CONVERGENCE ANALYSIS OF FINITE ELEMENT METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS IN NON-DIVERGENCE FORM

by

Lauren Hennings

B.A., University at Buffalo, 2011

Submitted to the Graduate Faculty of

the Dietrich Graduate School of Arts and Sciences in partial

fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

DIETRICH GRADUATE SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Lauren Hennings

It was defended on

April 17th 2014

and approved by

Dr. Michael Neilan, University of Pittsburgh, Mathematics

Dr. William Layton, University of Pittsburgh, Mathematics

Dr. Hisham Sati, University of Pittsburgh, Mathematics

Thesis Advisor: Dr. Michael Neilan, University of Pittsburgh, Mathematics

CONVERGENCE ANALYSIS OF FINITE ELEMENT METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS IN NON-DIVERGENCE FORM

Lauren Hennings, M.S.

University of Pittsburgh, 2014

The purpose of this project is to derive stability estimates for a finite element method for linear, elliptic partial differential equation in non-divergence form. The thesis begins by introducing basic definitions of the Sobolev spaces used and the corresponding norms of these spaces. We then define the finite element method of the problem. We dedicate one chapter to working out what the finite element method reduces to when we are in one dimension. Chapter 4 involves preliminary lemmas which will lead up to the proof of the main result. The proof of these lemmas, including the main result, involve a common theme of using inverse estimates, interpolation estimates, and various other inequalities. Once we prove the main result, we then prove existence, uniqueness, and error estimates of the solution of the finite element method. The last chapter is dedicated to numerical experiments. We choose three test problems in the one dimensional case and discuss the error and convergence rates of each, as well as whether each problem supports the theoretical estimates.

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

1.1 DESCRIPTION OF THE PROBLEM

Consider the Model Elliptic Problem:

$$-\nabla \cdot (A\nabla u) = f \text{ in } \Omega \subset \mathbb{R}^d \text{ where } d \geq 1, \tag{1.1}$$
$$u = g \text{ in } \partial\Omega,$$

where f, g, and A are assumed from to be sufficiently smooth given data, Ω is a convex polytope domain, and $A : \Omega \to \mathbb{R}^{d\times d}$ is symmetric and uniformly positive definite. The variational formulation problem reads: Find $u \in H^1_g(\Omega) : = \{v \in H^1(\Omega) : v|_{\partial\Omega} = g\}$ such that

$$\int_\Omega (A\nabla u) \cdot \nabla v dx = \int_\Omega fv dx \quad \forall\, v \in H^1_0(\Omega), \tag{1.2}$$

where

$$\int_\Omega (A\nabla u) \cdot \nabla v dx = \sum_{i,j=1}^{d} \int_\Omega A_{ij} \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} dx. \tag{1.3}$$

We will define the Hilbert space $H^1(\Omega)$ in chapter 2. By the Lax-Milgram Theorem, this is a well posed problem [6].

In this project, we consider finite element methods for the following linear problem:

$$-A : D^2 u = f \quad \text{in } \Omega, \tag{1.4}$$
$$u = 0 \quad \text{on } \partial\Omega,$$

where $A \in [C^{0,\alpha}(\bar{\Omega})]^{d\times d}$, with f given, and where

$$A : D^2 u = \sum_{i,j=1}^{d} A_{i,j} \frac{\partial^2 u}{\partial x_i \partial x_j}. \tag{1.5}$$

Several problems arise when discussing second order elliptic operators in non-divergence form. Defining what it means to have a weak solution explicitly is sometimes difficult. Also, variational formulation of the problem generally does not exist. Despite these difficulties, well-posedness of solutions in $H^2(\Omega) \cap H_0^1(\Omega)$ is recovered for problems with a convex domain [14]. Since the domain in convex, there exists a strong solution to the problem (1.4) in H^2 [4]. Furthermore, the existence of a strong solution is also guaranteed if the boundary $\partial\Omega$ is smooth and therefore the following estimate is satisfied with some constant c [10]:

$$\|u\|_{H^2(\Omega)} \leq c\|f\|_{L^2(\Omega)}.$$

For this problem we can define a strong solution as a solution that possesses second derivates, at least in a weak sense, and that satisfies (1.4) almost everywhere.

Since the standard trick of using integration by parts is not applicable, a conforming finite element method for problem (1.4) requires H^2 regularity of the approximate solution, i.e. a C^1 continuity condition of the finite element space [14]. The need for H^2 regularity is required because of the second derivatives present in the problem (1.4). C^1 finite element methods are not desirable because they require a higher polynomial degree, which is fairly complicated. Also, programming of the method is non-trivial.

1.2 APPLICATIONS

The Hamilton – Jacobi – Bellman equation is derived from optimal control problems. Let $\{L_\alpha\}_{\alpha \in E}$ be a family of second order elliptic operators, where E is a compact metric space. Let $\{f_\alpha\}_{\alpha \in E}$ be a family of functions. The Hamilton – Jacobi – Bellman equation reads: Find u such that

$$\sup_{\alpha \in E}(L_\alpha u - f_\alpha) = 0 \quad \text{on} \quad \Omega,$$

$$u = g \quad \text{on} \quad \partial\Omega.$$

The solution u is the *value function* that gives the minimum cost for a given dynamic system with an associated cost function [3]. Some of the difficulties of these equations are that they

are fully nonlinear, i.e., if we set $F[u] := \sup_{\alpha \in E}(L_\alpha u - f_\alpha)$ then $F[\alpha u + \beta y] \neq \alpha F[u] + \beta F[u]$ in general. Also, the structure of each L_α is unconventional, namely,

$$L_\alpha u := -A_\alpha : D^2 u + b_\alpha \cdot \nabla u + c_\alpha u, \tag{1.6}$$

where $A_\alpha : \Omega \to \mathbb{R}^{d \times d}$ is bounded, $b_\alpha : \Omega \to \mathbb{R}^d$, and $c_\alpha : \Omega \to \mathbb{R}$.

In general, the operators L_α are in non divergence form. If the coefficients are smooth then

$$L_\alpha u = -\nabla \cdot (A_\alpha \nabla u) + (\nabla \cdot A_\alpha + b_\alpha) \cdot \nabla u + c_\alpha u, \tag{1.7}$$

where $\nabla \cdot$ is applied to A_α row wise. However, $\nabla \cdot A_\alpha$ does not exist in general.

Notice the first term of equation (1.6) has the same non-divergence structure as problem (1.4). Therefore, the results of this project is a first step to constructing finite element methods for the Hamilton – Jacobi – Bellman equation.

1.3 GOALS AND PURPOSE

The purpose of this paper is to formulate and prove convergence estimates of a finite element method for problem (1.4). We do this in several steps:

(1) Define the Sobolev Spaces we will use and formulate the finite element method for partial differential equations in divergence and non-divergence form.

(2) Formulate the finite element method in one dimension and derive an equivalent finite difference scheme.

(3) State commonly used inequalities with respect to the norms defined in step (1).

(4) Prove a discrete elliptic stability estimate for finite element methods for partial differential equations in divergence form.

(5) Prove the main result being a discrete stability estimate for finite element methods for partial differential equations in non-divergence form.

(6) Prove existence and uniqueness of the solution to the finite element method as well as prove an error estimate.

(7) Conduct numerical estimates which agree with the error estimates proven.

2.0 FORMULATION OF FINITE ELEMENT METHODS IN DIVERGENCE AND NON-DIVEGENCE FORM

This chapter discusses the derivation of the finite element method for problem (1.4). We first state definitions of the spaces and norms we will use. We then state the definition of the triangulation, the definition of the finite element method for partial differential equations in divergence form, and the derivation of the method for partial differential equations in non-divergence form.

2.1 PARTIAL DIFFERENTIAL EQUATIONS IN DIVERGENCE FORM

Definition 1. *Let $\bar{\Omega}$ denote the closure of Ω. The space, $C^{0,\alpha}(\bar{\Omega})$, is defined as the space of continuous functions on $\bar{\Omega}$ which are also α-Hölder continous, i.e.,*

$$C^{0,\alpha}(\bar{\Omega}) = \{v \in C^0(\bar{\Omega}) : \exists K > 0 \text{ such that } |v(x) - v(y)| \leq K|x-y|^\alpha \ \forall x, y \in \bar{\Omega}\}.$$

Definition 2. *The L^p space on domain Ω for $1 \leq p < \infty$ is defined as follows:*

$$L^p(\Omega) = \{v : \Omega \to \mathbb{R} \text{ is measurable } : \int_\Omega |v|^p \, dx < \infty\}.$$

The norm on this space is defined to be

$$\|v\|_{L^p(\Omega)} = (\int_\Omega |v|^p \, dx)^{1/p}.$$

Definition 3. *For $p = \infty$, the L^p space is defined as the set of all measurable functions on Ω which are bounded. The norm on this space is defined to be*

$$\|v\|_{L^\infty(\Omega)} = \textit{essential sup } |v|$$

Definition 4. *The Hilbert Space $H^m(\Omega)$ is defined as follows:*

$$H^m(\Omega) = \{u \in L^2(\Omega) : D^\alpha u \in L^2(\Omega) \ \forall |\alpha| \leq m\}. \tag{2.1}$$

The norm on this space is defined to be

$$\|v\|_{H^m(\Omega)} = (\sum_{|\alpha| \leqslant m} \|D^\alpha v\|^2_{L^2(\Omega)})^{1/2}.$$

We define $H^1_0(\Omega)$ as the set of elements in $H^1(\Omega)$ with zero trace.

Definition 5. *The Sobolev space $W^{1,\infty}(\Omega)$ is defined as follows:*

$$W^{1,\infty}(\Omega) = \{u \in L^\infty(\Omega) : D^\alpha u \in L^\infty(\Omega) \, \forall |\alpha| \leq 1\}. \tag{2.2}$$

The norm on this space is defined to be

$$\|v\|_{W^{1,\infty}(\Omega)} = \max_{|\alpha| \leq 1} \|D^\alpha v\|_{L^\infty(\Omega)}.$$

Definition 6. *We define the triangulation T_h as a partition of Ω (in our case, simplices) with the following properties [2]:*

(1): $\bar{\Omega} = \cup_{T \in T_h} T$

(2): $\forall T \in T_h, T$ is closed

(3): for any two distinct simplices T_i and T_j in T_h, $\overset{\circ}{T}_i \cap \overset{\circ}{T}_j = \emptyset$

(4): if $e = T_i \cap T_j \neq \emptyset$, then e is either a common face, side, or vertex of T_i and T_j.

We also denote by $\mathcal{E}_h^I$ the set of interior $(d-1)$ dimensional simplices of T_h (e.g., edges $(d = 2)$ or faces $(d = 3)$).

Definition 7. *Define X_h to be a space of globally continuous piecewise polynomials of degree $k \geq 1$ with respect to a partition space T_h of Ω, i.e.,*

$$X_h = \left\{ v_h \in H_0^1(\Omega) : v_h|_T \in \mathbb{P}_k(T) \quad \forall T \in T_h \right\},$$

where $\mathbb{P}_k(T)$ is the space of all polynomials of degree k with domain T and $v_h|_T$ denotes the restriction of v_h to T.

Remark 1. *It can be argued that $X_h \subset C^0(\bar{\Omega})$[6].*

A finite element method for problem (1.2) reads as follows:

$$\text{Find } u_h \in X_h \text{ such that } \int_\Omega (A\nabla u_h) \cdot \nabla v_h dx = \int_\Omega f v_h dx \ \ \forall v_h \in X_h. \tag{2.3}$$

Lemma 1. Céa's Lemma *[9]*

Let X be a real Hilbert space with the norm $\|\cdot\|$. Let $a : X \times X \to \mathbb{R}$ be a bilinear form with the following properties:

(1) $|a(v,w)| \leq \gamma \|v\| \|w\|$ for some constant $\gamma > 0$ and $\forall v, w \in X$

(2) $a(v,v) \geq \alpha \|v\|^2$ for some constant $\alpha > 0$ and $\forall v \in X$

Let X_h be a finite dimensional subspace of X, and let $L : X_h \to \mathbb{R}$ be a bounded linear operator. Consider the problem of finding an element $u_h \in X$ such that

$$a(u_h, v) = L(v) \ \ \forall v \in X_h.$$

Then if $u \in X$ satisfies $a(u,v) = L(v) \, \forall v \in X$, we have

$$\|u - u_h\| \leq \frac{\gamma}{\alpha} \|u - v\|, \ \forall v \in X_h.$$

Before we continue, we need an inequality.

Lemma 2. *Friedrichs' - Poincaré Inequality [1]*

There holds the following inequality.

$$\int_\Omega |u|^2 \, dx \leq C \int_\Omega |\nabla u|^2 \, dx \ \ \forall u \in H_0^1(\Omega). \tag{2.4}$$

Remark 2. *We can conclude by the above lemma that $u \to \|\nabla u\|_{L^2(\Omega)}$ is indeed a norm on $H_0^1(\Omega)$.*

Lemma 3. *Let $a(u,v) = \int_\Omega (A\nabla u)\cdot\nabla v\,dx$ and $L(v) = \int_\Omega fv\,dx$. Choosing the norm $\|\cdot\|$ to be $\|\nabla\cdot\|_{L^2(\Omega)}$, the bilinear form $a(\cdot,\cdot)$ satisfies the hypotheses of Céa's Lemma, and therefore the conclusions of Céa's Lemma hold. In particular, there exists a unique $u_h \in X_h$ satisfying (2.3), and*

$$\|\nabla(u-u_h)\|_{L^2(\Omega)} \le C\|\nabla(u-v)\|_{L^2(\Omega)}\ \forall v \in X_h.$$

Proof. Since A is uniformly positive definite, there exists positive constants m and M such that

$$m|y|^2 \le y^T A(x) y \le M|y|^2\ \ \forall y \in \mathbb{R}^d,\ \ \forall x \in \bar{\Omega}.$$

We will first show the first hypothesis of Céa's Lemma. We have

$$\begin{aligned}
|a(u,v)| = |\int_\Omega (A\nabla u)\cdot\nabla v\,dx| &\le \|A\|_{L^\infty(\Omega)} \int_\Omega |\nabla u||\nabla v|\,dx \\
&\le \|A\|_{L^\infty(\Omega)} (\int_\Omega |\nabla u|^2\,dx)^{1/2} (\int_\Omega |\nabla v|^2\,dx)^{1/2} \\
&= \|A\|_{L^\infty(\Omega)} \|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)}.
\end{aligned}$$

We now show the second hypothesis of Céa's Lemma. Using the uniform positive definiteness of A, we have

$$a(u,u) = \int_\Omega (A\nabla u)\cdot\nabla u\,dx = \int_\Omega (\nabla u)^T A \nabla u\,dx \ge m\int_\Omega |\nabla u|^2\,dx = m\|\nabla u\|^2_{L^2(\Omega)}.$$

Therefore, the hypotheses of Céa's Lemma hold, and hence the conclusion of Céa's Lemma holds. In particular,

$$\|\nabla(u-u_h)\|_{L^2(\Omega)} \le C\|\nabla(u-v)\|_{L^2(\Omega)}\ \forall v \in X_h,$$

where $C = \dfrac{\|A\|_{L^\infty(\Omega)}}{m}$. □

2.2 DERIVATION OF THE FINITE ELEMENT METHOD IN NON-DIVERGENCE FORM

In this section we derive the finite element method for the linear problem in non-divergence form (1.4).

To this end, suppose for the moment that the coefficient matrix A in (1.4) is smooth. We can then, using the product rule, write $-A : D^2u$ as

$$-A : D^2u = -\nabla \cdot (A\nabla u) + (\nabla \cdot A) \cdot \nabla u. \tag{2.5}$$

Expanding (2.5) we have

$$-\sum_{i,j=1}^{d} A_{i,j}\frac{\partial^2 u}{\partial x_i \partial x_j} = -\sum_{i,j=1}^{d}[\frac{\partial}{\partial x_j}(A_{i,j}\frac{\partial u}{\partial x_j}) - \frac{\partial}{\partial x_i}A_{i,j}\frac{\partial u}{\partial x_j}]. \tag{2.6}$$

A standard finite element method [6] for (2.5) is to find $u_h \in X_h$ such that

$$\int_\Omega A\nabla u_h \cdot \nabla v_h dx + \int_\Omega (\nabla \cdot A) \cdot \nabla u_h \, v_h \, dx = \int_\Omega f \, v_h \, dx \quad \forall \, v_h \in X_h. \tag{2.7}$$

Recall from Chapter 1 that we only assume $A \in [C^{0,\alpha}(\bar{\Omega})]^{d\times d}$. Therefore, the second integral in (2.7) is not well defined. For now, (2.7) will not suffice. We can write the first term of (2.7), $\int_\Omega A\nabla u_h \cdot \nabla v_h \, dx$, as the sum of the integral over each triangle in the triangulation, i.e., $\sum\limits_{T\in T_h} \int_T A\nabla u_h \cdot \nabla v_h \, dx$. Integrating by parts gives us

$$\sum_{T\in T_h}\int_T A\nabla u_h \cdot \nabla v_h \, dx = -\sum_{T\in T_h}[\int_T \nabla \cdot (A\nabla u_h)v_h \, dx - \int_{\partial T}(A\nabla u_h \cdot n_T)v_h \, ds], \tag{2.8}$$

where n_T is the outward unit normal of ∂T. Combining (2.8) and (2.5) gives us the following identity:

$$\begin{aligned}\sum_{T\in T_h}\int_T A\nabla u_h \cdot \nabla v_h \, dx &= -\sum_{T\in T_h}\int_T (A : D^2u_h)v_h \, dx - \sum_{T\in T_h}\int_T (\nabla \cdot A)\cdot \nabla u_h v_h \, dx \\ &\quad + \sum_{T\in T_h}\int_{\partial T}(A\nabla u_h \cdot n_T)v_h \, ds \\ &= -\sum_{T\in T_h}\int_T (A : D^2u_h)v_h \, dx - \int_\Omega (\nabla \cdot A)\cdot \nabla u_h v_h \, dx\end{aligned}$$

$$+\sum_{T\in T_h}\int_{\partial T}(A\nabla u_h\cdot n_T)v_h\,ds.$$

Plugging all this into the first term of equation (2.7), one term will cancel, and therefore the finite element method (2.7) is equivalent to

$$-\sum_{T\in T_h}\int_T (A:D^2u_h)v_h\,dx+\sum_{T\in T_h}\int_{\partial T}(A\nabla u_h)\cdot n_T v_h\,ds=\int_\Omega f v_h\,dx. \tag{2.9}$$

We now explain how the second term may be written as a sum of integrals over edges. Suppose $e=\partial T_e^{(1)}\cap\partial T_e^{(2)}\in\mathcal{E}_h^I$ (endpoint in 1D, edge in 2D, face in 3D), for some $T_e^{(1)}$ and $T_e^{(2)}$ in T_h. Define the jump of a vector-valued function as:

$$[v]|_e=v_{T_e^{(1)}}\cdot n_{T_e^{(1)}}|_e+v_{T_e^{(2)}}\cdot n_{T_e^{(2)}}|_e,\ \text{where } v_{T_e^{(i)}}:=v|_{T_e^{(i)}}\text{ for } i=1,2. \tag{2.10}$$

We see that

$$[A\nabla u_h]=(A\nabla u_h)_{T_e^{(1)}}\cdot n_{T_e^{(1)}}+(A\nabla u_h)_{T_e^{(2)}}\cdot n_{T_e^{(2)}}. \tag{2.11}$$

Then since v_h is continuous and vanishes on $\partial\Omega$, we have

$$\begin{aligned}\sum_{T\in T_h}\int_{\partial T}(A\nabla u_h)\cdot n_T v_h\,ds&=\sum_{e\in\mathcal{E}_h^I}\int_e\left((A\nabla u_h)|_{T_e^{(1)}}\cdot n_{T_e^{(1)}}v_h+(A\nabla u_h)|_{T_e^{(2)}}\cdot n_{T_e^{(2)}}v_h\right)ds\\&=\sum_{e\in\mathcal{E}_h^I}\int_e[A\nabla u_h]v_h\,ds.\end{aligned} \tag{2.12}$$

Plugging (2.12) into (2.9), the finite element method for (1.4) becomes:

$$-\sum_{T\in T_h}\int_T(A:D^2u_h)v_h\,dx+\sum_{e\in\mathcal{E}_h^I}\int_e[A\nabla u_h]v_h\,ds=\int_\Omega f v_h\,dx\quad\forall v_h\in X_h. \tag{2.13}$$

The method defined in (2.13) will be the method analyzed in the subsequent sections.

Remark 3. *The method is equivalent to* (2.7)*, provided* $A\in[W^{1,\infty}(\Omega)]^{d\times d}$*. Unlike* (2.7)*, the finite element method* (2.13) *makes sense for* $A\in[C^{0,\alpha}(\bar\Omega)]^{d\times d}$*.*

3.0 THE FINITE ELEMENT METHOD (2.13) IN ONE DIMENSION: A CONVERGENT FINITE DIFFERENCE SCHEME

In this chapter we will study the finite element method (2.13) in the one dimensional case, i.e., $d = 1$. Let X_h consist of globally continuous piecewise linear functions, i.e., $k = 1$. We let Ω be the open interval (a, b), and we let $\mathcal{E}_h^I = \{x_i\}_{i=1}^N$, where x_i are evenly spaced points on the interval (a, b), with $h = x_i - x_{i-1} = \dfrac{b-a}{N+1}$, $x_0 = a$, and $x_{N+1} = b$. Since the second derivative of a linear function is zero, the first term in (2.13) vanishes. The finite element method in 1D then reads

$$\sum_{k=1}^{N} A(x_k)[u_h'](x_k)v_h(x_k) = \int_a^b f v_h \, dx \quad \forall v_h \in X_h. \tag{3.1}$$

Let $\{\varphi_i\}_{i=1}^N$ be a basis of piecewise linear hat functions. Specifically, φ_i is piecewise linear and $\varphi_i(x_j) = \delta_{i,j}$. Setting $v_h = \varphi_i$ in (3.1) and since $\varphi_i = 0$ outside $[x_{i-1}, x_{i+1}]$, we have

$$\sum_{k=1}^{N} A(x_k)[u_h'](x_k)\varphi_i(x_k) = \int_a^b f\varphi_i \, dx = \int_{x_{i-1}}^{x_{i+1}} f\varphi_i \, dx. \tag{3.2}$$

Using the trapezoid rule to evaluate the right hand side of (3.2) we have

$$\int_{x_{i-1}}^{x_{i+1}} f\varphi_i \, dx \approx \frac{(x_{i+1} - x_{i-1})}{4}(f(x_{i-1})\varphi_i(x_{i-1}) + 2f(x_i)\varphi_i(x_i) + f(x_{i+1})\varphi_i(x_{i+1})) = \frac{h}{2}2f(x_i)$$

$$= hf(x_i).$$

The φ_i are a basis of piecewise linear functions and therefore we can write u_h as a linear combination of them. Therefore, there exists $\{c_j\}_{j=1}^N \subset \mathbb{R}$ such that

$$u_h = \sum_{j=1}^{N} c_j \varphi_j.$$

Also, since φ_k has support on $[x_{k-1}, x_{k+1}]$, we have

$$u\big|_{(x_{k-1},x_k)} = c_{k-1}\varphi_{k-1} + c_k\varphi_k, \text{ and } u\big|_{(x_k,x_{k+1})} = c_k\varphi_k + c_{k+1}\varphi_{k+1}.$$

Using the definition of the jump across interior nodes, we have

$$\begin{aligned}[u_h'](x_k) &= u_h'\big|_{(x_{k-1},x_k)}(x_k) - u_h'\big|_{(x_k,x_{k+1})}(x_k) \\ &= c_{k-1}\varphi_{k-1}'\big|_{(x_{k-1},x_k)}(x_k) + c_k\varphi_k'\big|_{(x_{k-1},x_k)}(x_k) - c_k\varphi_k'\big|_{(x_k,x_{k+1})}(x_k) - c_{k+1}\varphi_{k+1}'\big|_{(x_k,x_{k+1})}(x_k).\end{aligned} \tag{3.3}$$

Now the slope of φ_k on (x_{k-1}, x_k) is $\frac{1}{h}$, and therefore $\varphi_k' = \frac{1}{h}$ on (x_{k-1}, x_k). Similarly the slope of φ_k is $\frac{-1}{h}$ on (x_k, x_{k+1}), and therefore $\varphi_k' = \frac{-1}{h}$. Plugging these identities into (3.3), we get:

$$[u_h'](x_k) = -c_{k-1}\frac{1}{h} + 2c_k\frac{1}{h} - c_{k+1}\frac{1}{h}.$$

Therefore the linear problem reduces to finding $\{c_j\}_{i=1}^N \subset \mathbb{R}$ such that:

$$A(x_j)(-c_{j-1}\frac{1}{h} + 2c_j\frac{1}{h} - c_{j+1}\frac{1}{h}) = hf(x_j).$$

Multiplying both sides by h and dividing by $A(x_j)$ gives us the finite difference scheme:

$$(-c_{j-1} + 2c_j - c_{j+1}) = \frac{h^2 f(x_j)}{A(x_j)}. \tag{3.4}$$

Theorem 1. *[9] Suppose that f is continuous of $\bar{\Omega} = [a, b]$, and that the trapezoid rule is applied to the right hand side of (3.1). Then there exists a unique solution to (3.1). Moreover, there holds*

$$\max_{x\in[a,b]} |u(x) - u_h(x)| \leq Ch^2 \|f\|_{L^\infty([a,b])},$$

where C is some constant.

4.0 CONVERGENCE ANALYSIS OF THE FINITE ELEMENT METHOD (2.13)

In this section we derive crucial stability estimates of the finite element method (2.13). We first define a discrete L^2 norm we need in the convergence analysis as well as various important inequalities and estimates used to prove the stability estimates. We define an operator L_h associated with the finite element method (2.13) and then prove various lemmas and corollaries that illustrate properties and estimates with respect to the operator. Once we prove the main result, the stability estimate, we prove uniqueness and existence of the solution to the finite element method (2.13). These stability estimates naturally lead to error estimates of the finite element method (2.13).

4.1 PRELIMINARY RESULTS

Lemma 4. *For a function $v \in L^2(\Omega)$,*

$$\|v\|_{L^2(\Omega)} = \sup_{w \in L^2(\Omega)} \frac{\int_\Omega vw\, dx}{\|w\|_{L^2(\Omega)}}. \tag{4.1}$$

Proof. Using the Cauchy-Schwarz inequality, we have

$$\int_\Omega vw\, dx \leq (\int_\Omega v^2\, dx)^{1/2} (\int_\Omega w^2\, dx)^{1/2} = \|v\|_{L^2(\Omega)} \|w\|_{L^2(\Omega)}.$$

Therefore,

$$\frac{\int_\Omega vw\, dx}{\|w\|_{L^2(\Omega)}} \leq \|v\|_{L^2(\Omega)}.$$

Taking the supremum of both sides over all $w \in L^2(\Omega)$ we get

$$\sup_{w \in L^2(\Omega)} \frac{\int_\Omega vw\, dx}{\|w\|_{L^2(\Omega)}} \leq \|v\|_{L^2(\Omega)}.$$

Let $w = \dfrac{\mathrm{sgn}(v)|v|}{\|v\|_{L^2(\Omega)}}$. We see that $\|w\|_{L^2(\Omega)} = 1$. We then have

$$\frac{\int_\Omega vw\, dx}{\|w\|_{L^2(\Omega)}} = \int_\Omega vw\, dx = \int_\Omega v \frac{\mathrm{sgn}(v)|v|}{\|v\|_{L^2(\Omega)}}\, dx = \frac{1}{\|v\|_{L^2(\Omega)}} \int_\Omega |v|^2\, dx = \|v\|_{L^2(\Omega)}.$$

Therefore

$$\|v\|_{L^2(\Omega)} \leq \sup_{w \in L^2(\Omega)} \frac{\int_\Omega vw\, dx}{\|w\|_{L^2(\Omega)}}.$$

Both inequalities give us the desired equality. □

This identity motivates the following definition.

Definition 8. *The discrete L^2 norm is defined as*

$$\|r\|_{L_h^2(\Omega)} = \sup_{w_h \in X_h} \frac{\langle r, w_h \rangle}{\|w_h\|_{L^2(\Omega)}} \quad \forall r \in X_h', \tag{4.2}$$

where X_h' is the dual space of X_h, and $\langle \cdot, \cdot \rangle$ denotes the dual pairing between X_h' and X_h. The discrete H^2-type norm is defined as

$$\|v_h\|_{H_h^2(\Omega)}^2 = \sum_{T \in \mathcal{T}_h} \|v_h\|_{H^2(T)}^2 + \sum_{e \in \mathcal{E}_h^I} h_e^{-1} \|[\nabla v_h]\|_{L^2(e)}^2, \tag{4.3}$$

where $h_e = diam(e)$.

Lemma 5. *Trace Inequality [5]*
Let $T \in T_h$. Then for any $\varphi \in H^1(T)$ there holds

$$\|\varphi\|_{L^2(\partial T)}^2 \leq C\left(\frac{1}{h_T}\|\varphi\|_{L^2(T)}^2 + h_T \|\nabla \varphi\|_{L^2(T)}^2\right), \tag{4.4}$$

where h_T is the diameter of the triangle T, and C is independent of the size of T.

Lemma 6. *There holds the following inverse inequality [6]:*

$$\|v_h\|_{H^m(T)} \leq C h_T^{\ell - m} \|v_h\|_{H^\ell(T)} \; \forall v_h \in X_h, \forall T \in T_h, \text{ for } 0 \leq \ell \leq m. \tag{4.5}$$

Lemma 7. *There holds the following interpolation estimate [6]:*

$$\|\varphi - I_h\varphi\|_{H^m(T)} \leq Ch^{\ell-m}\|\varphi\|_{H^\ell(T)} \ \forall \varphi \in H^s(T), \tag{4.6}$$

where $\ell = min(k+1, s)$, $s \geq 2$, $I_h\varphi$ *is the interpolating polynomial of* φ, *and* $h = \max_{T\in T_h} h_T$.

Lemma 8. *For* $\varphi \in H^s(\Omega)$ *with* $s \geq 2$,

$$\|\varphi - I_h\varphi\|_{H_h^2(\Omega)} \leq Ch^{\ell-2}\|\varphi\|_{H^\ell(\Omega)}, \tag{4.7}$$

where $\ell = min\{k+1, s\}$.

Proof. Using the definition of the discrete H^2 norm from (4.3) we have

$$\begin{aligned}
\|\varphi - I_h\varphi\|^2_{H_h^2(\Omega)} &= \sum_{T\in\mathcal{T}_h} \|\varphi - I_h\varphi\|^2_{H^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla(\varphi - I_h\varphi)]\|^2_{L^2(e)} \\
&\leq \sum_{T\in\mathcal{T}_h} \|\varphi - I_h\varphi\|^2_{H^2(T)} + \sum_{T\in T_h} h_T^{-1}C\Big(\frac{1}{h_T}\|[\nabla(\varphi - I_h\varphi)]\|^2_{L^2(T)} + h_T\|\nabla[\nabla(\varphi - I_h\varphi)]\|^2_{L^2(T)}\Big) \\
&\qquad \text{(by the trace inequality (4.4))} \\
&= \sum_{T\in\mathcal{T}_h} \|\varphi - I_h\varphi\|^2_{H^2(T)} + C\sum_{T\in T_h} h_T^{-2}\|[\nabla(\varphi - I_h\varphi)]\|^2_{L^2(T)} + C\sum_{T\in T_h} \|\nabla[\nabla(\varphi - I_h\varphi)]\|^2_{L^2(T)} \\
&\leq C\sum_{T\in\mathcal{T}_h} \|\varphi - I_h\varphi\|^2_{H^2(T)} + C\sum_{T\in T_h} h_T^{-2}\|\varphi - I_h\varphi\|^2_{H^1(T)} \\
&\leq \sum_{T\in\mathcal{T}_h} (Ch^{\ell-2}\|\varphi\|_{H^\ell(T)})^2 + C\sum_{T\in T_h} h_T^{-2}(h^{\ell-1}\|\varphi\|_{H^\ell(T)})^2 \ \text{(by (4.6))} \\
&\leq (Ch^{\ell-2})^2 \sum_{T\in\mathcal{T}_h} \|\varphi\|^2_{H^\ell(T)} \\
&= (Ch^{\ell-2})^2(\|\varphi\|_{H^\ell(\Omega)})^2.
\end{aligned}$$

Taking the square root of both sides completes the proof. □

Lemma 9. *There holds the following equality:*

$$\|\varphi\|^2_{H_h^2(\Omega)} = \|\varphi\|^2_{H^2(\Omega)} \ \ \forall \varphi \in H^2(\Omega). \tag{4.8}$$

Proof. If $\varphi \in H^2(\Omega)$, then $[\nabla\varphi]|_e = 0 \ \ \forall e \in \mathcal{E}_h^I$. Therefore by (4.3),

$$\|\varphi\|^2_{H_h^2(\Omega)} = \sum_{T\in\mathcal{T}_h} \|\varphi\|^2_{H^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla\varphi]\|^2_{L^2(e)} = \sum_{T\in\mathcal{T}_h} \|\varphi\|^2_{H^2(T)} = \|\varphi\|^2_{H^2(\Omega)}.$$

□

Lemma 10. *There holds the following inverse inequality:*

$$\|v_h\|_{H_h^2(\Omega)} \le Ch^{-1}\|v_h\|_{H^1(\Omega)} \ \ \forall v_h \in X_h. \tag{4.9}$$

Proof. Using the identity from (4.3) we have,

$$\begin{aligned}
\|v_h\|^2_{H_h^2(\Omega)} &= \sum_{T\in\mathcal{T}_h} \|v_h\|^2_{H^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)} \\
&\le \sum_{T\in\mathcal{T}_h} \|v_h\|^2_{H^2(T)} + \sum_{T\in T_h} h_T^{-1}[C(\frac{1}{h_T}\|[\nabla v_h]\|^2_{L^2(T)} + h_T\|\nabla[\nabla v_h]\|^2_{L^2(T)})] \\
&\quad \text{(by the trace inequality (4.4))} \\
&= \sum_{T\in\mathcal{T}_h} \|v_h\|^2_{H^2(T)} + C\sum_{T\in T_h} h_T^{-2}\|[\nabla v_h]\|^2_{L^2(T)} + C\sum_{T\in T_h} \|\nabla[\nabla v_h]\|^2_{L^2(T)} \\
&\le C\sum_{T\in\mathcal{T}_h} \|v_h\|^2_{H^2(T)} + C\sum_{T\in T_h} h_T^{-2}\|v_h\|^2_{H^1(T)} \\
&\le C\sum_{T\in\mathcal{T}_h} (h_T^{-1}\|v_h\|_{H^1(T)})^2 \ \ \text{(by (4.5))} \\
&\le C(h^{-1}\|v_h\|_{H^1(\Omega)})^2.
\end{aligned}$$

Taking square roots of both sides we get the desired inequality. □

4.2 DISCRETE STABILITY ESTIMATES FOR PARTIAL DIFFERENTIAL EQUATIONS IN DIVERGENCE FORM

Define the operator $L_h : X_h \to X_h'$ such that

$$\langle L_h v_h, w_h \rangle = -\sum_{T\in\mathcal{T}_h} \int_T (A : D^2 v_h) w_h dx + \sum_{e\in\mathcal{E}_h^I} \int [A\nabla v_h] w_h \, ds.$$

Note that the finite element method (2.13) reads: Find $u_h \in X_h$ such that

$$\langle L_h u_h, v_h \rangle = \int_\Omega f v_h \, dx \;\; \forall v_h \in X_h.$$

For fixed $x_0 \in \Omega$, we define the operator $L_{h,0} : X_h \to X_h'$ such that

$$\begin{aligned}\langle L_{h,0} v_h, w_h \rangle &= -\sum_{T\in\mathcal{T}_h} \int_T (A(x_0) : D^2 v_h) w_h dx + \sum_{e\in\mathcal{E}_h^I} \int [A(x_0)\nabla v_h] w_h \, ds \\ &= \int_\Omega (A(x_0)\nabla v_h)\cdot \nabla w_h \, dx.\end{aligned}$$

where integration by parts was used to derive the last identity.

Lemma 11. *[12] There holds,* $\|v_h\|_{H_h^2(\Omega)} \le C\|L_{h,0} v_h\|_{L_h^2(\Omega)} \;\; \forall v_h \in X_h.$

Proof. Let $r = L_{h,0} v_h$, and let $P_2 r \in X_h$ be the unique minimizer of

$$v_h \to \frac{1}{2}\int_\Omega |v_h|^2 - \langle r, v_h \rangle$$

over X_h. Let $g(t) = \frac{1}{2}\int\limits_\Omega |P_2 r + t v_h|^2 - \langle r, P_2 r + t v_h\rangle$. Then, $g'(t) = \int\limits_\Omega v_h(P_2 r + t v_h) - \langle r, v_h\rangle$. We know $g'(0) = 0$ since $P_2 r$ is a minimizer. Evaluating $g'(0)$ and setting it equal to 0, we see that $P_2 r$ satisfies $\int_\Omega (P_2 r) v_h \, dx = \langle r, v_h\rangle \;\; \forall v_h \in X_h$.

Let $\varphi \in H_0^1(\Omega)$ solve:

$$\begin{aligned}\mathcal{L}_0\varphi := -\nabla\cdot(A(x_0)\nabla\varphi) &= P_2 r \text{ in } \Omega, \\ \varphi &= 0 \text{ on } \partial\Omega.\end{aligned}$$

Since Ω is convex, we then have $\varphi \in H^2(\Omega)$ [8]. We also have

$$\|\varphi\|_{H^m(\Omega)} \le C\|P_2 r\|_{H^{m-2}(\Omega)} \text{ for } m = 1, 2. \tag{4.10}$$

Note for $m = 1$ we have $H^{-1}(\Omega) := (H_0^1(\Omega))'$, with

$$\|r\|_{H^{-1}(\Omega)} = \sup_{v \in H_0^1(\Omega)} \frac{\langle r, v \rangle}{\|v\|_{H^1(\Omega)}}.$$

Now $\forall w_h \in X_h$, $\langle L_{h,0} v_h, w_h \rangle = \langle r, w_h \rangle = \int_\Omega (P_2 r) w_h \, dx = \langle \mathcal{L}_0 \varphi, w_h \rangle$ which implies that $v_h = L_{h,0}^{-1} \mathcal{L}_0 \varphi$. Therefore by Lemma 3 and Lemma 7, we have

$$\|\varphi - v_h\|_{H^1(\Omega)} \leq Ch \|\varphi\|_{H^2(\Omega)}. \tag{4.11}$$

We need the following inequality to complete the proof.

Claim:

$$\|v_h\|_{H_h^2(\Omega)} \leq C \|P_2 r\|_{L^2(\Omega)} \tag{4.12}$$

Proof of Claim: Let $I_h \varphi$ denote the interpolant of φ. We then have

$$\begin{aligned}
\|v\|_{H_h^2(\Omega)} &\leq \|v - \varphi\|_{H_h^2(\Omega)} + \|\varphi\|_{H_h^2(\Omega)} \text{ (by triangle inequality)} \\
&\leq \|v - I_h \varphi\|_{H_h^2(\Omega)} + \|\varphi - I_h \varphi\|_{H_h^2(\Omega)} + \|\varphi\|_{H_h^2(\Omega)} \text{ (by triangle inequality)} \\
&\leq Ch^{-1} \|v - I_h \varphi\|_{H^1(\Omega)} + C \|\varphi\|_{H^2(\Omega)} \text{ (by Lemma 8, Lemma 9, and Lemma 10)} \\
&\leq Ch^{-1} (\|v - \varphi\|_{H^1(\Omega)} + \|\varphi - I_h \varphi\|_{H^1(\Omega)}) + C \|\varphi\|_{H^2(\Omega)} \text{ (by triangle inequality)} \\
&\leq Ch^{-1} (Ch \|\varphi\|_{H^2(\Omega)} + Ch \|\varphi\|_{H^2(\Omega)}) + C \|\varphi\|_{H^2(\Omega)} \text{ (by (4.11) and Lemma 7)} \\
&\leq C \|\varphi\|_{H^2(\Omega)} \\
&\leq C \|P_2 r\|_{L^2(\Omega)} \text{ by (4.10).}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\|v_h\|_{H_h^2(\Omega)} &\leq C \|P_2 r\|_{L^2(\Omega)} \\
&= C \frac{\int_\Omega (P_2 r)(P_2 r) \, dx}{\|P_2 r\|_{L^2(\Omega)}} \\
&\leq C \sup_{0 \neq w_h \in V_h} \frac{\int_\Omega (P_2 r) w_h \, dx}{\|w_h\|_{L^2(\Omega)}} = C \sup_{0 \neq w_h \in V_h} \frac{\langle r, w_h \rangle}{\|w_h\|_{L^2(\Omega)}} \\
&= C \|r\|_{L_h^2(\Omega)} = C \|L_{h,0} v_h\|_{L_h^2(\Omega)}.
\end{aligned}$$

□

4.3 DISCRETE STABILITY ESTIMATES FOR PARTIAL DIFFERENTIAL EQUATIONS IN NON DIVERGENCE FORM

So far we have introduced and proven various inequalities that bound the norm of the Sobolev spaces and the discrete Sobolev spaces. We see that there is a common theme of using the trace inequality, the inverse inequality, and interpolation estimates from section 4.1 to prove these estimates. We also have proven a very important discrete stability estimate for partial differential equations in divergence form in the previous section. In particular, we know that the operator $L_{h,0}$ is very much like the operator L_h near x_0. In light of this, it seems likely that the following inequality can be derived. The proof of the following inequality will be saved for future work. For now, we will assume the following inequality to be valid. We need the following inequality to prove a lemma that will be used to prove the main result.

$$\|v_h\|_{H_h^2(\Omega)} \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + \|v_h\|_{H^1(\Omega)} + \|v_h\|_{L^2(\Omega)}) \ \forall v_h \in X_h. \tag{4.13}$$

Lemma 12. *There holds the following inequality:*

$$(\sum_{e\in\mathcal{E}_h^I} h_e\|v_h\|_{L^2(e)}^2)^{1/2} \leq C\|v_h\|_{L^2(\Omega)} \ \ \forall v_h \in X_h. \tag{4.14}$$

Proof. Using the trace inequality (4.4), we have

$$\begin{aligned}
(\sum_{e\in\mathcal{E}_h^I} h_e\|v_h\|_{L^2(e)}^2)^{1/2} &\leq (\sum_{T\in T_h} [Ch_T(h_T^{-1}\|v_h\|_{L^2(T)}^2 + h_T\|\nabla v_h\|_{L^2(T)}^2)])^{1/2} \\
&= C(\sum_{T\in T_h} [(\|v_h\|_{L^2(T)}^2 + h_T^2\|v_h\|_{H^1(T)}^2)])^{1/2} \\
&\leq C(\sum_{T\in T_h} [(\|v_h\|_{L^2(T)}^2 + h_T^2 h_T^{-2}\|v_h\|_{L^2(T)}^2)])^{1/2} \ \text{(by the inverse inequality (4.5))} \\
&= C(\sum_{T\in T_h} \|v_h\|_{L^2(T)}^2)^{1/2} \\
&= C\|v_h\|_{L^2(\Omega)}.
\end{aligned}$$

□

Lemma 13. *There exists a constant $C > 0$ independent of h such that*

$$\|v_h\|_{H^1(\Omega)} \leq C(\epsilon\|v_h\|_{H_h^2(\Omega)} + \frac{1}{\epsilon}\|v_h\|_{L^2(\Omega)}) \ \forall v_h \in X_h, \ \ \forall \epsilon > 0.$$

Proof. Integrating by parts, we have

$$\|\nabla v_h\|^2_{L^2(\Omega)} = \int_\Omega \nabla v_h \cdot \nabla v_h \, dx = -\int_\Omega \Delta v_h v_h \, dx + \sum_{e \in \mathcal{E}_h^I} \int_e [\nabla v_h] v_h \, ds = (A) + (B).$$

For term (B), using the Cauchy-Schwarz inequality, we have:

$$\begin{aligned}
(B) &\leq \sum_{e \in \mathcal{E}_h^I} \|[\nabla v_h]\|_{L^2(e)} \|v_h\|_{L^2(e)} \\
&\leq (\sum_{e \in \mathcal{E}_h^I} h_e^{-1} \|[\nabla v_h]\|^2_{L^2(e)})^{1/2} (\sum_{e \in \mathcal{E}_h^I} h_e \|v_h\|^2_{L^2(e)})^{1/2} \\
&\leq C(\sum_{e \in \mathcal{E}_h^I} h_e^{-1} \|[\nabla v_h]\|^2_{L^2(e)})^{1/2} (\|v_h\|_{L^2(\Omega)}) \ \text{(by Lemma 12)}.
\end{aligned}$$

For term (A) we just use Cauchy-Schwarz again. Putting this all together we have

$$\begin{aligned}
\|\nabla v_h\|^2_{L^2(\Omega)} &\leq \|\Delta v_h\|_{L^2(\Omega)} \|v_h\|_{L^2(\Omega)} + C(\sum_{e \in \mathcal{E}_h^I} h_e^{-1} \|[\nabla v_h]\|^2_{L^2(e)})^{1/2} (\|v_h\|_{L^2(\Omega)}) \\
&\leq C\|v_h\|_{H_h^2(\Omega)} \|v_h\|_{L^2(\Omega)}.
\end{aligned}$$

Now Cauchy-Schwarz says $ab \leq \frac{1}{2}a^2 + \frac{1}{2}b^2$. For $\epsilon > 0$, let $a = \epsilon\|v_h\|_{H_h^2(\Omega)}$ and let $b = \frac{1}{\epsilon}\|v_h\|_{L^2(\Omega)}$. We then have

$$\|\nabla v_h\|^2_{L^2(\Omega)} \leq C\|v_h\|_{H_h^2(\Omega)} \|v_h\|_{L^2(\Omega)} \leq C(\frac{\epsilon^2}{2}\|v_h\|^2_{H_h^2(\Omega)} + \frac{1}{2\epsilon^2}\|v_h\|^2_{L^2(\Omega)}).$$

We finally have

$$\begin{aligned}
\|v_h\|_{H^1(\Omega)} &= (\|v_h\|^2_{L^2(\Omega)} + \|\nabla v_h\|^2_{L^2(\Omega)})^{1/2} \\
&\leq C(\|v_h\|^2_{L^2(\Omega)} + \frac{\epsilon^2}{2}\|v_h\|^2_{H_h^2(\Omega)} + \frac{1}{2\epsilon^2}\|v_h\|^2_{L^2(\Omega)})^{1/2} \\
&\leq C(\epsilon\|v_h\|_{H_h^2(\Omega)} + \frac{1}{\epsilon}\|v_h\|_{L^2(\Omega)}).
\end{aligned}$$

□

Corollary 1. *There holds*

$$\|v_h\|_{H_h^2(\Omega)} \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + \|v_h\|_{L^2(\Omega)}) \ \ \forall v_h \in X_h. \tag{4.15}$$

Proof. Using the upper bound for $\|v_h\|_{H^1(\Omega)}$ from Lemma 13 and plugging it into (4.13), we have:

$$\begin{aligned}\|v_h\|_{H_h^2(\Omega)} &\leq C(\|L_h v_h\|_{L_h^2(\Omega)} + \|v_h\|_{H^1(\Omega)} + \|v_h\|_{L^2(\Omega)}) \\ &\leq C(\|L_h v_h\|_{L_h^2(\Omega)} + C(\epsilon\|v_h\|_{H_h^2(\Omega)} + \frac{1}{\epsilon}\|v_h\|_{L^2(\Omega)}) + \|v_h\|_{L^2(\Omega)}).\end{aligned}$$

Grouping terms and dividing we then have:

$$(1 - C^2\epsilon)\|v_h\|_{H_h^2(\Omega)} \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + \frac{C}{\epsilon}\|v_h\|_{L^2(\Omega)} + \|v_h\|_{L^2(\Omega)})$$

$$\|v_h\|_{H_h^2(\Omega)} \leq \frac{C}{1 - C^2\epsilon}(\|L_h v_h\|_{L_h^2(\Omega)} + (\frac{C}{\epsilon} + 1)\|v_h\|_{L^2(\Omega)}).$$

Taking ϵ to be sufficiently small completes the proof. □

Let $\bar{A}$ be the piecewise constant matrix defined by

$$\bar{A} = \frac{1}{|\ T\ |}\int_T A\, dx \ \ \forall T \in T_h.$$

Notice that $\bar{A}$ inherits the symmetry of A and if $A \in [C^{0,\alpha}(\Omega)]^{d\times d}$, then $\|A - \bar{A}\|_{L^\infty(\Omega)} \leq Ch^\alpha$. Define the operator $\bar{L}_h : X_h \to X_h'$ by

$$\langle \bar{L}_h v_h, w_h\rangle = -\sum_{T\in T_h}\int_T (\bar{A} : D^2 v_h) w_h\, dx + \sum_{e\in\mathcal{E}_h^I}\int_e [\bar{A}\nabla v_h] w_h\, ds, \ \ \forall v_h, w_h \in X_h. \tag{4.16}$$

Lemma 14. *The operator $\bar{L}_h$ is symmetric, i.e., $\langle \bar{L}_h v_h, w_h\rangle = \langle \bar{L}_h w_h, v_h\rangle\ \forall w_h, v_h \in X$, where $X = (\prod_{T\in T_h} H^2(T)) \cap H_0^1(\Omega)$.*

Proof. Using the definition from (4.16) we have

$$\begin{aligned}
\langle \bar{L}_h v_h, w_h\rangle &= -\sum_{T\in T_h}\int_T (\bar{A} : D^2 v_h) w_h \, dx + \sum_{e\in\mathcal{E}_h^I}\int_e [\bar{A}\nabla v_h] w_h \, ds \\
&= -\sum_{T\in T_h}\int_T \operatorname{div}(\bar{A}\nabla v_h) w_h \, dx + \sum_{e\in\mathcal{E}_h^I}\int_e [\bar{A}\nabla v_h] w_h \, ds \text{ (since } \bar{A} \text{ is piecewise constant)} \\
&= \int_\Omega \bar{A}\nabla v_h \cdot \nabla w_h \, dx \text{ (by divergence theorem)} \\
&= \int_\Omega \bar{A}\nabla w_h \cdot \nabla v_h \, dx \\
&= -\sum_{T\in T_h}\int_T \operatorname{div}(\bar{A}\nabla w_h) v_h \, dx + \sum_{e\in\mathcal{E}_h^I}\int_e [\bar{A}\nabla w_h] v_h \, ds \text{ (by divergence theorem)} \\
&= \langle \bar{L}_h w_h, v_h\rangle.
\end{aligned}$$

□

Lemma 15. *There holds* $|\langle L_h v_h, w_h\rangle| \le C\|A\|_{L^\infty(\Omega)}\|v_h\|_{H_h^2(\Omega)}\|w_h\|_{L^2(\Omega)}$ $\forall v_h, w_h \in X_h$.

Proof. Note that by the definition of the operator L_h, we have

$$|\langle L_h v_h, w_h\rangle| = |-\sum_{T\in T_h}\int_T (A : D^2 v_h) w_h \, dx + \sum_{e\in\mathcal{E}_h^I}\int_e [A\nabla v_h] w_h \, ds|. \tag{4.17}$$

Using the definition of the L^∞ norm, we can factor out $\|A\|_{L^\infty(\Omega)}$ from (4.17) and get the following inequality:

$$(4.17) \le \|A\|_{L^\infty(\Omega)}[\sum_{T\in T_h}\int_T |D^2 v_h||w_h| \, dx + \sum_{e\in\mathcal{E}_h^I}\int_e |[\nabla v_h]||w_h| \, ds].$$

Using Cauchy-Schwarz, we get:

$$\begin{aligned}
(4.17) &\le \|A\|_{L^\infty(\Omega)}[\sum_{T\in T_h}\|D^2 v_h\|_{L^2(T)}\|w_h\|_{L^2(T)} + \sum_{e\in\mathcal{E}_h^I}\|[\nabla v_h]\|_{L^2(e)}\|w_h\|_{L^2(e)}] \\
&\le \|A\|_{L^\infty(\Omega)}[(\sum_{T\in T_h}\|D^2 v_h\|_{L^2(T)}^2)^{1/2}(\sum_{T\in T_h}\|w_h\|_{L^2(T)}^2)^{1/2} \\
&\qquad + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|_{L^2(e)}^2)^{1/2}(\sum_{e\in\mathcal{E}_h^I} h_e\|w_h\|_{L^2(e)}^2)^{1/2}] \\
&\qquad \text{(from using Cauchy-Schwarz)}
\end{aligned}$$

$$\leq \|A\|_{L^\infty(\Omega)}[\|w_h\|_{L^2(\Omega)}(\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)})^{1/2} + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}(\sum_{e\in\mathcal{E}_h^I} h_e\|w_h\|^2_{L^2(e)})^{1/2}].$$

Using Lemma 12 and then factoring out $\|w_h\|_{L^2(\Omega)}$, we have

$$\begin{aligned}
(4.17) &\leq C\|A\|_{L^\infty(\Omega)}\|w_h\|_{L^2(\Omega)}[(\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)})^{1/2} + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}] \\
&\leq C\|A\|_{L^\infty(\Omega)}\|w_h\|_{L^2(\Omega)}[(\sum_{T\in T_h}\|v_h\|^2_{H^2(T)})^{1/2} + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}] \\
&\leq C\|A\|_{L^\infty(\Omega)}\|w_h\|_{L^2(\Omega)}(\sum_{T\in T_h}\|v_h\|^2_{H^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2} \\
&= C\|A\|_{L^\infty(\Omega)}\|w_h\|_{L^2(\Omega)}\|v_h\|_{H_h^2(\Omega)} \text{ (by(4.3))}.
\end{aligned}$$

□

Corollary 2. *There holds the following inequality*

$$|\langle (L_h - \bar{L}_h)v_h, w_h\rangle| \leq Ch^\alpha \|v_h\|_{H_h^2(\Omega)}\|w_h\|_{L^2(\Omega)} \ \forall v_h, w_h \in X_h.$$

Proof. Replacing L_h by $L_h - \bar{L}_h$ in Lemma 15 and using the definition of $\bar{L}_h$, we have

$$|\langle (L_h - \bar{L}_h)v_h, w_h\rangle| \leq C\|A - \bar{A}\|_{L^\infty(\Omega)}\|v_h\|_{H_h^2(\Omega)}\|w_h\|_{L^2(\Omega)},$$

and since $\|A - \bar{A}\|_{L^\infty(\Omega)} \leq Ch^\alpha$, we have

$$|\langle (L_h - \bar{L}_h)v_h, w_h\rangle| \leq Ch^\alpha \|v_h\|_{H_h^2(\Omega)}\|w_h\|_{L^2(\Omega)}.$$

□

Lemma 16. *There holds the following inequality.*

$$\|\varphi - I_h\varphi\|_{L^2(\Omega)} + (\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|^2_{L^2(e)})^{1/2} \leq Ch^2\|\varphi\|_{H^2(\Omega)}.$$

Proof. We first bound the first term by simply using Lemma 7,

$$\|\varphi - I_h\varphi\|_{L^2(\Omega)} \leq Ch^2\|\varphi\|_{H^2(\Omega)}.$$

For the second term, we will show $\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|^2_{L^2(e)} \leq Ch^4\|\varphi\|^2_{H^2(\Omega)}$ and then take the square root of both sides.

$$\begin{aligned}
\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|^2_{L^2(e)} &\leq C(\sum_{T\in T_h} \|\varphi - I_h\varphi\|^2_{L^2(T)} + h_T^2\|\nabla(\varphi - I_h\varphi)\|^2_{L^2(T)}) \text{ (by (4.4))} \\
&\leq C(\|\varphi - I_h\varphi\|^2_{L^2(\Omega)} + \sum_{T\in T_h} h_T^2\|\varphi - I_h\varphi\|^2_{H^1(T)}) \\
&\leq C(\|\varphi - I_h\varphi\|^2_{L^2(\Omega)} + h^4\sum_{T\in T_h} \|\varphi\|^2_{H^2(T)}) \text{ (by Lemma 7)} \\
&\leq C(h^4\|\varphi\|^2_{H^2(\Omega)} + h^4\|\varphi\|^2_{H^2(\Omega)}) \\
&= Ch^4\|\varphi\|^2_{H^2(\Omega)}.
\end{aligned}$$

Therefore, after taking the square roots of both sides, we have

$$(\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|^2_{L^2(e)})^{1/2} \leq Ch^2\|\varphi\|_{H^2(\Omega)},$$

and hence, both terms are bounded above by $Ch^2\|\varphi\|_{H^2(\Omega)}$. □

Theorem 2. *There holds for sufficiently small h,*

$$\|v_h\|_{H_h^2(\Omega)} \leq C\|L_h v_h\|_{L_h^2(\Omega)}. \tag{4.18}$$

Proof. Let $v_h \in X_h$, and let $\varphi \in H_0^1(\Omega)$ satisfy

$$\begin{aligned}
-A : D^2\varphi &= v_h \text{ in } \Omega, \\
\varphi &= 0 \text{ on } \partial\Omega.
\end{aligned}$$

Then by elliptic regularity, $\varphi \in H^2(\Omega)$ with $\|\varphi\|_{H^2(\Omega)} \leq C\|v_h\|_{L^2(\Omega)}$ [7]. For the purpose of this proof let's redefine what the operator L_h is. We redefine the domain of L_h to be

$X = (\prod_{T\in T_h} H^2(T)) \cap H_0^1(\Omega)$. Redefine $\bar{L}_h$ similarly. Now

$$\begin{aligned}
\|v_h\|^2_{L^2(\Omega)} &= -\int_\Omega (A : D^2\varphi) v_h \, dx = -\int_\Omega (A : D^2\varphi) v_h \, dx + \sum_{e\in\mathcal{E}_h^I} \int_e [A\nabla\varphi] v_h \, dx \\
&\quad (\text{ since } \varphi \in H^2(\Omega), [\nabla\varphi] = 0) \\
&= \langle L_h\varphi, v_h\rangle \text{ (by (4.17))} \\
&= \langle \bar{L}_h\varphi, v_h\rangle + \langle (L_h - \bar{L}_h)\varphi, v_h\rangle \text{ (from adding and subtracting } \bar{L}_h) \\
&= \langle \bar{L}_h v_h, \varphi\rangle + \langle (L_h - \bar{L}_h)\varphi, v_h\rangle \text{ (by Lemma 14)} \\
&= \langle L_h v_h, \varphi\rangle + \langle (L_h - \bar{L}_h)\varphi, v_h\rangle + \langle (\bar{L}_h - L_h)v_h, \varphi\rangle \\
&\quad \text{(from adding and subtracting } L_h) \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

We will first bound I_2.

$$\begin{aligned}
I_2 &\le \|A - \bar{A}\|_{L^\infty(\Omega)} [\sum_{T\in T_h} \|D^2\varphi\|_{L^2(T)} \|v_h\|_{L^2(T)} + \sum_{e\in\mathcal{E}_h^I} \|[\nabla\varphi]\|_{L^2(e)} \|v_h\|_{L^2(e)}] \\
&= \|A - \bar{A}\|_{L^\infty(\Omega)} \sum_{T\in T_h} \|D^2\varphi\|_{L^2(T)} \|v_h\|_{L^2(T)} \ \ (\text{since } [\nabla\varphi] = 0) \\
&\le \|A - \bar{A}\|_{L^\infty(\Omega)} (\sum_{T\in T_h} \|D^2\varphi\|^2_{L^2(T)})^{1/2} (\sum_{T\in T_h} \|v_h\|^2_{L^2(T)})^{1/2} \ \ (\text{by Cauchy-Schwarz}) \\
&\le \|A - \bar{A}\|_{L^\infty(\Omega)} (\sum_{T\in T_h} \|\varphi\|^2_{H^2(T)})^{1/2} (\sum_{T\in T_h} \|v_h\|^2_{L^2(T)})^{1/2} \\
&= \|A - \bar{A}\|_{L^\infty(\Omega)} \|\varphi\|_{H^2(\Omega)} \|v_h\|_{L^2(\Omega)} \\
&\le Ch^\alpha \|\varphi\|_{H^2(\Omega)} \|v_h\|_{L^2(\Omega)} \\
&\le Ch^\alpha \|v_h\|^2_{L^2(\Omega)}. \qquad (4.19)
\end{aligned}$$

We now bound I_3 using similar arguments as in the proof of Lemma 15.

$$\begin{aligned}
I_3 &\le \|A - \bar{A}\|_{L^\infty(\Omega)} [\sum_{T\in T_h} \int_T |D^2 v_h| |\varphi| \, dx + \sum_{e\in\mathcal{E}_h^I} \int_e |[\nabla v_h]| |\varphi| \, ds] \\
&\le \|A - \bar{A}\|_{L^\infty(\Omega)} [\sum_{T\in T_h} \|D^2 v_h\|_{L^2(T)} \|\varphi\|_{L^2(T)} + \sum_{e\in\mathcal{E}_h^I} \|[\nabla v_h]\|_{L^2(e)} \|\varphi\|_{L^2(e)}]
\end{aligned}$$

$$
\begin{aligned}
&\leq \|A-\bar{A}\|_{L^\infty(\Omega)}[(\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)})^{1/2}(\sum_{T\in T_h}\|\varphi\|^2_{L^2(T)})^{1/2} \\
&\quad + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}(\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}] \\
&= \|A-\bar{A}\|_{L^\infty(\Omega)}[\|\varphi\|_{L^2(\Omega)}(\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)})^{1/2} \\
&\quad + (\sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}(\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}].
\end{aligned}
$$

Adding extra terms under the square roots, we have

$$
\begin{aligned}
I_3 \leq &\|A-\bar{A}\|_{L^\infty(\Omega)}[\|\varphi\|_{L^2(\Omega)}(\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2} \\
&+ (\sum_{T\in T_h}\|D^2 v_h\|^2_{L^2(T)} + \sum_{e\in\mathcal{E}_h^I} h_e^{-1}\|[\nabla v_h]\|^2_{L^2(e)})^{1/2}(\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}] \\
= &\|A-\bar{A}\|_{L^\infty(\Omega)}[\|\varphi\|_{L^2(\Omega)}\|v_h\|_{H_h^2(\Omega)} + \|v_h\|_{H_h^2(\Omega)}(\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}] \quad \text{(by (4.3))}.
\end{aligned}
$$

Factoring out $\|v_h\|_{H_h^2(\Omega)}$, we now have

$$
I_3 = \langle(\bar{L}_h - L_h)v_h, \varphi\rangle \leq \|A-\bar{A}\|_{L^\infty(\Omega)}\|v_h\|_{H_h^2(\Omega)}[\|\varphi\|_{L^2(\Omega)} + (\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}]. \tag{4.20}
$$

And since $\|A-\bar{A}\|_{L^\infty(\Omega)} \leq Ch^\alpha$, we have

$$
I_3 = \langle(\bar{L}_h - L_h)v_h, \varphi\rangle \leq Ch^\alpha\|v_h\|_{H_h^2(\Omega)}[\|\varphi\|_{L^2(\Omega)} + (\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)})^{1/2}].
$$

By the trace inequality (4.4),

$$
\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi\|^2_{L^2(e)} \leq C\sum_{T\in T_h}(h_T^2\|\nabla\varphi\|^2_{L^2(T)} + \|\varphi\|^2_{L^2(T)}) \leq c\|\varphi\|^2_{H^1(\Omega)}.
$$

Therefore,

$$
I_3 \leq Ch^\alpha\|v_h\|_{H_h^2(\Omega)}\|\varphi\|_{H^1(\Omega)} \leq Ch^\alpha\|v_h\|_{H_h^2(\Omega)}\|\varphi\|_{H^2(\Omega)} \leq Ch^\alpha\|v_h\|_{H_h^2(\Omega)}\|v_h\|_{L^2(\Omega)}. \tag{4.21}
$$

Let $I_h\varphi \in X_h$ be such that

$$
\|I_h\varphi\|_{L^2(\Omega)} \leq C\|\varphi\|_{L^2(\Omega)} \leq C\|\varphi\|_{H^2(\Omega)} \leq C\|v_h\|_{L^2(\Omega)}. \tag{4.22}
$$

Finally, we look to bound I_1 by first splitting it up into two terms.

$$I_1 = \langle L_h v_h, I_h\varphi\rangle + \langle L_h v_h, \varphi - I_h\varphi\rangle.$$

Now,

$$\langle L_h v_h, I_h\varphi\rangle \leq \|L_h v_h\|_{L_h^2(\Omega)}\|I_h\varphi\|_{L^2(\Omega)}.$$

Replacing $\bar{L}_h - L_h$ with L_h and replacing φ with $\varphi - I_h\varphi$ in (4.20), we can bound the second term of I_1 as follows.

$$\begin{aligned}\langle L_h v_h, \varphi - I_h\varphi\rangle &\leq C\|A\|_{L^\infty(\Omega)}\|v_h\|_{H_h^2(\Omega)}(\|\varphi - I_h\varphi\|_{L^2(\Omega)} + (\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|_{L^2(e)}^2)^{1/2})\\ &= C\|v_h\|_{H_h^2(\Omega)}(\|\varphi - I_h\varphi\|_{L^2(\Omega)} + (\sum_{e\in\mathcal{E}_h^I} h_e\|\varphi - I_h\varphi\|_{L^2(e)}^2)^{1/2})\\ &\leq Ch^2\|v_h\|_{H_h^2(\Omega)}\|v_h\|_{L^2(\Omega)} \text{ (by Lemma 16 and (4.22)).}\end{aligned}$$

Combining these two bounds, we have

$$I_1 \leq \|L_h v_h\|_{L_h^2(\Omega)}\|\varphi_h\|_{L^2(\Omega)} + Ch^2\|v_h\|_{H_h^2(\Omega)}\|v_h\|_{L^2(\Omega)}. \tag{4.23}$$

Combining estimates from (4.23), (4.19), and (4.21) we get the following inequality:

$$\begin{aligned}\|v_h\|_{L^2(\Omega)}^2 \leq{}& \|L_h v_h\|_{L_h^2(\Omega)}\|v_h\|_{L^2(\Omega)} + Ch^2\|v_h\|_{H_h^2(\Omega)}\|v_h\|_{L^2(\Omega)}\\ &+ Ch^\alpha\|v_h\|_{H_h^2(\Omega)}\|v_h\|_{L^2(\Omega)} + Ch^\alpha\|v_h\|_{L^2(\Omega)}^2.\end{aligned} \tag{4.24}$$

Therefore

$$(1 - Ch^\alpha)\|v_h\|_{L^2(\Omega)}^2 \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + h^\alpha\|v_h\|_{H_h^2(\Omega)})\|v_h\|_{L^2(\Omega)}.$$

Taking h sufficiently small, we have

$$\|v_h\|_{L^2(\Omega)} \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + h^\alpha\|v_h\|_{H_h^2(\Omega)}). \tag{4.25}$$

Combining (4.25) and Corollary 1 then gives us

$$\|v_h\|_{H_h^2(\Omega)} \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + \|v_h\|_{L^2(\Omega)}) \leq C(\|L_h v_h\|_{L_h^2(\Omega)} + h^\alpha\|v_h\|_{H_h^2(\Omega)}).$$

Rearranging terms we get:

$$(1 - Ch^{\alpha})\|v_h\|_{H_h^2(\Omega)} \leq C\|L_h v_h\|_{L_h^2(\Omega)}.$$

For h sufficiently small we get the following stability estimate:

$$\|v_h\|_{H_h^2(\Omega)} \leq C\|L_h v_h\|_{L_h^2(\Omega)}.$$

□

4.4 EXISTENCE, UNIQUENESS AND ERROR ESTIMATES

Theorem 3. *Assume that (4.13) is satisfied. Then for h sufficiently small, there exists a unique solution to the finite element method (2.13).*

Proof. Since existence is equivalent to uniqueness for a linear operator in a finite dimensional setting, it is enough to show uniqueness. Suppose $u_{1,h}$ and $u_{2,h}$ are two solutions to (2.13). Setting $u_h = u_{1,h} - u_{2,h}$, we have $\langle L_h u_h, v_h\rangle = 0 \ \forall v_h \in X_h$. Then by Theorem 2, $\|u_h\|_{H_h^2(\Omega)} \leq C\|L_h u_h\|_{L_h^2(\Omega)} = 0$ and therefore $u_h \equiv 0$. □

Theorem 4. *Suppose $u \in H^s(\Omega)$ $(s \geq 2)$ is an exact solution of (1.4). Then*

$$\|u - u_h\|_{H_h^2(\Omega)} \leq Ch^{l-2}\|u\|_{H^l(\Omega)},$$

where $\ell = \min\{k+1, s\}$, and k is the polynomial degree of the finite element space.

Proof. Let $I_h u$ be the interpolant satisfying $\|u - I_h u\|_{H_h^2(\Omega)} \leq Ch^{l-2}\|u\|_{H^l(\Omega)}$. See Lemma 8. Note that $\langle L_h(u - u_h), v_h\rangle = 0 \ \forall v_h \in X_h$. Therefore

$$\begin{aligned}
\|u_h - I_h u\|_{H_h^2(\Omega)} &\leq C\|L_h(u_h - I_h u)\|_{L_h^2(\Omega)} \text{ (by (4.18))} \\
&= C \sup_{v_h \in X_h, v_h \neq 0} \frac{\langle L_h(u_h - I_h u), v_h\rangle}{\|v_h\|_{L^2(\Omega)}} \text{ (by (4.2))} \\
&= C \sup_{v_h \in X_h, v_h \neq 0} \frac{\langle L_h(u - I_h u), v_h\rangle}{\|v_h\|_{L^2(\Omega)}} \\
&\leq C\|A\|_{L^\infty(\Omega)}\|u - I_h u\|_{H_h^2(\Omega)} \text{ (by Lemma 15)}
\end{aligned}$$

$$= C\|u - I_h u\|_{H^2_h(\Omega)}$$
$$\leq Ch^{\ell-2}\|u\|_{H^\ell(\Omega)} \text{ (by Lemma 8).} \tag{4.26}$$

Using the triangle inequality to complete the proof of the theorem, we have:

$$\|u - u_h\|_{H^2_h(\Omega)} \leq \|u_h - I_h u\|_{H^2_h(\Omega)} + \|u - I_h u\|_{H^2_h(\Omega)}$$
$$\leq Ch^{\ell-2}\|u\|_{H^l(\Omega)} \text{ (by (4.26) and (4.7)).}$$

□

5.0 NUMERICAL EXPERIMENTS

In this chapter, we use Matlab [11] to implement the finite difference scheme discussed in Chapter 3 for three test problems in non-divergence form. We record the error estimates, the convergence rates, and compare the results to the theory developed in Chapters 3-4 if applicable. See the Appendix for the Matlab codes used to implement the finite difference scheme for each test problem.

5.0.1 Test 1

Consider the following example:

$$\begin{aligned} xu''(x) &= 1 \text{ on } (0,1), \\ u(0) &= u(1) = 0. \end{aligned} \tag{5.1}$$

We know the solution to be $u(x) = x\log(x)$. Using the calculations below:

$$\int_0^1 (x\log(x))^2\,dx < \infty \text{ and}$$

$$\int_0^1 (\frac{d}{dx}x\log(x))^2\,dx = \int_0^1 (\log(x)+1)^2\,dx < \infty. \text{ But,}$$

$$\int_0^1 (\frac{d^2}{dx^2}x\log(x))^2\,dx = \int_0^1 \frac{1}{x^2}\,dx = \infty,$$

we see that the solution u belongs to the space H^1 but not in H^2. We see that this problem does not fit within the framework of our research. In particular, the problem is degenerate

Table 5.1: Convergence Rates for (5.1)

N	h	error	rate
7	1.2500e-1	0.045198049852	
15	6.2500e-2	0.024839068556	0.8636
31	3.1250e-2	0.013380845425	0.8924
61	1.6129e-2	0.007232058477	0.9303
121	8.1967e-3	0.003794548940	0.9528

as $A(x) = -x$ is not uniformly positive. Nonetheless, according to Table 5.1, the method still converges with order $\mathcal{O}(h)$.

Here, the rate is the ratio of the difference between the log of error values with the difference between the log of h values. For example, The first rate value 0.8636 is computed using the following formula.

$$\frac{\log(0.045198049852) - \log(0.024839068556)}{\log(1.25e-1) - \log(6.25e-2)} \approx 0.8636.$$

Also, the *error* is calculated by taking the Euclidean norm of the difference between the approximated solution vector and the exact solution vector, i.e., the error at the gridpoints.
Observation: Note that we solved (5.1) using the following finite difference scheme

$$\frac{x_i u_{i-1} - 2x_i u_i + x_i u_{i+1}}{h^2} = 1. \tag{5.2}$$

Consider using the equivalent difference scheme

$$\frac{u_{i-1} - 2u_i + u_{i+1}}{h^2} = \frac{1}{x_i}. \tag{5.3}$$

As expected, we get the same solution. See Table 5.2 for the error values. The error here is the Euclidean norm of the difference between the approximated solution using (5.2) and the approximated solution using (5.3).

Table 5.2: Error values from using (5.2) and (5.3)

N	error
7	8.3267e-17
15	1.0007e-16
31	3.2469e-15
61	3.2021e-14
121	1.7757e-13

Also, it is noteworthy to point out that the condition number of the 2 matrices from the linear system developed from (5.2) and (5.3) are different. For example, for N=121 the condition number of the matrix generated by (5.2) is 1.7753e+04 and the condition number of the matrix generated by (5.3) is 6.0316e+03. See Table 5.3 and Table 5.4 for the rates of the condition number for both (5.2) and (5.3). In both cases the condition number is of order $\mathcal{O}(h^{-2})$.

Table 5.3: Condition Number for (5.2)

N	h	Condition Number	rate
7	1.2500e-1	50.035846467572995	
15	6.2500e-2	2.444135562222110e+02	-2.288290365585845
31	3.1250e-2	1.097554815559039e+03	-2.166896784044571
61	1.6129e-2	4.400602155220470e+03	-2.099569945491609
121	8.1967e-3	1.775346355450996e+04	-2.060667004336193

Table 5.4: Condition Number for (5.3)

N	h	Condition Number	rate
7	1.2500e-1	25.274142369088249	
15	6.2500e-2	1.030868689198178e+02	-2.028126532762339
31	3.1250e-2	4.143450622319003e+02	-2.006972153950930
61	1.6129e-2	1.557247895814273e+03	-2.001776984033977
121	8.1967e-3	6.031591333713156e+03	-2.000467234947457

5.0.2 Test 2

Consider a similar problem:

$$\begin{aligned} x^a u''(x) &= 1 \text{ for } a \neq 2, 1 \text{ on } (0,1), \\ u(0) &= u(1) = 0. \end{aligned} \tag{5.4}$$

The solution is given by

$$u(x) = \frac{1}{(1-a)(2-a)} x^{2-a} - \frac{1}{(1-a)(2-a)} x. \tag{5.5}$$

We see that this particular problem does not fit within the framework of research. One reason being, $A = x^\alpha$ is not uniformly positive definite. Nonetheless, let's explore what values of m does the solution of (5.4) belong to H^m. We can see that $u \approx x^{2-a}$. Differentiating, we have $u^{(m)} \approx x^{2-a-m}$. Plugging this expression into the integral we then have

$$\int_0^1 |u^{(m)}|^2 \, dx \approx \int_0^1 x^{2(2-a-m)} \, dx = C(1 - \lim_{x \to 0} x^{2(2-a-m)+1}).$$

Therefore we need $2(2-a-m)+1 > 0$ in order for the above quantity to be finite. For $m < \frac{5}{2} - a$, $u \in H^m$. See Tables 5.5, 5.6, and 5.7 for the convergence rates for specified values of a. For $a = 3/2$ we see the method converges with order $\mathcal{O}(h^{1/2})$. For $a = 1/2$ we see the method converges with order $\mathcal{O}(h^{3/2})$. Lastly, for $a = 1/3$ we see that the method converges with order $\mathcal{O}(h^{5/3})$.

Table 5.5: Convergence Rates for (5.2) for $a = 3/2$

N	h	error	rate
7	1.2500e-1	0.426222527483721	
15	6.2500e-2	0.324108453815603	0.3951
31	3.1250e-2	0.237234847711760	0.4502
61	1.6129e-2	0.175895283631180	0.4523
121	8.1967e-3	0.127539673580250	0.4749

Table 5.6: Convergence Rates for (5.2) for $a = 1/2$

N	h	error	rate
7	1.2500e-1	0.004634320300970	
15	6.2500e-2	0.001970169670021	1.2340
31	3.1250e-2	7.853900122705287e-04	1.3268
61	1.6129e-2	3.171045712105497e-04	1.3713
121	8.1967e-3	1.226996491915239e-04	1.4027

Table 5.7: Convergence Rates for (5.2) for $a = 1/3$

N	h	error	rate
7	1.2500e-1	0.001906638910787	
15	6.2500e-2	7.348467650385854e-04	1.3755
31	3.1250e-2	2.686587026141385e-04	1.4517
61	1.6129e-2	9.936768938939577e-05	1.5038
121	8.1967e-3	3.507334063203305e-05	1.5385

5.0.3 Test 3

Now consider the following ODE:

$$(|x-\frac{1}{2}|+1)u''(x) = -1 \quad \text{on } (0,1),$$
$$u(0) = u(1) = 0. \tag{5.6}$$

$$\text{Solution: } u(x) = \begin{cases} (x-\frac{3}{2})\ln|x-\frac{3}{2}| - x + \frac{3}{2}\ln(\frac{3}{2}), & \text{if } x < \frac{1}{2}, \\ -(x+\frac{1}{2})\ln|x+\frac{1}{2}| + x - 1 + \frac{3}{2}\ln(\frac{3}{2}), & \text{if } x \geq \frac{1}{2}. \end{cases}$$

This problem fits within the framework of our research. One reason being, $A(x)$ is α-Hölder continuous for $\alpha = 1$.

$$|A(x) - A(y)| = ||x-\frac{1}{2}| - |y-\frac{1}{2}|| \leq |x-y|.$$

$A(x)$ is also continuous since the absolute value function is continuous. $A(x)$ is also uniformly positive definite.

$$1 \leq |x-\frac{1}{2}| + 1 < \frac{3}{2}.$$

We also see that the solution belongs to H^2 since

$$\int_0^1 (u''(x))^2\,dx = \int_0^1 \frac{1}{(|x-\frac{1}{2}|+1)^2}\,dx < \infty.$$

See Table 5.8 for the rates of convergence. As seen from the table, the method converges with order $\mathcal{O}(h^2)$. However, as noted before, we calculate the error at the grid points. Therefore Theorem 4 is not applicable. See Table 5.9 for the condition numbers of the matrix generated by the finite difference scheme that implements (5.6). As can be seen, the condition number is of order $\mathcal{O}(h^{-2})$.

Table 5.8: Convergence Rates for (5.6)

N	h	error	rate
7	1.2500e-1	0.00108264086805	
15	6.2500e-2	2.711146245239493e-04	1.816748614206876
31	3.1250e-2	6.780727915188522e-05	1.909080460718710
61	1.6129e-2	1.806499358113223e-05	1.954084222366156
121	8.1967e-3	4.665665851930068e-06	1.976511141207573

Table 5.9: Condition Number for (5.6)

N	h	Condition Number	rate
7	1.2500e-1	26.284093272308411	
15	6.2500e-2	1.142618997077713e+02	-2.120082549410456
31	3.1250e-2	4.890808984918833e+02	-2.097728698634913
61	1.6129e-2	1.913120479391648e+03	-2.062234714037769
121	8.1967e-3	7.608571324558045e+03	-2.039541637292652

6.0 APPENDIX

6.1 MATLAB CODE FOR NUMERICAL TEST 1

```
function [condA,x,c]=findiffscheme(N);
%[x,u]=findiffscheme(N);
% N is the number of mesh points on (0,1) used
% corresponds to the ode x*c''=1

%Lauren Hennings

dx=1/(N+1);
x=dx*(0:N+1);
A=zeros(N);
for k=2:N
    A(k,k-1)=k*dx;
end
for k=1:N-1
    A(k,k+1)=k*dx;
end
for k=1:N
    A(k,k)=-2*k*dx;
end

CLEFT=0;
CRIGHT=0;

b=ones(N,1);
for n=1:N
    b(n,1)=dx^2;
end
b(1,1)=(dx^2)-(dx*CLEFT);
b(N,1)=(dx^2)-(N*dx*CRIGHT);

c=A\b;
c=[CLEFT;c;CRIGHT];

condA=cond(A);
```

6.1.1 Matlab Code for Numerical Test 1 using finite difference scheme (5.3)

```
function [condA,x,c]=findiffscheme2(N);
%[x,u]=findiffscheme(N);
% N is the number of mesh points on (0,1) used
% corresponds to the ode c''=1/x

%Lauren Hennings

dx=1/(N+1);
x=dx*(0:N+1);
A=-2*diag(ones(N,1))+diag(ones(N-1,1),1)+diag(ones(N-1,1),-1);

CLEFT=0;
CRIGHT=0;

b=ones(N,1);
for n=1:N
    b(n,1)=dx^2/(x(n+1));
end
b(1,1)=b(1)-(CLEFT);
b(N,1)=b(N)-(CRIGHT);

c=A\b;
c=[CLEFT;c;CRIGHT];

condA=cond(A);
```

6.2 MATLAB CODE FOR NUMERICAL TEST 2

```
function [x,c]=findiffschemea(N,a);
%[x,u]=findiffschemea(N);
%corresponds to the ode (x^a)*c''=1 where a is a fixed constant greater
%than 0.

%Lauren Hennings

dx=1/(N+1);
x=dx*(0:N+1);
A=zeros(N);
for k=2:N
    A(k,k-1)=(k*dx)^a;
end
for k=1:N-1
    A(k,k+1)=(k*dx)^a;
end
for k=1:N
```

```
    A(k,k)=-2*(k*dx)^a;
end

CLEFT=0;
CRIGHT=0;

b=ones(N,1);
for n=1:N
    b(n,1)=dx^2;
end
b(1,1)=(dx^2)-((dx^a)*CLEFT);
b(N,1)=(dx^2)-(((N*dx)^a)*CRIGHT);

c=A\b;
c=[CLEFT;c;CRIGHT];
```

```
function [c]=truesolution(x,a);
%calculates the true solution c of x^a*c''=1 and evaluates it at x.
%Lauren Hennings

c=(1/((1-a)*(2-a)))*[(x.^(2-a))-x];
```

6.3 MATLAB CODE FOR NUMERICAL TEST 3

```
function [condA,x,c]=piecewisefindiffscheme(N);
%[x,u]=findiffscheme(N);
%corresponds to the ode f(x)*c''=-1 where f(x) is |x-0.5|+1
%Lauren Hennings
dx=1/(N+1);
x=dx*(0:N+1);
A=zeros(N);
for k=2:N
    A(k,k-1)=-abs((k*dx)-0.5)-1;
end
for k=1:N-1
    A(k,k+1)=-abs((k*dx)-0.5)-1;
end
for k=1:N
    A(k,k)=2*abs((k*dx)-0.5)+2;
end

CLEFT=0;
CRIGHT=0;

b=ones(N,1);
for n=1:N
    b(n,1)=dx^2;
```

```
end
b(1,1)=(dx^2)+(abs(dx-0.5)+1)*CLEFT;
b(N,1)=(dx^2)+(abs((N*dx)-0.5)+1)*CRIGHT;

c=A\b;
c=[CLEFT;c;CRIGHT];

condA=cond(A);
```

BIBLIOGRAPHY

[1] R. Adams, and J.J.F Fournier, *Sobolev Spaces*, Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London, 1975.

[2] K.J. Bathe, *Numerical Methods in Finite Element Analysis*, Prentice Hall, 1976. Print.

[3] R. Bellman and S. Dreyfus, *An Application of Dynamic Programming to the Determination of Optimal Satellite Trajectories*. J. Brit. Interplanet. Soc. 17 1959 78-83.

[4] S. Bernstein, *Sur La Généralisation du Probléme de Dirichlet. (French)*, Springer, Math. Ann. 69 (1910), no. 1, 82-136.

[5] S.C. Brenner and L.R. Scott, *The Mathematical Theory of Finite Element Methods*, third edition, Springer, 2008.

[6] P.G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland, Amsterdam, 1978.

[7] D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*. Second edition, Springer, 1983.

[8] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*. Society for Industrial and Applied Mathematics, 2011.

[9] C. Johnson, *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press, 1987. Print.

[10] O.A. Ladyzhenskaya and N.N. Ural'tseva, *Linear and Quasilinear Elliptic Equations*. Acacemic Press, New York and London, 1968.

[11] MATLAB and Statistics Toolbox Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States.

[12] M. Neilan, *Quadratic Finite Element Approximations of the Monge-Ampere Equation*, Journal of Scientific Computing, Volume 54 Issue 1, January 2013. 200-226.

[13] A. Quarteroni, F. Saleri, and R. Sacco, *Numercal Mathematics*. New York: Springer, 2000. Print.

[14] I. Smears and E. Suli, *Discontinous Galerkin Finite Element Approximation of Nondivergence Form Elliptic Equations with Cordes Coefficients*, SIAM J. Numer. Anal. Vol. 51, No. 4, pp. 2088-2106. 2013.