DISCONTINUOUS GALERKIN METHODS FOR NONLINEAR ELLIPTIC PROBLEMS

Thesis
Submitted in partial fulfillment of the requirements
For the degree of

Doctor of Philosophy

by

Thirupathi Gudi
(01409006)

Under the Supervision of

Supervisor : Professor Amiya Kumar Pani

Co-supervisor : Professor Neela Nataraj

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY
2006

Approval Sheet

The thesis entitled

"DISCONTINUOUS GALERKIN METHODS FOR NONLINEAR ELLIPTIC PROBLEMS"

by

Thirupathi Gudi

is approved for the degree of

DOCTOR OF PHILOSOPHY

Examiners

_____

_____

Supervisors

_____

_____

Chairman

_____

Date : _____

Place : _____

Dedicated to

My Wife and My Son

Vijaya
&
Srivarshith

INDIAN INSTITUTE OF TECHNOLOGY, BOMBAY, INDIA

CERTIFICATE OF COURSE WORK

This is to certify that Mr. Thirupahi Gudi was admitted to the candidacy of the Ph.D. Degree on 07th January 2002, after successfully completing all the courses required for the Ph.D. Degree programme. The details of the course work done are given below.

Sr. No.	Course No.	Course Name	Credits
1.	MA 436	Partial Differential Equations	6.00
2.	MA 828	Functional Analysis	6.00
3.	MA 838	Special Topics in Mathematics II	6.00
4.	MA 529	Numerical Methods for Partial Differential Equations	6.00
5.	MA 827	Analysis	6.00
6.	MA 829	Mathematical Methods	6.00
7.	MA 525	Dynamical Systems	0.00

I. I. T. Bombay — Dy. Registrar (Academic)

Dated :

Abstract

The main focus of this thesis has been on the study of *hp*-discontinuous Galerkin (DG) methods for quasilinear and strongly nonlinear elliptic problems of nonmonotone-type. Amongst all the DG methods which are introduced in the literature, we concentrate on the Symmetric Interior Penalty Galerkin (SIPG), Non-symmetric Interior Penalty Galerkin (NIPG) and Local Discontinuous Galerkin (LDG) Finite Element Methods (FEM). The main emphasis is on the existence and uniqueness of the proposed DG schemes and their related error estimates in the broken H^1-norm and possibly in the L^2-norm. As a key tool in proving the well-posedness for each of the discrete DG schemes (SIPG, NIPG and LDG), the nonlinear system is rewritten in a fixed point form. For fixed point formulation, the corresponding nonlinear system is linearized and a map $S_h : \mathcal{O}_\delta(I_h u) \subset V_h \to V_h$, where $\mathcal{O}_\delta(I_h u)$ is a ball in the discontinuous finite element space V_h of radius $\delta = \delta(h)$ and centered at an interpolant $I_h u \in V_h$ of u is defined. Any fixed point of S_h is indeed a solution to the nonlinear system of DG or LDG method. It is then shown that S_h is Lipschitz continuous and it maps a ball $\mathcal{O}_\delta(I_h u)$ into itself. An appeal to Brouwer fixed point theorem yields the existence of a solution to the discrete problem and further a use of Lipschitz continuity of S_h implies uniqueness of the solution.

Then we proceed to derive *a priori* error estimates for the proposed DG and LDG schemes. The derived *a priori* error estimates in the broken energy norm are optimal in h (mesh size) and slightly suboptimal in p (degree of approximation). These estimates lead precisely to the same optimal in h and suboptimal in p in the case of the linear elliptic problems. The adjoint consistency plays a vital role in deriving optimal error estimates in the L^2-norm. Since the SIPG and LDG methods are adjoint consistent, *a priori* error estimates which are optimal in h and slightly suboptimal in p are derived for SIPG and LDG schemes. As the NIPG method is not adjoint consistent, it is difficult to derive optimal error estimate in L^2 as it stands. However, with additional assumptions that the mesh is regular and the Dirichlet boundary data is either a piecewise polynomial or zero, optimal order of convergence in h as well as in p for the NIPG method is derived.

Numerical experiments are presented for each of the method proposed in this thesis. The experiments confirm the theoretical order of convergence obtained in each of the Chapters through Chapter 2 to Chapter 5. Numerical experiments are also presented for super-penalty methods and they confirm the theoretical results. Further, some computational results are derived for mean curvature problem. Finally, the possible extensions with scope for future investigations are discussed in the concluding Chapter.

Contents

List of Figures

Chapter 1

Introduction

The main objective of this dissertation is to study *hp*-discontinuous Galerkin Finite Element Methods (DGFEM) for nonlinear elliptic problems. In this study, we mainly focus on the most popular DG schemes such as Symmetric Interior Penalty Galerkin (SIPG), Non-symmetric Interior Penalty Galerkin (NIPG) and Local Discontinuous Galerkin (LDG) methods for a class of non-monotone quasilinear and strongly nonlinear elliptic problems.

1.1 Motivation

In recent years, there has been a renewed interest in Discontinuous Galerkin (DG) methods for the numerical solution of a wide range of partial differential equations. This is due to their flexibility in local mesh adaptivity and in handling nonuniform degrees of approximation for solutions whose smoothness exhibit variation over the computational domain. Besides, they are elementwise conservative and easy to implement than the finite volume methods and the standard mixed finite element methods with high degree of piecewise polynomials.

The first work on DG methods for the elliptic and parabolic problems trace back to 1970's by Douglas *et al.* [38], Wheeler [66] and Arnold [4]. In 1971, Nitsche [54] introduced the concept of enforcing the Dirichlet boundary conditions weakly rather than incorporating into the finite element space by means of adding a penalty term to the variational formulation. In 1973, Babuska [6] introduced another penalty method to impose the Dirich-

let boundary condition weakly. Interior Penalty (IP) methods by Wheeler [66] and Arnold [4] arose from the observation that just as Dirichlet boundary conditions, inter element continuity could be imposed weakly instead of being built into the finite element space. This makes it possible and easier to use the spaces of discontinuous piecewise polynomials of higher degree. The IP methods are, presently, called Symmetric Interior Penalty Galerkin (SIPG) methods. The variational formulation of the SIPG method is symmetric and adjoint consistent, but the stabilizing penalty parameter in this method depends on the bounds of the coefficients of the problem and various constants in the inverse inequalities which are not known explicitly. To over come this shortcoming, Oden, Babuška and Baumann [55] proposed, recently, another DG method for the diffusion problems which is based on a non-symmetric formulation. This method is shown to be stable when the degree of approximation is greater than or equal to 2 [55], [61]. Riviére *et al.* [61] and Houston *et al.* [44] introduced and analyzed the Non-symmetric Interior Penalty Discontinuous Galerkin (NIPG) method which stabilizes the DG method of Oden, Babuška and Baumann [55] for any degree of approximation which is greater than or equal to 1. Subsequently, there are other variants of DG methods appeared in the literature to approximate diffusion and convection-diffusion problems, see [15], [5].

On the other hand, Reed and Hill [59] introduced the first DG method for hyperbolic equations. Around the same time, Lesiant and Raviart [51] derived *a priori* error estimates for the DG method when applied to the linear hyperbolic problems. Since then, there has been an active development on DG methods for hyperbolic and nearly hyperbolic problems such as problems with dominant convective part as well as nonnegligible diffusion part. Cockburn and coauthors developed the Runge-Kutta DG (RKDG) methods in a series of papers to achieve stability, higher order accuracy and convergence for the scalar conservation laws [24], [25], [26] and [27]. From the observation that mixed methods can handle elliptic operators very well and use a discontinuous approximation for potential, several methods which deal with discontinuous approximation to convective part and mixed method to diffusive part were proposed [33],[34]. Using the ideas of RKDG methods, Bassi - Rebay [11] introduced a new and completely discontinuous approximation for both convective part and diffusive part for the compressible Navier - Stokes equations.

In 1998, Cockburn and Shu [28] introduced Local Discontinuous Galerkin (LDG) method to approximate the general convection-diffusion problems by generalizing the original DG method of Bassi - Rebay [11]. The proposed LDG method is in mixed formulation and uses the approximation to displacement and to each component of velocity from the same space. Therefore, the coding of LDG method is simpler than the standard mixed method. The first attempt to apply the LDG method to purely elliptic problems was made in [18]. Since the LDG method in [18] is consistent and stable, this leads to an optimal order of convergence. Brezzi *et al.* [15] proposed a method which deals with the lifting operators and the primal formulation of the DG method of Bassi-Rebay [11]. In [5], a unified frame work of all the discontinuous Galerkin methods appeared in the literature was proposed and analyzed which, subsequently, became crucial in the unified adaptive methods, see [19].

Except for [45] and [17], there is hardly any result on DG methods for the nonlinear elliptic problems. In [45], the authors have applied a one parameter family of the DG methods to a class of monotone quasilinear elliptic problems while the authors of [17] have applied LDG method for a similar class of problems. In this dissertation, efforts are made to analyze the SIPG, NIPG and LDG methods for a class of quasilinear and strongly nonlinear elliptic problems which are of non-monotone type. Moreover, the analysis discussed in this thesis can be used to generalize the results of [45] and [17].

1.2 Preliminaries

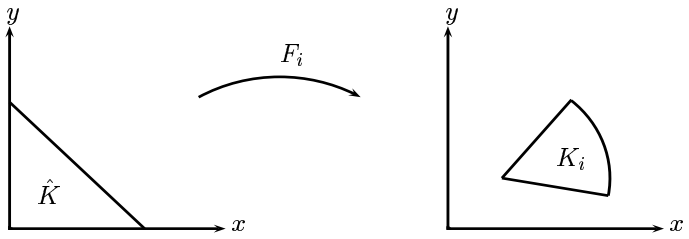
We split this Section into several subsections to introduce preliminaries which are used throughout this thesis. In the Subsection 1.2.1, we start with discontinuous finite element subdivision $\mathcal{T}_h$ of the computational domain Ω and introduce the interior Γ_I and boundary Γ_∂ edges which appear in the discontinuous Galerkin formulations. In the Subsection 1.2.2, we define the broken Sobolev spaces and discontinuous finite element spaces on the subdivision $\mathcal{T}_h$. We then, define jump and average of discontinuous functions. In the Subsection 1.2.3, we introduce some approximation properties of discontinuous finite element spaces, trace inequalities and inverse inequalities, etc., which are used in our subsequent analysis. In this subsection, we also derive an *hp*-approximation property which follows from an

interpolation of approximation properties derived in [1, 2]. We further modify the trace inequality of [58] so that it can be used in our *a priori* error estimates for the DG methods applied to nonlinear problems. In the Subsection 1.2.4, we have discussed the assumptions such as *local bounded variations* of the mesh $\mathcal{T}_h$, the degree of approximation of the finite element space V_h and *hp*-quasi uniform which are used in our subsequent Chapters.

1.2.1 Finite elements

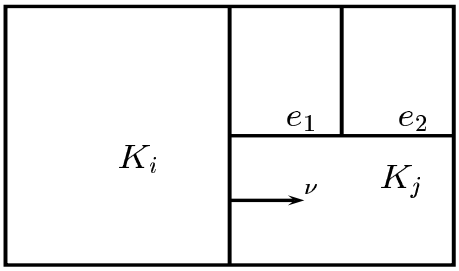
Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with boundary $\partial\Omega$. Let $\mathcal{T}_h = \{K_i : 1 \leq i \leq N_h\}$ be a shape regular finite element subdivision of Ω in the sense that there exists $\rho > 0$ such that if h_i is the diameter of K_i, then K_i contains a ball of radius ρh_i in its interior [20, p. 124]. Each element $K_i \in \mathcal{T}_h$ is either a triangle or a rectangle (possibly curvilinear) defined as follows. Let $\hat{K}$ be a shape regular master triangle or rectangle in $\mathbb{R}^2$, and let $\{F_i\}$ be a family of invertible maps such that F_i maps from $\hat{K}$ onto K_i, cf. Figure 1.1. Let

Figure 1.1: An example of construction of finite elements



h_i be the diameter of K_i and $h = \max\{h_i : 1 \leq i \leq N_h\}$. We denote the set of interior edges of $\mathcal{T}_h$ by $\Gamma_I = \{e_{ij} : e_{ij} = \partial K_i \cap \partial K_j,\ |e_{ij}| > 0\}$ and the set of boundary edges by $\Gamma_\partial = \{e_{i\partial} : e_{i\partial} = \partial K_i \cap \partial\Omega,\ |e_{i\partial}| > 0\}$, where $|\cdot|$ denotes the one-dimensional measure. Let $\Gamma = \Gamma_I \cup \Gamma_\partial$. Let Λ be the index set of elements of Γ. Since for each $e_k \in \Gamma_I,\ k \in \Lambda$, there exist $K_i,\ K_j \in \mathcal{T}_h$ such that $e_k = \partial K_i \cap \partial K_j (i > j)$, we associate with e_k a unit normal vector ν_k which is directed outward of K_i. For $e_k \in \Gamma_\partial$, let ν_k be the unit outward normal to the boundary $\partial\Omega$. For simplicity, we denote $\nu = \nu_k$. Note that our definition of e_k also admits hanging nodes along each side of the finite elements, cf. Figure 1.2.

Figure 1.2: Normal vector ν outward to K_i



1.2.2 Discontinuous spaces

In this subsection, we introduce the required broken Sobolev spaces with the associated norms and define the discontinuous finite element spaces which are used to in DG approximations. On the subdivision $\mathcal{T}_h$, we define the following broken Sobolev space of composite order $\mathbf{s} = \{s_i \geq 0 : 1 \leq i \leq N_h\}$ and exponent r, with $1 \leq r \leq \infty$:

$$W_r^{\mathbf{s}}(\Omega, \mathcal{T}_h) = \{v \in L^r(\Omega) : v|_{K_i} \in W_r^{s_i}(K_i), \text{ for all } K_i \in \mathcal{T}_h\},$$

where $W_r^{s_i}(K_i)$ is the standard Sobolev space of order s_i with exponent r, for each K_i. For $1 \leq r < \infty$, the associated broken norm and seminorm are defined, respectively, by

$$\|v\|_{W_r^{\mathbf{s}}(\Omega,\mathcal{T}_h)} = \left(\sum_{i=1}^{N_h} \|v\|^r_{W_r^{s_i}(K_i)}\right)^{1/r} \quad \text{and} \quad |v|_{W_r^{\mathbf{s}}(\Omega,\mathcal{T}_h)} = \left(\sum_{i=1}^{N_h} |v|^r_{W_r^{s_i}(K_i)}\right)^{1/r},$$

and for the case $r = \infty$, the associated broken norm and seminorm are defined, respectively, by

$$\|v\|_{W_\infty^{\mathbf{s}}(\Omega,\mathcal{T}_h)} = \max_{1\leq i\leq N_h} \|v\|_{W_\infty^{s_i}(K_i)} \quad \text{and} \quad |v|_{W_\infty^{\mathbf{s}}(\Omega,\mathcal{T}_h)} = \max_{1\leq i\leq N_h} |v|_{W_\infty^{s_i}(K_i)},$$

where $||v||_{W_r^{s_i}(K_i)}$ and $|v|_{W_r^{s_i}(K_i)}$ are the standard Sobolev norm and seminorm on K_i. When $r = 2$, we write $H^{\mathbf{s}}(\Omega, \mathcal{T}_h) = W_2^{\mathbf{s}}(\Omega, \mathcal{T}_h)$ and also write the norm and seminorm as

$$||v||_{\mathbf{s},h} = \|v\|_{W_r^{\mathbf{s}}(\Omega,\mathcal{T}_h)} \quad \text{and} \quad |v|_{\mathbf{s},h} = |v|_{W_r^{\mathbf{s}}(\Omega,\mathcal{T}_h)},$$

and when $s = s_i$ for all $1 \leq i \leq N_h$, we write $H^s(\Omega, \mathcal{T}_h)$,$\|v\|_{s,h}$ and $|v|_{s,h}$, respectively. For $s = 0$, we denote the norm by $\|.\|$ which is the standard L^2 norm. Denote the following

broken Sobolev spaces :

$$V = \{v \in L^2(\Omega) : v|_{K_i} \in H^1(K_i), \text{ for all } K_i \in T_h\}, \tag{1.1}$$

and

$$\mathbf{W} = \{\mathbf{w} \in (L^2(\Omega))^2 : \mathbf{w}|_{K_i} \in \left(H^1(K_i)\right)^2, \text{ for all } K_i \in T_h\}. \tag{1.2}$$

Discrete spaces: Let $\hat{P}_{p_i}(\hat{K})$ be the space of polynomials of total degree less than or equal to p_i on the triangle $\hat{K}$, and let $\hat{Q}_{p_i}(\hat{K})$ be the space of polynomials of degree less than or equal to p_i in each variable which are defined on the rectangle $\hat{K}$. Let $\mathrm{Z}_{p_i}(\hat{K})$ denote $\hat{P}_{p_i}(\hat{K})$ or $\hat{Q}_{p_i}(\hat{K})$ whenever $\hat{K}$ is a master triangle or a rectangle, respectively. Now, we set (see, [55], [39])

$$\mathrm{Z}_{p_i}(K_i) = \{v : v = \hat{v} \circ F_i^{-1}, \hat{v} \in \hat{\mathrm{Z}}_{p_i}(\hat{K})\}. \tag{1.3}$$

The discontinuous finite element spaces are defined as

$$V_h = \{v \in L^2(\Omega) : v|_{K_i} \in \mathrm{Z}_{p_i}(K_i)\}, \tag{1.4}$$

and

$$\mathbf{W}_h = \{\mathbf{w}_h \in (L^2(\Omega))^2 : \mathbf{w}_h|_{K_i} \in \mathrm{Z}_{p_i}(K_i)^2\}. \tag{1.5}$$

Let $p = \min\{p_i \geq 1 : 1 \leq i \leq N_h\}$.
We also need the following discontinuous finite element space of piecewise polynomials with uniform degree p:

$$V_h^* = \{v \in L^2(\Omega) : v|_{K_i} \in \mathrm{Z}_p(K_i), \ 1 \leq i \leq N_h\}. \tag{1.6}$$

We define the following Sobolev space with piecewise polynomial traces which is needed in our super-penalty results in Chapter 2.

$$H_p^s(\Omega) = \{v \in H^s(\Omega) : v|_{\partial\Omega} = w|_{\partial\Omega}, \quad \text{for some} \quad w \in V_h^*\}. \tag{1.7}$$

For $e_k \in \Gamma_I$, there are two elements K_i and K_j such that $e_k = \partial K_i \cap \partial K_j$. Hence, we define the 'degree' of polynomial in K_i and K_j restricted to e_k by p_k, by $p_k = (p_i + p_j)/2$. For

$e_k \in \Gamma_\partial$, we note that there is one element K_i with $e_k = \partial K_i \cap \partial\Omega$, and hence, we denote the degree of polynomial restricted to e_k by $p_k = p_i$.

Jump and Average of scalar function: We now define the jump and average of a function $v \in H^1(\Omega, \mathcal{T}_h)$ on an edge $e_k \in \Gamma$ as follows. If $e_k \in \Gamma_I$, that is $e_k = \partial K_i \cap \partial K_j$ $(i > j)$ for some i and j, then we set the jump and average as

$$[v] = v|_{K_i} - v|_{K_j}, \qquad \{v\} = \frac{v|_{K_i} + v|_{K_j}}{2}, \quad \text{respectively.} \tag{1.8}$$

In case $e_k \in \Gamma_\partial$, there exists K_i such that $e_k = \partial K_i \cap \partial\Omega$, and we then define, for notational convenience, the jump and average on e_k as

$$[v] = v|_{K_i \cap \partial\Omega}, \qquad \{v\} = v|_{K_i \cap \partial\Omega}, \quad \text{respectively.}$$

For the LDG method discussed in Chapter 3 and 5, we use the following jump: Let $e_k \in \Gamma_I$, that is $e_k = \partial K_i \cap \partial K_j$ for some i and j. Let ν_i and ν_j be the outward normals to the boundary ∂K_i and ∂K_j, respectively. On e_k, we now define the alternate jump of $v \in V$ as

$$[\![v]\!] = v|_{K_i}\nu_i + v|_{K_j}\nu_j.$$

In case $e_k \in \Gamma_\partial$, that is, there exists K_i such that $e_k = \partial K_i \cap \partial\Omega$, then we set, for notational convenience, the jump of $v \in V$ as

$$[\![v]\!] = v|_{K_i \cap \partial\Omega}\nu,$$

where ν is the outward normal to the boundary $\partial\Omega$.

Jump and Average of vector function: Let $e_k \in \Gamma_I$, that is $e_k = \partial K_i \cap \partial K_j$ for some i and j. Let ν_i and ν_j be the outward normals to the boundary ∂K_i and ∂K_j, respectively. On e_k, we now define the jump and average of $\mathbf{w} \in \mathbf{W}$ as

$$[\![\mathbf{w}]\!] = \mathbf{w}|_{K_i} \cdot \nu_i + \mathbf{w}|_{K_j} \cdot \nu_j, \qquad \{\mathbf{w}\} = \frac{\mathbf{w}|_{K_i} + \mathbf{w}|_{K_j}}{2}, \quad \text{respectively.}$$

In case $e_k \in \Gamma_\partial$, that is, there exists K_i such that $e_k = \partial K_i \cap \partial\Omega$, then we set for notational convenience, the jump and average of $\mathbf{w} \in \mathbf{W}$ as

$$[\![\mathbf{w}]\!] = \mathbf{w}|_{K_i \cap \partial\Omega} \cdot \nu, \qquad \{\mathbf{w}\} = \mathbf{w}|_{K_i \cap \partial\Omega}, \quad \text{respectively,}$$

where ν is the outward normal to the boundary $\partial\Omega$.

Broken energy norm: Let $v \in H^2(\Omega, \mathcal{T}_h)$. We define the following mesh-dependent norms which appear naturally in the analysis of interior penalty discontinuous Galerkin methods:

$$|||v|||^2 = \left(\sum_{i=1}^{N_h} \int_{K_i} |\nabla v|^2 \, \mathrm{dx} \; + \; \mathcal{J}^{\sigma,\beta}(v,\; v) \right), \tag{1.9}$$

and

$$|||v|||_+^2 = \left(\sum_{i=1}^{N_h} \int_{K_i} |\nabla v|^2 \, \mathrm{dx} \; + \; \sum_{e_k \in \Gamma} \frac{|e_k|^\beta}{p_k^2} \int_{e_k} \{\frac{\partial v}{\partial \nu}\}^2 \, \mathrm{ds} \; + \; \mathcal{J}^{\sigma,\beta}(v,\; v) \right), \tag{1.10}$$

where

$$\mathcal{J}^{\sigma,\beta}(v,\; w) = \left(\sum_{e_k \in \Gamma} \sigma_k \frac{p_k^2}{|e_k|^\beta} \int_{e_k} [v][w] \, \mathrm{ds} \right),$$

$\sigma|_{e_k} = \sigma_k$ and σ_k, β are positive real numbers.

1.2.3 Properties of finite element spaces

Approximation properties of finite element spaces. Below, we state a Lemma on some *hp*-approximation properties.

LEMMA 1.2.1 *For $\phi \in H^s(K_i)$, there exists a positive constant C_A (depending on s but independent of ϕ, p_i and h_i) and a sequence $\phi_{p_i}^{h_i} \in \mathrm{Z}_{p_i}(K_i)$, $p_i = 1, 2, ...$ such that :*
(i) for any $0 \le l \le s_i$,

$$\|\phi - \phi_{p_i}^{h_i}\|_{H^l(K_i)} \le C_A \frac{h_i^{\mu_i - l}}{{p_i}^{s_i - l}} \|\phi\|_{H^{s_i}(K_i)},$$

(ii) for $s_i > l + \frac{1}{2}$,

$$\|\phi - \phi_{p_i}^{h_i}\|_{H^l(e_k)} \le C_A \frac{h_i^{\mu_i - l - 1/2}}{p_i^{s_i - l - 1/2}} \|\phi\|_{H^{s_i}(K_i)},$$

(iii) for $0 \le l \le s_i - 1 + 2/r$,

$$\|\phi - \phi_{p_i}^{h_i}\|_{W_r^l(K_i)} \le C_A \frac{h_i^{\mu_i - l - 1 + 2/r}}{{p_i}^{s_i - l - 1 + 2/r}} \|\phi\|_{H^{s_i}(K_i)},$$

where $\mu_i = min(s_i,\ p_i + 1)$.

The proof of properties (i) and (ii) can be found in [9, Lemma 4.5]. Then using properties (1) and (3) in Lemma 1 of [1] and rescaling [2, Lemma 2], it is easy to derive the property (iii). However for completeness, we provide below a short proof of the property for $l = 0$. The proof follows by applying induction on $l \geq 1$.

Proof of property (iii) of Lemma 1.2.1 with $l = 0$. Let $\hat{K}$ be a shape regular reference triangle or rectangle. Let F_K be an invertible map which maps $\hat{K}$ onto a finite element K. Let its inverse be denoted by F_K^{-1}. Then for given $\phi \in H^s(K)$, define $\hat{\phi}(\hat{x}) = \phi \circ F(\hat{x})$. From [1, Lemma 1], we easily obtain the following approximation properties :

$$\|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^2(\hat{K})} \leq C \frac{1}{{p_i}^s} \|\hat{\phi}\|_{H^s(\hat{K})}, \tag{1.11}$$

and

$$\|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^\infty(\hat{K})} \leq C \frac{1}{{p_i}^{s-1}} \|\hat{\phi}\|_{H^s(\hat{K})}.$$

Let $2 < q < \infty$ and $r \geq 2$. Then, choose $\theta \in [0, 1]$ such that $r = 2\theta + (1 - \theta)q$. We easily see that

$$\theta = \frac{q - r}{q - 2} \quad \text{and} \quad (1 - \theta) = \frac{r - 2}{q - 2}.$$

Then, use Hölder's inequality to find that

$$\begin{aligned} \int_{\hat{K}} |v|^r \ dx &= \int_{\hat{K}} |v|^{2\theta + (1-\theta)q} \ dx \\ &= \int_{\hat{K}} |v|^{2\theta} \ |v|^{(1-\theta)q} \ dx \\ &\leq \left(\int_{\hat{K}} |v|^2 \ dx \right)^{\theta} \left(\int_{\hat{K}} |v|^q \ dx \right)^{(1-\theta)}. \end{aligned}$$

Thus, we obtain

$$\|v\|_{L^r(\hat{K})} \leq \|v\|_{L^2(\hat{K})}^{2\theta/r} \|v\|_{L^l(\hat{K})}^{q(1-\theta)/r}. \tag{1.12}$$

Note that

$$\|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^q(\hat{K})} \leq C(\text{measure } \hat{K})^{1/q} \|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^\infty(\hat{K})}. \tag{1.13}$$

Using $2\theta/r + q(1-\theta)/r = 1$ and the properties (1.11)-(1.12), we arrive at

$$\begin{aligned}
\|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^r(\hat{K})} &\leq C\|\hat{\phi} - \hat{\phi}_{p_i}\|^{2\theta/r}_{L^2(\hat{K})}\|\hat{\phi} - \hat{\phi}_{p_i}\|^{q(1-\theta)/r}_{L^q(\hat{K})} \\
&\leq C p_i^{-s2\theta/r} p_i^{-(s-1)q(1-\theta)/r}\|\hat{\phi}\|_{H^s(\hat{K})} \\
&\leq C p_i^{-s} p_i^{(r-2)/r(1-2/q)}\|\hat{\phi}\|_{H^s(\hat{K})}.
\end{aligned}$$

Since q is arbitrary and the constant C is independent of q, we obtain the estimate

$$\|\hat{\phi} - \hat{\phi}_{p_i}\|_{L^r(\hat{K})} \leq C p_i^{-s+1-2/r}\|\hat{\phi}\|_{H^s(\hat{K})}.$$

Now, scaling back to K implies the required result and this completes the rest of the proof. ■

For given $\phi \in H^{\mathbf{s}}(\Omega, \mathcal{T}_h)$, we define $I_h\phi \in V_h$ by $(I_h\phi)|_{K_i} = \phi^{h_i}_{p_i}(\phi|_{K_i}), \quad \forall\ 1 \leq i \leq N_h$. By virtue of Lemma 1.2.1, $I_h\phi$ satisfies the local approximation properties derived in Lemma 1.2.1, see [45, p. 737].

Trace inequalities. We need the following trace inequalities for our future use.

LEMMA 1.2.2 *Let $\phi \in H^{j+1}(K_i), \quad K_i \in \mathcal{T}_h$. Then, there exists a constant $C_{T_1} > 0$ such that*

$$\|\phi\|^r_{W^j_r(e_k)} \leq C_{T_1}\left(\frac{1}{h_i}\|\phi\|^r_{W^j_r(K_i)} + \|\phi\|^{r-1}_{W^j_{2r-2}(K_i)}\|\nabla^{(j+1)}\phi\|_{L^2(K_i)}\right), \tag{1.14}$$

where j=0, 1 and r=2, 4.

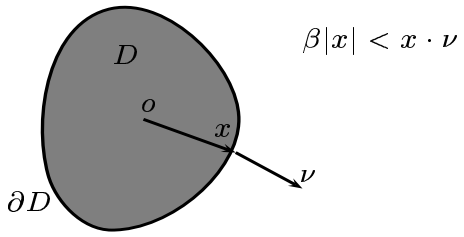
Proof. For $r = 2$, the inequality (1.14) is proved in [58, Appendix A.2]. We now extend the proof of the inequality to the case $r = 4$, $l = 0$. The proof for $l = 1$ follows by induction. Let D be a star-shaped domain [58] with piecewise smooth boundary ∂D, cf., Figure 1.3. Without loss of generality assume that $o \in D$ be the origin. Denote the outward normal to ∂D by ν and let $\mathbf{x}$ be the point vector on ∂D. From the definition of star-shaped domain, there is a $\beta > 0$ such that

$$\beta|\mathbf{x}| < \mathbf{x} \cdot \nu$$

Let $\phi \in H^2(D)$. Then, we apply the Green's theorem [48] to the vector field $\phi^4\mathbf{x}$ to obtain

$$\int_{\partial D} \phi^4 \mathbf{x} \cdot \nu \; ds = \int_D \nabla \cdot (\phi^4 \mathbf{x}) \; dx \tag{1.15}$$

Figure 1.3: Star-shaped domain D



Using the property of star-shaped domain, we first show that the term on the left hand side of (1.15) is bounded below

$$\int_{\partial D} \phi^4 \mathbf{x} \cdot \nu \; ds \geq \beta \inf_{x \in \partial D} |\mathbf{x}| \int_{\partial D} \phi^4 \; ds = \beta \inf_{x \in \partial D} |\mathbf{x}| \; \|\phi\|^4_{L^4(\partial D)}. \tag{1.16}$$

The term on the right hand side of (1.15) is shown bounded above as

$$\begin{aligned} \int_D \nabla \cdot (\phi^4 \mathbf{x}) \; dx &= \int_D \phi^4 \; \nabla \cdot \mathbf{x} \; dx + \int_D 4\phi^3 \nabla\phi \cdot \mathbf{x} \; dx \\ &= \int_D 2\phi^4 \; dx + \int_D 4\phi^3 \nabla\phi \cdot \mathbf{x} \; dx, \end{aligned}$$

and therefore,

$$\begin{aligned} \left| \int_D \nabla \cdot (\phi^4 \mathbf{x}) \; dx \right| &\leq 2\|\phi\|^4_{L^4(D)} + 4 \sup_{\mathbf{x} \in D} |\mathbf{x}| \int_D |\phi|^3 |\nabla\phi| \\ &\leq 2\|\phi\|^4_{L^4(D)} + 4 \sup_{\mathbf{x} \in D} |\mathbf{x}| \; \|\phi\|^3_{L^6(D)} \|\nabla\phi\|_{L^2(D)}. \end{aligned} \tag{1.17}$$

We now combine (1.16)-(1.17) to obtain

$$\|\phi\|^4_{L^4(\partial D)} \leq \frac{2}{\beta \inf_{\mathbf{x} \in \partial D} |\mathbf{x}|} \left(\|\phi\|^4_{L^4(D)} + 2 \sup_{\mathbf{x} \in D} |\mathbf{x}| \; \|\phi\|^3_{L^6(D)} \|\nabla\phi\|_{L^2(D)} \right). \tag{1.18}$$

To complete the proof, let o be the center of the inscribed circle in K_i with radius ρh_i. Then

$$\sup_{\mathbf{x} \in D} |\mathbf{x}| \leq h_i \quad \text{and} \quad \inf_{\mathbf{x} \in \partial D} |\mathbf{x}| \geq \rho h_i.$$

We obtain from (1.18)

$$\|\phi\|^4_{L^4(\partial K_i)} \leq C \frac{1}{h_i} \left(\|\phi\|^4_{L^4(D)} + h_i \|\phi\|^3_{L^6(D)} \|\nabla\phi\|_{L^2(D)} \right),$$

where C is a positive constant which depends on ρ but independent of h_i. This completes the rest of the proof. ∎

We recall the following trace inequality on finite element spaces for our future use. For a proof, we refer to [61, Lemma 2.1].

LEMMA 1.2.3 *Let $v_h \in \mathrm{Z}_{p_i}(K_i)$. Then, there exists a constant $C_{T_2} > 0$ such that*

$$\|\nabla^l v_h\|_{L^2(e_k)} \leq C_{T_2} p_i h_i^{-1/2} \|\nabla^l v_h\|_{L^2(K_i)}, \quad l = 0, 1. \tag{1.19}$$

Inverse inequalities. Below, we state without proof a Lemma on inverse inequalities. For a proof, we refer to [49, p. 6], [12, Theorem 6.1].

LEMMA 1.2.4 *Let $v_h \in \mathrm{Z}_{p_i}(K_i)$. Then, for $r \geq 2$, there exists a constant $C_I > 0$ such that*

$$\|v_h\|_{L^r(K_i)} \leq C_I p_i^{1-2/r} h_i^{(2/r-1)} \|v_h\|_{L^2(K_i)}, \tag{1.20}$$

$$|v_h|_{H^l(K_i)} \leq C_I p_i^2 h_i^{-1} |v_h|_{H^{l-1}(K_i)}, \quad l \geq 1 \tag{1.21}$$

and

$$\|v_h\|_{L^r(e_k)} \leq C_I p_i^{1-2/r} |e_k|^{(1/r-1/2)} \|v_h\|_{L^2(e_k)}, \tag{1.22}$$

where $e_k \subset \partial K_i$ is an edge.

LEMMA 1.2.5 *(L^2-Projection Π): Let $\boldsymbol{\psi} \in H^s(K_i)^2$ and $\boldsymbol{\psi}_h = \Pi\boldsymbol{\psi} \in \mathrm{Z}_{p_i}(K_i)^2$ be the L^2 projection of $\boldsymbol{\psi}$ onto $\mathrm{Z}_{p_i}(K_i)$. Then, the following approximation properties hold :*

$$\|\boldsymbol{\psi} - \boldsymbol{\psi}_h\|_{L^2(K_i)^2} + \frac{h_i^{1/2}}{p_i} \|\boldsymbol{\psi} - \boldsymbol{\psi}_h\|_{L^2(\partial K_i)^2} \leq C \frac{h_i^{\mu}}{p_i^s} \|\boldsymbol{\psi}\|_{H^s(K_i)^2},$$

and

$$\|\boldsymbol{\psi} - \boldsymbol{\psi}_h\|_{L^4(K_i)^2} \leq C \frac{h_i^{\mu-1/2}}{p_i^{s-1/2}} \|\boldsymbol{\psi}\|_{H^s(K_i)^2},$$

where $\mu = min\{s,\ p_i + 1\}$.

Proof. First inequality of the lemma follows from the Lemma 1.2.1 and the trace inequality (1.19). For the estimate of $\|\boldsymbol{\psi} - \boldsymbol{\psi}_h\|_{L^4(K_i)^2}$, we use inverse inequality (1.20). This completes the rest of the proof. ∎

For our future use, we state the following Poincaré type inequalities on $H^1(\Omega, \mathcal{T}_h)$. For a proof, we refer to [49, Theorem 3.7]; see also [13] for the case of $r = 2$.

LEMMA 1.2.6 **(Poincaré type inequalities)**. *Let $v \in H^1(\Omega, \mathcal{T}_h)$. Then, there exists a constant $C_P > 0$ independent of h and v such that, for $1 \leq r < \infty$,*

$$\|v\|_{L^r(\Omega)} \leq C_P |||v|||.$$

1.2.4 Assumptions on the mesh and degree of approximation

Assumption (P):

1. The finite element subdivision $\mathcal{T}_h$ satisfies the *bounded local variation* condition in the sense that if $|\partial K_i \cap \partial K_j| > 0$, for any K_i and $K_j \in \mathcal{T}_h$, then there exists a constant κ independent of $h_i,\ h_j$ such that
$$\frac{h_i}{h_j} \leq \kappa.$$
In particular, this implies that for any element K_i the number of neighboring elements $K_j \in \mathcal{T}_h$ with $|\partial K_i \cap \partial K_j| > 0$ is bounded by N_κ uniformly, for some positive integer N_κ.

2. The discontinuous finite element space V_h satisfies the following *bounded local variation* : If $|\partial K_i \cap \partial K_j| > 0$, for any K_i and $K_j \in \mathcal{T}_h$, then there exists a constant ϱ independent of p_i and p_j such that
$$\frac{p_i}{p_j} \leq \varrho\ .$$
Here, $|.|$ denotes the one dimensional Euclidean measure.

We now present some examples which satisfy the assumption $\mathbf{P}(1)$.
(i) **Regular subdivision:** A subdivision of Ω into shape regular elements $K_i,\ 1 \leq i \leq N_h$, is such that for any two elements K_i and K_j, the common portion $\partial K_i \cap \partial K_j$ is either empty or a vertex of K_i or an entire edge e of K_i, that is, $e = \partial K_i \cap \partial K_j$ and there is no other element $K_l \in \mathcal{T}_h$, $(l \neq j,\ i)$ such that $|e \cap \partial K_l| > 0$, [20, p. 132].
(ii) **1-irregular subdivision:** A shape regular subdivision $\{K_i\}_{i=1}^{N_h}$ of Ω is such that for any side of an element K_i, there can be at most one hanging node (cf. Figure 1.2), see [44] and [45, p. 5].

From the assumption (**P**), it is easy to see that if $e_k \subset \partial K_i$ then there exist constants $c_1(\kappa)$, $c_2(\kappa)$, $c_3(\varrho)$ and $c_4(\varrho)$ which are independent of h and p such that

$$c_1(\kappa)h_i \leq |e_k| \leq c_2(\kappa)h_i, \qquad c_3(\varrho)p_i \leq p_k \leq c_4(\varrho)p_i. \tag{1.23}$$

Assumption (Q) (hp-quasiuniformity):
Along with the assumption (**P**), we also assume that the subdivision $\mathcal{T}_h$ and discontinuous space V_h satisfy the following hp quasi-uniformity condition

$$\left(\max_{1\leq i\leq N_h} \frac{h_i}{p_i}\right) \leq C_Q \left(\min_{1\leq i\leq N_h} \frac{h_i}{p_i}\right), \tag{1.24}$$

where C_Q is a positive constant which is independent of h and p.
Observe that under the assumption (1.24), the following holds

$$\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h} \frac{h_i}{p_i}\right) = \left(\min_{1\leq i\leq N_h} \frac{h_i}{p_i}\right)^{-1}\left(\max_{1\leq i\leq N_h} \frac{h_i}{p_i}\right) \leq C_Q. \tag{1.25}$$

1.2.5 Some results from functional analysis

We need some well known results from the functional analysis, which we state without proof in this subsection.

THEOREM 1.2.1 **(Brouwer Fixed Point Theorem, [48, p. 218])** *Let V be a finite dimensional Hilbert space and S be a continuous map from a nonempty, compact and convex subset K of V which maps into K. Then, there is a $v \in K$ such that $S(v) = v$.*

LEMMA 1.2.7 **(Hölder's inequality, [48])** *Let $1 \leq p, q < \infty$ be such that $1/p + 1/q = 1$ and $D \subset \mathbb{R}^n$. Suppose that $\phi \in L^p(D)$ and $\psi \in L^q(D)$. Then*

$$\left|\int_D \phi\psi \, dx\right| \leq \left(\int_D |\phi|^p \, dx\right)^{1/p}\left(\int_D |\psi|^q \, dx\right)^{1/q}.$$

LEMMA 1.2.8 **(Generalized Hölder's inequality, [48])** *Let $1 \leq p,\ q,\ r < \infty$ be such that $1/p+1/q+1/r = 1$ and $D \subset \mathbb{R}^n$. Suppose that $\phi \in L^p(D)$, $\psi \in L^q(D)$ and $\chi \in L^r(D)$. Then*

$$\left|\int_D \phi\psi\chi \, dx\right| \leq \left(\int_D |\phi|^p \, dx\right)^{1/p}\left(\int_D |\psi|^q \, dx\right)^{1/q}\left(\int_D |\chi|^r \, dx\right)^{1/r}.$$

LEMMA 1.2.9 **(Cauchy-Schwarz inequality, [58])** *Let $1 \leq p, q < \infty$ be such that $1/p + 1/q = 1$. Suppose that $\{a_i\}$ and $\{b_i\}$ are two sequences of N positive real numbers. Then*

$$\left(\sum_{i=1}^{N} a_i b_i\right) \leq \left(\sum_{i=1}^{N} a_i^p \; dx\right)^{1/p} \left(\sum_{i=1}^{N} b_i^q \; dx\right)^{1/q}.$$

LEMMA 1.2.10 **(Generalized discrete Cauchy-Schwarz inequality, [58])** *Let $1 \leq p, q, r < \infty$ be such that $1/p + 1/q + 1/r = 1$. Suppose that $\{a_i\}$, $\{b_i\}$ and $\{c_i\}$ are three sequences of N positive real numbers. Then*

$$\left(\sum_{i=1}^{N} a_i b_i c_i\right) \leq \left(\sum_{i=1}^{N} a_i^p \; dx\right)^{1/p} \left(\sum_{i=1}^{N} b_i^q \; dx\right)^{1/q} \left(\sum_{i=1}^{N} c_i^r \; dx\right)^{1/r}.$$

1.3 Literature review

We refer the reader to the review article by C. Cockburn, G. E. Karniadakis and C. W. Shu, [29] for various motivations in developing the discontinuous Galerkin (DG) methods over the past 30 years. However, we provide here the results from some of the articles which play crucial role in developing new DG methods.

In 1973, Babuska [6] introduced the penalty method for incorporating the Dirichlet boundary condition weakly. Therein the following variational form, given $f \in L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$, find $u \in H^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \; dx + \int_{\partial\Omega} \sigma(u - g)v \; ds = \int_\Omega fv \; dx \tag{1.26}$$

is used to approximate the linear elliptic problem

$$-\triangle u = f \text{ in } \Omega, \tag{1.27}$$

$$u = g \text{ on } \partial\Omega. \tag{1.28}$$

In [6], for finite element space of degree p, then error estimate of order $h^{(2p+1)/3}$ is derived provided the penalty parameter σ is of order $h^{-(2p+1)/3}$. The lack of optimality in order

of convergence is due to the inconsistent formulation (1.26). Note that the solution u of (1.27)-(1.28) does not satisfy (1.26) and instead it satisfies

$$\int_\Omega \nabla u \cdot \nabla v \; dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \; ds + \int_{\partial\Omega} \sigma(u-g)v \; ds = \int_\Omega fv \; dx. \tag{1.29}$$

In 1971, Nitsche [54] introduced independently another penalty method to approximate (1.27)-(1.28) which is based on the following variational form

$$\int_\Omega \nabla u \cdot \nabla v \; dx - \int_{\partial\Omega} \frac{\partial u}{\partial n} v \; ds - \int_{\partial\Omega} \frac{\partial v}{\partial n} u \; ds + \int_{\partial\Omega} \sigma uv \; ds = \int_\Omega fv \; dx - \int_{\partial\Omega} \frac{\partial v}{\partial n} g \; ds + \int_{\partial\Omega} \sigma gv \; ds. \tag{1.30}$$

Note that the first two term on the left-hand side of (1.30) arise from integration by parts and, therefore, the formulation is consistent. The third term on the left-hand side of (1.30) is added to symmetrize the variational formulation and ensures the adjoint consistency. The fourth term on the left-hand side of (1.30) is to pose the Dirichlet boundary condition weakly. Nitsche [54] derived optimal order of convergence in both H^1-norm and L^2-norm provided the penalty parameter σ is of the form $\sigma = \eta/h$, where h mesh size and η is a sufficiently large constant.

Based on Nitsche's formulation, an Interior Penalty (IP) method is introduced by Wheeler in [66]. The IP methods arose from the observation that just as Dirichlet boundary condition, the inter element continuity could be attained weakly by adding a penalty term to the variational formulation. To be more precise, the weak form of IP method for (1.27)-(1.28) reads as : find $u \in H^2(\Omega, \mathcal{T}_h)$ such that

$$\sum_{i=1}^{N_h} \int_{K_i} \nabla u \cdot \nabla v \; dx - \sum_{e_k \in \Gamma} \int_{e_k} \{\frac{\partial u}{\partial n}\}[v] \; ds - \sum_{e_k \in \Gamma} \int_{e_k} \{\frac{\partial v}{\partial n}\}[u] \; ds + \sum_{e_k \in \Gamma} \int_{e_k} \sigma [u][v] \; ds = \int_\Omega fv \; dx - \int_{\partial\Omega} \frac{\partial v}{\partial n} g \; ds + \int_{\partial\Omega} \sigma gv \; ds.$$

The IP method of Wheeler [66] is presently called as SIPG (IP) method which is applied to a general linear elliptic problem. The variational form, therein, satisfies a Gärding type inequality on the finite element space provided the stabilizing penalty parameter σ is bounded below by σ_0 which depends on the coefficients of the problem and the constants involved in inverse and trace inequalities. Since this method is consistent as well as adjoint

consistent, optimal order of convergence are derived in the broken H^1-norm as well as in L^2-norm. Arnold [4] applied the SIPG method to a second order quasilinear parabolic problems and derived optimal error estimates. Where as, Douglas and Dupont [38] used the jump of normal derivative of the continuous approximation to obtain an approximate solution in between C^0 and C^1 for second order elliptic and parabolic problems. This method is known in the literature as C^0-interior penalty method. The results of [38] was further generalized to include the incompressible miscible displacement problems in reservoir studies Wheeler and Darlow [67], Pani and Das [56]. It is observed that using C^0-interior penalty method, a right amount of numerical diffusion is added to the numerical scheme to the concentration equation which is convection dominated diffusion problem, see Ali [3].

In the conforming FE approximation of the fourth order elliptic problems, the FE space has to be a subspace of C^1 which requires a higher degree polynomials and makes the scheme very expensive and complicated. The IP methods helped in approximating the fourth order problems using C^0 or discontinuous finite element spaces. In particular, Babuska and Zlámal [7] used the jump of normal derivative of continuous approximation as a penalty term to obtain an approximate solution in between C^0 and C^1. But the formulation, therein, is inconsistent, and hence, it is difficult to derive optimal error estimates. Subsequently using discontinuous FE spaces, an IP method is proposed by Baker in [10] and optimal error estimates are derived. Recently in Brenner and Sung [14], the idea of [7] has been used in approximating the fourth order problems using continuous FE spaces. While their formulation is consistent, the connections between C^0 finite elements and its C^1 relatives are explored. Moreover using a post precessing, they obtain C^1 approximate solution from the C^0 approximate solution.

Since it is observed in the theory and the experiments that the SIPG (IP) method is sensitive in choosing the stabilizing parameter, a variant of IP methods which is based on nonsymmetric formulation is proposed for diffusion problems in [55] by Oden, Babuska and Baumann. *A priori* error estimates, therein, which are nearly optimal in h (mesh size) and slightly suboptimal in p (degree of approximation) are derived in the broken H^1-norm. Rivière *et al.* [61] have analyzed the DG method of Oden *et al.* [55] for the self-adjoint elliptic problems. Using a new interpolant, the optimal estimate in h and suboptimal in p

is derived. A constrained DG method based on the Oden *et al.* nonsymmetric formulation has been proposed and analyzed in [61]. In the constrained DG method, the discontinuous discrete space satisfies a constraint that the jump of any element of the space is weakly zero on the edges. This is instrumental in proving the optimal order of convergence in L^2 norm though the formulation is not adjoint-consistent. Rivière *et al.* [61] and Houston *et al.* [44] have introduced and analyzed the NIPG method which is a stabilized DG method of Oden *et al.* [55] for linear elliptic problems. Therein *a priori* error estimates which are optimal in h and slightly suboptimal in p are derived. Since the NIPG formulation is not adjoint-consistent, it is difficult to prove optimal order of convergence in the L^2 norm. However, it is observed that the term which causes the loss of adjoint-consistency can be made small by imposing a super-penalty on regular mesh. This has made it possible in [61] to prove optimal order of convergence in L^2 norm on the regular mesh. For a review on the stability and *a priori* error estimates of SIPG and NIPG methods for linear elliptic problems, we refer to [58]. The NIPG method is applied using discontinuous FE spaces for fourth order elliptic problems [53] and *a priori* error estimates are derived. In [35], NIPG and SIPG methods are applied and analyzed for fourth order semilinear parabolic problems.

There is an extensive work on the DG methods for diffusion and convection-diffusion problems [28, 44]. Subsequently, SIPG and NIPG methods are applied to the linear elliptic systems such as Stokes [65, 62] and elasticity [43] problems. Using the extra degrees of freedom in DGFEM, it is shown in [43] that the proposed DG method for elastic problems is *locking free* which is not the case in the standard finite element methods. The applications of DG methods for Navier-Stokes equations can be found in [47].

The first attempt to study *hp*-DG methods for quasilinear elliptic problems has been made in [45]. The authors of [45] have applied a one parameter family of discontinuous Galerkin methods for a class of monotone quasilinear elliptic problems and have derived *a priori* error estimates in broken H^1-norm which are optimal in h and slightly suboptimal in p. They have also provided numerical experiments to illustrate the theoretical results. It is difficult to extend the analysis of [45] to a class of quasilinear or strongly nonlinear elliptic problems of non-monotone type, therefore, in the present study, we discuss and

analyze the SIPG and NIPG methods for such problems.

Apart from the work on DG methods for elliptic and parabolic problems, there is a substantial work on DG methods for the hyperbolic problems [64], [16] and second order partial differential equations with nonnegative characteristic form [63].

A parallel work on the LDG method has been pursued by several authors working on hyperbolic equations. The LDG method is proposed in [18] for linear elliptic problems in the mixed form in which the flux and displacement variables are approximated at the same time. This method is also adjoint consistent which is vital in deriving the optimal L^2 error estimate. In [18], the authors have discussed stability of the h-version LDG method applied to the Laplace equation and have derived *a priori* error estimates which are optimal. The work on hp-version of LDG method for the diffusion problems can be found in [57], where *a priori* error estimates which are optimal in h and slightly suboptimal in p are derived. Subsequently, there is an extensive work on the LDG methods for diffusion and convection-diffusion problems [28]. The LDG method is then applied to the Stokes systems [30]. Recently, the LDG method is extended to incompressible elastic materials [31] and related error estimates are discussed.

The first attempt on the h-version of the LDG method for the monotone quasilinear elliptic problems has been made in [17]. In [17], the authors have used the strongly monotone property of the associated bilinear form in a crucial way in their analysis and have derived optimal error estimates in h. Further, they have discussed some numerical results. It is difficult to extend the work of [17] to the quasilinear or strongly nonlinear elliptic problems of non-monotone type. Therefore, attempt has been made in this direction to analyze the hp-LDG method for such problems.

1.4 Organization of the thesis

While Chapter 1 deals with motivation, preliminaries and literature survey, Chapter 2 focuses on the SIPG and NIPG methods for a class of quasilinear elliptic problems of non-monotone type which lead to a system of nonlinear algebraic equations. For existence of a solution to this nonlinear system, we first reformulate the nonlinear system in a fixed

point form and this is then related to the solutions of the associated linearized problem which is a linear non-selfadjoint elliptic problem. Therefore, in the beginning of Chapter 2, we study both SIPG and NIPG methods for a general second order linear non-selfadjoint elliptic problems. The corresponding bilinear form satisfies a Gärding type inequality and, therefore, we need a bound for the error in L^2-norm. This estimate is derived using a discrete dual problem and, hence, *a priori* error estimates are derived in broken H^1-norm which are optimal in h and suboptimal in p. Then, using Brouwer fixed point theorem, we have shown that there is a fixed point which is a solution of the nonlinear system. Further using the Lipschitz continuity of the discrete solution map, we have proved that the solution is unique. *A priori* error estimates which are optimal in h and suboptimal in p are obtained as a by product of our fixed point arguments. In this Chapter, we also apply the super-penalty arguments for the nonlinear problem to prove optimal order of convergence in the L^2-norm when NIPG method is used. Finally, some numerical experiments are conducted to illustrate the theoretical results.

In Chapter 3, we analyze the LDG method for a class of quasilinear elliptic problems of non-monotone type. The key idea of proving the *a priori* error estimates is to rewrite the nonlinear system in the fixed point form. We define a fixed point map which maps the discrete functions to the solutions of a linear non-self adjoint elliptic problem. We then show that this fixed point map maps a ball into itself. Since, this map is continuous in the ball, an appeal to Brouwer fixed point theorem yields a solution to the nonlinear system. Moreover, as a by product of Lipschitz continuity of the solution map, uniqueness of the discrete solution is proved. *A priori* error estimates are immediate from the estimates of the fixed point map. These error estimates are optimal in h and slightly suboptimal in p which lead to precisely same optimal order of convergence in h and slightly suboptimal in p for the linear elliptic problems . Finally, we present some numerical experiments to illustrate the theoretical results.

Chapter 4 is devoted to the discontinuous Galerkin methods, in particular, the SIPG and NIPG methods for a class of strongly nonlinear elliptic problems. Under the assumption that the nonlinear operator is elliptic, we analyze these two DG methods. We consider the most general form as a model problem which covers many practical cases such as mean

curvature flow, subsonic flow and Bratu's problems, etc. The induced bilinear form includes also the cases linear elliptic [61] and quasilinear elliptic [41] problems. Therefore, the results of this Chapter can be thought of as an extension of the results presented in Chapter 2. *A priori* error estimates are derived in broken H^1-norm which are optimal in h and nearly suboptimal in p, when the degree of approximation is greater than or equal to 2.

In Chapter 5, we extend the results of Chapter 3 to a general strongly nonlinear elliptic problem. By adding a stabilizing term containing the jumps of the flux variable, we prove *a priori* error estimates, when the degree of approximation is greater than or equal to 1. Dropping of these terms needs the degree of approximation greater than or equal to 2. In both the cases, *a priori* estimates are optimal in h and slightly suboptimal in p. These results are precisely the same optimal order of convergence in h and slightly suboptimal in p, when the method is applied to linear [18] or quasilinear [17] elliptic problems.

Finally, we present, in Chapter 6, some critical assessments of our results. Further, we discuss the possible extensions and also the scope for future problems.

Chapter 2

DG Methods for Quasilinear Elliptic Problems of Non-monotone Type

2.1 Introduction

In literature, optimal *a priori* error estimates are derived in broken H^1-norm for SIPG and NIPG methods applied to linear self-adjoint elliptic problems, see [5], [61]. Except for [45], there are hardly any results on the discontinuous Galerkin (DG) approximation to the nonlinear elliptic problems. In [45], a one parameter family of DG methods is applied to the quasilinear elliptic problems which are strongly monotone and Lipschitz continuous. In particular, the authors have considered a class of elliptic problems of the form

$$-\nabla \cdot (\mu(x,\ |\nabla u|)\nabla u) = f(x)$$

subject to mixed Dirichlet-Neumann boundary conditions. Under the structural conditions on $\mu \in C(\bar{\Omega} \times [0, \infty))$:

$$m_\mu(t-s) \leq \mu(x,t)t - \mu(x,s)s \leq M_\mu(t-s) \quad \text{for } \ t \geq s \geq 0 \tag{2.1}$$

and for some positive constants m_μ and M_μ, it is shown that the DG formulation is monotone and hence, *a priori* error estimates in broken H^1-norm are derived. For nonlinear

problems of the following type

$$-\nabla \cdot (a(u)\nabla u) = f \quad \text{in } \Omega, \tag{2.2}$$

$$u = g \quad \text{on } \partial\Omega, \tag{2.3}$$

where $0 < \alpha \leq a(u) \leq M$, for some positive constants α and $M \in \mathbb{R}^+$, the nonlinearity may not satisfy (2.1) and hence, it is difficult to extend the analysis of [45]. Therefore, an attempt has been made in this Chapter to study DG methods for the problem (2.2)-(2.3). The results presented in this chapter can be thought of an extension to discontinuous Galerkin methods of the results established for the nonlinear Dirichlet problem (2.2)-(2.3) by using a Galerkin method in [37]. Both SIPG and NIPG methods are discussed for the problem (2.2)-(2.3) and *a priori* error estimates are derived in the broken H^1-norm which are optimal in h. These results lead precisely the same h-optimal and p-suboptimal rates of convergence in the broken H^1-norm as in the case of linear elliptic problems, when it is approximated by a NIPG method, see [61, Theorem 3.1].

The organization of this Chapter is as follows. In Section 2.2, DG methods are applied to linear non-selfadjoint elliptic problems and *a priori* error estimates are derived in the broken H^1-norm, which are optimal in h and suboptimal in p. Section 2.3 is devoted to SIPG and NIPG methods for the quasilinear elliptic problems (2.2)-(2.3). Using Brouwer's fixed point theorem, existence of a discrete solution is proved. Then *a priori* error estimates are derived in the broken H^1-norm, which are optimal in h and suboptimal in p. Further, an *a priori* error estimate in the L^2-norm is established on regular meshes for (2.2)-(2.3) with piecewise polynomial or zero Dirichlet boundary datum. In Section 2.4, we provide some numerical experiments to illustrate the theoretical results obtained in this chapter.

2.2 Non-selfadjoint Linear Elliptic Problems

For our error analysis of DG methods applied to the nonlinear elliptic problem (2.2)-(2.3), we need some results on the corresponding linearized problems. Since the linearized problem is a non-selfadjoint elliptic problem, in this section, we consider the following second-order

linear non-selfadjoint elliptic partial differential equation:

$$\begin{aligned} -\nabla \cdot (a(x)\nabla u) + \vec{b}(x) \cdot \nabla u + a_0(x)u &= f(x) \quad \text{in } \Omega, \\ u &= g \quad \text{on } \partial\Omega. \end{aligned} \tag{2.1}$$

We adopt the following assumptions on the problem (2.1).

Assumptions (R):

1. There exists $\alpha > 0$ such that $0 < \alpha \leq a(x)$ and $a_0(x) \geq 0,\ \forall x \in \bar{\Omega}$.
2. $a \in W^1_\infty(\Omega)$ and $b,\ a_0 \in L^\infty(\Omega)$ with $M = \max\{\|a\|_{L^\infty(\Omega)},\ \|b\|_{L^\infty(\Omega)},\ \|a_0\|_{L^\infty(\Omega)}\}$.
3. $f \in L^2(\Omega)$ and $g \in H^{3/2}(\partial\Omega)$.

Then, from [40, Lemma 9.17] it is well known that there exists a unique solution $u \in H^2(\Omega)$ to the problem (2.1) satisfying

$$\|u\|_{H^2(\Omega)} \leq C \left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{3/2}(\partial\Omega)} \right). \tag{2.2}$$

2.2.1 Weak formulation

For $w,\ v \in H^2(\Omega, \mathcal{T}_h)$, we consider the following bilinear form

$$\begin{aligned} B(w,\ v) = \sum_{i=1}^{N_h} \int_{K_i} (a\nabla w \cdot \nabla v + a_0 wv + (\vec{b} \cdot \nabla w)v)\ dx - \sum_{e_k \in \Gamma} \int_{e_k} \{a\frac{\partial w}{\partial \nu}\}[v]\ ds \\ -\theta \sum_{e_k \in \Gamma} \int_{e_k} \{a\frac{\partial v}{\partial \nu}\}[w]\ ds + \mathcal{J}^{\sigma,\beta}(w,\ v) + \sum_{e_k \in \Gamma} \int_{e_k} \{\vec{b} \cdot \nu v\}[w]\ ds, \end{aligned}$$

and the linear form

$$L(v) = \int_\Omega fv\ dx - \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} (a\frac{\partial v}{\partial \nu})g\ ds + \sum_{e_k \in \Gamma_\partial} \int_{e_k} \sigma_k \frac{p_k^2}{|e_k|^\beta} vg\ ds + \sum_{e_k \in \Gamma_\partial} \int_{e_k} \vec{b} \cdot \nu v g\ ds.$$

where $\theta = \pm 1$. When $\vec{b} = 0$, we note that $\theta = +1$ corresponds to a symmetric and $\theta = -1$ to a non-symmetric interior penalty method.

We define a weak formulation which is suitable for the discontinuous Galerkin methods as follows: Find $u \in H^2(\Omega, \mathcal{T}_h)$ such that

$$B(u,\ v) = L(v) \quad \forall v \in H^2(\Omega, \mathcal{T}_h). \tag{2.3}$$

Let $V_h = \{v \in L^2(\Omega) : v|_{K_i} \in \mathrm{Z}_{p_i}(K_i)\}$, where $\mathrm{Z}_{p_i}(K_i) = \{v : v = \hat{v} \circ F_i^{-1}, \hat{v} \in \hat{\mathrm{Z}}_{p_i}(\hat{K})\}$. Now the discontinuous Galerkin approximation of u is to seek $u_h \in V_h$ such that

$$B(u_h,\ v_h) = L(v_h) \quad \forall v_h \in V_h. \tag{2.4}$$

Below, we examine the consistency of the above scheme (2.4).

THEOREM 2.2.1 *If the solution u of the problem (2.1) is in $H^2(\Omega)$, then u satisfies the problem (2.3). Conversely, if the solution u of the problem (2.3) is in $H^1(\Omega) \cap H^2(\Omega, \mathcal{T}_h)$, then u satisfies the problem (2.1) weakly.*

The proof techniques of Rivière *et al.* [61, Lemma 2.2] or [58, Theorem 3.1] can be easily modified to prove Theorem 2.2.1 and hence, the proof is omitted. The solvability of (2.4) will be discussed at the end of Section 3. From the equations (2.3)-(2.4) and Theorem 2.2.1, it is easy to check that

$$B(u - u_h,\ v_h) = 0 \quad \forall v_h \in V_h. \tag{2.5}$$

Following [58] and [66], we state the following Gärding type inequality.

LEMMA 2.2.1 *Let $\beta \geq 1$ and $0 < \sigma_0 \leq \sigma_k \leq \sigma_m$. Further, assume that $\sigma_0 \geq C(\alpha,\ M,\ C_{T_2},\ N_\kappa)$ when $\theta = 1$, and $\sigma_0 > 0$ when $\theta = -1$. Then, there exist two constants $C_1 = C(\alpha,\ \sigma_0) > 0$ and $C_2 = C(\alpha,\ \sigma_0,\ M,\ C_{T_2},\ N_\kappa) > 0$ which are independent of h and p such that*

$$B(v_h,\ v_h) \geq C_1 |||v_h|||^2 - C_2 \|v_h\|^2 \qquad \forall v_h \in V_h.$$

A straightforward modification of the analysis of Prodhomme *et al.* [58, Theorem 3.4 and Theorem 3.5] and of Wheeler [66, Lemma 3] yields the proof of the Lemma 2.2.1 and hence, we omit the proof. Throughout this chapter, C denotes a generic constant which is independent of h, p and u_h, but may depend on κ, ϱ, σ_0, σ_m, α, M, C_A, C_{T_1}, C_{T_2}, C_I, C_P, C_1 and C_2.

Using the trace inequality (1.19) and (1.23), it is an easy exercise to prove the following Lemma 2.2.2. For details, see [58, Theorem 3.3].

LEMMA 2.2.2 *Let $\beta \geq 1$ and $\phi \in H^2(\Omega, \mathcal{T}_h)$. If σ_k is bounded above by a positive number σ_m, then there exists a positive constant C, independent of h and p, such that*

$$|B(\phi,\ v_h)| \leq C |||\phi|||_+\ |||v_h||| \qquad \forall v_h \in V_h.$$

LEMMA 2.2.3 *Let $\beta = 1$. Then, there exists a positive constant C which depends on C_A, but is, independent of h and p, such that*

$$|||\phi - I_h\phi|||_+ \leq C\left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i-2}}{p_i^{2s_i-3}}\|\phi\|^2_{H^{s_i}(K_i)}\right)^{1/2},$$

where $\mu_i = min\{p_i + 1, s_i\}$.

Proof. Let $\eta^* = \phi - I_h\phi$. Then, using (1.10), Lemma 1.2.1 and (1.23), we obtain

$$\begin{aligned} |||\eta^*|||^2_+ &= \sum_{i=1}^{N_h}\int_{K_i} |\nabla\eta^*|^2\, dx + \sum_{e_k\in\Gamma}\int_{e_k}\frac{|e_k|^\beta}{p_i^2}\{\frac{\partial\eta^*}{\partial\nu}\}^2\, ds + \sum_{e_k\in\Gamma}\int_{e_k}\frac{p_i^2}{|e_k|^\beta}[\eta^*]^2\, ds \\ &\leq C\sum_{i=1}^{N_h}\left(\frac{h_i^{2\mu_i-2}}{p_i^{2s_i-2}}\|\phi\|^2_{H^{s_i}(K_i)} + \frac{h_i^{2\mu_i-3+\beta}}{p_i^{2s_i-1}}\|\phi\|^2_{H^{s_i}(K_i)} + \frac{h_i^{2\mu_i-1-\beta}}{p_i^{2s_i-3}}\|\phi\|^2_{H^{s_i}(K_i)}\right) \quad (2.6) \end{aligned}$$

Since $\beta = 1$, the lemma is proved by taking a square root on both the sides of (2.6). ■

We prove the following lemma which will be used in the proof of *a priori* error estimates.

LEMMA 2.2.4 *Let $\beta = 1$ and $q \in L^2(\Omega)$. Then, for sufficiently small h, there exists a unique $\phi_h \in V_h$ satisfying*

$$B(v_h,\ \phi_h) = \int_\Omega q v_h\, dx \qquad \forall v_h \in V_h. \quad (2.7)$$

Moreover, ϕ_h satisfies

$$|||\phi_h||| \leq C\|q\|. \quad (2.8)$$

Proof. Note that (2.7) leads to a system of linear algebraic equations. So it is enough to prove uniqueness. Set $v_h = \phi_h$ in (2.7) and use Lemma 2.2.1 to obtain

$$\begin{aligned} C_1|||\phi_h|||^2 - C_2\|\phi_h\|^2 &\leq B(\phi_h,\ \phi_h) = \int_\Omega q\phi_h\, dx \\ &\leq \|q\|\ \|\phi_h\|. \end{aligned}$$

Therefore, we arrive at

$$|||\phi_h||| \leq C_1\|q\| + C_2\|\phi_h\|. \quad (2.9)$$

To estimate $\|\phi_h\|$ in terms of $|||\phi_h|||$, we apply the standard Aubin-Nitsche duality argument. For $\phi_h \in V_h$, we consider the following auxiliary problem:

$$\begin{aligned} -\nabla \cdot (a(x)\nabla\psi) + \vec{b}(x) \cdot \nabla\psi + a_0(x)\psi &= \phi_h \quad \text{in } \Omega, \\ \psi &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.10}$$

Then, assumption (**R**) implies that ψ satisfies the following elliptic regularity

$$\|\psi\|_{H^2(\Omega)} \leq C\|\phi_h\|_{L^2(\Omega)}. \tag{2.11}$$

We multiply (2.10) by ϕ_h and integrate over Ω, apply integration by parts and then use Lemmas 2.2.2-2.2.3 to obtain

$$\begin{aligned} \|\phi_h\|^2 &= B(\psi,\ \phi_h) = B(\psi - I_h\psi,\ \phi_h) + B(I_h\psi,\ \phi_h) = B(\psi - I_h\psi,\ \phi_h) + \int_\Omega qI_h\psi\ dx \\ &\leq C\left(h\ |||\phi_h||| + \|q\|\right)\|\psi\|_{H^2(\Omega)}. \end{aligned}$$

From the elliptic regularity (2.11), we now arrive at

$$\|\phi_h\| \leq Ch|||\phi_h||| + \|q\|. \tag{2.12}$$

Substituting (2.12) in (2.9), we obtain the estimate (2.8) for sufficiently small h. Hence, the problem (2.7) has a unique solution and this completes the rest of the proof. ■

2.2.2 A priori error estimates

Let $\beta = 1$. Since Lemma 2.2.1 holds for elements in V_h, we split $e = u - u_h$ into $e = \eta + \chi$, where $\eta = u - I_h u$ and $\chi = I_h u - u_h$. Then using Lemma 2.2.1, Lemma 2.2.2 and (2.5), we obtain

$$\begin{aligned} C_1|||\chi|||^2 - C_2\|\chi\|^2 \leq B(\chi,\ \chi) &= B((I_h u - u) + (u - u_h),\ \chi) \\ &= B(I_h u - u,\ \chi) = B(\eta,\ \chi) \\ &\leq C|||\eta|||_+\ |||\chi|||. \end{aligned}$$

Therefore,

$$|||\chi||| \leq C|||\eta|||_+ + C_2\|\chi\|. \tag{2.13}$$

In order to estimate $\|\chi\|$, we set $q = \chi$ and $v_h = \chi$ in Lemma 2.2.4. Using (2.5) and Lemma 2.2.2, we now obtain

$$\begin{aligned} \|\chi\|^2 &= B(\chi,\ \phi_h) = B(I_h u - u_h,\ \phi_h) = B(I_h u - u,\ \phi_h) \\ &\leq C|||\eta|||_+ \ |||\phi_h|||. \end{aligned}$$

Using (2.8), we arrive at

$$\|\chi\| \leq C|||\eta|||_+. \tag{2.14}$$

From the estimates (2.13) and (2.14), we obtain

$$|||\chi||| \leq C|||\eta|||_+. \tag{2.15}$$

Now using Lemma 2.2.3, the inequality (2.15) and the triangle inequality, we deduce the following theorem.

THEOREM 2.2.2 *Let $\beta = 1$; then for sufficiently small h, there exists a positive constant C which is independent of h and p such that*

$$|||u - u_h||| \ \leq \ C \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i - 2}}{p^{2s_i - 3}} \|u\|^2_{H^{s_i}(K_i)} \right)^{1/2},$$

where $\mu_i = min\{s_i,\ p_i + 1\}$.

Existence and uniqueness. We now prove existence of a unique solution to the problem (2.4) using the discrete dual problem (2.7) stated in Lemma 2.2.4. Assume that there exist two distinct solutions u_h^1 and u_h^2 for the problem (2.4). Let $\xi = u_h^1 - u_h^2$ and set $q = \xi$, $v_h = \xi$ in (2.7). Since $B(u_h^1 - u_h^2, v_h) = 0 \ \ \forall v_h \in \mathcal{D}_p(\mathcal{T}_h)$, we obtain

$$\|\xi\|^2 \ = \ B(\xi,\ \phi_h) = B(u_h^1 - u_h^2,\ \phi_h) = 0.$$

Therefore, $u_h^1 = u_h^2$ and this leads to a contradiction. Hence, we conclude that there exists a unique solution u_h for the problem (2.4). Now uniqueness implies existence of a discrete solution u_h to the problem (2.4).

2.3 Quasilinear Elliptic Problems

In this section, we consider the following nonlinear elliptic boundary value problem :

$$-\nabla \cdot (a(x,u)\nabla u) = f(x) \quad \text{in } \Omega, \tag{2.1}$$

$$u(x) = g(x) \quad \text{on } \partial\Omega, \tag{2.2}$$

where Ω is a bounded domain in $\mathbb{R}^2$ with smooth boundary $\partial\Omega$. As in [37], we make the following assumptions for the problem (2.1)-(2.2). There exist positive constants $\alpha,\ M$ such that $0 < \alpha \le a(x,u) \le M,\ x \in \bar{\Omega},\ \ a(\cdot,\cdot) \in C_b^2(\bar{\Omega} \times \mathbb{R})$, where $C_b^2(\bar{\Omega} \times \mathbb{R})$ is the class of twice continuously differentiable functions on $\bar{\Omega} \times \mathbb{R}$ such that all derivatives of $a(\cdot,\cdot)$ up to and including second order are bounded in $\bar{\Omega} \times \mathbb{R}$. Further for some $\delta \in (0,1)$, $f \in C^{\delta}(\Omega)$ and g can be extended to Ω to be in $C^{2+\delta}(\Omega)$, then, it follows from [36] that there exists a unique weak solution u to (2.1)-(2.2) and $u \in C^{2+\delta}(\bar{\Omega})$, where $C^{m+\delta}(\bar{\Omega})$ consists of all functions whose mth order derivatives are Hölder continuous of order δ on $\bar{\Omega}$.

2.3.1 Weak formulation

For $\psi,\ w$ and $v \in H^2(\Omega, \mathcal{T}_h)$, we define the form $B(\psi; w,\ v)$ which is linear in $w,\ v$ for fixed ψ by

$$\begin{aligned} B(\psi; w,\ v) = & \sum_{i=1}^{N_h} \int_{K_i} a(\psi)\nabla w \cdot \nabla v \ dx - \sum_{e_k \in \Gamma_I} \int_{e_k} (\{a(\psi)\frac{\partial w}{\partial \nu}\}[v] \ ds + \theta\{a(\psi)\frac{\partial v}{\partial \nu}\}[w]) \ ds \\ & - \sum_{e_k \in \Gamma_\partial} \int_{e_k} (a(g)\frac{\partial w}{\partial \nu} v + \theta a(g)\frac{\partial v}{\partial \nu} w) \ ds + \mathcal{J}^{\sigma,\beta}(w,\ v), \end{aligned}$$

and the linear functional L on $H^2(\Omega,\ \mathcal{T}_h)$ by

$$L(v) = \int_{\Omega} fv \ dx + \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} a(g)\frac{\partial v}{\partial \nu} g \ ds + \sum_{e_k \in \Gamma_\partial} \int_{e_k} \sigma_k \frac{p_k^2}{|e_k|^\beta} vg \ ds,$$

where $\theta = \pm 1$. Since for each fixed ψ, $B(\psi; \cdot,\ \cdot)$ is a bilinear form, we note that $\theta = +1$ corresponds to a symmetric and $\theta = -1$ to a non-symmetric method. We define the weak formulation of (2.1)-(2.2) which is suitable for applying a discontinuous Galerkin method as: Find $u \in H^2(\Omega, \mathcal{T}_h)$ such that

$$B(u; u,\ v) = L(v) \quad \forall v \in H^2(\Omega, \mathcal{T}_h). \tag{2.3}$$

Now the discontinuous Galerkin (SIPG and NIPG) approximation of u is to seek $u_h \in V_h$ such that

$$B(u_h; u_h,\ v_h) = L(v_h) \quad \forall v_h \in V_h. \tag{2.4}$$

Below, we state without proof the consistency of the above scheme (2.4).

THEOREM 2.3.1 *(Equivalence of (2.1)-(2.2) and (2.3)). If the solution u of (2.1)-(2.2) is in $H^2(\Omega)$, then u satisfies (2.3). Conversely, if $u \in H^1(\Omega) \cap H^2(\Omega, \mathcal{T}_h)$ is a solution of (2.3) , then u satisfies (2.1)-(2.2) weakly.*

The proof follows the lines of the proof given in [61, Lemma 2.2] or [58, Theorem 3.1], so we omit it. With $v = v_h \in V_h \subset H^2(\Omega, \mathcal{T}_h)$ in (2.3), we obtain using (2.4)

$$B(u; u,\ v_h) = B(u_h; u_h,\ v_h) \quad \forall v_h \in V_h. \tag{2.5}$$

Following Taylor series expansion, we write

$$a(w) = a(u) + \tilde{a}_u(w)(w - u), \tag{2.6}$$

where $\tilde{a}_u(w) = \displaystyle\int_0^1 a_u(w + t(u - w))dt$, and

$$a(w) = a(u) + a_u(u)(w - u) + \tilde{a}_{uu}(w)(w - u)^2, \tag{2.7}$$

where $\tilde{a}_{uu}(w) = \displaystyle\int_0^1 (1 - t)a_{uu}(w + t(w - u))dt$.

Note that since $a_u \in C_b^1(\bar{\Omega} \times \mathbb{R})$ and $a_{uu} \in C_b^0(\bar{\Omega} \times \mathbb{R})$, it is easy to see that $\tilde{a}_u \in L^\infty(\Omega \times \mathbb{R})$ and $\tilde{a}_{uu} \in L^\infty(\Omega \times \mathbb{R})$. We use the following notation throughout this section :

$$C_a = \max\left[\ \|\tilde{a}_u\|_{L^\infty(\Omega\times\mathbb{R})},\ \|\tilde{a}_{uu}\|_{L^\infty(\Omega\times\mathbb{R})}\right]. \tag{2.8}$$

REMARK 2.3.1 *For our subsequent analysis, it is sufficient to assume that a is locally bounded. In fact, it is enough to assume that a along with its derivatives are bounded in a ball around u, see Remark 2.3.5.*

For simplicity, we consider the following form $\tilde{B}(\cdot\ ;\ \cdot\ ,\ \cdot)$

$$\tilde{B}(\psi; w,\ v) = B(\psi; w,\ v) + \sum_{i=1}^{N_h} \int_{K_i} (a_u(\psi)\nabla\psi)w \cdot \nabla v\ dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{a_u(\psi)\frac{\partial \psi}{\partial \nu} w\}[v]\ ds.$$

Note that $\tilde{B}$ is linear in w and $v \in H^2(\Omega, \mathcal{T}_h)$ for a fixed ψ. It is clear from the assumptions on $a(u)$ that Lemma 2.2.1 and Lemma 2.2.2 hold for $\tilde{B}$. Since $a \in C_b^2(\bar{\Omega}\times\mathbb{R})$ and $u \in C^2(\bar{\Omega})$, there is a unique solution $\psi \in H^2(\Omega)$ to the following elliptic problem:

$$\begin{aligned} -\nabla \cdot (a(u)\nabla\psi + a_u(u)\nabla u\psi) &= \phi_h \quad \text{in } \Omega, \\ \psi &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{2.9}$$

and ψ satisfies the elliptic regularity $\|\psi\|_{H^2(\Omega)} \leq C\|\phi_h\|$, see [36, Theorem 2], [37, p. 692]. Hence, Lemma 2.2.4 holds as well for $\tilde{B}$. Now we linearize the problem (2.4) around $I_h u$ for our subsequent analysis. Set $e = u - u_h$. Subtracting $B(u; u_h,\ v_h)$ from both the sides of (2.5), we obtain

$$\begin{aligned} B(u; e,\ v_h) &= \sum_{i=1}^{N_h} \int_{K_i} (a(u_h) - a(u))\nabla u_h \cdot \nabla v_h \, dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{(a(u_h) - a(u))\frac{\partial u_h}{\partial \nu}\}[v_h] \, ds \\ &\quad -\theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{(a(u_h) - a(u))\frac{\partial v_h}{\partial \nu}\}[u_h] \, ds. \end{aligned} \tag{2.10}$$

Since $[u] = 0$ on each $e_k \in \Gamma_I$, we rewrite (2.10) as

$$\begin{aligned} B(u; e,\ v_h) = &\sum_{i=1}^{N_h} \int_{K_i} (a(u_h) - a(u))\nabla(u_h - u) \cdot \nabla v_h \, dx - \int_{\Gamma_I} \{(a(u_h) - a(u))\frac{\partial u}{\partial \nu}\}[v_h] \, ds \\ &- \sum_{e_k \in \Gamma_I} \int_{e_k} \{(a(u_h) - a(u))\frac{\partial (u_h - u)}{\partial \nu}\}[v_h] \, ds + \sum_{i=1}^{N_h} \int_{K_i} (a(u_h) - a(u))\nabla u \cdot \nabla v_h \, ds \\ &-\theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{(a(u_h) - a(u))\frac{\partial v_h}{\partial \nu}\}[u_h - u] \, ds. \end{aligned} \tag{2.11}$$

Finally, we add the following terms to both the sides of (2.11)

$$-\sum_{i=1}^{N_h} \int_{K_i} a_u(u)(u_h - u)\nabla u \cdot \nabla v_h \, dx + \sum_{e_k \in \Gamma_I} \int_{e_k} \{a_u(u)(u_h - u)\frac{\partial u}{\partial \nu}\}[v_h] \, ds.$$

We split $e = u - u_h = u - I_h u + I_h u - u_h$. Now using the Taylor formulae (2.6)-(2.7), the equation (2.11) takes the form

$$\tilde{B}(u; I_h u - u_h,\ v_h) = \tilde{B}(u; I_h u - u,\ v_h) + \mathcal{F}(u_h; u_h - u,\ v_h), \tag{2.12}$$

where

$$\begin{aligned}\mathcal{F}(u_h; -e,\ v_h) &= \sum_{i=1}^{N_h}\int_{K_i}\tilde{a}_u(u_h)e\nabla e\cdot\nabla v_h\ dx - \sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_u(u_h)e\frac{\partial e}{\partial\nu}\}[v_h]\ ds \\ &\quad -\theta\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_u(u_h)e\frac{\partial v_h}{\partial\nu}\}[e]\ ds + \sum_{i=1}^{N_h}\int_{K_i}\tilde{a}_{uu}(u_h)e^2\nabla u\cdot\nabla v_h\ dx \\ &\quad -\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_{uu}(u_h)e^2\frac{\partial u}{\partial\nu}\}[v_h]\ ds. \end{aligned} \tag{2.13}$$

Note that (2.4) is equivalent to (2.12).

2.3.2 Existence and Uniqueness

For a given $z \in V_h$, let $S_h : V_h \to V_h$ be a map $y = S_h z \in V_h$ satisfying

$$\tilde{B}(u; I_h u - y,\ v_h) = \tilde{B}(u; I_h u - u,\ v_h) + \mathcal{F}(z; z-u,\ v_h) \quad \forall v_h \in V_h. \tag{2.14}$$

For a given z, the problem (2.14) leads to a system of linear algebraic equations. So using Lemma 2.2.1 and Lemma 2.2.4, it is easy to show that the map S_h is well defined. Now consider the following ball

$$\mathcal{O}_\delta(I_h u) = \{z \in V_h :\ |||I_h u - z||| \le \delta\}$$

of radius δ, where δ will be chosen later. We first show that for some $\delta > 0$, S_h maps $\mathcal{O}_\delta(I_h u)$ into itself and S_h is continuous. Then an appeal to Brouwer's fixed point theorem yields the existence of a solution to the problem (2.12) and hence, there exists a solution to the problem (2.4). The following lemma is a key result for proving existence of a unique solution to the discrete problem (2.4). Throughout this section, we use the following notation to denote the Sobolev norm of u:

$$C_u = \max\left[\ \|u\|_{H^2(\Omega)},\ \|u\|_{H^1(\Omega)}|u|_{W^1_\infty(\Omega)}\right]. \tag{2.15}$$

LEMMA 2.3.1 *Let $\beta \ge 1$, and $z,\ v_h \in V_h$. Set $\chi = z - I_h u$ and $\eta = u - I_h u$. Then, there exists a constant $C > 0$ which is independent of h and p such that*

$$|\mathcal{F}(z; z-u,\ v_h)| \le CC_a\left[\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}|||\chi|||^2 + C_u h^{1/2}(|||\chi||| + |||\eta|||)\right]|||v_h|||.$$

Proof. Let $z \in V_h$ and set $\zeta = z - u$. In (2.13), we now replace u_h by z and e by $z - u$ to obtain

$$\mathcal{F}(z;\zeta,\ v_h) = \sum_{i=1}^{N_h} \int_{K_i} \tilde{a}_u(z)\zeta\nabla\zeta\cdot\nabla v_h\ dx - \sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_u(z)\zeta\frac{\partial\zeta}{\partial\nu}\}[v_h]\ ds$$
$$-\theta\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_u(z)\zeta\frac{\partial v_h}{\partial\nu}\}[\zeta]\ ds + \sum_{i=1}^{N_h}\int_{K_i}\tilde{a}_{uu}(z)\zeta^2\nabla u\cdot\nabla v_h\ dx$$
$$-\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{a}_{uu}(z)\zeta^2\frac{\partial u}{\partial\nu}\}[v_h]\ ds. \qquad (2.16)$$

We split $\zeta = \chi - \eta$, where $\chi = z - I_h u$ and $\eta = u - I_h u$. Then, we estimate from below the bound for each term on the right-hand side of (2.16). For the first term on the right-hand side of (2.16), we split and then bound it as

$$\sum_{i=1}^{N_h}\int_{K_i}|\tilde{a}_u(z)\zeta\nabla\zeta\cdot\nabla v_h|\ dx \le C_a\sum_{i=1}^{N_h}\int_{K_i}|\chi\nabla\chi\cdot\nabla v_h|\ dx + C_a\sum_{i=1}^{N_h}\int_{K_i}|\chi\nabla\eta\cdot\nabla v_h|\ dx$$
$$+C_a\sum_{i=1}^{N_h}\int_{K_i}|\eta\nabla\chi\cdot\nabla v_h|\ dx + C_a\sum_{i=1}^{N_h}\int_{K_i}|\eta\nabla\eta\cdot\nabla v_h|\ dx. \qquad (2.17)$$

Using Hölder's inequality and the inverse inequality (1.20), we estimate the first term on the right-hand side of (2.17) as

$$\begin{aligned}
\sum_{i=1}^{N_h}\int_{K_i}|\chi\nabla\chi\cdot\nabla v_h|dx &\le \sum_{i=1}^{N_h}\|\chi\|_{L^6(K_i)}\|\nabla\chi\|_{L^3(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\le C\sum_{i=1}^{N_h}\|\chi\|_{L^6(K_i)}\frac{p_i^{1/3}}{h_i^{1/3}}\|\nabla\chi\|_{L^2(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\le C\|\chi\|_{L^6(\Omega)}\left(\sum_{i=1}^{N_h}\frac{p_i}{h_i}\|\nabla\chi\|^3_{L^2(K_i)}\right)^{1/3}|v_h|_{1,h}\\
&\le C\|\chi\|_{L^6(\Omega)}\left[\max_{K_i}\|\nabla\chi\|^{1/3}_{L^2(K_i)}\left(\sum_{i=1}^{N_h}\frac{p_i}{h_i}\|\nabla\chi\|^2_{L^2(K_i)}\right)^{1/3}\right]|v_h|_{1,h}\\
&\le C|||\chi|||\left(\sum_{i=1}^{N_h}\|\nabla\chi\|^2_{L^2(K_i)}\right)^{1/6}\left(\sum_{i=1}^{N_h}\frac{p_i}{h_i}\|\nabla\chi\|^2_{L^2(K_i)}\right)^{1/3}|||v_h|||\\
&\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/3}|||\chi|||^2\ |||v_h|||. \qquad (2.18)
\end{aligned}$$

For the second term on the right-hand side of (2.17), we use Hölder's inequality, and Lemma 1.2.1 to obtain

$$\begin{aligned}
\sum_{i=1}^{N_h}\int_{K_i}|\chi\nabla\eta\cdot\nabla v_h|dx &\leq C\sum_{i=1}^{N_h}\|\chi\|_{L^6(K_i)}\|\nabla\eta\|_{L^3(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\leq C\sum_{i=1}^{N_h}\|\chi\|_{L^6(K_i)}\frac{h_i^{2/3}}{p_i^{2/3}}\|u\|_{H^2(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\leq C\frac{h^{2/3}}{p^{2/3}}\|\chi\|_{L^6(\Omega)}\left(\sum_{i=1}^{N_h}\|u\|^3_{H^2(K_i)}\right)^{1/3}|v_h|_{1,h}
\end{aligned}$$

$$\begin{aligned}
&\leq C\frac{h^{2/3}}{p^{2/3}}|||\chi|||\left[\max_{K_i}\|u\|^{1/3}_{H^2(K_i)}\left(\sum_{i=1}^{N_h}\|u\|^2_{H^2(K_i)}\right)^{1/3}\right]|||v_h|||\\
&\leq C\frac{h^{2/3}}{p^{2/3}}|||\chi|||\left(\sum_{i=1}^{N_h}\|u\|^2_{H^2(K_i)}\right)^{1/2}|||v_h|||\\
&\leq C\frac{h^{2/3}}{p^{2/3}}\|u\|_{H^2(\Omega)}|||\chi|||\ |||v_h|||. \qquad (2.19)
\end{aligned}$$

To estimate the third term on the right-hand side of (2.17), apply Hölder's inequality and the inverse inequality (1.20) to find, following the estimate (2.20), that

$$\begin{aligned}
\sum_{i=1}^{N_h}\int_{K_i}|\eta\nabla\chi\cdot\nabla v_h|dx &\leq C\sum_{i=1}^{N_h}\|\eta\|_{L^6(K_i)}\|\nabla\chi\|_{L^3(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\leq C\sum_{i=1}^{N_h}\frac{h_i^{4/3}}{p_i^{4/3}}\|u\|_{H^2(K_i)}\frac{p_i^{2/3}}{h_i^{2/3}}\|\nabla\chi\|_{L^2(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\leq C\frac{h^{2/3}}{p^{2/3}}\|u\|_{H^2(\Omega)}|||\chi|||\ |||v_h|||. \qquad (2.20)
\end{aligned}$$

For the last term on the right-hand side of (2.17), we use Lemma 1.2.1 and Lemma 1.2.6 to estimate it as

$$\begin{aligned}
\sum_{i=1}^{N_h}\int_{K_i}|\eta\nabla\eta\cdot\nabla v_h|dx &\leq C\sum_{i=1}^{N_h}\|\eta\|_{L^6(K_i)}\|\nabla\eta\|_{L^3(K_i)}\|\nabla v_h\|_{L^2(K_i)}\\
&\leq C\frac{h^{2/3}}{p^{2/3}}\|u\|_{H^2(\Omega)}|||\eta|||\ |||v_h|||. \qquad (2.21)
\end{aligned}$$

We substitute the estimates (2.18)-(2.21) into (2.17) to obtain

$$\sum_{i=1}^{N_h}\int_{K_i}|\tilde{a}_u(z)\zeta\nabla\zeta\cdot\nabla v_h|dx \;\leq\; CC_a\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}|||\chi|||^2\,|||v_h|||$$
$$+CC_a\frac{h^{2/3}}{p^{2/3}}\|u\|_{H^2(\Omega)}\left(|||\chi|||+|||\eta|||\right)|||v_h|||. \qquad (2.22)$$

As in (2.17) the second term on the right-hand side of (2.16) becomes

$$\sum_{e_k\in\Gamma_I}\int_{e_k}|\{\tilde{a}_u(z)\zeta\frac{\partial\zeta}{\partial\nu}\}[v_h]|ds\leq C\left(C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\chi\frac{\partial\chi}{\partial\nu}||[v_h]|ds+C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\chi\frac{\partial\eta}{\partial\nu}||[v_h]|ds\right.$$
$$\left.+C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\frac{\partial\chi}{\partial\nu}||[v_h]|ds+C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\frac{\partial\eta}{\partial\nu}||[v_h]|ds\right). \qquad (2.23)$$

Using Hölder's inequality, the inverse inequality (1.22), the trace inequalities (1.14)-(1.19) and (1.23), the first term on the right-hand side of (2.23) is estimated as

$$\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\chi\frac{\partial\chi}{\partial\nu}||[v_h]|ds &\leq C\sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{\beta/2}}{p_k}\|\chi\|_{L^4(e_k)}\|\nabla\chi\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\right)\\
&\leq C\sum_{e_k\in\Gamma_I}\frac{|e_k|^{\beta/2-1/4}}{p_k^{1/2}}\|\chi\|_{L^4(e_k)}\|\nabla\chi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{p_k^{1/2}}{|e_k|^{1-\beta/2}}\|\nabla\chi\|_{L^2(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\\
&\qquad\left(\|\chi\|^4_{L^4(K_i)}+h_i\|\chi\|^3_{L^6(K_i)}\|\nabla\chi\|_{L^2(K_i)}\right)^{1/4}\\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i^{2-\beta}}\right)^{1/2}|||\chi|||^2\,|||v_h|||. \qquad (2.24)
\end{aligned}$$

Similarly, we use Hölder's inequality, (1.23), the trace inequality (1.14) and Lemma 1.2.1 to estimate the second term on the right-hand side of (2.23) as

$$\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\chi\frac{\partial\eta}{\partial\nu}||[v_h]|ds &\leq C\sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{\beta/2}}{p_k}\|\chi\|_{L^4(e_k)}\|\nabla\eta\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{|e_k|^{\beta/2-1/2}}{p_k}\left(\|\chi\|^4_{L^4(K_i)}+h_i\|\chi\|^3_{L^6(K_i)}\|\nabla\chi\|_{L^2(K_i)}\right)^{1/4}\\
&\qquad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}
\end{aligned}$$

$$
\begin{aligned}
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{h_i^{\beta/2-1/2}}{p_k}\left(\frac{h_i^2}{p_i^2}\|u\|^4_{H^2(K_i)}+h_i\frac{h_i}{p_i}\|u\|^4_{H^2(K_i)}\right)^{1/4}\\
&\quad\left(\|\chi\|^4_{L^4(K_i)}+h_i\|\chi\|^3_{L^6(K_i)}\|\nabla\chi\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h^{\beta/2}}{p^{3/2}}+\frac{h^{\beta/2}}{p^{5/4}}\right)|||\chi|||\;|||v_h|||. \qquad (2.25)
\end{aligned}
$$

For the third on the right-hand side of (2.23), apply Hölder's inequality, the trace inequalities (1.14)-(1.19) and Lemma 1.2.1 to find that

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\frac{\partial\chi}{\partial\nu}||[v_h]|ds &\leq C\sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{\beta/2}}{p_k}\|\eta\|_{L^4(e_k)}\|\nabla\chi\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{|e_k|^{\beta/2-1/2}}{p_k^{1/2}}\|\nabla\chi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\\
&\qquad\left(\|\eta\|^4_{L^4(K_i)}+h_i\|\eta\|^3_{L^6(K_i)}\|\nabla\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h^{1/2+\beta/2}}{p}+\frac{h^{1/2+\beta/2}}{p^{3/4}}\right)|||\chi|||\;|||v_h|||. \qquad (2.26)
\end{aligned}
$$

For the last term on the right-hand side of (2.23), we use a similar argument as in (2.26) to obtain

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\frac{\partial\eta}{\partial\nu}||[v_h]|ds &\leq C\sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{\beta/2}}{p_k}\|\eta\|_{L^4(e_k)}\|\nabla\eta\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{|e_k|^{\beta/2-1/2}}{p_k}\left(\|\eta\|^4_{L^4(K_i)}+h_i\|\eta\|^3_{L^6(K_i)}\|\nabla\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\qquad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[v_h]^2ds\right)^{1/2}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h^{\beta/2}}{p^{3/2}}+\frac{h^{\beta/2}}{p^{5/4}}\right)|||\eta|||\;|||v_h|||. \qquad (2.27)
\end{aligned}
$$

We substitute the estimates (2.24)-(2.27) into (2.23). Since $\beta\geq 1$, we obtain

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\zeta\frac{\partial\zeta}{\partial\nu}||[v_h]| &\leq CC_a\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}|||\chi|||^2\;|||v_h|||\\
&\quad+CC_ah^{1/2}\|u\|_{H^2(\Omega)}(|||\chi|||+|||\eta|||)|||v_h|||. \qquad (2.28)
\end{aligned}
$$

In a similar way, we find the following estimates for the third, fourth and the last terms on the right-hand side of (2.16):

$$\begin{aligned}\sum_{e_k\in\Gamma_I}\int_{e_k}|\zeta\frac{\partial v_h}{\partial\nu}|||[\zeta]|ds &\le CC_a\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}|||\chi|||^2\,|||v_h||| \\ &\quad +CC_ah^{1/2}\|u\|_{H^2(\Omega)}(|||\chi|||+|||\eta|||)|||v_h|||, \qquad (2.29)\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{N_h}\int_{K_i}|(\tilde{a}_{uu}(z)\nabla u)\zeta^2\cdot\nabla v_h|dx &\le CC_a|||\chi|||^2\,|||v_h||| \\ &+ CC_ah^{1/2}\|u\|_{H^1(\Omega)}|u|_{W^1_\infty(\Omega)}(|||\chi|||+|||\eta|||)|||v_h|||, \qquad (2.30)\end{aligned}$$

and

$$\begin{aligned}\sum_{e_k\in\Gamma}\int_{e_k}|\{(\tilde{a}_{uu}(z)\frac{\partial u}{\partial\nu})\zeta^2\}[v_h]|ds &\le CC_a|||\chi|||^2\,|||v_h||| \\ &+ CC_ah^{1/2}\|u\|_{H^1(\Omega)}|u|_{W^1_\infty(\Omega)}(|||\chi|||+|||\eta|||)|||v_h|||. \qquad (2.31)\end{aligned}$$

Substituting the estimates (2.22) and (2.28)-(2.31) into (2.16), we complete the rest of the proof. ■

LEMMA 2.3.2 *Let $\beta\ge 1$ and $z\in V_h$. Set $y=S_hz$. Then, there exists a positive constant C which is independent of h and p such that*

$$\begin{aligned}|||I_hu-y|||\ &\le\ CC_a\left[\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}|||I_hu-z|||^2+C_uh^{1/2}|||I_hu-z|||\right] \\ &\quad +CC_a\left[(1+C_uh^{1/2})|||I_hu-u|||_+\right].\end{aligned}$$

Proof. Let $\chi=I_hu-z$, $\eta=I_hu-u$ and $\xi=I_hu-y$. Set $v_h=\xi$ in (2.14). Then for the first term on the right-hand side of (2.14), use Lemma 2.2.2 to obtain

$$|\tilde{B}(u;\eta,\xi)|\ \le\ C|||\eta|||_+\,|||\xi|||. \qquad (2.32)$$

Set $v_h=\xi$ in Lemma 2.3.1 to arrive at

$$\begin{aligned}|\mathcal{F}(z;z-u,\ \xi)|\ &\le\ CC_a\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}|||\chi|||^2\,|||\xi||| \qquad (2.33)\\ &\quad +CC_aC_uh^{1/2}\,(|||\chi|||+|||\eta|||)\ |||\xi|||.\end{aligned}$$

Substituting the estimates (2.32)-(2.33) into (2.14) and using the fact that $|||\eta||| \leq |||\eta|||_+$, we obtain

$$\begin{aligned}
|\tilde{B}(u;\xi,\ \xi) &\leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + C_u h^{1/2} \left(|||\chi||| + |||\eta||| \right) + |||\eta|||_+ \right] |||\xi||| \\
&\leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + C_u h^{1/2} |||\chi||| \right] \ |||\xi||| \\
&\quad + CC_a (1 + C_u h^{1/2}) |||\eta|||_+ \ |||\xi|||.
\end{aligned}$$

Then using Gärding inequality, that is Lemma 2.2.1, we obtain

$$\begin{aligned}
C_1 |||\xi||^2 - C_2 \|\xi\|^2 &\leq \tilde{B}(u;\xi,\ \xi). \\
&\leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + C_u h^{1/2} |||\chi||| \right] |||\xi||| \\
&\quad + CC_a \left[(1 + C_u h^{1/2}) |||\eta|||_+ \right] |||\xi|||,
\end{aligned}$$

and hence,

$$\begin{aligned}
|||\xi||| &\leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + C_u h^{1/2} |||\chi||| \right] \\
&\quad + CC_a \left[(1 + C_u h^{1/2}) |||\eta|||_+ \right] + C\|\xi\|.
\end{aligned} \tag{2.34}$$

In order to complete the proof of the lemma, it is now sufficient to obtain an estimate for $\|\xi\|$. Setting $q = \xi$ and $v_h = \xi$ in the Lemma 2.2.4, it follows that

$$\begin{aligned}
\|\xi\|^2 &= \tilde{B}(u; I_h u - y,\ \phi_h) = \tilde{B}(u; I_h u - u,\ \phi_h) + \mathcal{F}(z; z - u, \phi_h) \\
&\leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + C_u h^{1/2} |||\chi||| + (1 + C_u h^{1/2}) |||\eta|||_+ \right] |||\phi_h|||.
\end{aligned}$$

Therefore, using the fact from Lemma 2.2.4 that $|||\phi_h||| \leq C\|\xi\|$, we obtain

$$\|\xi\| \leq CC_a \left[\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i} \right)^{1/2} |||\chi|||^2 + h^{1/2} C_u |||\chi||| + (1 + C_u h^{1/2}) |||\eta|||_+ \right]. \tag{2.35}$$

We combine the inequalities (2.34) and (2.35) to complete the rest of the proof. ■

THEOREM 2.3.2 *Let $\beta = 1$. Then, for sufficiently small h, there is a $\delta = \delta(h,p)$ such that the map S_h maps $\mathcal{O}_\delta(I_h u)$ into itself.*

Proof. Let $z \in \mathcal{O}_\delta(I_h u)$ and set $y = S_h z$. Choose $\delta = \dfrac{1}{h^\epsilon}|||I_h u - u|||_+$, for some $0 < \epsilon \leq 1/4$. Then using the fact that $z \in \mathcal{O}_\delta(I_h u)$, Lemma 2.2.3 with $s_i \geq 2,\ p_i \geq 1$ and (1.25), we obtain

$$\begin{aligned}
\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)^{1/2} |||I_h u - z|||^2 &\leq \left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)^{1/2} \delta^2 \\
&\leq \left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)^{1/2} \frac{1}{h^\epsilon}|||I_h u - u|||_+ \ \delta \\
&\leq C\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)^{1/2} \left[\sum_{i=1}^{N_h} \frac{h_i^2}{p_i}\|u\|^2_{H^2(K_i)}\right]^{1/2} \delta \\
&\leq C\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)^{1/2} \left(\max_{1\leq i\leq N_h} \frac{h_i^2}{p_i}\right)^{1/2} \|u\|_{H^2(\Omega)}\ \delta \\
&\leq CC_Q C_u h^{1/2-\epsilon}\ \delta. \qquad (2.36)
\end{aligned}$$

We substitute (2.36) in Lemma 2.3.2 to arrive at

$$\begin{aligned}
|||I_h u - y||| &\leq CC_a\left(C_Q C_u h^{1/2-\epsilon}\ \delta + C_u h^{1/2}\delta + (1 + C_u h^{1/2})h^\epsilon \delta\right) \qquad (2.37)\\
&\leq CC_a\left(C_Q C_u h^{1/2-\epsilon} + C_u h^{1/2} + (1 + C_u h^{1/2})h^\epsilon\right)\ \delta.
\end{aligned}$$

Choose h small so that $CC_a\left(C_Q C_u h^{1/2-\epsilon} + C_u h^{1/2} + (1 + C_u h^{1/2})h^\epsilon\right) \leq 1$ and hence, S_h maps $\mathcal{O}_\delta(I_h u)$ into itself. This completes the rest of the proof. ■

THEOREM 2.3.3 *Let $\beta = 1$. There is a $\delta = \delta(h,\ p) > 0$ and a positive constant C such that the following holds for any given $z_1,\ z_2 \in \mathcal{O}_\delta(I_h u)$ and $0 < \epsilon \leq \frac{1}{4}$*

$$|||S_h z_1 - S_h z_2||| \leq CC_a C_Q C_u h^\epsilon\ |||z_1 - z_2|||.$$

Proof. Set $y_1 = S_h z_1$ and $y_2 = S_h z_2$. Using the definition (2.14) of S_h, it is clear that

$$\tilde{B}(u; y_2 - y_1,\ v_h) \ = \ \mathcal{F}(z_1; z_1 - u,\ v_h) - \mathcal{F}(z_2; z_2 - u,\ v_h). \qquad (2.38)$$

Choose $\delta = \frac{1}{h^\epsilon}|||\eta|||_+$, for some $0 < \epsilon \leq 1/4$ with $\eta = u - I_h u$. Set $\chi = z_1 - z_2$. Using Taylor's formulae (1.14)-(1.19) and (2.16), we rewrite the first terms from each of the terms

on the right-hand side of (2.38) on each K_i as

$$\begin{aligned}\tilde{a}_{uu}(z_1)(z_1-u)^2-\tilde{a}_{uu}(z_2)(z_2-u)^2 &= a(z_1)-a(z_2)+a_u(u)(z_2-z_1)\\ &= a(z_1)-a(z_2)-a_u(z_2)\chi+(a_u(z_2)-a_u(u))\,\chi\\ &= \tilde{R}(z_1,z_2)\chi^2+\tilde{a}_{uu}(z_2)(z_2-u)\chi,\end{aligned}$$

where $\tilde{R}(z_1,z_2)=\int_0^1(1-t)a_{uu}(z_1+t[z_1-z_2])dt$. Similarly, other terms on the right-hand side of (2.38) can be rewritten in a similar fashion. Now, a similar argument as in Lemma 2.3.1 implies that

$$\begin{aligned}|\mathcal{F}(z_1;z_1-u,\ v_h)-\mathcal{F}(z_2;z_2-u,\ v_h)| &\le CC_a\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}\Big[\ |||\chi|||^2\ |||v_h|||\\ &\qquad +|||z_1-I_hu|||\ |||\chi|||\ |||v_h|||+\ \ |||I_hu-z_2|||\ |||\chi|||\ |||v_h|||\ \Big]\\ &\le CC_aC_QC_uh^{\epsilon}\ |||\chi|||\ |||v_h|||. \qquad (2.39)\end{aligned}$$

We set $v_h=(y_2-y_1)$ in (2.38) and (2.39). Then, using Lemma 2.2.1 and Lemma 2.2.4, we obtain

$$|||y_1-y_2|||\le CC_aC_QC_uh^{\epsilon}\ |||z_1-z_2|||,$$

and this completes the proof. ■

For sufficiently small h, we deduce from Theorem 2.3.3 that there is a $\delta>0$ such that the map $S_h:\mathcal{O}_\delta(I_hu)\to\mathcal{O}_\delta(I_hu)$ is continuous. Hence, using Theorem 2.3.2, Theorem 2.3.3 and Brouwer's fixed point theorem, we conclude for small h that there exists a $u_h\in\mathcal{O}_\delta(I_hu)$ such that $S_hu_h=u_h$. Then, from Theorem 2.3.3, it is clear that u_h is the unique fixed point of S_h. Hence, we have proved that there exists a unique solution u_h to the problem (2.4).

2.3.3 A priori error Estimates

Note that, from (2.37), u_h satisfies

$$|||I_hu-u_h|||\ \ \le\ \ CC_a\left(C_QC_uh^{1/2-\epsilon}+C_uh^{1/2}+(1+C_uh^{1/2})h^{\epsilon}\right)\ \delta.$$

Since $\delta = \frac{1}{h^\epsilon}|||\eta|||_+$, for any $0 < \epsilon \leq 1/4$, we obtain

$$\begin{aligned} |||I_h u - u_h||| &\leq CC_a \left(C_Q C_u h^{1/2-\epsilon} + C_u h^{1/2} + (1 + C_u h^{1/2})h^\epsilon\right) \frac{1}{h^\epsilon}|||\eta|||_+ \\ &\leq CC_a C_Q C_u |||\eta|||_+. \end{aligned} \tag{2.40}$$

Using Lemma 2.2.3, estimate (2.40) and a triangle inequality, we have obtained the following estimate which is optimal in h and suboptimal in p.

THEOREM 2.3.4 *Let $\beta = 1$. Then, for sufficiently small h, there exists a constant $C = C(\alpha, M)$ which is independent of h and p such that the solution u_h of the problem (2.4) satisfies*

$$|||u - u_h||| \leq CC_a C_Q C_u \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i - 2}}{p_i^{2s_i - 3}} \|u\|^2_{H^{s_i}(K_i)}\right)^{1/2},$$

where $\mu_i = min\{p_i + 1,\ s_i\}$, C_a, C_Q and C_u are as in (2.8), (1.24) and (2.15), respectively.

REMARK 2.3.2 *Note that the estimate obtained in the Theorem 2.3.4 is optimal in h and suboptimal in p. However, this results leads precisely the same h-optimal and p-suboptimal rate of convergence in the broken H^1-norm as in the case of linear elliptic problem, when it is approximated by NIPG method [61, Theorem 3.1] and [44].*

2.3.4 Optimal error estimates in the broken energy norm and the L^2-norm, when $u \in H^s_p(\Omega),\ s \geq 2$

In the following, with the additional assumptions on the mesh and g, we prove optimal error estimates in the broken energy norm as well as in the L^2-norm. Therefore, along with the assumption (**Q**), we also assume that $\mathcal{T}_h$ is a regular subdivision of Ω into triangles or rectangles and $V_h = V_h^*$. We also assume that there is a $v \in V_h^*$ such that $g = v|_{\partial\Omega}$.

We note from [61, p. 908-913] that by using a continuous interpolant $I_h^c u \in V_h \cap \bar{C}^0(\Omega)$ of u instead of $I_h u$ which may be discontinuous across the edges in Γ_I, the optimal rate of convergence can be recovered. Since the construction of I_h^c is not discussed in [61], we present below the results related to the construction of I_h^c. The idea of constructing

$I_h^c u \in V_h^* \cap C^0(\bar{\Omega})$ is to modify the sequence $u_p^{h_i} \in \mathcal{D}_p^*(\mathcal{T}_h)$ in Lemma 1.2.1, by adding suitable piecewise polynomials on each K_i. For more on the construction of I_h^c, we refer to [8, Theorem 4.1], [9, Theorem 4.6], [1, Theorem 4] and [2, Theorem 3]. Following these constructions, we prove the following lemma.

LEMMA 2.3.3 *Let $\mathcal{T}_h$ be a regular subdivision. Then, for a given $\phi \in H_p^s(\Omega)$, $s \geq 2$, there exists a positive constant C_{A_c} (depending on s but independent of ϕ, p and h) and an $I_h^c\phi \in V_h^* \cap C^0(\bar{\Omega})$ such that for all K_i and e_k:*

(i) $I_h^c\phi|_{\partial\Omega} = \phi|_{\partial\Omega}$.
(ii) for any $0 \leq l \leq s$ *and* $0 \leq l \leq 2$,

$$\|\phi - I_h^c\phi\|_{H^l(K_i)} \leq C_{A_c} \frac{h_i^{\mu-l}}{p^{s-l-\delta_1}} \left(\sum_{K_j \in K_i^*} \|\phi\|^2_{H^s(K_j)} \right)^{1/2},$$

where $\delta_1 = 0$, *if l=0,1 and* $\delta_1 = 1$, *if l=2.*
(iii) for $s > l + \frac{1}{2}$ *and* $l = 0, 1$,

$$|\phi - I_h^c\phi|_{H^l(e_k)} \leq C_{A_c} \frac{h_i^{\mu-l-1/2}}{p^{s-l-1/2-\delta_2}} \left(\sum_{K_j \in K_i^*} \|\phi\|^2_{H^s(K_j)} \right)^{1/2},$$

where $\delta_2 = 0$, *if l=0 and* $\delta_2 = 1/2$, *if l=1.*
(iv) for $0 \leq l \leq s - 1 + 2/r$ *and* $l = 0,\ 1$,

$$\|\phi - I_h^c\phi\|_{W_r^l(K_i)} \leq C_{A_c} \frac{h_i^{\mu-l-1+2/r}}{p^{s-l-1+2/r}} \left(\sum_{K_j \in K_i^*} \|\phi\|^2_{H^s(K_j)} \right)^{1/2},$$

where $\mu = min(s,\ p+1)$, $K_i^* = \{K_j : |\partial K_i \cap \partial K_j| > 0\}$ *and* e_k *is an edge on* ∂K_i.

REMARK 2.3.3 *Note that the assumption* $\mathbf{P}$*(1) implies that the cardinality of* K_i^* *is bounded by* N_κ, *for all i.*

Proof of Lemma 2.3.3. The statement (i) in the lemma is proved in [9, Theorem 4.6]. For $0 \leq l \leq 1$, the approximation property in (ii) is proved in [2, Theorem 3], [9, Theorem 4.6]. Using the inverse inequality (1.21) and Lemma 1.2.1, we present the proof of the property

(ii) for $l = 2$, as follows:

$$\begin{aligned}
\|\phi - I_h^c\phi\|_{H^2(K_i)} &\leq \|\phi - I_h\phi\|_{H^2(K_i)} + \|I_h\phi - I_h^c\phi\|_{H^2(K_i)} \\
&\leq \|\phi - I_h\phi\|_{H^2(K_i)} + C\frac{p^2}{h_i}\|I_h\phi - I_h^c\phi\|_{H^1(K_i)} \\
&\leq \|\phi - I_h\phi\|_{H^2(K_i)} + C\frac{p^2}{h_i}\|\phi - I_h\phi\|_{H^1(K_i)} + \frac{p^2}{h_i}\|\phi - I_h^c\phi\|_{H^1(K_i)} \\
&\leq C\frac{h_i^{\mu-2}}{p^{s-3}}\left(\sum_{K_j\in K^*}\|\phi\|^2_{H^s(K_j)}\right)^{1/2}.
\end{aligned}$$

Then, using the trace inequality (1.14), we deduce the property (iii) of the lemma. Finally, using similar arguments as in [2, Theorem 3], the property (iv) can be easily proved. This completes the rest of the proof. ■

REMARK 2.3.4 *The approximation property (ii) for $l = 2$ and the property (iii) for $l = 1$ of Lemma 2.3.3 are not optimal in terms of p. But as we see in our next analysis, these properties do not affect the accuracy of the approximation u_h.*

LEMMA 2.3.4 *Let $\mathcal{T}_h$ be a regular subdivision and $V_h = V_h^*$. Then, for any $\beta \geq 1$ and given any $\phi \in H_p^s(\Omega)$, $s \geq 2$, there exists a constant C independent of h and p such that*

$$|||\phi - I_h^c\phi|||_+ \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^{2\mu-2}}{p^{2s-2}}\|\phi\|^2_{H^s(K_i)}\right)^{1/2}, \tag{2.41}$$

where $\mu = min\{p+1, s\}$.

Proof. Let $\eta^* = \phi - I_h^c\phi$. Since $I_h^c\phi \in V_h^* \cap C^0(\bar{\Omega})$ and $I_h^c\phi|_{\partial\Omega} = \phi|_{\partial\Omega}$, the jump $[\phi - I_h^c\phi] = 0$ on each $e_k \in \Gamma$. Hence, using (1.10) and Lemma 2.3.3, we obtain

$$\begin{aligned}
|||\phi - I_h^c\phi|||_+^2 &= \sum_{i=1}^{N_h}\int_{K_i}|\nabla\eta^*|^2dx + \sum_{e_k\in\Gamma}\int_{e_k}\frac{|e_k|^\beta}{p^2}\{\frac{\partial\eta*}{\partial\nu}\}^2ds \\
&\leq C\sum_{i=1}^{N_h}\sum_{K_j\in K_i^*}\left(\frac{h_i^{2\mu-2}}{p^{2s-2}}\|\phi\|^2_{H^s(K_j)} + \frac{h_i^\beta}{p^2}\frac{h_i^{2\mu-3}}{p^{2s-4}}\|\phi\|^2_{H^s(K_j)}\right) \\
&\leq C\sum_{i=1}^{N_h}\left(\frac{h_i^{2\mu-2}}{p^{2s-2}}\|\phi\|^2_{H^s(K_i)} + \frac{h_i^{2\mu-2}}{p^{2s-2}}\|\phi\|^2_{H^s(K_i)}\right).
\end{aligned}$$

This completes the rest of the proof. ■

THEOREM 2.3.5 *Let $\mathcal{T}_h$ be a regular subdivision and $V_h = V_h^*$. Suppose that $u \in H_p^s(\Omega)$, $s \geq 2$. Then, for any $\beta \geq 1$ and for sufficiently small h, there exists a constant $C = C(\alpha, M)$ which is independent of h and p such that the solution u_h of the problem (2.4) satisfies*

$$|||u - u_h||| \leq CC_aC_QC_u \left(\sum_{i=1}^{N_h} \frac{h^{2\mu-2}}{p^{2s-2}} \|u\|^2_{H^s(K_i)} \right)^{1/2},$$

where $\mu = min\{p+1,\ s\}$, C_a, C_Q and C_u are as in (2.8), (1.24) and (2.15), respectively.

Proof. Under the hypotheses on the mesh, there is an $I_h^c\phi \in V_h^* \cap C^0(\bar{\Omega})$ such that $I_h^c\phi|_{\partial\Omega} = \phi|_{\partial\Omega}$. Hence, the jump $[\phi - I_h^c\phi] = 0$ on each $e_k \in \Gamma$. Then, using Lemma 2.2.1 and Lemma 2.3.3, it is easy to prove Lemma 2.2.4 for any $\beta \geq 1$. Now, note that estimates (2.25) and (2.27) in the Lemma 2.3.1 depend on the approximation property (ii) for $l = 2$. Though there is a suboptimality in this property, we still obtain the results in Lemma 2.3.1, by replacing I_hu by I_h^cu and for any $\beta \geq 1$. Below, we only indicate the changes to be made in the text of the proof of Lemma 2.3.1. We now consider the term on the left-hand side of (2.25) with $\eta = u - I_h^cu$. Then, using Hölder's inequality, the trace inequality (1.14) and Lemma 2.3.3, we estimate this term as follows:

$$\begin{aligned}
\sum_{e_k \in \Gamma_I} \int_{e_k} |\{\chi \frac{\partial \eta}{\partial \nu}\}[v_h]| ds &\leq C \sum_{e_k \in \Gamma_I} \left(\frac{|e_k|^{\beta/2}}{p} \|\chi\|_{L^4(e_k)} \|\nabla\eta\|_{L^4(e_k)} \left(\int_{e_k} \frac{p^2}{|e_k|^\beta} [v_h]^2 ds \right)^{1/2} \right) \\
&\leq C \sum_{i=1}^{N_h} \sum_{e_k \in \partial K_i} \frac{|e_k|^{\beta/2-1/2}}{p} \left(\|\chi\|^4_{L^4(K_i)} + h_i \|\chi\|^3_{L^6(K_i)} \|\nabla\chi\|_{L^2(K_i)} \right)^{1/4} \\
&\quad \left(\|\nabla\eta\|^4_{L^4(K_i)} + h_i \|\nabla\eta\|^3_{L^6(K_i)} \|\nabla^2\eta\|_{L^2(K_i)} \right)^{1/4} \left(\int_{e_k} \frac{p^2}{|e_k|^\beta} [v_h]^2 \right)^{1/2} \\
&\leq C \sum_{i=1}^{N_h} \sum_{e_k \in \partial K_i} \frac{h^{\beta/2-1/2}}{p} \left(\frac{h^2}{p^2} \|u\|^4_{H^2(K_i)} + h \frac{h}{p} \frac{1}{p^{-1}} \|u\|^4_{H^2(K_i)} \right)^{1/4} \\
&\quad \left(\|\chi\|^4_{L^4(K_i)} + h_i \|\chi\|^3_{L^6(K_i)} \|\nabla\chi\|_{L^2(K_i)} \right)^{1/4} \left(\int_{e_k} \frac{p^2}{|e_k|^\beta} [v_h]^2 ds \right)^{1/2} \\
&\leq C \|u\|_{H^2(\Omega)} \left(\frac{h^{\beta/2}}{p^{3/2}} + \frac{h^{\beta/2}}{p} \right) |||\chi|||\ |||v_h|||. \qquad (2.42)
\end{aligned}$$

Similarly, a replacement of I_hu by I_h^cu in Lemma 2.3.2 and an application of Lemma 2.3.4, yield the proof of Theorem 2.3.2 and Theorem 2.3.3 for any $\beta \geq 1$. Hence, the estimate (2.40) holds for any $\beta \geq 1$ with $\eta = u - I_h^cu$. Then, an application of Lemma 2.3.4 completes the rest of the proof. ■

Now, we proceed to derive the L^2-norm error estimate. Since, the SIPG method is adjoint consistent, one can expect optimal L^2-norm error estimates in terms of h. But for the NIPG method, the bilinear form is not adjoint consistent. In general, it may be difficult to prove the optimal L^2 error estimate in terms of h. However, if $u \in H_p^s(\Omega),\ s \geq 2$, it is possible to obtain optimal L^2-norm error estimates in terms of both h and p by increasing the penalty on the uniform regular subdivision. Assume that the hypotheses of Theorem 2.3.5 hold. Of course, these assumptions are not necessary to derive optimal L^2-norm error estimate in terms of h for the SIPG method. Below, we appeal to the Aubin-Nitsche duality argument to estimate $\|u - u_h\|$.

THEOREM 2.3.6 *Let $a \in C_b^2(\bar{\Omega} \times \mathbb{R})$ and $u \in W_\infty^1(\Omega)$. Suppose that $\beta \geq 3$ when $\theta = -1$, and $\beta \geq 1$ when $\theta = 1$. Further, assume that the hypotheses of Theorem 2.3.5 hold. Then, there exists a constant $C = C(\alpha,\ M)$ such that for small h*

$$\|u - u_h\| \leq CC_Q C_a C_u^2 \frac{h^\mu}{p^s} \|u\|_{\mathbf{s},h},$$

where $\mu = min\{p+1,\ s\}$, $C_a,\ C_Q$ and C_u are as in (2.8), (1.24) and (2.15), respectively.

Proof. Our assumptions on a and u imply that there is a unique solution $\phi \in H^2(\Omega)$ to the following linear elliptic problem

$$\begin{aligned} -\nabla \cdot (a(u)\nabla\phi) + (a_u(u)\nabla u) \cdot \nabla\phi &= e \quad \text{on } \Omega, \\ \phi &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

and ϕ satisfies the following elliptic regularity, see [40, Lemma 9.17]

$$\|\phi\|_{H^2(\Omega)} \leq C\|e\|_{L^2(\Omega)}. \tag{2.43}$$

Note that

$$\|e\|^2 = B(u; e,\ \phi) + \int_\Omega (a_u(u)\nabla u) \cdot e\nabla\phi\ dx + (\theta - 1) \sum_{e_k \in \Gamma} \int_{e_k} \{a(u)\frac{\partial\phi}{\partial\nu}\}[e]\ ds. \tag{2.44}$$

The first term on the right-hand side of (2.44) is rewritten as

$$\begin{aligned} B(u; e,\ \phi) &= B(u; u,\ \phi) - B(u_h; u_h,\ \phi) + B(u_h; u_h,\ \phi) - B(u; u_h,\ \phi) \\ &= (B(u; u,\ \phi - \chi) - B(u_h; u_h,\ \phi - \chi)) + (B(u_h; u_h,\ \phi) - B(u; u_h,\ \phi)) \\ &= \qquad I \qquad + \qquad II, \end{aligned} \tag{2.45}$$

where $\chi = I_h^c \phi$ such that $\chi|_{\partial\Omega} = 0$. For the first term on the right-hand side of (2.45), we note that

$$\begin{aligned} I &= B(u;u,\ \phi-\chi) - B(u_h;u,\ \phi-\chi) + B(u_h;u,\ \phi-\chi) - B(u_h;u_h,\ \phi-\chi) \\ &= \sum_{i=1}^{N_h}\int_{K_i}(a(u)-a(u_h))\nabla u\cdot\nabla(\phi-\chi)dx + \sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla e\cdot\nabla(\phi-\chi)dx \\ &\quad - \sum_{e_k\in\Gamma}\int_{e_k}\{(a(u_h)-a(u))\frac{\partial(\phi-\chi)}{\partial\nu}\}[e]ds + \sum_{i=1}^{N_h}\int_{K_i}a(u)\nabla e\cdot\nabla(\phi-\chi)dx \\ &\quad - \sum_{e_k\in\Gamma}\int_{e_k}\{a(u)\frac{\partial(\phi-\chi)}{\partial\nu}\}[e]ds. \end{aligned} \tag{2.46}$$

Since $u \in W^{1,\infty}(\Omega)$, we use the Cauchy-Schwarz inequality, Lemma 2.3.4, to bound the first and fourth term of on the right-hand side of (2.46) as

$$\begin{aligned} |\sum_{i=1}^{N_h}\int_{K_i}(a(u)-a(u_h))\nabla u\cdot\nabla(\phi-\chi)|dx &\le C_u\|e\|\ \|\phi-\chi\|_{H^1(\Omega)} \\ &\le CC_u\frac{h}{p}\|e\|\ \|\phi\|_{H^2(\Omega)}, \end{aligned} \tag{2.47}$$

and

$$\begin{aligned} |\sum_{i=1}^{N_h}\int_{K_i}a(u)\nabla(u-u_h)\cdot\nabla(\phi-\chi)|dx &\le M|||e|||\ \|\phi-\chi\|_{H^1(\Omega)} \\ &\le C\frac{h}{p}|||e|||\ \|\phi\|_{H^2(\Omega)}. \end{aligned} \tag{2.48}$$

Now, using Hölder's inequality, Lemma 2.3.4 and Lemma 1.2.6, we estimate the second term on the right-hand side of (2.46) as

$$\begin{aligned} |\sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla(u_h-u)\cdot\nabla(\phi-\chi)|dx &\le C\ \|e\|_{L^3(\Omega)}|||e|||\ \|\phi-\chi||_{W_6^1(\Omega)} \\ &\le C\ |||e|||^2\ \|\phi\|_{H^2(\Omega)}. \end{aligned} \tag{2.49}$$

Next, using similar arguments as in (2.42), we bound the third term on the right-hand side

of (2.46) as

$$\begin{aligned} |\sum_{e_k\in\Gamma}\int_{e_k}\{(a(u_h)-a(u))\frac{\partial(\phi-\chi)}{\partial\nu}\}[u-u_h]|ds &\leq CC_a\sum_{e_k\in\Gamma}\frac{|e_k|^{\beta/2}}{p_k}\|e\|_{L^4(e_k)}|\phi-\chi|_{W_4^1(e_k)} \\ &\quad \left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[e]^2ds\right)^{1/2} \\ &\leq CC_a|||e|||^2\ \|\phi\|_{H^2(\Omega)}. \end{aligned} \tag{2.50}$$

Then, using Lemma 2.3.3, the fifth term on the right-hand side of (2.46) is estimated as

$$|\sum_{e_k\in\Gamma}\int_{e_k}\{a(u)\frac{\partial(\phi-\chi)}{\partial\nu}\}[u-u_h]|ds \leq C\frac{h^{\beta/2+1/2}}{p}|||e|||\ \|\phi\|_{H^2(\Omega)}. \tag{2.51}$$

Hence using (2.43), we obtain for any $\beta\geq 1$,

$$|I| \leq CC_uC_a\left(|||e|||^2+\frac{h}{p}|||e||||+\|e\|\right)\|e\|. \tag{2.52}$$

For the second term on the right-hand side of (2.45), that is, II, we note that $[u]=0$ on $e_k\in\Gamma_I$. Thus,

$$\begin{aligned} II &= \sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla u_h\cdot\nabla\phi\ dx-\int_{\Gamma_I}\{((a(u_h)-a(u))\frac{\partial\phi}{\partial\nu}\}[u_h-u]ds \\ &= \sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla(u_h-u)\cdot\nabla\phi\ dx+\sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla u\cdot\nabla\phi\ dx \\ &\quad -\int_{\Gamma_I}\{((a(u_h)-a(u))\frac{\partial\phi}{\partial\nu}\}[u_h-u]\ ds. \end{aligned} \tag{2.53}$$

Use Hölder's inequality, the Sobolev imbedding theorem and Lemma 1.2.6 to estimate the first term on the right-hand side of (2.53) as

$$\begin{aligned} |\sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u))\nabla(u_h-u)\cdot\nabla\phi|\ dx &\leq C_a\|e\|_{L^3(\Omega)}|e|_{1,h}\|\phi||_{W_6^1(\Omega)} \\ &\leq CC_a|||e|||^2\ \|\phi\|_{H^2(\Omega)}. \end{aligned} \tag{2.54}$$

Now for the third term on the right-hand side of (2.53), using Hölder's inequality, the trace

inequality (1.14) and Lemma 1.2.6, we arrive at

$$\begin{aligned}
|\sum_{e_k\Gamma_I}\int_{e_k}\{((a(u_h)-a(u))\frac{\partial\phi}{\partial\nu}\}[u_h-u]|\ ds &\leq CC_a\sum_{e_k\Gamma_I}\int_{e_k}|e\frac{\partial\phi}{\partial\nu}|\ |[e]|\ ds \\
&\leq CC_a\sum_{e_k\Gamma_I}\frac{|e_k|^{\beta/2}}{p_k}\|e\|_{L^4(e_k)}\|\phi\|_{W_4^1(e_k)}\mathcal{J}^{1,\beta}(e,e)^{1/2} \\
&\leq CC_a\frac{h^{\beta/2-1/2}}{p}|||e|||^2\ \|\phi\|_{H^2(\Omega)}. \qquad (2.55)
\end{aligned}$$

We rewrite the second term on the right-hand side of (2.53) together with the second term on the right-hand side of (2.44) as

$$\sum_{i=1}^{N_h}\int_{K_i}(a(u_h)-a(u)+a_u(u)(u-u_h))\nabla u\cdot\nabla\phi\ dx = \sum_{i=1}^{N_h}\int_{K_i}\tilde{a}_{uu}(u_h)(u-u_h)^2\nabla u\cdot\nabla\phi\ dx,$$

and then we use Lemma 1.2.6 to obtain

$$\begin{aligned}
|\sum_{i=1}^{N_h}\int_{K_i}\tilde{a}_{uu}(u_h)(u-u_h)^2\nabla u\cdot\nabla\phi|dx &\leq C_aC_u|\sum_{i=1}^{N_h}\int_{K_i}(u-u_h)^2\cdot\nabla\phi|dx \\
&\leq CC_aC_u|||e|||^2\ \|\phi\|_{H^2(\Omega)}. \qquad (2.56)
\end{aligned}$$

Hence using (2.43), we obtain for any $\beta\geq 1$

$$|II|\leq CC_uC_a|||e|||^2\ \|e\|. \qquad (2.57)$$

For $\theta=1$, the third term on the right-hand side of (2.44) becomes zero. For $\theta=-1$, using the trace inequality (1.14) and (2.43), the third term on the right-hand side of (2.10) is estimated as

$$\begin{aligned}
\sum_{e_k\in\Gamma}\int_{e_k}\{a(u)\frac{\partial\phi}{\partial\nu}[e]\}\ ds &\leq C\sum_{e_k\in\Gamma}\left(\int_{e_k}\frac{|e_k|^\beta}{p_k^2}|\frac{\partial\phi}{\partial\nu}|^2ds\right)^{1/2}\left(\int_{e_k}\frac{p_k^2}{|e_k|^\beta}[e]^2ds\right)^{1/2} \\
&\leq C\frac{h^{\beta/2-1/2}}{p}|||e|||\ \|e\|. \qquad (2.58)
\end{aligned}$$

We combine the estimates (2.52)-(2.58) to obtain

$$\|u-u_h\| \leq C\left(|||u-u_h|||+\frac{h}{p}+|\theta-1|\frac{h^{\beta/2-1/2}}{p}\right)|||u-u_h|||.$$

A use of Theorem 2.3.5 completes the rest of the proof. ■

REMARK 2.3.5 *In the proof of Lemma 2.3.1 and the subsequent results in Section 4, we have assumed that the range of* $\frac{\partial^l a}{\partial u^l}(x,v),\ x \in \bar{\Omega},\ v \in \mathbb{R},\ l = 0,1,2$ *is a compact set, say* $[m, M] \subset \mathbb{R}$. *But, if* $u \in H^{5/2}(\Omega)$, *we note that asymptotically only the values of* $v \in [m_u - \delta^*, M_u + \delta^*] \subset \mathbb{R}$, *where* $0 < \delta^* < 1$, $m_u = inf\{u(x) : x \in \bar{\Omega}\}$ *and* $M_u = sup\{u(x) : x \in \bar{\Omega}\}$ *are considered to derive the proof of Lemma 2.3.1 and the subsequent results. To be more precise, the terms* $\tilde{a}_u(z)$ *and* $\tilde{a}_{uu}(z),\ z \in \mathcal{O}_\delta(I_h u)$ *in (2.16) (see the estimates (2.17)-(2.31)) can be estimated as follows. Since* $z \in \mathcal{O}_\delta(I_h u)$, *where* $\delta = h^{-\epsilon}|||u - I_h u|||_+,\ 0 < \epsilon \leq 1/4$, *using inverse inequality (1.20), Lemma 1.2.6 and Lemma 1.2.1, we find that*

$$\begin{aligned}
\|z-u\|_{L^\infty(\Omega)} &\leq \|z - I_h u\|_{L^\infty(\Omega)} + \|I_h u - u\|_{L^\infty(\Omega)} \\
&\leq C\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)\|z - I_h u\|_{L^2(\Omega)} + \|I_h u - u\|_{L^\infty(\Omega)} \\
&\leq C\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)|||z - I_h u||| + \|I_h u - u\|_{L^\infty(\Omega)} \\
&\leq Ch^{-\epsilon}\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)|||u - I_h u|||_+ + \|I_h u - u\|_{L^\infty(\Omega)} \\
&\leq Ch^{-\epsilon}\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)\left(\sum_{i=1}^{N_h} \frac{h_i^3}{p_i^2}\|u\|^2_{H^{5/2}(K_i)}\right)^{1/2} + C\frac{h}{p}\|u\|_{H^2(\Omega)} \\
&\leq Ch^{-\epsilon}\left(\max_{1\leq i\leq N_h} \frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h} \frac{h_i^{3/2}}{p_i}\right)\|u\|_{H^{5/2}(\Omega)} + C\frac{h}{p}\|u\|_{H^2(\Omega)} \\
&\leq Ch^{1/2-\epsilon}\|u\|_{H^{5/2}(\Omega)}. \qquad (2.59)
\end{aligned}$$

Therefore, for sufficiently small h, $\|z\|_{L^\infty(\Omega)} \leq \delta^* + \|u\|_{L^\infty(\Omega)}$, *where* $0 < \delta^* < 1$. *Now, since the nonlinear functions* a_u *and* a_{uu} *are continuous, they map the compact set* $[m_u - \delta^*, M_u + \delta^*]$ *into a compact set in* $\mathbb{R}$ *and hence, the results in Lemma 2.3.1 and the subsequent results in Section 4 remain valid when* $a(v),\ a_u(v)$ *and* $a_{uu}(v)$ *are bounded for bounded* u. *Finally, we remark that when* $u \in H^2(\Omega)$, *it may be possible to show the boundedness of* $a(v)$ *and its derivatives for* $v \in [m_u - \delta^*, M_u + \delta^*] \subset \mathbb{R}$ *by using better inverse inequalities, say in the first line of (2.59), apply* $\|z - I_h u\|_{L^\infty(K_i)} \leq Cp_i^{1/2}h_i^{-1/4}\|z - I_h u\|_{L^8(K_i)}$ *(see [52, p. 916]) and use the Poincaré inequality in Lemma 1.2.6 to complete the estimate (2.59).*

2.4 Numerical Experiments

In this section, we discuss the performance of the proposed NIPG and SIPG methods for the numerical approximation of the quasilinear elliptic problem (2.1)-(2.2). For this, we consider the following nonlinear elliptic problem:

$$\begin{aligned} -\nabla \cdot ((1+u)\nabla u) &= f, \quad \text{in} \quad \Omega, \\ u &= 0, \quad \text{on} \quad \partial\Omega, \end{aligned}$$

where $\Omega = (0,1) \times (0,1)$ and f is taken in such a way that the exact solution is $u = x(1-x)y(1-y)$. We divide Ω into regular uniform triangles. We denote the total number of triangles by N_h. The stabilizing parameter σ_k, appearing in the penalty term $\mathcal{J}^{\sigma,\beta}$ is taken as follows: $\sigma_k = 10 \ \forall e_k$. We investigate the convergence of NIPG($\theta = -1$) and SIPG($\theta = 1$) on a sequence of uniform triangular meshes for each of degree of approximation $p = 1, 2$ and 3 ($p_i = p, \ 1 \leq i \leq N_h$). Similarly, we also investigate the convergence of both methods by enriching the polynomial degree p on a fixed mesh.

Since the discrete space V_h can have piecewise polynomials which may be discontinuous across the edges of elements, we choose basis functions as follows:

For piecewise linear, that is, $p = 1$ and $1 \leq i \leq N_h$,

$$\Phi_{(i-1)\times 3+1} = \begin{cases} 1 & \text{on } K_i, \\ 0 & \text{elsewhere}, \end{cases} \tag{2.60}$$

$$\Phi_{(i-1)\times 3+j} = \begin{cases} \lambda_j & \text{on } K_i, \ j = 2, \ 3, \\ 0 & \text{elsewhere}. \end{cases} \tag{2.61}$$

For piecewise quadratic, that is, $p = 2$ and $1 \leq i \leq N_h$,

$$\Phi_{(i-1)\times 6+1} = \begin{cases} 1 & \text{on } K_i, \\ 0 & \text{elsewhere}, \end{cases} \tag{2.62}$$

$$\Phi_{(i-1)\times 6+j} = \begin{cases} \lambda_j & \text{on } K_i, \ j = 2, \ 3, \\ 0 & \text{elsewhere}, \end{cases} \tag{2.63}$$

$$\Phi_{(i-1)\times 6+4} = \begin{cases} \lambda_2^2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.64}$$

$$\Phi_{(i-1)\times 6+5} = \begin{cases} \lambda_3^2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.65}$$

$$\Phi_{(i-1)\times 6+6} = \begin{cases} \lambda_2\lambda_3 & \text{on } K_i, \\ 0 & \text{elsewhere.} \end{cases} \tag{2.66}$$

For picewise cubic, that is, $p = 3$ and $1 \le i \le N_h$,

$$\Phi_{(i-1)\times 10+1} = \begin{cases} 1 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.67}$$

$$\Phi_{(i-1)\times 10+j} = \begin{cases} \lambda_j & \text{on } K_i,\ j = 2,\ 3, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.68}$$

$$\Phi_{(i-1)\times 10+4} = \begin{cases} \lambda_2^2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.69}$$

$$\Phi_{(i-1)\times 10+5} = \begin{cases} \lambda_3^2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.70}$$

$$\Phi_{(i-1)\times 10+6} = \begin{cases} \lambda_2\lambda_3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.71}$$

$$\Phi_{(i-1)\times 10+7} = \begin{cases} \lambda_2^3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.72}$$

$$\Phi_{(i-1)\times 10+8} = \begin{cases} \lambda_3^3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.73}$$

$$\Phi_{(i-1)\times 10+9} = \begin{cases} \lambda_2^2\lambda_3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.74}$$

$$\Phi_{(i-1)\times 10+10} = \begin{cases} \lambda_2\lambda_3^2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases} \tag{2.75}$$

where λ_2 and λ_3 are barycentric coordinates of K_i. We note that each of the basis functions takes support only on the corresponding finite element K_i. Let N denote the dimension of V_h. Denote the basis of V_h by $\{\Phi_i : 1 \le i \le N\}$, where $N = N_h * \frac{(p+1)(p+2)}{2}$. The discrete solution u_h is written as

$$u_h = \sum_{i=1}^{N} \alpha_i \Phi_i. \tag{2.76}$$

In order to derive the nonlinear algebraic system corresponding to (2.4), we set $v_h = \Phi_j$ in (2.4) and obtain

$$F_j(\alpha) = B(u_h; u_h, \Phi_j) - L(\Phi_j) = 0, \quad 1 \le j \le N,$$

where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]$,

$$\begin{aligned} B(u_h; u_h, \Phi_j) &= \sum_{i=1}^{N_h} \int_{K_i} (1+u_h)\nabla u_h \cdot \nabla \Phi_j \, dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{(1+u_h)\nabla u_h \cdot \nu\}[\Phi_j] \, ds \\ &+ \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{(1+u_h)\nabla \Phi_j \cdot \nu\}[u_h] \, ds - \sum_{e_k \in \Gamma_\partial} \int_{e_k} \nabla u_h \cdot \nu \Phi_j \, ds \\ &+ \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \nabla \Phi_j \cdot \nu u_h \, ds + \mathcal{J}^{\sigma,\beta}(u_h, \Phi_j) \\ &= I_1 - I_2 + \theta I_3 - I_4 + \theta I_5 + I_6 \end{aligned} \tag{2.77}$$

and $L(\Phi_j) = (f, \Phi_j)$. The resulting nonlinear system is then denoted by

$$\mathbf{F}(\alpha) = [F_1(\alpha), F_2(\alpha), \cdots, F_N(\alpha)]^T = [0, 0, \cdots, 0]^T. \tag{2.78}$$

We then apply the Newton's method to find the solution α to (2.78). The Jacobian matrix of the system $\mathbf{F}(\alpha)$ is computed as follows :

$$\mathbf{J} = \left[\frac{\partial F_j}{\partial \alpha_m}\right]_{1 \le j,m \le N} = \left[\frac{\partial B(u_h, \Phi_j)}{\partial \alpha_m}\right]_{1 \le j,m \le N}. \tag{2.79}$$

We substitute (2.76) in (2.79) and then using (2.77), we first compute

$$\begin{aligned}\frac{\partial I_1}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m}\left(\sum_{i=1}^{N_h}\int_{K_i}\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\left(1+\sum_{l=1}^{N}\alpha_l\nabla\Phi_l\right)\cdot\nabla\Phi_j\ dx\right)\\ &= \sum_{i=1}^{N_h}\int_{K_i}\Phi_m\left(1+\sum_{l=1}^{N}\alpha_l\nabla\Phi_l\right)\cdot\nabla\Phi_j dx+\sum_{i=1}^{N_h}\int_{K_i}\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\nabla\Phi_m\cdot\nabla\Phi_j dx.\end{aligned}$$

Next, we note that

$$\begin{aligned}\frac{\partial I_2}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\left(1+\sum_{l=1}^{N}\alpha_l\nabla\Phi_l\right)\cdot\nu\}[\Phi_j]\ ds\right)\\ &= \frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\Phi_m\left(1+\sum_{l=1}^{N}\alpha_l\nabla\Phi_l\right)\cdot\nu\}[\Phi_j]\ ds\right)\\ &\quad+\frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\nabla\Phi_m\cdot\nu\}[\Phi_j]\ ds\right).\end{aligned}$$

Similarly,

$$\begin{aligned}\frac{\partial I_3}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\nabla\Phi_j\cdot\nu\}\left(1+\sum_{l=1}^{N}\alpha_l[\Phi_l]\right)\ ds\right)\\ &= \frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\Phi_m\nabla\Phi_j\cdot\nu\}\left(1+\sum_{l=1}^{N}\alpha_l[\Phi_l]\right)\ ds\right)\\ &\quad+\frac{\partial}{\partial \alpha_m}\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(1+\sum_{l=1}^{N}\alpha_l\Phi_l\right)\nabla\Phi_j\cdot\nu\}[\Phi_m]\ ds\right).\end{aligned}$$

It is easy to see that

$$\frac{\partial I_6}{\partial \alpha_m}=\frac{\partial}{\partial \alpha_m}\sum_{e_k\in\Gamma}\int_{e_k}\sigma_k\frac{p_k^2}{|e_k|^\beta}\left(\sum_{l=1}^{N}\alpha_l[\Phi_l]\right)[\Phi_j]\ ds=\sum_{e_k\in\Gamma}\int_{e_k}\sigma_k\frac{p_k^2}{|e_k|^\beta}[\Phi_m][\Phi_j]ds.$$

For the boundary integrals I_4 and I_5, we find the derivative as :

$$\frac{\partial I_4}{\partial \alpha_m}=\frac{\partial}{\partial \alpha_m}\sum_{e_k\in\Gamma_\partial}\int_{e_k}\left(\sum_{l=1}^{N}\alpha_l\nabla\Phi_l\right)\cdot\nu\Phi_j\ ds=\sum_{e_k\in\Gamma_\partial}\int_{e_k}\nabla\Phi_m\cdot\nu\Phi_j ds,$$

and

$$\frac{\partial I_5}{\partial \alpha_m} = \frac{\partial}{\partial \alpha_m} \sum_{e_k \in \Gamma_\partial} \int_{e_k} \nabla \Phi_j \cdot \nu \left(\sum_{l=1}^{N} \alpha_l \Phi_l \right) \, ds = \sum_{e_k \in \Gamma_\partial} \int_{e_k} \nabla \Phi_j \cdot \nu \Phi_m ds.$$

Now, we use the following algorithm for Newton's method to solve the system (2.78). For given α^0, find α^k, for $1 \leq k \leq k_{max}$, such that

$$\alpha^k = \alpha^{k-1} - \mathbf{J}^{-1}\mathbf{F}(\alpha^{k-1}),$$

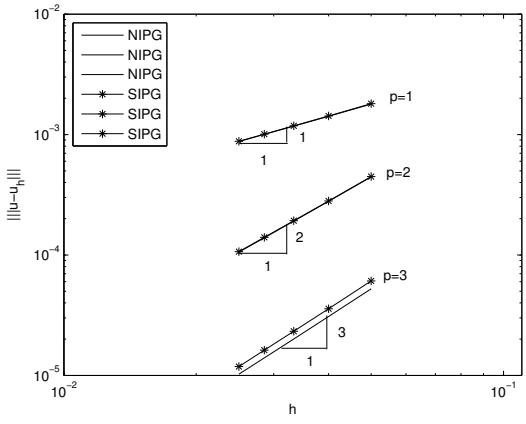
where $\mathbf{J}$ is given by (2.79). The intial iterate α^0 is chosen as the solution of the following linearized problem: Find $u_h^0 \in V_h$ such that

$$B(0; u_h^0, v_h) = L(v_h) \quad \forall \; v_h \in V_h.$$

The maximum number of iterations k_{max} is set to be $k_{max} = 10$.

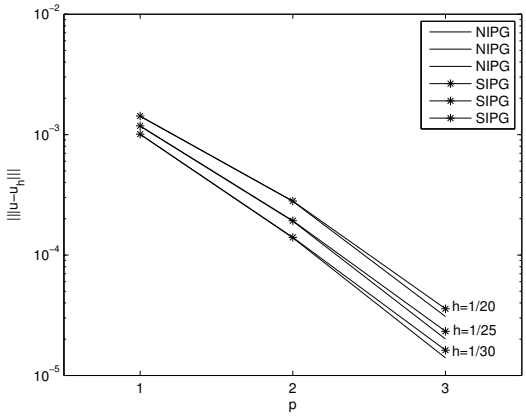
Convergence in the broken H^1-norm: We set $\beta = 1$ for both NIPG and SIPG methods. In Figure 5.1, we plot the broken H^1-norm of the error against the mesh function h for polynomial degrees $p = 1, 2$ and 3. Here, we observe that for each p, $|||u - u_h|||$ converges to zero at the rate $\mathcal{O}(h^p)$ as the mesh is refined. These experiments illustrate the theoretical results obtained in the Theorem 2.3.4. In Figure 5.2, we present the convergence of the broken H^1-norm of the error as the degree of the polynomials increases on a fixed mesh.

Figure 2.1: convergence of NIPG and SIPG with h-refinement



Convergence in the L^2-norm: According to Theorem 2.3.6, the NIPG method gives optimal L^2 order of convergence provided the jump term is super-penalized. We take

Figure 2.2: convergence of NIPG and SIPG with p-refinement



$\beta = 3$, when $\theta = -1$. Since the SIPG method is optimal in the L^2-norm, we take $\beta = 1$, when $\theta = 1$. We investigate the theoretical results obtained in Theorem 2.3.6 by performing the experiments with the above values of β. In Figure 5.3, we plot the L^2-norm of the error against the mesh function h for polynomial degrees $p = 1, 2$ and 3. We note that for each p, $||u - u_h||$ converges to zero at the rate $\mathcal{O}(h^{p+1})$ as the mesh is refined. The convergence lines are almost same for both NIPG and SIPG methods. These results show that the NIPG method exhibits an optimal order of convergence in the L^2-norm on a regular mesh, by imposing the super-penalty. In Figure 5.4, we also plot the L^2-norm of the error against

Figure 2.3: convergence of NIPG and SIPG with h-refinement

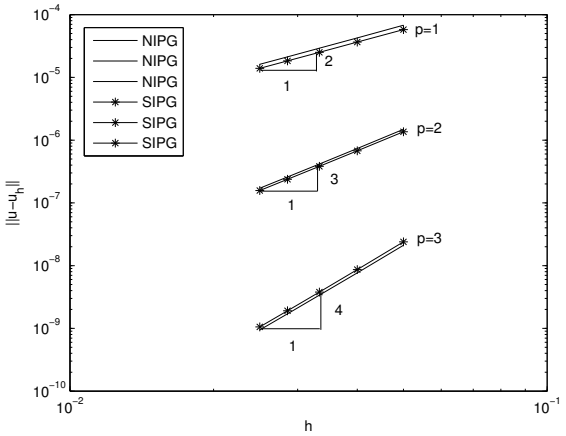
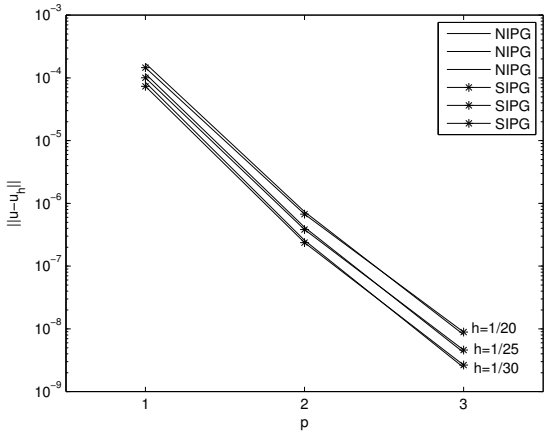


Figure 2.4: convergence of NIPG and SIPG with p-refinement



the degree of the polynomial p on a fixed mesh. The L^2-norm of the error converges exponentially to zero as p increases. These experiments illustrate the theoretical results obtained in the Theorem 2.3.6.

Chapter 3

LDG Method for Quasilinear Elliptic Problems of Non-monotone-Type

3.1 Introduction

In [17], the LDG method is applied to a strongly nonlinear elliptic problem of the following type:

$$-\nabla \cdot \mathbf{a}(\cdot, \nabla u) = f \quad \text{in } \Omega$$

with mixed boundary conditions. Under the assumption that the nonlinear operator induced by $\mathbf{a}$ is strongly monotone, it is shown that the primal form of the LDG method is monotone. Then, existence of an approximate solution to the LDG method is shown and *a priori* error estimates of only h-version are derived. But for nonmonotone nonlinear elliptic problem of following type:

$$-\nabla \cdot (a(u)\nabla u) = f \quad \text{in } \Omega, \tag{3.1}$$

$$u = g \quad \text{on } \partial\Omega, \tag{3.2}$$

it is difficult to extend the analysis of [17]. Therefore, an attempt has been made in this Chapter to study the LDG method for the problem (3.1)-(3.2). We assume that Ω is a bounded domain in $\mathbb{R}^2$ with boundary $\partial\Omega$, and there exist positive constants α, M such that $0 < \alpha \leq a(x,u) \leq M$, $a(\cdot,\cdot)$ is a twice continuously differentiable function in $\bar{\Omega} \times \mathbb{R}$

and all the derivatives of $a(\cdot,\cdot)$ through second order are bounded in $\bar{\Omega} \times \mathbb{R}$. Further, assume that $f \in L^2(\Omega)$, g can be extended to Ω so that the extended g is in $H^2(\Omega)$ and there exists a unique weak solution u of (3.1) -(3.2) such that $u \in H^2(\Omega) \cap W^{1,\infty}(\Omega)$.

In this Chapter, an *hp*-LDG method is applied to the problem (3.1) -(3.2) and error estimates which are optimal in h (mesh size) and slightly suboptimal in p (degree of approximation) are derived. The results proved in this Chapter are same as in the linear case, see [57]. Assuming *hp*-quasiuniformity condition on the mesh, existence of a solution to the discrete problem is proved using Brouwer fixed point theorem for small h. Moreover, the Lipschitz continuity of the discrete solution map shows the uniqueness of the discrete problem.

The rest of the Chapter is organized as follows. Section 3.2 is devoted to the LDG method, and *a priori* error estimates for the method. In Section 3.3, we provide numerical experiments to illustrate the theoretical results for two different nonlinear elliptic problems.

3.2 Local Discontinuous Galerkin (LDG) method

The LDG methods were originally initiated for the system of first order hyperbolic problems. To define the method, we rewrite the equation (3.1) as a problem of first order system of equations. We introduce auxiliary variable $\mathbf{q} = \nabla u$ and $\boldsymbol{\sigma} = a(u)\mathbf{q}$ and rewrite (3.1) -(3.2) as

$$\begin{aligned} \mathbf{q} &= \nabla u \ \text{ in } \ \Omega, && (3.1)\\ \boldsymbol{\sigma} &= a(u)\mathbf{q} \ \text{ in } \ \Omega, && (3.2)\\ -\nabla\cdot\boldsymbol{\sigma} &= f \ \text{ in } \ \Omega, && (3.3)\\ u &= g \ \text{ on } \ \partial\Omega. && (3.4) \end{aligned}$$

We multiply the equation (3.1) by $\mathbf{w} \in \mathbf{W}$, the equation (3.2) by $\boldsymbol{\tau} \in \mathbf{W}$ and the equation (3.3) by $v \in V$ and integrate over the element $K \in \mathcal{T}_h$. Then using the integration by parts formula, we obtain

$$\int_K \mathbf{q}\cdot\mathbf{w}dx + \int_K u\nabla\cdot\mathbf{w}dx - \int_{\partial K} u\mathbf{w}\cdot\nu_K ds = 0, \quad \mathbf{w}\in\mathbf{W}, \tag{3.5}$$

$$\int_K a(u)\mathbf{q}\cdot\boldsymbol{\tau}dx - \int_K \boldsymbol{\sigma}\cdot\boldsymbol{\tau}dx = 0, \quad \boldsymbol{\tau}\in\mathbf{W}, \tag{3.6}$$

and

$$\int_K \boldsymbol{\sigma}\cdot\nabla v dx - \int_{\partial K}\boldsymbol{\sigma}\cdot\nu_K v ds = \int_K f v dx, \quad v\in V. \tag{3.7}$$

Note that there may be difficulty in defining u and $\mathbf{q}$ on ∂K. Therefore, this is just an initial formulation which is helpful in defining the following approximate method given below. The approximate solution $(u_h,\ \mathbf{q}_h,\ \boldsymbol{\sigma}_h)\in \mathrm{Z}_p(K)\times\mathrm{Z}_p(K)^2\times\mathrm{Z}_p(K)^2$ is defined using above weak formulation, that is by imposing that for all K,

$$\int_K \mathbf{q}_h\cdot\mathbf{w}_h dx + \int_K u_h\nabla\cdot\mathbf{w}_h dx - \int_{\partial K}\hat{u}\mathbf{w}_h\cdot\nu_K ds = 0, \quad \mathbf{w}_h\in\mathrm{Z}_p(K)^2, \tag{3.8}$$

$$\int_K a(u_h)\mathbf{q}_h\cdot\boldsymbol{\tau}_h - \int_K \boldsymbol{\sigma}_h\cdot\boldsymbol{\tau}_h dx = 0, \quad \boldsymbol{\tau}_h\in\mathrm{Z}_p(K)^2, \tag{3.9}$$

and

$$\int_K \boldsymbol{\sigma}_h\cdot\nabla v_h dx - \int_{\partial K}\hat{\boldsymbol{\sigma}}\cdot\nu_K v_h ds = \int_K f v_h dx, \quad v_h\in\mathrm{Z}_p(K), \tag{3.10}$$

where the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ have to be suitably chosen in order to ensure the stability of the method and also to improve the order of convergence. The following choice of numerical fluxes are used in solving the linear elliptic problems. If $e_k\in\Gamma_I$, then the numerical fluxes are defined on e_k as :

$$\hat{u}(u_h) = \{u_h\} + C_{12}\cdot[\![u_h]\!], \tag{3.11}$$

$$\hat{\boldsymbol{\sigma}}(u_h,\ \boldsymbol{\sigma}_h) = \{\boldsymbol{\sigma}_h\} - C_{11}[\![u_h]\!] - C_{12}[\![\boldsymbol{\sigma}_h]\!], \tag{3.12}$$

and if $e_k\in\Gamma_\partial$, then the numerical fluxes are taken as :

$$\hat{u} = g, \tag{3.13}$$

$$\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_h - C_{11}(u_h - g)\nu, \tag{3.14}$$

where $C_{11}|_{e_k} = \beta p_k^2/h_k$, $\beta>0$ and $C_{12}\in\mathbb{R}^2$ on $e_k\in\Gamma_I$, We set $C_{12}=0$ on $e_k\in\Gamma_\partial$. The numerical fluxes are *conservative* since they are single valued on $e_k\in\Gamma_I$, that is, on $e_k\in\Gamma_I$,

$$[\![\hat{u}]\!] = 0, \qquad [\![\hat{\boldsymbol{\sigma}}]\!] = 0. \tag{3.15}$$

and *consistent* since the following holds for smooth u and $\mathbf{q}$:

$$\hat{u}(u) = u, \tag{3.16}$$

$$\hat{\boldsymbol{\sigma}}(u,\ \boldsymbol{\sigma}) = \boldsymbol{\sigma}. \tag{3.17}$$

We sum (3.8)-(3.10) over all elements $K \in \mathcal{T}_h$. Then using the conservative property (3.15) and the definition of numerical fluxes, we obtain the following equations :

$$\int_\Omega \mathbf{q}_h \cdot \mathbf{w}_h dx + \sum_{i=1}^{N_h} \int_{K_i} u_h \nabla \cdot \mathbf{w}_h dx - \int_{\Gamma_I} (\{u_h\} + C_{12}.[u_h])[\![\mathbf{w}_h]\!] ds \tag{3.18}$$
$$= \int_{\Gamma_\partial} g \mathbf{w}_h \cdot \nu ds,\ \mathbf{w}_h \in \mathbf{W}_h,$$

$$\sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\sigma}_h \cdot \nabla v_h dx - \int_\Gamma (\{\boldsymbol{\sigma}_h\} - C_{11}[\![u_h]\!] - C_{12}[\![\boldsymbol{\sigma}_h]\!])[\![v_h]\!] ds \tag{3.19}$$
$$= \int_\Omega f v_h dx + \int_{\Gamma_\partial} C_{11} g v_h ds, \quad v_h \in V_h,$$

$$\int_\Omega a(u_h) \mathbf{q}_h \cdot \boldsymbol{\tau}_h dx - \int_\Omega \boldsymbol{\sigma}_h \cdot \boldsymbol{\tau}_h dx = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \tag{3.20}$$

Let $z \in L^2(\Omega)$ and $(\phi,\ \mathbf{p}),\ (v,\ \mathbf{w}) \in V \times \mathbf{W}$. We define the following bilinear functional $A_1 : \mathbf{W} \times \mathbf{W} \to \mathbb{R}$ as

$$A_1(\mathbf{p}, \mathbf{w}) = \int_\Omega \mathbf{p} \cdot \mathbf{w}\ dx,$$

$A_2 : \mathbf{W} \times V \to \mathbb{R}$ as

$$\begin{aligned} A_2(\mathbf{p}; v) &= \sum_{i=1}^{N_h} \int_{K_i} \mathbf{p} \cdot \nabla v\ dx - \int_\Gamma (\{\mathbf{p}\} - C_{12}[\![\mathbf{p}]\!])[\![v]\!]\ ds \\ &= -\sum_{i=1}^{N_h} \int_{K_i} v \nabla \cdot \mathbf{p}\ dx + \int_{\Gamma_I} (\{v\} + C_{12} \cdot [\![v]\!])[\![\mathbf{p}]\!]\ ds, \end{aligned}$$

$J : V \times V \to \mathbb{R}$ as

$$J(\phi, v) = \int_\Gamma C_{11}[\![\phi]\!][\![v]\!]\ ds,$$

and $B : \mathbf{W} \times \mathbf{W} \to \mathbb{R}$ as

$$B(z; \mathbf{p}, \mathbf{w}) = \int_\Omega a(z)\mathbf{p} \cdot \mathbf{w}\ dx.$$

We also define the linear functionals $L_1 : \mathbf{W} \to \mathbb{R}$ and $L_2 : V \to \mathbb{R}$ as

$$L_1(\mathbf{w}) = \int_{\Gamma_\partial} g\mathbf{w}.\nu \; ds \qquad \text{and} \qquad L_2(v) = \int_\Omega fv \; dx + \int_{\Gamma_\partial} C_{11} gv \; ds.$$

Using the above definitions, we write the LDG method for the problem (3.1)-(3.2) in compact form as : Find $(u_h,\ \mathbf{q}_h,\ \boldsymbol{\sigma}_h) \in V_h \times \mathbf{W}_h \times \mathbf{W}_h$ such that

$$A_1(\mathbf{q}_h, \mathbf{w}_h) - A_2(\mathbf{w}_h, u_h) = L_1(\mathbf{w}_h), \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{3.21}$$

$$A_2(\boldsymbol{\sigma}_h, v_h) + J(u_h, v_h) = L_2(v_h), \quad v_h \in V_h, \tag{3.22}$$

$$B(u_h; \mathbf{q}_h, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \tag{3.23}$$

Since the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are consistent, we note that the following hold :

$$A_1(\mathbf{q}, \mathbf{w}) - A_2(\mathbf{w}, u) = L_1(\mathbf{w}), \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{3.24}$$

$$A_2(\boldsymbol{\sigma}, v) + J(u, v) = L_2(v), \quad v_h \in V_h, \tag{3.25}$$

$$B(u; \mathbf{q}, \boldsymbol{\tau}) - A_1(\boldsymbol{\sigma}, \boldsymbol{\tau}) = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \tag{3.26}$$

In order to derive the *a priori* error estimates and to prove existence of a unique approximate solution to the problem (3.21)-(3.23), we proceed as follows: Using the equations (3.21)-(3.26), we obtain

$$A_1(\mathbf{q} - \mathbf{q}_h, \mathbf{w}_h) - A_2(\mathbf{w}_h, u - u_h) = 0, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{3.27}$$

$$A_2(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, v_h) + J(u - u_h, v_h) = 0, \quad v_h \in V_h, \tag{3.28}$$

$$B(u; \mathbf{q}, \boldsymbol{\tau}_h) - B(u_h; \mathbf{q}_h, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \tag{3.29}$$

Adding and subtracting $B(u; \mathbf{q}_h, \boldsymbol{\tau}_h)$, we rewrite (3.29) as

$$B(u; \mathbf{q} - \mathbf{q}_h, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = \int_\Omega (a(u_h) - a(u))\mathbf{q}_h \cdot \boldsymbol{\tau}_h dx,$$

and using (2.7), we now arrive at

$$B(u; \mathbf{q} - \mathbf{q}_h, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) + \int_\Omega (a_u(u)(u - u_h))\mathbf{q} \cdot \boldsymbol{\tau}_h dx$$
$$= \int_\Omega (a(u_h) - a(u))(\mathbf{q}_h - \mathbf{q}) \cdot \boldsymbol{\tau}_h dx + \int_\Omega (a(u_h) - a(u) - a_u(u)(u_h - u))\mathbf{q} \cdot \boldsymbol{\tau}_h dx.$$

For notational simplicity, we introduce for $\boldsymbol{\tau},\ \mathbf{p},\ \mathbf{q} \in \mathbf{W}$ and $\phi,\ v \in V$

$$\begin{aligned} N(u, \mathbf{q}; \phi, \boldsymbol{\tau}) &= \int_\Omega (a_u(u)\mathbf{q})\phi \cdot \boldsymbol{\tau}\ dx, \\ N_1(v-u; \mathbf{p}-\mathbf{q}, \boldsymbol{\tau}) &= \int_\Omega (a(v)-a(u))(\mathbf{p}-\mathbf{q}) \cdot \boldsymbol{\tau} dx \\ &= \int_\Omega \tilde{a}_u(v)(v-u)(\mathbf{p}-\mathbf{q}) \cdot \boldsymbol{\tau} dx, \end{aligned}$$

and

$$\begin{aligned} N_2(v-u; \mathbf{q}, \boldsymbol{\tau}) &= \int_\Omega (a(v)-a(u)-a_u(u)(v-u))\mathbf{q} \cdot \boldsymbol{\tau} dx \\ &= \int_\Omega \tilde{a}_{uu}(v)(v-u)^2\mathbf{q} \cdot \boldsymbol{\tau} dx. \end{aligned}$$

Hence, the equations (3.27)-(3.29) take the form

$$A_1(\mathbf{q}-\mathbf{q}_h, \mathbf{w}_h) - A_2(\mathbf{w}_h, u-u_h) = 0, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{3.30}$$

$$A_2(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, v_h) + J(u-u_h, v_h) = 0, \quad v_h \in V_h, \tag{3.31}$$

$$\begin{aligned} B(u, \mathbf{q}-\mathbf{q}_h, \boldsymbol{\tau}_h) + N(u, \mathbf{q}; u-u_h, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = \\ N_1(u_h-u; \mathbf{q}_h-\mathbf{q}, \boldsymbol{\tau}_h) + N_2(u_h-u; \mathbf{q}, \boldsymbol{\tau}_h), \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \end{aligned} \tag{3.32}$$

We state the following lemma without proof. The proof follows by an appeal to Cauchy-Schwarz inequality and using the assumption on u and $a(u)$.

LEMMA 3.2.1 *There exists positive constants C_1 and C_2 such that for all $(v,\ \mathbf{w}) \in V \times \mathbf{W}$,*

$$B(u; \mathbf{w}, \mathbf{w}) + N(u, \mathbf{q}; v, \mathbf{w}) + J(v, v) \geq C_1 \left(\|\mathbf{w}\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} C_{11}[v]^2 ds \right) - C_2 \|v\|^2.$$

3.2.1 Existence and Uniqueness of the Discrete Problem.

For a given $z \in V_h$, we define a map $S_h : V_h \to V_h$ by $S_h(z) = y \in V_h$ and $\mathbf{q}_z,\ \boldsymbol{\sigma}_z \in \mathbf{W}_h$ satisfying

$$A_1(\mathbf{q}-\mathbf{q}_z, \mathbf{w}_h) - A_2(\mathbf{w}_h, u-y) = 0, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{3.33}$$

$$A_2(\boldsymbol{\sigma}-\boldsymbol{\sigma}_z, v_h) + J(u-y, v_h) = 0, \quad v_h \in V_h, \tag{3.34}$$

$$\begin{aligned} B(u; \mathbf{q}-\mathbf{q}_z, \boldsymbol{\tau}_h) + N(u, \mathbf{q}; u-y, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\sigma}-\boldsymbol{\sigma}_z, \boldsymbol{\tau}_h) = \\ N_1(z-u; \mathbf{q}_z-\mathbf{q}, \boldsymbol{\tau}_h) + N_2(z-u; \mathbf{q}, \boldsymbol{\tau}_h), \quad \boldsymbol{\tau}_h \in \mathbf{W}_h. \end{aligned} \tag{3.35}$$

We write $e_y = u - y = \xi_y - \eta_u$ where $\xi_y = I_h u - y$ and $\eta_u = I_h u - u$. Similarly, $\mathbf{e}_q = \mathbf{q} - \mathbf{q}_z = \boldsymbol{\xi}_q - \boldsymbol{\eta}_q$ and $\mathbf{e}_\sigma = \boldsymbol{\sigma} - \boldsymbol{\sigma}_z = \boldsymbol{\xi}_\sigma - \boldsymbol{\eta}_\sigma$, where $\boldsymbol{\xi}_q = I_h \mathbf{q} - \mathbf{q}_z$, $\boldsymbol{\eta}_q = I_h \mathbf{q} - \mathbf{q}$, $\boldsymbol{\xi}_\sigma = \Pi\boldsymbol{\sigma} - \boldsymbol{\sigma}_z$ and $\boldsymbol{\eta}_\sigma = \Pi\boldsymbol{\sigma} - \boldsymbol{\sigma}$. With these notations, we rewrite (3.33)-(3.35) as

$$A_1(\boldsymbol{\xi}_q, \mathbf{w}_h) - A_2(\mathbf{w}_h, \xi_y) = A_1(\boldsymbol{\eta}_q, \mathbf{w}_h) - A_2(\mathbf{w}_h, \eta_u), \quad \mathbf{w} \in \mathbf{W}, \tag{3.36}$$

$$A_2(\boldsymbol{\xi}_\sigma, v_h) + J(\xi_y, v_h) = A_2(\boldsymbol{\eta}_\sigma, v_h) + J(\eta_u, v_h), \quad v_h \in V_h, \tag{3.37}$$

$$B(u; \boldsymbol{\xi}_q, \boldsymbol{\tau}_h) + N(u, \mathbf{q}; \xi_y, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\xi}_\sigma, \boldsymbol{\tau}_h) = B(u; \boldsymbol{\eta}_q, \boldsymbol{\tau}_h) \tag{3.38}$$
$$+ N(u, \mathbf{q}; \eta_u, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\tau}_h) + N_1(z-u; \mathbf{q}_z - \mathbf{q}, \boldsymbol{\tau}_h) + N_2(z-u; \mathbf{q}, \boldsymbol{\tau}_h), \quad \boldsymbol{\tau}_h \in \mathbf{W}_h.$$

First we show that S_h maps from a ball $O_\delta(I_h u)$ to itself, where

$$O_\delta(I_h u) = \{z \in V_h : |||z - I_h u||| \leq \delta\}$$

and

$$\delta = \frac{1}{h^\epsilon} \left(|||\eta_u|||^2 + \|\boldsymbol{\eta}_q\|^2 + \|\boldsymbol{\eta}_\sigma\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} \frac{h_k}{p_k^2} \{|\boldsymbol{\eta}_\sigma|\}^2 ds \right)^{1/2}.$$

REMARK 3.2.1 *In our subsequent analysis, it is enough to assume that $a(\cdot)$ is bounded in a ball around u. This fact follows from Remark 2.3.5 in Chapter 1.*

The following lemma will prove to be useful in our error analysis. The proof is an easy consequence of Lemma 1.2.1 and Lemma 1.2.5.

LEMMA 3.2.2 *There is a constant C which is independent of h and p such that*

$$\begin{aligned} \left(|||\eta_u|||^2 + \|\boldsymbol{\eta}_q\|^2 + \|\boldsymbol{\eta}_\sigma\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} \frac{h_k}{p_k^2} \{|\boldsymbol{\eta}_\sigma|\}^2 ds \right) &\leq C \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i^+}}{p_i^{2s_i}} \|\nabla u\|_{s_i}^2 \right) \\ &+ C \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i^*}}{p_i^{2s_i - 1}} \|u\|_{s_i+1}^2 \right), \end{aligned}$$

where $\mu_i^+ = min\{s_i, p_i + 1\}$ and $\mu_i^* = min\{s_i, p_i\}$.

Since $u \in H^2(\Omega)$, using Lemma 3.2.2, it is easy to see that

$$\delta \leq C \left(\|u\|_{H^2(\Omega)} \right) \frac{1}{h^\epsilon} \left(\max_{1 \leq i \leq N_h} h_i / p_i^{1/2} \right). \tag{3.39}$$

LEMMA 3.2.3 *Let the assumption (**Q**) holds. For any $0 < \epsilon < 1/2$, there exists a constant C such that*

$$|N_1(z-u;\mathbf{e}_q,\boldsymbol{\tau}) + N_2(z-u;\mathbf{q},\tau)| \le C\left(h^{1/2-\epsilon}\|\boldsymbol{\xi}_q\| + h^{1/2-\epsilon}\ \delta\right)\|\boldsymbol{\tau}\|. \tag{3.40}$$

Proof. First, we consider the first term on the left-hand side of (3.40) and rewrite it as

$$\begin{aligned} N_1(z-u;\mathbf{q}_z-\mathbf{q},\boldsymbol{\tau}) &= \int_\Omega \tilde{a}_u(z)(z-u)(\mathbf{q}_z-\mathbf{q})\cdot\boldsymbol{\tau}dx \\ &= -\int_\Omega \tilde{a}_u(z)(z-I_hu)\boldsymbol{\xi}_q\cdot\boldsymbol{\tau}dx + \int_\Omega \tilde{a}_u(z)(z-I_hu)\boldsymbol{\eta}_q\cdot\boldsymbol{\tau}dx \\ &\quad -\int_\Omega \tilde{a}_u(z)\eta_u\boldsymbol{\xi}_q\cdot\boldsymbol{\tau}dx + \int_\Omega \tilde{a}_u(z)\eta_u\boldsymbol{\eta}_q\cdot\boldsymbol{\tau}dx. \end{aligned} \tag{3.41}$$

Using the inverse inequality (1.20) and Lemma 1.2.1, we estimate the first term on the right-hand side of (3.41) as

$$\begin{aligned} |\int_\Omega \tilde{a}_u(z)(z-I_hu)\boldsymbol{\xi}_q\cdot\boldsymbol{\tau}dx| &\le C\sum_{i=1}^{N_h}\|z-I_hu\|_{L^4(K_i)}\|\boldsymbol{\xi}_q\|_{L^4(K_i)^2}\|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\le C\sum_{i=1}^{N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\|z-I_hu\|_{L^4(K_i)}\ \|\boldsymbol{\xi}_q\|_{L^2(K_i)^2}\|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)|||z-I_hu|||\ \|\boldsymbol{\xi}_q\|\ \|\boldsymbol{\tau}\| \\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\ \frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{h_i}{p_i^{1/2}}\right)\|\boldsymbol{\xi}_q\|\ \|\boldsymbol{\tau}\| \\ &\le Ch^{1/2-\epsilon}\ \|\boldsymbol{\xi}_q\|\ \|\boldsymbol{\tau}\|. \end{aligned} \tag{3.42}$$

For the second term on the right-hand of (3.41), we use Lemma 1.2.5 and the trace inequality (1.19) to obtain

$$\begin{aligned} |\int_\Omega \tilde{a}_u(z)(z-I_hu)\boldsymbol{\eta}_q\cdot\boldsymbol{\tau}dx| &\le C\sum_{i=1}^{N_h}\|z-I_hu\|_{L^4(K_i)}\|\boldsymbol{\eta}_q\|_{L^4(K_i)^2}\|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\le C\sum_{i=1}^{N_h}\frac{h_i^{1/2}}{p_i^{1/2}}\|z-I_hu\|_{L^4(K_i)}\ \|\mathbf{q}\|_{H^1(K_i)^2}\ \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\le C\left(\max_{1\le i\le N_h}\frac{h_i^{1/2}}{p_i^{1/2}}\right)|||z-I_hu|||\ \|\mathbf{q}\|_{H^1(\Omega)^2}\ \|\boldsymbol{\tau}\| \\ &\le Ch^{1/2}\delta\ \|\boldsymbol{\tau}\|. \end{aligned} \tag{3.43}$$

Similarly, using the inverse inequality (1.20) and Lemma 1.2.1, the third term on the right-hand of (3.41) is estimated as

$$\begin{aligned} |\int_\Omega \tilde{a}_u(z)\eta_u \boldsymbol{\xi}_q \cdot \boldsymbol{\tau} dx| &\leq \sum_{i=1}^{N_h} \|\eta_u\|_{L^4(K_i)} \|\boldsymbol{\xi}_q\|_{L^4(K_i)^2} \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq C \sum_{i=1}^{N_h} \frac{h_i^{3/2}}{p_i^{3/2}} \|u\|_{H^2(K_i)} \frac{p_i^{1/2}}{h_i^{1/2}} \|\boldsymbol{\xi}_q\|_{L^2(K_i)^2} \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq Ch \|\boldsymbol{\xi}_q\| \|\boldsymbol{\tau}\|. \end{aligned} \tag{3.44}$$

For the fourth term on the right-hand side of (3.41), we apply Lemma 1.2.1 to find that

$$\begin{aligned} |\int_\Omega \tilde{a}_u(z)\eta_u \boldsymbol{\eta}_q \cdot \boldsymbol{\tau} dx| &\leq \sum_{i=1}^{N_h} \|\eta_u\|_{L^4(K_i)} \|\boldsymbol{\eta}_q\|_{L^4(K_i)^2} \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq C \sum_{i=1}^{N_h} \frac{h_i^{1-1/2}}{p_i^{1-1/2}} \|\eta_u\|_{L^4(K_i)} \|\mathbf{q}\|_{H^1(K_i)} \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq Ch^{1/2} |||\eta_u||| \|\boldsymbol{\tau}\| \\ &\leq Ch^{1/2+\epsilon} \delta \|\boldsymbol{\tau}\|. \end{aligned} \tag{3.45}$$

Finally, consider the second term on the left-hand side of (3.40). We bound this term as follows:

$$\begin{aligned} |\int_\Omega \tilde{a}_{uu}(z)\eta_u^2 \mathbf{q} \cdot \boldsymbol{\tau} dx| &\leq C \sum_{i=1}^{N_h} \|\eta_u\|_{L^4(K_i)}^2 \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq C |||\eta_u|||^2 \|\boldsymbol{\tau}\|_{L^2(K_i)^2} \\ &\leq Ch |||\eta_u||| \|\boldsymbol{\tau}\| \\ &\leq Ch\delta \|\boldsymbol{\tau}\|. \end{aligned} \tag{3.46}$$

Now, we combine (3.41)-(3.46) to complete the rest of the proof for any $0 < \epsilon < 1/2$. ∎

LEMMA 3.2.4 *There exists a positive constant C such that*

$$|N_1(z-u; \mathbf{e}_q, \boldsymbol{\tau}) + N_2(z-u; \mathbf{q}, \boldsymbol{\tau})| \leq C \left(|||z-u|||^2 + \|\mathbf{e}_q\| \, |||z-u||| \right) \|\boldsymbol{\tau}\|_{L^4(\Omega)}. \tag{3.47}$$

Proof. First, we consider the first term on the left-hand side of (3.47). Using the arguments as in Lemma 3.2.3, we arrive at

$$\begin{aligned} |N_1(z-u;\mathbf{q}_z-\mathbf{q},\boldsymbol{\tau})| &= |\int_\Omega \tilde{a}_u(z)(z-u)(\mathbf{q}_z-\mathbf{q})\cdot\boldsymbol{\tau}dx| \\ &\leq C|||z-u|||\ \|\mathbf{e}_q\|\ \|\boldsymbol{\tau}\|_{L^4(\Omega)^2}. \end{aligned} \tag{3.48}$$

Next, consider the second term on the left-hand side of (3.47). We bound this term as follows:

$$\begin{aligned} |N_2(z-u;\mathbf{q},\boldsymbol{\tau})| &= |\int_\Omega \tilde{a}_{uu}(z)(z-u)^2\mathbf{q}\cdot\boldsymbol{\tau}dx| \\ &\leq C\|z-u\|^2_{L^4(\Omega)}\ \|\mathbf{q}\|_{L^4(\Omega)^2}\ \|\boldsymbol{\tau}\|_{L^4(\Omega)^2} \\ &\leq C|||z-u|||^2\ \|\boldsymbol{\tau}\|_{L^4(K_i)^2}. \end{aligned} \tag{3.49}$$

Now, we combine (3.48)-(3.49) to complete the rest of the proof. ∎

LEMMA 3.2.5 *There exists a positive constant C such that*

$$|B(u;\boldsymbol{\eta}_q,\boldsymbol{\tau}_h)+N(u,\mathbf{q};\eta_u,\boldsymbol{\tau}_h)-A_1(\boldsymbol{\eta}_\sigma,\boldsymbol{\tau}_h)|\leq C\left(\|\boldsymbol{\eta}_q\|+\|\eta_u\|\right)\|\boldsymbol{\tau}_h\|,\quad \boldsymbol{\tau}_h\in\mathbf{W}_h$$

and for $\mathbf{w}_h\in\mathbf{W}_h$

$$|A_1(\boldsymbol{\eta}_q,\mathbf{w}_h)-A_2(\mathbf{w}_h,\eta_u)|\leq C\left(\|\boldsymbol{\eta}_q\|^2+|||\eta_u|||^2\right)^{1/2}\|\mathbf{w}_h\|.$$

Proof. Since $\boldsymbol{\eta}_\sigma=\Pi\boldsymbol{\sigma}-\boldsymbol{\sigma}$, where $\Pi\boldsymbol{\sigma}$ is the L^2 projection of $\boldsymbol{\sigma}$, an appeal to the Cauchy-Schwarz inequality yields the proof of the first inequality of the Lemma. For the second inequality, we note from the definition that

$$\begin{aligned} A_1(\boldsymbol{\eta}_q,\mathbf{w}_h)-A_2(\mathbf{w}_h,\eta_u) &= \int_\Omega \boldsymbol{\eta}_q\cdot\mathbf{w}_h dx+\sum_{i=1}^{N_h}\int_{K_i}\eta_u\nabla\cdot\mathbf{w}_h dx-\int_{\Gamma_I}\{\eta_u\}[\![\mathbf{w}_h]\!]ds \\ &\quad -\int_{\Gamma_I}C_{12}.[\![\eta_u]\!][\![\mathbf{w}_h]\!]ds. \end{aligned} \tag{3.50}$$

For the second term on the right-hand side of (3.50), we integrate by parts to obtain

$$\sum_{i=1}^{N_h}\int_{K_i}\eta_u\nabla\cdot\mathbf{w}_h dx-\int_{\Gamma_I}\{\eta_u\}[\![\mathbf{w}_h]\!]ds=-\sum_{i=1}^{N_h}\int_{K_i}\nabla\eta_u\cdot\mathbf{w}_h+\int_\Gamma[\![\eta_u]\!]\{\mathbf{w}_h\}ds,$$

and hence, using the trace inequality (1.19) for $l = 0$, we arrive at

$$|\sum_{i=1}^{N_h} \int_{K_i} \eta_u \nabla \cdot \mathbf{w}_h dx - \int_{\Gamma_I} \{\eta_u\}[\![\mathbf{w}_h]\!] ds| \leq C|||\eta_u||| \; \|\mathbf{w}_h\|. \tag{3.51}$$

A use of Cauchy-Schwarz inequality yields

$$|\int_{\Omega} \boldsymbol{\eta}_q \cdot \mathbf{w}_h dx| \leq \|\boldsymbol{\eta}_q\| \; \|\mathbf{w}_h\|. \tag{3.52}$$

Next, using the trace inequality (1.19), we bound the last term on the right-hand side of (3.50) as :

$$|\int_{\Gamma_I} C_{12}.[\![\eta_u]\!])[\![\mathbf{w}_h]\!] ds| \;\; \leq \;\; C \left(\sum_{e_k \in \Gamma_I} \int_{e_k} \frac{p_k^2}{h_k} [\![\eta_u]\!]^2 \right)^{1/2} \|\boldsymbol{\xi}_\sigma\|. \tag{3.53}$$

We combine (3.50)-(3.53) to complete the rest of proof. ∎

THEOREM 3.2.1 *There is a positive constant C such that for any $0 < \epsilon < 1/2$*

$$\|\boldsymbol{\xi}_\sigma\| \leq C \left(\|\boldsymbol{\xi}_q\| + \|\xi_y\| + \|\eta_u\| + \|\boldsymbol{\eta}_q\| + h^{1/2-\epsilon} \delta \right).$$

Proof. Using (3.35), we write

$$\begin{aligned} B(u; \boldsymbol{\xi}_q, \boldsymbol{\tau}_h) + N(u, \mathbf{q}; \xi_y, \boldsymbol{\tau}_h) - A_1(\boldsymbol{\xi}_\sigma, \boldsymbol{\tau}_h) = B(u; \boldsymbol{\eta}_q, \boldsymbol{\tau}_h) + N(u, \mathbf{q}; \eta_y, \boldsymbol{\tau}_h) \\ -A_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\tau}_h) + N_1(z-u; \mathbf{e}_q, \boldsymbol{\tau}_h) + N_2(z-u; \mathbf{q}, \boldsymbol{\tau}_h). \end{aligned} \tag{3.54}$$

Set $\boldsymbol{\tau}_h = \boldsymbol{\xi}_\sigma$ in (3.54) to obtain

$$\begin{aligned} \int_{\Omega} \boldsymbol{\xi}_\sigma \cdot \boldsymbol{\xi}_\sigma dx \;\; = \;\; & B(u, \boldsymbol{\xi}_q, \boldsymbol{\xi}_q) + N(u, \mathbf{q}; \xi_y, \boldsymbol{\xi}_\sigma) - B(u; \boldsymbol{\eta}_q, \boldsymbol{\xi}_\sigma) - N(u, \mathbf{q}; \eta_y, \boldsymbol{\xi}_\sigma) \\ & + A_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_\sigma) - N_1(z-u; \mathbf{e}_q, \boldsymbol{\xi}_\sigma) - N_2(z-u; \mathbf{q}, \boldsymbol{\xi}_\sigma). \end{aligned} \tag{3.55}$$

Then, a use of Lemma 3.2.3 and Lemma 3.2.5 completes the proof of the Theorem. ∎

THEOREM 3.2.2 *Let $z \in O_\delta(I_h u)$ and $(y, \; \mathbf{q}_z, \; \boldsymbol{\sigma}_z) \in V \times \mathbf{W} \times \mathbf{W}$ be the corresponding solution of (3.33)-(3.35). For any $0 < \epsilon < 1/2$ the following estimate holds :*

$$\begin{aligned} \|e_y\| \;\; \leq \;\; & C_1 \left(\max_{1 \leq i \leq N_h} \frac{h_i}{p_i^{1/2}} \right) \left(\|\boldsymbol{\xi}_q\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} C_{11} [\![\xi_y]\!]^2 ds \right)^{1/2} + C_2 |||z-u|||^2 \\ & + C_3 (h^\epsilon + h^{1/2-\epsilon}) \left(\max_{1 \leq i \leq N_h} \frac{h_i}{p_i^{1/2}} \right) \delta + C_4 \|\mathbf{e}_q\| \; |||z-u|||. \end{aligned}$$

Proof. We now appeal to the Aubin-Nitsche duality argument. Consider the following auxiliary problem :

$$\begin{aligned} -\nabla \cdot (a(u)\nabla\phi) + a_u(u)\nabla u \cdot \nabla\phi &= e_y \quad \text{in } \Omega, \\ \phi &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

which satisfies the elliptic regularity

$$\|\phi\|_{H^2(\Omega)} \leq C\|e_y\|. \tag{3.56}$$

In order to write the mixed weak formulation, let $\mathbf{p} = \nabla\phi$ and $-\boldsymbol{\psi} = a(u)\mathbf{p}$. Then, we obtain

$$\mathbf{p} = \nabla\phi \quad \text{in } \Omega, \tag{3.57}$$

$$-\boldsymbol{\psi} = a(u)\mathbf{p} \quad \text{in } \Omega, \tag{3.58}$$

$$\nabla \cdot \boldsymbol{\psi} + a_u(u)\mathbf{q} \cdot \mathbf{p} = e_y \quad \text{in } \Omega, \tag{3.59}$$

We multiply (3.59) by e_y, (3.58) by $\mathbf{e}_q$ and (3.57) by $\mathbf{e}_\sigma$, then integrate to arrive at

$$\begin{aligned} \|e_y\|^2 &= \int_\Omega e_y \nabla \cdot \boldsymbol{\psi} dx + \int_\Omega a_u(u)\mathbf{q}e_y \cdot \mathbf{p} dx + \int_\Omega a(u)\mathbf{p} \cdot \mathbf{e}_q dx + \int_\Omega \boldsymbol{\psi} \cdot \mathbf{e}_q \\ &\quad - \int_\Omega \mathbf{p} \cdot \mathbf{e}_\sigma dx + \int_\Omega \nabla\phi \cdot \mathbf{e}_\sigma dx. \end{aligned}$$

Since $[\![\phi]\!] = 0$, $[\![\boldsymbol{\psi}]\!] = 0$ on $e_k \in \Gamma_I$ and $\phi = 0$ on $\partial\Omega$, we now arrive at

$$\begin{aligned} \|e_y\|^2 &= A_1(\mathbf{e}_q, \boldsymbol{\psi}) - A_2(\boldsymbol{\psi}, e_y) + B(u; \mathbf{e}_q, \mathbf{p}) + N(u, \mathbf{q}; e_y, \mathbf{p}) - A_1(\mathbf{e}_\sigma, \mathbf{p}) \\ &\quad + A_2(\mathbf{e}_\sigma, \phi) + J(e_y, \phi). \end{aligned}$$

Then, using (3.33)-(3.35), we obtain

$$\begin{aligned} \|e_y\|^2 &= A_1(\mathbf{e}_q, \boldsymbol{\eta}_\psi) - A_2(\boldsymbol{\eta}_\psi, e_y) + A_2(\mathbf{e}_\sigma, \eta_\phi) + B(u; \mathbf{e}_q, \boldsymbol{\eta}_p) - A_1(\mathbf{e}_\sigma, \boldsymbol{\eta}_p) \\ &\quad + N(u, \mathbf{q}; e_y, \boldsymbol{\eta}_p) + J(e_y, \eta_\phi) + N_1(z - u; \mathbf{e}_q, I_h\mathbf{p}) + N(\mathbf{q}; z - u, I_h\mathbf{p}), \end{aligned} \tag{3.60}$$

where $\eta_\phi = \phi - I_h\phi$, $\boldsymbol{\eta}_p = \mathbf{p} - I_h\mathbf{p}$ and $\boldsymbol{\eta}_\psi = \boldsymbol{\psi} - \Pi\boldsymbol{\psi}$. We now expand (3.60) to find that

$$\begin{aligned}
\|e_y\|^2 &= \int_\Omega \mathbf{e}_q \cdot \boldsymbol{\eta}_\psi dx - \int_\Omega \mathbf{e}_\sigma \cdot \boldsymbol{\eta}_p dx + \sum_{i=1}^{N_h} \int_{K_i} e_y \nabla \cdot \boldsymbol{\eta}_\psi dx - \int_{\Gamma_I} \{e_y\}[\![\boldsymbol{\eta}_\psi]\!] ds \\
&+ \sum_{i=1}^{N_h} \int_{K_i} \mathbf{e}_\sigma \cdot \nabla \eta_\phi dx - \int_\Gamma (\{\mathbf{e}_\sigma\} - C_{11}[\![e_y]\!] - C_{12}[\![\mathbf{e}_\sigma]\!])[\![\eta_\phi]\!] ds \\
&+ \int_\Omega a(u)\mathbf{e}_q \cdot \boldsymbol{\eta}_p dx - \int_{\Gamma_I} C_{12}.[\![e_y]\!][\![\boldsymbol{\eta}_\psi]\!] ds + \int_\Omega a_u(u)\mathbf{q}e_y \cdot \boldsymbol{\eta}_p dx \\
&- N_1(z-u; \mathbf{e}_q, I_h\mathbf{p}) + N_2(z-u; \mathbf{q}, I_h\mathbf{p}). \qquad (3.61)
\end{aligned}$$

Since $\Pi\boldsymbol{\psi}$ is the L^2 projection of $\boldsymbol{\psi}$, we bound the following terms using Lemma 1.2.5 as:

$$|\sum_{i=1}^{N_h} \int_{K_i} e_y \nabla \cdot \boldsymbol{\eta}_\psi dx - \int_{\Gamma_I} \{e_y\}[\![\boldsymbol{\eta}_\psi]\!] ds| = |-\sum_{i=1}^{N_h} \int_{K_i} \nabla e_y \cdot \boldsymbol{\eta}_\psi dx + \int_\Gamma [\![e_y]\!]\{\boldsymbol{\eta}_\psi\} ds|$$

$$\begin{aligned}
&= |-\sum_{i=1}^{N_h} \int_{K_i} \nabla \eta_u \cdot \boldsymbol{\eta}_\psi dx - \sum_{i=1}^{N_h} \int_{K_i} \nabla \xi_y \cdot \boldsymbol{\eta}_\psi dx + \int_\Gamma [\![e_y]\!]\{\boldsymbol{\eta}_\psi\} ds| \\
&\leq C \left(\sum_{i=1}^{N_h} \frac{h_i^2}{p_i^2} \|\nabla \eta_u\|^2_{L^2(K_i)} \right)^{1/2} \|\boldsymbol{\psi}\|_{H^1(\Omega)^2} \\
&\quad + \sum_{e_k \in \Gamma} \left(\int_{e_k} \frac{p_k^2}{h_k} [\![e_y]\!]^2 ds \right)^{1/2} \left(\int_{e_k} \frac{h_k}{p_k^2} \{\boldsymbol{\eta}_\psi\}^2 ds \right)^{1/2} \\
&\leq C \left(\sum_{i=1}^{N_h} \frac{h_i^2}{p_i^2} \|\nabla \eta_u\|^2_{L^2(K_i)} \right)^{1/2} \|\boldsymbol{\psi}\|_{H^1(\Omega)^2} \\
&\quad + \left(\sum_{e_k \in \Gamma} \int_{e_k} \frac{h_k^2}{p_k^2} \frac{p_k^2}{h_k} [\![e_u]\!]^2 ds \right)^{1/2} \|\boldsymbol{\psi}\|_{H^1(\Omega)^2}. \qquad (3.62)
\end{aligned}$$

Next, using Lemma 1.2.1 we find that

$$|\int_\Omega \mathbf{e}_q \cdot \boldsymbol{\eta}_\psi dx + \int_\Omega a(u)\mathbf{e}_q \cdot \boldsymbol{\eta}_p dx| \leq C \left(\sum_{i=1}^{N_h} \frac{h_i^2}{p_i^2} \|\mathbf{e}_q\|^2_{L^2(K_i)^2} \right)^{1/2} \|\mathbf{p}\|_{H^1(\Omega)^2}. \qquad (3.63)$$

Again a use of Lemma 1.2.1 yields

$$|\sum_{i=1}^{N_h} \int_{K_i} \mathbf{e}_\sigma \cdot \nabla \eta_\phi dx - \int_\Omega \mathbf{e}_\sigma \cdot \boldsymbol{\eta}_p dx| \leq C \left(\sum_{i=1}^{N_h} \frac{h_i^2}{p_i^2} \|\mathbf{e}_\sigma\|^2_{L^2(K_i)^2} \right)^{1/2} \|\phi\|_{H^2(\Omega)} \qquad (3.64)$$

and

$$|\int_{\Omega} a_u(u)\mathbf{q}e_y \cdot \boldsymbol{\eta}_p dx| \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|e_y\|^2_{L^2(K_i)}\right)^{1/2}\|\mathbf{p}\|_{H^1(\Omega)^2}. \tag{3.65}$$

Using Lemma 1.2.5, we obtain

$$\begin{aligned}|\int_{\Gamma_I} C_{12}.[e_y][\boldsymbol{\eta}_\psi]ds| &\leq C\sum_{e_k\in\Gamma_I}\left(\int_{e_k} C_{11}[e_y]^2ds\right)^{1/2}\left(\int_{e_k}\frac{h_k}{p_k^2}\{\boldsymbol{\eta}_\psi\}^2ds\right)^{1/2}\\ &\leq C\left(\sum_{e_k\in\Gamma_I}\int_{e_k}\frac{h_k^2}{p_k^2}C_{11}[\![e_y]\!]^2ds\right)^{1/2}\|\boldsymbol{\psi}\|_{H^1(\Omega)}. \end{aligned}\tag{3.66}$$

Now an application of Lemma 1.2.1 with the trace inequality (1.19) implies that

$$\begin{aligned}|\int_{\Gamma}(\{\mathbf{e}_\sigma\} &- C_{12}[\![\mathbf{e}_\sigma]\!])[\![\eta_\phi]\!]ds| \leq C\sum_{e_k\in\Gamma}\left(\int_{e_k}\{|\boldsymbol{\xi}_\sigma|\}|[\eta_\phi]|ds+\int_{e_k}\{|\boldsymbol{\eta}_\sigma|\}|[\eta_\phi]|ds\right)\\ &\leq C\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\xi}_\sigma|\}^2ds+\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\eta}_\sigma|\}^2ds\right)^{1/2}\|\phi\|_{H^2(\Omega)}\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\|\boldsymbol{\xi}_\sigma\|\ \|\phi\|_{H^2(\Omega)}\\ &\quad+\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\eta}_\sigma|\}^2ds\right)^{1/2}\|\phi\|_{H^2(\Omega)}, \end{aligned}\tag{3.67}$$

and

$$\begin{aligned}|\int_{\Gamma} C_{11}[\![e_y]\!][\eta_\phi]ds| &\leq C\sum_{e_k\in\Gamma}\left(\int_{e_k}C_{11}[\![e_y]\!]^2ds\right)^{1/2}\left(\int_{e_k}\frac{p_k^2}{h_k}[\eta_\phi]^2ds\right)^{1/2}\\ &\leq C\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^2}{p_k}C_{11}[e_y]^2ds\right)^{1/2}\|\phi\|_{H^2(\Omega)}. \end{aligned}\tag{3.68}$$

Finally, using Lemma 3.2.4 and the stability of I_h that is, $\|I_h\mathbf{p}\|_{L^4(\Omega)^2}\leq C\|\mathbf{p}\|_{H^1(\Omega)^2}$, we find that

$$\begin{aligned}|N_1(z-u;\mathbf{e}_q,I_h\mathbf{p}) &+ N_2(z-u;\mathbf{q},I_h\mathbf{p})|\\ &\leq C|||z-u|||^2\ \|\mathbf{p}\|_{H^1(\Omega)^2}+\|\mathbf{e}_q\|\ |||z-u|||\ \|\mathbf{p}\|_{H^1(\Omega)^2}. \end{aligned}\tag{3.69}$$

We combine the estimates (3.62)-(3.69) and then use elliptic regularity (3.56) to obtain

$$\begin{aligned}\|e_y\| \;\leq\; & C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\boldsymbol{\xi}_q\|^2_{L^2(K_i)^2}+\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^2}{p_k}C_{11}[\![\xi_y]\!]^2ds\right)^{1/2}+C_2|||z-u|||^2\\
&+C_3\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\left(\|\nabla\eta_u\|^2_{L^2(K_i)}+\|\boldsymbol{\eta}_q\|^2_{L^2(K_i)^2}+\|\boldsymbol{\eta}_\sigma\|^2_{L^2(K_i)^2}\right)\right)^{1/2}\\
&+C_4\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^2}{p_k}C_{11}[\![\eta_u]\!]^2ds+\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\eta}_\sigma|\}^2ds\right)^{1/2}+C_5\|\mathbf{e}_q\|\;|||z-u|||\\
&+C_6\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\|\boldsymbol{\xi}_\sigma\|. \qquad (3.70)\end{aligned}$$

Then, a use of Theorem 3.2.1 completes the rest of the proof. ∎

THEOREM 3.2.3 *For all $0<h<h_0$ where $h_0<1$, there is a $\delta=\delta(h)>0$ such that S_h maps from $O_\delta(I_hu)$ into itself.*

Proof. Set $v_h=\xi_y$, $\boldsymbol{\tau}_h=\boldsymbol{\xi}_q$ and $\mathbf{w}_h=\boldsymbol{\xi}_\sigma$ in (3.36) -(3.38). Using Lemma 3.2.1, we obtain

$$\begin{aligned}C_1\left(\|\boldsymbol{\xi}_q\|^2+\int_\Gamma C_{11}[\![\xi_y]\!]^2\right)-C_2\|\xi_y\|^2 &\leq A_1(\boldsymbol{\xi}_q,\boldsymbol{\xi}_\sigma)-A_2(\boldsymbol{\xi}_\sigma,\xi_y)+B(u;\boldsymbol{\xi}_q,\boldsymbol{\xi}_q)\\
&\quad -A_2(\boldsymbol{\xi}_\sigma,\boldsymbol{\xi}_q)+N(u,\mathbf{q};\xi_y,\boldsymbol{\xi}_q)+A_2(\boldsymbol{\xi}_\sigma,\xi_y)+J_1(\xi_y,\xi_y)\\
&= A_1(\boldsymbol{\eta}_q,\boldsymbol{\xi}_\sigma)-A_2(\boldsymbol{\xi}_\sigma,\eta_u)+B(u;\boldsymbol{\eta}_q,\boldsymbol{\xi}_q)-A_1(\boldsymbol{\eta}_\sigma,\boldsymbol{\xi}_q)\\
&\quad +N(u,\mathbf{q};\eta_u,\boldsymbol{\xi}_q)+A_2(\boldsymbol{\eta}_\sigma,\xi_y)+J_1(\eta_u,\xi_y)\\
&\quad +N_1(z-u;\mathbf{e}_q,\boldsymbol{\xi}_q)+N_2(z-u;\mathbf{q},\boldsymbol{\xi}_q). \qquad (3.71)\end{aligned}$$

From the definition of A_2 and J, we write

$$\begin{aligned}A_2(\boldsymbol{\eta}_\sigma,\xi_y)+J_1(\eta_u,\xi_y) \;=\; & \sum_{i=1}^{N_h}\int_{K_i}\boldsymbol{\eta}_\sigma\cdot\nabla\xi_y dx\\
&-\int_\Gamma(\{\boldsymbol{\eta}_\sigma\}-C_{11}[\![\eta_u]\!]-C_{12}[\![\boldsymbol{\eta}_\sigma]\!])[\![\xi_y]\!]ds. \qquad (3.72)\end{aligned}$$

Since $\Pi\boldsymbol{\sigma}$ is L^2 projections of $\boldsymbol{\sigma}$ onto $\mathbf{W}_h$, we obtain

$$\sum_{i=1}^{N_h}\int_{K_i}\boldsymbol{\eta}_\sigma\cdot\nabla\xi_y dx=0. \qquad (3.73)$$

Next, using the trace inequality (1.19) and the assumption that $C_{11}|_{e_k} = \beta p_k^2/h_k$, we bound the following terms as:

$$\begin{aligned} |\int_\Gamma (\{\boldsymbol{\eta}_\sigma\} - C_{11}[\![\eta_u]\!] - C_{12}[\![\boldsymbol{\eta}_\sigma]\!])[\![\xi_y]\!] ds| \;\; &\leq \;\; C \left(\sum_{e_k \in \Gamma} \int_{e_k} \frac{h_k}{p_k^2} \{|\boldsymbol{\eta}_\sigma|\}^2 ds \right)^{1/2} J(\xi_y, \xi_y) \\ &\quad + C \; J(\eta_u, \eta_u) \; J(\xi_y, \xi_y). \end{aligned}$$

An appeal to Lemma 3.2.5 with $\boldsymbol{\tau}_h = \boldsymbol{\xi}_q$ and $\mathbf{w}_h = \boldsymbol{\xi}_\sigma$ yields

$$|B(u; \boldsymbol{\eta}_q, \boldsymbol{\xi}_q) + N(u, \mathbf{q}; \eta_u, \boldsymbol{\xi}_q) - A_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_q)| \leq C(\|\boldsymbol{\eta}_q\| + \|\eta_u\|)\|\boldsymbol{\xi}_q\|, \tag{3.74}$$

and

$$|A_1(\boldsymbol{\eta}_q, \boldsymbol{\xi}_\sigma) - A_2(\boldsymbol{\xi}_\sigma, \eta_u)| \leq C|||\eta_u||| \; \|\boldsymbol{\xi}_\sigma\|. \tag{3.75}$$

For the last two terms on the right-hand side of (3.71), we set $\boldsymbol{\tau} = \boldsymbol{\xi}_q$ in Lemma 3.2.3 to obtain

$$\begin{aligned} |N_1(z-u; \mathbf{e}_q, \boldsymbol{\xi}_q) + N_2(z-u; \mathbf{q}, \boldsymbol{\xi}_q)| \;\; &\leq \;\; Ch^{1/2-\epsilon} \|\boldsymbol{\xi}_q\|^2 \\ &\quad + h^{1/2-\epsilon} \; \delta \; \|\boldsymbol{\xi}_q\|. \end{aligned} \tag{3.76}$$

An appeal to Theorem 3.2.1 implies that

$$\|\boldsymbol{\xi}_\sigma\| \leq C \left(\|\boldsymbol{\xi}_q\| + \|\xi_y\| + \|\eta_u\| + \|\boldsymbol{\eta}_q\| + \|\boldsymbol{\eta}_\sigma\| + h^{1/2-\epsilon}\delta \right). \tag{3.77}$$

We combine the estimates (3.71)-(3.77) to obtain for sufficiently small h

$$\begin{aligned} \left(\|\boldsymbol{\xi}_q\|^2 + \int_\Gamma C_{11}[\![\xi_y]\!]^2 \right) \;\; &\leq \;\; C_1 \left(|||\eta_u|||^2 + \|\boldsymbol{\eta}_q\|^2 + \|\boldsymbol{\eta}_\sigma\|^2 + h^{1-2\epsilon}\delta^2 \right. \\ &\quad \left. + \sum_{e_k \in \Gamma} \int_{e_k} \frac{h_k}{p_k^2} \{|\boldsymbol{\eta}_\sigma|\}^2 ds \right) + C_2 \|\xi_y\|^2 \\ &\leq \;\; C_1 \left(h^2\epsilon + h^{1-2\epsilon} \right) \delta^2 + C_2 \|\xi_y\|^2. \end{aligned} \tag{3.78}$$

Using Theorem 3.2.2 and the estimate (3.78), we find that for sufficiently small h

$$\left(\|\boldsymbol{\xi}_q\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} C_{11}[\![\xi_y]\!]^2 ds \right)^{1/2} \leq C \left(h^\epsilon + h^{1/2-\epsilon} + \delta \right) \delta. \tag{3.79}$$

Next, set $\mathbf{w}_h = \nabla\xi_y$ in (3.36) to obtain

$$\begin{aligned}\sum_{i=1}^{N_h}\int_{K_i}\boldsymbol{\xi}_q\cdot\nabla\xi_y dx+\sum_{i=1}^{N_h}\int_{K_i}\xi_y\nabla\cdot\nabla\xi_y dx &- \int_{\Gamma_I}(\{\xi_y\}+C_{12}[\![\xi_y]\!])[\![\nabla\xi_y]\!]ds\\ &= A_1(\boldsymbol{\eta}_q,\nabla\xi_y)-A_2(\nabla\xi_y,\eta_u).\end{aligned}$$

An integration by parts yields

$$\begin{aligned}\sum_{i=1}^{N_h}\int_{K_i}\nabla\xi_y\cdot\nabla\xi_y dx &= -\int_\Omega\boldsymbol{\xi}_q\cdot\nabla\xi_y dx+\int_\Gamma[\xi_y]\{\nabla\xi_y\}ds+\int_{\Gamma_I}C_{12}[\![\xi_y]\!][\![\nabla\xi_y]\!]ds\\ &\quad -A_1(\boldsymbol{\eta}_q,\nabla\xi_y)+A_2(\nabla\xi_y,\eta_u).\end{aligned}$$

Then, using the trace inequality (1.19) and Lemma 3.2.5, we obtain

$$\left(\sum_{i=1}^{N_h}\int_{K_i}\|\nabla\xi_y\|^2_{L^2(K_i)}dx\right)^{1/2} \le C\left(\|\boldsymbol{\xi}_q\|^2+\int_\Gamma C_{11}[\![\xi_y]\!]^2ds+|||\eta_u|||^2\right)^{1/2}. \quad (3.80)$$

Substituting (3.79) in (3.80), we obtain for small h and $0<\delta<1$ with $0<\epsilon<1/2$

$$|||\xi_y|||\le C\left(h^\epsilon+h^{1/2-\epsilon}+\delta\right)\delta\le\delta. \quad (3.81)$$

This completes the rest of the proof. ∎

THEOREM 3.2.4 *Let $z_1,\ z_2\in O_\delta(I_hu)$ with $0<\delta<1$. Then for sufficiently small h and $0<\epsilon<1/2$, there exists a constant C such that*

$$|||S_hz_1-S_hz_2|||\le Ch^{1/2-\epsilon}|||z_1-z_2|||. \quad (3.82)$$

Proof. Let $y_i=S_hz_i$, $\mathbf{q}_{z_i}=\mathbf{q}_i$ and $\boldsymbol{\sigma}_{z_i}=\boldsymbol{\sigma}_i$, for $i=1,\ 2$. From Theorem 3.2.3 and the estimates (3.77) as well as (3.79), it follows that

$$(\ |||y_i-I_hu|||+\|\mathbf{q}_i-I_h\mathbf{q}\|+\|\boldsymbol{\sigma}_i-\Pi\boldsymbol{\sigma}\|\ )\le C\left(h^\epsilon+h^{1/2-\epsilon}+\delta\right)\delta.$$

Using (3.33)-(3.35), we note that for any $(\mathbf{w}_h,v_h,\boldsymbol{\tau}_h)\in\mathbf{W}_h\times V_h\times\mathbf{W}_h$

$$\begin{aligned}A_1(\mathbf{q}_1-\mathbf{q}_2,\mathbf{w}_h)-A_2(\mathbf{w}_h,y_1-y_2)&=0,\\ A_2(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,v_h)+J(y_1-y_2,v_h)&=0,\end{aligned}$$

and

$$B(u;\mathbf{q}_1-\mathbf{q}_2,\boldsymbol{\tau}_h)+N(u,\mathbf{q};y_1-y_2,\boldsymbol{\tau}_h)-A_1(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,\boldsymbol{\tau}_h) \quad (3.83)$$
$$=\int_\Omega (a(z_1)-a(u))(\mathbf{q}_1-\mathbf{q})\cdot\boldsymbol{\tau}_h dx+\int_\Omega (a(z_1)-a(u)-a_u(u)(z_1-u))\mathbf{q}\cdot\boldsymbol{\tau}_h dx$$
$$-\int_\Omega (a(z_2)-a(u))(\mathbf{q}_2-\mathbf{q})\cdot\boldsymbol{\tau}_h dx-\int_\Omega (a(z_2)-a(u)-a_u(u)(z_2-u))\mathbf{q}\cdot\boldsymbol{\tau}_h dx.$$

We rewrite (3.83) as

$$\begin{aligned}
&B(u;\mathbf{q}_1-\mathbf{q}_2,\boldsymbol{\tau}_h)+N(u,\mathbf{q};y_1-y_2,\boldsymbol{\tau}_h)-A_1(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,\boldsymbol{\tau}_h)\\
&=\int_\Omega (a(z_1)-a(z_2))(\mathbf{q}_1-\mathbf{q})\cdot\boldsymbol{\tau}_h dx-\int_\Omega (a(z_2)-a(u))(\mathbf{q}_1-\mathbf{q}_2)\cdot\boldsymbol{\tau}_h dx\\
&\quad+\int_\Omega (a(z_1)-a(z_2)-a_u(z_2)(z_1-z_2))\mathbf{q}\cdot\boldsymbol{\tau}_h dx\\
&\quad-\int_\Omega (a_u(z_2)-a_u(u))(z_1-z_2)\mathbf{q}\cdot\boldsymbol{\tau}_h dx.
\end{aligned}$$

Now using similar arguments as in Theorem 3.2.3, we first obtain

$$\left(\|\mathbf{q}_1-\mathbf{q}_2\|^2+\sum_{e_k\in\Gamma}\int_{e_k} C_{11}[\![y_1-y_2]\!]^2 ds\right)^{1/2} \leq C_1 h^{1/2-\epsilon}|||z_1-z_2||| \quad (3.84)$$
$$+C_2\|y_1-y_2\|,$$

and

$$\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\| \leq C_1 h^{1/2-\epsilon}|||z_1-z_2|||+C_2\|y_1-y_2\|. \quad (3.85)$$

Then, an application of duality argument as in Theorem 3.2.2 yields

$$\|y_1-y_2\|\leq Ch^{1/2-\epsilon}\left(\|\mathbf{q}_1-\mathbf{q}_2\|^2+\sum_{e_k\in\Gamma}\int_{e_k} C_{11}[\![y_1-y_2]\!]^2 ds\right)^{1/2}. \quad (3.86)$$

Since

$$|||y_1-y_2|||\leq C\left(\|\mathbf{q}_1-\mathbf{q}_2\|^2+\sum_{e_k\in\Gamma}\int_{e_k} C_{11}[\![y_1-y_2]\!]^2 ds\right)^{1/2}, \quad (3.87)$$

we combine the estimates (3.84)-(3.87) to complete the rest of the proof. ∎

Now, we conclude from Theorem 3.2.4 that the map S_h is well defined, that is, the linearized problem (3.33)-(3.34) is well-posed and continuous in the ball $O_\delta(I_h u)$. Hence,

an appeal to Brouwer fixed point theorem implies that S_h has a fixed point u_h in $O_\delta(I_h u)$. Then, using Theorem 3.2.4, it is easy to see that u_h is the unique fixed point in $O_\delta(I_h u)$ for small h. Moreover, $(u_h, \mathbf{q}_h = \mathbf{q}_{u_h}, \boldsymbol{\sigma}_h = \boldsymbol{\sigma}_{u_h})$ is the unique solution of the problem (3.30)-(3.32).

3.2.2 A priori error estimates.

Note that $u - u_h$ satisfies the estimate (3.81) and Theorem 3.2.2 by replacing y by u_h, Further, from the estimate (3.79) replacing $\mathbf{q}_z$ by $\mathbf{q}_h$ and using the property of I_h, we arrive at an estimate of $\mathbf{q} - \mathbf{q}_h$. Hence, by choosing $\epsilon = 1/4$, we easily prove the following theorem.

THEOREM 3.2.5 *There exists a constant C such that for sufficiently small h the following estimates hold:*

$$\begin{aligned} |||u - u_h|||^2 &\leq C \sum_{i=1}^{N_h} \left(\frac{h_i^{2\mu_i^+}}{p_i^{2s_i}} \|\nabla u\|^2_{H^{s_i}(K_i)} + \frac{h_i^{2\mu_i^*}}{p_i^{2s_i-1}} \|u\|^2_{H^{s_i+1}(K_i)} \right), \\ \|\mathbf{q} - \mathbf{q}_h\|^2 &\leq C \sum_{i=1}^{N_h} \left(\frac{h_i^{2\ \mu_i^+}}{p_i^{2s_i}} \|\nabla u\|^2_{H^{s_i}(K_i)} + \frac{h_i^{2\mu_i^*}}{p_i^{2s_i-1}} \|u\|^2_{H^{s_i+1}(K_i)} \right), \end{aligned}$$

and

$$\|u - u_h\|^2 \leq C \left(\max_{1 \leq i \leq N_h} \frac{h_i^2}{p_i} \right) \sum_{i=1}^{N_h} \left(\frac{h_i^{2\mu_i^+}}{p_i^{2s_i}} \|\nabla u\|^2_{H^{s_i}(K_i)} + \frac{h_i^{2\mu_i^*}}{p_i^{2s_i-1}} \|u\|^2_{H^{s_i+1}(K_i)} \right),$$

where $\mu_i^+ = min\{s_i,\ p_i + 1\}$ and $\mu_i^* = min\{s_i,\ p_i\}$.

REMARK 3.2.2 . *Note that the error estimates obtained in the above theorem are optimal in h and suboptimal in p. These estimates are exactly same as in the case of linear elliptic problems, see [57].*

3.3 Numerical experiments

In this section, we discuss some numerical results to illustrate the performance of the LDG method applied to two different types of nonlinear elliptic problems. Since the scheme deals

with discontinuous finite element spaces, the global basis functions can have support only on a single finite element. Hence, the assembly of the local matrices to the corresponding global matrices is easier than in the case of conforming finite element method.

For both the examples, we take $\Omega = (0,\ 1) \times (0,\ 1)$ and $g = 0$. The finite element subdivision $\mathcal{T}_h$ is of uniform triangles and the discontinuous finite element spaces of degree $p = 1$ and $p = 2$ ($p_i = p\ \forall i$). Take the stabilizing parameter $\beta = 1$ and set $C_{12} = (1,\ 1)$. The LDG method (3.18)-(3.20) has three unknowns, namely; u_h, $\mathbf{q}_h$ and $\boldsymbol{\sigma}_h$. Using (3.18), we first solve $\mathbf{q}_h$ in terms of u_h to write the system (3.19)-(3.20) in two unknowns u_h and $\boldsymbol{\sigma}_h$. Then, we apply the Newton's method to solve the resulting nonlinear system.
Let N_v and N_w be the dimensions of V_h and W_h. If $p = 1$, we choose the basis $\{\phi_i\}_{i=1}^{N_v}$ for V_h as in (2.60)-(2.61) and if $p = 2$, we then choose the basis as in (2.62)-(2.66). Now, let $\{\boldsymbol{\psi}_l\}_{l=1}^{N_w}$ denote bases for $\mathbf{W}_h$, which is obtained by taking tensor product of the basis of V_h. Then, we define the following matrices

$$A = [a_{ml}]_{1\le m,l\le N_w}\,, \qquad B = [b_{li}]_{1\le l\le N_w,\ 1\le i\le N_v}\,, \qquad D = [d_{ij}]_{1\le i,j\le N_v} \tag{3.88}$$

and the vector

$$L = [l_i]_{1\le i\le N_v,1}\,,$$

where

$$a_{ml} = \int_\Omega \boldsymbol{\psi}_m \cdot \boldsymbol{\psi}_l dx, \qquad b_{li} = \sum_{i=1}^{N_h} \int_{K_i} \phi_i \nabla\cdot \boldsymbol{\psi}_l dx - \sum_{e_k\in\Gamma_I} \int_{e_k} (\{\phi_i\} + C_{12}[\phi_i])[\boldsymbol{\psi}_l] ds,$$
$$d_{ij} = \sum_{e_k\in\Gamma} \int_{e_k} C_{11}[\phi_i][\phi_j] ds, \qquad \text{and} \quad l_i = \int_\Omega f\phi_i dx.$$

Write

$$u_h = \sum_{i=1}^{N_v} \alpha_i\phi_i, \quad \mathbf{q}_h = \sum_{l=1}^{N_w} b_l\boldsymbol{\psi}_l \ \text{ and } \ \boldsymbol{\sigma}_h = \sum_{l=1}^{N_w} \gamma_l\boldsymbol{\psi}_l, \tag{3.89}$$

where $\alpha = [\alpha_1,\ \alpha_2, \cdots, \alpha_{N_w}]$, $\mathbf{b} = [b_1,\ b_2, \cdots, b_{N_w}]$ and $\gamma = [\gamma_1,\ \gamma_2, \cdots, \gamma_{N_w}]$. Using the bases for V_h and $\mathbf{W}_h$, (3.18) can be reduced to the following matrix equation

$$A\mathbf{b} + B\alpha = 0, \qquad \text{with } A,\ B \ \text{ defined as in (3.88).} \tag{3.90}$$

Since the basis functions $\{\boldsymbol{\psi}_l\}_{l=1}^{N_w}$ can be assumed independently in each triangle $K \in T_h$, the symmetric positive definite global matrix $[A]$ has the following block diagonal form

$$[A] = \left[[A_{K_1}], ..., [A_{K_{N_h}}] \right],$$

where only the diagonal entries are shown. The other entries in $[A]$ are null matrices. The element matrices $[A_{K_i}]$ are symmetric and positive definite for $i = 1,\ 2, ..., N_h,\ \ [A]^{-1}$ has the block diagonal form

$$[A]^{-1} = \left[[A_{K_1}]^{-1}, ..., [A_{K_{N_h}}]^{-1} \right].$$

From (3.90), it is easy to see that $\mathbf{b} = -A^{-1}B\alpha$. Substituting $\mathbf{b} = -A^{-1}B\alpha$ in (3.19)-(3.20), using (3.88)-(3.89) and the bases for V_h and $\mathbf{W}_h$, (3.19)-(3.20) can be reformulated as : Find $[\gamma, \alpha]^T$ such that

$$F_i^1(\gamma, \alpha) = 0 \quad \text{for} \ \ 1 \le i \le N_v,$$
$$F_l^2(\gamma, \alpha) = 0 \quad \text{for} \ \ 1 \le l \le N_w,$$

where

$$F_i^1(\gamma, \alpha) \ = \ \sum_{m=1}^{N_w} \gamma_m(-b_{mi}) + \sum_{j=1}^{N_v} \alpha_j d_{ji} - l_i,$$

$$F_l^2(\gamma, \alpha) \ = \ \int_\Omega a \left(\sum_{j=1}^{N_v} \alpha_j \phi_j \right) \left(\sum_{m=1}^{N_w} [-A^{-1}B\alpha]_m \boldsymbol{\psi}_m \right) \cdot \boldsymbol{\psi}_l dx - \sum_{m=1}^{N_w} \gamma_m a_{ml}$$

and $[-A^{-1}B\alpha]_m = -\sum_{j=1}^{N_v} (A^{-1}B)_{m,j}\alpha_j$.

In order to solve the nonlinear algebraic system, we apply the Newton's method. The Jacobian Matrix J of the system takes the form

$$J = \begin{bmatrix} -B^T & D \\ -A & G \end{bmatrix},$$

where $G = [g_{li}] = [\partial F_l^2 / \partial \alpha_i]$ and B^T is the transpose of B.

Example 1. In this example, we set the nonlinear term $a(u)$ as $1 + u^2$, and choose the load

function f suitably so that the exact solution is $u = x(e - e^x)y(e - e^y)$. The initial guess for the Newton's iteration is taken to be the solution of the LDG method corresponding to the linearized problem, *i.e.*, by setting $a(u) = 1$. For this example, we consider the approximate solution obtained after 10 iterations. The order of convergence for $e_u = u - u_h$ and $\mathbf{e}_q = \mathbf{q} - \mathbf{q}_h$ is computed for the cases $p = 1$ and 2. Figures 3.1 and 3.2 show the computed order of convergences for $\|e_u\|$ and $\|e_{\mathbf{q}}\|$, respectively, in the log-log scale. These computed order of convergences match with the theoretical order of convergence derived in the Theorem 3.2.5.

Figure 3.1: Order of convergence for $||e_u||$ in Example 1.

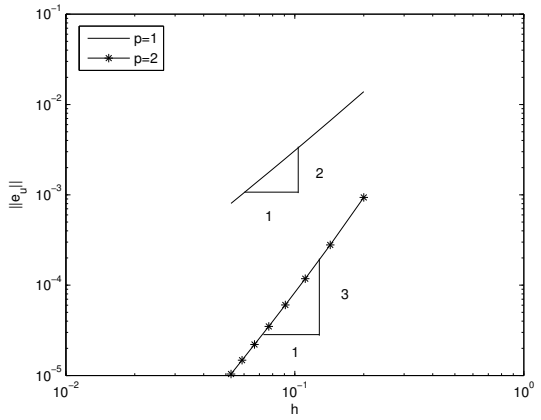
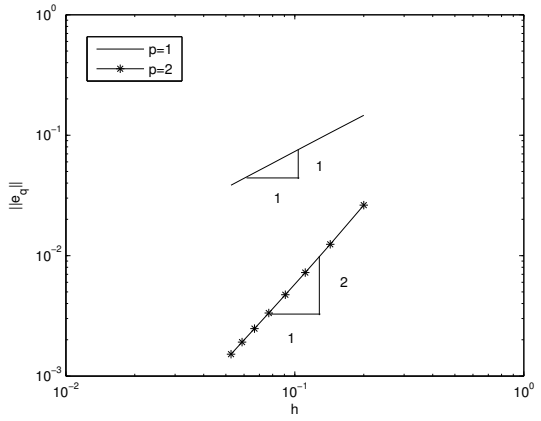


Figure 3.2: Order of convergence for $||\mathbf{e}_q||$ in Example 1.



Example 2. Set the nonlinear term $a(u)$ as $1 + u$ and choose the load function f so that

the exact solution is $u = x^{7/2}(1-x)y^{7/2}(1-y)$. The initial guess and the number of iterations for the Newton's method are taken as in Example 1. We then compute the order of convergence for $e_u = u - u_h$ and $\mathbf{e}_q = \mathbf{q} - \mathbf{q}_h$ for the cases $p = 1$ and 2.

Figure 3.3: Order of convergence for $\|e_u\|$ in Example 2.

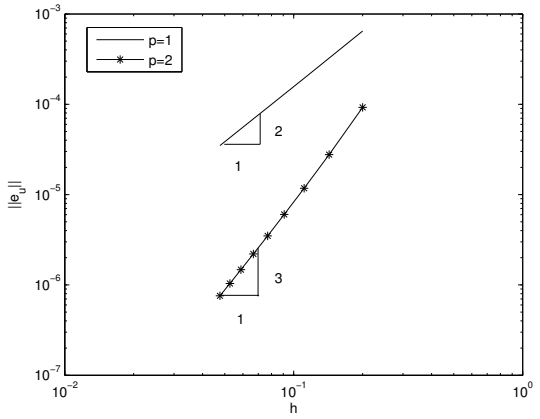
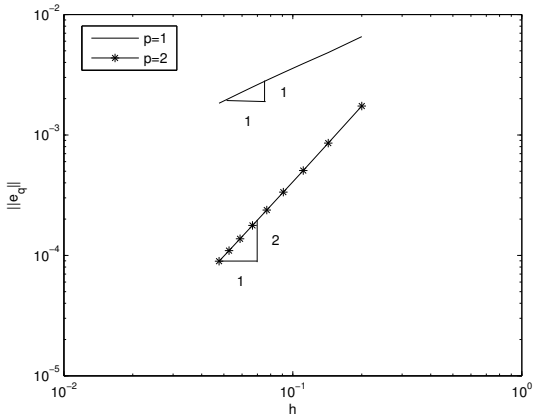


Figure 3.4: Order of convergence for $\|\mathbf{e}_q\|$ in Example 2.



Figures 3.3 and 3.4 show the computed order of convergences for $\|e_u\|$ and $\|e_{\mathbf{q}}\|$, respectively, in the log-log scale. These computed order of convergences match with the theoretical order of convergence obtained in Theorem 3.2.5.

Chapter 4

DG Methods for Strongly Nonlinear Elliptic Problems

4.1 Introduction

In this Chapter, we discuss the *hp*-discontinuous Galerkin methods for the following strongly nonlinear elliptic problem:

$$-\nabla \cdot \mathbf{a}(\mathbf{x}, u, \nabla u) + f(\mathbf{x}, u, \nabla u) = 0, \qquad \mathbf{x} \in \Omega, \tag{4.1}$$

$$u = g, \qquad \mathbf{x} \in \partial\Omega, \tag{4.2}$$

where Ω is a bounded domain with boundary $\partial\Omega$. Problems of the type (4.1)-(4.2) arise in several areas of applications such as the mean curvature, subsonic flow and Bratu's problem. In particular, we have the following for examples which fall under this form.

1. In mean curvature flow, we note that

$$\mathbf{a}(\mathbf{x}, u, \nabla u) = \left[1 + \|\nabla u\|^2\right]^{-1/2} \nabla u, \quad f(\mathbf{x}, u, \nabla u) = f(x).$$

2. In subsonic flow of an irrotational, ideal and compressible gas problem, we have

$$\mathbf{a}(\mathbf{x}, u, \nabla u) = \left[1 - \frac{\gamma - 1}{2}\|\nabla u\|^2\right]^{1/(\gamma-1)} \nabla u, \ \gamma > 1, \quad f(\mathbf{x}, u, \nabla u) = f(x).$$

3. In Bratu's problem, we observe that

$$\mathbf{a}(\mathbf{x}, u, \nabla u) = \nabla u, \quad f(\mathbf{x}, u, \nabla u) = \lambda e^u, \ \lambda > 0.$$

The work on the discontinuous Galerkin (DG) methods for the linear elliptic problems can be found in [61], [58] and [5]. Except for [45], there is hardly any result on the DG methods for the nonlinear elliptic problems. In [45], the authors have discussed DG methods for a class of monotone nonlinear elliptic problems. Therefore, in this chapter, an attempt has been made to study the DG methods for the strongly nonlinear elliptic boundary value problems (4.1)-(4.2) in which the principal part may not satisfy the strongly monotonicity and uniformly Lipschitz continuity conditions. Note that principal part in the examples 1 and 2 satisfy one of these conditions only.

We have developed a one parameter of DG methods for the problems (4.1)-(4.2) and have derived error estimates in broken H^1-norm which are optimal in h and mildly suboptimal in p. These estimates are precisely the same estimates, that is, optimal in h and suboptimal in p estimates in the case of linear elliptic problems.

The rest of the chapter is organized as follows. Section 4.3 is devoted to the discontinuous Galerkin formulations and error estimates. In this section, we have also introduced a one parameter family of discontinuous formulation which is parametrized by $\theta \in [-1, 1]$, which are exactly same as nonsymmetric when $\theta = +1$ and symmetric when $\theta = -1$ in the case of $\mathbf{a}(u, \nabla u) = \nabla u$ and $f(u, \nabla u) = 0$, that is, in the case of Laplace equation. We have discussed the existence of a discrete solution using Brouwer fixed point theorem and have derived the hp-error estimates in broken H^1-norm. In section 4.4, we have derived optimal h and suboptimal p error estimates in the L^2-norm when $\theta = -1$.

4.2 Discontinuous Galerkin Methods

In this section, we again recall the following nonlinear elliptic boundary value problem :

$$-\sum_{i=1}^{2} \frac{\partial a_i}{\partial x_i}(\mathbf{x}, u, \nabla u) + f(\mathbf{x}, u, \nabla u) = 0, \qquad \mathbf{x} \in \Omega, \tag{4.1}$$

$$u = g, \qquad \mathbf{x} \in \partial\Omega, \tag{4.2}$$

where Ω is a bounded domain in $\mathbb{R}^2$ with smooth boundary $\partial\Omega$ or piecewise smooth as required by our regularity results. We now make the following assumptions on the coefficients a_i, forcing function and f and the boundary function g. We assume that g can be extended to Ω to be in $H^{5/2}(\Omega)$ and there exists a unique weak solution u of (4.1)-(4.2) such that $u \in H^{5/2}(\Omega)$. The functions $f,\ a_i : \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R},\ i = 1,\ 2$, are twice continuously differentiable functions with all the derivatives through the second order are bounded. Further, assume that the matrix $[a^{ij}(\mathbf{x}, u, \mathbf{z})] = \left[\frac{\partial a_i}{\partial z_j}\right]_{i,j=1,2}$ is symmetric and if $\lambda(\mathbf{x}, u, \mathbf{z})$, $\Lambda(x, u, \mathbf{z})$ are minimum and maximum eigenvalues of the matrix $[a^{ij}]$, then for all $\xi \in \mathbb{R}^2 - \{0\}$ and for all $(\mathbf{x}, u, \mathbf{z}) \in \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^2$

$$0 < \lambda(\mathbf{x}, u, \mathbf{z})|\xi|^2 \leq a^{ij}(\mathbf{x}, u, \mathbf{z})\xi_i\xi_j \leq \Lambda(x, u, \mathbf{z})|\xi|^2. \tag{4.3}$$

Finally, assume that if $\|u\|_{W^1_\infty(\Omega)} \leq \alpha$, then there is a positive constant C_α such that

$$0 < C_\alpha \leq \lambda(\mathbf{x}, u, \nabla u). \tag{4.4}$$

Here, onwards we do not specify the dependence of the functions $a_i,\ f$ on $\mathbf{x}$ and we set $\mathbf{a} = (a_1, a_2)$. Note that we define $f_{\mathbf{z}}$ (or $\mathbf{a}_{\mathbf{z}}$) by the partial derivative of $f(u, \mathbf{z})$(or $\mathbf{a}(u, \mathbf{z})$) with respect to its second argument.

In our subsequent analysis, we use the following integral form of the Taylor's formula for $(v, \mathbf{p}) \in V \times \mathbf{W}$ in terms of $(u, \mathbf{q}) \in V \times \mathbf{W}$:

$$\begin{aligned} f(v, \mathbf{p}) - f(u, \mathbf{q}) &= -f_u(u, \mathbf{q})(u - v) - f_{\mathbf{q}}(u, \mathbf{q})(\mathbf{q} - \mathbf{p}) + \tilde{R}_f(u - v, \mathbf{q} - \mathbf{p}), \\ &= -\tilde{f}_u(u, \mathbf{q})(u - v) - \tilde{f}_{\mathbf{q}}(u, \mathbf{q})(\mathbf{q} - \mathbf{p}). \end{aligned} \tag{4.5}$$

with $v^t = u + t(v - u)$, $\mathbf{p}^t = \mathbf{q} + t(\mathbf{p} - \mathbf{q})$, the reminder terms in (4.5) are given by

$$\tilde{f}_u(u, \mathbf{q}) = \int_0^1 f_u(v^t, \mathbf{p}^t)dt, \qquad \tilde{f}_{\mathbf{q}}(u, \mathbf{q}) = \int_0^1 f_{\mathbf{q}}(v^t, \mathbf{p}^t)dt$$

and

$$\begin{aligned} \tilde{R}_f(u - v, \mathbf{q} - \mathbf{p}) &= \tilde{f}_{uu}(v, \mathbf{p})(u - v)^2 + (\mathbf{q} - \mathbf{p})^T \tilde{f}_{\mathbf{qq}}(v, \mathbf{p})(\mathbf{q} - \mathbf{p}) \\ &\quad + 2\tilde{f}_{u\mathbf{q}}(v, \mathbf{p}) \cdot (\mathbf{q} - \mathbf{p})(u - v), \end{aligned}$$

where

$$\tilde{f}_{uu}(v,\mathbf{p}) = \int_0^1 (1-t) f_{uu}(v^t,\mathbf{p}^t)dt,$$
$$\tilde{f}_{\mathbf{qq}}(v,\mathbf{p}) = \int_0^1 (1-t) f_{\mathbf{qq}}(v^t,\mathbf{p}^t)dt,$$
$$\tilde{f}_{u\mathbf{q}}(v,\mathbf{p}) = \int_0^1 (1-t) f_{u\mathbf{q}}(v^t,\mathbf{p}^t)dt.$$

Similarly, we note that (4.5) can be modified for vector valued function $\mathbf{a} = (a_1, a_2)$ as follows :

$$\begin{aligned}\mathbf{a}(v,\mathbf{p}) - \mathbf{a}(u,\mathbf{q}) &= -\mathbf{a}_u(u,\mathbf{q})(u-v) - \mathbf{a}_{\mathbf{q}}(u,\mathbf{q})(\mathbf{q}-\mathbf{p}) + \tilde{R}_{\mathbf{a}}(u-v,\mathbf{q}-\mathbf{p}) \\ &= -\tilde{\mathbf{a}}_u(u,\mathbf{q})(u-v) - \tilde{\mathbf{a}}_{\mathbf{q}}(u,\mathbf{q})(\mathbf{q}-\mathbf{p}), \qquad (4.6)\end{aligned}$$

where

$$\tilde{\mathbf{a}}_u(u,\mathbf{q}) = \int_0^1 \mathbf{a}_u(v^t,\mathbf{p}^t)dt, \qquad \tilde{\mathbf{a}}_{\mathbf{q}}(u,\mathbf{q}) = \int_0^1 \mathbf{a}_{\mathbf{q}}(v^t,\mathbf{p}^t)dt$$

and

$$\tilde{R}_{\mathbf{a}}(u-v,\mathbf{q}-\mathbf{p}) = \left(\tilde{R}_{a_1}(u-v,\mathbf{q}-\mathbf{p}), \tilde{R}_{a_2}(u-v,\mathbf{q}-\mathbf{p})\right).$$

Since $\mathbf{a}$ and f are twice continuously differentiable functions with all the derivatives through the second order are bounded, we note that $\tilde{\mathbf{a}}_u$, $\tilde{\mathbf{a}}_{\mathbf{q}}$, $\tilde{\mathbf{a}}_{uu}$, $\tilde{\mathbf{a}}_{\mathbf{qq}}$, $\tilde{\mathbf{a}}_{u\mathbf{q}}$, $\tilde{\mathbf{a}}_{\mathbf{q}u}$, $\tilde{f}_u$, $\tilde{f}_{\mathbf{q}}$, $\tilde{f}_{uu}$, $\tilde{f}_{\mathbf{qq}}$, $\tilde{f}_{\mathbf{q}u}$ and $\tilde{f}_{u\mathbf{q}} \in L^\infty(\Omega \times \mathbb{R} \times \mathbb{R}^2)$. We now denote C_a by

$$C_a = \max\{\|\mathbf{a}\|_{W^2_\infty(\Omega\times\mathbb{R}\times\mathbb{R}^2)},\ \|f\|_{W^2_\infty(\Omega\times\mathbb{R}\times\mathbb{R}^2)}\}. \qquad (4.7)$$

REMARK 4.2.1 *For our subsequent analysis, it is sufficient to assume that* $\mathbf{a}$ *and* f *are locally bounded. In fact, it is enough to assume that* $\mathbf{a}$ *and* f *along with its derivatives are bounded in a ball around* u*, see Remark 4.2.3.*

DG methods. For given w and $v \in H^2(\Omega, \mathcal{T}_h)$, we define the form $B(w,v)$ as

$$\begin{aligned}B(w,v) &= \sum_{i=1}^{N_h} \int_{K_i} \mathbf{a}(w,\nabla w)\cdot\nabla v\ dx + \sum_{i=1}^{N_h} \int_{K_i} f(w,\nabla w)v\ dx + \mathcal{J}^\sigma(w,\ v) \\ &- \sum_{e_k\in\Gamma_I} \int_{e_k} \{\mathbf{a}(w,\nabla w)\cdot\nu\}[v]\ ds + \theta \sum_{e_k\in\Gamma_I} \int_{e_k} \{\mathbf{a}_{\mathbf{z}}(w,\nabla w)\nabla v\cdot\nu\}[w]\ ds\end{aligned}$$

(4.8)

$$- \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}(g, \nabla w) \cdot \nu v \; ds + \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}_{\mathbf{z}}(g, \nabla w) \nabla v \cdot \nu (w - g) \; ds$$

$$- \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{ f_{\mathbf{z}}(w, \nabla w) v \cdot \nu \} [w] \; ds \tag{4.9}$$

where $\theta \in [-1, 1]$, $\gamma = 0$ when $\theta \in (-1, 1]$, and $\gamma = 1$ when $\theta = -1$. Further, let

$$L(v) = \sum_{e_k \in \Gamma_\partial} \int_{e_k} \sigma_k \frac{p_k^2}{|e_k|} g v \; ds,$$

We now define the *hp*-discontinuous Galerkin method for the problem (4.1)-(4.2) as : Find $u_h \in V_h$ such that

$$B(u_h, v_h) = L(v_h) \qquad \forall v_h \in V_h. \tag{4.10}$$

REMARK 4.2.2 *Note that in the formulation of $B(\cdot, \cdot)$, there are three special terms that is fifth, seventh and eight terms appearing on the right-hand side of (4.9). In fact, this weak formulation boils down to the earlier weak formulations appeared in the literature for the following cases:*

1. *Linear case, i.e.,* $\mathbf{a}(u, \nabla u) = \nabla u$ *and* $f(u, \nabla u) = 0$, *see [61].*
   Since $\mathbf{a}_{\mathbf{z}}(u, \nabla u) = I_{2\times 2}$, *where* $I_{2\times 2}$ *is the identity matrix,* $\mathbf{a}_{\mathbf{z}}(u, \nabla u) \nabla v \cdot \nu = \nabla v \cdot \nu$.

2. *Quasilinear case, i.e., when* $\mathbf{a}(u, \nabla u) = a(u) \nabla u$ *and* $f(u, \nabla u) = 0$, *see [41]:*
   we note that $\mathbf{a}_{\mathbf{z}}(u, \nabla u) \nabla v \cdot \nu = a(u) I_{2\times 2} \nabla v \cdot \nu = a(u) \nabla v \cdot \nu$.

Further for the linearized problem (4.18) (see, in the next subsection) of (4.10), these special terms help us to preserve the unconditional stability of the method when $\theta = 1$, and, on the other hand, they also preserve the adjoint consistency of the method, when $\theta = -1$.

4.2.1 Existence and uniqueness of the Discrete Problem

We note that the solution $u \in H^2(\Omega)$ of the problem (4.1)-(4.2) satisfies $[u] = 0$ on each $e_k \in \Gamma_I$ and

$$\sum_{i=1}^{N_h} \int_{K_i} \mathbf{a}(u, \nabla u) \cdot \nabla v \, dx + \sum_{i=1}^{N_h} \int_{K_i} f(u, \nabla u) v \, dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a}(u, \nabla u) \cdot \nu\}[v] \, ds$$
$$- \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}(g, \nabla u) \cdot \nu v \, ds + \mathcal{J}^\sigma(u, v) = L(v) \quad \forall v \in H^2(\Omega, \mathcal{T}_h). \quad (4.11)$$

With $v = v_h \in V_h \subset H^2(\Omega, \mathcal{T}_h)$ in (4.11), we subtract (4.10) from (4.11). Then add the following terms

$$\theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a}_\mathbf{z}(u, \nabla u) \nabla v_h \cdot \nu\}[u - u_h] ds + \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}_\mathbf{z}(g, \nabla u) \nabla v_h \cdot \nu (u - u_h) ds$$
$$- \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{f_\mathbf{z}(u, \nabla u) v_h \cdot \nu\}[u - u_h] \, ds$$

to the both sides of the resulting equation. Hence, we arrive at

$$\sum_{i=1}^{N_h} \int_{K_i} (\mathbf{a}(u, \nabla u) - \mathbf{a}(u_h, \nabla u_h)) \cdot \nabla v_h dx + \sum_{i=1}^{N_h} \int_{K_i} (f(u, \nabla u) - f(u_h, \nabla u_h)) v_h dx$$
$$- \sum_{e_k \in \Gamma_I} \int_{e_k} \{(\mathbf{a}(u, \nabla u) - \mathbf{a}(u_h, \nabla u_h)) \cdot \nu\}[v_h] ds - \sum_{e_k \in \Gamma_\partial} \int_{e_k} (\mathbf{a}(g, \nabla u) - \mathbf{a}(g, \nabla u_h)) \cdot \nu v_h \, ds$$
$$+ \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a}_\mathbf{z}(u, \nabla u) \nabla v_h \cdot \nu\}[u - u_h] ds + \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}_\mathbf{z}(g, \nabla u) \nabla v_h \cdot \nu (u - u_h) ds$$
$$+ \mathcal{J}^\sigma(u - u_h, v_h) - \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{f_\mathbf{z}(u, \nabla u) v_h \cdot \nu\}[u - u_h] \, ds =$$
$$\theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a}_\mathbf{z}(u, \nabla u) \nabla v_h \cdot \nu\}[u - u_h] ds + \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a}_\mathbf{z}(u_h, \nabla u_h) \nabla v_h \cdot \nu\}[u_h] ds$$
$$+ \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}_\mathbf{z}(g, \nabla u) \nabla v_h \cdot \nu (u - u_h) ds + \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a}_\mathbf{z}(g, \nabla u_h) \nabla v_h \cdot \nu (u_h - g) ds$$
$$- \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{f_\mathbf{z}(u, \nabla u) v_h \cdot \nu\}[u - u_h] \, ds - \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{f_\mathbf{z}(u, \nabla u) v_h \cdot \nu\}[u_h] \, ds. \quad (4.12)$$

Using Taylor series expansions (4.5)-(4.6), we rewrite (4.12) as

$$\sum_{i=1}^{N_h}\int_{K_i}\mathbf{a_z}(u,\nabla u)\nabla(u-u_h)\cdot\nabla v_h dx+\sum_{i=1}^{N_h}\int_{K_i}\mathbf{a}_u(u,\nabla u)(u-u_h)\cdot\nabla v_h dx$$
$$-\sum_{e_k\in\Gamma_I}\int_{e_k}\{\mathbf{a_z}(u,\nabla u)\nabla(u-u_h)\cdot\nu\}[v_h]ds-\sum_{e_k\in\Gamma_I}\int_{e_k}\{\mathbf{a}_u(u,\nabla u)(u-u_h)\cdot\nu\}[v_h]ds$$
$$-\sum_{e_k\in\Gamma_\partial}\int_{e_k}\mathbf{a_z}(g,\nabla u)\nabla(u-u_h)\cdot\nu v_h\ ds+\theta\sum_{e_k\in\Gamma_I}\int_{e_k}\{\mathbf{a_z}(u,\nabla u)\nabla v_h\cdot\nu\}[u-u_h]ds$$
$$+\theta\sum_{e_k\in\Gamma_\partial}\int_{e_k}\mathbf{a_z}(g,\nabla u)\nabla v_h\cdot\nu(u-u_h)ds+\sum_{i=1}^{N_h}\int_{K_i}f_{\mathbf{z}}(u,\nabla u)\cdot\nabla(u-u_h)v_h dx$$
$$+\sum_{i=1}^{N_h}\int_{K_i}f_u(u,\nabla u)(u-u_h)v_h dx+\mathcal{J}^\sigma(u-u_h,v_h)$$
$$-\gamma\sum_{e_k\in\Gamma_I}\int_{e_k}\{f_{\mathbf{z}}(u,\nabla u)v_h\cdot\nu\}[u-u_h]\ ds=\mathcal{N}(u,u_h;v_h),\tag{4.13}$$

where

$$\begin{aligned}\mathcal{N}(u,u_h;v_h) &= \sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_{\mathbf{a}}(u-u_h,\nabla(u-u_h))\cdot\nabla v_h dx\\ &- \sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{R}_{\mathbf{a}}(u-u_h,\nabla(u-u_h))\cdot\nu\}[v_h]\ ds\\ &+ \sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_f(u-u_h,\nabla(u-u_h))v_h dx\\ &- \sum_{e_k\in\Gamma_\partial}\int_{e_k}\nabla(u-u_h)^T\tilde{\mathbf{a}}_{\mathbf{zz}}(g,\nabla u)\nabla(u-u_h)\cdot\nu v_h\ ds\\ &+ \theta\sum_{e_k\in\Gamma_I}\int_{e_k}\{(\mathbf{a_z}(u,\nabla u)-\mathbf{a_z}(u_h,\nabla u_h))\nabla v_h\cdot\nu\}[u-u_h]ds\\ &+ \theta\sum_{e_k\in\Gamma_\partial}\int_{e_k}(\mathbf{a_z}(g,\nabla u)-\mathbf{a_z}(g,\nabla u_h))\nabla v_h\cdot\nu(g-u_h)ds\\ &- \gamma\sum_{e_k\in\Gamma_I}\int_{e_k}\{(f_{\mathbf{z}}(u,\nabla u)-f_{\mathbf{z}}(u_h,\nabla u_h)v_h\cdot\nu\}[u-u_h]\ ds.\end{aligned}\tag{4.14}$$

We now introduce some notations which simplifies the expressions in (4.13). Set

$$A(u)=\mathbf{a_z}(u,\nabla u),\quad \mathbf{b}(u)=\mathbf{a}_u(u,\nabla u),\quad \mathbf{f}(u)=f_{\mathbf{z}}(u,\nabla u),\quad F(u)=f_u(u,\nabla u).\tag{4.15}$$

For given $\psi,\ w$ and $v \in H^2(\Omega, \mathcal{T}_h)$, we now define the following forms

$$\begin{aligned} a(\psi; w, v) &= \sum_{i=1}^{N_h} \int_{K_i} A(\psi)\nabla w \cdot \nabla v dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{A(\psi)\nabla w \cdot \nu\}[v] ds \\ &- \sum_{e_k \in \Gamma_\partial} \int_{e_k} A(\psi)\nabla w \cdot \nu v \ ds + \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{A(\psi)\nabla v \cdot \nu\}[w] ds \\ &+ \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} A(\psi)\nabla v \cdot \nu w ds + \mathcal{J}^\sigma(w, v) \end{aligned}$$

and

$$\begin{aligned} b(\psi; w, v) &= \sum_{i=1}^{N_h} \int_{K_i} \mathbf{b}(\psi) w \cdot \nabla v dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{b}(\psi) w \cdot \nu\}[v] ds \\ &+ \sum_{i=1}^{N_h} \int_{K_i} \mathbf{f}(\psi) \cdot \nabla w v dx + \sum_{i=1}^{N_h} \int_{K_i} F(\psi) w v dx - \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{f}(\psi) v_h \cdot \nu\}[w] \ ds. \end{aligned}$$

Note that for fixed ψ, the form $a(\psi; \cdot, \cdot)$ and the form $b(\psi; \cdot, \cdot)$ are bilinear. The identity (4.13) takes the form

$$a(u; u - u_h, v_h) + b(u; u - u_h, v_h) = \mathcal{N}(u, u_h; v_h) \quad \forall v_h \in V_h. \tag{4.16}$$

Using interpolant $I_h u$ of u, we rewrite (4.16) for all $v_h \in V_h$ as

$$a(u; I_h u - u_h, v_h) + b(u; I_h u - u_h, v_h) = a(u; \eta, v_h) + b(u; \eta, v_h) + \mathcal{N}(u, u_h; v_h), \tag{4.17}$$

where $\eta = I_h u - u$. Now, for proving the existence of a solution u_h to the problem (4.10), we put the problem in the fixed point formulation and hence, define a map $S_h : V_h \to V_h$ as follows. For a given $y \in V_h$, find $S_h(y) = u_y \in V_h$ such that for all $v_h \in V_h$

$$a(u; I_h u - u_y, v_h) + b(u; I_h u - u_y, v_h) = a(u; \eta, v_h) + b(u; \eta, v_h) + \mathcal{N}(u, y; v_h). \tag{4.18}$$

Below, in Theorem 4.2.1, we have shown that the map S_h is well-defined. Note that the existence of a fixed point of the map S_h is equivalent to the existence of a solution to the discrete problem (4.10). The following lemmas are useful for our subsequent analysis. Using the techniques of [58] and [66], it is easy to prove the following 3 lemmas that is Lemma 4.2.1, Lemma 4.2.2 and Lemma 4.2.3. Hence, the proofs are omitted.

LEMMA 4.2.1 *(Gärding Type Inequality) Let* $0 < \sigma_0 \le \sigma_k \le \sigma_m$. *Let* $\sigma_0 \ge C(C_T, C_a)$, *in case* $\theta \in [-1, 1)$, *and* $\sigma_0 > 0$, *when* $\theta = 1$. *Then, there are two positive constants* C_1 *and* C_2 *such that*

$$a(u; v_h, v_h) + b(u; v_h, v_h) \ge C_1 \, |||v_h|||^2 - C_2 \|v_h\|^2 \quad \forall v_h \in V_h. \tag{4.19}$$

LEMMA 4.2.2 *Let* $\sigma_k \le \sigma_m$ *and* $w \in H^2(\Omega, \mathcal{T}_h)$. *Then, there is a positive constant* C *such that*

$$|a(u; w, v_h)| + |b(u; w, v_h)| \le C|||w||| \, |||v_h||| \quad \forall v_h \in V_h. \tag{4.20}$$

LEMMA 4.2.3 *Let* $\sigma_k \le \sigma_m$ *and* $\phi \in H^{s_i}(K_i)$, $s_i \ge 2$, $1 \le i \le N_h$. *Then, there is a positive constant* C *such that*

$$|||\phi - I_h\phi||| \le C \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i - 2}}{p_i^{2s_i - 3}} \|\phi\|^2_{H^{s_i}(K_i)} \right)^{1/2}, \tag{4.21}$$

where $\mu_i = min\{s_i, p_i + 1\}$.

LEMMA 4.2.4 *Assume that* $(f_u(u, \nabla u) - \nabla \cdot \mathbf{a}_u(u, \nabla u)) \ge 0$ *and* $0 < h < h_0 < 1$. *Then, for given* $\xi \in L^2(\Omega)$, *there is a unique* $\phi_h \in V_h$ *satisfying*

$$a(u; v_h, \phi_h) + b(u; v_h, \phi_h) = (\xi, v_h)\,. \tag{4.22}$$

Moreover, there is a positive constant C *such that*

$$|||\phi_h||| \le C\|\xi\|. \tag{4.23}$$

Proof. An appeal to Lemma 4.2.1 yields

$$\begin{aligned} C_1|||\phi_h|||^2 - C_2\|\phi_h\|^2 &\le Ca(u; \phi_h, \phi_h) + b(u, \phi_h, \phi_h) \\ &= (\xi, \phi_h) \\ &\le C\|\xi\| \, \|\phi\|. \end{aligned}$$

Therefore, we obtain

$$|||\phi_h||| \le C_1\|\phi_h\| + C_2\|\xi\|. \tag{4.24}$$

In order to estimate $\|\phi_h\|$, we appeal to the Aubin-Nitsche duality argument. We now consider the following auxiliary problem :

$$\begin{aligned} -\nabla\cdot(A(u)\nabla\psi+\mathbf{b}(u)\psi)+\mathbf{f}(u)\cdot\nabla\psi+F(u)\psi &= \phi_h \quad \text{on } \Omega, \\ \psi &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{4.25}$$

Using the form of F and $\mathbf{b}$, we note that $F(u)\geq\nabla\cdot\mathbf{b}(u)$. Hence, it follows from [40, Lemma 9.17] that there exists a unique solution $\psi\in H^2(\Omega)$ to the problem (4.25) satisfying the elliptic regularity

$$\|\psi\|_{H^2(\Omega)}\leq C\|\phi_h\|. \tag{4.26}$$

Note that $[\psi]=0$ on each $e_k\in\Gamma$. Now, form an L^2-inner product between (4.25) and ϕ_h. Then, apply integration by parts and use Lemma 4.2.3 to arrive at

$$\begin{aligned} \|\phi_h\|^2 &= a(u;\psi,\phi_h)+b(u;\psi,\phi_h) \\ &= a(u;\psi-I_h\psi,\phi_h)+b(u;\psi-I_h\psi,\phi_h)+(\xi,I_h\psi) \\ &\leq C|||\psi-I_h\psi|||\;|||\phi|||+\|\xi\|\;\|I_h\psi\| \\ &\leq C\left(h|||\phi_h|||+\|\xi\|\right)\|\psi\|_{H^2(\Omega)}. \end{aligned}$$

A use of the elliptic regularity (4.26) yields

$$\|\phi_h\|\leq Ch|||\phi_h|||+C\|\xi\|. \tag{4.27}$$

Substituting (4.27) in (4.24), we obtain for sufficiently small h the required estimate (4.23). Being a finite dimensional problem, existence the solution of ϕ_h to the problem (4.22) follows from the uniqueness. Uniqueness now follows trivially from (4.23) and this completes the rest of the proof. ■

THEOREM 4.2.1 *Let $0<h<h_0<1$. Then, for given $y\in V_h$, there is a unique solution $u_y\in V_h$ to (4.18), that is $S_h(y)=u_y$.*

Proof. Suppose that for given y, there are two distinct solutions u_y^1 and u_y^2 for the problem (4.18). Then, it is easy to check that

$$a(u;u_y^1-u_y^2,v_h)+b(u;u_y^1-u_y^2,v_h)=0\quad\forall v_h\in V_h. \tag{4.28}$$

We set $\xi = u_y^1 - u_y^2$ and $v_h = u_y^1 - u_y^2$ in (4.22) to obtain

$$\|u_y^1 - u_y^2\| = a(u; u_y^1 - u_y^2, \phi_h) + b(u; u_y^1 - u_y^2, \phi_h). \tag{4.29}$$

Now, we set $v_h = \phi_h$ in (4.28) to complete the rest of the proof. ■

We show that the map S_h maps from a ball $\mathcal{O}_\delta(I_h u) \subset V_h$ to itself and it is continuous in the ball. Our choice of ball is

$$\mathcal{O}_\delta(I_h u) = \{y \in V_h : |||I_h u - y||| \leq \delta\}, \quad \text{where } \delta = h^{-\epsilon}|||\eta|||,\ 0 < \epsilon < 1/4. \tag{4.30}$$

Since $u \in H^{5/2}(\Omega)$ and $p_i \geq 2,\ 1 \leq i \leq N_h$, we note that

$$\delta \leq C_u h^{-\epsilon} \left(\max_{1 \leq i \leq N_h} \frac{h_i^{3/2}}{p_i} \right), \tag{4.31}$$

where

$$C_u = \|u\|_{H^{5/2}(\Omega)}, \tag{4.32}$$

Further, there exists $0 < h_0 < 1$ such that for $0 < h < h_0 < 1$, δ can be made less than 1, that is $\delta < 1$.

REMARK 4.2.3 *In our subsequent analysis, we only need the bounds of* $\frac{\partial^l a}{\partial u^l}(\mathbf{x}, y, \nabla y)$, $\mathbf{a_z}(\mathbf{x}, y, \nabla y)$ *and* $\mathbf{a_{zz}}(\mathbf{x}, y, \nabla y)$ *for* $\mathbf{x} \in \bar{\Omega},\ y \in \mathcal{O}_\delta(I_h u),\ l = 0, 1, 2$. *If* $u \in H^{5/2}(\Omega)$, *we note that asymptotically only the values of* $y(\mathbf{x}) \in [m_u - \delta^*, M_u + \delta^*]$, *where* $0 < \delta^* < 1$, $m_u = inf\ \{u(\mathbf{x}) : \mathbf{x} \in \bar{\Omega}\}$ *and* $M_u = sup\ \{u(\mathbf{x}) : \mathbf{x} \in \bar{\Omega}\}$ *are considered to derive the bounds. Similarly, asymptotically the values of* $\frac{\partial y}{\partial x_i}(\mathbf{x}) \in [m_u^1 - \delta^*, M_u^1 + \delta^*]$, *where* $m_u^1 = inf\ \{\nabla u(\mathbf{x}) : \mathbf{x} \in \bar{\Omega}\}$ *and* $M_u^1 = sup\ \{\nabla u(\mathbf{x}) : \mathbf{x} \in \bar{\Omega}\}$ *are considered. To be more precise, the terms* $\tilde{\mathbf{a}}_u(\mathbf{x}, y, \nabla y)$ *and* $\tilde{\mathbf{a}}_{uu}(\mathbf{x}, y, \nabla y),\ y \in \mathcal{O}_\delta(I_h u)$ *can be estimated as follows. Since* $y \in \mathcal{O}_\delta(I_h u)$, *where* $\delta = h^{-\epsilon}|||u - I_h u|||_+,\ 0 < \epsilon \leq 1/4$, *using the inverse inequality (1.20), Lemma 1.2.6 and Lemma 1.2.1, we find that*

$$\begin{aligned}
\|y - u\|_{W_\infty^1(\Omega)} &\leq \|y - I_h u\|_{W_\infty^1(\Omega)} + \|I_h u - u\|_{W_\infty^1(\Omega, \mathcal{T}_h)} \\
&\leq C \left(\max_{1 \leq i \leq N_h} \frac{p_i}{h_i} \right) \|y - I_h u\|_{1,h} + \|I_h u - u\|_{W_\infty^1(\Omega, \mathcal{T}_h)} \\
&\leq C \left(\max_{1 \leq i \leq N_h} \frac{p_i}{h_i} \right) |||z - I_h u||| + \|I_h u - u\|_{W_\infty^1(\Omega, \mathcal{T}_h)} \\
&\leq C h^{-\epsilon} \left(\max_{1 \leq i \leq N_h} \frac{p_i}{h_i} \right) |||u - I_h u|||_+ + \|I_h u - u\|_{W_\infty^1(\Omega, \mathcal{T}_h)}
\end{aligned}$$

$$
\begin{aligned}
&\leq Ch^{-\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\left(\sum_{i=1}^{N_h}\frac{h_i^3}{p_i^2}\|u\|^2_{H^{5/2}(K_i)}\right)^{1/2}+C\frac{h^{1/2}}{p^{1/2}}\|u\|_{H^{5/2}(\Omega)}\\
&\leq Ch^{-\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\|u\|_{H^{5/2}(\Omega)}+C\frac{h^{1/2}}{p^{1/2}}\|u\|_{H^{5/2}(\Omega)}\\
&\leq Ch^{1/2-\epsilon}\|u\|_{H^{5/2}(\Omega)}. \qquad (4.33)
\end{aligned}
$$

Therefore, for sufficiently small h, $\|y\|_{W^1_\infty(\Omega)}\leq\delta^*+\|u\|_{W^1_\infty(\Omega)}$, where $0<\delta^*<1$. Now, since the nonlinear functions $\mathbf{a}_u$, $\mathbf{a}_{uu}$, $\mathbf{a}_{\mathbf{z}}$ and $\mathbf{a}_{\mathbf{zz}}$ are continuous, they map the compact set $[m_u-\delta^*,M_u+\delta^*]\times[m_u^1-\delta^*,M_u^1+\delta^*]$ into a compact set. Hence, the results in the subsequent Sections remain valid when $\mathbf{a}(\mathbf{x},y,\nabla y)$, $\mathbf{a}_u(\mathbf{x},y,\nabla y)$, $\mathbf{a}_{uu}(\mathbf{x},y,\nabla y)$, $\mathbf{a}_{\mathbf{z}}(\mathbf{x},y,\nabla y)$ and $\mathbf{a}_{\mathbf{zz}}(\mathbf{x},y,\nabla y)$ are bounded for bounded $u\in W^1_\infty(\Omega)$. Similar arguments can also be applied to f.

LEMMA 4.2.5 *Let $y\in\mathcal{O}_\delta(I_hu)$ and $v_h\in V_h$. Set $\zeta=u-y$. Then, there is a positive constant C such that*

$$
|\sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_{\mathbf{a}}(\zeta,\nabla\zeta)\cdot\nabla v_h dx| \leq C_a\|\zeta\|_{W^1_4(\Omega)}\|\zeta\|_{W^1_2(\Omega)}|v_h|_{W^1_4(\Omega,\mathcal{T}_h)}.
$$

Proof. To prove the inequality of the lemma, we first expand it as :

$$
\begin{aligned}
\sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_{\mathbf{a}}(\zeta,\nabla\zeta)\cdot\nabla v_h dx &= \sum_{i=1}^{N_h}\int_{K_i}\tilde{\mathbf{a}}_{uu}(y,\nabla y)\zeta^2\cdot\nabla v_h dx\\
&+2\sum_{i=1}^{N_h}\int_{K_i}\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\zeta\nabla\zeta\cdot\nabla v_h dx+\sum_{i=1}^{N_h}\int_{K_i}\nabla\zeta^T\tilde{\mathbf{a}}_{\mathbf{zz}}(y,\nabla y)\nabla\zeta\cdot\nabla v_h dx. \qquad (4.34)
\end{aligned}
$$

Now, using Hölder's inequality and Lemma 1.2.6, the first term on the right-hand side of (4.34) is estimated as

$$
\begin{aligned}
|\sum_{i=1}^{N_h}\int_{K_i}\tilde{\mathbf{a}}_{uu}(y,\nabla y)\zeta^2\cdot\nabla v_h dx| &\leq C_a\sum_{i=1}^{N_h}\|\zeta\|_{L^4(K_i)}\|\zeta\|_{L^2(K_i)}\|\nabla v_h\|_{L^4(K_i)}\\
&\leq C_a\|\zeta\|_{L^4(\Omega)}\|\zeta\|_{L^2(\Omega)}|v_h|_{W^1_4(\Omega,\mathcal{T}_h)}. \qquad (4.35)
\end{aligned}
$$

Then, for the second term on the right hand side of (4.34), we use Hölder's inequality and the inverse inequality to obtain

$$
\begin{aligned}
|\sum_{i=1}^{N_h}\int_{K_i}\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\zeta\nabla\zeta\cdot\nabla v_h dx| &\leq C_a\sum_{i=1}^{N_h}\|\zeta\|_{L^r(K_i)}\|\nabla\zeta\|_{L^2(K_i)}\|\nabla v_h\|_{L^q(K_i)}\\
&\leq C_a\|\zeta\|_{L^4(\Omega)}|\zeta|_{W^1_2(\Omega,\mathcal{T}_h)}|v_h|_{W^1_4(\Omega,\mathcal{T}_h)}. \qquad (4.36)
\end{aligned}
$$

Finally, for the third term on the right-hand side of (4.34), we use Hölder's inequality to obtain

$$\begin{aligned} |\sum_{i=1}^{N_h}\int_{K_i}\nabla\zeta^T\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\nabla\zeta\cdot\nabla v_h dx| &\leq C_a\sum_{i=1}^{N_h}\|\nabla\zeta\|^2_{L^4(K_i)}\|\nabla\zeta\|_{L^2(K_i)}\|\nabla v_h\|_{L^4(K_i)} \\ &\leq C_a|\zeta|_{W_4^1(\Omega,\mathcal{T}_h)}|\zeta\|_{W_2^1(\Omega,\mathcal{T}_h)}|v_h|_{W_4^1(\Omega,\mathcal{T}_h)}. \qquad (4.37)\end{aligned}$$

Now, we substitute (4.35)-(4.37) in (4.34) to complete the proof. ■

LEMMA 4.2.6 *Let $y\in\mathcal{O}_\delta(I_hu)$ and $v_h\in V_h$. Then, there is a positive constant C such that*

$$|\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{R}_{\mathbf{a}}(u-y,\nabla(u-y))\cdot\nu\}[v_h]\ ds|\leq CC_aC_uC_Qh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.38)$$

Proof. Let $\zeta=u-y=\eta+\xi$, where $\eta=u-I_hu$ and $\xi=I_hu-y$. Then, we expand the term on the left hand side of the inequality of (4.38) as

$$\begin{aligned}\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{R}_{\mathbf{a}}(\zeta,\nabla\zeta)\cdot\nu\}[v_h]\ ds &= \sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{\mathbf{a}}_{uu}(y,\nabla y)\zeta^2\cdot\nu\}[v_h]\ ds \qquad (4.39)\\ &+ 2\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\zeta\nabla\zeta\cdot\nu\}[v_h]\ ds \\ &+ \sum_{e_k\in\Gamma_I}\int_{e_k}\{\nabla\zeta^T\tilde{\mathbf{a}}_{\mathbf{zz}}(y,\nabla y)\nabla\zeta\cdot\nu\}[v_h]\ ds.\end{aligned}$$

For the first term on the right hand side of (4.39), we use Hölder's inequality, the trace inequality (1.14), the Cauchy-Schwarz inequality and Lemma 1.2.6 to obtain

$$\begin{aligned}|\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{\mathbf{a}}_{uu}(y,\nabla y)\zeta^2\cdot\nu\}[v_h]\ ds| &\leq CC_a\sum_{e_k\in\Gamma_I}\frac{|e_k|^{1/2}}{p_k}\|\zeta\|^2_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2ds\right)^{1/2}\\ &\leq CC_a\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{h_i^{1/2}}{p_i}\|\zeta\|^2_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2ds\right)^{1/2}\\ &\leq CC_a\sum_{i=1}^{N_h}\frac{h_i^{1/4}}{p_i}\left(\|\zeta\|^4_{L^4(K_i)}+h_i\|\zeta\|^3_{L^6(K_i)}\|\nabla\zeta\|_{L^2(K_i)}\right)^{\frac{1}{2}}\\ &\qquad\sum_{e_k\subset\partial K_i}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2ds\right)^{1/2}\\ &\leq CC_ah^{1/4}|||\zeta|||^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\ &\leq CC_ah^{1/4}\delta^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\ &\leq CC_aC_uh^{3/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.40)\end{aligned}$$

We now split the second term on the right-hand side of (4.39) as

$$
\begin{aligned}
|\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\zeta\nabla\zeta\cdot\nu\}[v_h]\ ds| &= C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\nabla\eta||[v_h]|\ ds\\
&+C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\nabla\xi||[v_h]|\ ds + C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\xi\nabla\eta||[v_h]|\ ds\\
&+C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\xi\nabla\xi||[v_h]|\ ds. \qquad (4.41)
\end{aligned}
$$

For the first term on the right hand side of (4.41), we use Hölder' inequality, the trace inequality (1.14), Lemma 1.2.1 and Lemma 1.2.6 to obtain

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\nabla\eta||[v_h]| &\leq \sum_{e_k\in\Gamma_I}\frac{|e_k|^{1/2}}{p_k}\|\eta\|_{L^4(e_k)}\|\nabla\eta\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{1}{p_i}\left(\|\eta\|^4_{L^4(K_i)}+h_i\|\eta\|^3_{L^6(K_i)}\|\nabla\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\quad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\leq \sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{1}{p_i}\left(\|\eta\|^4_{L^4(K_i)}+h_i\|\eta\|^3_{L^6(K_i)}\|\nabla\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\quad\left(\frac{h_i^2}{p_i^2}\|u\|^4_{H^2(K_i)}+h_i\frac{h_i}{p_i}\|u\|^4_{H^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h^{1/2}}{p^{3/2}}+\frac{h^{1/2}}{p^{5/4}}\right)|||\eta|||\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_u h^{1/2+\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.42)
\end{aligned}
$$

For the second term on the right hand side of (4.41), apply Hölder's inequality, the trace inequalities (1.14)-(1.19), Lemma 1.2.1 and Lemma 1.2.6 to find that

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\eta\nabla\xi||[v_h]| &\leq \sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{1/2}}{p_k}\|\eta\|_{L^4(e_k)}\|\nabla\xi\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{1}{p_i^{1/2}}\|\nabla\xi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\qquad\left(\|\eta\|^4_{L^4(K_i)}+h_i\|\eta\|^3_{L^6(K_i)}\|\nabla\eta\|_{L^2(K_i)}\right)^{1/4}
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{p_i^{1/2}}{h_i^{1/2}}\|\nabla\xi\|_{L^2(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\qquad\left(\frac{h_i^6}{p_i^6}\|u\|^4_{H^2(K_i)}+h_i\frac{h_i^5}{p_i^5}\|u\|^4_{H^2(K_i)}\right)^{1/4}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h}{p}+\frac{h}{p^{3/4}}\right)|||\xi|||\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_uh\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.43)
\end{aligned}
$$

Similarly, use Hölder's inequality, (1.23), the trace inequality (1.14) and Lemma 1.2.1 to estimate the third term on the right hand side of (4.41) as

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\xi\nabla\eta||[v_h]| &\leq C\sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{1/2}}{p_k}\|\xi\|_{L^4(e_k)}\|\nabla\eta\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{1}{p_k}\left(\|\xi\|^4_{L^4(K_i)}+h_i\|\xi\|^3_{L^6(K_i)}\|\nabla\xi\|_{L^2(K_i)}\right)^{1/4}\\
&\qquad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{1}{p_k}\left(\frac{h_i^2}{p_i^2}\|u\|^4_{H^2(K_i)}+h_i\frac{h_i}{p_i}\|u\|^4_{H^2(K_i)}\right)^{1/4}\\
&\qquad\left(\|\xi\|^4_{L^4(K_i)}+h_i\|\xi\|^3_{L^6(K_i)}\|\nabla\xi\|_{L^2(K_i)}\right)^{1/4}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\leq C\|u\|_{H^2(\Omega)}\left(\frac{h^{1/2}}{p^{3/2}}+\frac{h^{1/2}}{p^{5/4}}\right)|||\xi|||\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_uh^{1/2}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.44)
\end{aligned}
$$

Similarly for the fourth term on the right-hand side of (4.41), Hölder's inequality with the inverse inequality (1.22), the trace inequalities (1.14)-(1.19) and (1.23) yields

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\xi\nabla\xi||[v_h]| &\leq \sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{1/2}}{p_k}\|\xi\|_{L^4(e_k)}\|\nabla\xi\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{e_k\in\Gamma_I}\frac{|e_k|^{1/4}}{p_k^{1/2}}\|\xi\|_{L^4(e_k)}\|\nabla\xi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\in\partial K_i}\frac{p_i^{1/2}}{h_i^{1/2}}\|\nabla\xi\|_{L^2(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\quad\left(\|\xi\|^4_{L^4(K_i)}+h_i\|\xi\|^3_{L^6(K_i)}\|\nabla\xi\|_{L^2(K_i)}\right)^{1/4}\\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)|||\xi|||^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_u\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\delta^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_uh^{-\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_QC_uh^{1-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.45)
\end{aligned}
$$

To estimate the third term on the right hand of (4.39), we first split it as

$$
\begin{aligned}
|\sum_{e_k\in\Gamma_I}\int_{e_k}\{\nabla\zeta^T\tilde{\mathbf{a}}_{u\mathbf{z}}(y,\nabla y)\nabla\zeta\cdot\nu\}[v_h]\ ds| &= 2C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\eta|^2|[v_h]|\ ds\\
+4C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\eta|\ |\nabla\xi||[v_h]|\ ds &+ C_a\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\xi|^2|[v_h]|\ ds. \qquad (4.46)
\end{aligned}
$$

For the first term on the right-hand side of (4.46), we use Hölder' inequality, the trace inequality (1.14), the inverse inequality (1.22) and Lemma 1.2.1 to obtain

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\eta|^2|[v_h]|ds &\leq C\sum_{e_k\in\Gamma_I}\|\nabla\eta\|_{L^2(e_k)}\|\nabla\eta\|_{L^4(e_k)}\left(\int_{e_k}[v_h]^4ds\right)^{1/4}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{p_k^{1/2}}{|e_k|^{1/2}}\|\nabla\eta\|_{L^2(e_k)}\left(\int_{e_k}[v_h]^2ds\right)^{1/2}\\
&\quad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{1}{p_k^{1/2}}\|\nabla\eta\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\quad\left(\frac{h_i^4}{p_i^4}\|u\|^4_{H^{5/2}(K_i)}+h_i\frac{h_i^3}{p_i^3}\|u\|^4_{H^{5/2}(K_i)}\right)^{1/4}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{|e_k|}{p_k^{5/4}}\|\nabla\eta\|_{L^2(e_k)}\|u\|_{H^{5/2}(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\frac{h^{1/2+\epsilon}}{p^{3/4}}\left(\sum_{e_k\in\Gamma_I}\frac{|e_k|}{p_k}\|\nabla\eta\|^2_{L^2(e_k)}\right)^{1/2}\|u\|_{H^{5/2}(\Omega)}\mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_u h^{1/2+\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.47)
\end{aligned}
$$

For the second term on the right hand side of (4.46), apply Hölder's inequality, trace inequality (1.14)-(1.19), Lemma 1.2.1 and Lemma 1.2.6 to find that

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\eta|\ |\nabla\xi||[v_h]| &\leq \sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{1/2}}{p_k}\|\nabla\eta\|_{L^4(e_k)}\|\nabla\xi\|_{L^4(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{1}{p_i^{1/2}}\|\nabla\xi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\qquad\left(\|\nabla\eta\|^4_{L^4(K_i)}+h_i\|\nabla\eta\|^3_{L^6(K_i)}\|\nabla^2\eta\|_{L^2(K_i)}\right)^{1/4}\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\frac{p_i^{1/2}}{h_i^{1/2}}\|\nabla\xi\|_{L^2(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\\
&\qquad\left(\frac{h_i^4}{p_i^4}\|u\|^4_{H^{5/2}(K_i)}+h_i\frac{h_i^3}{p_i^3}\|u\|^4_{H^{5/2}(K_i)}\right)^{1/4}
\end{aligned}
$$

$$
\begin{aligned}
&\leq C\|u\|_{H^{5/2}(\Omega)}\left(\frac{h^{1/2}}{p^{1/2}}+\frac{h^{1/2}}{p^{1/4}}\right)|||\xi|||\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_u h^{1/2}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.48)
\end{aligned}
$$

Finally, using Hölder's inequality, the inverse inequality (1.22), the trace inequalities (1.14)-(1.19) and (1.23), the first term on the right hand side of (4.46) is estimated as

$$
\begin{aligned}
\sum_{e_k\in\Gamma_I}\int_{e_k}|\nabla\xi|^2|[v_h]| &\leq \sum_{e_k\in\Gamma_I}\left(\frac{|e_k|^{1/2}}{p_k}\|\nabla\xi\|_{L^\infty(e_k)}\|\nabla\xi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\left(\frac{|e_k|^{1/2}}{p_k}\|\nabla\xi\|_{L^\infty(K_i)}\|\nabla\xi\|_{L^2(e_k)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\sum_{i=1}^{N_h}\sum_{e_k\subset\partial K_i}\left(\frac{p_i}{h_i}\|\nabla\xi\|^2_{L^2(K_i)}\left(\int_{e_k}\frac{p_k^2}{|e_k|}[v_h]^2\right)^{1/2}\right)\\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)|||\xi|||^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_u\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\delta^2\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}
\end{aligned}
$$

$$
\begin{aligned}
&\leq CC_u h^{-\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}\\
&\leq CC_Q C_u h^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.49)
\end{aligned}
$$

Altogether, we complete the rest of the proof. ■

LEMMA 4.2.7 *Let $y\in\mathcal{O}_\delta(I_h u)$ and $v_h\in V_h$. Then, there is a positive constant C such that*

$$
|\mathcal{N}(u,y;v_h)| \ \leq\ CC_a\|\zeta\|_{W_4^1(\Omega,\mathcal{T}_h)}\|\zeta\|_{W_2^1(\Omega,\mathcal{T}_h)}\|v_h\|_{W_4^1(\Omega,\mathcal{T}_h)}+CC_aC_QC_uh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}.
$$

Moreover, there is positive constant C which is independent of h and p such that

$$
|\mathcal{N}(u,y;v_h)|\leq CC_aC_uC_Qh^{1/2-\epsilon}\delta\ |||v_h|||.
$$

Proof. From (4.14), we note using definition of $\mathcal{N}(\cdot,\cdot;\cdot)$ that

$$
\begin{aligned}
\mathcal{N}(u,y;v_h) &= \sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_{\mathbf{a}}(u-y,\nabla(u-y))\cdot\nabla v_h dx\\
&- \sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{R}_{\mathbf{a}}(u-y,\nabla(u-y))\cdot\nu\}[v_h]\ ds\\
&+ \sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_f(u-y,\nabla(u-y))v_h dx\\
&- \sum_{e_k\in\Gamma_\partial}\int_{e_k}\nabla(u-y)^T\tilde{\mathbf{a}}_{\mathbf{zz}}(g,\nabla u)\nabla(u-y)\cdot\nu v_h\ ds\\
&+ \theta\sum_{e_k\in\Gamma_I}\int_{e_k}\{(\mathbf{a}_{\mathbf{z}}(u,\nabla u)-\mathbf{a}_{\mathbf{z}}(y,\nabla y))\nabla v_h\cdot\nu\}[u-y]ds\\
&- \gamma\sum_{e_k\in\Gamma_I}\int_{e_k}\{(f_{\mathbf{z}}(u,\nabla u)-f_{\mathbf{z}}(y,\nabla y))v_h\cdot\nu\}[u-y]ds\\
&+ \theta\sum_{e_k\in\Gamma_\partial}\int_{e_k}(\mathbf{a}_{\mathbf{z}}(g,\nabla u)-\mathbf{a}_{\mathbf{z}}(g,\nabla y))\nabla v_h\cdot\nu(u-y)ds. \qquad (4.50)
\end{aligned}
$$

Using a similar arguments as in Lemma 4.2.5, it is easy to see that

$$
|\sum_{i=1}^{N_h}\int_{K_i}\tilde{R}_f(u-y,\nabla(u-y))v_h dx|\leq C_a\|\zeta\|_{W_4^1(\Omega,\mathcal{T}_h)}\|\zeta\|_{W_2^1(\Omega,\mathcal{T}_h)}\|v_h\|_{L^4(\Omega)}, \qquad (4.51)
$$

and a similar argument of Lemma 4.2.6 yields

$$
|\sum_{e_k\in\Gamma_\partial}\int_{e_k}\nabla(u-y)^T\tilde{\mathbf{a}}_{\mathbf{zz}}(g,\nabla u)\nabla(u-y)\cdot\nu v_h\ ds|\leq CC_aC_QC_uh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2}. \qquad (4.52)
$$

Using Taylor's expansions, we rewrite the fifth term on the right-hand side of (4.50) as

$$\sum_{e_k\in\Gamma_I}\int_{e_k}\{(\mathbf{a_z}(u,\nabla u)-\mathbf{a_z}(y,\nabla y))\nabla v_h\cdot\nu\}[u-y]ds=\sum_{e_k\in\Gamma_I}\int_{e_k}\{\tilde{\mathbf{a}}_{\mathbf{z}u}(y,\nabla y)\zeta\nabla v_h\cdot\nu\}[\zeta]ds$$
$$+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\nabla\zeta^T\tilde{\mathbf{a}}_{\mathbf{zz}}(y,\nabla y)\nabla v_h\cdot\nu\}[\zeta]ds. \quad (4.53)$$

Then, using repeated arguments as in Lemma 4.2.6, we arrive at

$$|\sum_{e_k\in\Gamma_I}\int_{e_k}\{(\mathbf{a_z}(u,\nabla u)-\mathbf{a_z}(y,\nabla y))\nabla v_h\cdot\nu\}[u-y]ds|\le CC_aC_QC_uh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2} \quad (4.54)$$

Similarly we obtain

$$|\sum_{e_k\in\Gamma_I}\int_{e_k}\{(f_{\mathbf{z}}(u,\nabla u)-f_{\mathbf{z}}(y,\nabla y))\nabla v_h\cdot\nu\}[u-y]ds|\le CC_aC_QC_uh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2} \quad (4.55)$$

and

$$|\sum_{e_k\in\Gamma_\partial}\int_{e_k}(\mathbf{a_z}(g,\nabla u)-\mathbf{a_z}(g,\nabla y))\nabla v_h\cdot\nu(u-y)ds|\le CC_aC_QC_uh^{1/2-\epsilon}\delta\ \mathcal{J}^\sigma(v_h,v_h)^{1/2} \quad (4.56)$$

We substitute (4.52)-(4.56) in (4.50). Then use Lemma 4.2.5 and Lemma 4.2.6 to complete the proof of the first inequality. Now, for the second inequality of the lemma, we use the inverse inequality (1.20) to find that

$$\begin{aligned}
\|v_h\|_{W_4^1(\Omega,\mathcal{T}_h)} &= \left(\sum_{i=1}^{N_h}\|v_h\|^4_{W_4^1(K_i)}\right)^{1/4}\\
&\le C\left(\sum_{i=1}^{N_h}\frac{p_i}{h_i}\|v_h\|^4_{W_2^1(K_i)}\right)^{1/4}\\
&\le C\left(\max_{1\le i\le N_h}\frac{p_i^2}{h_i^2}\right)^{1/4}\left(\sum_{i=1}^{N_h}\|v_h\|^4_{W_2^1(K_i)}\right)^{1/4}\\
&\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}\max_{1\le i\le N_h}\|v_h\|^{1/2}_{W_2^1(K_i)}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{W_2^1(K_i)}\right)^{1/4}\\
&\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{W_2^1(K_i)}\right)^{1/4}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{W_2^1(K_i)}\right)^{1/4}\\
&\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)^{1/2}\|v_h\|_{W_2^1(\Omega,\mathcal{T}_h)}. \quad (4.57)
\end{aligned}$$

Let $\zeta = u - y = \eta + \xi$, where $\eta = u - I_h u$ and $\xi = I_h u - y$. Now, a use of triangle inequality implies that

$$\|\zeta\|_{W_4^1(\Omega,\mathcal{T}_h)} \leq \|\eta\|_{W_4^1(\Omega,\mathcal{T}_h)} + \|\xi\|_{W_4^1(\Omega,\mathcal{T}_h)}. \tag{4.58}$$

Then, we obtain using Lemma 1.2.1

$$\begin{aligned}\|\eta\|_{W_4^1(\Omega,\mathcal{T}_h)} &\leq C\left(\sum_{i=1}^{N_h}\|\eta\|^4_{W_4^1(K_i)}\right)^{1/4} \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^4}{p_i^4}\|u\|^4_{H^{5/2}(K_i)}\right)^{1/4}\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right)\|u\|_{H^{5/2}(\Omega)},\end{aligned} \tag{4.59}$$

and using the inverse inequalities (1.20) and (4.30), we now find that

$$\begin{aligned}\|\xi\|_{W_4^1(\Omega,\mathcal{T}_h)} &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}\|\xi\|_{W_2^1(\Omega,\mathcal{T}_h)}\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}\delta\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\\ &\leq CC_Q\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right).\end{aligned} \tag{4.60}$$

Now using (4.31), we arrive at

$$\|\zeta\|_{W_2^1(\Omega,\mathcal{T}_h)} \leq \|\eta\|_{W_2^1(\Omega,\mathcal{T}_h)} + \|\xi\|_{W_2^1(\Omega,\mathcal{T}_h)} \leq h^\epsilon\delta + \delta \leq 2\delta. \tag{4.61}$$

Substitute (4.59)-(4.60) in (4.58) and use (4.57) to find that

$$\begin{aligned}\|\zeta\|_{W_4^1(\Omega,\mathcal{T}_h)}\|v_h\|_{W_4^1(\Omega,\mathcal{T}_h)} &\leq C\left(\|\eta\|_{W_4^1(\Omega,\mathcal{T}_h)} + \|\xi\|_{W_4^1(\Omega,\mathcal{T}_h)}\right)\|v_h\|_{W_4^1(\Omega,\mathcal{T}_h)}\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)^{1/2}\left(\|\eta\|_{W_4^1(\Omega,\mathcal{T}_h)} + \|\xi\|_{W_4^1(\Omega,\mathcal{T}_h)}\right)\|v_h\|_{W_2^1(\Omega,\mathcal{T}_h)}\\ &\leq CC_QC_u h^{1/2}\|v_h\|_{W_2^1(\Omega,\mathcal{T}_h)}.\end{aligned} \tag{4.62}$$

We substitute (4.61)-(4.62) in (4.51) to obtain the required estimate and this completes the rest of the proof. ■

THEOREM 4.2.2 *Let $0 < h < h_0 < 1$ and δ be given by (4.30). Then, S_h maps $\mathcal{O}_\delta(I_h u)$ into itself.*

Proof. Let $y \in \mathcal{O}_\delta(I_h u)$ and $S_h(y) = u_y$ be the solution given by (4.18). Set $\xi = I_h u - u_y$. Then, using Lemma 4.2.1, Lemma 4.2.2 and Lemma 4.2.7, we find that

$$\begin{aligned} C_1 |||\xi|||^2 - C_2 \|\xi\|^2 &\leq a(u;\xi,\xi) + b(u;\xi,\xi) \\ &= a(u;\eta,\xi) + b(u;\eta,\xi) + \mathcal{N}(u,y;\xi) \\ &\leq C\left(|||\eta||| + CC_aC_uC_Qh^{1/2-\epsilon}\delta\right)|||\xi||| \\ &\leq C\left(h^\epsilon\delta + CC_aC_uC_Qh^{1/2-\epsilon}\delta\right)|||\xi||| \\ &\leq CC_aC_uC_Qh^\epsilon\delta\ |||\xi|||. \end{aligned}$$

Hence, we obtain

$$|||\xi||| \leq CC_aC_uC_Qh^\epsilon\delta + C_1\|\xi\|. \tag{4.63}$$

Now, we use Lemma 4.2.4 to estimate $\|\xi\|$. Setting $v_h = \xi$ in (4.22), we arrive at

$$\begin{aligned} \|\xi\|^2 &= a(u;\xi,\phi_h) + b(u;\xi,\phi_h) \\ &= a(u;\eta,\phi_h) + b(u;\eta,\phi_h) + \mathcal{N}(u,y;\phi_h) \\ &\leq C\left(|||\eta||| + CC_aC_uC_Qh^{1/2-\epsilon}\delta\right)|||\phi_h||| \\ &\leq C\left(h^\epsilon\delta + CC_aC_uC_Qh^{1/2-\epsilon}\delta\right)\|\xi\|. \end{aligned}$$

Therefore, we obtain

$$\|\xi\| \leq CC_aC_uC_Qh^\epsilon\delta. \tag{4.64}$$

We substitute (4.64) in (4.63). Then, choose h sufficiently small so that $CC_aC_uC_Qh^\epsilon \leq 1$, and hence, S_h maps $\mathcal{O}_\delta(I_h u)$ to itself. This now completes the proof of the theorem. ■

Below, we discuss the continuity of the map S_h in the ball $\mathcal{O}_\delta(I_h u)$.

THEOREM 4.2.3 *For $y_1,\ y_2 \in \mathcal{O}_\delta(I_h u)$, let $u_{y_1} = S_h(y_1),\ u_{y_2} = S_h(y_2)$ be the corresponding solutions of (4.18). Then, there is a positive constant C such that for any $0 < h < h_0 < 1$,*

$$|||u_{y_1} - u_{y_2}||| \leq CC_aC_uC_Qh^{1/2-\epsilon}|||y_1 - y_2|||. \tag{4.65}$$

Proof. The solutions u_{y_1} and u_{y_2} of the linearized problem (4.18) satisfy

$$a(u; I_h u - u_{y_1}, v_h) + b(u; I_h u - u_{y_1}, v_h) = a(u; \eta, v_h) + b(u; \eta, v_h) + \mathcal{N}(u, u_{y_1}; v_h) \quad (4.66)$$

and

$$a(u; I_h u - u_{y_2}, v_h) + b(u; I_h u - u_{y_2}, v_h) = a(u; \eta, v_h) + b(u; \eta, v_h) + \mathcal{N}(u, u_{y_2}; v_h), \quad (4.67)$$

respectively, where $\eta = I_h u - u$. We subtract (4.66) from (4.67) to arrive at

$$a(u; u_{y_2} - u_{y_1}, v_h) + b(u; u_{y_2} - u_{y_1}, v_h) = \left(\mathcal{N}(u, u_{y_2}; v_h) - \mathcal{N}(u, u_{y_1}; v_h)\right). \quad (4.68)$$

Let $\zeta_1 = u - y_1$ and $\zeta_2 = u - y_2$. We then expand the term on the right-hand side of (4.68) as follows :

$$\begin{aligned}
\mathcal{N}(u, y_2; v_h) - \mathcal{N}(u, y_1; v_h) &= \sum_{i=1}^{N_h} \int_{K_i} \left(\tilde{R}_{\mathbf{a}}(\zeta_2, \nabla\zeta_2) - \tilde{R}_{\mathbf{a}}(\zeta_1, \nabla\zeta_1)\right) \cdot \nabla v_h dx \\
&- \sum_{e_k \in \Gamma_I} \int_{e_k} \{\left(\tilde{R}_{\mathbf{a}}(\zeta_2, \nabla\zeta_2) - \tilde{R}_{\mathbf{a}}(\zeta_1, \nabla\zeta_1)\right) \cdot \nu\}[v_h]\ ds \\
&+ \sum_{i=1}^{N_h} \int_{K_i} \left(\tilde{R}_f(\zeta_2, \nabla\zeta_2) - \tilde{R}_f(\zeta_1, \nabla\zeta_1)\right) v_h dx \\
&- \sum_{e_k \in \Gamma_\partial} \int_{e_k} \left(\nabla\zeta_2^T \tilde{\mathbf{a}}_{\mathbf{zz}}(g, \nabla u)\nabla\zeta_1 - \nabla\zeta_1^T \tilde{\mathbf{a}}_{\mathbf{zz}}(g, \nabla u)\nabla\zeta_1\right) \cdot \nu v_h\ ds \\
&+ \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{(\mathbf{a}_{\mathbf{z}}(u, \nabla u) - \mathbf{a}_{\mathbf{z}}(y_2, \nabla y_2))\nabla v_h \cdot \nu\}[y_2] ds \\
&- \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{(\mathbf{a}_{\mathbf{z}}(u, \nabla u) - \mathbf{a}_{\mathbf{z}}(y_1, \nabla y_1))\nabla v_h \cdot \nu\}[y_1] ds \\
&+ \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} (\mathbf{a}_{\mathbf{z}}(g, \nabla u) - \mathbf{a}_{\mathbf{z}}(g, \nabla y_2))\nabla v_h \cdot \nu(g - y_2) ds \\
&- \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} (\mathbf{a}_{\mathbf{z}}(g, \nabla u) - \mathbf{a}_{\mathbf{z}}(g, \nabla y_1))\nabla v_h \cdot \nu(g - y_1) ds \\
&+ \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{(f_{\mathbf{z}}(u, \nabla u) - f_{\mathbf{z}}(y_2, \nabla y_2))v_h \cdot \nu\}[y_2] ds \\
&- \gamma \sum_{e_k \in \Gamma_I} \int_{e_k} \{(f_{\mathbf{z}}(u, \nabla u) - f_{\mathbf{z}}(y_1, \nabla y_1))v_h \cdot \nu\}[y_1] ds. \quad (4.69)
\end{aligned}$$

Now, using Taylor's expansions, we rewrite each term on the right hand side of (4.69) so that every term contains either $y_1 - y_2$ or $\nabla(y_1 - y_2)$. For example, the first term can be rewritten as

$$\begin{aligned}
&\tilde{R}_{\mathbf{a}}(\zeta_2, \nabla\zeta_2) - \tilde{R}_{\mathbf{a}}(\zeta_1, \nabla\zeta_1) \\
&= \mathbf{a}(y_2, \nabla y_2) - \mathbf{a}(u, \nabla u) + \mathbf{a}_u(u, \nabla u)(u - y_2) + \mathbf{a}_{\mathbf{z}}(u, \nabla u)\nabla(u - y_2) \\
&-\mathbf{a}(y_1, \nabla y_1) + \mathbf{a}(u, \nabla u) - \mathbf{a}_u(u, \nabla u)(u - y_1) - \mathbf{a}_{\mathbf{z}}(u, \nabla u)\nabla(u - y_1) \\
&= \mathbf{a}(y_2, \nabla y_2) - \mathbf{a}(y_1, \nabla y_1) + \mathbf{a}_u(y_1, \nabla y_1)(y_1 - y_2) + \mathbf{a}_{\mathbf{z}}(y_1, \nabla y_1)\nabla(y_1 - y_2) \\
&+ \left(\mathbf{a}_u(u, \nabla u) - \mathbf{a}_u(y_1, \nabla y_1)\right)(y_1 - y_2) + \left(\mathbf{a}_{\mathbf{z}}(u, \nabla u) - \mathbf{a}_{\mathbf{z}}(y_1, \nabla y_1)\right)\nabla(y_1 - y_2) \\
&= \mathbf{a}(y_2, \nabla y_2) - \mathbf{a}(y_1, \nabla y_1) + \mathbf{a}_u(y_1, \nabla y_1)(y_1 - y_2) + \mathbf{a}_{\mathbf{z}}(y_1, \nabla y_1)\nabla(y_1 - y_2) \\
&+ \left(\mathbf{a}_u(u, \nabla u) - \mathbf{a}_u(y_1, \nabla y_1)\right)(y_1 - y_2) + \left(\mathbf{a}_{\mathbf{z}}(u, \nabla u) - \mathbf{a}_{\mathbf{z}}(y_1, \nabla y_1)\right)\nabla(y_1 - y_2).
\end{aligned}$$

Then, we use Taylor's expansions (4.5)-(4.6) with integral reminders. Hence, using the arguments as in Lemma 4.2.5 and Lemma 4.2.6, the term on the right hand side of (4.68) is estimated as

$$|\mathcal{N}(u, u_{y_2}; v_h) - \mathcal{N}(u, u_{y_1}; v_h)| \leq CC_aC_uC_Qh^{1/2-\epsilon}\ |||y_1 - y_2|||\ |||v_h|||. \tag{4.70}$$

Setting $v_h = u_{y_2} - u_{y_1}$ in (4.68), we obtain

$$\begin{aligned}
C_1|||u_{y_2} - u_{y_1}|||^2 - C_2\|u_{y_2} - u_{y_1}\|^2 &\leq a(u; u_{y_2} - u_{y_1}, u_{y_2} - u_{y_1}) + b(u; u_{y_2} - u_{y_1}, u_{y_2} - u_{y_1}) \\
&= \mathcal{N}(u, u_{y_2}; u_{y_2} - u_{y_1}) - \mathcal{N}(u, u_{y_1}; u_{y_2} - u_{y_1}).
\end{aligned} \tag{4.71}$$

Similarly with $v_h = u_{y_2} - u_{y_1}$ in (4.70), we use (4.71) to find that

$$|||u_{y_2} - u_{y_1}||| \leq CC_aC_uC_Qh^{1/2-\epsilon}|||y_1 - y_2||| + C_2\|u_{y_2} - u_{y_1}\|. \tag{4.72}$$

To estimate $\|u_{y_2} - u_{y_1}\|$, we set $\xi = u_{y_2} - u_{y_1}$ and $v_h = u_{y_2} - u_{y_1}$ in (4.22). We then use (4.22), (4.68) and (4.70) to obtain

$$\begin{aligned}
\|u_{y_2} - u_{y_1}\|^2 &= a(u; u_{y_2} - u_{y_1}, \phi_h) + b(u; u_{y_2} - u_{y_1}, \phi_h) \\
&= \mathcal{N}(u, u_{y_2}; \phi_h) - \mathcal{N}(u, u_{y_1}; \phi_h) \\
&\leq CC_aC_uC_Qh^{1/2-\epsilon}|||y_1 - y_2|||\ |||\phi_h||| \\
&\leq CC_aC_uC_Qh^{1/2-\epsilon}|||y_1 - y_2|||\ \|u_{y_2} - u_{y_1}\|.
\end{aligned} \tag{4.73}$$

We now substitute (4.73) in (4.72) to complete the rest of the proof. ■

Now using Theorem 4.2.2, Theorem 4.2.3 and Brouwer fixed point theorem, we can conclude that for all $0 < h < h_0 < 1$, the map S_h has a fixed point u_h in the ball $\mathcal{O}_\delta(I_h u)$. Theorem 4.2.3 also implies that there is atmost one fixed point of S_h in that ball which is the solution u_h for the nonlinear system (4.10).

4.2.2 A priori error estimates.

From the estimates (4.63)-(4.64), we note that the following estimate holds for u_h

$$\begin{aligned} |||I_h u - u_h||| &\leq CC_a C_u C_Q h^\epsilon \delta \\ &\leq CC_a C_u C_Q |||I_h u - u|||. \end{aligned} \tag{4.74}$$

Now, the proof of the following theorem is a consequence of (4.74) and Lemma 4.2.3.

THEOREM 4.2.4 *Let $0 < h < h_0 < 1$. Then there is a positive constant C such that*

$$|||u - u_h||| \leq CC_a C_u C_Q \left(\sum_{i=1}^{N_h} \frac{h_i^{2\mu_i - 2}}{p_i^{2s_i - 3}} \|u\|^2_{H^{s_i}(K_i)} \right)^{1/2}, \tag{4.75}$$

where $\mu_i = min\{s_i, p_i + 1\}$.

Note that the estimate obtained in Theorem 4.2.4 is optimal in h and mildly suboptimal in p which leads to the same optimal order of convergence in h and suboptimal in p in case of linear elliptic boundary value problems.

4.2.3 L^2-norm error estimate when $\theta = -1$

In this section, we estimate the error in the L^2-norm. Since the linearized problem (4.18) is adjoint consistent when $\theta = -1$, we expect to obtain optimal order of convergence in the L^2-norm by applying the standard Aubin-Nitsche duality argument. But, when $\theta \in (-1, 1]$, we note that the linearized problem is not adjoint consistent. Therefore, it may not be possible to improve the error estimate in the L^2-norm. However, if g is either zero or a piecewise polynomial of degree less than or equal to p on the boundary, then it is possible to obtain optimal rate of convergence in the L^2-norm on regular mesh by imposing the super-penalty, see [41], [61]. Presently, we restrict ourself only to the case $\theta = -1$.

In order to apply the standard Aubin-Nitsche duality argument, we consider the following auxiliary problem. Assume that there is a unique solution $\psi \in H^2(\Omega)$ of

$$-\nabla \cdot (A(u)\nabla\psi + \mathbf{f}(u)\psi) + \mathbf{b}(u)\cdot\nabla\psi + F(u)\psi = u - u_h \quad \text{in } \Omega, \tag{4.76}$$

$$\psi = 0 \quad \text{on } \partial\Omega, \tag{4.77}$$

which satisfies the following elliptic regularity

$$\|\psi\|_{H^2(\Omega)} \leq C\|u - u_h\|. \tag{4.78}$$

This would be guaranteed if $f_u(u, \nabla u) \geq \nabla \cdot f_{\mathbf{z}}(u, \nabla u)$, see [40]. Note that $[\psi] = 0$ on each $e_k \in \Gamma$. Let $e = u - u_h = \eta + \xi$, where $\eta = u - I_h u$ and $\xi = I_h u - u_h$. Let $\eta_\psi = \psi - \psi_h$, where $\psi_h = I_h\psi$. Now, we multiply (4.76) by e, integrate over Ω and apply integration by parts to obtain

$$\begin{aligned}
\|e\|^2 &= a(u; e, \psi) + b(u; e, \psi) \\
&= a(u; e, \eta_\psi) + b(u; e, \eta_\psi) + a(u; e, \psi_h) + b(u; e, \psi_h) \\
&= a(u; e, \eta_\psi) + b(u; e, \eta_\psi) + \mathcal{N}(u, u_h; \psi_h).
\end{aligned} \tag{4.79}$$

Using Lemma 4.2.3 and (4.74), the first two terms on the right-hand side of (4.79) are estimated as

$$\begin{aligned}
|a(u; e, \eta_\psi) + b(u; e, \eta_\psi)| &\leq CC_a|||e||| \; |||\eta_\psi||| \\
&\leq CCC_a\frac{h}{p^{1/2}}|||e||| \; \|\psi\|_{H^2(\Omega)} \\
&\leq CC_aC_QC_u\frac{h}{p^{1/2}}|||\eta||| \; \|\psi\|_{H^2(\Omega)}.
\end{aligned} \tag{4.80}$$

Before proceeding to estimate the other term, we note from Lemma 1.2.1 that the following stability condition holds for ψ_h. For all $1 \leq i \leq N_h$,

$$\|\psi_h\|_{W_4^1(K_i)} \leq C\|\psi\|_{H^2(K_i)}. \tag{4.81}$$

Now, using Lemma 4.2.7, we obtain

$$\begin{aligned}
|\mathcal{N}(u, u_h; \psi_h)| &\leq C_a\|e\|_{W_4^1(\Omega,\mathcal{T}_h)}\|e\|_{W_2^1(\Omega,\mathcal{T}_h)}\|\psi_h\|_{W_4^1(\Omega,\mathcal{T}_h)} \\
&\quad + CC_aC_uC_Q|||\eta||| \; \mathcal{J}^\sigma(\psi_h, \psi_h).
\end{aligned} \tag{4.82}$$

We use (4.57) and Theorem 4.2.4 to find that

$$\begin{aligned}
\|\xi\|_{W_4^1(\Omega)} &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|\xi\|_{W_2^1(\Omega,\mathcal{T}_h)}\\
&\leq CC_uC_Q\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\sum_{i=1}^{N_h}\frac{h_i^3}{p_i^2}\|u\|^2_{H^{5/2}(\Omega)}\right)^{1/2}\\
&\leq CC_u^2C_Q\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\\
&\leq CCaC_u^2C_Q\frac{h}{p^{1/2}}. \qquad (4.83)
\end{aligned}$$

An appeal to Lemma 1.2.1 yields

$$\|\eta\|_{W_4^1(\Omega)}\leq C\frac{h}{p^{1/2}}\|u\|_{H^{5/2}(\Omega)}. \qquad (4.84)$$

Hence, we obtain

$$\|e\|_{W_4^1(\Omega)}\leq\|\eta\|_{W_4^1(\Omega)}+\|\xi\|_{W_4^1(\Omega)}\leq CC_u^2C_QC_a\frac{h}{p^{1/2}}. \qquad (4.85)$$

Since $[\psi]=0$ on each $e_k\in\Gamma$, we use Lemma 4.2.3 to find that

$$\mathcal{J}^\sigma(\psi_h,\psi_h)=-\mathcal{J}^\sigma(\eta_\psi,\eta_\psi)\leq C\frac{h}{p^{1/2}}\|\psi\|_{H^2(\Omega)}. \qquad (4.86)$$

We combine the estimates (4.79)-(4.86) to prove the following theorem.

THEOREM 4.2.5 *Let $0<h<h_0<1$ and $f_u(u,\nabla u)\geq\nabla\cdot f_{\mathbf{z}}(u,\nabla u)$. then, there is a positive constant C which is independent of h and p such that*

$$\|u-u_h\|\leq CC_aC_u^2C_Q\frac{h^u}{p^{s-1}}\|u\|_{H^s(\Omega)}.$$

where $\mu=min\{s,p+1\}$.

Note that the estimate obtained in the Theorem 4.2.5 is optimal in h and suboptimal in p.

REMARK 4.2.4 *When $\theta=-1$, to estimate $\|\xi\|$ in the Theorem 4.2.2 and subsequently, one can directly apply the standard Aubin-Nitsche duality argument instead of using Lemma 4.2.4. Hence, the assumption in the Lemma 4.2.4 that $f_u(u,\nabla u)-\nabla\cdot\mathbf{a}_u(u,\nabla u)\geq 0$ can be replaced by the assumption $f_u(u,\nabla u)-\nabla\cdot f_{\mathbf{z}}(u,\nabla u)\geq 0$ which is used in the Theorem 4.2.5.*

4.3 Application to the mean curvature problem.

We note that for mean curvature flow problem $\mathbf{a}(\mathbf{x}, u, \mathbf{z}) = (1 + |\mathbf{z}|^2)^{-1/2}\,\mathbf{z}$ and $f(\mathbf{x}, u, \mathbf{z}) = f(x)$. We verify the conditions stated in Lemma 4.2.4 and in Theorem 4.2.5. First, we verify the ellipticity condition for this problem. We calculate the matrix $\mathbf{a}_{\mathbf{z}}(u, \mathbf{z}) = [a^{ij}(u, \mathbf{z})]_{1 \le i,j \le 2}$, where $\mathbf{z} = (z_1, z_2)$, as

$$\mathbf{a}_{\mathbf{z}}(u, \mathbf{z}) = R(\mathbf{z}) \begin{bmatrix} 1 + z_2^2 & -z_1 z_2 \\ -z_1 z_2 & 1 + z_1^2 \end{bmatrix},$$

where $R(\mathbf{z}) = 1/(1 + z_1^2 + z_2^2)^{3/2}$. Now, for any $\xi = (\xi_1, \xi_2) \in \mathbb{R}^2 - \{0\}$, we note that

$$\begin{aligned} \sum_{i,j=1}^{2} a^{ij}(u, \mathbf{z})\xi_i\xi_j &= R(\mathbf{z}) \left[(1 + z_2^2)\xi_1^2 - 2z_1 z_2 \xi_1 \xi_2 + (1 + z_1^2)\xi_2^2\right] \\ &= R(\mathbf{z}) \left[(z_1\xi_2 - z_2\xi_1)^2 + (\xi_1^2 + \xi^2)\right] \end{aligned}$$

Then, it is easy to check that

$$R(\mathbf{z})|\xi|^2 \le \sum_{i,j=1}^{2} a^{ij}(u, \mathbf{z})\xi_i\xi_j \le R(\mathbf{z})(1 + 4|\mathbf{z}|^2)|\xi|^2.$$

For bounded $\mathbf{z}$, $R(\mathbf{z})$ can be made bounded below by a positive constant C_α. Further, $R(\mathbf{z}) \le 1$ and each a^{ij} is bounded for bounded u. Since $f_u(u, \mathbf{z}) = 0$, $f_{\mathbf{z}}(u, \mathbf{z}) = 0$ and $\nabla \cdot \mathbf{a}_u(u, \mathbf{z}) = 0$, the conditions $f_u(u, \mathbf{z}) - \nabla \cdot \mathbf{a}_u(u, \mathbf{z}) \ge 0$ and $f_u(u, \mathbf{z}) - f_{\mathbf{z}}(u, \mathbf{z}) \ge 0$ are satisfied which are used in Lemma 4.2.4 and Theorem 4.2.5, respectively. Therefore, we conclude that the results of this chapter are well applicable to this problem.

4.4 Numerical Experiments

In this section, we present some numerical experiments to illustrate the theoretical order of convergence obtained in Theorem 4.2.4 and Theorem 4.2.5. We consider mean curvature flow as a model problem which is given by

$$-\nabla \cdot \left(\frac{\nabla u}{(1 + |\nabla u|^2)^{1/2}} \right) = f \quad \text{in } \Omega, \tag{4.87}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{4.88}$$

where $\Omega = (0,1) \times (0,1)$. We take the forcing function f in such a way that the exact solution is $u = x(1-x)y(1-y)$.

We compute the approximate solution u_h on a sequence of finite element subdivisions $\mathcal{T}_h$ of Ω, where $\mathcal{T}_h$ is formed by uniform triangles. We use discrete space V_h with piecewise polynomials of uniform degree $p = 2$. Since the discrete space V_h can have piecewise polynomials which may be discontinuous across the edges of elements, we choose basis functions as follows. For $1 \leq i \leq N_h$,

$$\Phi_{(i-1)\times 6+j} = \begin{cases} \lambda_j & \text{on } K_i,\ j = 1,\ 2,\ 3, \\ 0 & \text{elsewhere,} \end{cases}$$

$$\Phi_{(i-1)\times 6+4} = \begin{cases} \lambda_1\lambda_2 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases}$$

$$\Phi_{(i-1)\times 6+5} = \begin{cases} \lambda_2\lambda_3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases}$$

$$\Phi_{(i-1)\times 6+6} = \begin{cases} \lambda_1\lambda_3 & \text{on } K_i, \\ 0 & \text{elsewhere,} \end{cases}$$

where λ_1, λ_2 and λ_3 are barycentric coordinates of K_i. We note that each of the basis functions takes support only on the corresponding finite element K_i. Let $N = N_h * 6$ denotes the dimension of V_h. Denote the basis of V_h by $\{\Phi_i : 1 \leq i \leq N\}$. The penalty parameter σ_k is chosen as $\sigma_k = 10$, on each $e_k \in \Gamma$. We choose $\theta = -1$, $\theta = 0$ and $\theta = 1$ which corresponds to the symmetric, incomplete and nonsymmetric interior penalty methods, respectively. The discrete solution u_h is written as

$$u_h = \sum_{i=1}^{N} \alpha_i \Phi_i. \tag{4.89}$$

In order to derive the nonlinear algebraic system corresponding to (4.10), we set $v_h = \Phi_j$ in (4.10) and obtain for each j

$$F_j(\alpha) = B(u_h, \Phi_j) - L(\Phi_j) = 0,\ 1 \leq j \leq N,$$

where $\alpha = [\alpha_1, \alpha_2, \cdots, \alpha_N]$,

$$\begin{aligned} B(u_h, \Phi_j) &= \sum_{i=1}^{N_h} \int_{K_i} \frac{\nabla u_h \cdot \nabla \Phi_j}{(1+|\nabla u_h|)^{1/2}}\, dx - \sum_{e_k \in \Gamma_I} \int_{e_k} \{\frac{\nabla u_h \cdot \nu}{(1+|\nabla u_h|)^{1/2}}\}[\Phi_j]\, ds \\ &+ \theta \sum_{e_k \in \Gamma_I} \int_{e_k} \{\mathbf{a_z}(\nabla u_h)\nabla \Phi_j \cdot \nu\}[u_h]\, ds - \sum_{e_k \in \Gamma_\partial} \int_{e_k} \frac{\nabla u_h \cdot \nu}{(1+|\nabla u_h|)^{1/2}} \Phi_j\, ds \\ &+ \theta \sum_{e_k \in \Gamma_\partial} \int_{e_k} \mathbf{a_z}(\nabla u_h)\nabla \Phi_j \cdot \nu u_h\, ds + \mathcal{J}^\sigma(u_h, \Phi_j) \qquad (4.90) \\ &= I_1 + I_2 + I_3 + I_4 + I_5 + I_6 \end{aligned}$$

with

$$\mathbf{a_z}(\nabla u_h) = \frac{1}{(1+|\nabla u_h|)^{3/2}} \begin{bmatrix} 1+\left(\dfrac{\partial u_h}{\partial x_2}\right)^2 & -\dfrac{\partial u_h}{\partial x_1}\dfrac{\partial u_h}{\partial x_2} \\ -\dfrac{\partial u_h}{\partial x_2}\dfrac{\partial u_h}{\partial x_1} & 1+\left(\dfrac{\partial u_h}{\partial x_1}\right)^2 \end{bmatrix},$$

and $L(\Phi_j) = (f, \Phi_j)$. The resulting nonlinear system is then denoted by

$$\mathbf{F}(\alpha) = [F_1(\alpha), F_2(\alpha), \cdots, F_N(\alpha)]^T = [0, 0, \cdots, 0]^T. \qquad (4.91)$$

We then apply the Newton's method to find the solution α to (4.91). The Jacobian matrix of the system $\mathbf{F}(\alpha)$ is computed as follows.

$$\mathbf{J} = \left[\frac{\partial F_j}{\partial \alpha_m}\right]_{1 \le j,m \le N} = \left[\frac{\partial B(u_h, \Phi_j)}{\partial \alpha_m}\right]_{1 \le j,m \le N}. \qquad (4.92)$$

We substitute (4.89) in (4.92) and then using (4.90), we first compute

$$\begin{aligned} \frac{\partial I_1}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m} \left(\sum_{i=1}^{N_h} \int_{K_i} \frac{\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_j}{D^{1/2}}\, dx \right) \\ &= \sum_{i=1}^{N_h} \int_{K_i} \left(\frac{\nabla \Phi_m \cdot \nabla \Phi_j}{D^{1/2}} - \frac{\left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_j\right)\left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_m\right)}{D^{3/2}} \right) dx, \end{aligned}$$

where

$$D = 1 + \left(\sum_{l=1}^{N} \alpha_l \frac{\partial \Phi_l}{\partial x_1} \right)^2 + \left(\sum_{l=1}^{N} \alpha_l \frac{\partial \Phi_l}{\partial x_2} \right)^2 . \tag{4.93}$$

Next, we note that

$$\begin{aligned} \frac{\partial I_2}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m} \left(\sum_{e_k \in \Gamma_I} \int_{e_k} \{ \frac{\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nu}{D^{1/2}} \} [\Phi_j] \; ds \right) \\ &= \sum_{e_k \in \Gamma_I} \int_{e_k} \{ \frac{\nabla \Phi_m \cdot \nu}{D^{1/2}} - \frac{\left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nu \right) \left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_m \right)}{D^{3/2}} \} [\Phi_j] ds. \end{aligned}$$

Similarly,

$$\frac{\partial I_4}{\partial \alpha_m} = \sum_{e_k \in \Gamma_\partial} \int_{e_k} \left(\frac{\nabla \Phi_m \cdot \nu}{D^{1/2}} - \frac{\left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nu \right) \left(\sum_{l=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_m \right)}{D^{3/2}} \right) \Phi_j ds.$$

It is easy to see that

$$\frac{\partial I_6}{\partial \alpha_m} = \frac{\partial}{\partial \alpha_m} \sum_{e_k \in \Gamma} \int_{e_k} \sigma_k \frac{p_k^2}{|e_k|} \left(\sum_{l=1}^{N} \alpha_l [\Phi_l] \right) [\Phi_j] \; ds = \sum_{e_k \in \Gamma} \int_{e_k} \sigma_k \frac{p_k^2}{|e_k|} [\Phi_m][\Phi_j] ds.$$

For the integral I_3, we expand it as as in the following.

$$\begin{aligned} I_3 &= \sum_{e_k \in \Gamma_I} \int_{e_k} \{ \frac{1}{D^{3/2}} \left(\frac{\partial \Phi_j}{\partial x_1} \nu_1 + \left(\frac{\partial u_h}{\partial x_2} \right)^2 \frac{\partial \Phi_j}{\partial x_1} \nu_1 - \frac{\partial u_h}{\partial x_1} \frac{\partial u_h}{\partial x_2} \frac{\partial \Phi_j}{\partial x_2} \nu_2 - \frac{\partial u_h}{\partial x_2} \frac{\partial u_h}{\partial x_1} \frac{\partial \Phi_j}{\partial x_1} \nu_1 \right. \\ &\quad \left. + \frac{\partial \Phi_j}{\partial x_2} \nu_2 + \left(\frac{\partial u_h}{\partial x_1} \right)^2 \frac{\partial \Phi_j}{\partial x_2} \nu_2 \right) \} [u_h] ds \\ &= E_1 + E_2 - E_3 - E_4 + E_5 + E_6. \end{aligned} \tag{4.94}$$

First, we compute the derivative of E_1

$$\begin{aligned} \frac{\partial E_1}{\partial \alpha_m} &= \sum_{e_k \in \Gamma_I} \int_{e_k} \{ \frac{-3}{D^{5/2}} \left(\sum_{i=1}^{N} \alpha_l \nabla \Phi_l \cdot \nabla \Phi_m \right) \frac{\partial \Phi_j}{\partial x_1} \nu_1 \} \sum_{i=1}^{N} \alpha_l [\Phi_l] ds \\ &\quad + \sum_{e_k \in \Gamma_I} \int_{e_k} \{ \frac{1}{D^{3/2}} \frac{\partial \Phi_j}{\partial x_1} \nu_1 \} [\Phi_m] ds. \end{aligned}$$

Similarly, we compute for E_5

$$\begin{aligned}\frac{\partial E_5}{\partial \alpha_m} &= \sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{-3}{D^{5/2}}\left(\sum_{i=1}^{N}\alpha_l\nabla\Phi_l\cdot\nabla\Phi_m\right)\frac{\partial\Phi_j}{\partial x_2}\nu_2\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\frac{\partial\Phi_j}{\partial x_2}\nu_2\}[\Phi_m]ds.\end{aligned}$$

Next, we consider E_2 and compute the derivative as

$$\begin{aligned}\frac{\partial E_2}{\partial \alpha_m} &= \sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{-3}{D^{5/2}}\left(\sum_{i=1}^{N}\alpha_l\nabla\Phi_l\cdot\nabla\Phi_m\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)^2\frac{\partial\Phi_j}{\partial x_1}\nu_1\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{2}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_m}{\partial x_2}\frac{\partial\Phi_j}{\partial x_1}\nu_1\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)^2\frac{\partial\Phi_j}{\partial x_1}\nu_1\}[\Phi_m]ds.\end{aligned}$$

Similarly, the derivative of E_6 is computed as

$$\begin{aligned}\frac{\partial E_6}{\partial \alpha_m} &= \sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{-3}{D^{5/2}}\left(\sum_{i=1}^{N}\alpha_l\nabla\Phi_l\cdot\nabla\Phi_m\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)^2\frac{\partial\Phi_j}{\partial x_2}\nu_2\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{2}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\frac{\partial\Phi_m}{\partial x_1}\frac{\partial\Phi_j}{\partial x_2}\nu_2\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)^2\frac{\partial\Phi_j}{\partial x_2}\nu_2\}[\Phi_m]ds.\end{aligned}$$

Now, we consider E_3 and compute its derivative as

$$\begin{aligned}\frac{\partial E_3}{\partial \alpha_m} &= -3\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(\sum_{i=1}^{N}\frac{\alpha_l\nabla\Phi_l\cdot\nabla\Phi_m}{D^{5/2}}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_j}{\partial x_2}\nu_2\}[u_h]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\frac{\partial\Phi_m}{\partial x_2}\frac{\partial\Phi_j}{\partial x_2}\nu_2\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_m}{\partial x_1}\frac{\partial\Phi_j}{\partial x_2}\nu_2\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_j}{\partial x_2}\nu_2\}[\Phi_m]ds.\end{aligned}$$

Finally, we consider E_4 and compute its derivative as

$$\begin{aligned}\frac{\partial E_4}{\partial \alpha_m} &= -3\sum_{e_k\in\Gamma_I}\int_{e_k}\{\left(\sum_{i=1}^{N}\frac{\alpha_l\nabla\Phi_l\cdot\nabla\Phi_m}{D^{5/2}}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_j}{\partial x_1}\nu_1\}[u_h]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\frac{\partial\Phi_m}{\partial x_2}\frac{\partial\Phi_j}{\partial x_1}\nu_1\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_m}{\partial x_1}\frac{\partial\Phi_j}{\partial x_1}\nu_1\}\sum_{i=1}^{N}\alpha_l[\Phi_l]ds \\ &+\sum_{e_k\in\Gamma_I}\int_{e_k}\{\frac{1}{D^{3/2}}\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_1}\right)\left(\sum_{i=1}^{N}\alpha_l\frac{\partial\Phi_l}{\partial x_2}\right)\frac{\partial\Phi_j}{\partial x_1}\nu_1\}[\Phi_m]ds.\end{aligned}$$

We use 13 point Gaussian quadrature formula [32] to evaluate the element-wise integrals and 8 point Gaussian quadrature formula to evaluate the edge-wise integrals.

Now, we use the following algorithm for Newton's method to solve the system (4.91) : For given α^0, find α^k, for $1 \le k \le k_max$, such that

$$\alpha^k = \alpha^{k-1} - \mathbf{J}^{-1}\mathbf{F}(\alpha^{k-1}),$$

where $\mathbf{J}$ is given by (4.92). The intial iterate α^0 is chosen as $\alpha^0 = [0, 0, \cdots, 0]$. We set the maximum number of iterations as $k_{max} = 10$.

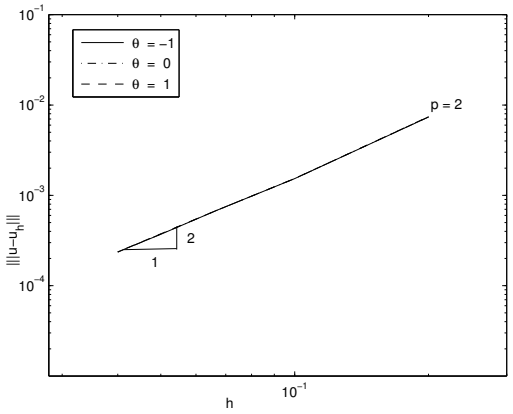
Convergence in the broken H^1-norm. On the sequence of triangulations $\mathcal{T}_h$, we compute error $u - u_h$ in the broken H^1-norm for three values of θ, that is, for $\theta = -1,\ 0$ and 1. The error $|||u - u_h|||$ is plotted against h. We have then computed the numerical order of convergence which illustrate the theoretical order of convergence, see Fig 4.1.

Table 4.1 Convergence of DG schemes in $|||u - u_h|||$

h	$\theta = 1$	$\theta = 0$	$\theta = -1$
1/10	1.52334651e-003	1.55194889e-003	1.52579542e-003
1/15	6.80041075e-004	6.80905440e-004	6.80139245e-004
1/20	3.70429508e-004	3.72832165e-004	3.74638916e-004
1/25	2.37048656e-004	2.35671519e-004	2.36096113e-004

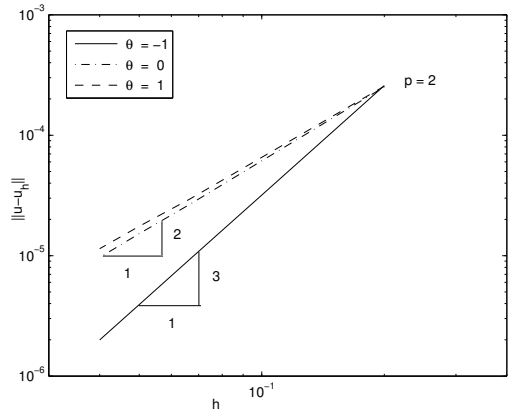
We note that the convergence lines are almost same for each of the method, that is, for $\theta = -1,\ 0$ and 1. We also have presented in Table 4.1 the convergence of DG schemes in the broken H^1-norm for $\theta = -1,\ 0$ and 1. Note that the computed order of convergence in broken H^1-norm is two which confirms the theoretical order of convergence as has been stated in Theorem 4.2.4.

Figure 4.1: convergence of NIPG and SIPG with p-refinement



Convergence in the L^2-norm: We have plotted the L^2-norm of the error $u - u_h$ against h for $\theta = -1,\ 0$ and 1. The computed order of convergence is three which is illustrating the theoretical order of convergence obtained in Theorem 4.2.5, when $\theta = -1$. But for other

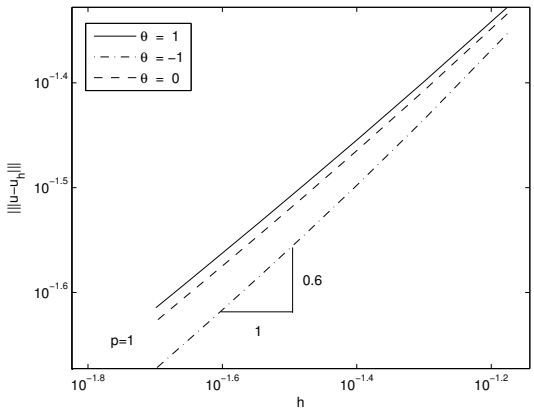
Figure 4.2: convergence of NIPG and SIPG with p-refinement



two methods, that is, when $\theta = 0$ and $\theta = 1$, we have observed that the convergence lines deviate from the convergence line for $\theta = -1$ and the computed order of convergence is two in the L^2-norm, see Fig 4.2. We remark that when $\theta = 0$ and $\theta = 1$, the linearized problem 4.18 is not adjoint consistent and hence the corresponding DG schemes exhibit the suboptimal order of convergence in the L^2-norm.

Numerical Experiments using piecewise linear polynomials: Although, the theoretical results obtained in this chapter require the degree of approximation $p \geq 2$, we have performed some numerical experiments using piecewise linear polynomials, that is, when $p = 1$. The results obtained using piecewise linear polynomials show that there is a sub-optimal convergence in h in this case, see Fig 4.3.

Figure 4.3: convergence of NIPG and SIPG with p-refinement



However, it is difficult to derive the suboptimal order of convergence using the present analysis of this paper and hence, it is subject of our future research.

Chapter 5

LDG Method for Strongly Nonlinear Elliptic Problems

5.1 Introduction

In this Chapter, we study the LDG method for the strongly nonlinear elliptic problem of following type :

$$-\sum_{i=1}^{2} \frac{\partial a_i}{\partial x_i}(\mathbf{x}, u, \nabla u) + f(\mathbf{x}, u, \nabla u) = 0, \qquad \mathbf{x} \in \Omega, \tag{5.1}$$

$$u = g, \qquad \mathbf{x} \in \partial\Omega, \tag{5.2}$$

where Ω is a bounded domain in $\mathbb{R}^2$ with boundary $\partial\Omega$, Problems of this type arise in many practical cases such as mean curvature, subsonic flow problems, etc. We assume that g can be extended to Ω so as to be in $H^{5/2}(\Omega)$ and there exists a unique weak solution u to the problem (5.1)-(5.2) in such a way that $u \in H^{5/2}(\Omega) \cap W^{1,\infty}(\Omega)$. The functions $f,\ a_i : \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R},\ i = 1,\ 2$, are twice continuously differentiable with all the derivatives through the second order are bounded. Further, assume that the matrix $[a^{ij}(\mathbf{x}, u, \mathbf{z})] = \left[\frac{\partial a_i}{\partial z_j}\right]_{i,j=1,2}$ is symmetric and if $\lambda(\mathbf{x}, u, \mathbf{z})$, $\Lambda(x, u, \mathbf{z})$ are minimum and maximum eigenvalues of the matrix $[a^{ij}]$, then for all $\xi \in \mathbb{R}^2 - \{0\}$ and for all $(\mathbf{x}, u, \mathbf{z}) \in \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^2$,

$$0 < \lambda(\mathbf{x}, u, \mathbf{z})|\xi|^2 \leq \sum_{i,j=1}^{2} a^{ij}(\mathbf{x}, u, \mathbf{z})\xi_i\xi_j \leq \Lambda(x, u, \mathbf{z})|\xi|^2. \tag{5.3}$$

In this Chapter, an *hp*-LDG method is applied to the problem (5.1) -(5.2) and error estimates which are optimal in h and slightly suboptimal in p are derived. The results proved in this Chapter are same as in the linear case, see [57]. Assuming *hp*-quasiuniformity condition on the mesh, existence of a solution to the discrete problem is proved using Brouwer fixed point theorem for small h (mesh size). Moreover, the Lipschitz continuity of the discrete solution map shows the uniqueness of the discrete problem.

The rest of the Chapter is organized as follows. Section 5.3 is devoted to the LDG method, and a priori error estimates for the method.

We assume that for given $y \in L^2(\Omega)$, there is a unique $\phi \in H^2(\Omega)$ satisfying the following elliptic problem

$$-\nabla \cdot (\mathbf{a_q}(u,\mathbf{q})\nabla\phi + f_{\mathbf{q}}(u,\mathbf{q})\phi) + \mathbf{a}_u(u,\mathbf{q}) \cdot \nabla\phi + f_u(u,\mathbf{q})\phi = y \quad \text{in } \Omega, \tag{5.4}$$

$$\phi = 0 \quad \text{on } \partial\Omega. \tag{5.5}$$

Further, assume that ϕ satisfies the following elliptic regularity

$$\|\phi\|_{H^2(\Omega)} \leq C\|y\|_{L^2(\Omega)}. \tag{5.6}$$

This would be guaranteed if $(f_u(u,\mathbf{z}) - \nabla \cdot f_{\mathbf{q}}(u,\mathbf{z})) \geq 0,\ \forall\ (u,\mathbf{z}) \in \mathbb{R} \times \mathbb{R}^2$ and Ω is of the class $C^{1,1}$, see [40].

Since $\mathbf{a}$, f are twice continuously differentiable functions with all the derivatives through the second order are bounded, the remainder terms in the Taylor series expansions (4.5)-(4.6) that $\tilde{\mathbf{a}}_u$, $\tilde{\mathbf{a}}_{\mathbf{q}}$, $\tilde{\mathbf{a}}_{uu}$, $\tilde{\mathbf{a}}_{\mathbf{qq}}$, $\tilde{\mathbf{a}}_{u\mathbf{q}}$, $\tilde{\mathbf{a}}_{\mathbf{q}u}$, $\tilde{f}_u$, $\tilde{f}_{\mathbf{q}}$, $\tilde{f}_{uu}$, $\tilde{f}_{\mathbf{qq}}$, $\tilde{f}_{\mathbf{q}u}$ and $\tilde{f}_{u\mathbf{q}}$ are in $L^\infty(\Omega \times \mathbb{R} \times \mathbb{R}^2)$. We denote C_a by

$$C_a = \max\{\|\mathbf{a}\|_{W^2_\infty(\Omega\times\mathbb{R}\times\mathbb{R}^2)},\ \|f\|_{W^2_\infty(\Omega\times\mathbb{R}\times\mathbb{R}^2)}\}. \tag{5.7}$$

REMARK 5.1.1 *For our subsequent analysis, it is sufficient to assume that* $\mathbf{a}$ *and* f *are locally bounded. In fact, it is enough to assume that* $\mathbf{a}$ *and* f *along with its derivatives are bounded in a ball around* u*, see Remark 5.2.1.*

5.2 Local Discontinuous Galerkin (LDG) method

To define the LDG method, we first rewrite the equation (5.1) as a problem of first order system of equations. In order to achieve this, we now introduce auxiliary variable $\mathbf{q} = \nabla u$

and $\boldsymbol{\sigma} = a(u)\mathbf{q}$ and rewrite (5.1) -(5.2) as

$$\begin{aligned} \mathbf{q} &= \nabla u \ \text{ in } \ \Omega, && (5.1)\\ \boldsymbol{\sigma} &= \mathbf{a}(u,\mathbf{q}) \quad \text{in } \ \Omega, && (5.2)\\ -\nabla\cdot\boldsymbol{\sigma} + f(u,\mathbf{q}) &= 0 \quad \text{in } \ \Omega, && (5.3)\\ u &= g \quad \text{on} \quad \partial\Omega. && (5.4) \end{aligned}$$

We multiply the equation (5.1) by $\mathbf{w} \in \mathbf{W}$, the equation (5.2) by $\boldsymbol{\tau} \in \mathbf{W}$ and the equation (5.3) by $v \in V$ and then integrate over the element $K \in T_h$. Using the integration by parts formula, we obtain

$$\int_K \mathbf{q}\cdot\mathbf{w}dx + \int_K u\nabla\cdot\mathbf{w}dx - \int_{\partial K} u\mathbf{w}\cdot\nu_K ds = 0, \tag{5.5}$$

$$\int_K \mathbf{a}(u,\mathbf{q})\cdot\boldsymbol{\tau}dx - \int_K \boldsymbol{\sigma}\cdot\boldsymbol{\tau}dx = 0, \tag{5.6}$$

and

$$\int_K \boldsymbol{\sigma}\cdot\nabla v dx - \int_{\partial K} \boldsymbol{\sigma}\cdot\nu_K v ds + \int_K f(u,\mathbf{q})v dx = 0. \tag{5.7}$$

Note that there may be difficulty in defining u and $\boldsymbol{\sigma}$ on ∂K. Therefore, this is just an initial formulation which help us in defining the approximate method given below. The following approximate solution $(u_h,\ \mathbf{q}_h,\ \boldsymbol{\sigma}_h) \in \mathrm{Z}_p(K) \times \mathrm{Z}_p(K)^2 \times \mathrm{Z}_p(K)^2$ is defined using above weak formulation, that is by imposing that for all K,

$$\int_K \mathbf{q}_h\cdot\mathbf{w}_h dx + \int_K u_h\nabla\cdot\mathbf{w}_h dx - \int_{\partial K} \hat{u}\mathbf{w}_h\cdot\nu_K ds = 0, \quad \mathbf{w}_h \in \mathrm{Z}_p(K)^2, \tag{5.8}$$

$$\int_K \mathbf{a}(u_h,\mathbf{q}_h)\cdot\boldsymbol{\tau}_h - \int_K \boldsymbol{\sigma}_h\cdot\boldsymbol{\tau}_h dx = 0, \quad \boldsymbol{\tau}_h \in \mathrm{Z}_p(K)^2, \tag{5.9}$$

and

$$\int_K \boldsymbol{\sigma}_h\cdot\nabla v_h dx - \int_{\partial K} \hat{\boldsymbol{\sigma}}\cdot\nu_K v_h ds + \int_K f(u_h,\mathbf{q}_h)v_h dx = 0, \quad v_h \in V_h, \tag{5.10}$$

where the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ have to be suitably chosen in order to ensure the stability of the method and also to improve the order of convergence. The following choice

of numerical fluxes are used in solving the linear elliptic problems. If $e_k \in \Gamma_I$, then the numerical fluxes are defined on e_k as :

$$\hat{u}(u_h, \boldsymbol{\sigma}_h) = \{u_h\} + \mathbf{C}_{12} \cdot [\![u_h]\!] - C_{22}[\![\boldsymbol{\sigma}_h]\!], \tag{5.11}$$

$$\hat{\boldsymbol{\sigma}}(u_h,\ \boldsymbol{\sigma}_h) = \{\boldsymbol{\sigma}_h\} - C_{11}[\![u_h]\!] - \mathbf{C}_{12}[\![\boldsymbol{\sigma}_h]\!], \tag{5.12}$$

and if $e_k \in \Gamma_\partial$, then the numerical fluxes are taken as :

$$\hat{u} = g, \tag{5.13}$$

$$\hat{\boldsymbol{\sigma}} = \boldsymbol{\sigma}_h - C_{11}(u_h - g)\nu, \tag{5.14}$$

with $C_{11} \in \mathbb{R}^+$ on each $e_k \in \Gamma$, $C_{22} \in \mathbb{R}^+$ and $\mathbf{C}_{12} \in \mathbb{R}^2$ on $e_k \in \Gamma_I$. We set $\mathbf{C}_{12} = 0$ on $e_k \in \Gamma_\partial$. The numerical fluxes are *conservative* since they are single valued on $e_k \in \Gamma_I$, that is, on $e_k \in \Gamma_I$,

$$[\![\hat{u}]\!] = 0, \qquad [\![\hat{\boldsymbol{\sigma}}]\!] = 0. \tag{5.15}$$

and *consistent* since the following holds for smooth u and $\boldsymbol{\sigma}$:

$$\hat{u}(u) = u, \tag{5.16}$$

$$\hat{\boldsymbol{\sigma}}(u,\ \boldsymbol{\sigma}) = \boldsymbol{\sigma}. \tag{5.17}$$

We sum (5.8)-(5.10) over all elements $K \in T_h$. Then using the conservative property (5.15) and the definition of numerical fluxes, we obtain

$$\int_\Omega \mathbf{q}_h \cdot \mathbf{w}_h dx + \sum_{i=1}^{N_h} \int_{K_i} u_h \nabla \cdot \mathbf{w}_h dx - \int_{\Gamma_I} (\{u_h\} + \mathbf{C}_{12}.[u_h] - C_{22}[\![\boldsymbol{\sigma}_h]\!])[\![\mathbf{w}_h]\!] ds$$
$$= \int_{\Gamma_\partial} g \mathbf{w}_h \cdot \nu ds, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{5.18}$$

$$\int_\Omega \mathbf{a}(u_h, \mathbf{q}_h) \cdot \boldsymbol{\tau}_h dx - \int_\Omega \boldsymbol{\sigma}_h \cdot \boldsymbol{\tau}_h dx = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h, \tag{5.19}$$

$$\sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\sigma}_h \cdot \nabla v_h dx - \int_\Gamma (\{\boldsymbol{\sigma}_h\} - C_{11}[\![u_h]\!] - \mathbf{C}_{12}[\![\boldsymbol{\sigma}_h]\!])[\![v_h]\!] ds + \int_\Omega f(u_h, \mathbf{q}_h) v_h dx$$
$$= \int_{\Gamma_\partial} C_{11} g v_h ds, \quad v_h \in V_h. \tag{5.20}$$

Let $z \in L^2(\Omega)$ and $(\phi,\ \mathbf{p}),\ (v,\ \mathbf{w}) \in V \times \mathbf{W}$. Set $B_1 : \mathbf{W} \times \mathbf{W} \to \mathbb{R}$ as

$$B_1(\mathbf{p}, \mathbf{w}) = \int_\Omega \mathbf{p} \cdot \mathbf{w}\ dx,$$

$B_2 : \mathbf{W} \times V \to \mathbb{R}$ as

$$\begin{aligned} B_2(\mathbf{p}, v) &= \sum_{i=1}^{N_h} \int_{K_i} \mathbf{p} \cdot \nabla v\ dx - \int_\Gamma (\{\mathbf{p}\} - \mathbf{C}_{12}[\![\mathbf{p}]\!])[\![v]\!]\ ds \\ &= -\sum_{i=1}^{N_h} \int_{K_i} v \nabla \cdot \mathbf{p}\ dx + \int_{\Gamma_I} (\{v\} + \mathbf{C}_{12} \cdot [\![v]\!])[\![\mathbf{p}]\!]\ ds, \end{aligned}$$

and $J_1 : V \times V \to \mathbb{R}$, $J_2 : \mathbf{W} \times \mathbf{W} \to \mathbb{R}$ as

$$J_1(\phi, v) = \int_\Gamma C_{11}[\![\phi]\!][\![v]\!]\ ds, \quad J_2(\mathbf{p}, \mathbf{w}) = \int_\Gamma C_{22}[\![\mathbf{p}]\!][\![\mathbf{w}]\!]\ ds.$$

We also define the following linear functionals $L_1 : \mathbf{W} \to \mathbb{R}$ and $L_2 : V \to \mathbb{R}$ as

$$L_1(\mathbf{w}) = \int_{\Gamma_\partial} g\mathbf{w}.\nu\ ds, \qquad \text{and} \qquad L_2(v) = \int_{\Gamma_\partial} C_{11} g v\ ds.$$

Using the above definitions, we write the LDG method for the problem (5.1)-(5.2) in compact form : Find $(u_h,\ \mathbf{q}_h,\ \boldsymbol{\sigma}_h) \in V_h \times \mathbf{W}_h \times \mathbf{W}_h$ such that

$$B_1(\mathbf{q}_h, \mathbf{w}_h) - B_2(\mathbf{w}_h, u_h) + J_2(\boldsymbol{\sigma}_h, \mathbf{w}_h) = L_1(\mathbf{w}_h), \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{5.21}$$

$$(\mathbf{a}(u_h, \mathbf{q}_h), \boldsymbol{\tau}_h) - B_1(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = 0, \quad \boldsymbol{\tau}_h \in \mathbf{W}_h, \tag{5.22}$$

$$B_2(\boldsymbol{\sigma}_h, v_h) + J_1(u_h, v_h) + (f(u_h, \mathbf{q}_h), v_h) = L_2(v_h), \quad v_h \in V_h. \tag{5.23}$$

Since the numerical fluxes $\hat{u}$ and $\hat{\boldsymbol{\sigma}}$ are consistent, we note that the following identity holds for all $(v,\ \boldsymbol{\tau},\ \mathbf{w}) \in V \times \mathbf{W} \times \mathbf{W}$.

$$B_1(\mathbf{q}, \mathbf{w}) - B_2(\mathbf{w}, u) + J_2(\boldsymbol{\sigma}, \mathbf{w}) = L_1(\mathbf{w}), \quad \mathbf{w} \in \mathbf{W}, \tag{5.24}$$

$$(\mathbf{a}(u, \mathbf{q}), \boldsymbol{\tau}) - B_1(\boldsymbol{\sigma}, \boldsymbol{\tau}) = 0, \quad \boldsymbol{\tau} \in \mathbf{W}, \tag{5.25}$$

$$B_2(\boldsymbol{\sigma}, v) + J_1(u, v) + (f(u, \mathbf{q}), v) = L_2(v), \quad v \in V. \tag{5.26}$$

In order to derive *a priori* error estimates and to prove existence of a unique approximate solution to the problem (5.21)-(5.23), we first note that using (5.21)-(5.26)

$$B_1(\mathbf{q} - \mathbf{q}_h, \mathbf{w}_h) - B_2(\mathbf{w}_h, u - u_h) + J_2(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{w}_h) = 0, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{5.27}$$

$$\int_\Omega (\mathbf{a}(u, \mathbf{q}) - \mathbf{a}(u_h, \mathbf{q}_h)) \cdot \boldsymbol{\tau}_h - B_1(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\tau}_h) = 0, \quad \boldsymbol{\tau}_h \in \boldsymbol{\tau}_h, \tag{5.28}$$

$$B_2(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, v_h) + J_1(u - u_h, v_h) + \int_\Omega (f(u, \mathbf{q}) - f(u_h, \mathbf{q}_h))\, v_h = 0, \quad v_h \in V_h. \tag{5.29}$$

Using (4.6), we rewrite (5.28) as

$$\int_\Omega \mathbf{a}_u(u,\mathbf{q})(u-u_h)\cdot\boldsymbol{\tau}_h + \int_\Omega \mathbf{a}_{\mathbf{q}}(u,\mathbf{q})(\mathbf{q}-\mathbf{q}_h)\cdot\boldsymbol{\tau}_h - B_1(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\boldsymbol{\tau}_h) = \int_\Omega \tilde{R}_{\mathbf{a}}(u-u_h,\mathbf{q}-\mathbf{q}_h)\cdot\boldsymbol{\tau}_h dx.$$

Similarly, using (4.5), we obtain from (5.29)

$$B_2(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,v_h) + J_1(u-u_h,v_h) + \int_\Omega f_u(u,\mathbf{q})(u-u_h)v_h + \int_\Omega f_{\mathbf{q}}(u,\mathbf{q})\cdot(\mathbf{q}-\mathbf{q}_h)v_h = \int_\Omega \tilde{R}_f(u-u_h,\mathbf{q}-\mathbf{q}_h)v_h dx.$$

For notational simplicity, we introduce the following for $\boldsymbol{\tau},\ \mathbf{p}\in\mathbf{W}$ and $\phi,\ v\in V$

$$\begin{aligned} A_u(u,\mathbf{q};\phi,\boldsymbol{\tau}) &= \int_\Omega \mathbf{a}_u(u,\mathbf{q})\phi\cdot\boldsymbol{\tau}\ dx,\\ A_{\mathbf{q}}(u,\mathbf{q};\mathbf{p},\boldsymbol{\tau}) &= \int_\Omega \mathbf{a}_{\mathbf{q}}(u,\mathbf{q})\mathbf{p}\cdot\boldsymbol{\tau}dx,\\ F_u(u,\mathbf{q};\phi,v) &= \int_\Omega f_u(u,\mathbf{q})\phi v\ dx \end{aligned}$$

and

$$F_{\mathbf{q}}(u,\mathbf{q};\mathbf{p},v) = \int_\Omega f_{\mathbf{q}}(u,\mathbf{q})\cdot\mathbf{p}v dx$$

Hence, the equations (5.27)-(5.29) take the form

$$B_1(\mathbf{q}-\mathbf{q}_h,\mathbf{w}_h) - B_2(\mathbf{w}_h,u-u_h) + J_2(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\mathbf{w}_h) = 0,\quad \mathbf{w}_h\in\mathbf{W}_h, \tag{5.30}$$

$$A_u(u,\mathbf{q};u-u_h,\boldsymbol{\tau}_h) + A_{\mathbf{q}}(u,\mathbf{q};\mathbf{q}-\mathbf{q}_h,\boldsymbol{\tau}_h) - B_1(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\boldsymbol{\tau}_h) = \left(\tilde{R}_{\mathbf{a}}(u-u_h,\mathbf{q}-\mathbf{q}_h),\boldsymbol{\tau}_h\right),\quad \boldsymbol{\tau}_h\in\boldsymbol{\tau}_h, \tag{5.31}$$

$$B_2(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,v_h) + J_1(u-u_h,v_h) + F_u(u,\mathbf{q};u-u_h,v_h) + F_{\mathbf{q}}(u,\mathbf{q};\mathbf{q}-\mathbf{q}_h,v_h) = \left(\tilde{R}_f(u-u_h,\mathbf{q}-\mathbf{q}_h),v_h\right),\quad v_h\in V_h. \tag{5.32}$$

We state the following lemma without proof. Since, the matrix $[\mathbf{a}^{ij}(u,\mathbf{q})]$ is positive definite, the proof follows by an appeal to Cauchy-Schwarz inequality and using the assumption on u and $\mathbf{a}(u,\mathbf{q})$.

LEMMA 5.2.1 *There exist positive constants C_1 and C_2 such that for all $(v,\ \mathbf{w})\in V\times\mathbf{W}$,*

$$A_u(u,\mathbf{q};v,\mathbf{w}) + A_{\mathbf{q}}(u,\mathbf{q};\mathbf{w},\mathbf{w}) + F_u(u,\mathbf{q};v,v) + F_{\mathbf{q}}(u,\mathbf{q};\mathbf{w},v) \geq C_1\left(\|\mathbf{w}\|^2 + J_1(v,v)\right) - C_2\|v\|^2.$$

5.2.1 Existence and Uniqueness of the Discrete Problem.

For a given $(z, \boldsymbol{\theta}) \in V_h \times \mathbf{W}_h$, we define a map $S_h : V_h \to V_h$ by $S_h((z, \boldsymbol{\theta})) = (u_l, \mathbf{q}_l) \in V_h \times \mathbf{W}_h$ and $\boldsymbol{\sigma}_l \in \mathbf{W}_h$ satisfying

$$B_1(\mathbf{q} - \mathbf{q}_l, \mathbf{w}_h) - B_2(\mathbf{w}_h, u - u_l) + J_2(\boldsymbol{\sigma} - \boldsymbol{\sigma}_l, \mathbf{w}_h) = 0, \quad \mathbf{w}_h \in \mathbf{W}_h, \tag{5.33}$$

$$\begin{aligned} &A_u(u, \mathbf{q}; u - u_l, \boldsymbol{\tau}_h) + A_{\mathbf{q}}(u, \mathbf{q}; \mathbf{q} - \mathbf{q}_l, \boldsymbol{\tau}_h) - B_1(\boldsymbol{\sigma} - \boldsymbol{\sigma}_l, \boldsymbol{\tau}_h) \\ &\quad = \left(\tilde{R}_{\mathbf{a}}(u - z, \mathbf{q} - \boldsymbol{\theta}), \boldsymbol{\tau}_h\right), \quad \boldsymbol{\tau}_h \in \mathbf{W}_h, \end{aligned} \tag{5.34}$$

$$\begin{aligned} &B_2(\boldsymbol{\sigma} - \boldsymbol{\sigma}_l, v_h) + J_1(u - u_l, v_h) + F_u(u, \mathbf{q}; u - u_l, v_h) \\ &\quad + F_{\mathbf{q}}(u, \mathbf{q}; \mathbf{q} - \mathbf{q}_l, v_h) = \left(\tilde{R}_f(u - z, \mathbf{q} - \boldsymbol{\theta}), v_h\right), \quad v_h \in V_h. \end{aligned} \tag{5.35}$$

We write $e_u = u - u_l = \xi_u - \eta_u$, where $\xi_u = \Pi u - u_l$ and $\eta_u = \Pi u - u$. Similarly $\mathbf{e}_q = \mathbf{q} - \mathbf{q}_l = \boldsymbol{\xi}_q - \boldsymbol{\eta}_q$ and $\mathbf{e}_\sigma = \boldsymbol{\sigma} - \boldsymbol{\sigma}_l = \boldsymbol{\xi}_\sigma - \boldsymbol{\eta}_\sigma$, where $\boldsymbol{\xi}_q = I_h\mathbf{q} - \mathbf{q}_l$, $\boldsymbol{\eta}_q = I_h\mathbf{q} - \mathbf{q}$, $\boldsymbol{\xi}_\sigma = \Pi\boldsymbol{\sigma} - \boldsymbol{\sigma}_l$ and $\boldsymbol{\eta}_\sigma = \Pi\boldsymbol{\sigma} - \boldsymbol{\sigma}$. With these notations rewrite (5.33)-(5.35) as

$$\begin{aligned} &B_1(\boldsymbol{\xi}_q, \mathbf{w}_h) - B_2(\mathbf{w}_h, \xi_u) + J_2(\boldsymbol{\xi}_\sigma, \mathbf{w}_h) = B_1(\boldsymbol{\eta}_q, \mathbf{w}_h) - B_2(\mathbf{w}_h, \eta_u) \\ &\qquad + J_2(\boldsymbol{\eta}_\sigma, \mathbf{w}_h), \quad \mathbf{w}_h \in \mathbf{W}_h, \end{aligned} \tag{5.36}$$

$$\begin{aligned} &A_u(u, \mathbf{q}; \xi_u, \boldsymbol{\tau}_h) + A_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\xi}_q, \boldsymbol{\tau}_h) - B_1(\boldsymbol{\xi}_\sigma, \boldsymbol{\tau}_h) = A_u(u, \mathbf{q}; \eta_u, \boldsymbol{\tau}_h) + \\ &\quad A_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\eta}_q, \boldsymbol{\tau}_h) - B_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\tau}_h) + \left(\tilde{R}_{\mathbf{a}}(u - z, \mathbf{q} - \boldsymbol{\theta}), \boldsymbol{\tau}_h\right), \quad \boldsymbol{\tau}_h \in \mathbf{W}_h, \end{aligned} \tag{5.37}$$

$$\begin{aligned} &B_2(\boldsymbol{\xi}_\sigma, v_h) + J_1(\xi_u, v_h) + F_u(u, \mathbf{q}; \xi_u, v_h) + F_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\xi}_q, v_h) = B_2(\boldsymbol{\eta}_\sigma, v_h) + \\ &\quad J_1(\eta_u, v_h) + F_u(u, \mathbf{q}; \eta_u, v_h) + F_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\eta}_q, v_h) + \left(\tilde{R}_f(u - u_h, \mathbf{q} - \mathbf{q}_h), v_h\right), \quad v_h \in V_h. \end{aligned} \tag{5.38}$$

First we show that there is a δ such that S_h maps $O_\delta(\Pi u, I_h\mathbf{q})$ into itself, where

$$O_\delta(\Pi u, I_h\mathbf{q}) = \{(z, \boldsymbol{\theta}) \in V_h \times \mathbf{W}_h : \|(z, \boldsymbol{\theta}) - (\Pi u, I_h\mathbf{q})\|_+ \leq \delta\},$$

where $\|(\cdot, \cdot)\|_+$ is defined as for given any $(v, \mathbf{w}) \in V \times \mathbf{W}$, $\|(v, \mathbf{w})\|_+ = \|v\| + \|\mathbf{w}\|$. We mention here that δ is defined as

$$\begin{aligned} \delta = \frac{1}{h^\epsilon} \Bigg[&\left((J_1(\eta_u, \eta_u) + J_2(\boldsymbol{\eta}_\sigma, \boldsymbol{\eta}_\sigma)\right)^{1/2} + \left(\max_{1 \leq i \leq N_h} \frac{h_i^{1/2}}{p_i^{1/2}}\right) |u - I_h u|_{1,h} \\ &+ \|\boldsymbol{\eta}_q\| + \|\eta_u\|_{L^4(\Omega)} + \left(\sum_{e_k \in \Gamma} \int_{e_k} \left(|\{\boldsymbol{\eta}_\sigma\}|^2 + |\{\eta_u\}|^2\right) ds\right)^{1/2} \Bigg]. \end{aligned} \tag{5.39}$$

The following lemma is useful in our subsequent analysis. The proof is an easy consequence of Lemma 1.2.1 and Lemma 1.2.5.

LEMMA 5.2.2 *There is a constant C which is independent of h and p such that the following estimates hold :*

$$\begin{aligned} J_1(\eta_u,\eta_u)+J_2(\boldsymbol{\eta}_\sigma,\boldsymbol{\eta}_\sigma) &\le C\left(\sum_{i=1}^{N_h}\frac{h_i^{2\mu_i^+-1}}{p_i^{2s_i-2}}\|\boldsymbol{\sigma}\|^2_{H^{s_i}(K_i)^2}+\frac{h_i^{2\mu_i^*+1}}{p_i^{2s_i}}\|u\|^2_{H^{s_i+1}(K_i)}\right), \\ |u-I_hu|_{1,h} &\le C\left(\sum_{i=1}^{N_h}\frac{h_i^{2\mu_i^*}}{p_i^{2s_i}}\|u\|^2_{H^{s_i+1}(K_i)}\right), \\ ||\eta_u||^2_{L^4(\Omega)}+\sum_{e_k\in\Gamma}\int_{e_k}|\{\eta_u\}|^2ds &\le \left(\sum_{i=1}^{N_h}\frac{h_i^{2\mu_i^*+1}}{p_i^{2s_i}}\|u\|^2_{H^{s_i+1}(K_i)}\right), \end{aligned}$$

and

$$\sum_{e_k\in\Gamma}\int_{e_k}|\{\boldsymbol{\eta}_\sigma\}|^2ds \le C\left(\sum_{i=1}^{N_h}\frac{h_i^{2\mu_i^+-1}}{p_i^{2s_i-2}}\|\boldsymbol{\sigma}\|^2_{H^{s_i}(K_i)^2}\right),$$

where $\mu_i^+=min\{s_i,p_i+1\}$ and $\mu_i^*=min\{s_i,p_i\}$. ■

Using Lemma 5.2.2, the δ which is defined in (5.39) is bounded as follows. If $u\in H^{5/2}(\Omega)$ and $\boldsymbol{\sigma}\in H^2(\Omega)$, then

$$\delta \le CC_u\frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{h_i^{3/2}}{p_i}\right), \qquad \text{where } C_u=\max\{\|u\|_{H^{5/2}(\Omega)},\ \|\boldsymbol{\sigma}\|_{H^2(\Omega)^2}\}. \tag{5.40}$$

LEMMA 5.2.3 *Let $v_h\in V_h$ and $\mathbf{w}_h\in\mathbf{W}_h$. Then, there is a constant $C>0$ such that*

$$\|v_h\|_{L^4(\Omega)} \le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|v_h\|_{L^2(\Omega)}.$$

and

$$\|\mathbf{w}_h\|_{L^4(\Omega)^2} \le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|\mathbf{w}_h\|_{L^2(\Omega)^2}.$$

Proof. Using the inverse inequality (1.20), we obtain

$$\begin{aligned}
\|v_h\|_{L^4(\Omega)} &= \left(\sum_{i=1}^{N_h}\|v_h\|^4_{L^4(K_i)}\right)^{1/4} \leq C\left(\sum_{i=1}^{N_h}\frac{p_i^2}{h_i^2}\|v_h\|^4_{L^2(K_i)}\right)^{1/4} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\sum_{i=1}^{N_h}\|v_h\|^4_{L^2(K_i)}\right)^{1/4} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\max_{1\leq i\leq N_h}\|v_h\|^{1/2}_{L^2(K_i)}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{L^2(K_i)}\right)^{1/4} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\max_{1\leq i\leq N_h}\left(\int_{K_i}v_h^2dx\right)^{1/4}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{L^2(K_i)}\right)^{1/4} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\sum_{i=1}^{N_h}\int_{K_i}v_h^2dx\right)^{1/4}\left(\sum_{i=1}^{N_h}\|v_h\|^2_{L^2(K_i)}\right)^{1/4} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|v_h\|_{L^2(\Omega)}.
\end{aligned}$$

A similar proof yields the second inequality of the lemma. This completes the rest of the proof. ■

REMARK 5.2.1 *In our subsequent analysis, it is enough to assume that the remainder terms* $\tilde{\mathbf{a}}_u(z,\boldsymbol{\theta})$, $\tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})$, $\tilde{\mathbf{a}}_{\mathbf{q}}(z,\boldsymbol{\theta})$ *and* $\tilde{\mathbf{a}}_{\mathbf{qq}}(z,\boldsymbol{\theta})$ *in the Taylor series expansions (4.5)-(4.6) of* $\mathbf{a}$ *and* f *are bounded for* $(z,\boldsymbol{\theta})\in O_\delta(\Pi u, I_h\mathbf{q})$. *We show, in the following that asymptotically only the values of* $z(\mathbf{x})\in[m_u-\delta^*, M_u+\delta^*]$ *where* $m_u = inf\,\{u(\mathbf{x}):\mathbf{x}\in\bar{\Omega}\}$ *and* $M_u = sup\,\{u(\mathbf{x}):\mathbf{x}\in\bar{\Omega}\}$ *are considered to derive these bounds. Similarly, asymptotically the values of* $\boldsymbol{\theta}_i(\mathbf{x})\in[m_u^1-\delta^*, M_u^1+\delta^*]$, *where* $m_u^1 = inf\,\{\nabla u(\mathbf{x}):\mathbf{x}\in\bar{\Omega}\}$ *and* $M_u^1 = sup\,\{\nabla u(\mathbf{x}):\mathbf{x}\in\bar{\Omega}\}$ *are considered. To prove this, we use the inverse inequality (1.20), Lemma 1.2.6 and Lemma 1.2.1 to find for* $y\in\mathcal{O}_\delta(I_hu)$ *that*

$$\begin{aligned}
\|z-u\|_{L^\infty(\Omega)} &\leq \|z-I_hu\|_{L^\infty(\Omega)} + \|I_hu-u\|_{L^\infty(\Omega)} \\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\|z-I_hu\|_{L^2(\Omega)} + \|I_hu-u\|_{L^\infty(\Omega)} \\
&\leq CC_u\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right) + \|I_hu-u\|_{L^\infty(\Omega)} \\
&\leq CC_uh^{1/2-\epsilon} + \frac{h^{1/2}}{p^{1/2}}\|u\|_{H^{5/2}}.
\end{aligned}$$

Similarly, it is easy to show that

$$\begin{aligned}\|\boldsymbol{\theta}-\mathbf{q}\|_{L^\infty(\Omega)} &\leq |\boldsymbol{\theta}-I_h\mathbf{q}|_{L^\infty(\Omega)}+|I_h\mathbf{q}-\mathbf{q}|_{L^\infty(\Omega)}\\ &\leq CC_u h^{1/2-\epsilon}+\frac{h^{1/2}}{p^{1/2}}\|u\|_{H^{5/2}}.\end{aligned}$$

Therefore, for sufficiently small h, $\|z\|_{L^\infty(\Omega)}\leq\delta^*+\|u\|_{L^\infty(\Omega)}$ *and* $\|\boldsymbol{\theta}\|_{L^\infty(\Omega)}\leq\delta^*+\|\nabla u\|_{L^\infty(\Omega)}$, *where* $0<\delta^*<1$. *Now, since the nonlinear functions* $\mathbf{a}_u$, $\mathbf{a}_{uu}$, $\mathbf{a}_{\mathbf{z}}$ *and* $\mathbf{a}_{\mathbf{zz}}$ *are continuous, they map the compact set* $[m_u-\delta^*,M_u+\delta^*]\times[m_u^1-\delta^*,M_u^1+\delta^*]$ *into a compact set. Hence, the results in the subsequent Sections remain valid when* $\mathbf{a}(z,\boldsymbol{\theta})$, $\mathbf{a}_u(z,\boldsymbol{\theta})$, $\mathbf{a}_{uu}(z,\boldsymbol{\theta})$, $\mathbf{a}_{\mathbf{z}}(z,\boldsymbol{\theta})$ *and* $\mathbf{a}_{\mathbf{zz}}(z,\boldsymbol{\theta})$ *are bounded for bounded* $u\in W_\infty^1(\Omega)$. *Similar arguments can be applied to* f.

LEMMA 5.2.4 *Let the assumption* ($\mathbf{Q}$), *that is, hp-quasiuniformity hold. Then, for given any* $(z,\theta)\in O_\delta(\Pi u, I_h\mathbf{q})$ *and* $(v,\boldsymbol{\tau})\in V_h\times\mathbf{W}_h$, *there exists a positive constant* C *such that for any* $0<\epsilon\leq 1/4$

$$|\int_\Omega \tilde{R}_{\mathbf{a}}(u-z,q-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx|\leq CC_aC_uC_Qh^{1/2-\epsilon}\ \delta\ \|\boldsymbol{\tau}\|,$$

and

$$|\int_\Omega \tilde{R}_f(u-z,q-\boldsymbol{\theta})vdx|\leq CC_aC_uC_Qh^{1/2-\epsilon}\ \delta\ \|v\|.$$

Proof. In order to prove the the first inequality, we rewrite using the definition of $\tilde{R}_{\mathbf{a}}$

$$\begin{aligned}\int_\Omega \tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx &= \int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx \qquad (5.41)\\ &+\int_\Omega \tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(u-z)(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx+\int_\Omega(\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{\mathbf{qq}}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx.\end{aligned}$$

For the first term on the right-hand side of (5.41), we bound it as

$$\begin{aligned}|\int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx| &\leq C_a\|z-u\|^2_{L^4(\Omega)}\ \|\boldsymbol{\tau}\|\\ &\leq C_a\left(\|z-\Pi u\|^2_{L^4(\Omega)}+\|\eta_u\|^2_{L^4(\Omega)}\right)\|\boldsymbol{\tau}\|. \qquad (5.42)\end{aligned}$$

Then, using Lemma 5.2.3 and the assumption (**Q**), we obtain

$$\begin{aligned}
\|z-\Pi u\|^2_{L^4(\Omega)} &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\|z-\Pi u\|^2_{L^2(\Omega)} \qquad (5.43)\\
&\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\delta^2\\
&\leq CC_u\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i}{h_i}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\delta\\
&\leq CC_uC_Qh^{1/2-\epsilon}\delta, \qquad (5.44)
\end{aligned}$$

and using the identity (5.39), we find that

$$\begin{aligned}
\|\eta_u\|^2_{L^4(\Omega)} &\leq CC_uh^{2\epsilon}\delta^2\\
&\leq CC_uh^\epsilon\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\delta\\
&\leq CC_uh^{3/2+\epsilon}\delta. \qquad (5.45)
\end{aligned}$$

We substitute (5.44)-(5.45) in (5.42) to obtain

$$|\int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx| \leq CC_aC_uC_Qh^{1/2-\epsilon}\delta\|\boldsymbol{\tau}\|. \qquad (5.46)$$

Using the inverse inequality (1.20), the second term on the right-hand side of (5.41) is estimated as follows:

$$\begin{aligned}
|\int_\Omega \tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(u-z)(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx| &\leq CC_a\sum_{i=1}^{N_h}\|u-z\|_{L^4(K_i)}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(K_i)^2}\|\boldsymbol{\tau}\|_{L^4(K_i)^2}\\
&\leq CC_a\sum_{i=1}^{N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\|u-z\|_{L^4(K_i)}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(K_i)^2}\|\boldsymbol{\tau}\|_{L^2(K_i)^2}\\
&\leq CC_a\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|u-z\|_{L^4(\Omega)}\|\mathbf{q}-\boldsymbol{\theta}\|\ \|\boldsymbol{\tau}\|. \qquad (5.47)
\end{aligned}$$

Then, using Lemma 5.2.3 and the assumption (**Q**), we obtain

$$\begin{aligned}\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)||z-\Pi u||_{L^4(\Omega)} &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)||z-\Pi u|| \\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)\delta \qquad (5.48)\\ &\le CC_u\frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)\left(\max_{1\le i\le N_h}\frac{h_i^{3/2}}{p_i}\right)\\ &\le CC_uC_Qh^{1/2-\epsilon}, \qquad (5.49)\end{aligned}$$

and using the Lemma 1.2.5 and assumption (**Q**), we find that

$$\begin{aligned}\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|\eta_u\|_{L^4(\Omega)} &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\sum_{i=1}^{N_h}\frac{h_i^6}{p_i^6}\|u\|^4_{H^2(K_i)}\right)^{1/4}\\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\le i\le N_h}\frac{h_i^{3/2}}{p_i^{3/2}}\right)\|u\|_{H^2(\Omega)}\\ &\le CC_uC_Qh. \qquad (5.50)\end{aligned}$$

A use of triangle inequality yields

$$\|\mathbf{q}-\boldsymbol{\theta}\|\le C\left(\|\mathbf{q}-I_h\mathbf{q}\|+\|I_h\mathbf{q}-\boldsymbol{\theta}\|\right)\le C\delta. \qquad (5.51)$$

We substitute (5.49)-(5.51) in (5.47) to obtain

$$\left|\int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})(u-z)\cdot\boldsymbol{\tau}dx\right| \le CC_aC_uC_Qh^{1/2-\epsilon}\delta\|\boldsymbol{\tau}\|. \qquad (5.52)$$

Finally, using the inverse inequality (1.20), the third term on the right-hand side of (5.41) is estimated as

$$\begin{aligned}\left|\int_\Omega(\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx\right| &\le CC_a\sum_{i=1}^{N_h}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^4(K_i)}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(K_i)^2}\|\boldsymbol{\tau}\|_{L^4(K_i)^2}\\ &\le CC_a\sum_{i=1}^{N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^4(K_i)^2}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(K_i)^2}\|\boldsymbol{\tau}\|_{L^2(K_i)^2}\\ &\le CC_a\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|\mathbf{q}-\boldsymbol{\theta}\|_{L^4(\Omega)^2}\|\mathbf{q}-\boldsymbol{\theta}\|\ \|\boldsymbol{\tau}\| \qquad (5.53)\end{aligned}$$

Then, using Lemma 5.2.3, (5.40) and the assumption (**Q**), we obtain

$$\begin{aligned}\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|I_h\mathbf{q}-\boldsymbol{\theta}\|_{L^4(\Omega)^2} &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)||I_h\mathbf{q}-\boldsymbol{\theta}|| \\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)\delta \\ &\le CC_u\frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{p_i}{h_i}\right)\left(\max_{1\le i\le N_h}\frac{h_i^{3/2}}{p_i}\right) \\ &\le CC_uC_Qh^{1/2-\epsilon}, \end{aligned} \tag{5.54}$$

and from Lemma 1.2.5, we arrive at

$$\begin{aligned}\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|\boldsymbol{\eta}_q\|_{L^4(\Omega)} &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\sum_{i=1}^{N_h}\frac{h_i^4}{p_i^4}\|\mathbf{q}\|^4_{H^{3/2}(K_i)^2}\right)^{1/4} \\ &\le C\left(\max_{1\le i\le N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\le i\le N_h}\frac{h_i}{p_i}\right)\|u\|_{H^{5/2}(\Omega)} \\ &\le CC_uC_Qh^{1/2}. \end{aligned} \tag{5.55}$$

We substitute (5.54)-(5.55) in (5.51) to obtain

$$|\int_\Omega(\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx| \le CC_aC_uC_Qh^{1/2-\epsilon}\delta\ \|\boldsymbol{\tau}\|. \tag{5.56}$$

We now combine (5.42), (5.52) and (5.56) to complete the proof of the first inequality of the lemma. A similar argument yields the second inequality. This completes the rest of the proof. ■

LEMMA 5.2.5 *Let the assumption (**Q**) hold. Then, for given $(z,\theta)\in O_\delta(\Pi u, I_h\mathbf{q})$ and $(v,\boldsymbol{\tau})\in V_h\times\mathbf{W}_h$, there exists a positive constant C such that for $0<\epsilon\le 1/4$*

$$|\int_\Omega\tilde{R}_{\mathbf{a}}(u-z,q-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx|\le CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{h_i}{p_i^{1/2}}\right)\ \delta\ \|\boldsymbol{\tau}\|_{L^4(\Omega)^2}$$

and

$$|\int_\Omega\tilde{R}_f(u-z,q-\boldsymbol{\theta})vdx|\le CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\le i\le N_h}\frac{h_i}{p_i^{1/2}}\right)\ \delta\ \|v\|_{L^4(\Omega)}.$$

Proof. For the first inequality, we now consider

$$\int_\Omega \tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx = \int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx \tag{5.57}$$
$$+\int_\Omega \tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(u-z)(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx + \int_\Omega (\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{\mathbf{qq}}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx$$

Using Hölder's inequality, the first term on the right-hand side of (5.57) is estimated as

$$\left|\int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx\right| \leq CC_a\|z-u\|_{L^4(\Omega)}\|z-u\|_{L^2(\Omega)}\ \|\boldsymbol{\tau}\|_{L^4(\Omega)^2}. \tag{5.58}$$

Then we apply the inverse inequality (1.20), Lemma 5.2.3 and the assumption (**Q**) to obtain

$$\begin{aligned}\|z-\Pi u\|_{L^4(\Omega)} &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\|z-\Pi u\| \leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\delta\\ &\leq CC_u\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right)\\ &\leq CC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right),\end{aligned} \tag{5.59}$$

and using Lemma 1.2.5, we arrive at

$$\begin{aligned}\|u-\Pi u\|_{L^4(\Omega)} &\leq C\left(\sum_{i=1}^{N_h}\frac{h_i^3}{p_i^3}\|u\|^2_{H^2(K_i)}\right)^{1/2}\\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i^{3/2}}\right)\|u\|_{H^2(\Omega)}\\ &\leq CC_u\left(\max_{1\leq i\leq}\frac{h_i^{3/2}}{p_i^{3/2}}\right).\end{aligned} \tag{5.60}$$

Note that a use triangle inequality yields

$$\|z-u\| \leq \|z-\Pi u\| + \|\Pi u-u\| \leq 2\delta. \tag{5.61}$$

We substitute (5.59)-(5.61) in (5.58) to obtain

$$\left|\int_\Omega \tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(u-z)^2\cdot\boldsymbol{\tau}dx\right| \leq CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta. \tag{5.62}$$

Using (5.59), (5.60) and (5.51), the second term on the right-hand side of (5.57) is estimated as

$$\begin{aligned} |\int_\Omega \tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(u-z)(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx| &\leq C_a\|u-z\|_{L^4(\Omega)}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(\Omega)^2}\|\boldsymbol{\tau}\|_{L^4(\Omega)^2} \\ &\leq CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta\|\boldsymbol{\tau}\|_{L^4(\Omega)^2}. \quad (5.63) \end{aligned}$$

Finally for third term on the right-hand side of (5.57), we apply Hölder's inequality to find that

$$\int_\Omega(\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{u\mathbf{q}}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx \leq C_a\|\mathbf{q}-\boldsymbol{\theta}\|_{L^4(\Omega)^2}\|\mathbf{q}-\boldsymbol{\theta}\|_{L^2(\Omega)^2}\|\boldsymbol{\tau}\|_{L^4(\Omega)^2}. \quad (5.64)$$

Then, using the inverse inequality (1.20), Lemma 5.2.3 and the assumption ($\mathbf{Q}$), we obtain

$$\begin{aligned} \|I_h\mathbf{q}-\boldsymbol{\theta}\|_{L^4(\Omega)^2} &\leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)||I_h\mathbf{q}-\boldsymbol{\theta}|| \leq C\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\delta \\ &\leq CC_u\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{p_i^{1/2}}{h_i^{1/2}}\right)\left(\max_{1\leq i\leq N_h}\frac{h_i^{3/2}}{p_i}\right) \\ &\leq CC_uC_Q\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right), \quad (5.65) \end{aligned}$$

and from Lemma 1.2.5, we arrive at

$$\begin{aligned} \|\boldsymbol{\eta}_q\|_{L^4(\Omega)} &\leq \left(\sum_{i=1}^{N_h}\frac{h_i^4}{p_i^4}\|\mathbf{q}\|^4_{H^{3/2}(K_i)^2}\right)^{1/4} \\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right)\|u\|_{H^{5/2(\Omega)}} \\ &\leq CC_u\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right). \quad (5.66) \end{aligned}$$

We substitute (5.65)-(5.66) in (5.64). A use of (5.51) yields

$$|\int_\Omega(\mathbf{q}-\boldsymbol{\theta})^T\tilde{\mathbf{a}}_{uu}(z,\boldsymbol{\theta})(\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}dx| \leq CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta\ \|\boldsymbol{\tau}\|_{L^4(\Omega)^2}. \quad (5.67)$$

We now combine (5.58), (5.63) and (5.67) to complete the proof of the first inequality of the lemma. A similar argument yields the second inequality. This completes the rest of the proof. ■

LEMMA 5.2.6 *There exists a positive constant C such that*

$$|A_{\mathbf{q}}(u,\mathbf{q};\boldsymbol{\eta}_q,\boldsymbol{\tau}_h)+A_u(u,\mathbf{q};\eta_u,\boldsymbol{\tau}_h)-B_1(\boldsymbol{\eta}_\sigma,\boldsymbol{\tau}_h)|\leq CC_a\left(\|\boldsymbol{\eta}_q\|+\|\eta_u\|\right)\|\boldsymbol{\tau}_h\|,\quad \boldsymbol{\tau}_h\in\mathbf{W}_h$$

$$|F_{\mathbf{q}}(u,\mathbf{q};\boldsymbol{\eta}_q,v_h)+F_u(u,\mathbf{q};\eta_u,v_h)|\leq CC_a\left(\|\boldsymbol{\eta}_q\|+\|\eta_u\|\right)\|v_h\|,\quad v_h\in V_h$$

and for $\mathbf{w}_h\in\mathbf{W}_h$,

$$|B_1(\boldsymbol{\eta}_q,\mathbf{w}_h)-B_2(\mathbf{w}_h,\eta_u)+J_2(\boldsymbol{\eta}_\sigma,\mathbf{w}_h)|\leq$$
$$C\left(\|\boldsymbol{\eta}_q\|^2+\sum_{e_k\in\Gamma}\int_{e_k}|\{\eta_u\}|^2ds+J_1(\eta_u,\eta_u)+J_2(\boldsymbol{\eta}_\sigma,\boldsymbol{\eta}_\sigma)\right)^{1/2}\left(\|\mathbf{w}_h\|^2+J_2(\mathbf{w}_h,\mathbf{w}_h)\right)^{1/2}.$$

Proof. Since $\boldsymbol{\eta}_\sigma=\Pi\boldsymbol{\sigma}-\boldsymbol{\sigma}$, where $\Pi\boldsymbol{\sigma}$ is the L^2 projection of $\boldsymbol{\sigma}$, an appeal to the Cauchy-Schwarz inequality yields the proof of the first inequality. Similarly, the second inequality follows from Cauchy-Schwarz inequality. For the third inequality, we note that

$$B_1(\boldsymbol{\eta}_q,\mathbf{w}_h)-B_2(\mathbf{w}_h,\eta_u)+J_2(\boldsymbol{\eta}_\sigma,\mathbf{w}_h)=\int_\Omega\boldsymbol{\eta}_q\cdot\mathbf{w}_hdx+\sum_{i=1}^{N_h}\int_{K_i}\eta_u\nabla\cdot\mathbf{w}_hdx$$
$$-\int_{\Gamma_I}\{\eta_u\}[\![\mathbf{w}_h]\!]ds-\int_{\Gamma_I}\mathbf{C}_{12}.[\eta_u][\![\mathbf{w}_h]\!]ds+\int_{\Gamma_I}-C_{22}.[\![\boldsymbol{\eta}_\sigma]\!][\![\mathbf{w}_h]\!]ds.\qquad(5.68)$$

Since $\nabla\cdot\mathbf{w}_h\in V_h$, using the definition of L^2-projection, the second term on the right-hand side of (5.68) becomes zero. The third term on the right-hand side of (5.68) is estimated as

$$|\int_{\Gamma_I}\{\eta_u\}[\![\mathbf{w}_h]\!]ds|\ \leq\ C\left(\sum_{e_k\in\Gamma}\int_{e_k}|\{\eta_u\}|^2ds\right)^{1/2}J_2(\mathbf{w}_h,\mathbf{w}_h)^{1/2}.\qquad(5.69)$$

For the first term on the right-hand side of (5.68), a use of Cauchy-Schwarz inequality yields

$$|\int_\Omega\boldsymbol{\eta}_q\cdot\mathbf{w}_hdx|\leq\|\boldsymbol{\eta}_q\|\ \|\mathbf{w}_h\|.\qquad(5.70)$$

We bound the fourth term on the right-hand side of (5.68) as

$$|\int_{\Gamma_I}\mathbf{C}_{12}.[\![\eta_u]\!][\![\mathbf{w}_h]\!]ds|\ \leq\ CJ_1(\eta_u,\eta_u)^{1/2}\ J_2(\mathbf{w}_h,\mathbf{w}_h)^{1/2}.\qquad(5.71)$$

Finally, the last term on the right-hand side of (5.68) is estimated as

$$|\int_{\Gamma_I} C_{22}.[\![\boldsymbol{\eta}_\sigma]\!][\![\mathbf{w}_h]\!]ds| \;\leq\; CJ_2(\boldsymbol{\eta}_\sigma,\boldsymbol{\eta}_\sigma)^{1/2}\; J_2(\mathbf{w}_h,\mathbf{w}_h)^{1/2}. \tag{5.72}$$

We combine (5.68)-(5.72) to complete the rest of the proof. ■

THEOREM 5.2.1 *There is a positive constant C such that*

$$\|\boldsymbol{\xi}_\sigma\| \leq CC_a\left(\|\mathbf{e}_q\| + \|e_u\| + h^{1/2-\epsilon}\delta\right).$$

Proof. Using (5.34), we write

$$A_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\boldsymbol{\tau}_h) + A_u(u,\mathbf{q};e_u,\boldsymbol{\tau}_h) - B_1(\mathbf{e}_\sigma,\boldsymbol{\tau}_h) = \int_\Omega \tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}_h dx. \tag{5.73}$$

Set $\boldsymbol{\tau}_h = \boldsymbol{\xi}_\sigma$ in (5.73) to obtain

$$\int_\Omega \boldsymbol{\xi}_\sigma\cdot\boldsymbol{\xi}_\sigma dx = A_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\boldsymbol{\xi}_\sigma) + A_u(u,\mathbf{q};e_u,\boldsymbol{\xi}_\sigma) + B_1(\boldsymbol{\eta}_\sigma,\boldsymbol{\tau}_h) - \int_\Omega \tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta})\cdot\boldsymbol{\tau}_h dx.$$

Using Cauchy-Schwartz inequality and Lemma 5.2.4, we complete the proof. ■

THEOREM 5.2.2 *The following error estimate holds for $0<h<h_0<1$:*

$$\begin{aligned}\|e_u\| \;\leq\; & C_1h^{1/2}\left(J_1(e_u,e_u)^{1/2} + J_2(\mathbf{e}_\sigma,\mathbf{e}_\sigma)^{1/2}\right) + C_2C_a\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right)\|\mathbf{e}_q\| \\ & + C_3C_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i^{1/2}}{p_i^{1/2}}\right)\delta.\end{aligned}$$

Proof. We now appeal to the duality argument. Consider the following auxiliary problem :

$$\begin{aligned}-\nabla\cdot(\mathbf{a}_{\mathbf{q}}(u,\mathbf{q})\nabla\phi + f_{\mathbf{q}}(u,\mathbf{q})\phi) + \mathbf{a}_u(u,\mathbf{q})\cdot\nabla\phi + f_u(u,\mathbf{q})\phi &= e_u \quad \text{in } \Omega,\\ \phi &= 0 \quad \text{on } \partial\Omega,\end{aligned}$$

which satisfies the elliptic regularity

$$\|\phi\|_{H^2(\Omega)} \leq C\|e_u\|. \tag{5.74}$$

In order to write the mixed weak formulation, we introduce $\mathbf{p}$ and $\boldsymbol{\psi}$ such that

$$\mathbf{p} = \nabla\phi \quad \text{in } \Omega, \tag{5.75}$$

$$-\boldsymbol{\psi} = \mathbf{a}_{\mathbf{q}}(u,\mathbf{q})\mathbf{p} + f_{\mathbf{q}}(u,\mathbf{q})\phi \quad \text{in } \Omega, \tag{5.76}$$

$$\nabla\cdot\boldsymbol{\psi} + \mathbf{a}_u(u,\mathbf{q})\cdot\mathbf{p} + f_u(u,\mathbf{q})\phi = e_u \quad \text{in } \Omega. \tag{5.77}$$

We multiply (5.77) by e_u, (5.76) by $\mathbf{e}_q$ and (5.75) by $\mathbf{e}_\sigma$, then integrate to arrive at

$$\begin{aligned}\|e_u\|^2 &= \int_\Omega e_u \nabla\cdot\boldsymbol{\psi} dx + \int_\Omega \mathbf{a}_u(u,\mathbf{q}) e_u\cdot\mathbf{p} dx + \int_\Omega f_u(u,\mathbf{q}) e_u\phi + \int_\Omega \boldsymbol{\psi}\cdot\mathbf{e}_q dx \\ &\quad + \int_\Omega \mathbf{a}_{\mathbf{q}}(u,\mathbf{q})\mathbf{p}\cdot\mathbf{e}_q dx + \int_\Omega f_{\mathbf{q}}(u,\mathbf{q})\phi\cdot\mathbf{e}_q dx - \int_\Omega \mathbf{p}\cdot\mathbf{e}_\sigma dx + \int_\Omega \nabla\phi\cdot\mathbf{e}_\sigma dx.\end{aligned}$$

Since $[\![\phi]\!] = 0$, $[\![\boldsymbol{\psi}]\!] = 0$ on $e_k \in \Gamma_I$ and $\phi = 0$ on $\partial\Omega$, we write

$$\begin{aligned}\|e_y\|^2 &= B_1(\mathbf{e}_q,\boldsymbol{\psi}) - B_2(\boldsymbol{\psi},e_u) + J_2(\mathbf{e}_\sigma,\boldsymbol{\psi}) + A_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\mathbf{p}) + A_u(u,\mathbf{q};e_u,\mathbf{p}) \\ &\quad -B_1(\mathbf{e}_\sigma,\mathbf{p}) + B_2(\mathbf{e}_\sigma,\phi) + J_1(e_u,\phi) + F_u(u,\mathbf{q};e_u,\phi) + F_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\phi).\end{aligned}$$

Then, we use (5.33)-(5.35) to obtain

$$\begin{aligned}\|e_u\|^2 &= B_1(\mathbf{e}_q,\boldsymbol{\eta}_\psi) - B_2(\boldsymbol{\eta}_\psi,e_u) + J_2(\mathbf{e}_\sigma,\boldsymbol{\eta}_\psi) + A_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\boldsymbol{\eta}_p) + A_u(u,\mathbf{q};e_u,\boldsymbol{\eta}_p) \\ &\quad -B_1(\mathbf{e}_\sigma,\boldsymbol{\eta}_p) + B_2(\mathbf{e}_\sigma,\eta_\phi) + J_1(e_u,\eta_\phi) + F_u(u,\mathbf{q};e_u,\eta_\phi) + F_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\eta_\phi) \\ &\quad - \left(\tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta}), I_h\mathbf{p}\right) - \left(\tilde{R}_f(u-z,\mathbf{q}-\boldsymbol{\theta}), I_h\phi\right),\end{aligned}$$

where $\eta_\phi = \phi - I_h\phi$, $\boldsymbol{\eta}_p = \mathbf{p} - I_h\mathbf{p}$ and $\boldsymbol{\eta}_\psi = \boldsymbol{\psi} - \Pi\boldsymbol{\psi}$. We now apply integration by parts for the following term to find that

$$\begin{aligned}-B_2(\boldsymbol{\eta}_\psi,e_u) &= \sum_{i=1}^{N_h}\int_{K_i} e_u\nabla\cdot\boldsymbol{\eta}_\psi dx - \int_{\Gamma_I}(\{e_u\} + \mathbf{C}_{12}[\![e_u]\!])[\![\boldsymbol{\eta}_\psi]\!] ds \\ &= -\sum_{i=1}^{N_h}\int_{K_i}\nabla e_u\cdot\boldsymbol{\eta}_\psi dx + \int_{\Gamma_I}(1-\mathbf{C}_{12})[\![e_u]\!]\{\boldsymbol{\eta}_\psi\} ds \\ &= \sum_{i=1}^{N_h}\int_{K_i}\nabla(u-I_hu)\cdot\boldsymbol{\eta}_\psi dx - \sum_{i=1}^{N_h}\int_{K_i}\nabla(I_hu-u_l)\cdot\boldsymbol{\eta}_\psi dx \\ &\quad + \int_\Gamma (1-\mathbf{C}_{12})[e_u]\{\boldsymbol{\eta}_\psi\} ds. \qquad (5.78)\end{aligned}$$

Since $\Pi\boldsymbol{\psi}$ is the L^2 projection of $\boldsymbol{\psi}$, the second term on the right hand side of (5.78) becomes zero. Then, for the first term on the right hand side of (5.78), we use Lemma 1.2.5 to obtain

$$\begin{aligned}|\sum_{i=1}^{N_h}\int_{K_i}\nabla(u-I_hu)\cdot\boldsymbol{\eta}_\psi dx| &\le C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\nabla(u-I_hu)\|^2_{L^2(K_i)}\right)^{1/2}\|\boldsymbol{\psi}\|_{H^1(\Omega)^2} \\ &\le C\left(\max_{1\le i\le N_h}\frac{h_i}{p_i}\right)|u-I_hu|_{1,h}\|\boldsymbol{\psi}\|_{H^1(\Omega)^2}. \qquad (5.79)\end{aligned}$$

Similarly, using Lemma 1.2.5, we estimate the third term as

$$\begin{aligned} |\int_\Gamma [\![e_u]\!]\{\boldsymbol{\eta}_\psi\}ds| &\leq C\sum_{e_k\in\Gamma}\left(\int_{e_k} C_{11}[\![e_u]\!]^2 ds\right)^{1/2}\left(\int_{e_k}\{\boldsymbol{\eta}_\psi\}^2 ds\right)^{1/2} \\ &\leq Ch^{1/2}J_1(e_u,e_u)^{1/2}\ \|\boldsymbol{\psi}\|_{H^1(\Omega)^2}. \end{aligned} \tag{5.80}$$

Using Lemma 1.2.5, we find that

$$|B_1(\mathbf{e}_q,\boldsymbol{\eta}_\psi)| = |\int_\Omega \mathbf{e}_q\cdot\boldsymbol{\eta}_\psi dx| \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\mathbf{e}_q\|^2_{L^2(K_i)^2}\right)^{1/2}\|\mathbf{p}\|_{H^1(\Omega)^2}. \tag{5.81}$$

We first note that

$$B_2(\mathbf{e}_\sigma,\eta_\phi) = \sum_{i=1}^{N_h}\int_{K_i}\mathbf{e}_\sigma\cdot\nabla\eta_\phi dx - \int_\Gamma(\{\mathbf{e}_\sigma\}-C_{12}[\![\mathbf{e}_\sigma]\!])[\![\eta_\phi]\!]ds. \tag{5.82}$$

Then, we use Lemma 1.2.1 to estimate the following term

$$|\sum_{i=1}^{N_h}\int_{K_i}\mathbf{e}_\sigma\cdot\nabla\eta_\phi dx| \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\mathbf{e}_\sigma\|^2_{L^2(K_i)^2}\right)^{1/2}\|\phi\|_{H^2(\Omega)}, \tag{5.83}$$

and use Lemma 1.2.1 with trace inequality (1.19) to find that

$$\begin{aligned} |\int_\Gamma(\{\mathbf{e}_\sigma\}-C_{12}[\![\mathbf{e}_\sigma]\!])[\![\eta_\phi]\!]ds| &\leq C\sum_{e_k\in\Gamma}\left(\int_{e_k}\{|\boldsymbol{\xi}_\sigma|\}|[\![\eta_\phi]\!]|ds+\int_{e_k}\{|\boldsymbol{\eta}_\sigma|\}|[\![\eta_\phi]\!]|ds\right) \\ &\leq C\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\xi}_\sigma|\}^2ds+\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\eta}_\sigma|\}^2ds\right)^{1/2}\|\phi\|_{H^2(\Omega)} \\ &\leq C\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\|\boldsymbol{\xi}_\sigma\|\ \|\phi\|_{H^2(\Omega)}+\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k^3}{p_k^3}\{|\boldsymbol{\eta}_\sigma|\}^2ds\right)^{1/2}\|\phi\|_{H^2(\Omega)} \end{aligned} \tag{5.84}$$

Now, using Cauchy-Schwarz inequality and Lemma 1.2.1, we obtain

$$|B_1(\mathbf{e}_\sigma,\boldsymbol{\eta}_p)| \leq C\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\mathbf{e}_\sigma\|^2_{L^2(K_i)^2}\right)^{1/2}\|\mathbf{p}\|_{H^1(\Omega)^2}, \tag{5.85}$$

$$|A_u(u,\mathbf{q};e_u,\boldsymbol{\eta}_p)| \leq CC_a\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|e_u\|^2_{L^2(K_i)}\right)^{1/2}\|\mathbf{p}\|_{H^1(\Omega)^2}, \tag{5.86}$$

$$|A_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\boldsymbol{\eta}_p)| \leq CC_a\left(\sum_{i=1}^{N_h}\frac{h_i^2}{p_i^2}\|\mathbf{e}_q\|^2_{L^2(K_i)}\right)^{1/2}\|\mathbf{p}\|_{H^1(\Omega)^2}, \tag{5.87}$$

$$|F_u(u,\mathbf{q};e_u,\eta_\phi)| \leq CC_a\left(\sum_{i=1}^{N_h}\frac{h_i^4}{p_i^4}\|e_u\|^2_{L^2(K_i)}\right)^{1/2}\|\phi\|_{H^2(\Omega)^2}, \tag{5.88}$$

and

$$|F_{\mathbf{q}}(u,\mathbf{q};\mathbf{e}_q,\eta_\phi)| \leq CC_a \left(\sum_{i=1}^{N_h} \frac{h_i^4}{p_i^4}\|\mathbf{e}_q\|^2_{L^2(K_i)}\right)^{1/2} \|\phi\|_{H^2(\Omega)^2}. \tag{5.89}$$

Using Lemma 1.2.5, we obtain

$$\begin{aligned} |J_2(\mathbf{e}_\sigma,\boldsymbol{\eta}_\psi)| &\leq CJ_2(\mathbf{e}_\sigma,\mathbf{e}_\sigma)^{1/2} \left(\int_{\Gamma_I} [\![\boldsymbol{\eta}_\psi]\!]^2 ds\right)^{1/2} \\ &\leq Ch^{1/2}J_2(\mathbf{e}_\sigma,\mathbf{e}_\sigma)^{1/2} \|\boldsymbol{\psi}\|_{H^1(\Omega)^2}, \end{aligned} \tag{5.90}$$

and using Lemma 1.2.1, we find that

$$\begin{aligned} |J_1(e_u,\eta_\phi)| &\leq C\sum_{e_k\in\Gamma} \left(\int_{e_k} C_{11}[\![e_u]\!]^2 ds\right)^{1/2} \left(\int_{e_k} C_{11}[\![\eta_\phi]\!]^2 ds\right)^{1/2} \\ &\leq C\left(\max_{1\leq i\leq N_h} \frac{h_i^{3/2}}{p_i^{3/2}}\right) J_1(\eta_u,\eta_u)^{1/2}\|\phi\|_{H^2(\Omega)}. \end{aligned} \tag{5.91}$$

Finally, from Lemma 5.2.5, $\|I_h\phi\|_{L^4(\Omega)^2} \leq C\|\phi\|_{H^1(\Omega)}$ and $\|I_h\mathbf{p}\|_{L^4(\Omega)^2} \leq C\|\mathbf{p}\|_{H^1(\Omega)^2}$, we find that

$$|\left(\tilde{R}_{\mathbf{a}}(u-z,\mathbf{q}-\boldsymbol{\theta}), I_h\mathbf{p}\right)| \leq CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta\|\mathbf{p}\|_{H^1(\Omega)^2}. \tag{5.92}$$

$$|\left(\tilde{R}_f(u-z,\mathbf{q}-\boldsymbol{\theta}), I_h\phi\right)| \leq CC_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta\|\phi\|_{H^1(\Omega)}. \tag{5.93}$$

We combine the estimates (5.80)-(5.93) and then use elliptic regularity (5.74) to obtain for small h

$$\begin{aligned} \|e_u\| \leq\ & C_1h^{1/2}\left(J_1(e_u,e_u)^{1/2}+J_2(\mathbf{e}_\sigma,\mathbf{e}_\sigma)^{1/2}\right) + C_2\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right)(C_a\|\mathbf{e}_q\|+|u-I_hu|_{1,h}) \\ &+C_4\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\left(\sum_{e_k\in\Gamma}\int_{e_k}\frac{h_k}{p_k^2}\{|\boldsymbol{\eta}_\sigma|\}^2 ds + \|\boldsymbol{\xi}_\sigma\|^2\right)^{1/2} \\ &+C_3C_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i^{1/2}}\right)\delta. \end{aligned}$$

Then, from Theorem 5.2.1 and the estimate (5.39), we arrive at

$$\begin{aligned} \|e_u\| \leq\ & C_1h^{1/2}\left(J_1(e_u,e_u)^{1/2}+J_2(\mathbf{e}_\sigma,\mathbf{e}_\sigma)^{1/2}\right) + C_2C_a\left(\max_{1\leq i\leq N_h}\frac{h_i}{p_i}\right)\|\mathbf{e}_q\| \\ &+C_3C_aC_uC_Q\frac{1}{h^\epsilon}\left(\max_{1\leq i\leq N_h}\frac{h_i^{1/2}}{p_i^{1/2}}\right)\delta. \end{aligned}$$

This completes the rest of the proof. ■

THEOREM 5.2.3 *There exists constants $C_1 > 0$ and $C_2 > 0$ such that for any $0 < \epsilon \leq 1/4$,*

$$\|\mathbf{e}_q\|^2 + J_1(e_u, e_u) + J_2(\mathbf{e}_\sigma, \mathbf{e}_\sigma) \;\leq\; C_1 C_a C_u C_Q h^\epsilon \delta_2 + C_2 C_a \|e_u\|^2. \tag{5.94}$$

Proof. Set $v_h = \xi_u$, $\boldsymbol{\tau}_h = \boldsymbol{\xi}_q$ and $\mathbf{w}_h = \boldsymbol{\xi}_\sigma$ in (5.36) -(5.38), respectively. Using Lemma 5.2.1, we obtain

$$\begin{aligned}
C_1 \left(\|\boldsymbol{\xi}_q\|^2 + J_1(\xi_u, \xi_u) + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma) \right) - C_2 \|\xi_u\|^2 &\leq J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma) + A_u(u, \mathbf{q}; \boldsymbol{\xi}_q, \boldsymbol{\xi}_q) \\
&+ A_{\mathbf{q}}(u, \mathbf{q}; \xi_u, \boldsymbol{\xi}_q) + F_u(u, \mathbf{q}; \xi_u, \boldsymbol{\xi}_q) + F_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\xi}_q, \xi_u) + J_1(\xi_u, \xi_u) \\
&= B_1(\boldsymbol{\eta}_q, \boldsymbol{\xi}_\sigma) - B_2(\boldsymbol{\xi}_\sigma, \eta_u) + J_2(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_\sigma) + A_u(u, \mathbf{q}; \boldsymbol{\eta}_q, \boldsymbol{\xi}_q) + A_{\mathbf{q}}(u, \mathbf{q}; \eta_u, \boldsymbol{\xi}_q) \\
&- B_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_q) + B_2(\boldsymbol{\eta}_\sigma, \xi_u) + F_u(u, \mathbf{q}; \eta_u, \xi_u) + F_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\eta}_q, \xi_u) \\
&+ J_1(\eta_u, \xi_u) + \left(\tilde{R}_{\mathbf{a}}(u - z, \mathbf{q} - \boldsymbol{\theta}), \boldsymbol{\xi}_q \right) + \left(\tilde{R}_f(u - z, \mathbf{q} - \boldsymbol{\theta}), \xi_u \right).
\end{aligned} \tag{5.95}$$

From the definition of B_2 and J_1, we write

$$B_2(\boldsymbol{\eta}_\sigma, \xi_u) + J_1(\eta_u, \xi_u) = \sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\eta}_\sigma \cdot \nabla \xi_u dx - \int_\Gamma (\{\boldsymbol{\eta}_\sigma\} - \beta[\eta_u] - C_{12}[\![\boldsymbol{\eta}_\sigma]\!])[\![\xi_u]\!] ds. \tag{5.96}$$

Since $\Pi\boldsymbol{\sigma}$ is L^2 projections of $\boldsymbol{\sigma}$ onto $\mathbf{W}_h$ and $\nabla \xi_u \in \mathbf{W}_h$, we note that

$$\sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\eta}_\sigma \cdot \nabla \xi_u dx = 0. \tag{5.97}$$

Next using the trace inequality (1.19), we bound the following term as:

$$\begin{aligned}
\left| \int_\Gamma (\{\boldsymbol{\eta}_\sigma\} - C_{11}[\eta_u] - C_{12}[\![\boldsymbol{\eta}_\sigma]\!])[\![\xi_u]\!] ds \right| \;\leq\; & C \left(\sum_{e_k \in \Gamma} \int_{e_k} \{|\boldsymbol{\eta}_\sigma|\}^2 ds \right)^{1/2} (J_1(\xi_u, \xi_u))^{1/2} \\
& + C \;\; (J_1(\eta_u, \eta_u))^{1/2} \;\; (J_1(\xi_u, \xi_u))^{1/2}.
\end{aligned}$$

An appeal to Lemma 5.2.6 with $\boldsymbol{\tau}_h = \boldsymbol{\xi}_q$, $v_h = \xi_u$ and $\mathbf{w}_h = \boldsymbol{\xi}_\sigma$ yields

$$|A_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\eta}_q, \boldsymbol{\xi}_q) + A_u(u, \mathbf{q}; \eta_u, \boldsymbol{\xi}_q) - B_1(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_q)| \leq C C_a (\|\boldsymbol{\eta}_q\| + \|\eta_u\|) \|\boldsymbol{\xi}_q\|, \tag{5.98}$$

and

$$|F_{\mathbf{q}}(u, \mathbf{q}; \boldsymbol{\eta}_q, \xi_u) + F_u(u, \mathbf{q}; \eta_u, \xi_u)| \leq C C_a (\|\boldsymbol{\eta}_q\| + \|\eta_u\|) \|\xi_u\|. \tag{5.99}$$

Using Lemma 5.2.6, we now arrive at

$$|B_1(\boldsymbol{\eta}_q, \boldsymbol{\xi}_\sigma) - B_2(\boldsymbol{\xi}_\sigma, \eta_u) + J_2(\boldsymbol{\eta}_\sigma, \boldsymbol{\xi}_\sigma)| \leq$$
$$C\left(\|\boldsymbol{\eta}_q\|^2 + \sum_{e_k \in \Gamma} \int_{e_k} |\{\eta_u\}|^2 ds + J_1(\eta_u, \eta_u) + J_2(\boldsymbol{\eta}_\sigma, \boldsymbol{\eta}_\sigma)\right)^{1/2} \left(\|\boldsymbol{\xi}_\sigma\|^2 + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma)\right)^{1/2}.$$

For the last two terms on the right-hand side of (5.95), we set $\boldsymbol{\tau} = \boldsymbol{\xi}_q$ and $v = \xi_u$ in Lemma 5.2.4 to obtain

$$|\left(\tilde{R}_{\mathbf{q}}(u - z, \mathbf{q} - \boldsymbol{\theta}), \boldsymbol{\xi}_q\right)| \leq CC_aC_uC_Qh^{1/2-\epsilon}\delta \ \|\boldsymbol{\xi}_q\|,$$

and

$$|\left(\tilde{R}_u(u - z, \mathbf{q} - \boldsymbol{\theta}), \xi_u\right)| \leq CC_aC_uC_Qh^{1/2-\epsilon}\delta \ \|\xi_u\|. \tag{5.100}$$

We now combine the estimates (5.95)-(5.100) with the estimate of $\|\boldsymbol{\xi}\|$ from Theorem 5.2.1 to obtain for sufficiently small h

$$\begin{aligned}
\|\boldsymbol{\xi}_q\|^2 + J_1(\xi_u, \xi_u) + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma) &\leq C_1C_a\left(J_1(\eta_u, \eta_u) + J_2(\boldsymbol{\eta}_\sigma, \boldsymbol{\eta}_\sigma) + \|\eta_u\|^2 + \|\boldsymbol{\eta}_q\|^2\right) \\
&\quad + C_2C_aC_uC_Qh^{1-2\epsilon}\delta^2 + C_3C_a\|e_u\|^2 \\
&\quad + C_4\sum_{e_k \in \Gamma} \int_{e_k} \left(\{|\boldsymbol{\eta}_\sigma|\}^2 + \{|\eta_u|\}^2\right) ds \\
&\leq C_1C_aC_uC_Q\left(h^{2\epsilon} + h^{1-2\epsilon}\right)\delta^2 + C_2C_a\|e_u\|^2. \qquad (5.101)
\end{aligned}$$

This completes the rest of the proof. ■

Now in the following theorem, we examine the range of S_h, for $0 < h < h_0 < 1$.

THEOREM 5.2.4 *For all $0 < h < h_0 < 1$, the map S_h maps $O_\delta(\Pi u, I_h\mathbf{q})$ into itself, where δ is as in (5.40).*

Proof. Using Theorem 5.2.3 and Theorem 5.2.2, we obtain

$$\begin{aligned}
\|\boldsymbol{\xi}_q\| + J_1(\xi_u, \xi_u)^{1/2} + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma)^{1/2} &\leq C\left(C_1, C_2, C_a, C_u, C_Q\right)\left(h^\epsilon\delta + \|e_u\|\right) \\
&\leq C\left(C_1, C_a, C_u, C_Q\right)\left(h^\epsilon\delta + \left(\max_{1 \leq i \leq N_h} \frac{h_i}{p_i}\right)\|\boldsymbol{\xi}_q\|\right) \\
&\quad + C_3h^{1/2}\left(J_1(\xi_u, \xi_u)^{1/2} + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma)^{1/2}\right).
\end{aligned}$$

For sufficiently small h, we find that

$$\|\boldsymbol{\xi}_q\| + J_1(\xi_u, \xi_u)^{1/2} + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma)^{1/2} \leq C\left(C_a, C_u, C_Q\right) h^\epsilon \delta. \tag{5.102}$$

Using Theorem 5.2.2 and Theorem 5.2.3, we arrive at

$$\begin{aligned}
\|\xi_u\| &\leq C\left(C_a, C_u, C_Q\right) \left(\left(\max_{1\leq i\leq N_h} \frac{h_i}{p_i^{1/2}} \right) \|\boldsymbol{\xi}_q\| + h^{1/2} \left(J_1(\xi_u, \xi_u)^{1/2} + J_2(\boldsymbol{\xi}_\sigma, \boldsymbol{\xi}_\sigma)^{1/2} + h^\epsilon \delta \right) \right) \\
&\leq C\left(C_a, C_u, C_Q\right) h^\epsilon \left(\max_{1\leq i\leq N_h} \frac{h_i}{p_i^{1/2}} \right) \delta_2 + C\left(C_a, C_u, C_Q\right) h^{1/2+\epsilon} \delta \\
&\leq C\left(C_a, C_u, C_Q\right) h^{1/2+\epsilon} \delta.
\end{aligned} \tag{5.103}$$

Hence, $\|S_h((z,\boldsymbol{\theta})) - (\Pi u, I_h\mathbf{q})\|_+ = ||\xi_u|| + \|\boldsymbol{\xi}_q\| \leq \delta$ and we conclude that S_h maps $\mathcal{O}_\delta(\Pi u, I_h\mathbf{q})$ into itself. This completes the proof. ■

THEOREM 5.2.5 *Let $(z_1, \boldsymbol{\theta}_1)$ and $(z_2, \boldsymbol{\theta}_2) \in O_\delta(\Pi u, I_h\mathbf{q})$ with δ as in (5.40). Then for sufficiently small h and $0 < \epsilon \leq 1/4$, there exists a positive constant C such that*

$$||S_h((z_1, \boldsymbol{\theta}_1)) - S_h((z_2, \boldsymbol{\theta}_2))||_+ \leq C\left(C_a, C_u, C_Q\right) h^\epsilon ||(z_1, \boldsymbol{\theta}_1) - (z_2, \boldsymbol{\theta}_2)||_+. \tag{5.104}$$

Proof. Let $\{u_i,\ \mathbf{q}_i, \boldsymbol{\sigma}_i\},\ i = 1,\ 2$ be two distinct solution of (5.33)-(5.35) when $(z, \boldsymbol{\theta})$ is replaced by $(z_i, \boldsymbol{\theta}_i)$, for $i = 1,\ 2$. From Theorem 5.2.3 and Theorem 5.2.1, it follows that

$$\left(\ ||u_i - \Pi u|| + \|\mathbf{q}_i - I_h\mathbf{q}\| + \|\boldsymbol{\sigma}_i - \Pi\boldsymbol{\sigma}\| \ \right) \leq Ch^\epsilon\delta.$$

Using (5.33)-(5.35), we note that for any $(\mathbf{w}_h, v_h, \boldsymbol{\tau}_h) \in \mathbf{W}_h \times V_h \times \mathbf{W}_h$

$$B_1(\mathbf{q}_1 - \mathbf{q}_2, \mathbf{w}_h) - B_2(\mathbf{w}_h, y_1 - y_2) = 0, \tag{5.105}$$

$$\begin{aligned}
A_{\mathbf{q}}(u, \mathbf{q}; \mathbf{q}_1 - \mathbf{q}_2, \boldsymbol{\tau}_h) + A_u(u, \mathbf{q}; u_1 - u_2, \boldsymbol{\tau}_h) - B_1(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2, \boldsymbol{\tau}_h) = \\
\int_\Omega \left(\tilde{R}_{\mathbf{a}}(u - z_1, \mathbf{q} - \boldsymbol{\theta}_1) - \tilde{R}_{\mathbf{a}}(u - z_2, \mathbf{q} - \boldsymbol{\theta}_2) \right) \cdot \boldsymbol{\tau}_h dx
\end{aligned} \tag{5.106}$$

and

$$\begin{aligned}
B_2(\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2, v_h) + J(u_1 - u_2, v_h) + F_{\mathbf{q}}(u, \mathbf{q}; \mathbf{q}_1 - \mathbf{q}_2, v_h) - F_u(u, \mathbf{q}; u_1 - u_2, v_h) \\
= \int_\Omega \left(\tilde{R}_f(u - z_1, \mathbf{q} - \boldsymbol{\theta}_1) - \tilde{R}_f(u - z_2, \mathbf{q} - \boldsymbol{\theta}_2) \right) v_h dx.
\end{aligned} \tag{5.107}$$

We rewrite the following term from the right hand side of (5.106) as

$$\begin{aligned}
&\tilde{R}_{\mathbf{a}}(u-z_1,\mathbf{q}-\boldsymbol{\theta}_1)-\tilde{R}_{\mathbf{a}}(u-z_2,\mathbf{q}-\boldsymbol{\theta}_2)\\
&=\mathbf{a}(z_1,\boldsymbol{\theta}_1)-\mathbf{a}(z_2,\boldsymbol{\theta}_2)+\mathbf{a}_u(u,\mathbf{q})(z_2-z_1)+\mathbf{a}_{\mathbf{q}}(u,\mathbf{q})(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)\\
&=\mathbf{a}(z_1,\boldsymbol{\theta}_1)-\mathbf{a}(z_2,\boldsymbol{\theta}_2)+\mathbf{a}_u(z_2,\boldsymbol{\theta}_2)(z_2-z_1)+\mathbf{a}_{\mathbf{q}}(z_2,\boldsymbol{\theta}_2)(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)-\mathbf{a}_u(z_2,\boldsymbol{\theta}_2)(z_2-z_1)\\
&\quad-\mathbf{a}_{\mathbf{q}}(z_2,\boldsymbol{\theta}_2)(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)+\mathbf{a}_u(u,\mathbf{q})(z_2-z_1)+\mathbf{a}_{\mathbf{q}}(u,\mathbf{q})(\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)\\
&=\tilde{R}_{\mathbf{a}}(z_2-z_1,\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)++\left[\mathbf{a}_u(u,\mathbf{q})-\mathbf{a}_u(z_2,\boldsymbol{\theta}_2)\right](z_2-z_1)\\
&\quad+\left[\mathbf{a}_{\mathbf{q}}(u,\mathbf{q})-\mathbf{a}_{\mathbf{q}}(z_2,\boldsymbol{\theta}_2)\right](\boldsymbol{\theta}_2-\boldsymbol{\theta}_1).
\end{aligned}$$

Using Taylor series expansion (4.6), we obtain

$$\begin{aligned}
&\tilde{R}_{\mathbf{a}}(u-z_1,\mathbf{q}-\boldsymbol{\theta}_1)-\tilde{R}_{\mathbf{a}}(u-z_2,\mathbf{q}-\boldsymbol{\theta}_2)\\
&=\tilde{R}_{\mathbf{a}}(z_2-z_1,\boldsymbol{\theta}_2-\boldsymbol{\theta}_1)+\left[-\tilde{\mathbf{a}}_{uu}(z_2,\boldsymbol{\theta}_2)(u-z_2)-\tilde{\mathbf{a}}_{u\mathbf{q}}(z_2,\boldsymbol{\theta}_2)(\mathbf{q}-\boldsymbol{\theta}_2)\right](z_2-z_1)\\
&\quad+\left[-\tilde{\mathbf{a}}_{\mathbf{q}u}(z_2,\boldsymbol{\theta}_2)(u-z_2)-\tilde{\mathbf{a}}_{\mathbf{qq}}(z_2,\boldsymbol{\theta}_2)(\mathbf{q}-\boldsymbol{\theta}_2)\right](\boldsymbol{\theta}_2-\boldsymbol{\theta}_1).
\end{aligned}$$

We can write similar expressions for $\tilde{R}_f(u-z_1,\mathbf{q}-\boldsymbol{\theta}_1)-\tilde{R}_f(u-z_2,\mathbf{q}-\boldsymbol{\theta}_2)$. Now using similar arguments as in Theorem 5.2.4, we first obtain

$$\begin{aligned}
&\left(\|\mathbf{q}_1-\mathbf{q}_2\|^2+\sum_{e_k\in\Gamma}\int_{e_k}C_{11}[\![u_1-u_2]\!]^2ds\right)^{1/2}\\
&\qquad\leq C_1C_a\left(C_uC_Qh^{\epsilon}||z_1-z_2||+\|u_1-u_2\|\right)
\end{aligned}\tag{5.108}$$

and

$$\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\| \;\leq\; C_1C_ah^{\epsilon}||z_1-z_2||+C_2C_a\|u_1-u_2\|. \tag{5.109}$$

Then, an application of duality argument as in Theorem 5.2.2 yields

$$\|u_1-u_2\|\leq CC_ah^{\epsilon}\left(\|\mathbf{q}_1-\mathbf{q}_2\|^2+\sum_{e_k\in\Gamma}\int_{e_k}C_{11}[\![u_1-u_2]\!]^2ds\right)^{1/2}. \tag{5.110}$$

We combine the estimates (5.108)-(5.110) to complete the rest of the proof. ■

Now, we can conclude from Theorem 5.2.5 that the map S_h is well defined, that is, the linearized problem (5.33)-(5.34) is well-posed and continuous in the ball $O_\delta(\Pi u)$. Hence, an appeal to Brouwer fixed point theorem implies that S_h has a fixed point u_h in $O_\delta(\Pi u)$. Then, using Theorem 5.2.5, it is easy to see that u_h is the unique fixed point in $O_\delta(\Pi u)$ for small h. Moreover, $(u_h,\mathbf{q}_h,\boldsymbol{\sigma}_h)$ is the unique solution for the problem (5.30)-(5.32).

5.2.2 A priori error estimates

Note that by replacing $\Pi u - u_l$ by $\Pi u - u_h$ in (5.103) and Theorem 5.2.2, $\Pi u - u_h$ satisfies the estimate (5.103). Further replacing $I_h\mathbf{q} - \mathbf{q}_l$ by $I_h\mathbf{q} - \mathbf{q}_h$, $I_h\mathbf{q} - \mathbf{q}_h$ satisfies the estimate (5.102). Hence, we easily prove the following theorem.

THEOREM 5.2.6 *There exists a constant C such that for sufficiently small h the following estimates hold:*

$$\|\mathbf{q} - \mathbf{q}_h\|^2 \leq CC_uC_Q\sum_{i=1}^{N_h}\left(\frac{h_i^{2\mu_i^+-1}}{p_i^{2s_i-2}}\|\mathbf{q}\|^2_{H^{s_i}(K_i)^2} + \frac{h_i^{2\mu_i^*+1}}{p_i^{2s_i}}\|u\|^2_{H^{s_i+1}(K_i)}\right)$$

and

$$\|u - u_h||^2 \leq CC_uC_Qh^{1/2}\sum_{i=1}^{N_h}\left(\frac{h_i^{2\mu_i^+-1}}{p_i^{2s_i-2}}\|\mathbf{q}\|^2_{H^{s_i}(K_i)^2} + \frac{h_i^{2\mu_i^*+1}}{p_i^{2s_i}}\|u\|^2_{H^{s_i+1}(K_i)}\right),$$

where $\mu_i^+ = min\{s_i,\ p_i + 1\}$ and $\mu_i^* = min\{s_i,\ p_i\}$.

REMARK 5.2.2 *Note that the error estimates obtained in the above theorem are optimal in h and suboptimal in p. These estimates are exactly same as in the case of linear elliptic problems, see [57].*

5.3 Numerical Experiments

This section is devoted to some numerical experiments to illustrate the theoretical results obtained in Theorem 5.2.6. The model problem is a mean curvature flow as in (4.87)-(4.88) which is governed by

$$\begin{aligned} -\nabla\cdot\left(\frac{\nabla u}{(1+|\nabla u|^2)^{1/2}}\right) &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

where $\Omega = (0,1)\times(0,1)$. The forcing function f is taken in such a way that the exact solution is $u = x(1-x)y(1-y)$.

We compute the approximate solutions u_h and $\mathbf{q}_h$ on a sequence of finer subdivisions $\mathcal{T}_h$ of Ω, where $\mathcal{T}_h$ is formed by uniform triangles. Let N_h denote the number of triangles. The discrete space V_h is constructed by using piecewise polynomials of uniform degree $p = 1$ and uniform degree $p = 2$. Since the discrete space V_h can have piecewise polynomials which may be discontinuous across the edges of elements, we choose basis functions as follows. If $p = 1$, then we choose the basis as in (2.60)-(2.61) and if $p = 2$, we choose the basis as in (2.62)-(2.66). We note that each basis function takes support only on the corresponding element K_i. The space $\mathbf{W}_h$ is constructed by taking the tensor product of elements of V_h, *i.e.*, any element $\mathbf{w}_h \in \mathbf{W}_h$ is obtained by taking $\mathbf{w}_h = (v_h, \phi_h)$, for some $v_h, \phi_h \in V_h$. Let $N = N_h * \frac{(p+1)(p+2)}{2}$ and $2N$ be the dimensions of V_h and $\mathbf{W}_h$, respectively. Denote the basis of V_h by $\{\Phi_j : 1 \leq j \leq N\}$ and the basis of $\mathbf{W}_h$ by $\{\Psi_l : 1 \leq l \leq 2N\}$. We choose the parameters $\mathbf{C}_{12} = [0, 0]$, $C_{11} = 10$ and $C_{22} = 1$.

In order to derive the nonlinear system corresponding to (5.18)-(5.20), we set $\mathbf{w}_h = \Psi_l$, $\boldsymbol{\tau}_h = \Psi_k$ and $v_h = \Phi_j$ in (5.18), (5.19) and (5.20), respectively. Then, we arrive at

$$\int_\Omega \mathbf{q}_h \cdot \Psi_l dx + \sum_{i=1}^{N_h} \int_{K_i} u_h \nabla \cdot \Psi_l dx - \int_{\Gamma_I} (\{u_h\} - C_{22}[\boldsymbol{\sigma}_h])[\Psi_l] ds = 0, \ 1 \leq l \leq 2N, \quad (5.111)$$

$$\int_\Omega \frac{\mathbf{q}_h}{(1 + |\mathbf{q}_h|^2)^{1/2}} \cdot \Psi_k dx - \int_\Omega \boldsymbol{\sigma}_h \cdot \Psi_k dx = 0, \ 1 \leq k \leq 2N, \quad (5.112)$$

$$\sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\sigma}_h \cdot \nabla \Phi_j dx - \int_\Gamma (\{\boldsymbol{\sigma}_h\} - C_{11}[u_h])[\Phi_j] ds = \int_\Omega f \Phi_j ds, \ 1 \leq j \leq N. \quad (5.113)$$

Since the nonlinear term appearing only in one term which is integrated over elements K_i, we apply one-step fixed point iteration method to solve the system (5.111)-(5.113). For given $\mathbf{q}_h^0 \equiv 0$, find $\mathbf{q}_h^k$, $\boldsymbol{\sigma}_h^k$ and u_h^k, for $k \geq 1$, such that

$$\int_\Omega \mathbf{q}_h^k \cdot \Psi_l dx + \sum_{i=1}^{N_h} \int_{K_i} u_h^k \nabla \cdot \Psi_l dx - \int_{\Gamma_I} (\{u_h^k\} - C_{22}[\boldsymbol{\sigma}_h^k])[\Psi_l] ds = 0, \ 1 \leq l \leq 2N, \quad (5.114)$$

$$\int_\Omega \frac{\mathbf{q}_h^k}{(1 + |\mathbf{q}_h^{k-1}|^2)^{1/2}} \cdot \Psi_k dx - \int_\Omega \boldsymbol{\sigma}_h^k \cdot \Psi_k dx = 0, \ 1 \leq k \leq 2N, \quad (5.115)$$

$$\sum_{i=1}^{N_h} \int_{K_i} \boldsymbol{\sigma}_h^k \cdot \nabla \Phi_j dx - \int_\Gamma (\{\boldsymbol{\sigma}_h^k\} - C_{11}[u_h^k])[\Phi_j] ds = \int_\Omega f \Phi_j dx, \ 1 \leq j \leq N. \quad (5.116)$$

We now define the following matrices

$$A = [a_{ml}]_{1 \leq m,l \leq 2N}, \ B = [b_{li}]_{1 \leq l \leq 2N, \ 1 \leq i \leq N}, \ D = [d_{ij}]_{1 \leq i,j \leq N}, \ S = [s_{ij}]_{1 \leq i,j \leq 2N} \quad (5.117)$$

and the vector

$$L = [l_i]_{1 \le i \le N_v, 1}\,,$$

where

$$a_{ml} = \int_\Omega \Psi_m \cdot \Psi_l dx, \quad b_{li} = \sum_{i=1}^{N_h} \int_{K_i} \Phi_i \nabla \cdot \Psi_l dx - \sum_{e_k \in \Gamma_I} \int_{e_k} (\{\Phi_i\}) [\![\Psi_l]\!] ds,$$
$$d_{ij} = \sum_{e_k \in \Gamma} \int_{e_k} C_{11} [\![\Phi_i]\!] [\![\Phi_j]\!] ds, \quad s_{ij} = \sum_{e_k \in \Gamma} \int_{e_k} C_{22} [\![\Psi_i]\!] [\![\Psi_j]\!] ds, \quad \text{and} \quad l_i = \int_\Omega f \Phi_i dx.$$

Write

$$u_h = \sum_{i=1}^{N} \alpha_i \Phi_i, \quad \mathbf{q}_h = \sum_{l=1}^{2N} b_l \Psi_l \ \text{ and } \ \boldsymbol{\sigma}_h = \sum_{l=1}^{2N} \gamma_l \Psi_l, \tag{5.118}$$

where $\alpha = [\alpha_1,\ \alpha_2, \cdots, \alpha_{2N}]$, $\mathbf{b} = [b_1,\ b_2, \cdots, b_{2N}]$ and $\gamma = [\gamma_1,\ \gamma_2, \cdots, \gamma_{2N}]$. We define the following matrix which is generated from the first term on the left-hand side of (5.115).

$$R^m = [r_{ij}^m]_{1 \le i,j \le 2N},$$

where

$$r_{ij}^m = \int_\Omega \frac{\Psi_i \cdot \Psi_j}{1 + |\mathbf{q}_h^m|} dx.$$

Using the bases for V_h and $\mathbf{W}_h$, now the system (5.114)-(5.116) reduce to the form : Find α^k, $\mathbf{b}^k$ and γ^k, for $k \ge 1$, such that

$$A\mathbf{b}^k + B\alpha^k + S\gamma^k = 0, \tag{5.119}$$
$$R^{k-1}\mathbf{b}^k - A\gamma^k = 0, \tag{5.120}$$
$$-B^T\gamma^k + D\alpha^k = L, \tag{5.121}$$

where B^T is the transpose of B. Since the basis functions $\{\Psi_l\}_{l=1}^{2N}$ can be assumed independently in each triangle $K \in T_h$, the symmetric positive definite global matrix $[A]$ has the following block diagonal form

$$[A] = \left[[A_{K_1}], ..., [A_{K_{N_h}}] \right],$$

where only the diagonal entries are shown. The other entries in $[A]$ are null matrices. The element matrices $[A_{K_i}]$ are symmetric and positive definite for $i = 1,\ 2, ..., N_h,\ \ [A]^{-1}$ has the block diagonal form

$$[A]^{-1} = \left[[A_{K_1}]^{-1}, ..., [A_{K_{N_h}}]^{-1} \right].$$

From (5.119), we find that

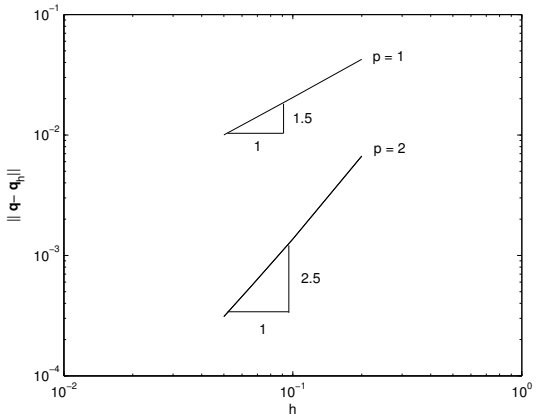
$$\mathbf{b}^k = -A^{-1}(B\alpha^k + S\gamma^k). \tag{5.122}$$

Substituting $\mathbf{b}^k = -A^{-1}(B\alpha^k + S\gamma^k)$ in (5.120)-(5.121), the system (5.119)-(5.121) reduce to the form : For given $\mathbf{b}^0 = 0$, find α^k and γ^k, for $k \geq 1$, such that

$$\begin{aligned} -R^{k-1}A^{-1}(B\alpha^k + S\gamma^k) - A\gamma^k &= 0, \\ -B^T\gamma^k + D\alpha^k &= L. \end{aligned}$$

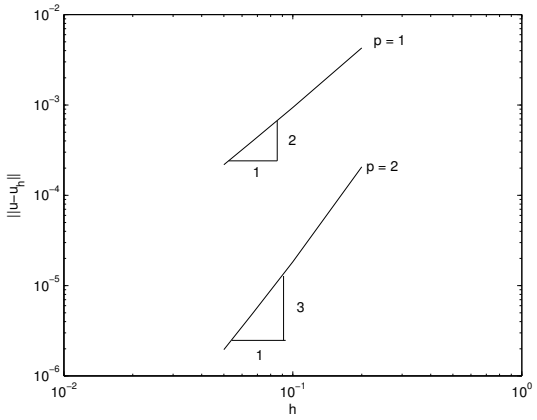
We plot the L^2-norm of the error $\mathbf{e}_q = \mathbf{q} - \mathbf{q}_h$ against mesh size h in Fig 5.1, for each degree of approximation $p = 1$ and $p = 2$. We observe that the convergence lines are straight lines and converge to 0 as h goes to 0. We compute the order of convergence which is 1.5 when $p = 1$, and 2.5 when $p = 2$. Therefore, the computed order of convergence illustrate the theoretical order of convergence obtained in Theorem 5.2.6.

Figure 5.1: Order of convergence for $\|\mathbf{e}_q\|$.



We now plot the L^2-norm of the error $e_u = u - u_h$ against mesh size h in Fig 5.2, for each degree of approximation $p = 1$ and $p = 2$. We also observe that the convergence lines are straight lines and converge to 0 as h goes to 0. We compute the order of convergence which is

Figure 5.2: Order of convergence for $\|e_u\|$.



2 when $p = 1$, and 3 when $p = 2$. Therefore, the computed order of convergence illustrate the theoretical order of convergence obtained in Theorem 5.2.6.

Chapter 6

Conclusions

In this concluding Chapter, we highlight the main results obtained in the present dissertation. Further, we discuss the possible extensions and the scope for further investigations in this direction.

6.1 Summary and Critical Assessment of the Results

In this thesis, we have studied *hp*-discontinuous Galerkin finite element methods (DGFEM) in the primal form, namely; Symmetric Interior Penalty Galerkin (SIPG) and Non-symmetric Interior Penalty Galerkin (NIPG) methods for a class of quasilinear and strongly nonlinear elliptic problems of nonmonotone-type. We also have analyzed the Local Discontinuous Galerkin (LDG) method which is in the mixed form for both quasilinear and strongly nonlinear elliptic problems. The emphasis throughout this study is on the existence and uniqueness of the DG and LDG approximate solutions and the order of convergence in the broken energy norm and L^2-norm. The error estimates have been illustrated with numerical experiments for each of these methods.

We have discussed *hp*-DGFEM (SIPG, NIPG and LDG) to approximate the solution of the following quasilinear elliptic problem :

$$-\nabla \cdot (a(u)\nabla u) = f \text{ in } \Omega, \tag{6.1}$$

$$u = g \text{ on } \partial\Omega, \tag{6.2}$$

where $0 < \alpha \leq a(u) \leq M$, for some $\alpha,\ M \in \mathbb{R}^+$. In this context, we note that it is not possible for the induced variational form B of the DGFEM for (6.1)-(6.2) to satisfy the following strong monotone property

$$B(v_h, v_h - w_h) - B(w_h, v_h - w_h) \geq C|||v_h - w_h|||^2 \quad v_h,\ w_h \in V_h, \tag{6.3}$$

and (or) uniform Lipschitz continuity

$$|B(v_h, \phi_h) - B(w_h, \phi_h)| \leq C|||v_h - w_h||| \ |||\phi_h||| \quad v_h,\ w_h,\ \phi_h \in V_h \tag{6.4}$$

which play a crucial role in deriving the error estimates, see [45, 17]. An attempt has been made in this thesis to discuss DG methods for the nonlinear elliptic problems (6.1)-(6.2) with non-monotone and (or) nonuniform Lipschitz continuous principal part. The basic tool used here is to work on a linearized problem in a ball $O_\delta(I_h u)$ around an interpolant $I_h u$ of u so that the nonlinear operator a can be assumed locally elliptic and (or) locally bounded in the ball $O_\delta(I_h u)$. This has made it possible to allow more general nonlinear coefficient a in (6.1). As a tool for deriving *a priori* error estimates, a discrete solution map is defined via the elements of the discontinuous finite element space V_h to the set of discrete solutions of a linearized problem which is a non-selfadjoint linear elliptic partial differential equation. Since the associated operator satisfies the Gärding type inequality, special care is taken while deriving the error estimates. Specially, a discrete dual problem is introduced to discuss these estimate. It is also shown that the discrete solution map is Lipschitz continuous and maps a ball into itself. An appeal to Brouwer fixed point theorem yields existence of a solution which is also uniquebecause of Lipschitz continuity of the discrete solution map under an *hp*-quasiuniformity condition on the mesh. The derived *a priori* error estimates in the broken H^1 norm are optimal in h and suboptimal in p for both NIPG and SIPG methods. In case of NIPG method, these results lead precisely to the same h-optimal and slightly p-suboptimal rates of convergence as in the case of linear elliptic boundary value problems [61], [44]. It is well noted in the literature that the NIPG method is not adjoint consistent, and hence, it is difficult to expect optimal L^2 error estimate. However on a regular mesh using super-penalty, an optimal error estimate in the L^2-norm is established. The numerical experiments are presented to illustrate the performance of the SIPG and NIPG methods applied to nonlinear elliptic problems (6.1)-(6.2) in Chapter

2. Further, the numerical experiments show the improved order of convergence in the L^2-norm when NIPG method is applied with super penalty. In Chapter 3, the LDG method is analyzed for (6.1)-(6.2) and optimal estimates in h and slightly suboptimal estimates in p are also derived. By and large, the numerical experiments confirm the theoretical results established in Chapter 2 and 3.

We then extend the results obtained for nonlinear elliptic problems (6.1)-(6.2) to a class of strongly nonlinear elliptic problems :

$$-\sum_{i=1}^{2} \frac{\partial a_i}{\partial x_i}(\mathbf{x}, u, \nabla u) + f(\mathbf{x}, u, \nabla u) = 0, \qquad \mathbf{x} \in \Omega, \tag{6.5}$$

$$u = g, \qquad \mathbf{x} \in \partial\Omega, \tag{6.6}$$

where the matrix $[a^{ij}(\mathbf{x}, u, \mathbf{z})] = \left[\frac{\partial a_i}{\partial z_j}(\mathbf{x}, u, \mathbf{z})\right]_{i,j=1,2}$ is symmetric and if $\lambda(\mathbf{x}, u, \mathbf{z})$, $\Lambda(x, u, \mathbf{z})$ are minimum and maximum eigenvalues of the matrix $[a^{ij}]$, then for all $\xi \in \mathbb{R}^2 - \{0\}$ and for all $(\mathbf{x}, u, \mathbf{z}) \in \bar{\Omega} \times \mathbb{R} \times \mathbb{R}^2$

$$0 < \lambda(\mathbf{x}, u, \mathbf{z})|\xi|^2 \leq a^{ij}(\mathbf{x}, u, \mathbf{z})\xi_i\xi_j \leq \Lambda(x, u, \mathbf{z})|\xi|^2. \tag{6.7}$$

We have assumed that if $\|u\|_{W^1_\infty(\Omega)} \leq \alpha$, then there is a positive constant C_α such that

$$0 < C_\alpha \leq \lambda(\mathbf{x}, u, \nabla u). \tag{6.8}$$

We have applied a one parameter family of DG methods to (6.5)-(6.6), which is parametrized by $\theta \in [-1, 1]$, where $\theta = -1$ corresponds to the symmetric and $\theta = 1$ corresponds to the non-symmetric interior penalty methods when $(a_1(u, \nabla u), a_2(u, \nabla u)) = \nabla u$ and $f(u, \nabla u) = 0$. The induced variational form also includes the cases linear elliptic problem [61] and quasilinear elliptic problems [41]. To prove the well-posedness of the DGFEM, we have worked on a corresponding linearized problem which is a non-self adjoint linear elliptic problem. We then define a discrete solution map $S_h : y \in \mathcal{O}_\delta(I_h u) \to V_h$, where $\mathcal{O}_\delta(I_h u)$ is a ball with radius $\delta = \delta(h)$ and centered at an interpolant $I_h u$ of u, and $S_h(y)$ is the solution to the corresponding linearized problem. The map S_h is defined in such a way that any fixed point of S_h is a solution to the nonlinear system of DGFEM. We then find

$$|||S_h(y) - I_h u||| \leq \delta, \text{ for given } y \in \mathcal{O}_\delta(I_h u), \tag{6.9}$$

and

$$|||S_h(y_1) - S_h(y_2)||| \leq Ch^{\epsilon}|||y_1 - y_2|||, \text{ for } y_1,\ y_2 \in \mathcal{O}_{\delta}(I_h u),\ 0 < \epsilon \leq 1/4. \qquad (6.10)$$

In order to establish (6.9)-(6.10), we choose δ in such a way that if $y \in \mathcal{O}_{\delta}(I_h u)$, then $||y||_{W^1_{\infty}(\Omega,\mathcal{T}_h)} \leq \|u\|_{W^1_{\infty}(\Omega)} + Ch^{\epsilon}\|u\|_{H^{5/2}(\Omega)}$. This has made it possible to assume that the nonlinear terms a_i and f along with their derivatives are bounded in a ball around u. Using (6.9)-(6.10) and Brouwer fixed point theorem, we have shown that the discrete problem has a unique solution under hp-quasiuniformity assumption on the mesh when the degree of approximation ≥ 2. The error estimates obtained in broken H^1-norm are optimal in h and suboptimal in p. These results lead precisely to the same h-optimal and mildly p-suboptimal rate of convergence as in the case of linear elliptic problems, see [61]. Finally, we have derived *a priori* estimates in the L^2 norm which are optimal in h and slightly optimal in p when $\theta = -1$. We have discussed some numerical experiments which confirms the theoretical results obtained for (6.5)-(6.6) in Chapter 4. It is also noted from the numerical experiments using piecewise linear elements that is $p = 1$ that there is a deterioration in the computed order of convergence. However, we do not have theoretical justification to this observation.

In Chapter 5, we have discussed the hp-local discontinuous Galerkin method (LDG) for strongly nonlinear elliptic problems (6.5)-(6.6). We use the results of LDG method for quasilinear elliptic problems (6.2)-(6.2) and the results of DG methods for the strongly nonlinear elliptic problems (6.5)-(6.6) to analyze the LDG method for strongly nonlinear elliptic problems (6.5)-(6.6). In this context, it has been possible to derive error estimates for the nonlinear operators (a_1, a_2) which are locally elliptic and (or) locally Lipschitz continuous. Due to the technicalities in using inverse estimates, we add the extra stabilizing terms containing the jumps of the flux variable which allow us to use the piecewise linear polynomials. In the absence of these jump terms, we need the degree of approximation ≥ 2. Using Brouwer fixed point theorem and Lipschitz continuity of the discrete solution map, it is shown that the discrete problem has a unique solution under hp-quasiuniformity assumption on the mesh. The error estimates obtained are optimal in h and suboptimal in p. These results lead precisely to the same h-optimal and slightly p-suboptimal rate of convergence as in the case of linear elliptic problems, see [17].

6.2 Possible Extensions and Future Problems

The results of this thesis can be easily extended to the problems in three space-dimensions by making appropriate changes in the analysis. Moreover, it is not difficult to extend our analysis to the problem $-\nabla \cdot (a(u)\nabla u) + f(u) = 0$, where $f(u) \in C_b^2(\bar{\Omega} \times \mathbb{R})$. With appropriate modifications in the analysis, it is possible to extend the theoretical results of this thesis to the problem (2.1)-(2.2) when $a(u)$ is a positive-definite matrix.

In this thesis, we have derived a priori error estimates in the broken H^1-norm and in the L^2-norm. Establishing error estimates in the L^∞ (maximum) norm for the nonlinear elliptic problems is an interesting and more challenging task. In literature, the maximum norm error estimates are derived for the h-version of the SIPG and LDG methods applied to linear elliptic problems, [21, 22, 23, 42, 46].

As in the unified frame work of all the existing DG methods for linear elliptic problems [5], an attempt can be made to derive a unified frame work for the nonlinear elliptic problems which are studied in the present thesis. As in [5], the concept of lifting operators and L^2-projections may be used to derive the unified frame work. This may make it possible to study a large class of DG methods such as Oden [55], Brezzi [15], Bassi-Rebay [11], etc., and useful to derive a unified *a posteriori* error estimates. In this context, an attempt has been made in [19] to discuss unified a posteriori estimates for the unified DG methods when applied to linear elliptic problems. Given a tolerance ϵ and norm $|||\cdot|||$, the objective of the adaptive DG methods is to compute an approximate solution u_h to the exact solution say u such that

$$|||u - u_h||| \leq \epsilon$$

is achieved efficiently in minimal computational cost. We expect that the unified *a posteriori* estimates proposed in [19] for unified DG methods applied to linear elliptic problems will be useful in deriving *a posteriori* error analysis for the nonlinear elliptic problems.

The analysis of this dissertation may be extended to study the DG and LDG methods for

the following time dependent problems :

$$\begin{aligned} u_t - \nabla \cdot (a(u)\nabla u) &= f \ \text{ on } \ \Omega \times (0,T], \\ u(\mathbf{x},t) &= g_1(t) \ \text{ on } \ \partial\Omega \times [0,T], \\ u(\mathbf{x},0) &= g_2(\mathbf{x}) \ \text{ on } \ \Omega, \end{aligned}$$

and

$$\begin{aligned} u_t - \nabla \cdot \mathbf{a}(u,\nabla u) &= f \ \text{ on } \ \Omega \times (0,T], \\ u(\mathbf{x},t) &= g_1(t) \ \text{ on } \ \partial\Omega \times [0,T], \\ u(\mathbf{x},0) &= g_2(\mathbf{x}) \ \text{ on } \ \Omega, \end{aligned}$$

where the nonlinear operator a or $\mathbf{a}$ may not be uniformly monotone and (or) uniformly Lipschitz continuous. Problems of these type occur in variety of applications, for example, in nonlinear convection-diffusion, flow of gases in porous media, etc. Therefore, it is worthwhile to generalize the results of this thesis to such problems. Finally, it is possible to expand the present analysis to include the coupled systems of equations in either incompressible miscible or slightly compressible miscible problems in Oil reservoir studies. Specially, an application of LDG method may prove effect to the Concentration equation which is a convection dominated diffusion equation.

Bibliography

[1] M. AINSWORTH AND D. KAY, *The approximation theory for the p-version finite element method and application to the nonlinear elliptic PDEs,* Numer. Math., 82 (1999), pp. 351-388.

[2] M. AINSWORTH AND D. KAY, *Approximation theory for the hp-version finite element method and application to the nonlinear Laplacian,* Applied Numerical Mathematics, 34 (2000), pp. 329-344.

[3] M. ALI, M. ALI, *Numerical methods for enhanced oil recovery in reservoir simulation,* Ph. D. Thesis, IIT Bombay (1997).

[4] D. N. ARNOLD, *An interior penalty finite element method with discontinuous elements,* SIAM J. Numer. Anal., 19 (1982), pp. 742-760.

[5] D. N. ARNOLD, F. BREZZI, B. COCKBURN AND L. D. MARINI, *Unified analysis of discontinuous Galerkin methods for elliptic problems,* SIAM J. Numer. Anal., 39 (2002), pp. 1749-1779.

[6] I. BABUŠKA, *The finite element method with penalty,* Math. Comput., 27 (1973), pp. 221-228.

[7] I. BABUŠKA AND M. ZLÁMAL, *Nonconforming elements in the finite element method with penalty,* SIAM J. Numer. Anal., 10 (1973), pp. 863-875.

[8] I. BABUŠKA AND M. SURI, *The optimal convergence rate of the p version of the finite element method,* SIAM J. Numer. Anal., 24 (1987), pp. 750-776.

[9] I. BABUŠKA AND M. SURI, *The h-p version of the finite element method with quasiuniform meshes,* RAIRO Model. Math. Anal. Nume., 21 (1987), pp. 199-238.

[10] G. A. BAKER, *Finite element methods for elliptic equations using nonconforming elements,* Math. Comput., 31 (1977), pp. 45-59.

[11] F. BASSI AND S. REBAY, *A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations,* J. Comput. Phys., 131 (1997), pp. 267-279.

[12] C. BERNARDI, M. DAUGE AND Y. MADAY, *Polynomials in the Sobolev world,* Preprint of the Laboratoire Jacques-Louis Lions, No. R03038, (2003).

[13] S. C. BRENNER, *Poincare-Friedrichs inequalities for piecewise H^1 functions,* SIAM J. Numer. Anal., 41 (2003), pp. 306-324.

[14] S. C. BRENNER AND L. Y. SUNG, *C^0 interiori penalty methods for fourth order elliptic boundary value problems on polygonal domains,* J. Sci. Comput., 22-23 (2005), pp. 83-118.

[15] F. BREZZI, M. MANZINI, L. D. MARINI AND P. PIETRA, *Discontinuous Galerkin approximations for elliptic problems,* Numer. Meth PDE., 16 (2000), pp. 265-278.

[16] F. BREZZI, L. D. MARINI AND E. Süli, *Discontinuous Galerkin methods for hyperbolic problems,* M3AS., 14 (2004), pp. 1893-1903.

[17] R. BUSTINZA AND G. GATICA, *A local discontinuous Galerkin method for nonlinear diffusion problems with mixed boundary conditions,* SIAM J. Sci. Comput., 26 (2004), pp. 152-177.

[18] P. CASTILLO, B. COCKBURN, I. PERUGIA AND D. SCHÖTZAU, *An a priori error analysis of the local discontinuous Galerkin method for elliptic problems,* SIAM J. Numer. Anal., 38 (2000), pp. 1676-1706.

[19] C. CARSTENSEN, T. GUDI AND M. JENSEN, *A unifying theory of a posteriori error control for discontinuous Galerkin methods,* Preprint, Department of Mathematics, Humboldt University, Berlin, 2006.

[20] P. G. CIARLET, *The Finite Element Method for Elliptic Problems,* NORTH-HOLLAND PUBLISHING COMPANY, (1978).

[21] H. CHEN, *Pointwise estimates for discontinuous Galerkin methods with penalty for second order elliptic problems,* SIAM. J. Numer. Anal., 42 (2004), pp. 1146-1166.

[22] H. CHEN, *Local error estimates of mixed discontinuous Galerkin methods for elliptic problems,* J. Numer. Math., 12 (2004), pp. 1-22.

[23] H. CHEN, *Pointwise error estimates of the local discontinuous Galerkin method for a second order elliptic problems,* Math. Comput., 74 (2005), pp. 1097-1116.

[24] B. COCKBURN AND C. W. SHU, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws II, General frame work,* Math. Comput., 52 (1989), pp. 411-435.

[25] B. COCKBURN, S. Y. LIN AND C. W. SHU, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws III, One dimensional system,* J. Comput. Phys., 84 (1989), pp. 90-113.

[26] B. COCKBURN, S. HOU AND C. W. SHU, *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for scalar conservation laws IV, The multidimensional case,* Math. Comput., 54 (1990), pp. 545-581.

[27] B. COCKBURN AND C. W. SHU, *The Runge - Kutta discontinuous Galerkin finite element for conservation laws V: Multidimensional systems,* J. Comput. Phys., 141 (1998), pp. 199-224.

[28] B. COCKBURN AND C. W. SHU, *The local discontinuous Galerkin time-dependent method for convection-diffusion systems,* SIAM J. Numer. Anal., 35 (1998), pp. 2440-2463.

[29] B. COCKBURN, G. E. KARNIADAKIS AND C. W. SHU, *The development of discontinuous Galerkin methods,* in Discontinuous Galerkin Methods. Theory, Computation and Applications, B. Cockbur, G. E. Karniadakis and C. W. Shu, eds, Lecture Notes in Comput. Sci. Engrg., Springer-Verlag, 11 (2000), pp. 3-50.

[30] B. COCKBURN, G. KANCHAT, D. SCHÖTZAU AND C. SCHWAB, *Local discontinuous Galerkin methods for the Stokes systems,* SIAM J. Numer. Anal., 40 (2002), pp. 319-343.

[31] B. COCKBURN, G. KANSCHAT, D. SCHÖTZAU, AND C. SCHWAB, *Discontinuous Galerkin methods for incompressible elastic materials,* Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 3184-3204.

[32] G. R. COWPER, *Gaussian quadrature formulas for triangles,* Inter. J. Numer. Meth. Engrg., 7 (1973), pp. 405-408.

[33] C. DAWSON, *Godunov-mixed methods for advection flow problems equation in one space dimension,* SIAM J. Numer. Anal., 30 (1993), pp. 1315-1332.

[34] C. DAWSON, *Analysis of an upwind-mixed finite element method for contaminant transport equations,* SIAM J. Numer. Anal., 35 (1998), pp. 1709-1724.

[35] P. DHANUMJAYA, *Finite Element Methods for Extended Fisher-Kolmogorov (EFK) equation,* PH. D. Dissertation, IIT Bombay, (2003).

[36] J. DOUGLAS, Jr., T. DUPONT AND J. SERRIN, *Uniqueness and comparison theorems for nonlinear elliptic equations in divergence form,* Arch. Rational Mech. Anal., 42 (1971), pp. 157-168.

[37] J. DOUGLAS, Jr. AND T. DUPONT, *A Galerkin method for a nonlinear Dirichlet problem,* Math. Comput., 29 (1975), pp. 689-696.

[38] J. DOUGLAS, Jr. AND T. DUPONT, *Interior penalty procedures for elliptic and parabolic Galerkin methods,* in Computing Methods in Applied Sciences, Lecture Notes in Phys. 58, Springer-Verlag, Berlin, (1976), pp. 207-216.

[39] E. H. GEORGOULIS AND E. SÜLI, *Optimal error estimates for the hp-version interior penalty discontinuous Galerkin finite element method,* IMA J. Numer. Anal., 25 (2005), pp. 205-220.

[40] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order,* Springer-Verlag, (1983).

[41] T. GUDI AND A. K. PANI, *Discontinuous Galerkin methods for quasilinear elliptic problems of nonmonotone-type,* accepted in SIAM J. Numer. Anal., (2006).

[42] J. GUZMAN, *Pointwise error estimates for discontinuous Galerkin methods with lifting operators for elliptic problems,* Math. Comput., 75 (2006), pp. 1067-1085.

[43] P. HANSBO AND M. G. LARSON, *Discontinuous Galerkin methods for incompressible and nearly incompressible elassticity by Nitsche's method,* Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 1895-1908.

[44] P. HOUSTON, C. SCHWAB, AND E. SÜLI, *Discontinuous hp-finite element methods for advection-diffusion-reaction problems,* SIAM J. Numer. Anal., 39 (2002), pp. 2133-2163.

[45] P. HOUSTON, J. A. ROBSON AND E. SÜLI, *Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems I: the scalar case,* IMA J. Numer. Anal., 25 (2005), pp. 726-749.

[46] G. KANSCHAT AND R. RANNACHER, *Local error analysis of the interior penalty discontinuous Galerkin methods for second order elliptic problems,* J. Numer. Math., 4 (2002), pp. 249-274.

[47] O. A. KARAKASHIAN AND W. N. JUREIDINI, *A nonconforming finite element method for the stationary Navier-Stokes equations,* SIAM J. Numer. Anal., 35 (1998), pp. 93-120.

[48] S. KESAVAN, *Topics in Functional Analysis and Applications,*Wiley-Eastern Ltd., (1989).

[49] A. LASIS AND E. SÜLI, *Poincare-Type inequalities for Broken Sobolev spaces,*Isaac Newton Institute for Mathematical Sciences, Preprint No. NI03067-CPD, (2003).

[50] A. LASIS AND E. SÜLI, *One-parameter discontinuous Galerkin finite element discretisation of quasilinear parabolic problems,* Oxford Univ. Comp. Lab., Research Report NA-04/25 (2004).

[51] P. LESIANT AND P. A. RAVIART, *On a finite element method for solving the neutron transport equation,* Mathematical aspects of finite elements in partial differential equationsi: proccedings of symposium, Madison, 1974, pp. 89-123.

[52] F. A. MILNER AND M. SURI, *Mixed finite element methods for quasilinear elliptic problems: The p-version* M^2AN, 26 (1992), pp. 913-931.

[53] I. MOZOLEVSKI, E. SÜLI AND P. BÖSING, *hp-version a priori error analysis of interior penalty discontinuous Galerkin finite element approximations to the biharmonic equation,* accepted in J. Sci. Comput., (2006).

[54] J. A. NITSCHE, *Über ein Variationprinzip zur Lösung Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unteworfen sind,* Abh. Math. Sem. Univ. Hamburg, 36 (1971), pp. 9-15.

[55] J. T. ODEN, I. BABUSKA AND C. E. BAUMANN, *A discontinuous hp finite element method for diffusion problems,* J. Comput. Phys., 146 (1998), pp. 491-519.

[56] A. K. PANI AND P. C. DAS, *C^0-interior-penalty Galerkin method for slightly compressible miscible displacement in porous media,* Proceedings of International Conference on Nonlinear Mechanics, Sanghai (China), Chein Wei Zang (ed.), Science Press, Sanghai (1985), pp. 1186-1191.

[57] I. PERUGIA AND D. SCHOETZAU, *An hp-analysis of the local discontinuous Galerkin method for diffusion problems,* J. Sci. Comput., 6 (2001), pp. 411-433.

[58] S. PRODHOMME, F. PASCAL AND J. T. ODEN, *Review of error estimation for discontinuous Galerkin methods,* TICAM REPORT 00-27, October 17, (2000).

[59] W. H. REED AND T. R. HILL, *Triangular mesh methods for the neutron transport equation,* Technical Report LA-UR-73-479, Las Alamos Scientific Laboratory, Las Alamos, NM, 1973.

[60] B. RIVIÉRE AND M. F. WHEELER, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems I,* Comput. Geosci., 3 (1999), pp. 337-360.

[61] B. RIVIÉRE, M. F. WHEELER AND V. GIRAULT, *A priori error estimates for finite element methods based on discontinuous approximation spaces for elliptic problems,* SIAM J. Numer. Anal., 39 (2001), pp. 902-931.

[62] D. SCHÖTZAU, C. SCHWAB, AND A. TOSELLI, *Mixed hp-DGFEM for incompressible flows,* SIAM J. Numer. Anal., 40 (2003), pp. 2171-2194.

[63] E. SÜLI, C. SCHWAB AND P. HOUSTON *hp-DGFEM for second order partial differential equations with nonnegative characteristic from*, In : B. Cockburn, G. Karniadakis, and C.-W. Shu (Eds.) Discontinuous Galerkin Finite Element Methods, Lecture Notes in Computational Science and Engineering, Springer-Verlag, 11 (1999), pp 221-230.

[64] E. SÜLI, P. HOUSTON AND B. SENIOR, *hp- discontinuous Galerkin finite element methods for hyperbolic problems: error analysis and adaptivity,* In M J Baines, editor, Numerical Methods for Fluid Dynamics VII. ICFD, Oxford, (2001) pp. 347-353.

[65] A. TOSELLI, *hp-discontinuous Galerkin approximations for Stokes problems,* Math. Models Methods Appl. Sci., 12 (2002), pp. 1565-1616.

[66] M. F. WHEELER, *An elliptic collocation-finite element method with interior penalties,* SIAM J. Numer. Anal., 15 (1978), pp. 152-161.

[67] M. F. WHEELER AND B. L. DARLOW, *Interior penalty Galerkin procedures for miscible displacements in porous media,* Computational Methods in Nonlinear Mechanics, (1980), pp. 485-506.

Acknowledgments

It is my great pleasure to take this opportunity to acknowledge the people who have supported me during the period of research.

First and foremost, I would like to express my heartiest gratitude to my supervisor Prof. Amiya Kumar Pani for his enthusiastic supervision and kind support. I cannot imagine to have completed this research without his intuitive suggestions and enlightened lectures. I owe lots of gratitude to him for showing me the way of research. I am also very thankful to his wife Mrs. Tapaswani Pani for her support.

Secondly, I am deeply grateful to my co-supervisor Prof. Neela Nataraj for giving me invaluable suggestions throughout this research. Her detailed and constructive lectures on 'MATLAB implementations for finite element methods', have helped me a lot in understanding the basics of numerical implementations. I thank her for helping in relating the numerical results to the theoretical findings.

I am greatly thankful to my RPC (Research Progress Committee) members Prof. Rekha P. Kulkarni and Prof. K. Suresh Kumar for their valuable suggestions and comments.

I am very thankful to the Department of Mathematics, IIT Bomabay for providing me the working place and computing facility during my research work.

I acknowledge the DST-DAAD for providing me with financial support to visit the Department of Mathematics, Humboldt University, Berlin during December 2005 and July 2006. I am greatly thankful to Prof. Amiya Kumar Pani for giving me a wonderful opportunity to work in the DST-DAAD (PPP-05) Project based Personal Exchange Programme.

I would like to express my heartfelt thanks to Prof. Carsten Carstensen, who invited me to visit the Department of Mathematics, Humboldt University, Berlin during December 2005 and July 2006 under the DST-DAAD (PPP-05) project. The discussions with him on a posteriori error control for FEM helped me to increase my knowledge. I also had fruitful discussions with Dr. Sören Bartles and Dr. Max Jensen on adaptive finite elements and

DGFEM. Thanks to them for their help during my stay in Berlin.

Thanks are also due to Prof. T. Ram Reddy, Department of Mathematics, Kakatiya University, Warangal without whose help I would not have come to Mumbai.

Many thanks to my parents Shri Laxma Reddy and Smt. Laxmi, who taught me the value of hard work by their own example. I am also thankful to my brothers Mohan and Ravi for their encouragement and moral support to continue my studies. My special thanks to my father-in-law Shri Sammi Reddy, my mother-in-law Smt. Varadevi and my brother-in-law Paapi Reddy who have taken care of my family in my absence at Warangal.

I am very grateful to my wife Vijaya who has provided me an extra hand with her patience, love and support throughout the research. Without her encouragement and understanding it would not have been possible to work till late hours during the nights. My little son Srivarshith has provided me with a lot of joy and love during the past three years, ever since he was born.

Thanks to my dear friends and colleagues for their help and also for creating a homely environment at IIT Bombay.

Finally, I would like to thank all whose direct and indirect support helped me in completing my thesis in time.

IIT Bombay **Thirupathi Gudi**