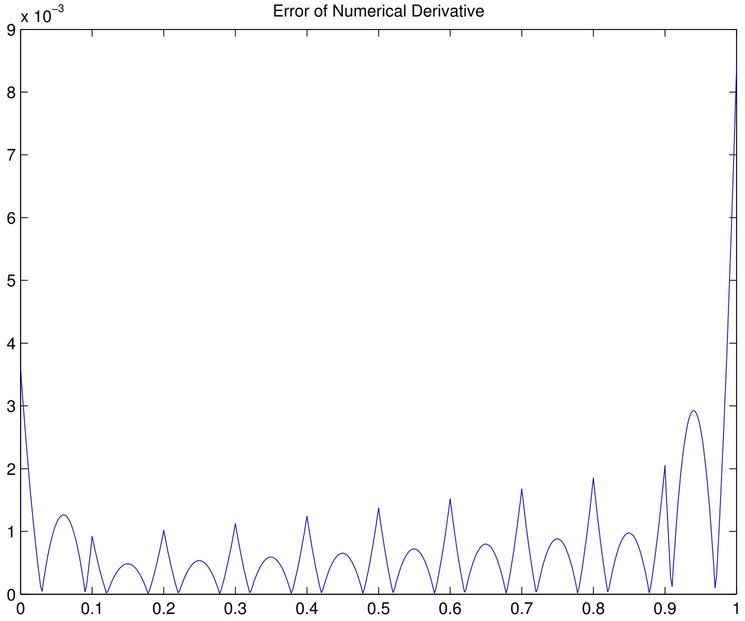


Derivative of e^x using a discrete input

Here we will compute the numerical derivative of e^x using only a discrete number of points. This time we will use cubic approximation. We will also plot the error between the numerical derivative and the true derivative (e^x).

```
>> mesh = [0:.1:1];
>> vp = exp(mesh);
>> poly_data = meshNumericalDerivative(vp,3,mesh);
>> error = @(x) abs(postProcessing(poly_data,mesh,0,x)-exp(x));
>> fplot(error,[0,1]);title('Error of Numerical Derivative');
```

Below is the graph of the error:

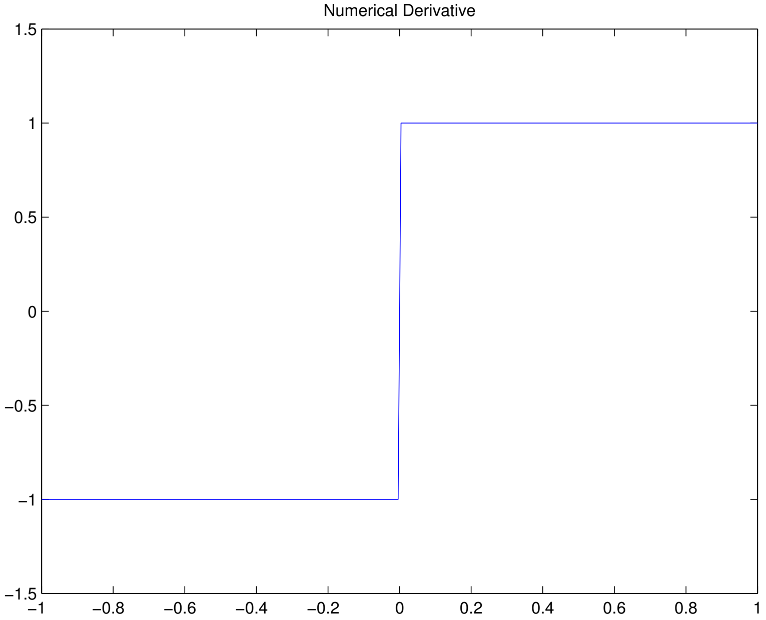


Derivative of $|x|$ with jumps

Here we will compute the numerical derivative of $|x|$ on the interval $[-1, 1]$. We will use a linear approximation and will adjust the `'accuracy'` to `'low'`. Note the derivative has a discontinuity at $x = 0$, with the left-hand limit being -1 and the right-hand limit being 1. We will show how the `'format'` argument in `postProcessing` allows the user to choose which value to use at $x = 0$.

```
>> mesh = [-1:.1:1];
>> poly_data = meshNumericalDerivative('abs',1,mesh,'accuracy','low');
>> poly = @(x) postProcessing(poly_data,mesh,1,x);
>> poly(0)


ans =


   -1.0000

    1.0000

   -0.0000


>> poly2 = @(x) postProcessing(poly_data,mesh,0,x);
>> fplot(poly2,[-1,1]);title('Numerical Derivative');
```

Also the plot of the numerical derivative.



Numerical Derivative of Functions with no Weak Derivative

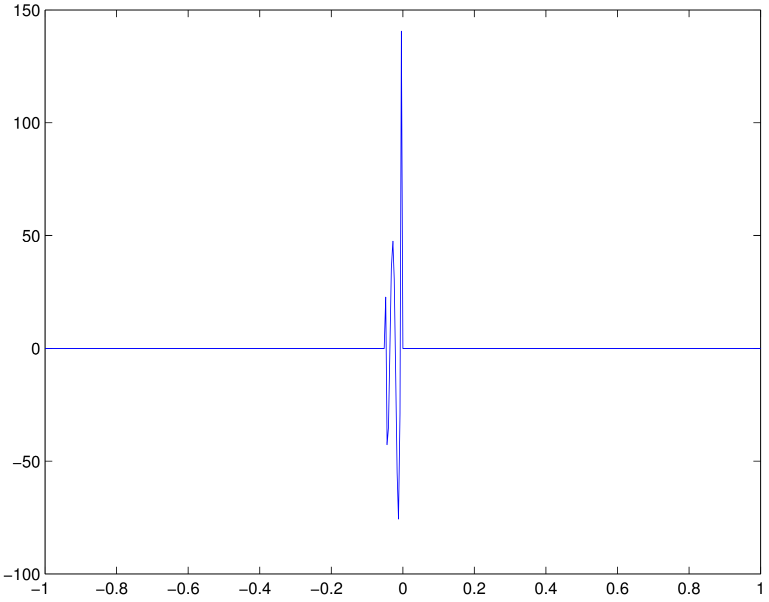
This example demonstrates what information the numerical derivative possesses when the inputted function does not have a weak derivative. Take for example, the

Heaviside function

$$H(x) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases}$$

This function has a distributional derivative, namely the Dirac-Delta function δ, but the Dirac-Delta function is not $L^1_{\text{loc}}([-1,1])$ so $H(x)$ does not have a weak derivative. We will calculate a high order numerical derivative of the Heaviside function.

```
>> mesh = [-1:.1:1];
>> poly_data = meshNumericalDerivative('heaviside',4,mesh,...
'accuracy','high');
>> poly = @(x) postProcessing(poly_data,mesh,0,x);
>> fplot(poly,[-1,1])
```



Since the Dirac-Delta function is not $L^1_{\text{loc}}([-1,1])$, no information is gained by looking at the plot. However, if we look at our numerical derivative, call it ψ, in the sense of distributions, our output does approximate the Dirac-Delta function for appropriate functions $\varphi \in C^1(\Omega)$ with compact support. Indeed

$$\int_{-1}^{1} \psi(x)\varphi(x)\,\mathrm{d}x = \langle \psi, \varphi \rangle \approx \langle \delta, \varphi \rangle = \varphi(0)$$

as shown in a few examples below. For our candidate φ we will use a cubic order polynomial `c1Cpt(x,x_0,x_1,y)`, which is supported inside $[x_0, x_1]$ and satisfies $\varphi((x_0+x_1)/2) = y$. Note that each integral should be the candidate function evaluated at 0.

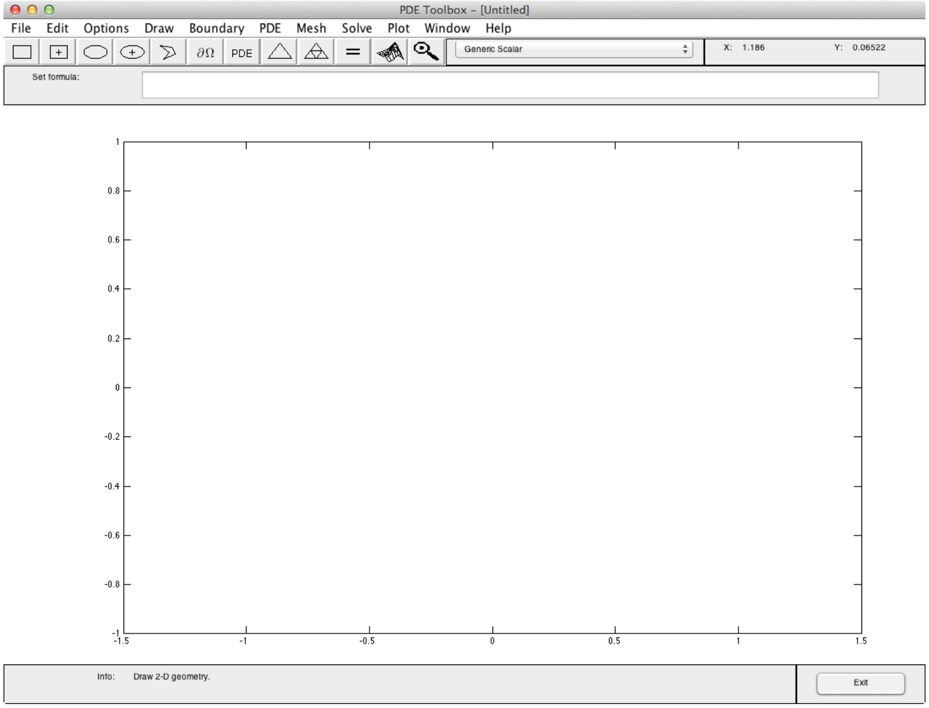
```
>> format long
>> test1 = @(x) c1Cpt(x,-.5,.5,2) .* poly(x);
>> integral(test1,-1,1)


ans =


   1.999999254385638


>> c1Cpt(0,-.5,.5,2)


ans =


     2


>> test2 = @(x) c1Cpt(x,-.43,.75,-30) .* poly(x);
>> integral(test2,-1,1)


ans =


 -57.304586160017060
```

```
>> c1Cpt(0,-.43,.75,-30)

ans =

 -57.304610703052091
```

6.3.3 Two Dimension Case

Prerequisites

The mesh in the 1-dimensional case is trivial to create since it is a vector of points and geometrically the edge of each element is the boundary points. The 2-D case requires some additional prep work before we can dive into the numerical derivative. This tollbox is designed to work with MATLAB's Partial Differential Equation Toolbox. Specifically, we will be using the `pdetool` command to create an appropriate mesh. The Partial Differential Equation Toolbox documentation is a great place to look for more information, but, for now, we will create a simple mesh. First we will type `pdetool` into the command window and this screen below will appear.

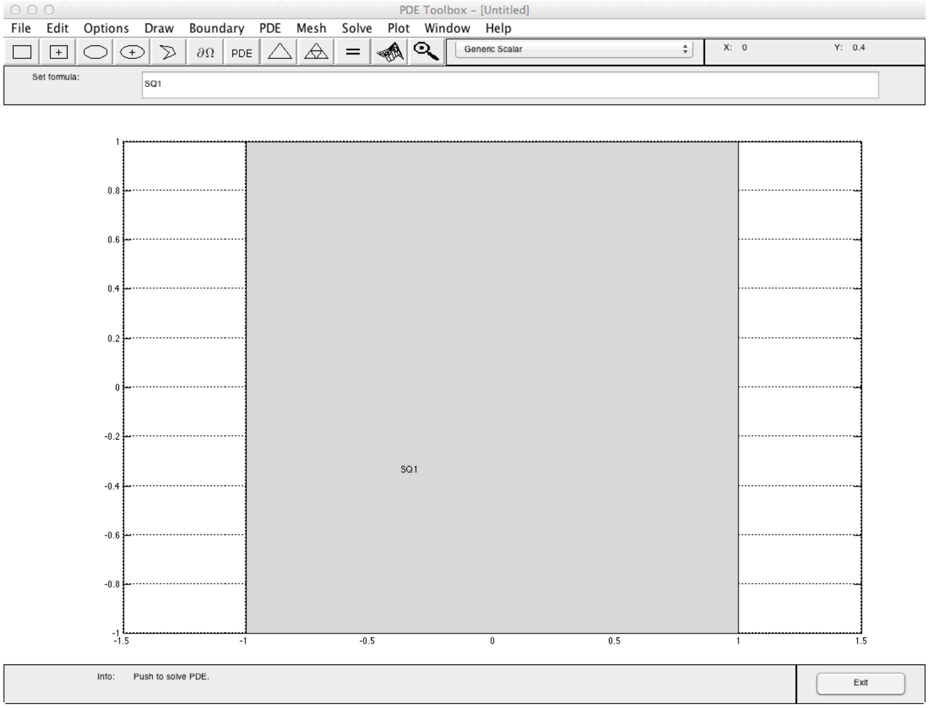


Generic Scalar

X: 1.186 Y: 0.06522

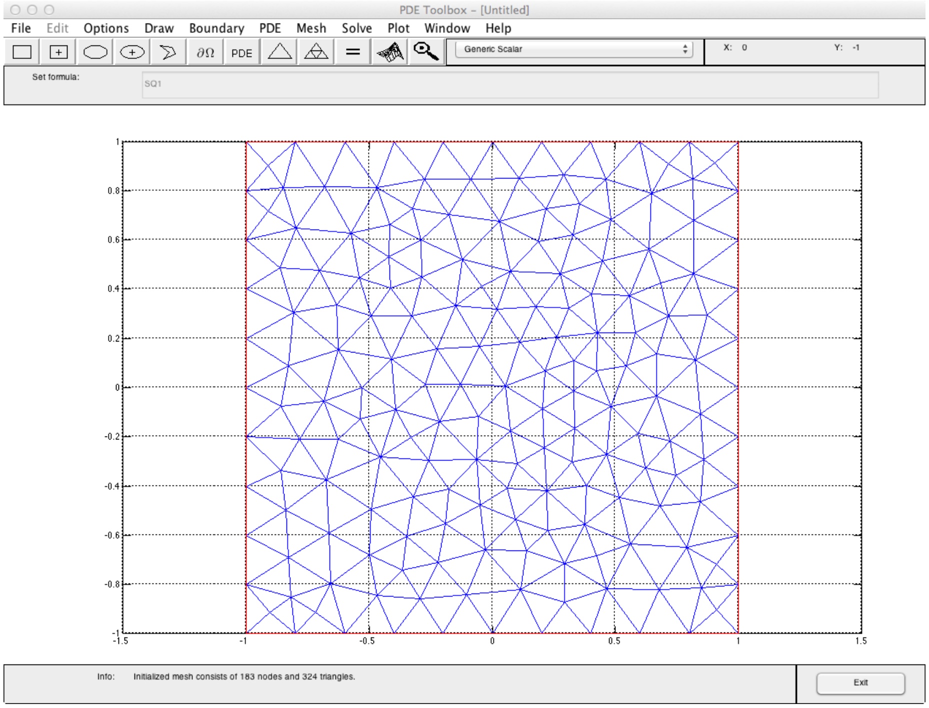
Info: Draw 2-D geometry.

Next we will enable “Grid” and “Snap” from the Options menu. Finally using any of the five geometry tools from the row below the menu bar we will create our shape. In this case we will create a square with a side length of two and centered at the origin.



Info: Push to solve PDE.

To create the mesh, we will click “Initialize Mesh” under the Mesh menu. The “Parameters…” option is useful for further mesh properties.



Info: Initialized mesh consists of 183 nodes and 324 triangles.

Finally we will click "Initialize Mesh" under the Mesh menu and confirm with "OK". We have now created a point matrix (p), an edge matrix (e), and a triangle matrix (t) from the mesh generator. The point and triangle matrices are required for the numerical derivative toolbox; however, since this toolbox is for the Discontinuous Galerkin framework we need the edge data for every element instead of just the boundary data (which is what `e` provides). To get this data run

```
ee = gatherEdgeData( p, e, t);
```

where `p, e, t` are from above. This command will output a cell with 6×3 matrix for each triangle in (t). Each column in the 6×3 matrix corresponds to the following edge data:

```
% | 1 first point of edge                         |
% | 2 second point of edge                        |
% | 3 triangle # of the edges shared partner      |
% | 4 x-coordinate of unit outward normal vector  |
% | 5 y-coordinate of unit outward normal vector  |
% | 6 element normal vector == edge normal vector |
% |   is the edge on the boundary of the mesh     |
```

Syntax

Similar to 1-D, the main components are the following two functions. `meshNumericalDerivative` is the function called to compute the numerical derivative data (plus, minus, or full) and `postProcessing` interpolates this data and can be used to evaluate the derivative a specified point.

`meshNumericalDerivative` is the function called to compute the numerical derivative (plus, minus, or full). The proper syntax for calling this function is:

```
polydata = meshNumericalDerivative( v, p, ee, t, degree,
                    position,...)
```

where polydata is the numerical derivative outputted as a matrix with each column being the coefficient vector $a_{K'}$ associated with the polynomial basis on K'. The required arguments are

- v: The input function. Note that the function must be continuous and have a weak derivative on each element in order to have a numerical derivative. v can be several classes including:

  A discrete function represented as a vector. This input should correspond to $\{v(x_j, y_j)\}_{j=0}^{N}$ where $[x_j; y_j]$ is the jth column of p and N is the total number of points in the mesh. Note a cubic interpolation of the data will be computed before calculating the numerical derivative. See griddata in the MATLAB documentation for more information.

  A matrix having the same size as polydata which contains the representation coefficients in the polynomial basis on K'. This is the preferred and fastest method when computing second, third, and higher ordered derivatives.

  A function_handle loaded into the workspace.

  A string of the function_handle that is not loaded into the workspace, e.g., 'harmonic' (See appendix).

  Note: Any non-discrete function inputed must be a function of exactly two arguments. If the inputted function has more than two inputs - func2(x,y,a), then the user should create an anonymous function to handle the extra argument - testfunc = @(x) func2(x,y,3). Also, these functions must be vectorized, accept column vectors for x and y, and output solutions as a matrix with number of rows the same as the number of rows in x.

  This algorithm does not explicitly use the function values at each vertex since there can be five or more elements that share the same vertex. However, we cannot avoid edges as they are exclusively used in the calculation of the edge integrals. Since the inputted function can be discontinuous over an edge, the output can be a 1×2 vector where the first output corresponds with the element which has the lower global labeling of the two. Inside each element, the inputted function is continuous can output a scalar or 1×2 vector (having the same value in both entries); the program can accept both options.

- `t,p,ee`: The triangle matrix, point matrix, and edge data respectively created from the `pdetool` mesh generator and `gatherEdgeData`. Please see 6.3.3 for more information.

- `degree`: A non-negative integer that specifies the degree of polynomial space. Here the sum of the degrees of each component specifies the polynomial space. For example, the polynomials x^3, x^2y, xy^2, y^3 are all degree 3. Increase this number to increase the accuracy of the numerical derivative or if the user wants to take multiple derivatives of the inputed function. The `degree` should be a non-negative integer not exceeding 4.

- `position`: This flag of `1` or `2` specifies which the program will compute $\frac{\partial}{\partial x}$ or $\frac{\partial}{\partial y}$ respectfully.

There are also optional arguments which must come in pairs. The first argument is the argument identifier, and the second is the argument value. For example,

```
poly_derv = meshNumericalDerivative('exp', 3,
            [-1:.1:1],'format','first').
```

Here `'format'` is the argument identifier and `'first'` is the argument value.

- `'format'`: (Default value `'average'`) A string that determines what value of the numerical derivative uses on the boundary of an element. The numerical derivative may have a discontinuity on the boundary of an element (for example, the numerical derivative of $f(x) = |x|$ is discontinuous at 0), but must be continuous on each element. Because of this there can be "two" values of the numerical derivative at each boundary point. Since this does not create a function suitable for plotting, the `'format'` option allows us to weight the value of each point. The possible options are `'first'`, `'last'`, `'average'`, and `same`. Below is a description of each option with an example point z and example function $\partial_h v$.

`'left'` specifies the left-hand value of the function, that is, $\lim_{x \to z^-} \partial_h v(x)$.

`'right'` specifies the right-hand value of the function, that is, $\lim_{x \to z^+} \partial_h v(x)$.

`'average'` specifies the average value of the function, that is,

$$\frac{1}{2}\left(\lim_{x \to z^-} \partial_h v(x) + \lim_{x \to z^+} \partial_h v(x)\right).$$

`'same'` handles the case where the polynomial derivative has discontinuities over the edge. The polynomial outputted is column vector-valued. If evaluated on an edge the top element is the left-hand limit and the bottom is the right-hand limit. If evaluated on the interior of an element both the top and bottom values are the same.

- `'derivative'`: (Default value: `'full'`). Specifies whether the output is $\partial_h^+ v$, $\partial_h^- v$, or $\partial_h v$. The possible argument values are `'plus'`, `'minus'`, and `'full'` which will represent $\partial_h^+ v$, $\partial_h^- v$, or $\partial_h v$ respectively.

- `'accuracy'`: (Default value: `'medium'`). Specifies the accuracy of the numerical quadrature used when computing the numerical derivative. This toolbox takes advantage of a Gaussian quadrature scheme.

  `'low'`: A low order, 2 point, Gaussian quadrature scheme. While this is the fastest of the three options, only use it with a `degree` of 0, 1, or 2.

  `'medium'`: A medium order, 3 point, Gaussian quadrature scheme. Use it with a `degree` of 3 or 4.

  `'high'`: A high order, 5 point, Gaussian quadrature scheme. Use it with a `degree` of 5 through 8.

postProcessing

The proper syntax for calling this function is:

```
pointvalues = postProcessing(polydata,p,t,format,x,y),
```

where `pointvalues` is a matrix of derivative values for input `x`. The 4 required arguments are

- `polydata`: The derivative polynomial data outputted from `meshNumericalDerivative`
- `p,t`: The same p and t used in the creation of `polydata`.
- `format`: This should be either a 0 or 1 depending on what the user wants outputted when one of the values in `x` is on the boundary of an element.

  0 will give only scalar outputs with values on the edge being the average of the left and right-hand value, that is,

  $$\frac{1}{2}\left(\partial_h v|_{K^+}(x) + \partial_h v|_{K^-}(x)\right).$$

  Note that values on the interior of an element will only have one value regardless. When 0 is specified, `x` can be any size matrix and `size(pointvalues) == size(x)`.

  1 will give a 3×1 column vector/matrix output. If the inputed value is on an edge of an element, then the vector will be

  $$\begin{bmatrix} \partial_h v|_{K^-}(x) \\ \partial_h v|_{K^+}(x) \\ \frac{1}{2}\left(\partial_h v|_{K^+}(x) + \partial_h v|_{K^-}(x)\right) \end{bmatrix},$$

  giving the left, right, and average derivatives values. If the given value is on the interior of an element then the vector will be the same value repeated thrice.

Note that if 1 is specified `x` may only be inputted as a row vector and not a matrix with two or more rows.

- `x,y`: The specified x and y values where the derivative will be evaluated. Look at the `format` paragraph above for how `x,t` should be entered.

Examples

Partial Derivative of $x^3 + xy + y^2$ in x direction

We will first convert the polynomial $f(x, y) = x^3+xy+y^2$ into an anonymous function and then compute its numerical partial derivative $\frac{\partial f}{\partial x}$ which is $\frac{\partial f}{\partial x} = 3x^2 + y$ using a quadratic approximation. We will then compute the L^2 error of the derivative and its approximation. Our domain in this problem is a square centered at (.5,.5) with side length of 1.

```
>> f = @(x,y) x.^3 + x.*y + y.^2;
>> polydata = meshNumericalDerivative(f,p,ee,t,2,1);
>> error = @(x,y) (postProcessing(polydata,p,t,0,x,y) ...
- derivative(x,y)).^2;
>> integral2(error,0,1,0,1)^(1/2)


ans =


   1.6531e-12
```

Laplacian of a Harmonic Function

This example shows the accuracy of second order numerical derivatives. We will take the numerical Laplacian of the harmonic function $\ln(|x|)$. Note the Laplacian of $\ln(|x|)$ is identically 0. In order to speed up the computation, we will not use the first derivatives' global function, instead we use the local functions defined on each element. Our mesh is a square centered at $(.75, .75)$ and side lengths of one. We will output the time it takes to compute each derivative, then to test accuracy, we will measure the error with the discrete norm defined by

$$||f|| = \frac{1}{N}\left(\sum_{j=0}^{N} f(x_j, y_j)^2\right)^{\frac{1}{2}},$$

where (x_j, y_j) are the points given in the point matrix `p` and N is the number of points in the point matrix.

```
>> tic;[dx] = meshNumericalDerivative(’harmonic’,p,ee,t,3,1);toc;
Elapsed time is 0.121740 seconds.
>> tic;[dxx] = meshNumericalDerivative(dx,p,ee,t,3,1);toc;
Elapsed time is 0.246442 seconds.
>> tic;[dy] = meshNumericalDerivative(’harmonic’,p,ee,t,3,2);toc;
Elapsed time is 0.109847 seconds.
>> tic;[dyy] = meshNumericalDerivative(dy,p,ee,t,3,2);toc;
Elapsed time is 0.251037 seconds.
>> f = @(x,y) postProcessing(dxx,p,t,0,x,y) ...
+ postProcessing(dyy,p,t,0,x,y);
>> norm(f(p(1,:)’,p(2,:)’))/188

ans =

   8.0858e-05
```

Chapter 7

Future Directions

The research in this dissertation has extensions into other topics and directions. The following are a few projects that I intend to pursue in the near future.

- *Extension of the linear elliptic PDEs in non-divergence form to Hamilton-Jacobi-Bellman (HJB) equations.* As previously mentioned, a direct application of linear elliptic PDEs in non-divergence form is their use in HJB equations. As shown in Section 3.6, it is expected that this technique may work for some simple problems but will fail to approximate more complex problems. The addition of the vanishing moment method of Chapter 3 will most likely be needed to ensure convergence. I aspire to study the well-posedness of the numerical framework given in Section 3.6 and its convergence to the viscosity solutions of HJB equations.

- *Lower order norm error estimates for the IP-DG solutions to linear elliptic PDEs in non-divergence form.* As of yet both papers [22, 28] fail to give any L^2 or H^1 error estimates. Since the formal PDE adjoint to non-divergence form PDEs with continuous coefficients is not well understood, deriving these error estimates is not easy using the standard Nitsche's duality argument. Some nonstandard duality argument must be developed; one such argument involves an approximate dual problem. I intend to follow this path to see if any meaningful results may arise from it.

- *Local discontinuous Galerkin methods for linear elliptic PDEs in non-divergence form.* The IP-DG methods developed in Chapter 2 require the degree of polynomials in V_h to be at least two. This is to ensure the space V_h approximates the space $W^{2,p}(\Omega)$ in the discrete $W^{2,p}$ norm as $h \to 0$. The local discontinuous Galerkin (LDG) framework; however, does not normally require quadratic polynomials since the primal problem is converted into a mixed formulation. I intend to construct and analyze LDG schemes for linear elliptic PDEs in non-divergence form that hopefully converge on piecewise constant and linear DG spaces.

- L^∞ *error estimates for the IP-DG solutions to linear elliptic PDEs in non-divergence form.* For some problems, such as finite element approximations the the obstacle problem, L^2 estimates have yet to be proven; however, maximum norm error estimates have been shown for linear finite elements on highly structured meshes. The key to this proof is that both the PDE problem and the discrete problem preserve a maximum principle. Since the non-divergence PDE (2.1.1) possesses a maximum principle, I want to see if the same technique can be applied to the IP-DG or finite element discretizations of linear elliptic PDEs in non-divergence form.

- Γ-*convergence of the enhanced finite element method to the Maniá example and other functionals exhibiting the Lavrentiev Gap Phenomenon.* Chapter 4 gives the Γ-convergence of $\mathcal{J}_h^\alpha$ to $\mathcal{J}$ for the Maniá example only if we minimize over the infinite dimensional space $\mathcal{A}_\infty$. In practice this is not good enough since we want to minimize over the finite element subspace V_h. I plan to continue this line of work and show that the sequence of finite element minimizers u_h of $\mathcal{J}_h^\alpha$ do indeed converge to minimizer for the Maniá example and others examples that show the gap phenomenon.

- *Extension of the enhanced finite element method to non-conforming discretizations.* Ortner (cf. [48]) showed that for some functions exhibiting the Lavrentiev gap phenomenon the use of non-conforming discretization, such as Crouzeix-Raviart, is enough to overcome the gap phenomenon and no change in the functional needed. I aim to combine the powers of both non-conforming discretizations, for example, discontinuous Galerkin finite element method, and our enhanced finite element method to create robust schemes to tackle many, if not all, of the problems exhibiting the gap phenomenon.

- *Extension of the discontinuous Ritz methods to other problems.* While we have proven the convergence of the discontinuous Ritz method for the class of convex and coercive problems, I would like to implement the discontinuous Ritz framework for other problems. One such example is the total variation problem for image denoising by Rudin, Osher, and Fatemi [51], wihch minimizes an energy over BV(Ω), the space of functions having bounded total variation.

Bibliography

[1] Alnæs, M. S., Blechta, J., Hake, J., Johansson, A., Kehlet, B., Logg, A., Richardson, C., Ring, J., Rognes, M. E., and Wells, G. N. (2015). The fenics project version 1.5. *Archive of Numerical Software*, 3(100). Available from: https://fenicsproject.org/. 23, 27

[2] Angermann, L. and Henke, C. (2015). Interpolation, Projection and Hierarchical Bases in Discontinuous Galerkin Methods. *Numer. Math. Theory Methods Appl.*, 8(3):425–450. Available from: http://dx.doi.org/10.4208/nmtma.2015.m1305. 32

[3] Babuška, I., Caloz, G., and Osborn, J. E. (1994). Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM Journal on Numerical Analysis*, 31(4):945–981. 8

[4] Bai, Y. and Li, Z.-P. (2006). A truncation method for detecting singular minimizers involving the Lavrentiev phenomenon. *Math. Models Methods Appl. Sci.*, 16(6):847–867. Available from: http://dx.doi.org/10.1142/S0218202506001376. 18, 118, 119, 121, 122, 123, 134

[5] Ball, J. M. and Knowles, G. (1987). A numerical method for detecting singular minimizers. *Numer. Math.*, 51(2):181–197. Available from: http://dx.doi.org/10.1007/BF01396748. 17, 18, 118, 121, 122

[6] Braides, A. (2002). *Gamma-convergence for Beginners*, volume 22. Clarendon Press. 15

[7] Brenner, S. C. (2011). C 0 interior penalty methods. In *Frontiers in Numerical Analysis-Durham 2010*, pages 79–147. Springer. 97

[8] Brenner, S. C. and Scott, L. R. (2008). *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition. Available from: http://dx.doi.org/10.1007/978-0-387-75934-0. 28, 36, 38, 72, 153

[9] Brezis, H. (2010). *Functional analysis, Sobolev spaces and partial differential equations*. Springer Science & Business Media. 31

[10] Buffa, A. and Ortner, C. (2009). Compact embeddings of broken sobolev spaces and applications. *IMA journal of numerical analysis*, 29(4):827–855. 14, 20, 140, 141, 143, 146, 147, 148, 151

[11] Caffarelli, L. A. and Gutiérrez, C. E. (1997). Properties of the solutions of the linearized monge-ampere equation. *American Journal of Mathematics*, pages 423–465. 6

[12] Carstensen, C. and Ortner, C. (2009). Computation of the lavrentiev phenomenon. OxMOS Preprint. 17, 18, 116, 118

[13] Carstensen, C. and Ortner, C. (2010). Analysis of a class of penalty methods for computing singular minimizers. *Computational Methods in Applied Mathematics Comput. Methods Appl. Math.*, 10(2):137–163. 118

[14] Chen, Z. and Chen, H. (2004). Pointwise error estimates of discontinuous galerkin methods with penalty for second-order elliptic problems. *SIAM Journal on Numerical Analysis*, 42(3):1146–1166. 38, 39, 45, 46, 48, 49

[15] Chiarenza, F., Frasca, M., and Longo, P. (1993). $W^{2,p}$-solvability of the Dirichlet problem for nondivergence elliptic equations with VMO coefficients. *Transactions of the American Mathematical Society*, 336(2):841–853. 8

[16] Ciarlet, P. (2002). *3. Conforming Finite Element Methods for Second-Order Problems*, chapter 3, pages 110–173. Society for Industrial and Applied Mathematics (SIAM). Available from: `http://epubs.siam.org/doi/abs/10.1137/1.9780898719208.ch3`. 56

[17] Cockburn, B. and Shu, C.-W. (1998). The local discontinuous galerkin method for time-dependent convection-diffusion systems. *SIAM Journal on Numerical Analysis*, 35(6):2440–2463. 161

[18] Crandall, M. G. and Lions, P.-L. (1983). Viscosity solutions of hamilton-jacobi equations. *Transactions of the American Mathematical Society*, 277(1):1–42. 81

[19] Dacorogna, B. (2015). *Introduction to the calculus of variations*. Imperial College Press, London, third edition. 4, 9, 11

[20] Douglas Jr, J., Dupont, T., Percell, P., and Scott, R. (1979). A family of $c1$ finite elements with optimal approximation properties for various galerkin methods for 2nd and 4th order problems. *RAIRO-Analyse numérique*, 13(3):227–255. 30

[21] Evans, L. C. (2010). *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition. 2, 130

[22] Feng, X., Hennings, L., and Neilan, M. (2017). Finite element methods for second order linear elliptic partial differential equations in non-divergence form. *Mathematics of Computation*. 13, 19, 23, 25, 27, 66, 72, 73, 74, 183

[23] Feng, X. and Jensen, M. (2017). Convergent semi-lagrangian methods for the monge–ampère equation on unstructured grids. *SIAM Journal on Numerical Analysis*, 55(2):691–712. 6

[24] Feng, X., Kao, C.-Y., and Lewis, T. (2013). Convergent finite difference methods for one-dimensional fully nonlinear second order partial differential equations. *Journal of Computational and Applied Mathematics*, 254:81–98. 82

[25] Feng, X., Lewis, T., and Neilan, M. (2016a). Discontinuous galerkin finite element differential calculus and applications to numerical solutions of linear and nonlinear partial differential equations. *Journal of Computational and Applied Mathematics*, 299:68–91. 20, 144, 146, 157, 158, 159, 160

[26] Feng, X. and Neilan, M. (2009a). Mixed finite element methods for the fully nonlinear monge–ampère equation based on the vanishing moment method. *SIAM Journal on Numerical Analysis*, 47(2):1226–1250. 82

[27] Feng, X. and Neilan, M. (2009b). Vanishing moment method and moment solutions forfully nonlinear second order partial differential equations. *Journal of Scientific Computing*, 38(1):74. Available from: http://dx.doi.org/10.1007/s10915-008-9221-9. 82

[28] Feng, X., Neilan, M., and Schnake, S. (2016b). Interior penalty discontinuous galerkin methods for second order linear non-divergence form elliptic pdes. *ArXiv e-prints*. arXiv preprint. 27, 38, 183

[29] Feng, X. and Schnake, S. (2016). An enhanced finite element method for a class of variational problems exhibiting the lavrentiev gap phenomenon. *ArXiv e-prints*. arXiv preprint. 120

[30] Fleming, W. H. and Soner, H. M. (2010). *Controlled Markov Processes and Viscosity Solutions (Stochastic Modelling and Applied Probability)*. Springer, softcover reprint of hardcover 2nd ed. 2006 edition. Available from: http://amazon.com/o/ASIN/1441920781/. 6

[31] Foss, M. (2003). Examples of the Lavrentiev phenomenon with continuous Sobolev exponent dependence. *J. Convex Anal.*, 10(2):445–464. 130, 134, 135

[32] Foss, M., Hrusa, W. J., and Mizel, V. J. (2003). The Lavrentiev gap phenomenon in nonlinear elasticity. *Arch. Ration. Mech. Anal.*, 167(4):337–365. Available from: http://dx.doi.org/10.1007/s00205-003-0249-6. 17, 116

[33] Furihata, D. and Matsuo, T. (2010). *Discrete variational derivative method: A structure-preserving numerical method for partial differential equations*. CRC Press. 14

[34] Georgoulis, E. H., Houston, P., and Virtanen, J. (2009). An a posteriori error indicator for discontinuous galerkin approximations of fourth-order elliptic problems. *IMA journal of numerical analysis*, page drp023. 30

[35] Gilbarg, D. and Trudinger, N. S. (2001). *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin. Reprint of the 1998 edition. 8, 36, 66, 90

[36] Han, Q. and Lin, F. (1997). *Elliptic partial differential equations*, volume 1 of *Courant Lecture Notes in Mathematics*. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI. 9

[37] Lewis, T. and Neilan, M. (2014). Convergence analysis of a symmetric dual-wind discontinuous galerkin method. *Journal of Scientific Computing*, 59(3):602–625. 158, 161

[38] Lewis, T. L. (2013). *Finite Difference and Discontinuous Galerkin Finite Element Methods for Fully Nonlinear Second Order Partial Differential Equations*. PhD thesis, The University of Tennessee. 82, 110

[39] Li, Z. (1995). A numerical method for computing singular minimizers. *Numerische Mathematik*, 71(3):317–330. 118, 121

[40] Li, Z. (1996). A theorem on lower semicontinuity of integral functionals. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, 126(02):363–374. 122

[41] Li, Z. P. (1992). Element removal method for singular minimizers in variational problems involving Lavrentiev phenomenon. *Proc. Roy. Soc. London Ser. A*, 439(1905):131–137. Available from: http://dx.doi.org/10.1098/rspa.1992.0138. 18, 118, 121

[42] Maniá, B. (1934). Soppa un esempio di lavrentieff. *Ball. Unione Mat. Ital*, 13:147–153. 17, 116

[Mathworks] Mathworks. Matlab. Available from: https://www.mathworks.com/products/matlab.html. 23

[44] Maugeri, A., Palagachev, D. K., and Softova, L. G. (2000). *Elliptic and parabolic equations with discontinuous coefficients*. Wiley Online Library. 8

[45] Negrón Marrero, P. V. (1990). A numerical method for detecting singular minimizers of multidimensional problems in nonlinear elasticity. *Numerische Mathematik*, 58(1):135–144. Available from: http://dx.doi.org/10.1007/BF01385615. 18

[46] Nitsche, J. A. and Schatz, A. H. (1974). Interior estimates for ritz-galerkin methods. *Mathematics of Computation*, 28(128):937–958. 32

[47] Nochetto, R. H. and Zhang, W. (2014). Discrete abp estimate and convergence rates for linear elliptic equations in non-divergence form. *arXiv preprint arXiv:1411.6036*. 12, 25

[48] Ortner, C. (2011). Nonconforming finite-element discretization of convex variational problems. *IMA J. Numer. Anal.*, 31(3):847–864. Available from: http://dx.doi.org/10.1093/imanum/drq004. 18, 117, 118, 185

[49] Rivière, B. (2008). *Discontinuous Galerkin methods for solving elliptic and parabolic equations*, volume 35 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Theory and implementation. Available from: http://dx.doi.org/10.1137/1.9780898717440. 37, 73

[50] Rodgers, S. (2007). 2d weighted polynomial fitting and evaluation. Available from: https://www.mathworks.com/matlabcentral/fileexchange/13719-2d-weighted-polynomial-fitting-and-evaluation. 162

[51] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268. 185

[52] Schatz, A. H. (1998). Pointwise error estimates, superconvergence, and extrapolation. *LECTURE NOTES IN PURE AND APPLIED MATHEMATICS*, pages 237–248. 38, 39

[53] Schatz, A. H. and Wang, J. P. (1996). Some new error estimates for Ritz-Galerkin methods with minimal regularity assumptions. *Math. Comp.*, 65(213):19–27. Available from: http://dx.doi.org/10.1090/S0025-5718-96-00649-7. 61, 67

[54] Schnake, S. (2014). A matlab toolbox for the discontinuous galerkin finite element numerical calculus. Available from: https://bitbucket.org/stefanschnake/dgfenumericcalculus. 162

[55] Smears, I. and Süli, E. (2013). Discontinuous galerkin finite element approximation of nondivergence form elliptic equations with cordes coefficients. *SIAM Journal on Numerical Analysis*, 51(4):2088–2106. 8, 12, 25, 74

[56] Wang, C. and Wang, J. (2015). A Primal-Dual Weak Galerkin Finite Element Method for Second Order Elliptic Equations in Non-Divergence Form. *ArXiv e-prints*. 13, 25, 74

[57] Winter, M. (1996). Lavrentiev phenomenon in microstructure theory. *Electron. J. Differential Equations*, pages No. 06, approx. 12 pp. (electronic). 17, 116

Vita

Stefan Schnake was born in the town of Carterville, Illinois in 1990. He obtained his Bachelor of Science in Applied Mathematics with an Emphasis in Computer Science from Murray State University in the Spring of 2012. He graduated from Murray State with an honors diploma and his honors thesis is titled "Dynamics of Differential Operators" written under the direction of Professor John E. Porter. That same fall, he began his doctoral studies at the University of Tennessee, Knoxville. In the Fall of 2014, he revived a concurrent Master of Science degree from the University of Tennessee.